TESE DE DOUTORAMENTO

A CORPUS-BASED STUDY ON NEAR-SYNONYMY: THE CONCEPT PLEASANT SMELLING IN 19TH- AND 20TH-CENTURY AMERICAN ENGLISH

Daniela Beatriz Pettersson-Traba

ESCOLA DE DOUTORAMENTO INTERNACIONAL

PROGRAMA DE DOUTORAMENTO EN ESTUDOS INGLESES AVANZADOS: LINGÜÍSTICA, LITERATURA E CULTURA

SANTIAGO DE COMPOSTELA

2021

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DECLARACIÓN DO AUTOR/A DA TESE

D./Dna. **Daniela Beatriz Pettersson-Traba**

Título da tese: **A corpus-based study on near-synonymy: The concept PLEASANT SMELLING in 19th- and 20th-century American English**

Presento a miña tese, seguindo o procedemento axeitado ao Regulamento, e declaro que:

1) A tese abarca os resultados da elaboración do meu traballo.
2) De ser o caso, na tese faise referencia ás colaboracións que tivo este traballo.
3) Confirmo que a tese non incorre en ningún tipo de plaxio doutros autores nin de traballos presentados por min para a obtención doutros títulos.

E comprométome a presentar o Compromiso Documental de Supervisión no caso de que o orixinal non estea na Escola.

En **Santiago de Compostela**, **7 de xaneiro de 2021**.

Sinatura electrónica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AUTORIZACIÓN DO DIRECTOR/TITOR DA TESE

D./Dna. **María José López Couso**

En condición de: **Titor/a e director/a**

Título da tese: **A corpus-based study on near-synonymy: The concept PLEASANT SMELLING in 19th- and 20th-century American English**

INFORMA:

Que a presente tese, correspóndese co traballo realizado por D/Dna Daniela Beatriz Pettersson-Traba, baixo a miña dirección/titorización, e autorizo a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como director/titor desta non incorre nas causas de abstención establecidas na Lei 40/2015.

En **Santiago de Compostela**, **7 de xaneiro de 2021.**

Sinatura electrónica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ABSTRACT

The last couple of decades have witnessed a renewed interest in the semantic phenomenon of near-synonymy (e.g. Gries 2001; 2003; Taylor 2003; Divjak 2010; Liu 2010; 2013). In particular, recent distributional corpus-based approaches and techniques used for semantic analysis, such as Behavioral Profiles (e.g. Divjak & Gries 2006; 2008) and correspondence analysis (e.g. Desagulier 2014; Krawczak 2018), have successfully uncovered subtle distinctions in meaning between near-synonyms by analyzing, among other factors, their collocational and stylistic preferences. Some semantic domains have received particular attention, for instance, those of SIZE and AMOUNT (cf. Biber, Conrad & Reppen 1998: Section 2.6 on *big*, *large*, and *great*; Taylor 2003 on *high* and *tall*; and Gries & Otani 2010 on *big*, *large*, and *great* and *little*, *small*, and *tiny*), while other near-synonym sets have been relatively underresearched. Moreover, most studies so far have dealt with the semantic structure of sets of near-synonyms from a synchronic perspective (e.g. Divjak & Gries 2006; 2008; Liu 2010; 2013), whereas their diachronic evolution has generally been neglected, with only a handful of investigations adopting a historical approach (e.g. Kaunisto 2001; Primahadi-Wijaya-R & Rajeg 2014).

Against this backdrop, the aim of the present dissertation is to examine five adjectival near-synonyms in the history of American English from the understudied semantic domain of SMELL, namely *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*, which designate the concept PLEASANT SMELLING. To this end, instances of these five adjectives are retrieved from a large historical corpus of this variety of English, to wit, the *Corpus of Historical American English* (Davies 2010–), which covers the timespan 1810–2009. The distribution of the adjectives is analyzed over time across a wide range of contexts and functions, including semantic, morphosyntactic, stylistic, and collocational variables, since distributional patterns of this type have been shown to serve as a proxy for semantic (dis)similarity (e.g. Divjak & Gries 2006; 2008; Gries & Otani 2010). The data is submitted to various univariate and multivariate statistical techniques in order to uncover fine-grained

(dis)similarities among the members of this near-synonym set, as well as possible changes in their prototypical structures, from both an onomasiological and a semasiological perspective.

Three separate analyses are conducted, each focusing on a different aspect of the internal semantic structure of the near-synonym set. The first one (Chapter 5) examines the distribution of the five adjectives across senses and semantic categories of the nouns that they modify to discover their prototypical uses in both semasiological and onomasiological terms. The second analysis (Chapter 6) delves deeper into the competition between the three most common of these near-synonyms (*fragrant*, *perfumed*, and *scented*) by expanding the range of contexts considered to cover also non-semantic and stylistic factors (e.g. syntactic function of the adjectives, degree, text-type). Finally, Chapter 7 accounts for the idiosyncratic collocational behavior of these three adjectives, hence zooming in on their specific collocational preferences by using techniques which are specifically geared towards this issue.

The results demonstrate that the near-synonym set under analysis is undergoing a process of semantic convergence, whereby the adjectives are progressively used in more similar semantic contexts, becoming more and more frequent over time to designate artificial smells as opposed to natural ones. This change is here claimed to be motivated by extralinguistic factors, to wit, the social and technological transformations experienced by American society after the First and Second Industrial Revolutions, which have led to an ever-increasing need to refer to artificially scented products rather than to naturally fragrant plants and flowers. Moreover, this process of convergence is accompanied by one of substitution, in which the initially most dominant adjective of the set (*fragrant*) is gradually being replaced by another one (*scented*). Therefore, the findings obtained shed valuable light on the diachronic development of lexical near-synonyms, a dimension that has up to now been relatively disregarded in the specialized literature and that is yet to receive the attention it certainly deserves.

Keywords: (near-)synonymy, historical linguistics, semantic change, distributional corpus-based approach, collocation, PLEASANT SMELLING, American English

ACKNOWLEDGEMENTS

I would like to express my most sincere and profound gratitude to all the people who, in one way or another, have played a part in the successful completion of this PhD dissertation. Without their wholehearted support and commitment during these years, this would undoubtedly not have been possible.

First of all, I am greatly indebted to my supervisor Dr. María José López Couso. From the time when I started working under her supervision six years ago to write my BA thesis on this same topic, she has always encouraged me to follow my own research interests and to continue on the *dark path* of semantics. Her invaluable advice and guidance ever since are the main reasons why I decided to continue in the academic world, hence aiding me in discovering my vocation. Her help has transcended academic matters, as she has also been a source of moral support on a more personal level, always ready to lend a helping hand with any problem that might arise. I feel immensely privileged to have learnt —and still be learning— from such an exceptional professional.

I would also like to acknowledge the useful feedback received from Dr. Belén Méndez Naya, Dr. Paloma Núñez Pertejo, and Dr. Susana Doval Suárez, who served as examiners of a pilot study of this work, my MA thesis. Thanks to their insightful suggestions, I have been able to continue polishing this piece of work. Additionally, I would like to show my genuine gratitude to Dr. Kathryn Allan, Dr. Dirk Speelman, and Dr. Augusto Soares da Silva for their expertise and wise counsel received during my research stays at University College London, KU Leuven, and the Catholic University of Portugal. The time they all generously devoted to discussing my research with me has certainly enhanced its quality. Special thanks are due to Dr. Dirk Speelman for everything he has taught me about statistics, both concerning the use of the program *R* and concerning the multiple techniques and methods for semantic analysis. His continued patience for bearing with me, a statistics rookie, every week for three months in Leuven is gratefully acknowledged. I would also like to extend my thanks to all those scholars I have had the pleasure to meet and learn from at conferences, seminars, and research meetings

or through peer-reviews. In this respect, I want to mention with gratitude the particularly useful comments and recommendations of the following people: Dirk Geeraerts, Belén Méndez Naya, Gerold Schneider, Mario Serrano Losada, and Stefanie Wulff.

I am also indebted to the research team *Variation, Linguistic Change and Grammaticalization* (*VLCG*) and to the network *English Linguistics Circle* (*ELC*) and all their members for their support during these last five years. I have benefited greatly from the conferences and seminars they organized, in which I have not only been able to attend many fascinating talks by leading researchers in the field, but where I have also had the opportunity to present my own work and discuss it with different scholars. I also thank the Department of English and German Philology of the University of Santiago de Compostela and all its members, especially for offering me the possibility to teach at the university. Throughout these years I have had the privilege to learn from many of my colleagues when teaching several BA courses; special thanks are in this regard due to Cecilia, Paloma, María José, Elsa, Naara, Monika, Sandra, Eithne, Carmen, Fátima, and Idoya. Since a few months back, I have been working at the Department of English at the University of Extremadura. I must acknowledge that changing jobs and moving across the country in the midst of a pandemic is not easy. However, a few people have contributed greatly to making this change a bit less challenging. In this regard, I want to give a special thanks to Cristina Blanco García for all support and for being a shoulder to cry on during the difficult moments.

For generous financial support, I am grateful to the European Regional Development Fund and the following institutions: The Regional Government of Galicia (grants ED481A–2016/168, ED431B 2017/12, and ED431D 2017/09) and the Spanish Ministry of Innovation, Science, and Universities (grants FFI2014–52188-P and FFI2017–86884-P).

En un plano más personal, quiero agradecer a todas las personas que he conocido durante los más de diez años que he pasado en la Facultad de Filología de la USC, desde el grado hasta el fin de la etapa doctoral. Muchas de estas personas se han convertido con el tiempo en verdaderas amigas y amigos. Al grupo de chicas del grado, por todas las risas y buenos ratos, tanto durante las clases como fuera de ellas: Regina, Tatiana, Marta, Silvia, Celina, Natalia, Leila e Irene. A las compañeras y compañeros del doctorado, tanto del *VLCG* y/o el *ELC*, y en general del Departamento de Filología Inglesa y Alemana, por todos los sabios consejos y apoyo, y como no, por esta experiencia compartida que se ha hecho indudablemente más llevadera gracias a tod@s vosotr@s: Iván, Mario, Cristina, Iría, Tania, Tamara, Idoya, Paula,

Elsa, Carla, Sofía, Yolanda, Noelia, Nerea, Regina y Cecilia. Al resto de investigadores en formación de la facultad y compañeros en general que he tenido el placer de conocer durante estos últimos dos años, por esa pequeña comunidad que hemos construido durante las comidas en la sala de microondas y los múltiples cafés tomados en la cafetería de la facultad: Fernando, Antía, Lúa, Lucio, Martha, Pablo, Iria, Miki, Sara, Miriam, Noa, Estrella y Bea. Y, finalmente, a mi pequeña familia dentro de la facultad, por todos los ánimos recibidos, especialmente en momentos de crisis. Sin vosotros esto no hubiese sido posible: Iván, Regina, Nerea R., Nerea L., María y Fernando.

Fuera del ámbito académico, hay un gran número de personas a las que les debo un mundo, por motivarme constantemente y hacerme ver que el esfuerzo y la determinación al final siempre valen la pena. En primer lugar, quiero agradecer a todas mis amistades, tanto aquí en Galicia como en Suecia, por recorrer todo o parte del camino conmigo. Ett stort och hjärtligt tack till alla mina vänner i Sverige, speciellt till Josefine och Carro och såklart till tjejgruppen i SP3C: Neda, Rebecka H., Rebecka B., Elin och Tanja. En segundo lugar, quiero agradecer a mi familia, a los que están y a los que ya no están, por permanecer siempre a mi lado, no solamente durante estos cinco años, sino durante toda la vida. En especial, quiero recordar todo el cariño recibido por parte de mi madre, gracias a la cual soy hoy la persona que soy. Eres y has sido siempre mi fuente de inspiración por ser la persona más fuerte y determinada que conozco, recordándome siempre a seguir hacia delante pese a los obstáculos que pueda encontrar en mi camino. También quiero dar las gracias a mi pareja Iván, por la infinita paciencia que muestras día tras día, así como por tu constante amor y apoyo, que nunca flaquean. Tu energía y positividad ha llenado de luz esta etapa y por ello te estaré siempre agradecida. Finalmente, a Poe y a Kompis, por todos los abrazos y besos caninos. No hay día, por muy oscuro que sea, en el que la alegría de estos pequeños amigos peludos no me saque una sonrisa.

THANK YOU – GRACIAS – GRAZAS – TACK

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

TABLE OF CONTENTS

List of figures xi
List of tables xiii
Abbreviations and conventions xvii
1 Introduction 1
1.1 Aims and scope of the study 3
1.2 Structure 5
2 Approaches to and notions in lexical semantics 9
2.1 Historical-philological semantics 11
2.2 Structuralist and generative semantics 18
2.3 Neostructuralist semantics 26
2.3.1 The compositional orientation 27
2.3.2 The relational orientation 29
2.4 Cognitive semantics 39
2.4.1 Prototype theory: Prototypicality, gradience, entrenchment, and salience 42
2.4.2 Usage-based theories of semantic change 45
2.5 Summary 49
3 Synonymy 51
3.1 Classifications of synonymy 52
3.2 Usage-based corpus studies on specific synonyms 61
3.2.1 Two waves of research 62
3.2.2 Near-synonyms and POS: Studies on adjectives 72
3.3 Synonymy and diachrony 77
3.4 Summary 87
4 The concept PLEASANT SMELLING: Data selection and annotation 89
4.1 Introducing the synonym set 90
4.1.1 Main motivations for the selection of the concept PLEASANT SMELLING 91

4.1.2 Revising reference works 92
4.2 Corpus description and data retrieval 103
4.3 Data annotation 108
4.3.1 Language-internal semantic variables 112
4.3.1.1 Sense 112
4.3.1.2 Semantic category 115
4.3.1.3 Other semantic variables: Animacy, Concreteness, and Countability 132
4.3.2 Language-internal non-semantic variables 137
4.3.3 Language-external variables 143
4.4 Summary 144
5 Semasiological and onomasiological analyses of the synonym set 147
5.1 Overall frequency and frequency over time 150
5.2 The concept PLEASANT SMELLING 155
5.2.1 Sense 155
5.2.2 Semantic category 159
5.2.3 Ongoing semantic change in a modernizing society 165
5.3 Semasiological analysis 170
5.3.1 Sense 171
5.3.2 Semantic category 174
5.3.3 Interim summary and discussion 179
5.4 Onomasiological analysis 182
5.4.1 Sense 183
5.4.2 Semantic category 185
5.4.3 Interim summary and discussion 189
5.5 Chapter summary and discussion 190
6 In-depth onomasiological analysis of the synonym set: A multivariate approach 195
6.1 Dataset and methodology 196
6.2 Results 202
6.2.1 Multinomial models: Fixed effect structure 203
6.2.2 Mixed-effects models: The effects of individual noun collocates 222
6.3 Discussion 230
6.4 Summary 238

7 Idiosyncratic collocational preferences of the near-synonyms 239
7.1 Dataset and methodology 241
7.2 Results 249
7.2.1 Semantic Vector Space modeling 249
7.2.2 Collocational networks 259
7.3 Summary 275
8 Concluding remarks and suggestions for future research 279
List of references and sources 291
Appendix A: Univariate analyses 317
Appendix B: Multivariate analyses 335
Appendix C: Collocational analyses 357
Relación de publicacións da doutoranda mencionadas na tese de doutoramento 363
Resumo en galego 369

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

LIST OF FIGURES

Figure 1. Main types of semantic change distinguished in the historical-philological tradition 13
Figure 2. Development of corpus-based research on lexical synonymy over the last 30 years 63
Figure 3. Proportions of the five near-synonyms across periods 152
Figure 4. Proportions of the concept PLEASANT SMELLING in each sense across periods..... 156
Figure 5. Proportions of the concept PLEASANT SMELLING in each semantic category across periods 161
Figure 6. Average frequency per million words in *COHA* of the semantic domains INDUSTRY AND TECHNOLOGY, ADVERTISING AND MEDIA, and COSMETICS AND HYGIENE in the period 1810–2009 169
Figure 7. Overall sense distribution (in proportions) per near-synonym 171
Figure 8. Sense distribution (in proportions) per near-synonym across periods 173
Figure 9. Overall semantic category distribution (in proportions) per near-synonym 175
Figure 10. Semantic category distribution (in proportions) per near-synonym across periods 177
Figure 11. Continuum of senses of the five adjectives..... 181
Figure 12. Overall proportions of the five near-synonyms in each sense 183
Figure 13. Proportions of the five near-synonyms in each sense across periods 184
Figure 14. Overall proportions of the five near-synonyms in each semantic category..... 186
Figure 15. Proportions of the five near-synonyms in each semantic category across periods 187
Figure 16. Variable importance of predictors according to random forest analysis..... 207
Figure 17. Effect of Concreteness rating on the probabilities of the adjectives *fragrant*, *perfumed*, and *scented*..... 212
Figure 18. Effect of Degree on the probabilities of the adjectives *fragrant*, *perfumed*,

and *scented*.....213
Figure 19. Effect of Countability on the probabilities of the adjectives *fragrant*, *perfumed*, and *scented*.....214
Figure 20. Effect of Syntactic function on the probabilities of the adjectives *fragrant*, *perfumed*, and *scented*.....215
Figure 21. Effect of Text-type on the probabilities of the adjectives *fragrant*, *perfumed*, and *scented* across periods.....216
Figure 22. Effect of Sense on the probabilities of the adjectives *fragrant*, *perfumed*, and *scented* across periods.....217
Figure 23. Effect of Semantic category on the probabilities of the adjectives *fragrant*, *perfumed*, and *scented* across periods.....219
Figure 24. Collocational preferences of *fragrant*.....225
Figure 25. Collocational preferences of *perfumed*.....226
Figure 26. Collocational preferences of *scented*.....228
Figure 27. The use of *fragrant*, *perfumed*, and *scented* in the artificial, indeterminate, and natural senses over time.....233
Figure 28. Dendrogram of the collocational preferences of *fragrant*, *perfumed*, and *scented* throughout the period 1810–2009.....250
Figure 29. Two-dimensional MDS map of the collocational preferences of *fragrant*, *perfumed*, and *scented* throughout the period 1810–2009.....252
Figure 30. Significant correlations between MDS dimensions and semantic (sub-)categories in *USAS*.....257
Figure 31. Collocational network of *fragrant*, *perfumed*, and *scented* in P1 (1810–1859).....260
Figure 32. Collocational network of *fragrant*, *perfumed*, and *scented* in P2 (1860–1909).....262
Figure 33. Collocational network of *fragrant*, *perfumed*, and *scented* in P3 (1910–1959).....264
Figure 34. Collocational network of *fragrant*, *perfumed*, and *scented* in P4 (1960–2009).....266
Figure C.1. Three-dimensional MDS map of the collocational preferences of *fragrant*, *perfumed*, and *scented* throughout the period 1810–2009.....358

LIST OF TABLES

Table 1. Classifications of synonymy by Lyons and Cruse 56
Table 2. Sense division of the five adjectives according to nine dictionaries and thesauri..... 97
Table 3. Number of words in *COHA* across decade and genre 104
Table 4. Number of instances retrieved per query..... 107
Table 5. Major semantic classes in *USAS*..... 116
Table 6. Major semantic classes in the *HTOED*..... 117
Table 7. Classification of nouns into semantic categories 127
Table 8. Variables in the first dataset and their levels 145
Table 9. Absolute and relative frequencies of the five near-synonyms..... 151
Table 10. Frequency clines of the five adjectives in the period 1810–2009 153
Table 11. Distribution of the concept PLEASANT SMELLING across senses..... 157
Table 12. Sense distribution of the semantic categories BODY AND PEOPLE, OBJECT, SENSATION, SPACE, and SUBSTANCE AND MATERIAL 160
Table 13. Distribution of the concept PLEASANT SMELLING across semantic categories..... 162
Table 14. Distribution of animate and inanimate referents per Semantic category 198
Table 15. Absolute and relative frequencies of *fragrant*, *perfumed*, and *scented*..... 201
Table 16. Hypothetical scenario of reordering of levels in random forest analysis 202
Table 17. Model summary of multinomial logistic regression Model A (Sense) 203
Table 18. Model summary of multinomial logistic regression Model B (Semantic category)..... 205
Table 19. Model summary of multinomial logistic regression with only semantic predictors (Sense)..... 205
Table 20. Model summary of multinomial logistic regression with only semantic predictors (Semantic category)..... 205
Table 21. Model A coefficients (Sense) 209
Table 22. Model B coefficients (Semantic category)..... 210

Table 23. Model summary of the binary mixed-effects logistic regression (*fragrant*).....222
Table 24. Model summary of the binary mixed-effects logistic regression (*perfumed*).....222
Table 25. Model summary of the binary mixed-effects logistic regression (*scented*).....222
Table 26. Structure of the dataset.....241
Table 27. Similarity matrix based on cosine similarity scores for the data 245
Table 28. Distance matrix for the data.....246
Table 29. Settings for the construction of collocational networks.....249
Table 30. AU *p*-values for the five-cluster solution in Figure 28 251
Table 31. Significant correlations between the dimensions in Figure 29 and semantic (sub-)categories in *USAS* 254
Table 32. Examples of types of noun collocates in the semantic (sub)-categories in *USAS* 255
Table 33. Number and percentage of consistent, initiating, terminating, and transient collocates of *fragrant*, *perfumed*, and *scented*.....268
Table 34. Consistent, initiating, terminating, and transient noun collocates of *fragrant*.....269
Table 35. Consistent, initiating, terminating, and transient noun collocates of *perfumed*.....272
Table 36. Consistent, initiating, terminating, and transient noun collocates of *scented*.....273
Table A.1. Chi-square residuals of the frequency of the five adjectives in the period 1810–2009 317
Table A.2. Chi-square residuals of the concept PLEASANT SMELLING across senses in the period 1810–2009 317
Table A.3. Diachronic development of the semantic category SPACE according to sense.....317
Table A.4. Chi-square residuals of the concept PLEASANT SMELLING across semantic categories in the period 1810–2009 318
Table A.5. Average frequency per million words in *COHA* of the semantic domains INDUSTRY AND TECHNOLOGY, ADVERTISING AND MEDIA, and COSMETICS AND HYGIENE in the period 1810–2009 318
Table A.6. Sense distribution of *fragrant* 319
Table A.7. Sense distribution of *perfumed* 319
Table A.8. Sense distribution of *scented*.....320
Table A.9. Sense distribution of *sweet-scented* 320
Table A.10. Sense distribution of *sweet-smelling*.....321

Table A.11. Semantic category distribution of *fragrant* 321
Table A.12. Semantic category distribution of *perfumed* 323
Table A.13. Semantic category distribution of *scented* 324
Table A.14. Semantic category distribution of *sweet-scented* 325
Table A.15. Semantic category distribution of *sweet-smelling* 326
Table A.16. Overall absolute and relative frequencies of the five near-synonyms in each sense 327
Table A.17. Absolute and relative frequencies of the five near-synonyms in each sense in P1 (1810–1859) 327
Table A.18. Absolute and relative frequencies of the five near-synonyms in each sense in P2 (1860–1909) 328
Table A.19. Absolute and relative frequencies of the five near-synonyms in each sense in P3 (1910–1959) 328
Table A.20. Absolute and relative frequencies of the five near-synonyms in each sense in P4 (1960–2009) 329
Table A.21. Overall absolute and relative frequencies of the five near-synonyms in each semantic category 330
Table A.22. Absolute and relative frequencies of the five near-synonyms in each semantic category in P1 (1810–1859) 331
Table A.23. Absolute and relative frequencies of the five near-synonyms in each semantic category in P2 (1860–1909) 332
Table A.24. Absolute and relative frequencies of the five near-synonyms in each semantic category in P3 (1910–1959) 333
Table A.25. Absolute and relative frequencies of the five near-synonyms in each semantic category in P4 (1960–2009) 334
Table B.1. Distribution of the near-synonyms according to Animacy 335
Table B.2. Mean and median Concreteness rating 335
Table B.3. Distribution of the near-synonyms according to the Countability 336
Table B.4. Distribution of the near-synonyms according to Syntactic function 336
Table B.5. Distribution of the near-synonyms according to Degree 337
Table B.6. Distribution of the near-synonyms according to Text-type 337
Table B.7. Period contrasts in fiction 338

Table B.8. Period contrasts in non-fiction 339
Table B.9. Period contrasts in periodicals.....340
Table B.10. Period contrasts in the artificial sense 341
Table B.11. Period contrasts in the figurative sense 342
Table B.12. Period contrasts in the indeterminate value.....343
Table B.13. Period contrasts in the natural sense 344
Table B.14. Period contrasts in the semantic category ABSTRACT.....345
Table B.15. Period contrasts in the semantic category BODY AND PEOPLE 346
Table B.16. Period contrasts in the semantic category CLEANING 347
Table B.17. Period contrasts in the semantic category COSMETICS 348
Table B.18. Period contrasts in the semantic category EARTH, ATMOSPHERE, AND WEATHER 349
Table B.19. Period contrasts in the semantic category FOOD AND DRINK 350
Table B.20. Period contrasts in the semantic category OBJECT.....351
Table B.21. Period contrasts in the semantic category PLANTS AND FLOWERS 352
Table B.22. Period contrasts in the semantic category SUBSTANCE AND MATERIAL.....353
Table B.23. Period contrasts in the semantic category SENSATION.....354
Table B.24. Period contrasts in the semantic category SPACE 355
Table B.25. Period contrasts in the semantic category TEXTILE AND CLOTHING.....356
Table C.1. Goodness-of-fit statistics comparison of three SVS analyses.....357
Table C.2. Most frequent noun collocates of *sweet-scented* in the four periods 357
Table C.3. Most frequent noun collocates of *sweet-smelling* in the four periods.....358
Table C.4. PMI score of the noun collocates of *fragrant* in the four periods distinguished...359
Table C.5. PMI score of the noun collocates of *perfumed* in the four periods distinguished.360
Table C.6. PMI score of the noun collocates of *scented* in the four periods distinguished....361

ABBREVIATIONS AND CONVENTIONS

ABBREVIATIONS

Corpora, dictionaries, and thesauri

AHDOE = American Heritage Dictionary of the English Language

BNC = British National Corpus

CD = Cambridge Dictionaries

COCA = Corpus of Contemporary American English

CoD = Collins Dictionary

COHA = Corpus of Historical American English

HTOED = Historical Thesaurus of the Oxford English Dictionary

LDOCE = Longman Dictionary of Contemporary English

MD = MacMillan Dictionary

MW = Merriam-Webster

NHDAE = Newbury House Dictionary of American English

OED = Oxford English Dictionary

USAS = UCREL Semantic Analysis System

Periodization in the history of English

OE = Old English

ME = Middle English

eModE = Early Modern English

lModE = Late Modern English

PDE = Present-day English

Semantic categories

ABS = ABSTRACT

B&P = BODY & PEOPLE

CL = CLEANING

COS = COSMETICS

EAW = EARTH, ATMOSPHERE, AND EARTH

F&D = FOOD & DRINK

OBJ = OBJECTS

P&F = PLANTS & FLOWERS

S&M = SUBSTANCES & MATERIAL

SEN = SENSATIONS

SPACE = SPACES

T&C = TEXTILE MATERIAL & CLOTHING

USAS Semantic tags (only those referred to in the text)

B = THE BODY AND THE INDIVIDUAL

B1 = ANATOMY AND PHYSIOLOGY

B4 = CLEANING AND PERSONAL CARE

B5 = CLOTHES AND PERSONAL BELONGINGS

F1 = FOOD

F2 = DRINKS

F4 = FARMING AND HORTICULTURE

H = ARCHITECTURE, BUILDINGS, HOUSES AND THE HOME

H5 = FURNITURE AND HOUSEHOLD FITTINGS

L3 = PLANTS

M3 = MOVEMENT/TRANSPORTATION: LAND

M7 = PLACES

O1 = SUBSTANCES AND MATERIALS GENERALLY

O1.1 = SUBSTANCES AND MATERIALS GENERALLY: SOLID

O2 = OBJECTS GENERALLY

Q1 = COMMUNICATION

Q1.2 = PAPER DOCUMENTS AND WRITING

Q4 = THE MEDIA

S2 = PEOPLE

W = THE WORLD AND OUR ENVIRONMENT

W3 = GEOGRAPHICAL TERMS

W4 = WEATHER

X3 = SENSORY

Other

AmE = American English

AU = Approximately Unbiased *p*-values

BrE = British English

HAC = Hierarchical Agglomerative Cluster

HCFA = Hierarchical Configural Frequency Analysis

MDS = Multidimensional Scaling

NFs = Normalized frequencies

PMI = Pointwise Mutual Information

POS = Part of speech

SVS = Semantic Vector Space

CONVENTIONS

Italic type is employed for example words and phrases in the running text, as well as for the names of dictionaries and corpora, both in full and in abbreviated form.

SMALL CAPITALS are employed for semantic concepts or domains and for semantic categories, both those included in my own semantic classification and those provided by the *HTOED* and *USAS*. Additionally, SMALL CAPITALS are used for conceptual metaphors and metonymies.

Bold type is employed for the purpose of highlighting words in the examples provided.

Dictionary citations appear in the format (1538. *OED*, s.v. *perfumed* adj. 1), indicating year — if applicable—, name of the dictionary, headword, part of speech, and sense number.

Single inverted commas (‘’) are used for meanings and for search queries made in the corpus.

The first word of variable names is capitalized (e.g. Semantic category), whereas variable levels are not capitalized (e.g. animate).

The font `Lucida console` is used when referring to packages and functions in *R*.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

1 INTRODUCTION

What exactly is a synonym? This is a question that has excited much interest among linguists and laymen alike and for which slightly different answers have been proposed in the literature through the years. Even a cursory glance at definitions of the term *synonym* in dictionaries reveals that there is yet no agreement regarding exactly what constitutes synonymy. Despite the fact that all sources coincide in characterizing this semantic relation in terms of similarity of meaning, two groups of definitions emerge. While some dictionaries define the term *synonym* as "[…] a word or expression which means the same as another word or expression" (*Collins Dictionary*, s.v. *synonym* n.), in other sources a partial degree of similarity is considered sufficient for a word or expression to constitute a synonym of another item: "a word having the same or *nearly* the same meaning as another word or other words in a language" (emphasis mine; *American Heritage Dictionary of English*, s.v. *synonym* n. 1). As has been argued by many scholars (e.g. Cruse 2000: 157–158; Taylor 2003: 264; Divjak & Gries 2006: 24; Liu 2010: 56–57), absolute synonymy, that is, complete identity of meaning, is an extremely rare, if not a non-existent, phenomenon in language, as opposed to near-synonymy, which is exceedingly common. This can easily be experienced by considering for a moment the use of any pair or larger group of synonymous expressions in a given language. Regardless of the synonym set one selects, the most likely scenario is one in which there exist some contexts of use where one of the terms is either the only possible choice or at least a more suitable option.

Such an impressionistic experiment suggests that in all probability, the great majority of synonyms differ in one way or another, which prevents them from being completely interchangeable in all contexts of use. As argued by the philosopher Thomas Reid, "[m]ost synonimes [sic] have some minute distinction that deserves notice" (1785: 14; quoted in the *OED*, s.v. *synonym* n. 1a). Therefore, perhaps the most interesting question which arises when dealing with synonyms is not what makes them similar but, on the contrary, where exactly their differences reside. This is an issue which has fascinated not only linguists, but also ordinary speakers. As a matter of fact, as shown by Murphy (2013: Section 3), one of the most frequent

uses of dictionaries and thesauri is precisely to search for synonyms, especially in language production, in order to avoid repetition, achieve a more sophisticated writing, or adjust the style employed, among others. In fact, selecting the most appropriate word from a set of candidate synonymous expressions is a notoriously difficult task, both for non-native and for native language users, because synonyms can vary along a great number of dimensions —up to 35 according to Edmonds (1999: Chapter 3)— and differences are often multidimensional in nature (e.g. Cruse 1986: 284–289; 2000: 159–160; Murphy 2003: 154–157). In view of this, it is not surprising that research on synonymy has recently moved away from an excessive emphasis on the classification of synonyms into types and subtypes to a focus on the identification of the factors underlying their choice, thus zooming in on the differences between synonymous expressions. To this purpose, innovative methods and techniques have been employed in neostructuralist and cognitive semantics to capture precisely the dimensions along which particular pairs or sets of lexical synonyms vary, with successful results (e.g. Persson 1989; Arppe 2002; 2008; Taylor 2003; Divjak 2006; 2010; Divjak & Gries 2006; 2008; Gries & Otani 2010; Liu 2010; 2013b; Liu & Espino 2012; Desagulier 2014; Krawczak 2014; 2018). Such studies have proven the suitability of these approaches to uncover fine-grained differences among synonyms through the use of corpus data and in-depth analyses of the contexts in which these words are used, including semantic, lexical, morphosyntactic, and stylistic ones, among others.

Despite the usefulness of such approaches for the study of synonymy, most research of this type adopts a synchronic perspective, thus examining the set of factors that are thought to influence the alternation between synonymous lexical expressions at a particular point in time, most typically the present day. On the contrary, diachronic usage-based studies on how the relation between members of a lexical synonym set gradually changes over time are relatively scarce, with only a handful of such investigations over the last two decades (e.g. Kaunisto 2001; Primahadi-Vijaya-R & Rajeg 2014; Baker 2017: 95–101). In addition, the underlying motivations for changes in the internal semantic structure of such semantically related words have traditionally been reduced to the so-called *no synonymy rule* (e.g. Samuels 1972: 65; Bolinger 1977: ix–x, 9; Wierzbicka 1988: 13–14; Croft 2000: 176; Nuyts & Byloo 2015: 62–63), whereby languages are assumed to work against absolute synonymy. In particular, synonyms are expected to become increasingly differentiated over time or, if this is not the case, one or more of the terms from a given set are expected to drop out of use so as to ensure that

languages remain efficient and economic systems of communication. However, it has recently been claimed that this theory of competition is insufficient to account for all changes undergone by synonymous expressions and that other diachronic processes may also be at play (De Smet et al. 2018). Besides differentiation and substitution, it has been proposed that synonyms can also increasingly converge in functional terms as time passes. However, this process of attraction has thus far not received as much attention in semantic research as the more well-understood competition phenomena. Against this backdrop, the present dissertation aims at further contributing to our knowledge of the factors influencing lexical choice in English, as well as whether and how their effects fluctuate over time and the motivations behind such changes. In the remainder of this chapter, the aims and scope of the dissertation are first discussed in Section 1.1, followed by an outline of its structure in Section 1.2.

1.1 AIMS AND SCOPE OF THE STUDY

The main goal of the present dissertation is to fill a gap in the specialized literature on synonymy by conducting a usage-based corpus study of a set of near-synonymous adjectives from the relatively understudied semantic domain of SMELL in 19th and 20th century American English (AmE), to wit, *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*, which designate the concept PLEASANT SMELLING. The five adjectives are illustrated in examples (1)–(5):

(1) They made campfires with top-quality hardwood, which released **fragrant** smoke, to cook for their families. (*COHA*, 2009, FIC, FarawayWarNovel)

(2) Dellius had seen enough easterners to know that the curls were not dirty greasy; easterners pomaded their locks with **perfumed** creams. (*COHA*, 2007, FIC, ColorsInsulting)

(3) Avoid chemical irritants in bubble baths and heavily **scented** soaps, and don't let her sit in a wet bathing suit for too long. (*COHA*, 2009, MAG, Coca)

(4) As his eyes adjusted ever so slowly to the gloom, he saw looming shadows, blurred shapes like enormous trees that stirred not at all in the gentle **sweet-scented** wind. (*COHA*, 2005, FIC, FantasySciFi)

(5) Route 214 turned out to be a winding lane through **sweet-smelling** fields long abandoned to goldenrod and blackberry brambles. (*COHA*, 2008, FIC, BookOldHouses)

The reasons for choosing this specific synonym set are twofold. First, the *Historical Thesaurus of the Oxford English Dictionary* (*HTOED*) lists an extremely great amount of word senses in the semantic domain of SMELL, a large number of which are in fact adjectives. Many of these adjectives are defined in a similar way in dictionaries, with some minor variations. Therefore, this seems to be a particularly rich semantic field for the study of near-synonymy. Second, the rather low degree of polysemy of the near-synonymous adjectives at issue here constitutes an advantage for a study focusing on synonymy. If a highly polysemous set had been selected instead, the study would not have been focused exclusively on the semantic relation of synonymy, since polysemy would also play a prominent role in the discussion.

The data used is drawn from the *Corpus of Historical American English* (*COHA*; Davies 2010–), which covers the time span 1810–2009. The focus is on the distributional patterns of the selected near-synonymous expressions as a proxy to quantify their functional (dis)similarity. Therefore, following recent synchronic approaches to synonymy, this semantic relation is here measured in quantitative terms by examining a series of factors which have been shown to play an important role in lexical choice. The factors analyzed represent both language-internal semantic and non-semantic contexts of use, as well as language-external variables. The influence of these factors is examined over the time span covered by the corpus, which is divided into four 50-year periods, to wit, 1810–1859, 1860–1909, 1910–1959, and 1960–2009, in order to capture how each lexical item, as well as the relation between the members of the set, evolves over time. To this purpose, the data is submitted to both univariate and multivariate statistical techniques which have been previously employed to measure the semantic (dis)similarity of near-synonymous expressions from a synchronic perspective. Consequently, a secondary objective of the dissertation is to assess the usefulness for diachronic research on lexical synonymy of these primarily synchronic quantitative methods.

From a theoretical perspective, the present dissertation is grounded on the tenets of neostructuralist and cognitive semantics, both of which emphasize the utmost importance of actual use in language structure. In addition, the study draws on insights from theories of semantic change by considering the various processes at work in the development of the internal semantic structure of synonym sets, particularly those of differentiation and substitution, but also that of attraction, as well as the underlying motivations of such processes and their connection with the extralinguistic reality. Therefore, yet another aim of this dissertation is to

determine whether the traditional theory of competition and substitution is sufficient to account for the changes taking place in the selected synonym set, or whether, on the contrary, it is necessary to resort also to other processes of semantic change such as attraction.

In sum, by adding a diachronic dimension to the quantitative study of lexical near-synonymy, this dissertation provides one of the first applications of primarily synchronic methods to historical research on the semantic relation of synonymy. Moreover, the study considers a larger set of near-synonymous expressions from the same domain than is typically examined in the specialized literature and these near-synonyms are analyzed across a wide variety of contexts of use. Therefore, despite the relatively long research tradition on synonymy, this dissertation provides an innovative approach to the study of this semantic relation, as well as new and interesting findings regarding the development of near-synonyms over time.

1.2 STRUCTURE

The present dissertation is divided into eight chapters, organized as follows. Chapters 2 and 3 provide a review of the relevant literature on notions in and approaches to lexical semantics, in general, and to synonymy, in particular. Chapter 4 explains the data used for the analyses conducted, focusing on the near-synonyms object of study and the corpus employed, as well as the data extraction and annotation processes. Chapters 5–7 present the results of three separate corpus-based studies on the concept PLEASANT SMELLING and the items denoting this concept. Finally, Chapter 8 offers some concluding remarks and suggestions for future research.

The purpose of Chapter 2 is to provide an overview of the most important linguistic schools in research on lexical semantics over the last 200 years, focusing particularly on the origins and development of the ideas and notions underlying the theoretical background of the studies conducted in this dissertation. Section 2.1 summarizes the most important principles of historical-philological semantics, the framework that dominated the field from around the 1830s to the 1930s. A large part of this section is devoted to explaining the central tenets of this school, particularly its diachronic perspective and detailed classifications of semantic change, which were taken up by later schools and are still fundamental today. Section 2.2 moves on to the structuralist and generativist schools of linguistics. Special attention is paid here to the study of semantic relations in the structuralist tradition, which are thought to form the basis for the delineation and demarcation of meaning in the lexicon. In Section 2.3, the focus shifts to neostructuralist movements, in particular to the distributional corpus-based method set out by

Firth and Sinclair in the 1950s and 1960s. This approach and its emphasis on collocational behavior lie at the core of the methodological and theoretical assumptions of the present dissertation, as well as of much recent work in lexical semantics. Finally, Section 2.4 is concerned with cognitive semantics, a maximalist approach to the study of language, in general, and of meaning, in particular. Two of the most important contributions of this school figure prominently in this section, to wit, prototype theory and usage-based theories of semantic change, which turn out to be relevant for the interpretation of the findings of the analyses carried out in Chapters 5–7.

Chapter 3 zooms in on the semantic relation of synonymy. First, the most relevant classifications of synonyms into types and subtypes are discussed in Section 3.1. The main conclusion drawn from these classifications is that a distinction between absolute synonymy and near-synonymy is the most adequate for an empirical investigation into the semantic phenomenon of synonymy. Then, Section 3.2 moves on to an in-depth review of synchronic usage-based corpus studies on specific pairs or groups of near-synonyms that have been published over the last 30 years. Two separate waves of research are first identified and explained in Section 3.2.1, which suggest that considerable developments have recently taken place in this area of research. Then, in Section 3.2.2, the focus turns to studies on near-synonymous adjectives, since the items belonging to the synonym set analyzed in this dissertation are precisely adjectives. Finally, Section 3.3 provides an overview of diachronic research on lexical synonymy by first considering the few usage-based corpus studies on the historical development of particular pairs or groups of near-synonyms, and then shifting the attention to general processes of semantic change relevant for the diachronic study of synonymy.

Chapter 4 is concerned with issues related to the data employed in the analyses provided in subsequent chapters. The chapter opens with a description in Section 4.1 of the synonym set object of study; it outlines the main motivations for its selection and then reviews the information on these lexical items in various present-day and historical dictionaries and thesauri. Next, Section 4.2 discusses the corpus used as a source of data, as well as the data retrieval process. Lastly, Section 4.3 deals with the data annotation process. Considerable space is devoted here to the explanation of the language-internal and language-external variables used for the codification of the data.

Chapter 5 moves on to the first corpus-based study presented in the dissertation, which constitutes a descriptive univariate analysis of the data. The chapter is divided into four sections, each focusing on a different aspect of the synonym set. Section 5.1 provides the overall frequency of the five near-synonyms, as well as their frequency developments over the time span 1810–2009. Then, Section 5.2 turns to the concept PLEASANT SMELLING as a whole, that is, as designated by the five adjectives, and its distribution across senses and semantic categories of modified nouns. Next, Section 5.3 examines the behavior of each adjective separately, thus offering a semasiological perspective of the data. Finally, Section 5.4 deals with the competition between the near-synonyms in different semantic contexts, hence adopting an onomasiological orientation.

Chapter 6 delves deeper into the onomasiological structure of the synonym set. This is done by measuring the competition between the near-synonyms across a large number of intralinguistic and extralinguistic factors. Here, the perspective adopted is a multivariate one and hence a series of more sophisticated statistical techniques are employed. These statistical methods are explained in Section 6.1, together with the dataset used for this particular analysis. Section 6.2, in turn, discusses the results of the study by first focusing on the effects of the predictors included (Section 6.2.1) and then offering a preliminary enquiry into the collocational behavior of the adjectives (Section 6.2.2), where their individual noun collocates are taken into consideration. The chapter closes with a discussion of the main findings in Section 6.3, highlighting the importance of the idiosyncratic collocational preferences of the adjectives under scrutiny, a topic which is taken up in Chapter 7.

Chapter 7 opens with a description of the dataset and the methods used in the collocational analyses (Section 7.1). Section 7.2 deals with the results of these analyses and their discussion. A bird's-eye view of the collocational behavior of the near-synonymous adjectives is provided in Section 7.2.1 by having a look at their general co-occurring patterns before moving to a more qualitative examination of the most prominent individual noun collocates in Section 7.2.2.

The dissertation concludes with Chapter 8, a summary of the most relevant generalizations that can be drawn from the studies conducted in the preceding chapters and the identification of some areas that can be further improved and taken up in future research on the topic.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

2 APPROACHES TO AND NOTIONS IN LEXICAL SEMANTICS

Lexical semantics has been a more or less prominent discipline within linguistics since the beginning of the 19th century when it emerged as an established field of study. From a historical perspective, the research conducted in this field over approximately the last 200 years can be classified into five broad theoretical movements or traditions, to wit, historical-philological semantics, structuralist semantics, generative semantics, neostructuralist semantics, and cognitive semantics. The perspectives that these schools have adopted to the study of semantics are varied and there is yet no agreement on how best to describe lexical meaning. Some scholars (e.g. Geeraerts 2010: 277) suggest that the historical development of this field can be characterized on the basis of a series of oppositions that serve to describe the main interests of each of these five movements. First, the schools' approaches to the study of word meaning vary depending on whether meaning is conceptualized as a purely linguistic entity, separate from general human cognition, or whether the latter, in the form of encyclopedic knowledge, should also be contemplated in semantic description. Second, the five movements diverge in terms of the prominence they assign to language use (i.e. *parole* and *performance* in the parlance of Saussure 1916 and Chomsky 1965, respectively). A third dimension along which the traditions differ is that of *semasiology* (i.e. from word form to meaning) vs. *onomasiology* (i.e. from meaning to word form), with some schools focusing primarily on the study of the former and others on the latter. Lastly, the emphasis given to either synchronic or diachronic research varies from one school to another: while, for instance, historical-philological semantics is mainly interested in semantic change, embracing a full-fledged diachronic orientation, structuralist semantics generally opt for a synchronic approach.

These five schools should not be considered in isolation but rather in relation to one another, as the transition from one movement to the next can be characterized either in terms of continuation or of reaction along the four dimensions just described. For example, whereas the historical-philological tradition provides many ideas that are later revisited and further

developed within the framework of cognitive semantics, many of its main assumptions are flatly rejected by structuralist semantics. Therefore, to understand the concepts and principles present in one tradition, it is often necessary to have some knowledge of the key developments in the preceding ones.

In what follows, a brief overview of some of the main ideas of each of the five movements is presented. However, as the approach adopted in this dissertation follows the neostructuralist distributional corpus-based methodology, while at the same time embracing the usage-based onomasiological stance of cognitive semantics, only those aspects relevant for the development of the assumptions which lie at the core of both traditions are here discussed in depth. Therefore, the aim of this chapter is not to give an exhaustive account of the research conducted within the different schools, but rather to comment on those issues and concepts that have contributed significantly in one way or another to the principles that characterize the approach adopted here.[1] Due to the diverging foci of the five movements, some of them play a more crucial role than others in the evolution of the theoretical foundations which are of central importance throughout the analyses in Chapters 5–7 and are, consequently, described in greater detail. For instance, in view of the diachronic nature and primarily onomasiological orientation of the analyses conducted in this dissertation, many of the central issues of both historical-philological and structuralist semantics are highly relevant to understand subsequent developments within neostructuralist and cognitive semantics. On the other hand, the contributions of generative semantics are much less pertinent for our purposes.

This chapter is structured into four main sections, each of them focusing on the major developments of the five movements separately, with the exception of structuralist and generative semantics, which are treated together, given the great importance attributed to formalization in both movements. As such, the focus of Section 2.1 is on historical-philological semantics, while structuralist and generative semantics are discussed in Section 2.2, where more emphasis is placed on structuralist approaches than on generative ones. In Section 2.3 the neostructuralist methods are described, paying special attention to the principal assumptions of the distributional corpus-based approach. Section 2.4, in turn, offers an overview of the main theoretical concerns of cognitive semantics, focusing particularly on the concepts of prototypicality, salience, and entrenchment, as well as on its major contributions to diachronic

[1] The information included in Sections 2.1 and 2.2 of the present chapter is mainly drawn from Geeraerts (2010). For a detailed historical overview of the field of lexical semantics, the reader is referred to this monograph.

semantics. Finally, Section 2.5 provides a summary of the whole chapter, as well as some conclusions that can be drawn from it. The four oppositions mentioned earlier (i.e. semantic vs. encyclopedic knowledge, use vs. system, semasiology vs. onomasiology, and synchrony vs. diachrony) serve as a guiding theme in the remainder of this chapter in the sense that the five traditions are described in terms of their positions with respect to each of these dimensions.

2.1 HISTORICAL-PHILOLOGICAL SEMANTICS

Historical-philological semantics, sometimes also known as traditional diachronic semantics or prestructuralist semantics, dominated the field for approximately 100 years between 1830 and 1930. As mentioned in the preceding paragraphs, this school takes a diachronic perspective towards semantics and thus displays a great interest in meaning change, especially at the level of individual words. As the evolution of isolated words and their senses constitute the principal object of study, the main concern within this tradition is on semasiological structure in the lexicon. It is in order here to explain the difference of perspective between the study of semasiology and that of onomasiology. Whereas the former takes as its starting point particular words in order to study their senses and the concepts that they designate, the latter starts from specific concepts and explores the various expressions which can be used to represent them (Baldinger 1980: 278; Grondelaers, Speelman & Geeraerts 2007: 988–989; Kay & Allan 2015: 2–3, 185, 188). It is worth noting here that these two different views are commonly associated with the semantic phenomena of polysemy and synonymy, respectively (Levshina 2011: 11): whereas polysemy refers to the existence of various related word meanings or senses of one lexical item (i.e. semasiology), synonymy involves the existence of various lexical items denoting the same concept (i.e. onomasiology).

One of the major contributions of historical-philological semantics to contemporary diachronic approaches concerns the various classifications of semantic change (cf., for instance, Carnoy 1927; Stern 1931), which are in fact still in use today in refined and further developed forms.[2] In order to analyze the relations holding between different senses of particular words and the changes that they undergo over time, a great deal of the work within this movement

[2] Although various other classifications of semantic change within historical-philological semantics have been proposed, Carnoy's and Stern's are more elaborate and provide a more fine-grained depiction of the major types of changes. Another influential classification for contemporary research in semantics is that of Ullman (1957; 1962), which despite falling outside the period of the historical-philological tradition, is worth mentioning here.

revolve around the identification of regular mechanisms of semantic change. Without going into too much detail about the numerous classifications and subclassifications postulated by particular scholars, a brief overview of the main types of changes is provided in what follows (for a visual summary of the different types, see Figure 1 below). A first distinction is typically made between onomasiological and semasiological mechanisms of meaning change. The former involve changes that lead to the association of a newly coined expression to already existing ones in the lexicon. These changes, typically also known as lexicogenetic mechanisms, include morphological processes such as derivation, composition, clipping, and blending, as well as lexical borrowing from other languages (not to be confused with semantic borrowing, see below). However, since the focus within this tradition is primarily on semasiology, most of the types distinguished in these classifications fall within the semasiological group of changes, i.e. those that bring about new senses of already existing words or expressions. Changes of this kind can, in turn, be further divided into connotational and denotational changes, and within the latter scholars typically differentiate between non-analogical and analogical mechanisms. Analogical processes are those in which the new sense of a given word is established in relation to the semantics of a pre-existing expression, either in terms of similarity or in terms of opposition, as in the case of semantic borrowing and differentiation, respectively. Contrariwise, this is not the case in non-analogical changes, such as specialization and generalization, among others (cf. also Geeraerts 1997: 93–102).[3]

[3] Since a detailed description of the semantic changes relevant for the study of synonymy, and more specifically to the analyses conducted in the present dissertation, is provided in Section 3.3, the analogical changes included in Figure 1 (e.g. specialization, generalization, and differentiation) are not further discussed at this point.

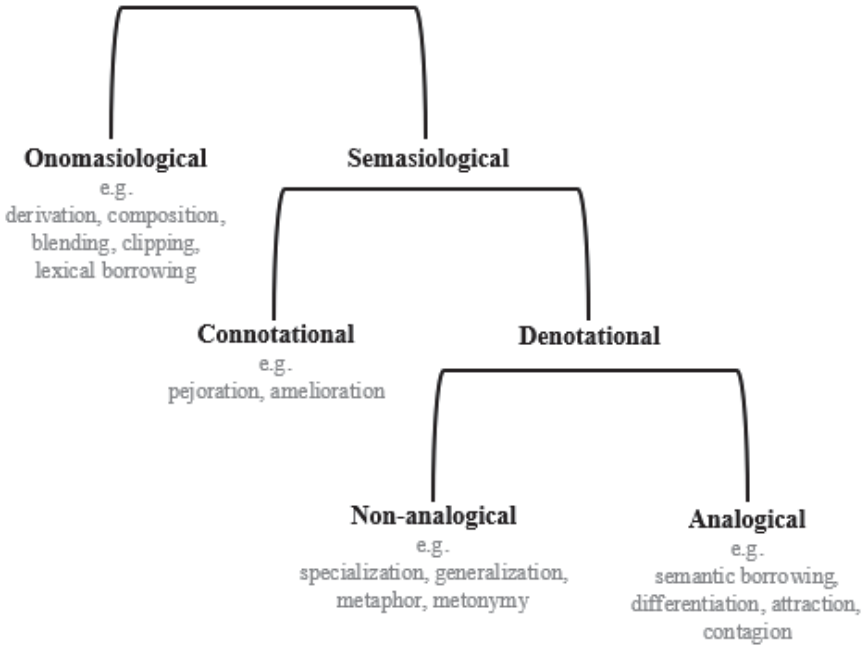


Figure 1. Main types of semantic change distinguished in the historical-philological tradition

Concerning the conceptualization of meaning, that is, the opposition between semantic vs. encyclopedic knowledge, the perspective adopted in traditional diachronic semantics is a psychological one, in which language is seen as being closely connected to our general cognitive abilities, which together help us understand the world we live in. Bréal (1897), who brought the mental position of linguistic meaning to the fore, is one of the most important scholars within this movement. Three of Bréal's main points should be mentioned here to shed further light on this perspective of meaning. First, Bréal contemplates linguistic meanings as thoughts or ideas, that is, as psychological entities, which permit language users to categorize their experiences of the world. It is in this sense that language is directly connected to, and thus not independent from, the rest of human cognition. Second, it follows from this conception of language in general, and of meaning in particular, that semantic changes are also cognitively motivated and thus result from mental processes. As such, the general mechanisms of semantic change identified above (cf. Figure 1) can also be understood as a result of the way in which human cognition works, for instance, by means of metaphorical and metonymical thought processes. Third, Bréal argues that the motivations behind semantic change are ultimately functional in nature because changes occur in order to adjust language to the communicative needs and experiences of its users, which are themselves ever evolving. This psychological conceptualization of meaning, which Bréal emphasized and explored already at the end of the

19th century, still figures prominently within cognitive semantics more than a century later, as we will see in Section 2.4.

A further relevant feature of historical-philological semantics is the standpoint it takes on the importance of language usage and of specific contexts of use in the conception of meaning, especially when it comes to semantic change. In this sense, Paul (1920: 75) argues that we need to differentiate between what he calls *usual* and *occasional* meanings. Usual meaning refers to the conventional senses of individual word forms, whereas occasional meaning is applied to the specific readings of the established senses in particular contexts of use. To further illustrate this idea, consider the following two made-up examples featuring the expression *red jacket*:

(6) I really like the **red jacket** that you are wearing today.
(7) Can you please give this book to the **red jacket** sitting in the corner over there?

In (6) we find an instance referring to one of the various usual meanings of *jacket* listed in the *OED*, namely '[a]n outer garment for the upper body […]' (*OED*, s.v. *jacket* n. I. 1a), whereas (7) displays an occasional meaning of the same word, in which it is used to designate the person *sitting in the corner over there* who is wearing a red jacket in that specific speech situation. In this case, the occasional meaning emerges from a metonymical pattern which can be characterized as PIECE OF CLOTHING FOR PERSON (e.g. Waag 1908; Nyrop 1913; Paul 1920; Esnault 1925). However, even though this particular reading of *jacket* can be inferred in this context, it has not been conventionalized as an established sense of the word in question. This example serves to demonstrate two fundamental aspects of Paul's conception of meaning. On the one hand, it reveals the significance that he ascribes to language use so as to understand the meaning of particular instances of words since the occasional meaning of a word may differ substantially from its usual meaning. In addition, context also plays a fundamental role, given that the realization of "the occasional meaning amounts to selecting the appropriate reading from among the multiple established senses of a [polysemous] word" (Geeraerts 2010: 15). Thus, for instance, the English word *hot* is bound to be interpreted differently if it serves as a modifier of nouns referring to liquids such as *water* or *coffee* than if it serves as a modifier of nouns denoting people. On the other hand, the distinction usual vs. occasional meaning shows that Paul views meaning as extremely dynamic and flexible due to the concrete interpretations that lexical items acquire in specific contexts of use.

According to Paul, language use is also essential to understand semantic changes taking place at the level of individual words, that is, semasiological changes. By making use of the distinction he establishes between the two types of meaning, he goes on to explain the way in which occasional meanings may develop into usual ones over time. Paul argues that this happens by means of repetition, in which case an occasional meaning of a word, by being used more and more frequently, comes to establish itself as another conventional sense of that word alongside others. In other words, the occasional meaning becomes independent from the usual meaning it derives from and can, consequently, be interpreted autonomously or without context. In this sense, we may say that it has become decontextualized. Moreover, Paul suggests that the opposite process, that is, the modulations of usual into occasional meanings, is directly linked to the classifications of semantic change introduced above (e.g. metonymy, metaphor, and specialization), since these mechanisms are precisely those through which an occasional meaning arises from a usual one. A case in point are examples (6) and (7), in which a metonymical extension resulted in a new occasional meaning derived from an established one. If contextually dependent readings such as this occur frequently enough in actual speech, they may become conventionalized and over time establish themselves in the language system. In fact, Paul's view of semantic change serves as a clear historical precedent of the usage-based model of change of cognitive semantics, which is explained in Section 2.4.2.

As we have seen, both Bréal's (1897) and Paul's (1920) conceptions of meaning as a cognitive and contextual phenomenon clearly emphasize the causes leading to semantic changes and can thus be said to go a step beyond the focus dominating the historical-philological movement until the 1860s, namely that of identifying and classifying semantic changes without paying much attention to the explanatory forces underlying such developments. Bréal's and Paul's psychological models of meaning prevailed in the historical-philological school from the 1860s onwards, although different views of these models coexisted. Scholars such as Waag (1908), Erdmann (1910), Nyrop (1913), and Esnault (1925) postulated their own views and opinions on the matter. A case in point, and perhaps a crucial one to understand later developments in the history of cognitive semantics, is Erdmann's claim about the widespread presence of vagueness, and not just polysemy, in the lexicon (1910: 5, quoted in Geeraerts 2010: 22):

> Worte sind vielmehr im allgemeinen Zeichen für ziemlich unbestimmte Komplexe von Vorstellungen, die in mehr oder minder loser Weise zusammenhängen […] Die Grenzen der Wortbedeutung sind verwaschen, verschwommen, zerfliessend. Treffender aber noch wird meines Erachtens der Sachverhalt gekennzeichnet, wenn man überhaupt nicht von Grenzlinien des Umfangs redet, sondern […] von einem Grenzgebuet, das einen Kern einschliesst. […] Den Kern denken wir uns dann alle Umständen die Benenung durch das fragliche Wort zukommt, während wir dem Grenzgebiet alle diejenigen Vorstellungen weisen, denen man die Benennung sowhol zu- wie absprechen kann.
>
> (Words in general are rather signs for fairly unspecific complexes of mental representations that belong together more or less loosely […] The boundaries of word meanings are vague, unclear, indeterminate. The situation is, I think, even more adequately described if one simply does not talk about the borderline of the range of a word, but […] if one talks about a border area that includes a central one. […] In the central area we situate those things and other representations that would under all circumstances deserve to be named by the word in question, while in the border area we situate all those representations for which the name may or may not hold.)

To illustrate this idea, he considers the concept GERMAN. He argues that there are three characteristics that a person needs to exhibit in order to be considered as a referent of this word in the extralinguistic world, namely citizenship, native tongue, and ancestry. If an individual complies with all three criteria, then there is no doubt that (s)he can be described as a German, but this is not so obvious in the case of a person who only exhibits one or two of these features. Rather than contemplating vagueness of meaning as a limitation of language, vagueness is viewed by Erdmann as typical of communication, as some features of meaning that are pertinent in one particular context are canceled out or simply disregarded in others. This idea is consistent with the prototype-based model of the semantic structure of linguistic categories and concepts in cognitive semantics (cf. Section 2.4.1), and particularly with the characteristic known as *non-discreteness*, which implies that the limits of linguistic categories and concepts are only clear-cut in the center, but not in the periphery (Geeraerts 2010: 183).

We can thus see how the historical-philological tradition understands meaning in general, and meaning changes in particular, as cognitive phenomena, in which a usage-based perspective is of utmost importance for semantic description. Nevertheless, despite the widespread acceptance of these ideas by scholars such as Bréal, Paul, and Erdmann, among others, this movement, while usage-based in theory, cannot in practice be characterized as a fully usage-

based approach to semantics due to the lack of systematic semantic analyses of concrete texts. In fact, on many occasions, cases of semantic change are examined and explained without explicit reference to the actual contexts in which they occur.

This is not the only shortcoming of the historical-philological tradition. Another crucial one, especially for our purposes in this dissertation, is that since this school takes a primarily semasiological stance, it does not pay enough attention to the onomasiological perspective of semantic change. This is apparent in most classifications of semantic change postulated in this tradition, including the most comprehensive ones, such as Carnoy's (1927) and Stern's (1931), as semasiological types of changes clearly predominate. While it is true that onomasiological changes were recognized as one of the major types of semantic developments in most classifications (cf. Figure 1 above), the work and discussions devoted to these mechanisms are minimal if compared to semasiological ones.

Moreover, research on semantic change in historical-philological semantics involves mainly those changes undergone by individual words, and thus generally disregards the effects these may have on neighboring words in the lexicon and, in consequence, how they influence the overall structure of a broader semantic category or domain. More importantly, the division between semasiological and onomasiological processes is sometimes not easily perceived, given that the latter can be said also to cover the former. In fact, as Geeraerts (2010: 30) suggests, "[…] the semasiological extension of the range of meanings of an existing word is itself one of the major mechanisms of onomasiological change." In other words, when a lexical item acquires a new sense through any of the semasiological processes mentioned above, it frequently becomes associated with a (set of) word(s) with which it comes to share that particular sense. This is one of the reasons why a comprehensive semantic description should include both semasiological and onomasiological processes, among other aspects (cf., for instance, Levshina 2011: 13–14).

In spite of these weaknesses, historical-philological semantics has greatly contributed to many of the most important principles and assumptions which constitute the foundations of cognitive semantics, and it is in this sense that we can certainly regard the latter tradition as either a continuation of, or a return to, the main ideas of the prestructuralist semantic movement.

2.2 STRUCTURALIST AND GENERATIVE SEMANTICS

In this section, both structuralist and generativist assumptions about meaning are discussed, since, as mentioned earlier, many scholars in both movements focus on the formalization of meaning. The first part is devoted to structuralist approaches, with special emphasis on Trier's (1931; 1934) lexical field theory and Lyons' (1963; 1968; 1977) relational model, which are both crucial for our purposes. Lexical field theory and its developments aids in the understanding of the semantic classification of modified nouns postulated in this dissertation (cf. Section 4.3.1.2) while the relational model introduces the systematic study of semantic relations within the lexicon, including synonymy, among others. Then, a bird's-eye view of generative semantics is presented, mainly in order to draw attention to those ideas brought to the fore by scholars such as Lakoff, Fillmore, and Langacker, who exerted an important influence on later developments in cognitive semantics, and who became the founding fathers of this theoretical approach of language.

Structuralist semantics is a branch of structuralist linguistics, the latter being identified mainly with Saussure's (1916) view of language as constituting an independent symbolic system that thus deserves to be studied separately from other systems, including further cognitive abilities of the language user. This tradition dominated the arena of semantic research from around 1930 to 1960, and its emergence can be contemplated as a direct reaction against many of the main assumptions of the prestructuralist semantic tradition introduced in Section 2.1. First, it takes a primarily synchronic approach, as the system that language constitutes is considered to be at work at a specific point in time. Therefore, diachronic research is only secondary and plays a rather insignificant role in semantic description according to this movement.[4] Second, the orientation adopted is an onomasiological one, since the focus is on the relations of concepts in the semantic system, that is, how signs within the system relate to and influence one another. This is one of the most important tenets of this school for the purposes of this dissertation, which has the phenomenon of synonymy as object of study. Structuralists believed that there existed a need to move away from the excessive emphasis placed in historical-philological semantics on the individual sign (i.e. concrete words or lexical

[4] However, some of the most important results of structuralist research have indirect diachronic implications. A case in point is Trier's (1934) study of the development of the field of KNOWLEDGE— involving the German terms *Wîsheit*, *Kunst*, *List*, and *Wizzen*—from the 13th century to the 14th century and the structural changes identified within this particular semantic field (for more details, cf. Lehrer 1974: 15–16; Geeraerts 2010: 54–56).

items), and instead pay special attention to the existing links between the various signs in the system. This is so because the semantics of a particular word cannot be adequately understood or described without considering its association to other words, especially to those with which it holds a close connection. Consequently, related words such as synonyms lie at the basis of the creation and demarcation of meanings and are, in turn, fundamental to the organization of the lexicon, since they have a great effect on the structure of lexical knowledge. As claimed by Saussure (1916: 114),

> [i]n a given language, all the words which express neighbouring ideas help define one another's meaning. Each of a set of synonyms […] has its particular value only because they stand in contrast to one another. […] So, the value of any given word is determined by what other words there are in that particular area of the vocabulary.

Fundamental to the onomasiological perspective just described is the emphasis on *paradigmatic* relations of most structuralists —as opposed to *syntagmatic* ones, which lie at the core of the distributional corpus-based methodology described in Section 2.3. In semantics, the term *paradigm* refers to a set of signs in the system that belong to a particular category or domain and, consequently, shares at least some, but typically many, of their semantic features (e.g. Crystal 1985: 220; Murphy 2003: 8).[5] Contrarily, syntagmatic relations are those which exist between two or more elements which occur simultaneously in the same linguistic unit, be it a phrase, a clause, a sentence, or a larger discourse unit (e.g. Trask 1993: 273; Kay & Allan 2015: 30).

Due to the Saussurean view of language as an autonomous system with its own rules and principles, in structuralism meaning is conceptualized as a non-psychological entity: it is purely a linguistic phenomenon and, therefore, semantic description should be limited to features internal to this system. In other words, there is no need for the linguist to resort to encyclopedic knowledge when trying to describe the meaning of words, as this emerges exclusively from the relations signs hold with other forms in the system. As will become clear, this rather restrictive view of semantic description is not devoid of problems and is, to say the least, highly controversial, especially because it raises the question of where to draw the dividing line

[5] Besides semantic paradigms, other types of paradigms can also be found in language, such as morphological paradigms, which are constituted by all the possible word forms of an item (for more details, cf. Trask 1993: 197).

between semantic and encyclopedic knowledge. In fact, as we will see below when discussing the most significant drawbacks of this movement, it could be argued that many of the descriptions of meaning postulated by structuralists are actually encyclopedic or extralinguistic in nature. A more or less direct consequence of this restrictive stance is the fact that language usage is generally disregarded in semantic description. Once again, this is due to the fact that the object of study in this tradition is the language system in isolation (i.e. *langue* and *competence* in the parlance of Saussure 1916 and Chomsky 1965, respectively), as opposed to the context in which linguistic signs are actually used. Therefore, occasional meanings such as those emphasized by Paul (1920; cf. Section 2.1) are here considered to lie beyond the scope of semantics.

This radically onomasiological, paradigmatical, and systemic orientation to meaning lies at the core of the three major branches of structuralist semantics that are typically distinguished, to wit, lexical field theory (cf., for instance, Trier 1931; 1934; Lehrer 1974), relational semantics (cf., for instance, Lyons 1963; 1968; 1977), and componential analysis (cf., for instance, Goodenough 1956; Coseriu 1964; Pottier 1964; 1965). Here, the focus is mostly on lexical field theory and relational semantics, which are more central for our purposes. However, a brief explanation of componential analysis is also provided.

Lexical field theory was first postulated by Trier (1931; 1934) and later refined by scholars such as Lehrer (1969; 1974), among others. The basic assumption of this theory is that the vocabulary of language can be divided into different areas, which Trier comes to call *lexical fields*, each of which contains semantically interconnected words or concepts that represent a specific domain of reality.[6] In other words, a semantic field, focused around a particular theme, constitutes a subgroup of the lexicon. The organization of semantic fields depends on the relations that the items making up a field hold to one another.

Whereas Trier's conception of fields is a paradigmatic, closed, and clearly delineated one, subsequent conceptions of field theory were critical of this perspective and improved several of its aspects. First, Trier only contemplates lexical items which exhibit paradigmatic relations

[6] Lexical field is not the only term employed to refer to this concept in semantics. Alternative labels include *word field*, *semantic field*, and *conceptual field*, depending on what these subsets of the lexicon are thought to comprise, be it words (as opposed to multiword lexical items), word senses (as opposed to words), and unlexicalized concepts (as opposed to only lexicalized concepts), respectively. For more information on this terminological fluctuation, see Murphy (2003: 92–93) and Geeraerts (2010: 56–57). Here, the term semantic field is used, given that it is the most habitual one in contemporary semantics and because fields are assumed to include specific senses of lexical items.

of similarity as constituting a field, thus disregarding syntagmatic relations. However, scholars such as Porzig (1934) or Harris (1954) point out that the latter type of relations are also important for the identification of meaning and that the combinatorial aspects of words should therefore be considered for inclusion in semantic fields. This emphasis on the importance of syntagmatic relations highlighted by Porzig and Harris, despite playing only a minor role in field theory, clearly invokes a distributional orientation towards meaning, similar to the assumptions of the corpus-based distributional approach (cf. Section 2.3). Nevertheless, only pure syntactic relations and combinations are referred to as being important here, whereas other types of co-occurrence, such as semantic or stylistic ones, are left unaccounted for.

Second, Trier seems to represent semantic fields as closed systems in which all possible concepts within a certain domain have already been lexicalized. However, Lehrer (1969; 1974), in her study of cooking terms (e.g. the verbs *bake*, *fry*, *roast*, and *steam*) demonstrates that this is in fact not the case, since lexical gaps are found. She illustrates her claim by highlighting that no verbs exist in English to denote the plausible notions 'elaborating food by means of a pan without using water or oil' or 'elaborating food by means of radiated warmth (i.e. fire) with oil (as opposed to water or vapor)'.

More importantly, Trier's original view is controversial because it seems to imply that both words within semantic fields and semantic fields themselves are clearly separated from one another. However, this is not often the case, as semantic fields can instead be said to exist on a semantic continuum. Here, Gipper's (1959) study on the German nouns *Stuhl* 'chair' and *Sessel* 'armchair or easy chair' comes in handy to illustrate the diffuse or fuzzy nature of concepts and fields, and as will become evident, clearly echoes Erdmann's (1910) idea of semantic vagueness introduced in Section 2.1. Gipper's analysis, an experimental naming task into the internal semantic structure of the two neighboring items *Stuhl* and *Sessel*, shows how it is not always possible to distinguish between the two words as their meanings partly coincide. This is demonstrated by the fact that even though in many cases *Stuhl* and *Sessel* refer to different types of 'individual seats with a back', thus generally not serving as substitutes for one another, on other occasions *Stuhl* and *Sessel* can be used interchangeably to refer to the same type of seat. Put differently, some items denoted by the concept INDIVIDUAL SEAT WITH A BACK can be named by any of the two terms. For this reason, Gipper postulates a structure of *Stuhl* and *Sessel* with central examples of each of the two neighboring concepts (i.e. those for which only one of the two terms apply) and peripheral examples with those sets of items which could be named by

either *Stuhl* or *Sessel* without any of the two prevailing over the other. The organization of fields, according to Gipper, is consequently based on central or focal areas, with the most clear-cut examples of a given field, and peripheral ones, where the borderline between related fields is not as obvious. Another important point made by Gipper is that the naming process is not based on random choice or personal whim, but rather displays a structured system. In the case of *Stuhl* and *Sessel*, for instance, the principal function of the seat to be named plays a major role, with 'comfort' being associated with *Sessel* and 'practicality' with *Stuhl*. Gipper's conclusion about the fuzziness of field structure, just as in the case of Erdmann's conception of semantic vagueness, is in agreement with the prototypical organization of categories which is later postulated within the theoretical framework of cognitive semantics (cf. Section 2.4.1).

Componential analysis, which evolved somewhat later than Trier's (1931; 1934) original field theory and can be regarded as a further development of it, focuses primarily on the formalization of the existing relations between the interconnected items in a semantic field, with particular emphasis on the oppositions between neighboring words (e.g. Coseriu 1964). As such, studies within this approach are devoted to the ways in which these relations and contrasts should be given formal expression. Alternative descriptive theories have been put forward to account for conceptual or semantic substance (cf., for instance, Nida 1951; Goodenough 1956; Pottier 1964; 1965), but essentially they all proceed by dividing the semantics of lexical items into smaller units called semantic components. This is done with the intention of trying to establish a fixed number of characteristics or *primitives* that could be used to distinguish between related lexical items. An example would be to decompose the words *man* and *woman* into the components [+HUMAN +MALE +ADULT] and [+HUMAN -MALE +ADULT], respectively, and to establish that the difference between them lies in their position regarding the component MALE (Kay & Allan 2015: 33–34). Given that the formalization of semantic content lies beyond the scope of this dissertation, componential analysis is not further described here.

The third branch of the structuralist movement especially relevant for our purposes is relational semantics, which is primarily associated with the work of Lyons (1963; 1977), although in-depth evaluative reviews of this approach are also provided in Cruse (1986) and Murphy (2003). Relational semantics, which builds on Trier's (1931; 1934) field theory and componential analysis, is devoted to the description of the structural connections that hold between neighboring words within semantic fields. According to Lyons, these relations are

more intricate than those of mere opposition present in compositional analysis and the work of scholars such as Coseriu (1962; 1964). Consequently, a more all-encompassing apparatus is needed to adequately explain the structure of language, particularly one in which a clear separation is made between the autonomous linguistic level and the extralinguistic, encyclopedic one. Lyons therefore postulates a broader model of meaning relations in which the structure of language is comprised by a complex network of multiple types of semantic links that connect word senses to one another. As such, it brings to the fore the systematic study of a wider range of lexical —or perhaps more accurately termed sense- or semantic— relations than just antonymy (i.e. semantic oppositeness), which are deemed equally relevant, including hyponymy, meronymy, and, more crucially for this dissertation, synonymy. To this purpose, much of the work within relational semantics is devoted to the identification and description of these four semantic relations, among others, as well as the different subtypes established within them: for instance, gradable as opposed to non-gradable antonymy or total as opposed to partial synonymy. For our goals, it is here enough to provide a brief definition of the various semantic relations distinguished without a detailed consideration of each of them. The different types of synonymy established principally in the works of Lyons (1963; 1968), Cruse (1986), and Murphy (2003) are further discussed in Chapter 3, which is devoted to synonymy in general and to the existing distributional corpus-based studies on the internal semantic structure of specific pairs and groups of synonyms in particular.

Starting with antonymy, this semantic relation is best described as an umbrella term for various and slightly diverging kinds of oppositions of meaning, such as those linking together terms which are (i) placed on opposite ends of a gradable cline (e.g. *good* vs. *bad*; *young* vs. *old*) or (ii) mutually and logically exclusive in the sense that they contradict one another rather than exist on a continuous cline (e.g. *mortal* vs. *immortal*; *single* vs. *married*).

Second, hyponymy is typically defined in terms of semantic inclusion which couples a more general with a more specific term—or rather concrete senses of the terms in question—belonging to the same semantic field. As an example consider the terms *tulip* and *flower*. In this case, the former term (i.e. *tulip*), typically called hyponym or subordinate, is included in the definition or description of the latter (i.e. *flower*), which is called hypernym or superordinate. Other lexical items which are also hyponyms of *flower* (e.g. *gardenia* and *forget-me-not*) are known as co-hyponyms of *tulip*, since they are situated at the same level within the hierarchy.

Contrasting with hyponymy, we find the relation of meronymy, a part-to-whole or whole-to-part relation such as that existing between the terms *eye* and *face* (i.e. a component part of a material whole) or between *captain* and *crew* (i.e. member meronymy). Although these relations can be conceptualized as representing inclusion, they are to be distinguished from hyponymy since the types of inclusive relations that hold in the two cases differ. To further elucidate this distinction, consider again the example of *flower* introduced above. As already mentioned, this term holds a relation of hyponymy with *tulip*, as the latter is a particular type of flower. In contrast, the relation between *flower* and *petal* is a meronymic one, as a petal is an integral element of a flower, thus designating a specific part of it.

Finally, synonymy is sometimes thought of, in layman's terms, as a relation of identity of meaning between two or more words or word senses. However, a somewhat less strict definition of synonymy is typically given in the specialized literature since, as will become clear in the course of Chapter 3, synonyms do not often have identical meaning and can thus not always substitute one another. If two words or word senses can always be used interchangeably, we have a case of total or absolute synonymy (as opposed to cognitive or near-synonymy). Moreover, a distinction is made in the specialized literature depending on whether the terms are semantically equivalent in all their senses or only in one or some of them. Near-synonyms, on the other hand, even though equivalent or even identical on the denotational dimension of meaning, often differ in other aspects of meaning such as connotation, style, or collocation, which prevents them from being used interchangeably in all contexts.[7]

The strong emphasis of componential analysis on formalization is also a prominent feature of generative semantics, which came after structuralism, that is, from around the 1960s onwards. In fact, componential analyses were frequent also within this movement, for instance, in the work of Katz & Fodor (1963; see also Katz 1972). However, despite the common ground with structuralism, the conception of meaning changed yet again in generativism, with meaning being once more conceived of as having a psychological nature. In other words, a mental orientation towards the study of meaning is now brought to the fore again. Two approaches to semantics can be distinguished within generativism, namely interpretative semantics and generative semantics. In interpretative semantics, semantics has a secondary role, since emphasis lies mainly on syntax. Issues of semantics considered important within this branch are in fact related to syntax, for instance, argument structure or quantification, which aid in

[7] Further details about the different types of synonymy and dimensions of meaning are provided in Section 3.1.

syntactic analysis. On the contrary, lexical semantics is completely disregarded. This interpretation of semantics is the one held by the majority of scholars in this tradition (Geeraerts 2006a: 143–144). Contrariwise, in the second branch of generative semantics, the goal is primarily to establish a semantically-based syntax.[8] This approach was favored by scholars such as Lakoff (e.g. 1970; 1971), Fillmore (e.g. 1968), and Langacker, who later became important figures in cognitive linguistics. One of their core assumptions concerns the impossibility of drawing a clear line between semantic and encyclopedic knowledge, a distinction which is rejected in cognitive semantics, as we will see in Section 2.4.

The major contribution of structuralist semantics has been to provide a theoretical framework to understand the relations between neighboring words in the language system, that is, onomasiology, which although not entirely lacking in historical-philological semantics, had not received sufficient attention before. However, several problems exist regarding how onomasiological structure is perceived and studied by structuralists. First, the emphasis put on semantic relations leads to an underestimation of the importance of semasiological processes. For instance, whereas Lyons clearly highlights that the semantic relations identified and established within the relational approach hold between senses rather than between words, almost no attention is paid to sense discrimination of particular words. Moreover, the strict division between a supposedly independent semantic level and an encyclopedic one emphasized in structuralist semantics cannot be maintained. Gipper's (1959) study on *Stuhl* and *Sessel* discussed above clearly demonstrates the need for extralinguistic criteria to adequately describe the differences between related concepts, since the factor 'functionality' (i.e. practicality vs. comfort), which is an aspect that relates to the actual referent of the concept INDIVIDUAL SEAT WITH A BACK, is crucial in order to explain the alternation between these two words. This is just one instance in which the encyclopedic level permeates what are supposed to be purely semantic descriptions.[9] Finally, and more importantly for our purposes, the excessive focus on the

[8] The heated debates and differences of opinion between the two strands of generativism, i.e. generative semantics vs. interpretative semantics, concerning the demarcation between grammar and semantics is often referred to as the "linguistics wars," which dominated part of the linguistic scene during the 1960s (cf., for instance, Harris 1993; Geeraerts & Cuyckens 2007: 12).

[9] In a recent study carried out within the framework of cognitive semantics, Levshina (2015: 373–384) shows that further factors related to the encyclopedic level of knowledge apart from 'functionality', such as 'material' (e.g. fabric, leather, and wood), 'texture' (e.g. soft vs. not soft), and 'seat height' (i.e. high, low, or adjustable), to mention a few, also play a major role in the choice between *Sessel* and *Stuhl*.

language system and the general lack of concern that this onomasiological orientation towards meaning shows for actual usage and language context is highly problematic. This creates major complications when it comes to explaining the existence of alternative forms such as synonyms, since it leads to descriptions which are too restrictive, as they merely focus on the identification and classification of types of semantic relations. However, for an in-depth analysis of the relations of particular synonym sets, it is certainly not sufficient to state that such relations exist between the members of the set, but an analysis of the multiple factors determining the choice between such expressions is also required. In other words, even more interesting than knowing about synonymous relations and the impact they have on the structure of language is knowing in which contexts we use one word instead of another and the reasons why we do so. This change of perspective, however, involves usage considerations and required a shift from a systemic onomasiological orientation, such as that postulated by structuralists, to a usage-based onomasiological orientation, which emerges in later traditions of semantics (cf., for instance, Grondelaers, Speelman & Geeraerts 2007). The lack of examination of the motivations underlying lexical variation is one of the main causes for the abandonment of the structuralist view of onomasiology by current scholars in the field.

With regard to the contributions of generativism, it is enough for our purposes to mention that since the major force within this framework was interpretative semantics, the study of meaning was relegated to a very minor topic, secondary to syntax. Nonetheless, one enduring and significant impact of generative semantics on the development of the field of lexical semantics is the postulation of several assumptions which lie at the core of contemporary theories, first put forward by scholars such as Lakoff, Fillmore, and Langacker, who rejected the division between an autonomous semantic level of knowledge and an encyclopedic, extralinguistic one.

2.3 NEOSTRUCTURALIST SEMANTICS

The fourth major school to the study of meaning, typically known as neostructuralist semantics, comprises a series of approaches which are possibly of a more heterogenous nature than those in the traditions discussed so far. This is because, as a matter of fact, these approaches do not adhere to any particular theoretical framework, but are instead grouped under the label of neostructuralism due to their affinity with the structuralist approaches and conceptions of meaning introduced in Section 2.2. As the prefix *neo-* suggests, they all can, in one way or

another, be contemplated as a renewed or at least modified form of the structuralist methodologies. At the same time, several of the neostructuralist approaches to semantics build on the generativist idea mentioned at the end of Section 2.2, according to which the need to incorporate the psychological nature of meaning into semantic description is again revived. As such, two broad groups are typically distinguished among the neostructuralist approaches, to wit, those with a compositional orientation and those with a relational orientation. A brief overview of the former (i.e. compositional approaches) is provided in Section 2.3.1, which examines the problems posed by such approaches for diachronic studies due to their failure to account adequately for semantic change. In turn, relational approaches are the focus of Section 2.3.2. Within the relational orientation, the distributional corpus-based analysis is discussed in more length, since it provides a methodological basis for the analyses in the practical part of the present dissertation (cf. Chapters 5–7).

2.3.1 The compositional orientation

Among the approaches taking a compositional orientation we find those that continue with the formalization of meaning set out in structuralism and generativism (cf. Section 2.2), but which explicitly do so by adopting a cognitive orientation to semantics in the sense of admitting that encyclopedic knowledge and contextual flexibility of meaning somehow need to be borne in mind. This, however, does not mean that scholars advocating such approaches abandon the restrictive tendencies of earlier types of compositional analyses, but rather that they try to find ways to accommodate this reduction while at the same time accounting for the fuzziness of meaning and its cognitive reality. Four well-known theories adopting this stance are Wierbicka's natural semantic metalanguage (e.g. 1972; 1996; 1997; cf. also Goddard & Wierzbicka 2002), Jackendoff's conceptual semantics (e.g. 1990; 1996; 2006), Bierwisch's two-level semantics (1983a; 1983b; cf. also Bierwisch & Lang 1989; Lang 1993), and Pustojevsky's generative lexicon (1995). Despite openly acknowledging the existence of different levels of knowledge and the importance of accounting for this multi-level reality, all four models arrive at the same conclusion, namely that the fuzziness, vagueness, and flexibility of meaning are due to the extralinguistic or encyclopedic level. In sum, what they imply is that whereas the semantic level is tidy and neat, the encyclopedic and contextual one is vague and fuzzy. Such a conclusion is what permits these models to keep semantic descriptions and definitions at a manageable size, by focusing only on those aspects of meaning of a word or

concept which are invariable —as opposed to those aspects which fluctuate— in the case of Wierzbicka, or only on those features of meaning which are supposed to be stored in the mental lexicon —as opposed to those features that can be contextually derived— in the case of Bierwisch. As a consequence, we can clearly see a parallelism with the problems mentioned above in relation to structuralist approaches (cf. Section 2.2), since the same limitations are applicable also in these cases. First, the research interest again lies not on the explanation of the inherent variability of concepts and categories of meaning, which is the most problematic to account for, but instead on the more stable characteristics of meaning. Second, these compositional approaches face the same difficulties in arriving at sufficiently general and satisfactory definitions that are able to encompass all members of a given concept or category while at the same time adequately distinguishing neighboring concepts and categories from one another.

Another point to highlight and which is possibly more important for our purposes is the disadvantages such approaches pose when employing a diachronic perspective to the study of meaning, especially when semantic changes taking place over time are accounted for. This is so because drawing such a clear division between the two levels of knowledge (i.e. semantic and encyclopedic), and limiting the study to what is invariant or stored in the lexicon, consequently disregarding contextually derived meanings in semantic descriptions, is not feasible when dealing with diachronic data, as what is considered an invariable feature of meaning today may in fact become a variable one in the future, or vice versa. This idea of a close connection —rather than a strict separation— between variable and invariable or between stored and derived aspects of meaning, which would allow us to adequately explain semantic changes, was already present in the prestructuralist approach discussed in Section 2.1, especially in Paul's (1920) notions of usual and occasional meanings (cf. examples (6) and (7) above). This conception of meaning in which the two levels of knowledge are brought together lies at the heart of cognitive semantics, which finds a much more suitable way of dealing with and explaining semantic changes by drawing on empirical data rather than introspection, as we will see in Section 2.4. However, before turning to this full-fledged psychological theory of meaning, we will have a look at how the relational approach postulated by Lyons (1963; 1968; 1977) within structuralist semantics (cf. Section 2.2) was further refined by neostructuralist scholars, especially within the method known as distributional corpus-based analysis.

2.3.2 The relational orientation

The second type of approach distinguished within neostructuralist semantics is said to develop the systematic description of sense relations first initiated by Lyons (1963; 1968; 1977). As in the case of the approaches building on the compositional method discussed in Section 2.3.1, we can distinguish two different branches of these more up-to-date relational traditions. On the one hand, we find two approaches with a strong focus on the formalization and the codification of semantic knowledge into a formal language, namely *WordNet* (cf., for instance, Miller 1990; Fellbaum 1998; 2005) and Mel'čuk's meaning-text theory (e.g. 1988; 1996; 1998). Due to their emphasis on formalization, these two approaches indeed share some common ground with the four neostructuralist traditions of componential analysis referred to above (i.e. natural semantic metalanguage, conceptual semantics, two-level semantics, and generative lexicon), but at the same time they go a step further by introducing computational devices to formalize meaning. *WordNet* is an online lexical database in which words are structured on the basis of the semantic relations they maintain with other words, including the traditional relations mentioned in Section 2.3, to wit, antonymy, hyponymy, meronymy, and synonymy.[10] As such, a search in this database of a particular lexical item yields a whole range of examples of other words with which it is connected in its different senses through any of the aforementioned relations. The meaning-text theory, put into practice in the *Explanatory Combinatorial Dictionary* (Mel'čuk, Clas & Arbatchewsky-Jumarie 1984), also links words which hold any of the four paradigmatic sense relations, but it includes additional types of relations such as grammatical and morphological ones. Moreover, syntagmatic relations of co-occurrence are also contemplated. In this way, the meaning-text theory moves further away from Lyons' original relational approach, whereas WordNet adopts a more classical perspective. As already mentioned, given that formalization is not an important issue in this dissertation, these two approaches are not discussed any further.

The second type of approach distinguished within the neostructuralist relational orientation consists in a full-blown usage-based distributional methodology which, instead of emphasizing the formalization of meaning, pays particular attention to the recurrent patterns of words. This is done by closely analyzing how lexical items are used in actual language and by examining different types of co-occurrence information, thus embracing a syntagmatic perspective. Although syntagmatic relations were not totally absent from the structuralist and the

[10] For more details about this database, see https://wordnet.princeton.edu/.

generativist views of meaning, comprehensive research adopting a full-fledged syntagmatic orientation only arrived with the advent of neostructuralist semantics in the form of the distributional approach advocated by scholars such as Levin (1993) and Sinclair (1966; 1987), among others.[11] However, we can distinguish various ways of implementing the distributionalist view in actual semantic research. The first form, which is of less relevance for our purposes, is embraced by Levin in her 1993 study of English verbs, in which she arrives at a classification of verbs depending on their syntactic patterns. In particular, she analyzes the patterns of alternations in which verbs can appear, demonstrating that these patterns are recurrent for a whole range of individual items. As we can see, although Levin's approach indeed assumes a close link between syntax and lexis, this type of distributional research mainly takes a syntactic perspective.

The second form of distributionalism is that set out by Sinclair (1966; 1987; 1991a; 2004) and later described and applied by a wide number of contemporary scholars working in the field of lexical semantics, including the pivotal works by Stubbs (1993; 2002a), Biber et al. (1998), and Partington (1998). Even though Sinclair is considered to be the pioneer of this approach, its main principles are based on ideas already present in earlier work, including that by Wittgenstein (1953), who asserts that meaning is equal to actual usage, and, more importantly, by Firth (e.g. 1935; 1957a). Firth was one of the first scholars to suggest that distributional patterns and, in particular, *collocation*, which are dealt with in detail below, are essential for the analysis of lexical semantics. Many of Firth's statements on the importance of contextual relations still enjoy wide currency in contemporary lexical semantic research. In two of his most cited works, he claims that "[…] the complete meaning of a word is always contextual […]" (1935: 37) and that language users "shall know a word by the company it keeps" (1957a: 11). In other words, the semantic and functional features of lexical items closely correlate with their distributional patterns. However, the methodology invoked by Firth in such claims did not start gaining much importance until the 1960s. Sinclair (1966) extends Firth's ideas by stating that the principal duty of researchers devoted to lexico-semantic analysis is to describe the collocational tendencies of lexical items by means of the examination of large collections of natural language data, i.e. corpora, which were starting to be compiled at that time. In fact, the

[11] Cf., for instance, the discussion of the work conducted by Porzig (1934) and Harris (1954) in Section 2.2 and Mel'čuk's meaning-text theory in the present section.

most important breakthroughs of the distributional method spring from its implementation in corpus analysis, which Sinclair vigorously advocates.

The tradition originating in Firth's and Sinclair's work is often said to rest on three major tenets, each of which is discussed individually below in further depth: (i) a radical usage-based orientation, (ii) a paramount importance of collocation, and (iii) a strong link with statistical analysis (cf. Geeraerts 2010: 167).

The first of these three principles is clearly reflected in both of Firth's claims cited above. The emphasis on a usage-based perspective in order to reach an adequate description of meaning remains a fundamental doctrine. The usage-based orientation was first highlighted by Firth in his explicit rejection of the introspective methods defended by both structuralists and generativists. In fact, he completely dismisses the analyses of invented sentences which are presented in isolation and states that it is unrealistic to examine meaning without access to the whole context in which these sentences occur if one wants to achieve serious and significant results. Despite his valuable critiques of the limitations of introspective data, Firth's own analyses are not based on much actual text, mainly because of the lack of resources to conduct such systematic research of usage patterns that corpus data would have allowed (cf. Stubbs 1993: 8–9). In this sense, much of his research suffers from the same limitations as that of historical-philological semantics (cf. Section 2.1), which prevents it from becoming a fully developed usage-based approach when put into practice. Consequently, even though Firth's conception of the study of meaning clearly reflects, in theory, the principles which later came to lie at the core of the distributional corpus-based theory, it is not until the arrival of Sinclair's work that the actual reappraisal of the Saussurean and Chomskyan dichotomies, that is, langue vs. parole and competence vs. performance, takes place. Firth, Sinclair, and other scholars such as Halliday instead opt for a view of language in which langue/competence and parole/performance are closely connected and truly interdependent, as they affect one another. Here, the weather and climate comparison provided by Halliday (1991: 33–34; 1992: 66–67) serves as a clear depiction of the interrelation of system and use —or *instance*, in Halliday's terminology (1991: 66):

> [T]here is language as a system, an abstract potential, and there are spoken and written texts, which are instances of language in use. But the "system" and the "instance" are not two distinct phenomena. There is only one phenomenon here, the phenomenon of language […]. If I may use again the analogy drawn from the weather: the instance-

> observer is the weather-man, whose texts are the day-to-day weather patterns displaying variations […], all of which can be observed, recorded and measured. The system-observer is the climatologist, who models the potential of a given climatic zone in terms of overall probabilities. What appears to the former as a long-term weather pattern becomes for the latter a defined climatic system […]. But one thing is clear: the more weather we observe, as instance-watchers, the better we shall perform as system-watchers when we turn to explaining the climate.

In sum, the system (i.e. climate in this metaphor) and the instance (i.e. weather) are not separate phenomena, but simply different aspects of one and the same phenomenon, and each individual instance has a direct effect, however small it may be, on the system of which it forms part. It is in this sense that the system is considered to be probabilistic in nature —as opposed to categorical— by scholars embracing this perspective of language in general, and of semantics in particular. This view of thinking has important methodological implications, since it implies a shift in focus from untypical data, as that gathered from introspective methods as well as those approaches based on limited corpus data, to typical and representative data. What is of most interest now, and becomes feasible with the advent of corpora, is the analysis of central and recurrent usage patterns of particular words. Consequently, the methodology set out by Sinclair makes it possible to corroborate actual claims made about the language system and about meaning. In fact, results of studies employing this radical usage-based methodology came to challenge previous assumptions about both linguistic data and form-meaning associations by demonstrating the untrustworthiness of intuitive judgements about the frequency and distribution of particular words and word senses.

Another significant finding of studies putting the contextual corpus-based approach into practice is the fact that the co-patterning of concrete words and word senses (i.e. patterns of co-occurrence with other particular words or structures) is much more detailed and idiosyncratic than previously assumed in grammars. Additionally, not only do words differ from each other with regard to the neighboring words or to the structures in which they occur, but so do the different forms and readings of a single word on many occasions. One frequently cited example is that mentioned by Sinclair (1991b: 494–496) with regard to the singular and plural forms of nouns, such as *eye* and *eyes*. He asserts that these two word forms of the lemma *eye* do not typically occur in the same structures or with the same words. For instance, whereas plural *eyes* is often used in the 'anatomical organ' sense, being modified by adjectives such as *gray*, *dark*,

or *deep*, the singular form *eye* hardly ever appears in this contextual environment. Results such as these, in which the co-selection of lexical and syntactic features is apparent, demonstrate the interdependence of lexis and morphosyntax. Therefore, they lead to the conclusion that lexical items are often jointly selected rather than singled out individually. Sinclair (cf., e.g., 1987; 1991a; 2004) thus proposes two principles to account for how meaning emerges in actual contexts of use by distinguishing between what he calls the *open-choice principle* and the *idiom-choice principle*. The former sees language production as a series of individual choices that are independent from one another. For instance, when we produce a unit of language, be it a phrase, a clause, a sentence, or a text, there are gaps that need to be filled with individual words that satisfy the grammatical requirements of that particular unit. However, according to Sinclair, this is typically not the way in which we as language users proceed in language production. Instead, we mostly follow the idiom-choice principle, under which "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices" (1991a: 110). This is clearly the case of idioms and phrasal verbs, but also of words that typically co-occur together and thus make up frequent and typical collocations. [12] This leads us to the second tenet of distributional corpus-based analysis mentioned above, namely the paramount importance of collocation.

As has already been mentioned previously, the importance of collocation was first emphasized by Firth during the latter part of the 1950s when he proposed "to bring forward as a technical term, meaning by *collocation*, and apply the test of *collocability*" (1957b: 194). The term collocation has since then been one of the central notions in lexical semantics. In order to discuss research on collocation, it is useful to introduce here some of the basic terminology associated with this concept. Whereas the term collocation refers to the co-occurrence of two or more lexical items that commonly appear juxtaposed in a text or utterance, the individual lexical items making up a collocation are known as *nodes* and *collocates*. The former term designates the lexical item which is the object of study, while the latter relates to the items that appear together with a node (e.g. Sinclair 1991a: 115; Stubbs 2002a: 29–30).

Even though the notion of collocation is often used in lexico-semantic research, it has received various slightly different interpretations over the years, which have led to the existence of a number of distinct definitions in the specialized literature. Here, we emphasize three of the most important types of definitions of the term which are, as we will see, closely interrelated.

[12] For more information on these two principles, cf. also Partington 1998: 19–21.

However, before discussing each of them in more detail, it is necessary to identify the dimensions or parameters along which these interpretations of the notion vary. As summarized in Gries (2013: 138–139), these parameters are the following:[13]

(i) The nature of the linguistic items analyzed, that is, words or more schematic categories such as parts of speech (POS) and constructions.
(ii) The number of items constituting a collocation, ranging from strings of two words to longer sequences.
(iii) The frequency threshold, typically established by means of a statistical association measure such as Pointwise Mutual Information or PMI (Church & Hanks 1990), of an expression co-occurring with a node in order to be considered a collocate.
(iv) The distance between the items making up a collocation, i.e. whether they are directly adjacent, syntactically related, or appear within a context window of *x* words.
(v) The specificity of the lexical items involved, to wit, word forms or entire lemmas.
(vi) The degree of compositionality or predictability of the meaning of the collocation.

These fluctuations in the use of the term collocation lead to three often cited types of definitions which have been labeled *textual*, *statistical*, and *associative* or *psychological* (cf., for instance, Partington 1998: 15–16). Starting with the textual definition, this interpretation views collocation as a syntagmatic relation of co-occurrence of lexical items appearing within a short span of words in any given text (e.g. Sinclair 1991a: 170; Partington 1998: 15; Stubbs 2002a: 24). As such, two words, lemmas, or more complex structures are said to form a collocation when they occur together in a textual unit such as a phrase, a clause, a sentence, or a larger discourse segment. As apparent from parameter (iv) above, a proximity limit is often set to establish which co-occurring items are to be contemplated as collocates of a node. Such a proximity limit typically involves a context window of a given number of words to the left and to the right of the node. For instance, a restriction of four or five words —often calculated in terms of lexical words, thus excluding function words— is a typical measurement and is often formalized as 4:4 and 5:5 or as L4–R4 and L5–R5. A further refinement of the textual interpretation is to use a more specific context window in which only immediately adjacent or

[13] Gries (2013) provides a comprehensive overview of the work done on collocation over the last half century and how research in this field can be improved.

syntactically related lexical items of the node are included. This is frequent in studies that make use of data extracted from POS-tagged or syntactically parsed corpora, which make it possible to search for more informative collocates. To illustrate this, if one is interested in researching adjectival nodes, especially when used in attributive function, it is probably more elucidating to analyze their noun collocates. This is so because, as several studies have demonstrated, looking at the nouns with which adjectives co-occur is one of the best ways to reveal the nature of their semantic content, as they mainly serve as modifiers of nouns. A case in point is Geeraerts' (1986: 282–284) study on the Dutch adjective *vers* 'fresh', which proves that the fine-grained aspects of meaning of polysemous adjectives such as this one can be discovered by examining their head nouns. Geeraerts provides evidence of *vers* having different meanings depending on the nouns it accompanies: when it modifies *wound*, *vers* means 'recent', but when it modifies the noun *air,* it takes on a slightly different meaning, namely 'pure, untainted, or optimal'.

Moving on to the statistical definition of collocation, which is closely linked to parameter (iii) above, this view assumes that only those lexical items that co-occur with a node often enough to be statistically significant, that is, those that occur more often than is expected by chance, should be included in collocational analysis (e.g. Hoey 1991: 7–8). This takes us back to one important aspect mentioned above in relation to the first tenet of the approach (i.e. assuming a radical usage-based orientation), to wit, the paramount importance of identifying the central and recurrent usage patterns of lexical items. As Sinclair notes (1991a: 170), this definition of collocation is perhaps the most widely used in contemporary semantic studies, especially since an ever-increasing quantity of corpus data is nowadays available to scholars working in the field, thus enabling the research community to calculate the association of strength between lexical items through measurements such as PMI, among others. Such means allow researchers to draw detailed conclusions about how lexical items behave in discourse and, more importantly, to explain their particular collocational preferences and idiosyncrasies (e.g. Partington 1998: 16).

The last definition of collocation to be discussed is the associative or psychological one. It is commonly known as the psychological or associative definition because, as native speakers of a specific language, our mental lexicon contains information about which collocations are typical and untypical in particular contexts of use. Put differently, and by making use of Firth's notion of *expectancies*, a language user, via his or her day-to-day experience and interaction

with language, develops clear intuitions about which lexical items go together in actual discourse. This is what Leech (1974: 17) calls *collocative meaning*, which for him constitutes one of the seven major types of meaning.[14] Collocative meaning contributes information regarding how lexical items are used, enabling speakers to correctly combine words and preventing them from making certain combinations. Consequently, even though lexical items in isolation do have a meaning, since speakers indeed have an idea about their semantic content even when they are not used in a particular context, a substantial part of their semantic properties arises from the context in which they are used. In other words, these properties are manifested in the relations they maintain with neighboring words in particular situations (Leith 1983: 73; Cruse 1986: 40–41; Partington 1998: 15–16). For instance, the near-synonyms *pretty* and *handsome*, which both share the conceptual meaning 'good-looking', are said to differ in collocative meaning in accordance with the nouns with which they typically collocate (Leech 1974: 17): while common noun collocates of *pretty* include *girl*, *boy*, *woman*, *flower*, *garden*, *color*, and *village*, those of *handsome* include *boy*, *man*, *car*, *vessel*, *overcoat*, *airliner*, and *typewriter*. Therefore, the collocational preferences of *pretty* and *handsome* clearly differ, even though they do share some collocates, for example, *boy*. Such collocational preferences are not to be confused with selectional restrictions, which are the co-occurrence incompatibilities of items in conceptual terms. To illustrate this difference, consider the selectional restrictions of the verb *die,* exemplified in (8) and (9) (adopted from Cruse 1986: 278):

(8) Arthur died.
(9) The spoon died.

Whereas the subject of *die* in example (8) is conceptually compatible with it, this is not the case in example (9), as only living beings can 'lose life or cease to live' (*OED*, s.v. *die* v.[1] I.1a) in the literal sense of the verb.

Besides differing in collocational preferences, lexical items can also diverge as regards their *collocational range*. This term refers to the amount of types —as opposed to tokens— of collocates that a given lexical item co-occurs with. For instance, the adjective *bad* can modify

[14] The seven major types of meaning are conceptual meaning, connotative meaning, stylistic meaning, social meaning, affective meaning, reflective meaning, and collocative meaning. The types of meaning which are relevant for the purposes of the present dissertation are discussed in more detail in Section 3.1.

a much wider range of nouns than the semantically similar but much more specific adjective *substandard*.

As we have seen, due to the various interpretations it has received in the specialized literature, the notion of collocation is a rather vast one and can be understood in different ways. As such, it also takes several different forms. Besides collocation, Sinclair (1991a; 1996) establishes three additional important levels of co-occurrence, to wit, *colligation*, *semantic preference*, and *semantic prosody*. Colligation refers to the co-occurrence of a particular lexical item with grammatical or function words, thus being opposed to collocation, which is a co-occurrence relation between lexical words.[15] Semantic preferences, in turn, have to do with the semantic features that the collocates of a particular node share with one another. For instance, many of the typical noun collocates of the adjective *large* refer to amounts and sizes, including the lemmas *number*, *scale*, *part*, and *quantities*. Many of the collocates of *large* can therefore be said to belong to the same semantic field or domain (Stubbs 2002a: 65). It is important to note here that two or more lexical items, for instance, synonyms, may have the same semantic preferences, but still differ as regards collocational preferences, that is, as regards the specific collocates with which they occur within a particular semantic category or domain. Finally, semantic prosody relates to the connotations of the typical collocates of a node, that is, whether they express negative or positive overtones. We thus speak about lexical items as having a positive, negative, or neutral semantic prosody depending on the evaluative attitudes conveyed by its neighboring words (cf., e.g., Stubbs 2002a: 64–66; 2002b; McEnery, Xiao & Tono 2006: 83–85; Geeraerts 2010: 170–173).

The third tenet of the distributional corpus-based approach is the importance given to the statistical analysis of data. This principle was already implied in the statistical definition of collocation mentioned above. However, during the latter half of the 20th century and the first two decades of the 21st century, major advancements concerning information technology and statistics have allowed the distributional approach to semantics to progress considerably. This is particularly true when it comes to identifying the important collocates of a given node word. While some scholars use percentages to calculate the collocational strength between co-occurring lexical items (e.g. Stubbs 2002a), so the higher the co-occurrence of a node with a collocate, the stronger the association between them, others apply more sophisticated techniques (e.g. Church et al. 1991; Clear 1993). Association measures such as PMI mentioned

[15] For further discussion on the difference between the two notions, cf. McEnery, Xiao & Tono 2006: 10–11.

earlier allow the quantification of the probability of association between collocates. PMI is a measure of collocational strength that compares the likelihood of two words co-occurring vis-à-vis the likelihood of the individual items (Church & Hanks 1990: 23): a score of zero implies that the words do not typically co-occur, whereas a score of three or higher indicates that they are strongly associated. There exists a variety of association measures to compute collocational strength, including *t*-score, log-likelihood, and ΔP (e.g. Gries 2013), among others, all of them tools to move away from introspective and impressionistic judgements about what constitutes a collocation. Additional statistical techniques have been used within the distributional corpus-baserd approach not only to measure collocational strength, but also to model the semantic dis(similarities) between related lexical items on the basis of their collocational behavior and other distributional patterns, including stylistic and syntactic ones. These include regression analyses (Speelman 2014; Levshina 2015: Chapters 7, 12–13), both linear and logistic models, various clustering techniques, such as hierarchical agglomerative cluster or HAC analysis (e.g. Gries 2009: 316–319) and conditional inference trees (e.g. Levshina 2015: 291–300), as well as semantic vector space (SVS) modeling (e.g. Sahlgren 2006; Heylen et al. 2008; Peirsman, Heylen & Geeraerts 2008; Hilpert & Correia Saavedra 2017). Some of these statistical techniques are used in the semantic analyses provided in Chapters 6 and 7, where they are explained in depth.

The distributional corpus-based approach has been successfully applied not only to lexical semantics, but also to research in fields such as language teaching and learning (cf., for instance, Nattinger & DeCarrico 1992; Nesselhauf 2005; Ellis, Simpson-Vlach & Maynard 2008; Paquot & Granger 2012; Gablasova, Brezina & McEnery 2017) and computational linguistics (cf., e.g., Kilgarriff 2006; Kilgarriff & Rychlý 2010; Schulz & Aziz 2016), among others, especially over the last two decades. Moreover, different linguistic phenomena have been analyzed following this approach, particularly polysemy and synonymy. For instance, by examining the distributional patterns of polysemic words, it is possible to distinguish their different senses (e.g. Church et al. 1991; Biber 1993). More importantly for our purposes, an ever-increasing number of studies have demonstrated the effectiveness of the distributional corpus-based approach to uncover fine-grained semantic and usage distinctions between sets of synonyms. A detailed overview of such research is provided in Sections 3.2 and 3.3, which deal with synonymy in depth. Undoubtedly, one of the major contributions of this radically distributionalist orientation to meaning has been the incorporation of collocation first into

semantic description and later into linguistic research in general. In fact, most software employed nowadays in corpus linguistics includes tools for the statistical analysis of collocational behavior, as is the case of *WordSmith Tools* (Scott 2020). Thus, due to its wide range of applications, the neostructuralist distributional approach has undoubtedly meant a significant improvement from a methodological viewpoint over structuralist theories. However, as argued by Geeraerts (2010: 177–178), the approach is ultimately a method rather than a theory, as it is not always clear how the distributional patterns uncovered relate to theoretical issues in lexical semantics. Recently, distributional methods have been adopted for the study of meaning under the theoretical framework of cognitive semantics, which embeds the study of meaning in the study of human cognition in general. As we will see in the next section (cf. Section 2.4), this theory converges with the original neostructuralist distributional method in many ways, in particular due to their common reliance on actual usage-based data and their quantitative orientation. Given the rapid increase of studies in cognitive semantics with a distributional stance over the last two decades, the importance and utility of this methodology cannot be underrated.

2.4 Cognitive semantics

The fifth and last tradition to be discussed and probably the approach that enjoys the widest currency in contemporary semantic research is that of cognitive semantics, a subfield within the framework of cognitive linguistics. As mentioned earlier (cf. Section 2.2), cognitive linguistics emerged in the 1980s with the work of Lakoff, Fillmore & Langacker, who are considered to be the founding fathers of this school (cf. Geeraerts & Cuyckens 2007: 3). Contrary to previous traditions, particularly generativism, this framework assumes that meaning is of chief importance, given that language is, as made clear in the editorial statement of the number one issue of the journal *Cognitive Linguistics* published in 1990, "an instrument for organizing, processing and conveying information" (quoted in Geeraerts 2006b: 3). Thus, semantic description is brought to the fore in linguistic research. As the name of the approach suggests, and as has been mentioned at several points throughout this chapter, meaning is conceptualized in this tradition as a psychological phenomenon whose study should therefore be integrated into the study of cognition in general. With regard to the oppositions or dichotomies that have characterized the development of the field of semantics and that have served as the guiding theme of the present chapter, to wit, semantic vs. encyclopedic

knowledge, use vs. system, semasiology vs. onomasiology, and synchrony vs. diachrony, cognitive semantics takes a maximalist approach. The maximalist perspective implies that some of these oppositions are completely abolished (namely, semantic vs. encyclopedic knowledge and use vs. system), whereas others are maintained (namely, semasiology vs. onomasiology and synchrony vs. diachrony) but none is given priority over the other. On the one hand, the elimination of the first two oppositions is crucial in order to go beyond the restrictive types of semantic description put forward by approaches in which a clear-cut separation is established between a pure semantic level and an encyclopedic one, and between language structure and language use. On the other, a reevaluation of the latter two oppositions is important to provide a more adequate description of meaning in which the interconnection of semasiological and onomasiological processes, and of the synchronic and diachronic orientations, are systematically examined, thus leading to a more comprehensive picture of meaning and of meaning relations.

This maximalist stance towards semantics brings about four central principles that lie at the core of the conception of meaning in cognitive semantics, which according to Geeraerts (2006b: 3–6), can be summarized as follows:

(i) Linguistic meaning implies perspectivization.
(ii) Linguistic meaning is characterized by its dynamicity and flexibility.
(iii) Linguistic meaning is encyclopedic and therefore not independent from the rest of human cognition.
(iv) Linguistic meaning is experiential in nature and heavily dependent on actual use.

First, given that the principal function of language is to categorize the world and our experiences in it, meaning indubitably involves perspectivization. This is so because, as a vehicle for categorization, language does not merely reproduce reality in an objective way, but rather reflects it from the subjective standpoint of language users. It is in this sense that language is said to "[…] impose a structure on the world" (Geeraerts & Cuyckens 2007: 5) or to "construe the world" (Geeraerts 2006b: 4). The way in which meaning imposes such a structure depends on the experiences, requirements, and desires of language users, which vary from one society to another, from one culture to another, and from one period to another. This leads to the second belief about meaning, which is directly related to the first one, namely its dynamicity and

flexibility. If language —and meaning— is a way of modeling the world, and since the world we live in is ever changing, language and meaning must change as well. As new realities and experiences arise, new categories and concepts need to be created or existing ones need to be modulated to make room for such adjustments. Therefore, language cannot be an inflexible structure, but must instead be flexible and dynamic, characterized by great and constant adaptability. Connected to the first principle, the third tenet implies that meaning necessarily involves encyclopedic knowledge and, therefore, a purely semantic level of knowledge does not exist. This links up with the idea that meaning is a cognitive phenomenon and, as such, it obviously holds a close relation with the rest of human cognitive abilities. Consequently, language in general and meaning in particular do not constitute autonomous entities that have to be analyzed independently of other systems, as is believed in structuralism (cf. Section 2.2). Rather, the study of meaning needs to be embedded into the study of cognition. Finally, because meaning is built upon the experiences of the world of language users, it is undoubtedly experiential in nature, and thus language use is key to understand the language system. The linguistic experiences of speakers do not directly involve abstract syntactic patterns, but instead concrete instantiations of those structures, that is, actual words which are recurrently used in certain patterns. This principle directly echoes the ideas postulated in the neostructuralist distributional corpus-based approach introduced in Section 2.3.2.

These four central tenets constitute the foundation of the major models and theories within cognitive semantics, such as the conceptual theories of metaphor and metonymy, idealized cognitive models, frame theory, prototype theory, and the various usage-based approaches to meaning change. In what follows, only the latter two contributions are discussed, as they are the most relevant for the subsequent analyses in Chapters 5–7.[16] Section 2.4.1 explains the

[16] The first three do not figure prominently in the present dissertation and are therefore only discussed briefly here. The conceptual theories of metaphor and metonymy focus on the interconnection between the different senses of words which are linked by means of the mechanisms of semantic extension of metaphor and metonymy, understood within this theory as cognitive rather than purely linguistic processes. A metaphor is a cognitive process whereby one domain (i.e. target domain) is understood in terms of a different domain (i.e. source domain) due to some resemblance perceived between them. Therefore, a mapping emerges from the source domain to the target domain following the pattern TARGET IS SOURCE, for instance, IDEAS ARE FOOD, as in *I just can't swallow that claim* (Lakoff & Johnson 2003: 43). Metonymy, in turn, is a cognitive process which connects related concepts within the same domain. In this case, we speak about vehicle and target, the former denoting the concept which is used to refer to the latter along the pattern VEHICLE FOR TARGET, for instance, AUTHOR FOR WORK, as in *I'm reading Shakespeare* (Kövecses 2010: 171–172; cf. also Lakoff 1987; 1993 for metaphor and Panther & Radden 1999 for metonymy). Idealized cognitive models refer to the set of assumptions and beliefs of the world, on which linguistic and other

concepts of *prototypicality*, *gradience*, *entrenchment*, and *salience*, while Section 2.4.2 deals with usage-based theories of semantic change, which constitute a continuation or reappraisal of the diachronic research set out in the historical-philological tradition to semantics (cf. Section 2.1).

2.4.1 Prototype theory: Prototypicality, gradience, entrenchment, and salience

As became evident in the introductory paragraphs to this section, in cognitive semantics meaning is essentially a means by which we categorize the world. As such, research concerning the internal structure of categories is of fundamental importance in this framework. One case in point is prototype theory, which since its origin in psycholinguistics in the 1970s in the work of Rosch (cf., for instance, 1973a; 1973b) has become extremely influential also in the fields of linguistics and semantics, where it has been applied to both semasiological and onomasiological structures and phenomena. The aim of this section is not to provide a detailed account of the theory, but merely to introduce some of the basic notions associated with it, to wit, prototypicality, gradience, salience, and entrenchment, concepts which will aid in the explanation of some of the changes undergone by the synonyms object of study in subsequent chapters.[17] However, before explaining these notions, a few words seem in order concerning some of the fundamental assumptions of prototype theory.

Prototype theory emerged and gained importance in the 1970s due to the limitations of the models based on the classical orientation to category structure postulated by Aristotle, which had dominated the field of semantics in the structuralist and the generativist compositional approaches, but also in Trier's (1931; 1934) original field theory (cf. Section 2.2). Such approaches assumed that categories —and the members belonging to them— can be defined by a set of necessary and sufficient conditions, thus reducing category membership of a particular object or experience to an 'all vs. nothing' or 'either vs. or phenomenon'. However, as we have

cognitive processes depend (Geeraerts 2010: 224; cf. also Lakoff 1987). This theory, therefore, emphasizes the strong connection between linguistic and encyclopedic knowledge. Frame semantics shares many of its main assumptions with idealized cognitive models, but goes beyond it by arguing that our interpretation of the world is filtered not only through cognitive models, but also through the different ways in which we give linguistic expression to these models (Fillmore 1977; 1985; Fillmore & Atkins 1992). Here, the notions of perspectivization and construal play a crucial role (cf., for instance, Verhagen 2007).

[17] For comprehensive accounts of prototype theory, see, for instance, Lakoff (1987), Lehmann (1988), and Taylor (1995). For an overview, see Geeraerts (2006a; 2010: 183–203).

seen when discussing Erdmann's (1910) example of GERMAN (cf. Section 2.1) and Gipper's (1959) study of *Stuhl* vs. *Sessel* (cf. Section 2.2), this is hardly ever the case in the real world, where it is frequently difficult to assign a particular object or experience, and consequently the lexical item(s) denoting it, either to category *X* or to category *Y*. Problems of this kind explain the shift from an interest in determining the essential features of members of categories to the establishment of the best or most typical examples of particular categories. This shift in focus, though, implied in conclusions such as those drawn by Erdmann and Gipper, among others, did not receive due attention until the emergence of prototype theory (Taylor 1995: Chapters 2 & 3; Geeraerts 2006b; 2010: 183–184; Kay & Allan 2015: 37–39).

The most essential notion in such a conception of category structure is that of prototypicality, which basically refers to the degree of representativeness of the members of categories, including both the actuals entities in the world as well as the lexical items used to refer to them. Whereas some members are more central and therefore more prototypical, others are more peripheral and therefore less prototypical. In this view, categories are said to exhibit the following four prototypicality effects that can help us explain how categories are organized (cf., for instance, Geeraerts, Grondelaers & Bakema 1994: 44–50; Geeraerts 1997: 11; 2006b: 146–147; 2010: 187–189; Grondelaers, Speelman & Geeraerts 2007: 989–990):

(i) Non-equality or gradience: There is no clear-cut delimitation between neighboring linguistic categories and concepts, but instead a cline with a set of central examples and a set of less central ones, as we saw in the case of *Stuhl* and *Sessel* in Section 2.2. Thus, rather than about absolute membership, we talk about degrees or gradience of membership within categories.

(ii) Family resemblance: Category membership depends on a series of features, but most members do not exhibit all of these features. On the contrary, most exhibit a subset of these features, and therefore share some of them with other members, but rarely coincide in all. For instance, if one category is defined by features *X*, *Y*, and *Z*, then one member may exhibit features *X* and *Y*, another *Y* and *Z*, and yet another *X* and *Z*. Here, all members are somewhat related, but none of them are entirely identical as they do not exhibit the same set of features.

(iii) Non-discreteness: The limits of linguistic categories and concepts are only clear-cut in the center, where the more prototypical members are located, not in the periphery, where membership is often uncertain.

(iv) Absence of necessary and sufficient conditions: No set of necessary and sufficient characteristics holds to define all members of a given category or concept, while at the same time adequately distinguishing that category or concept from neighboring ones.

Related to the notion of prototypicality just described are those of entrenchment and salience. Entrenchment refers to the process whereby a linguistic unit, such as a new word or a new sense of an already existing word, becomes more cognitively established in the grammar or lexicon due to its more frequent occurrence in actual language use. In the words of Langacker (1987: 59),

> [there exists a] continuous scale of entrenchment in cognitive organization. Every use of a structure has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact. With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; moreover, units are variably entrenched depending on the frequency of their occurrence.

Langacker thus equates entrenchment with frequency of use of a lexical item. However, nowadays the concept does not typically refer to the overall frequency of a word, but rather to its frequency with respect to a particular sense or function, especially when alternative choices exist to express that same sense or function (e.g. Schmid 2007: 119). Consequently, entrenchment is particularly relevant when dealing with onomasiological processes such as the relation of synonymy, since members within a synonym set commonly exhibit different degrees of entrenchment, with one or some being used more frequently than others to refer to the concept they denote. The degree of entrenchment of a lexical item in comparison with its alternatives, for which we can also use the term *onomasiological salience*,[18] can be calculated as follows, according to Geeraerts (2010: 202; cf. also Geeraerts 1997: 44–45): "the ratio

[18] Onomasiological salience is not to be confused with *semasiological salience*, which is defined by Grondelaers & Geeraerts (2003: 71) as "the degree of prototypicality of the referent with regard to the semasiological structure of the category." (cf. Chapter 5).

between (a) the frequency with which the members of a category are named with an item that is a unique name for that category, (b) and the total frequency with which the category occurs in a corpus."

Given that categorization processes are experientially based, we can easily observe how features such as prototypicality and entrenchment may vary across both space and time depending on the different experiences of language users. For instance, varying degrees of exposure to objects or situations lead to differences in prototypicality and entrenchment. A case in point is found in the examples mentioned by Levshina (2011: 15) regarding the nouns *kangaroo* and *kiwi* on websites from Australia and New Zealand. Whereas *kangaroo* is 8.5 times more common on the former sites, *kiwi* is 14 times more common on the latter. An example of the influence of time on changes in prototypicality is that of the concept or category SHIP, as its prototypical members nowadays probably differ substantially from those in the Middle Ages, even though the core meaning is still 'vehicle for transportation across masses of water' (Kay & Allan 2015: 39). This example demonstrates how sociocultural changes and technological advances lead to differences in degrees of prototypicality. Such changes and advances often bring about modifications of an existing category, the disappearance of some concepts associated with a given category, or the incorporation of new categories (Blank 1999: 71–73). Given the impact of experience on category structure, research within cognitive semantics has attempted to account for semantic variation and change by means of usage-based corpus approaches that allow the quantification of features such as the prototypicality and entrenchment of specific concepts and categories, as well as the items employed to refer to them (e.g. Geeraerts 1988; Geeraerts, Grondelaers & Bakema 1994; Soares da Silva 2015). This issue, among others related to semantic variation and change, are discussed in the next section.

2.4.2 Usage-based theories of semantic change

As mentioned earlier, both synchronic and diachronic perspectives to the study of meaning abound in cognitive semantics. Consequently, with the increasing popularity of this framework, the interest in semantic change is currently back on the agenda after having been somewhat disregarded in most structuralist and generativist approaches, especially during the 1950s and 1960s. This renewed interest is mainly due to the appeal of issues regarding the relation between cognition and linguistics, the flexibility and dynamicity of meaning as represented in actual usage, and the relation of meaning with society and culture, which, as Kay & Allan (2015: 73)

claim, "offer convincing ways of approaching the 'messiness' and unpredictability of lexical semantic change." Another likely reason for the resurgence of diachronic approaches to semantics is the existence of material such as electronic corpora which comprise a great amount of actual language use from previous historical periods, thus easing the process of studying diachronic developments in language.

Lexico-semantic changes are often considered to be more difficult to categorize than other types of linguistic change in the sense that each word follows idiosyncratic developmental patterns. The reasons for this are twofold. First, semantic changes are possibly caused by a more extensive range of factors than other types of change, including both intra- and extralinguistic determinants (Durkin 2009: 222–223; cf. also Kay & Allan 2015: 72–74). Secondly, extralinguistic factors do in fact play a major role in many meaning changes, which adds more complexity to the task of describing the evolution of particular lexical items. As such, non-linguistic history, including technological and cultural developments, is often resorted to in order to help explain the motivation underlying these linguistic changes. According to Durkin (2009: 222–223):

> Semantic changes are notoriously difficult to classify or systematize […]. [A]lthough some semantic changes occur in clusters, with a change in one word triggering another, we do not find anything comparable to a regular sound change […]. Additionally, semantic change is much more closely connected with change in the external, non-linguistic world, especially with developments in the spheres of culture and technology.

This does not mean, however, that recurrent patterns or mechanisms of semantic change cannot be identified. As we have already seen, comprehensive classifications of semantic change were already proposed in the historical-philological tradition (cf. Section 2.1), and many of these classifications still enjoy wide currency in contemporary semantic research. In fact, diachronic analyses conducted within cognitive semantics echo many of the principles of historical-philological semantics, as the conception of meaning in both frameworks is very similar (i.e. close relation between cognition and meaning, no separation between semantic and encyclopedic knowledge, flexibility and dynamicity of meaning, and importance of context and usage). Therefore, it can safely be maintained that the diachronic branch of cognitive semantics is both a continuation and an extension of the research set out in the prestructuralist theory of semantics (Geeraerts 2010: 229–230).

Language use figures prominently in explanations of semantic change in cognitive semantics. Building on Paul's (1920) distinction between usual and occasional meanings mentioned in Section 2.1, semantic scholars belonging to this tradition highlight that changes in meaning, for instance, the emergence of new senses and nuances of meaning, occur in actual language use. Therefore, they also distinguish between decontextualized and contextualized meanings. In fact, one explanation that has been postulated for the development of new meanings involves the concept of *invited inference* (cf. König & Traugott 1988; Traugott & Dasher 2002). This model argues that many utterances show a certain degree of ambiguity since, besides the more established and central sense of a linguistic expression, additional ones may also arise in a particular context of use. The theory of invited inference builds on Grice's (1989 [1975]) notion of *conversational implicature*, which refers to the meanings that go beyond the conventional meanings of words, but that are implied by speakers when using those words in a specific structure, either consciously or unconsciously. Conversational implicatures therefore arise when participants in a communicative exchange violate or flout the so-called cooperative principle by not adhering to one or more of its *maxims* (i.e. the maxims of quantity, quality, relation, and/or manner).[19] With time and repetition, such inferred meanings —or conversational implicatures in Grice's terminology— may become conventionalized and established as a new sense beside the original one. This is the case of temporal markers such as *since* and *after*, among others, which have developed causal meanings over time (Traugott & Dasher 2002: 16; Hopper & Traugott 2003: 80–81). Such inferred meanings, by being implied repeatedly, need to be stored at some point in the mental lexicon, thus leading to a higher degree of entrenchment and prototypicality. Consequently, there emerges a clear convergence between usage-based approaches to meaning change and prototype theory described above: while conventionalized senses correspond with central and prototypical senses, contextualized and less conventionalized senses correspond with peripheral and less entrenched ones. However, as we have already seen, features such as prototypicality, salience, and entrenchment are not constant, but rather vary across time and space.

However, not all semantic changes imply the appearance of new senses but may instead bring about alterations to the central and prototypical meaning so that its range of application becomes either enlarged or more restricted. Consider the example of Dutch *legging* 'leggings' discussed in Geeraerts (1997: 33–47; 2010: 234). When the form entered the language at the

[19] For an overview of inferential-based approaches, see, for instance, López-Couso (2017).

end of the 1980s it was mainly used to denote a specific type of women's pants, to wit, 'long, tight-fitting stretch pants'. Nevertheless, over the next few years, as more and more different models of leggings saw the light and became increasingly popular, the noun came also to be used to name other similar types of women's pants. These indeed had much in common with the prototypical and central model, but entailed an extension of the range of the term, which was also used to refer to more peripheral and less prototypical instances of pants which were short and tight-fitting, long but not as tight-fitting as the original, or long and tight-fitting but made of fabrics other than stretch material. Thus, we see how more peripheral and less conventionalized types gradually enter the category LEGGING, and over time become more entrenched through the repetition of uses of the term *legging* to refer to such variations of the original model.

The multiple contributions of cognitive semantics, of which only a few have been summarized in the preceding paragraphs, are primarily due to its maximalist stance to meaning, in which the dichotomies established in previous schools are fully neutralized. Therefore, semantic description does not usually suffer from the same limitations as many of the earlier traditions reviewed throughout this chapter. Moreover, the standpoint embraced by cognitive approaches is more balanced, in that it often considers both semasiological and onomasiological processes, as they are thought to be intimately related, and both synchronic and diachronic perspectives to meaning. This maximalist stance to meaning, according to which meaning is considered to be connected with the rest of human cognition, allows for new ways to systematically explain issues which had only exerted a minor influence in preceding schools, such as the apparent fuzziness of concepts and categories and their structure around prototypical centers. Similarly, a close link between meaning and sociocultural aspects, which is also a fundamental principle of cognitive semantics, offers a suitable framework for explaining semantic change by resorting to the study of non-linguistic history.

However, cognitive semantics is certainly not devoid of problems, but the many benefits of the approach demonstrate its great potential for semantic description. New developments are constantly appearing and improving semantic analysis within this tradition. In fact, Geeraerts (2010: 240–266) identifies a set of interesting research areas in need of further refinement. Among these, one particular line of research is related to the analyses provided in the present dissertation, namely usage-based approaches to the study of onomasiological variation (for an overview, see Grondelaers, Speelman & Geeraerts 2007). Such approaches focus on the

identification of factors that underlie the choice between synonymous expressions by employing electronic corpora. Although these factors are commonly of an internal nature (e.g. Divjak 2006; 2010; Divjak & Gries 2008; Liu 2010), many studies also focus on extralinguistic factors, and thus have a more sociolinguistic orientation (e.g. Speelman, Grondelaers & Geeraerts 2003; Levshina 2011; Soares da Silva 2013; 2014). Many of the internal factors studied in research of this type are related to co-occurrence features of the synonyms, including collocational patterns in the form of individual collocations or colligations, but also semantic prosody and wider semantic preferences. This, in turn, points to a close link between this line of research and the neostructuralist distributional corpus-based approach discussed in Section 2.3.2. Nonetheless, as stated earlier, the availability of more refined multivariate statistical methods such as logistic regression, among others, allows for the inclusion of a wider range of factors in one and the same analysis. A great majority of usage-based onomasiological studies focus on morphosyntactic phenomena, thus somewhat disregarding lexical semantics. Additionally, those that do examine lexical synonyms take primarily a synchronic perspective, while the diachronic evolution of synonyms has to date not received enough attention. It is, therefore, at the crossroads of the distributional approach, usage-based onomasiology, lexical semantics, and historical linguistics that the analyses in this dissertation are located.

2.5. SUMMARY

This chapter has provided a bird's-eye view of the five major theoretical frameworks in lexical semantics, as well as an explanation of key notions and concepts in the field. The focus has been on the major issues characterizing each theory, which have been explained on the basis of a series of dichotomies: semantic vs. encyclopedic knowledge, use vs. system, semasiology vs. onomasiology, and synchrony vs. diachrony. As such, the chapter has offered a historical account of the evolution of the field of lexical semantics from the 19th century to the present day, which has served to demonstrate how many aspects that are still highly relevant today were already present, albeit sometimes in a less developed form, in earlier schools of semantic description. We have also seen how many of the main ideas of previous traditions converged and were adopted into a maximalist approach to meaning in the form of cognitive semantics, which pays attention to the two sides of the aforementioned dichotomies, as they are now thought to be interrelated.

Both the main contributions and the most significant shortcomings of the various theories have been discussed, leading us to conclude that the most convenient framework for the purposes of the present dissertation lies at the interface of the neostructuralist distributional corpus-based methodology and cognitive semantics. In particular, the usage-based approach to onomasiological variation seems to be especially suitable for the examination of factors and causes underlying the choice between synonymous expressions. Nevertheless, studies adopting such an approach have to date concentrated primarily on constructional synonymy, while those that examine lexical synonyms do so mainly from a synchronic perspective. Therefore, the analyses provided in Chapters 5–7 of this dissertation attempt to contribute to this underresearched line of inquiry. The focus is on the phenomenon of synonymy and, in particular, on a set of adjectival near-synonyms from the olfactory domain, which are described in detail in Chapter 4. First, however, we turn in Chapter 3 to an in-depth theoretical description of the semantic relation of synonymy and, more specifically, to a review of studies of concrete synonym sets that have been carried out from the perspective of usage-based onomasiology.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3 SYNONYMY

In the previous chapter synonymy was defined as a relation of identity of meaning between two or more words or word senses (cf. Section 2.2). However, it was claimed that a less strict definition is more suitable for this semantic relation, as words are rarely —if ever— completely identical in meaning. Different types of synonyms were briefly introduced in Section 2.2 depending on which aspects of meaning they share, i.e. whether they match only in denotative meaning or whether they coincide also in other dimensions. such as connotation, style, and/or collocation. This implies that different degrees of synonymy exist, and it appears to be the case that competent language users have an intuition about some words or word senses being more synonymous than others (e.g. Lyons 1968: 447; Cruse 1986: 265–268). Even within a specific set of synonyms, speakers commonly have the feeling that some members of a synonym set share more semantic features than others and can thus be interchanged in more contexts of use. However, although synonymy is typically defined and viewed as a semantic relation of similarity, it is also crucial to consider the differences between synonymous words in order to provide a comprehensive picture of the semantic relation that holds between them. This is so because, as has already been pointed out, a great majority of synonyms, despite being semantically similar in one or more aspects, also differ in other respects, which prevents them from being freely interchangeable when used.

Once the focus shifts from concentrating solely on the similarities to considering also the level of contrast between synonymous words, it becomes evident that the way in which synonyms differ from each other varies from synonym set to synonym set, as well as from synonym to synonym. Edmonds (1999: Chapter 3) claims that there are at least 35 ways in which synonyms can differ and provides evidence of the complex and intriguing nature of synonymy. In fact, research on synonymy has been and continues to be a rich field of study in many areas of linguistics, including lexicography (e.g. Atkins & Levin 1995; Wiegand 1999; Bergenholtz & Gouws 2012; Adamska-Salaciak 2013; Moon 2013; Murphy 2013), computational linguistics (e.g. Ravin 1992; Edmonds 1999; Edmonds & Hirst 2002; Inkpen &

Hirst 2006; Inkpen 2007), and language learning and teaching (e.g. Martin 1984; Xiao & McEnery 2006; Yeh, Liou & Li 2007; Liu 2013a; Liu & Zong 2016), among others. Giving a detailed account of all the research conducted on this semantic relation therefore lies beyond the scope of the present dissertation, which analyzes a specific set of synonyms from the perspective of the distributional corpus-based approach introduced in Section 2.3.2. Instead, the focus of this chapter is exclusively on studies employing such a methodology to examine the internal semantic structure of pairs or sets of synonyms, either in a strict neostructuralist fashion or by resorting also to concepts and theories from cognitive semantics. Additionally, the emphasis is mainly on lexical synonymy, thus excluding most of the literature on constructional synonymy, though some studies of the latter kind are considered for methodological reasons.

However, before reviewing existing research on individual groups of synonyms, I describe in more detail the different types of synonyms identified mainly in structuralist work that were briefly introduced in the previous chapter (i.e. absolute synonyms and near-synonyms, among others). As such, this chapter is structured as follows. Section 3.1 is concerned with the classifications of types of synonymy, with a special emphasis on three types that are most recurrent: absolute synonymy, cognitive synonymy, and near-synonymy. Then, in Section 3.2, previous work on specific pairs and sets of synonyms employing the corpus-based contextual approach is summarized and assessed, focusing particularly on methodological aspects such as the number and types of factors analyzed to determine the choice between synonyms, the statistical techniques applied, but also relevant findings and contributions of such studies. While Section 3.2 deals exclusively with synonymy from a synchronic perspective, Section 3.3 considers the few corpus-based studies on lexical synonyms that have been conducted from a diachronic viewpoint. Lastly, Section 3.4 offers a summary of the issues discussed in the present chapter.

3.1 CLASSIFICATIONS OF SYNONYMY

As mentioned above, synonymy is a linguistic phenomenon which has been widely researched in different subfields of linguistics and which has also been studied from a broad variety of perspectives and theoretical frameworks. One early research area, primarily within the structuralist school, aimed at providing classifications of synonyms, in which different types and subtypes were identified and established depending on the extent and kind of semantic overlap that exists between two or more words or word senses. However, in order to fully

understand these classifications, a succinct summary of the different dimensions or types of meaning that are typically distinguished seems in order here.

Although at least seven types of meaning are distinguished in the specialized literature (cf., for instance, Leech 1974: Chapter 2), four of these types are especially relevant for our purposes: *denotational meaning*, *expressive meaning*, *stylistic meaning*, and *collocational meaning* (cf. also Leith 1983: 70–73; Cruse 2000: 46–61).[20] Denotational meaning refers to what a word designates in the extralinguistic world, be it an entity, an event, an action, or a process, which is said to be the reference or *denotatum* of that particular word. Expressive meaning, on the other hand, is used to refer to the overtones carried by words (i.e. connotations) as well as to speakers' positions and evaluative judgements toward words and their referents (i.e. affect). This dimension of meaning allows language users to regard lexical items as more or less neutral, negative, or positive. The third type, stylistic meaning, concerns aspects such as formality, discourse field, and mode. Although stylistic meaning is often reduced to the first of these three aspects (i.e. the formal/informal dichotomy), it also involves the fields of discourse in which words are typically used, as well as the medium in which they are employed, be it spoken or written language. Additionally, relationships between the participants in a linguistic exchange are also of fundamental importance for this dimension, since familiarity and social distance play a crucial role when it comes to stylistic choices. Collocational or collocative meaning, i.e. the dimension of meaning which concerns the way in which lexical items are used in combination with other items in the lexicon, has already been thoroughly discussed earlier (cf. Section 2.3.2) and is therefore not further explained here.[21]

Many scholars agree that the first of these four types of meaning, that is, denotational meaning, is the most relevant dimension (e.g. Murphy 2003: 137). In fact, two or more words must share the same denotation and thus refer to the same concept in the extralinguistic reality,

[20] Different labels have been used to refer to these dimensions of meaning in the specialized literature: (i) denotational meaning is sometimes also known as *referential*, *conceptual*, or *logical meaning*, (ii) expressive meaning can also be referred to with the labels *affective* or *emotive meaning*, and (iii) *social meaning* is another term for stylistic meaning (cf., for instance, Leech 1974: Chapter 2; Leith 1983: 70–73).

[21] Leech (1974: Chapter 2) distinguishes three additional major types of meaning, to wit, connotative meaning, reflected meaning, and thematic meaning. Connotative meaning is here included within expressive meaning, following Cruse (2000: 58–60). Reflected meaning arises in cases of polysemous words, when one sense evokes some of the characteristics of another sense (Leech 1974: 16–17). Thematic meaning, in turn, concerns the organization of information (e.g. focus and emphasis; Leech 1974: 19–20). Neither reflective meaning nor thematic meaning are relevant for the purposes of the present dissertation and are therefore not discussed in detail.

to be regarded as synonymous, while they are allowed to vary in other aspects of meaning. To provide an example, *woman* and *lady* are considered synonyms as they both denote an 'adult human female'. Nevertheless, the two words differ as regards expressive meaning: whereas *woman* is fairly neutral, connoting a vast set of properties, *lady* is expressively more restricted, since it is closely associated with features such as class, status, elegance, and manners, and thus carries a positive connotation (Leith 1983: 70–72). However, as shown by this example, all non-denotational dimensions of meaning are crucial when an onomasiological orientation is adopted, since as Grondelaers, Speelman & Geeraerts (2007: 994) point out, "[…] the very definition of nonreferential [i.e. non-denotational] meaning involves the concept of onomasiological alternatives". It is precisely along these other dimensions of meaning that the great majority of synonyms differ and, consequently, they must be accounted for in a semantic analysis of synonymy.

Different classifications of synonymy have been put forward in the literature and, as a result, a variety of labels have been used to refer to the same types of synonyms, in some cases even by the same author. Here, we focus exclusively on the different types of synonymy established in the most important and well-known works on the topic, namely those by Lyons (1968; 1981a; 1981b; 1995) and Cruse (1986; 2000; 2002), but also that of Murphy (2003).

Lyons' (1968) early classification is probably the most comprehensive one as it distinguishes between two dimensions: total synonymy, that is, those words which are interchangeable in all contexts of use, and complete synonymy, i.e. those words which are equivalent both on the denotative and connotative dimensions of meaning. By crossing these two dimensions, four types of synonyms can be established, namely (i) complete and total synonyms, (ii) complete and non-total synonyms, (iii) incomplete and total synonyms, and (iv) incomplete and non-total synonyms. For example, the first type, complete and total, includes those synonyms which share both their denotational and expressive meanings and are interchangeable in all contexts. This type refers to what is usually termed absolute synonymy.[22] However, although this classification is the most exhaustive one, it can also be said to provide unnecessary divisions. This is so because type (iii), incomplete and total synonymy, would refer to words or word senses which are interchangeable in all contexts of use without being denotationally and expressively equivalent. Such a relation is quite unlikely —if not

[22] Absolute synonymy is sometimes also known as *perfect synonymy* (e.g. Stern 1931: 225), *pure synonymy* (e.g. Ullmann 1957: 110), *exact synonymy* (e.g. Ziff 1964: 172), or *real synonymy* (e.g. Lyons 1963: 74).

impossible— to occur in language.[23] Lyons' (1995) latest classification differs considerably from his previous ones (i.e. 1968; 1981a; 1981b), as he establishes a different division into absolute synonymy, cognitive synonymy, near-synonymy, and partial synonymy. *Partial synonymy* roughly corresponds with type (ii) in his early classification (i.e. complete and non-total). The reason for drawing a distinction between absolute and partial synonyms, according to Lyons, is that whereas absolute synonyms are most certainly non-existent in language, partial synonyms, though infrequent, do exist. This is because, in contrast to absolute synonyms, partial synonyms are equivalent both as concerns denotational and expressive meaning, but need not be interchangeable in all contexts.

Cruse (1986), in turn, proposes a three-way split into absolute synonymy, cognitive synonymy, and plesionymy, which partially overlap with Lyons' (1995) typology.[24] In subsequent classifications by Cruse (2000; 2002), the same three types of synonyms are present, although the labels for the latter two vary. For instance, Cruse uses the labels *cognitive synonymy* and *plesionymy* in his 1986 monograph, while in later work (2000) he refers to the same types of synonymy as *propositional synonymy* and *near-synonymy*, respectively.

Table 1 below summarizes the various classifications of synonymy by the two authors and the terminology discussed so far:

[23] Some subsequent formulations of Lyons' classification overlap substantially with the original one while at the same time differing in certain respects, both regarding the labels used and the types identified (for more details, cf. 1981a; 1981b).

[24] Near-synonymy, rather than plesionymy (e.g. Hirst 1995; Storjohann 2009), is the label adopted in the present dissertation as it is the most frequent term employed in contemporary research on synonymy.

Table 1. Classifications of synonymy by Lyons and Cruse

Author / Definition	Lyons (1968)	Lyons (1995)	Cruse (1986)	Cruse (2000; 2002)
Identical on all dimensions of meaning and interchangeable in all contexts of use	Complete & total	Absolute synonymy	Absolute synonymy	Absolute synonymy
Identical on all dimensions of meaning but not interchangeable in all contexts of use	Complete & non-total	Partial synonymy	-	-
Not identical on all dimensions of meaning but interchangeable in all contexts of use	Incomplete & total	-	-	-
Identical on the denotative dimension of meaning, but not on other dimensions, and not interchangeable in all contexts of use	Incomplete & non-total	Cognitive synonymy	Cognitive synonymy	Propositional synonymy
Similar but not identical on the denotational dimension of meaning and not interchangeable in all contexts of use		Near-synonymy	Plesionymy	Near-synonymy

In addition to Lyons' and Cruse's classifications of synonymy, Murphy (2003) postulates yet another typology, in which she distinguishes between synonyms in terms of denotative meaning only and synonyms which share more than their denotation. As regards denotative meaning, she considers two dimensions, namely (i) how many senses two or more synonymous words share and (ii) how similar the shared senses are. *Full synonyms* are those which share all their senses and these senses are all identical (e.g. *groundhog* and *woodchuck*). *Sense synonyms*, in turn, are those which share one or more senses, but not all. As in the case of full synonyms, the shared senses are also identical. One example would be *couch* and *sofa*, which share the sense '[a] long, stuffed seat with a back and ends or end, used for reclining' (*OED* s.v. *sofa* n. 2), but not the meaning '[a] couch upon which a patient reclines when undergoing psychoanalysis or psychiatric treatment'(*OED*, s.v. *couch* n. 3b), which only the former term can designate. Full synonyms and sense synonyms are included within the general category of cognitive synonyms, or *logical synonyms* in Murphy's terminology. Finally, lexical items that are similar, but not identical in one or more of their senses are called near-synonyms or plesionyms, as in the case of *fog*(*gy*) and *mist*(*y*), which differ as regards the density of the weather phenomenon they denote (see below). Concerning non-denotative dimensions (e.g. register, dialect, and connotation), synonyms may also differ. According to Murphy (2003: 151), such aspects of meaning can either enhance or diminish their degree of similarity. For

instance, the sense synonyms *drunk* and *inebriated*, despite sharing the sense '[a]ffected by alcohol to the extent of losing control of one's faculties or behavior' (*Lexico*, s.v. *drunk* adj. 1), differ in register, the former being more informal than the latter.

Notwithstanding the existence of these and other classifications of synonymy, what is relevant here is that three types are recurrently mentioned in most classifications, though with different labels. These are absolute synonymy, cognitive synonymy, and near-synonymy. The first of these three types is thought to constitute one of the endpoints of the synonym continuum, namely that of absolute identity of meaning. Exactly what counts as absolute identity of meaning depends largely on the types of meaning that are considered to be relevant. Nevertheless, most scholars agree that absolute synonymy refers to those words or word senses which are identical on all four dimensions of meaning mentioned above (i.e. denotational meaning, expressive meaning, stylistic meaning, and collocational meaning). Ullmann (1957: 109–110), for instance, claims that two criteria need to be met for absolute synonymy to occur. On the one hand, absolute synonyms must be able to substitute for each other in all contexts of use and, on the other, they must not differ in neither denotational nor non-denotational aspects of meaning. Cruse (1986; 2000) goes one step further by stating that these two criteria are not sufficient, and adds yet a third one, to wit, that of contextual normality. This is a particularly useful criterion for a contextual approach such as the one adopted in the present dissertation. According to such a perspective, in addition to being interchangeable, absolute synonyms must also be *equinormal* in all contexts of use. Along these lines, Cruse (2000: 157) argues that

> […] absolute synonyms can be defined as items which are equinormal in all contexts: that is to say, for two lexical items X and Y, if they are to be recognized as absolute synonyms, in any context in which X is fully normal, Y is, too; in any context in which X is slightly odd, Y is also slightly odd, and in any context in which X is totally anomalous, the same is true of Y.

This definition evidences that very severe requirements are imposed on words or word senses in order for them to comply with the criteria for absolute synonymy and demonstrates why this type of synonymy is very rare —if not inexistent— in language. Following this definition, if a context exists in which two synonyms are not equally normal, that is, in which one is more natural or idiomatic than the other, absolute synonymy is ruled out. Cruse (2000: 268–269) tests some plausible candidates, for instance, *begin* and *commence*, *almost* and *nearly*, and

scandalous and *outrageous*, and draws the conclusion that for all these synonymous pairs contexts in which one of the members is more normal than the other can be observed. In fact, it is commonly claimed that there is no reason why a semantic relation of this type should exist and be maintained in language, as it goes against the economy principle. As Taylor (2003: 264) states, "[absolute] synonymy would have to be regarded as an extravagant luxury, even dysfunctional, in that limited symbolic resources get squandered on the designation of one and the same semantic unit." In this statement, Taylor indirectly refers to the well-known *no-synonymy* rule or *isomorphic* state of languages, i.e. one form, one meaning (e.g. Bolinger 1977: ix–x, 9; Wierzbicka 1988: 13–14; Croft 2000: 176; Nuyts & Byloo 2015: 62–63), which makes languages work against absolute synonymy if it at some point comes to exist through processes such as borrowing. This is an issue that has clear diachronic implications and is therefore further discussed in Section 3.3 below.

Concerning cognitive synonymy, this is a relation which holds between words or word senses that are identical on the denotational dimension —that is, which have the same conceptual content— and therefore mutually entail one another. In other words, cognitive synonyms, if substituted for one another in a particular context of use, generate clauses or sentences with the same truth-values. Nevertheless, such pairs or groups of synonyms do not meet the criteria for absolute synonymy because they differ in non-denotational traits, for instance, connotation (e.g. *firm* and *stubborn*), register (e.g. *drunk* and *pissed*), or style (e.g. *kick the bucket*, *die*, and *pass away*), or the language variety in which they occur, such as British English (BrE) *autumn* and AmE *fall* (Cruse 1986: 278–280; Desagulier 2014: 153). In many cases, cognitive synonyms differ in more than one of the non-denotational types of meaning simultaneously (e.g. both expressive and stylistic meaning). Given that cognitive synonyms differ in non-denotational traits, they are typically used in different speech situations, text-types, or fields of discourse, and using one lexical item in a context in which its synonym is more typical or frequent will most probably result in an odd sentence. Areas of vocabulary which are related to human experiences that stir strong emotions and opinions among people, for instance, death, religion, and sex in Western cultures and societies, tend to display a wide array of cognitive synonyms (Quine 1951: 28–31; Cruse 1986: 88, 270–285; 2000: 158–159).[25]

Finally, near-synonymy, which is undoubtedly the most frequent type of synonymy in language, refers to those words or word senses that differ slightly in conceptual content and are

[25] Cf. Cruse (1986: 285) for a comprehensive list of cognitive synonyms in these semantic domains.

thus not denotationally identical. As such, if substituted for one another, the generated sentences yield somewhat different truth-conditions. However, near-synonyms are still sufficiently semantically similar to be interchanged in many contexts of use. The denotational traits in which near-synonyms vary must, according to Cruse (2000: 159), "[…] be either minor or backgrounded". Differences between near-synonyms include, but are not limited to, aspects such as the following (examples taken from Cruse 2000: 157–160; Murphy 2003: 147; Divjak 2010: 3–4; Desagulier 2014: 153):

(i) Contiguity on a gradable continuum or degree, e.g. *fog*(*gy*) and *mist*(*y*): Both *fog* and *mist* refer to '[a] cloud of water substance present in the atmosphere at or close to the surface of the earth' (*OED*, s.v. *fog* n. 2, I 1a; *mist*. n. 1, I 1a). However, *fog* is denser and thicker than *mist*, the latter being lighter and therefore not limiting visibility to the same extent as the former.

(ii) Nuances involving specialization, e.g. *laugh* and *giggle*: the verb *laugh* denotes any type of spontaneous movement of the lower part of the face, accompanied by an explosive vocal sound, to display emotions such as joy or amusement, arising in a wide variety of situations (*OED*, s.v. *laugh* v. 1a) . In turn, the verb *giggle* is more specialized, as it refers to a particular way of laughing, namely a light and foolish type which occurs in a more restricted set of circumstances, including amusement but also uneasiness and embarrassment (*OED*, s.v. *giggle* v.1, a).[26]

(iii) Prototypicality, e.g. *brave* and *courageous*: While *brave* is used to describe a fearless person in physical terms, *courageous* emphasizes the moral or intellectual dimension.

(iv) Viewpoint or connotation, e.g. *slim*/*slender* and *skinny*: The three terms are all used to refer to people who are thin, but whereas *slim* and *slender* have positive connotations, being associated with grace and elegance, *skinny* has a clearly negative connotation, being linked to unattractiveness and negative esthetic qualities in general.

(v) Aspectual variation, e.g. *calm* and *placid*: The former adjective is used to denote a determinate state, which is ephemeral, while the latter refers to a more inherent feature of a person or entity.

[26] Some authors would argue that *giggle* is a hyponym of *laugh*, as *giggle* is included in the definition of *laugh*. This example demonstrates that the dividing line between the two semantic relations (i.e. synonymy and hyponymy) is not always clear-cut.

In contrast to absolute synonyms, near-synonyms, which are particularly frequent in language, cannot be considered uneconomical and unnecessary, especially since language users are more likely to devise differences between such related words than to abandon one or more of the terms. In fact, a great majority of scholars claim that near-synonyms add substantially to the lexical expressivity of language, as they enable speakers to convey different nuances of meaning in a very precise manner (cf., for instance, Edmonds & Hirst 2002: 107–108; Murphy 2003: 165–166). However, given that near-synonyms can vary on any dimension of meaning and since that variation is frequently multidimensional, even native speakers sometimes find it difficult to master the fine-grained differences that exist between them.

Although the distinction between cognitive synonyms and near-synonyms appears to be clear in theory, as the former entail each other while the latter do not, the boundary between the two types becomes much more problematic in practice. This is so because it is not often obvious whether two or more words differ solely in their non-denotational semantic traits or whether, on the contrary, slight differences in denotation also exist between them. Due to the difficulty in distinguishing between cognitive synonymy and near-synonymy when specific pairs or sets of synonyms are concerned, some scholars have argued for a two-fold, instead of a three-fold, division into synonym types, namely absolute synonymy vs. non-absolute synonymy or near-synonymy, thus dismissing the distinction between cognitive synonyms and near-synonyms, as this opposition is often troublesome. This is the standpoint adopted by Edmonds & Hirst (2002: 116–117) and Desagulier (2014: 153), among others, and the one followed in the present dissertation.

A great amount of research has been conducted on specific near-synonymous pairs or groups in the last few decades. Given that near-synonyms differ in one or more aspects of meaning, it is certainly not sufficient to state that a relation of partial equivalence exists between the members of a synonym set, but it is also necessary to examine the multiple factors determining the choice between them in order to reach an in-depth and comprehensive understanding of their semantic relation. Such an approach is the only one which allows us to know for certain in which contexts and the reasons why we use one word instead of another one. Disentangling the differences between near-synonyms, which are often multivariate and related to semantic, stylistic, and/or morphosyntactic dimensions, among others, has been and continues to be the main aim of many studies with a usage-based onomasiological orientation. The remainder of this chapter is concerned with a review of such studies. Section 3.2. discusses

synchronic research on lexical synonymy, while Section 3.3 considers studies with a diachronic perspective.

3.2 USAGE-BASED CORPUS STUDIES ON SPECIFIC SYNONYMS

Although synonymy in general has figured prominently in theories of lexical semantics, dating back to the structuralist school (cf. Section 2.2), the internal semantic structure of specific pairs or larger sets of lexical near-synonyms had until quite recently not received much attention, especially if compared to constructional near-synonyms and other semantic phenomena such as metaphor and polysemy, which have been the subject of an ever-increasing body of literature (Edmonds & Hirst 2002: 106; Taylor 2003: 264; Geeraerts 2010: 264; Liu 2010: 57). Consequently, several scholars have lately stressed the need for more research in this field, which has led to the emergence over the last 30 years of studies following diverse types of methodologies within the distributional approach originally set out by Firth and Sinclair in the middle of the 20th century (cf. Section 2.3.2). The aim of this section is to offer an overview of the existing synchronic corpus-based research on specific synonym sets that has been carried out since the end of the 1980s and the evolution of this line of research on methodological grounds. A detailed account of all the existing studies is not provided here because not all the methods proposed in the literature are equally relevant for our purposes. This is so because the approach adopted in such studies differs widely depending on the specific groups of synonyms analyzed, the POS the synonyms belong to, and the factors that significantly influence the choice between them.[27] This is to be expected since, as argued by Liu (2010: 61),

> […] the use of diverse procedures makes perfect sense in the study of internal structures of near-synonym sets because the way near synonyms differ often varies from synonym to synonym and from set to set. […] Thus a researcher has to decide, based on a close scrutiny of the features of the synonyms being examined, what micro-procedures to employ.[28]

[27] See Glynn (2010: Tables 1 and 2) for a list and a methodological summary of corpus-based studies in cognitive semantics dealing with lexical and constructional onomasiology and semasiology, including research on synonymy and polysemy, among other semantic relations.

[28] For a similar argument, cf. also Hanks (1996: 92; 96).

Therefore, the next section provides some generalizations that can be extracted from previous studies, primarily focusing on those dealing with adjectival near-synonyms, as the factors they consider have more in common with the analyses conducted in this dissertation (cf. Chapters 5–7).

3.2.1 Two waves of research

The research which is reviewed in this section can be classified into different groups depending on three parameters related to the scope of the analyses and to methodological issues:

(i) Number of synonyms examined: pairs vs. larger sets (i.e. three or more synonyms).
(ii) Amount and type of factors conditioning the choice between synonyms, ranging from collocational and/or stylistic patterns only to a wider array of determinants of variation which include semantic, morphosyntactic, and/or extralinguistic factors.
(iii) Statistical techniques and measures used: from mere frequencies and percentages, to association measures of collocational strength (e.g. PMI, *t*-score, log likelihood), multivariate tests (e.g. Hierarchical Configural Frequency Analysis (HCFA), logistic regression analysis, multiple correspondence analysis), and clustering techniques (e.g. HAC).

Along these parameters, corpus-based synonymy research can be divided, with some exceptions, into two main waves. The first one includes mainly those studies published approximately before the mid-2000s, most of which focus on pairs of near-synonyms by making use of raw frequencies, percentages, or association measures in order to examine their collocational and stylistic patterns. By contrast, in the second wave, which can be said to take off with the works of Divjak (2006) and Divjak & Gries (2006; 2008), the focus shifts to larger groups of near-synonyms. Here, in addition to accounting for collocational and stylistic behavior, further factors from different linguistic levels are also considered, including semantics and morphosyntax. A direct corollary of this multidimensional perspective is the use of more advanced and sophisticated techniques, which allow for the inclusion of a larger number of variables. The development of corpus-based research on lexical synonymy over the last 30 years can thus be summarized as shown in Figure 2:

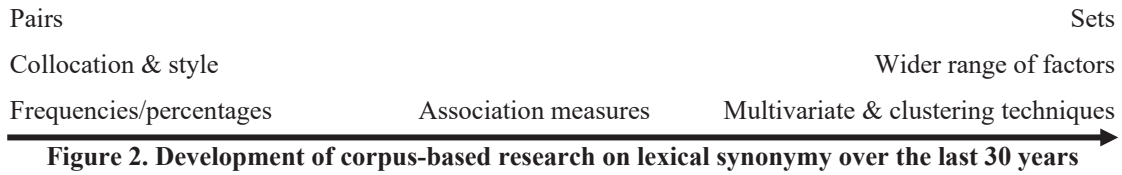


Figure 2. Development of corpus-based research on lexical synonymy over the last 30 years

As has already been mentioned, most of the early research in this field considered only pairs of near-synonyms rather than larger groups (cf., for instance, Persson 1989; Church et al. 1991; Kennedy 1991; Gries 2001; 2003; Arppe 2002; Kjellmer 2003; Taylor 2003). One case in point is Biber, Conrad & Reppen (1998: Section 4.3), who analyzed the grammatical preferences of the verbs *begin* and *start*. Although *begin* and *start* can be used in exactly the same syntactic structures, both intransitive and transitive, and even with the same type of objects when used transitively (i.e. noun phrases, *to*-clauses, and *-ing* clauses), closer inspection of their distributional patterns in such contexts reveals that the way in which they are typically used in actual discourse differs considerably. While the frequency of *start* almost doubles that of *begin* in intransitive structures in fictional texts, *begin* is three time more common than *start* with *to*-clauses as objects. In fact, in fictional texts more than half (60%) of the total examples of *begin* are transitive uses followed by *to*-clauses. Contrariwise, *start* occurs in this particular structure in only 17% of the cases, which demonstrates that it is preferred in other patterns, be it intransitive structures or transitive ones with other types of direct objects. Therefore Biber, Conrad & Reppen (1998) show that by examining the valency of the two verbs, remarkable differences in use emerge. Their results point to a division of labor between the two near-synonymous verbs: despite being interchangeable in most cases, each verb is clearly the preferred option in different contexts of use. Additionally, language users are not able to predict the individual preferred grammatical associations of the two verbs, which they generally consider to be completely interchangeable. It is therefore only by means of detailed analyses of usage data such as that in Biber, Conrad & Reppen, made possible by the availability of large corpora, that these preferred structures can be discovered, while native speaker intuitions are often unreliable.

Despite the valuable findings of studies on near-synonymous pairs such as the one conducted by Biber, Conrad & Reppen (1998), it is clear from the information presented in dictionaries and thesauri that most lexical items have more than one synonym and thus synonyms typically come in larger groups (e.g. Divjak 2006; Gries 2010). For instance, the verb

commence is also listed as a synonym of both *begin* and *start* in dictionaries and thesauri (e.g. *MW*). Analyzing only pairs of synonyms could therefore be considered reductive and insufficient in order to offer a complete picture of the internal semantic structure of specific concepts, which can in most cases be named by more than simply two lexical items. This is the principal motivation for the analysis of larger sets of near-synonyms in the studies belonging to the second wave of research mentioned above (e.g. Divjak 2006; 2010; Divjak & Gries 2006; 2008; Arppe & Järvikivi 2007; Arppe 2008; Gries & Otani 2010; Liu 2010; 2013b; Liu & Espino 2012; Desagulier 2014; Krawczak 2014; Yatandu Uba 2015).[29] For instance, Divjak & Gries (2006) examine the distributional profiles, particularly the syntactic and semantic patterns, of nine Russian verbs of trying: *probovat'* ('try'), *pytat'sja* ('try, attempt'), *starat'sja* ('try, endeavor'), *silit'sja* ('try, make efforts'), *norovit'* ('try, strive to, aim at'), *poryvat'sja* ('try, endeavor'), *tščit'sja* ('try, endeavor'), *pyžit'sja* ('go all out'), and *tužit'sja* ('make an effort, exert oneself'). This is done by means of cluster analysis so as to determine whether synonyms are structured in pairs, as assumed in earlier research, or in larger groups. Three clearly delineated clusters emerge from their analysis, each containing three verbs:

(i) *Probovat'*, *pytat'sja*, and *starat'sja* belong to the first cluster and are typically used to designate 'a human who is strongly encouraged to try to physically move himself or others'.

(ii) The verbs *tščit'sja*, *pyžit'sja*, and *tužit'sja* form the second cluster and are most commonly used to refer to 'an inanimate subject which tries very hard (without succeeding) to undertake "metaphorical extensions of physical actions"' (Divjak & Gries 2006: 42).

(iii) The third cluster, formed by the remaining three verbs, *silit'sja*, *norovit'*, and *poryvat'sja*, signals 'an inanimate entity that repeatedly, but not very intensely, tries but fails to perform a physical action'.

[29] Research before the mid-2000s that did focus on larger groups of near-synonyms include Hanks' (1996) on the verb *urge* and its near-synonyms (e.g. *request*, *order*, *command*, *incite*, and *encourage*), Biber, Conrad & Reppens' (1998: Section 2.6) analysis of the adjectives *big*, *large*, and *great*, and Partington's (1998: Sections 2.4 and 3.6) investigation into the adjectives *sheer*, *pure*, *complete*, and *absolute* and the adverbs *absolutely*, *completely*, and *entirely*.

Divjak & Gries (2008) test the corpus-based results in their 2006 paper by means of experimental data with the aim of validating their claim that humans categorize synonyms into groups rather than pairs. Their findings, for the most part, point to a significant correlation between the corpus-based analysis and the experimental data. They thus conclude that "[…] there seems to be a psychological reality corresponding to clusters of near synonyms" and that "[…] speakers group these near synonyms into clusters, not pairs. Hence near synonymy is not about pairs of words that entertain dichotomous, dyadic relations […] but about groups of words that are more similar to each other than to (words belonging to) other groups of semantically similar words" (2008: 208).

These two studies by Divjak & Gries constitute a clear break with the first wave of research on synonymy. Subsequent studies in the field focus on larger groups of lexical near-synonyms, with the exception of Chung's (2011) analysis of *create* and *produce* and Krawczak's (2018) contrastive investigation into the adjectival near-synonyms *ashamed* and *embarrassed* and their corresponding terms in French (i.e. *honteux* and *embarrassé*) and Polish (i.e. *zazenowany* and *zawstydzony*).

However, regardless of whether pairs or larger groups of near-synonyms are analyzed, most corpus-based onomasiological research conducted during the last three decades have the same main goal, namely, to identify how particular near-synonyms differ from each other. As such, the focus is on the factors determining the choice between the near-synonyms object of study. As has already been mentioned, the individual factors that are relevant depend largely on the nature of the synonym set examined, e.g. POS or semantic domain to which the members belong, among others. Therefore, as Liu (2010: 61; cf. quotation above) points out, the micro-procedures employed differ from study to study and one of the most important tasks of the researcher is precisely to tease out the particular factors that could have a significant impact on the choice between the near-synonyms at issue.

Closely related to the amount and types of factors included in the analyses are the specific methods used to evaluate the differences between near-synonyms. If only one or two factors, such as collocational and/or stylistic patterns are considered, it may be sufficient to use frequencies and percentages or association measures of collocational strength. However, including a wider range of factors in the analysis requires the use of more sophisticated multivariate techniques in order to disentangle the individual effects of the factors as well as the possible interactions between them. Consequently, parameters (ii) and (iii) mentioned

above, i.e. amount and types of factors and statistical techniques and measures, are here jointly discussed.

The main focus of most corpus-based studies on near-synonyms during the 1990s and early 2000s was on collocational preferences (e.g. Persson 1989; Church et al. 1991; Biber, Conrad & Reppen 1998; Partington 1998; Gries 2001; 2003; Kjellmer 2003). The paramount importance of collocation, as seen in Section 2.3.2, constitutes one of the three central tenets of the neostructuralist distributional approach (Geeraerts 2010: 167), and the term has since then been one of the most important notions in lexical semantics. We have already seen in the previous chapter that, in collocational analyses, a proximity limit is often determined to restrict the number of collocates of a node and that such a limit often takes the form of a context window of *x* words to the left and to the right of the node. However, whereas some studies use a rather wide context window, others consider only direct adjacent collocates of the node. Typically, adjacent collocates are examined to search for collocates which are considered to be more informative and thus more enlightening. Within the first wave of corpus-based studies on near-synonymy, we find collocational analyses of both types. For instance, Kjellmer (2003), in his work on *almost* and *nearly*, opts for a context window of L4–R4 and focuses on the type of individual collocates that occur with these adverbs in terms of their POS. Kjellmer draws the conclusion that *almost* and *nearly* differ considerably in collocational behavior: while the former is mainly followed by adverbs, adjectives, pronouns, and prepositions (e.g. *almost immediately*, *almost identical*, *almost all*, and *almost by definition*), post co-occurring collocates of *nearly* are primarily numerals (e.g. *nearly 50 years* or *nearly two thousand*). In turn, most studies on near-synonymous adjectives that focus on collocational behavior tend to consider exclusively their right-side collocates, particularly those immediately following the adjectives, i.e. R1 collocates. This is so because, in most cases, the nouns modified by adjectives appear in this particular slot, at least when used attributively. Such studies include, among others, Persson (1989) on *deep* and *profound*, Church et al. (1991) on *strong* and *powerful*, Biber, Conrad & Reppen (1998: Section 2.6) on *big*, *large*, and *great*, and Gries (2001; 2003) on *-ic* and *-ical* pairs, for instance, *alphabetic*(*al*), *economic*(*al*), and *classic*(*al*).

Much of this line of research focuses not only on the individual lexical words that co-occur with the near-synonyms object of study, but also on more schematic aspects of the notion of collocation. As discussed in Section 2.3.2, collocation can, on the one hand, be considered in a narrow or strict sense, including only individual collocates, thus signaling "[…] a lexical

relation between two or more [lexical] words which have a tendency to co-occur within a few words of each other in running text" (Stubbs 2002: 24). On the other hand, the notion can also be understood in a broader sense, thus extending to other types of co-occurrence information. It is here worth recalling that Sinclair (1991; 1996) distinguished three additional levels of co-occurrence, namely colligation, semantic preference, and semantic prosody.[30] In some studies, for instance Kjellmer (2003) on *almost* and *nearly*, function words such as pronouns and prepositions, but also determiners and conjunctions, are included in the collocational analysis; these are in fact colligations of the near-synonyms.

Other studies abstract away from individual lexical collocates by grouping them into semantic categories or domains, thus focusing on their semantic preferences. One case in point is Gries' (2001) analyses on adjectival *-ic* vs. *-ical* pairs. He claims that in order to obtain a clearer insight into differences in collocational behavior of near-synonyms, it is sometimes necessary not only to focus on their specific collocates but also to consider their discriminating features or indicator attributes. By doing so, for the pair *magic*(*al*), Gries is able to determine that while *magic* is commonly used to modify concrete nouns (e.g. *carpet*, *lantern*, and *sword*), *magical* is more typically used with abstract nouns (e.g. *power*, *mystery*, and *healing*). Similarly, in the pair *electric*(*al*), *electric* is significantly more frequent with specific or basic-level nouns of a concrete nature (e.g. *kettle*, *sunroof*, and *toothbrush*), while *electrical* appears to a greater extent with more general or superordinate nouns, including both concrete and abstract entities (e.g. *appliances* and *equipment*).

Finally, some scholars also include information pertaining to semantic prosody in their collocational analyses of near-synonyms. For example, Hanks (1996; cf. also 2013: Section 5.2.6) examines the contexts in which *urge* and two of its near-synonyms, *incite* and *encourage*, are used in the *British National Corpus* (*BNC*). He finds that *urge* and *encourage* are neutral, while *incite* has a negative semantic prosody, being used in the great majority of cases to refer to an action that someone is persuaded to carry out and which tends to have negative or unpleasant effects in society. The negative semantic prosody of *incite* is evident in examples such as (10) below, where words or expressions with a pejorative connotation clearly predominate (*abducting*, *rape*, *debauchery*, *violence*, and *assault*):

[30] As these three notions were already introduced and explained in the previous chapter (Section 2.3.2), they are not defined and described further here.

(10) Moroccan Chief Police Commissioner Haj Mohamed Mustapha Tabet, 54, was convicted of rape, deflowering virgins, abducting women, **inciting** debauchery, violence and assault in a flat he specially kept for staging sex orgies. (*BNC*, newspaper reports)

Another important distinction to draw in research on the collocational behavior of near-synonyms concerns how collocational preferences are measured: by means of frequencies and percentages only or by making use of association measures. This methodological issue is closely related to how collocation is defined. As mentioned earlier (cf. Section 2.3.2), three definitions of the concept are often identified in the specialized literature, namely the textual, the statistical, and the psychological or associative one (cf., for instance, Partington 1998: 15–16). As seen above (cf. quotation by Stubbs 2002), in the textual sense collocation refers to all the neighboring lexical items of a node appearing within a short span of words in a specific textual unit (e.g. clause or sentence). If such a definition is adopted, mere frequencies or percentages are enough to compute the collocational strength between a node and its collocates. Analyses on near-synonyms adopting this perspective include, for instance, Persson (1989), Kennedy (1991), Biber, Reppen & Conrad (1998), Taylor (2003), and Chung (2011). However, a great amount of research in the field embrace a statistical definition of collocation, thus considering only those collocates which occur with a node significantly more often than expected by chance. To this end, association measures such PMI, *t*-score, or log-likelihood are applied.[31] All these measures have in common that they compare the likelihood of two words co-occurring with the likelihood of the respective individual items. However, they differ in the specific calculations they perform, which makes each of them suitable for different purposes. For instance, PMI favors words with a generally low frequency in a corpus but that tend to collocate with a particular node; it is therefore more adequate for the analysis of content words, as opposed to *t*-score, which performs better with function words (Church et al. 1991; Church et al. 1994; Liu 2010).[32] Studies on near-synonyms using association measures are Church et al. (1991), Gries (2001; 2003), Arppe (2002), Kjellmer (2003), and Arppe & Järvikivi (2007), among others. The use of association measures undoubtedly meant an improvement over that of mere frequencies and percentages, as the former allow to distinguish between those words

[31] Cf., for instance, Evert 2005; Levshina 2015: Chapter 10 for more information about particular association measures.

[32] PMI is further discussed in Section 7.1, where it is used to measure collocational strength.

that significantly collocate with a node from those that merely have a high overall frequency in a corpus, and therefore also co-occur frequently with the node. This is the main motivation for the adoption in most studies in lexical semantics nowadays of the statistical orientation to collocation (Sinclair 1991: 170).

The main goal of collocational analysis is certainly to establish the collocational preferences of lexical items, in this case of particular near-synonyms, and whether they differ or not in this respect. However, by recurring to collocational analysis it is also possible to extract additional and valuable information about the node words. For instance, collocational patterns can serve as indicators of the distribution and frequencies of different senses of lexical items. A case in point is Taylor's (2003) study of the adjectives *high* and *tall.* While these two near-synonyms can be used in the same senses, both literal ones pertaining to the spatial domain (e.g. *high*/*tall building*) and non-spatial metaphorical ones (e.g. *high temperature* or *tall order*), a collocational analysis of the two terms reveals that *high* is much more frequent than *tall* in the latter sense. Moreover, collocational analysis can also serve as an indicator of the syntactic function of node words. Gries (2001) points out that by analyzing the R1 collocates of adjectives, it is possible to ascertain whether they are used attributively or in other functions, most commonly predicatively. He shows that attributive adjectives are frequently immediately followed by a noun while predicative uses can be identified because other POS, particularly function words, occur directly after them in textual units. In this way, Gries is able to identify significant differences in syntactic distribution between particular pairs of *-ic* and *-ical* adjectives. An example concerns the pair *politic* and *political* as the former is often followed by the function words *to*, *for*, and *not*, and is thus frequent in predicative function, whereas the latter is hardly ever followed by collocates other than nouns (Gries 2001: 88).

Many of the studies on collocational behavior that have been discussed so far also examine stylistic differences between near-synonyms. In fact, as pointed out in the previous section (cf. Section 3.1), style is one of the most important dimensions of meaning in which synonyms differ, and synonyms in emotive areas of the lexicon are therefore particularly abundant (Cruse 1986: 284–285). Analyses such as those by Persson (1989), Biber, Conrad & Reppen (1998: Section 2.6), Kjellmer (2003), Arppe (2008), Liu (2010), and Liu & Espino (2012) examine the stylistic preferences of pairs or larger groups of near-synonyms, mainly to determine whether they differ in formality. To this end, the distribution of near-synonyms in different text-types or genres is typically explored. For example example, on the basis of data from the *Corpus of*

Contemporary American English or *COCA* (Davies 2008–), which comprises five different text-types showing different degrees of formality, to wit, spoken, fiction, newspaper, magazine, and academic writing, Liu (2010) establishes a cline of formality for the near-synonymous adjectives *main*, *major*, *chief*, *principal*, and *primary*, with *main* being the most informal and *primary* the most formal.

While most corpus-based studies on near-synonyms during the 1990s and early 2000s exclusively examined collocational and/or stylistic preferences, a majority of studies from the second wave analyze a wider range of factors and therefore go far beyond previous corpus-based research on synonymy. Again, Divjak (2006) and Divjak & Gries (2006; 2008) represent a clear break with previous analyses, this time in the sense that they examine a large number of factors.[33] In fact, Divjak & Gries proposed a new approach to the study of semantic phenomena, mainly synonymy and polysemy, which they called the *Behavioral Profile* (BP) approach, following Hanks (1996: 79), the first author to use this label to refer to the typical patterns and contexts in which a given word is used. In brief, the BP approach consists in analyzing sets of near-synonymous words or senses of a polysemous word by considering many different types of co-occurrence information (e.g. morphological, syntactic, semantic, and stylistic) in order to determine their conventional uses. This is done by means of different types of statistical techniques, such as correlations, HAC, and logistic regression (Gries & Divjak 2009; Gries 2010).[34] In one of the earliest BP studies, namely Divjak & Gries (2006), referred to earlier on in this section, they analyze the distributional patterns of nine Russian verbs of trying, including 87 different variables —or as they call them, ID-tags, following Atkins (1987)— pertaining to the domains of semantics and morphosyntax. For instance, among the morphosyntactic contextual clues, they consider verb related characteristics, such as aspect, mood, and tense, as well as clause and subject related characteristics, such as sentence type and clause or subject structure. Semantic co-occurrence information, on the other hand, refers to features concerning the nominative subject paradigms, that is, oppositions such as animate vs. inanimate or concrete vs. abstract. Moreover, they analyze the adjacent verb collocates of the nine near-synonyms by grouping them into semantic categories such as PHYSICAL ACTION, PHYSICAL PERCEPTION,

[33] Even though some earlier studies (e.g. Hanks 1996; Biber, Conrad & Reppen 1998: Sections 4.2 and 4.3; Arppe 2002), also include some syntactic and morphological co-occurrence features, the amount of factors considered in these works cannot be compared to those by Divjak (2006) and Divjak & Gries (2006; 2008), who include 47 and 87 variables, respectively.

[34] These two papers constitute in-depth summaries of the BP approach. The reader is referred to them for further details on the steps and procedures used, as well as for examples of variables to be included in such analyses.

SPEECH, and INTELLECTUAL ACTIVITY, among others, thus paying special attention to the complementation patterns and argument structures of the verbs under analysis. As mentioned above, the results point to the existence of three clearly delineated clusters or groups within this synonym set, each of them containing three of the verbs. This study proved the usefulness of the BP approach for the delineation of the internal semantic structure of near-synonyms and can be said to have paved the way for further BP studies along the same lines on other POS: adjectives (e.g. Gries & Otani 2010; Liu 2010; Yatandu Uba 2015), nouns (e.g. Liu 2013b), and adverbs (e.g. Liu & Espino 2012).

The BP approach is not the only existing corpus-based distributional method in cognitive semantics that has been used to analyze a wider range of contextual factors that determine the choice between near-synonyms. Other methods can be found which share the same main goals, but differ in the specific procedures employed along three stages distinguished by Levshina (2011: 24–29). These stages are data collection, data exploration, and confirmatory testing. Within the first stage (i.e. data collection), we find methods in which the data is collected manually, as in most BP studies, as opposed to automatically, as in SVS modeling (e.g. Heylen et al. 2008). SVS can be used to model the behavior of polysemous or near-synonymous words by analyzing their collocational patterns with the help of association measures. In the second stage (i.e. data exploration), the data is explored. Depending on the method adopted, different statistical techniques can be used to visualize the results. For instance, BP studies tend to use dimensionality-reduction techniques, such as HAC, Multidimensional Scaling or MDS for short (Wickelmaier 2003; Levshina 2015: Chapter 17), and (multiple) correspondence analysis (e.g. Desagulier 2014; Krawczak 2018). Finally, in the third stage (i.e. confirmatory testing) statistical tests are applied to determine whether the results can be generalized to other data. Logistic regression analysis is most commonly employed to see which variables and variable interactions have a statistically significant effect on the choice between near-synonyms (e.g. Arppe 2008; Krawczak 2014; 2018), but other tests can also be used, for instance, HCFA (e.g. Liu 2010; Liu & Espino 2012; Liu 2013b). Many BP studies do not include this last stage, mainly due to the extremely large number of variables to be analyzed, and are therefore of a more exploratory nature (but cf., for instance, Divjak 2010 for an exception). Some of the techniques mentioned here are further explained in Chapters 6 and 7, where they are used to examine the near-synonymous adjectives from the olfactory domain *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*.

3.2.2 Near-synonyms and POS: Studies on adjectives

The distributional corpus-based research reviewed in the previous section reveals that synonym sets belonging to different POS have been analyzed over the last few decades, including verbs (Schmid 1993; Hanks 1996; 2013; Biber, Conrad & Reppen 1998: Section 4.3; Arppe 2002; 2008; Divjak 2006; 2010; Divjak & Gries 2006; 2008; Arppe & Järvikivi 2007; Chung 2011), nouns (Schmid 1993; Liu 2013b), adjectives (Persson 1989; Church et al. 1991; Biber, Conrad & Reppen 1998: Sections 2.6 and 4.2; Partington 1998: Section 2.4; Gries 2001; 2003; Taylor 2003; Gries & Otani 2010; Liu 2010; Krawczak 2014; 2018; Yatandu Uba 2015), adverbs (Partington 1998: Section 3.6; Liu & Espino 2012; Desagulier 2014), and prepositions (Kennedy 1991). As such, a wide variety of different types of factors from different linguistic levels (e.g. semantic, syntactic, and morphological) have been examined in order to determine their effect on the choice between particular near-synonyms. However, not all factors are equally relevant for all synonym studies, since the way synonyms differ from one another depends largely on their POS and conceptual nature. Thus, most studies on near-synonymous verbs, besides including collocational patterns and semantic features in the analysis, also consider a great amount of morphosyntactic contextual clues, for instance, tense, aspect, and mood, or sentence and clause type in which the verbs occur. In fact, in many of these cases the morphosyntactic variables have proven to be of utmost importance to distinguish between near-synonymous verbs (cf., for instance, Divjak 2010: 183–193). On the other hand, studies on adjectives and adverbs generally consider fewer morphosyntactic factors, whereas semantic features seem to play a more crucial role in determining the choice between near-synonyms belonging to these POS, for instance, semantic preference regarding the elements they modify (e.g. Gries 2001; Liu 2010; Liu & Espino 2012). In what follows, studies on near-synonymous adjectives are dealt with in more depth, concentrating on the types of factors that are typically included in these analyses, as such determinants are the most relevant for our purposes in the present dissertation.

A great majority of the existing distributional corpus-based research on pairs or sets of near-synonymous adjectives consider their collocational and/or stylistic behavior only. More specifically, as has already been mentioned in Section 3.2.1, such studies tend to focus on the R1 collocates of the adjectives in order to concentrate on the nouns they modify. This is so because, as has been demonstrated in previous research on adjectives in general, analyzing the nouns that adjectives modify is one of the best ways to reveal the nature of the semantic content

of the adjectives, and modified nouns are therefore considered to be more informative collocates (cf. Geeraerts 1986: 282–284 on the Dutch adjective *vers* 'fresh' in Section 2.3.2). However, other studies go beyond mere individual collocates and consider other features of the modified nouns, including primarily semantic aspects, but also morphological ones. For example, as mentioned earlier, Gries (2001; 2003) includes information concerning the dichotomies concrete vs. abstract and specific/basic level vs. general/superordinate level of the modified nouns of *-ic* and *-ical* adjectival pairs, but he also contemplates that of animacy (i.e. animate vs. inanimate) for those adjectives in which such a division is relevant, as in the case of *optic* vs. *optical*. In his study on *chief*, *main*, *major*, *primary*, and *principal*, Liu (2010) discovers a division of semantic labor between the five adjectives by classifying their noun collocates into six semantic categories:[35]

(i) ABSTRACT: *change*, *problem*, and *reason*.
(ii) CONCRETE: *dish*, *entrance*, and *street*.
(iii) DUAL (can be either concrete or abstract): *character*, *component*, and *source*.
(iv) INSTITUTION: *city*, *corporation*, and *school*.
(v) POSITION TITLE: *deputy*, *investigator*, and *officer*.
(vi) NON-POSITION TITLE: *author*, *owner*, and *sponsor*.

By doing so, Liu concludes that whereas dual and abstract nouns are often modified by all five near-synonymous adjectives, the other four semantic categories tend to be dominated by one or two of them: for instance, while concrete nouns prefer *main*, position title nouns are dominated by *chief*. Furthermore, as all five adjectives seem to converge in meaning when they modify dual and abstract nouns, Liu investigates whether other fine-grained differences exist between them when used with nouns belonging to these two specific semantic categories. To this end, he explores two morphological features of the modified nouns or noun phrases in general, to wit, number (singular vs. plural) and definiteness (indefinite vs. definite). This is relevant for the adjectival synonyms object of study in order to establish the degree of importance that they convey: *one main issue* is more important than *two main issues* and *the main issue* is more important than *a main issue.* Thus, Liu identifies a cline of importance, with *primary*, *chief*, and

[35] For further information about the nouns belonging to each category and the decisions regarding specific nouns, see Liu (2010: 84).

main being used to refer to entities that are considered to be more important by speakers, followed by *major*, and lastly *principal*.

By far the most comprehensive work on synonymous adjectives is the BP study conducted by Gries & Otani (2010), in which they examine two sets of near-synonymous from the semantic domain of SIZE, namely *big*, *great*, and *large*, on the one hand, and *little*, *small*, and *tiny*, on the other. They analyze a great amount of factors from different linguistic levels, including morphological, syntactic, and semantic ones. Among morphological factors, they consider the aspect, tense, voice, and transitivity marking of the finite verb of the clause in which the adjectives occur. As regards syntactic factors, these are similar to those considered in BP studies on near-synonymous verbs, for instance, the type (i.e. main vs. dependent) and function (e.g. direct object, noun phrase postmodifier) of the clauses where the adjectives are located. This syntactic information is automatically retrieved, as Gries & Otani use the British component of the *International Corpus of English*, which is a parsed corpus. Another syntactic feature that they examine, which is also present in other studies on adjectival near-synonyms (e.g. Biber, Conrad & Reppen 1998: Section 2.6; Gries 2001; 2003; Liu 2010), is the syntactic function of the adjectives, that is whether they are used in attributive, predicative, or adverbial function. Finally, Gries & Otani analyze several semantic features of the noun collocates, such as countability (count vs. non-count) and a wide range of semantic categories including, among others, CONCRETE, ABSTRACT, HUMAN, ORGANIZATION/INSTITUTION, and QUANTITY. Another semantic feature they consider is how SIZE is modified, i.e. literally, metaphorically, quantitatively, or evaluatively. The results for each of the variables are then aggregated by means of HAC and visualized with the use of dendrograms. Findings point to a clustering solution according to different parameters, including both sameness of meaning, so that *tiny* and *smallest* are grouped together, and oppositeness of meaning, with *big* and *little* forming one cluster, and *large* and *small* another. Although Gries & Otani's (2010) study includes a wide range of different types of factors, the effects of each of the levels of the variables are not discussed. This is so because it is practically impossible to statistically test for all the variables included in their analysis without an extremely large amount of data for each of the six adjectives at issue. This is, in fact, a recurrent problem in many BP studies, not just the one carried out by Gries & Otani, as the third step of analyses of this type, i.e. testing for significance, is unviable (cf. Section 3.2.1 above). Therefore, even though such studies provide valuable results concerning the general behavior of the synonyms examined, specific details

about exactly how they differ from one another are seldom offered. However, it must be borne in mind that the goal of many BP studies is not that of giving an in-depth description of each of the variables explaining the choice between near-synonyms, but rather that of presenting a bird's-eye view of their overall contextual behavior.

Finally, some studies on near-synonymous adjectives, besides considering intralinguistic factors pertaining to their co-occurrence preferences, also examine the effects of extralinguistic variables. This is the case of, for instance, Krawczak (2014; 2018), who analyzes adjectives from the semantic domain of SHAME. Her 2014 paper examines three lexical items, namely *ashamed*, *embarrassed*, and *humiliated*, in both AmE and BrE in order to verify Wierzbicka's (1992; 1999) claims as to their semasiological and onomasiological structures. To this end, Krawczak includes three different semantic features, to wit, cause of emotion (e.g. bodily causes, insecurity, social failure), type of emotion (i.e. internal vs. external), and temporal scope of the emotion (present, past, and general). Moreover, she includes the factor 'dialect' to ascertain whether there exist extralinguistic usage differences. By means of two multivariate methods, namely correspondence analysis and multinomial logistic regression, she establishes three clearly delineated usage-profiles for the three adjectives: *ashamed* is connected to internal and atemporal causes related to insecurities of social status, emotional or bodily problems, and social failures; *embarrassed*, in turn, is associated with interactive factors, for instance, self-esteem difficulties or deviation from social conventions regarding politeness; finally, *humiliated* is linked to damages to one's social status due to external motivations. With regard to the extralinguistic variable examined (i.e. dialect), no significant differences are found, as the lexical items display basically the same behavior in the two reference varieties of English.

Krawczak's (2018) study constitutes a continuation of the author's previous research. She adopts a cross-linguistic and cross-cultural perspective to the same social emotion, namely SHAME, but restricts the analysis to two of the adjectives: *ashamed* and *embarrassed*, and their counterparts in French (i.e. *honteux* and *embarrassé*) and Polish (i.e. *zazenowany* and *zawstydzony*). The same factors are examined, with the exception of dialect, which is replaced by country and language, and a variable coding for whether an audience is present or absent from the speech situation in which the social emotion emerges. The findings point to the existence of a cline of communities from Poland through France to the UK and the US that is linked to their respective cultures and, more particularly, to whether the societies are more collectivist, as in the case of Poland, or more individualistic, as in the case of the UK and the

US. France, in turn, occupies an intermediate position. Krawczak thus finds a significant effect of the extralinguistic factors on the choice between the two near-synonymous terms in the different languages. Consequently, although the concept EMBARRASSED in general is more interactive than that of ASHAMED, being closely linked to the presence of an audience, this is not the case in Polish, where the line between the two emotions is not as clear-cut. Krawczak's (2018) findings may indeed be relevant for the present dissertation since, in addition to intralinguistic factors of a primarily contextual nature, extralinguistic ones may need to be considered in order to reach a full understanding of onomasiological structure.[36]

As we have just seen, some studies resort to sociolinguistic and cultural aspects to explain onomasiological variation. However, many of the works that have been reviewed throughout this section, despite offering valuable descriptive information about differences between near-synonymous expressions, do not connect the patterns uncovered to a specific theoretical framework, nor do they discuss the implications of their findings. As mentioned in Section 2.3.2, Geeraerts (2010: 177–178) states that the distributional corpus-based approach is ultimately a method rather than a theory, given that it is not always evident how the distributional patterns relate to theoretical issues in lexical semantics. Similarly, Gries (2010: 324–325) points to three areas in which this line of research could be further improved, one of which has to do precisely with the lack of theoretical background in many of these studies.[37] Therefore, such distributional methods have recently tended to adopt the framework of cognitive semantics and make use of concepts and principles within this theory, such as prototypicality and entrenchment, to explain the findings obtained.

One rather early corpus-based study on near-synonymy which provides a cognitive explanation for its findings is Taylor (2003) on the adjectives *high* and *tall*. Both terms are positive polarity items indicating the extent of an entity on the vertical dimension. However, while *high* dates back to the Old English (OE) period, having been inherited from Germanic, the spatial sense of *tall* is rather recent, approximately from the mid-16th century according to

[36] As mentioned in Section 2.4.2, many other studies on usage-based onomasiological variation with a sociolinguistic orientation have been conducted in the domain of cognitive semantics over the last few decades, which demonstrate the usefulness and importance of extralinguistic variables related to cultural and social differences (cf., for instance, Speelman, Grondelaers & Geeraerts 2003; Levshina 2011; Soares da Silva 2013; 2014).

[37] The other two areas mentioned by Gries (2010: 324–325) concern (i) the number and range of synonyms object of study and (ii) issues related to data and methodology, more specifically to the type of co-occurrence information of the synonyms taken into consideration.

the *OED* (*OED*, s.v. *tall* n. II, 6). The adoption of *tall* to denote an entity of great vertical extent did not imply a decrease in the usage range of *high*, which was retained for all types of entities with which it was used before. The adjective *tall*, on the other hand, offered a new way of conceptualizing positive polarity verticality that is specific to the human body (e.g. *tall man*, *tall girl*). In order to explain this distribution, Taylor makes use of the so-called vantage theory (e.g. MacLaury 1987; 1995; 1997), originally applied to the semantic domain of COLOR, in which a distinction is made between *dominant* and *recessive* terms. A dominant term within a pair is that which can be applied to a wide range of entities or situations, as it emphasizes the similarities between them. On the other hand, a recessive term emphasizes their differences, and as such can only be applied to a narrow range of entities or situations (Taylor 2003: 280). By using this distinction, Taylor concludes that *high* is the dominant term in the pair, while *tall* is the recessive one, as it is limited to denote a specific type of verticality, i.e. that of the human body. Taylor (2003) thus demonstrates how findings from distributional corpus-based studies on near-synonymy can be embedded within a theoretical framework on language and human cognition in general, and lexical semantics, in particular.

Even though Taylor (2003) briefly considers the diachronic development of *tall* and *high* in his discussion of the findings, the rest of the studies reviewed until now adopt a fully synchronic orientation towards lexical onomasiological variation. The reason for this is the scarcity of diachronic studies on lexical near-synonymy from the perspective of usage-based onomasiology with a cognitive stance. The next section deals with the scant diachronic corpus-based analyses on pairs or sets of lexical near-synonyms, as well as with some generalizations about synonymy and diachrony that have been proposed in the specialized literature, including the types of semantic change that near-synonymous words or expressions have been said to undergo over time.

3.3 SYNONYMY AND DIACHRONY

As mentioned in Section 2.4.2, both synchronic and diachronic approaches to the study of meaning abound within cognitive semantics. However, while a considerable amount of diachronic work within this tradition has recently been devoted to constructional synonymy and other morphosyntactic changes (e.g. Shank, Plevoets & Cuyckens 2014; Hilpert 2016; De Smet et al. 2018; Breban & De Smet 2019; D'hoedt, De Smet & Cuyckens 2019), we find only a

handful of diachronic studies on lexical synonymy from a distributional corpus-based perspective. In what follows, a brief summary of this research is provided.

Kaunisto (2001) investigates the evolution of twelve pairs of *-ic* and *-ical* adjectives from the second half of the 16th century to the first half of the 19th century by drawing on data from the prose section of *Chadwyck-Healey Literature Online* (1996–2000): *angelic*(*al*), *authentic*(*al*), *comic*(*al*), *domestic*(*al*), *fantastic*(*al*), *heroic*(*al*), *magic*(*al)*, *majestic*(*al*), *philosophic*(*al*), *poetic*(*al*), *tragic*(*al*), and *tyrannic*(*al*). First, he analyzes the productivity of each of the two suffixes from the 1300s to the present-day by counting the number of first attestations of *-ic*/*-ical* adjectives in the *OED*. By doing so, Kaunisto is able to determine that, whereas until the 17th century both suffixes are approximately equally productive, from 1750 and until about 1900 *-ic* is used considerably more often to coin new adjectives, and thus seems to be the dominant suffix throughout this period, especially in scientific domains such as chemistry (e.g. *acidic*, *cyanic*). Nevertheless, in the 20th century the productivity of both suffixes becomes similar again. As regards the specific diachronic patterns of pairs of *-ic*/*-ical* near-synonyms, Kaunisto identifies five evolutionary trends:

(i) First, in some pairs, which are all characterized by exhibiting the semantic feature NOBILITY (e.g. *angelic*(*al*), *heroic*(*al*), *majestic*(*al*)), there is a move from *-ical* to *-ic* mainly in the 17th century. Moreover, in previous periods, when adjectives with both suffixes are common, there appears to be no change in meaning or function between them, as they are used with the same senses and in the same syntactic functions and even seem to collocate with the same types of nouns.

(ii) Second, pairs such as *comic*(*al*), *fantastic*(*al*), and *magic*(*al*) come to be more often used with *-ic* in the 18th century.

(iii) Third, in the case of *tyrannic*(*al*), the number of occurrences with *-ic* increase as in the preceding two cases, but the *-ical* form remains the most frequent.

(iv) Fourth, *philosophic* becomes more common at the expense of *philosophical* in the second half of the 18th century, when the two near-synonyms seem to undergo a process of differentiation: the former has a more popular meaning (e.g. *philosophic look*), while the latter occurs mainly with nouns referring to science, such as *research* and *theory*. Nevertheless, the two terms then converge again in meaning and *philosophic* became the dominant adjective in both uses.

(v) Finally, *tragic*(*al*) undergoes a change from *-ical* to *-ic*, but much later than the rest of the terms discussed, after the first half of the 19th century.

Thus, in almost all cases we witness a shift from *-ical* to *-ic*, although at different points in time. In some *-ic*/*-ical* pairs semantic differentiation takes place in a given period. However, Kaunisto concludes that, in the end, the main result of these changes seems to be that of discarding one of the options rather than maintaining both forms with different meanings and/or functions.

Primahadi-Vijaya-R. & Rajeg (2014) examine the nominal collocational profiles of two near-synonymous adjectives from the domain of TEMPERATURE, namely *hot* and *warm,* in AmE throughout the last one and a half centuries (i.e. 1860s–2000s) by drawing on data from *COHA* (Davies 2010–). To this purpose, they extract the top 100 R1 noun collocates of the two adjectives, excluding those that appear less than five times in each decade in the corpus. The data is visualized by means of motion charts (Hilpert 2011; 2013: 66–74), which is a method to display a series of graphs in order to plot the diachronic changes of a given phenomenon. Primahadi-Vijaya-R. & Rajeg show, for each decade in *COHA*, the absolute co-occurrence frequencies of the different R1 noun collocates and identify several diachronic trends that have contributed to the differential use of these two near-synonyms in present-day AmE. Whereas in the 1860s some nouns, for example, *blood*, *water*, and *weather*, are commonly modified by both *hot* and *warm* in the literal sense 'of high temperature', other nouns are already clearly associated with one of the two adjectives. For instance, *warm* typically collocates with nouns such as *smile*, *heart*, *affection*, and *friend*, thus pointing to its frequent use in the metaphorical extension of 'friendliness', although this sense seems to decrease in frequency over time, as shown by the decline of *warm* with some of these nouns (e.g. *heart* and *welcome*). On the other hand, *hot* seems to be typically used in the 1860s in another metaphorical sense, viz. that of 'excitement' and 'intensity', as in *hot haste* or *hot pursuit*. With the passing of time, the collocational profile of *hot* changes somewhat. First, from the 1920s onwards, it undergoes a lexicalization process with the noun *dog* (i.e. *hotdog*), which is today a compound noun rather than a combination of an adjective and a head noun. This is also the case of *hotspot*, in which the original meaning of *hot* is no longer transparent. Second, from the 2000s onwards, *hot* becomes strongly associated to nouns denoting people (e.g. *girl*, *guy*, and *woman*) in order to refer to 'sexually attractive' individuals, thus indicating a rise in this particular sense of the adjective. Primahadi-Vijaya-R. & Rajeg (2014) thus uncover some interesting differentiating

patterns of the two near-synonymous adjectives and how they develop through time, especially concerning the lexicalization processes of the nouns *hotdog* and *hotspot*.

Baker (2017) uses another useful visualization method for diachronic collocational changes of lexical near-synonyms, to wit, collocational networks (Brezina, McEnery & Wattam 2015). Put briefly, this approach consists in plotting the most important collocates of lexical items in networks in which the collocates are connected to their nodes by means of arrows. The size of the arrow indicates the strength of association between the node and its collocates: the shorter the line, the stronger the collocation. Baker (2017: 95–101) adapts this method to diachronic analysis by plotting different collocational networks in separate periods, thus revealing how particular nodes attract but also lose collocates over time. In particular, he analyzes the near-synonyms *on* and *upon*, on the one hand, and *round* and *around*, on the other, in two corpora of the *Brown* family, namely the *Before-LOB Corpus* or *BLOB* (Leech & Smith 2005) and the *British English 2006 Corpus* (Baker 2009), which contain data from BrE from the year 1931 and from 2006, respectively. In the case of *on* and *upon*, Baker finds that with time *on*, which was already more frequent in *BLOB*, draws away collocates from *upon*, for instance, the adjective *dependent*. In other words, several collocates which were previously shared by the two near-synonyms or which were only associated with *upon* in the 1930s are exclusive to *on* in 2006. This particular development is considered to illustrate a linguistic process termed densification, whereby shorter words, clauses, sentences, or other discourse units are preferred to longer ones (Baker 2017: 24). By contrast, the opposite process is identified in the development undergone by the lexical items *round* and *around*: whereas in 1931 *round* has a greater amount of collocates than *around*, with time the latter attracts many new collocates and becomes the preferred variant in 2006. As shown by Baker, diachronic collocational networks are valuable assets in order to zoom in on the competition between near-synonyms as they help to determine which collocates are shared by them at different points in time and to identify the potential variations in the relation between the near-synonyms. Therefore, this approach allows for the examination of changes at a lower level of granularity.

Finally, Pettersson-Traba (2018) conducts a pilot study on the diachronic development of the attributive uses of an adjectival synonym set which designates the concept PLEASANT SMELLING, i.e. *fragrant*, *perfumed*, *scented*, and *sweet-smelling*, in the latter part of Late Modern and present-day AmE. By drawing on data from *COHA* and applying an HCFA, she delineates the internal semantic structure of this set of near-synonyms, paying special attention to their

noun collocates, which are classified into nine semantic categories: BODY, CLEANING, EARTH, ESTHETICS, FOOD AND DRINK, MATTER, PLANTS AND FLOWERS, SENSATION, and TEXTILE AND CLOTHING. The results show that the concept PLEASANT SMELLING underwent major changes over the time span examined (1850–2009), from being used mostly to qualify entities which can exhibit a natural pleasant smell (e.g. flowers and trees) to modifying objects which are artificially sweet-smelling (e.g. oils and shampoos). This change is hypothesized to be a result of extralinguistic factors, to wit, socio-economic changes such as industrialization and mass production that took place during the period examined, in particular at the end of the 19th century, which led to an ever-increasing need to refer to artificially scented lotions and candles rather than naturally fragrant plants. Moreover, *fragrant* and *perfumed*, which were initially the most frequent adjectives, are gradually replaced by *scented*, thus reflecting a change in the relation between the near-synonyms over time. The present dissertation constitutes a continuation of the line of research initiated in Pettersson-Traba (2018), which is here considerably extended and refined.[38]

Though not a study on near-synonymy, it is worth mentioning here Jansegers & Gries' (2020) work on the semantic evolution of the polysemous verb *sentir* 'feel' in Peninsular Spanish, as it constitutes the first diachronic application of the BP approach, thus providing an interesting methodological avenue in distributional lexical semantics. Jansegers & Gries (2020) follow the ordinary steps of the BP approach (cf. Section 3.2), but repeat the procedure for each of the five historical periods that they examine, namely the 13th, 15th, 17th, 19th, and 21st centuries.[39] As in the case of Primahadi-Vijaya-R. & Rajeg (2014), they use motion charts to visualize their results, but do so by means of a more sophisticated statistical technique, namely MDS. This is a dimensionality-reduction technique common in many fields of linguistics to visualize, in a simple and straightforward way, the degree of (dis)similarity between two or more objects, either varieties of English (e.g. Szmrecsanyi & Kortmann 2009), constructions (Levshina 2011), or senses, as is the case in this particular study. The (dis)similarity between the objects is computed on the basis of a series of parameters selected by the researcher(s), for instance, a series of contextual cues of either a semantic or a morphosyntactic nature. The output of MDS is a two- or —in some cases— three-dimensional map on which these objects are situated, either close to or further apart from one another, on the basis of how (dis)similar they

[38] For additional preliminary work on this synonym set, cf. also Pettersson-Traba (2019; 2020a; 2020b; forthcoming).

[39] For further details about the data selection and extraction process, cf. Jansegers & Gries (2020: 149–150).

are in terms of the parameters selected for inclusion in the model: the closer two objects are to each other on the MDS map, the more similar. Following Hilpert (2011; 2013), who has recently adapted this technique to diachronic data in order to visualize language change, Jansegers & Gries (2020) use MDS instead of HAC, the more typical clustering technique in BP studies, which is employed in Divjak (2006), Divjak & Gries (2006; 2008), and Gries & Otani (2010), among others. The reason for this methodological shift is that MDS is semantically more realistic than HAC, since it does not suppose a categorical either/or split between individual clusters, but instead displays a more precise and continuous cline of semantic (dis)similarity (Jansegers & Gries 2020: 158). A further benefit of the application of MDS to diachronic data is that it makes it possible to compare not only one particular element with others, but also that same element with itself at different points in time. In this way, Jansegers & Gries (2020) discover the evolution of the three different broad semantic domains of *sentir*, namely direct physical, cognitive, and emotional perception. First, they identify the distributional patterns of these three uses of the verb in different periods and then trace their frequency development over time, concluding that *sentir* evolves from being used primarily as a physical perception verb in the earliest periods, specifically in the 13th and the 15th centuries, to being used mostly in the emotional sense in later stages. Moreover, the MDS maps also display an interesting development in recent times, namely the emergence and rise of *lo siento* 'I'm sorry' as a discourse marker. In this way, the MDS maps also serve to ascertain from which particular senses of *sentir* this discourse marker originates, namely its use with the meaning 'regret', belonging to the emotional perception domain, on the one hand, and its cognitive use 'consider, judge', on the other. Consequently, Jansegers & Gries (2020) demonstrate the usefulness of examining the diachronic evolution of distributional patterns of lexical items by means of a technique which had until then been primarily used for morphosyntactic research, both of a synchronic and a diachronic nature.

The usage-based studies reviewed in this section so far show that lexical semantic relations do not, in many cases, remain stable over time, neither when it comes to the association between semantically related words such as near-synonyms nor in the case of the senses of a particular polysemous word. In other words, by using a distributional corpus-based approach, both onomasiological and semasiological changes can be successfully tracked over time, and this method can even help reveal their possible underlying motivations. Nevertheless, the idea that modulations of meaning occur over time is a well-established one in the specialized literature,

and much research has been devoted to the identification of the principal mechanisms of semantic change, as well as to classifications of such regular patterns into different types. This reminds us of the works mentioned in Section 2.1 in relation to Carnoy (1927) and Stern (1931), who are undoubtedly pioneers of this line of enquiry, but also to the work of Ullmann (1962). The types of changes these authors identify are commonly only discussed in relation to individual words and their semasiological evolution. However, many of these semantic changes are relevant also for the diachronic evolution of the internal semantic structure of synonyms if an onomasiological perspective is adopted.

As mentioned earlier in this chapter (Section 3.1), a well-established belief in the specialized literature is that languages work against absolute synonymy and that this leads to competition between semantically related words to ensure that what has been labeled the no synonymy rule or isomorphic state is maintained (e.g. Bolinger 1977: ix–x, 9; Wierzbicka 1988: 13–14; Croft 2000: 176; Nuyts & Byloo 2015: 62–63). As Samuels (1972: 65) argues, "if […] two exact synonyms exist for a time in the spoken chain, either one of them will be less and less selected and eventually discarded, or a difference of meaning, connotation, nuance or register will arise to distinguish them". In other words, synonyms are expected to become functionally more dissimilar over time, thus undergoing processes of substitution or differentiation. In the former case, one of the alternatives progressively takes over more and more of the original shared semantic ground and, consequently, the other variant becomes less frequently selected until it falls into disuse. An example drawn from the history of English concerns the lexical items *guard* (*OED*, s.v. *guard* n. 7a) and *ward* (*OED*, s.v. *ward* n.1), which shared the sense '[a] person who keeps, protects, or defends; a watchman' during the latter part of the Midde English (ME) period. *Ward* has its origins in the OE noun *weard* and was used in this sense until the end of the 15th century, as its last attestation from the year 1472 in the *OED* demonstrates. Interestingly, this is precisely the same century in which the noun *guard*, coming from French *garde*, was borrowed into English with the same meaning. The two terms coexisted in ME for a while, but eventually *guard* replaced *ward*, as the latter fell into disuse in this sense. In the case of differentiation, on the other hand, one of the synonyms will experience some change in meaning, such as *specialization* or *generalization* (cf., for instance, Kay & Allan 2015: 75–76). Specialization, also termed *narrowing*, refers to the process whereby a word or a particular word sense becomes increasingly narrower or more specific, so that the number of referents which the word can denote is reduced. On the other hand, generalization, sometimes

also referred to as *widening* or *broadening*, is the opposite process, that is, when a word or a word sense becomes increasingly broader or more general, thus coming to designate a wider number of referents. The nouns *bird* and *fowl*, which could at one point in history be used interchangeably, exemplify the two processes. *Bird*, which dates back to OE, was originally used to denote a specific type of bird, namely a young one (*OED*, s.v. *bird* n. 1a). Nevertheless, in the ME period it underwent a process of generalization and came to be used to refer to any type of bird, including both young and adult members of the species. In turn, *fowl*, of Germanic origin, was a general term during the OE period, being used for '[a]ny feathered vertebrate animal' (*OED*, s.v. *fowl* n. 1a), much as its present-day counterparts in other Germanic languages (e.g. German *Vogel* and Swedish *fågel*). Once *bird* began to be used in the more general sense, the two terms progressively became differentiated, as *fowl* underwent a process of specialization and was established with a narrower sense in the Early Modern English (eModE) period, namely that of '[a] domestic bird (e.g. turkey or hen)' (*OED*, s.v. *fowl* n. 3). Consequently, the relationship between *bird* and *fowl* can be characterized as hyponymy in OE and part of ME (*bird* being a hyponym of *fowl*), followed by a period of synonymy in ME and eModE, and then again hyponymy in Present-day English (PDE) (*fowl* being a hyponym of *bird*).[40]

Two additional and related semantic changes often mentioned in the literature are *amelioration* and *pejoration* (cf., for instance, Kay & Allan 2015: 77–81), which pertain to the connotative dimension of meaning explained earlier (cf. Section 3.1). Amelioration refers to a change from a more negative to a more positive meaning, while pejoration entails the opposite development, i.e. a change from a more positive to a more negative meaning. The PDE terms *quean* and *queen*, which can be traced back to the same Indo-European root, could be used with the meaning 'woman' in OE, thus being interchangeable sometimes, although *queen* was more commonly employed to denote a specific type of woman, namely a noble one. With time the nouns underwent differentiation, as *quean* became more negative (i.e. pejoration) and *queen* more positive (i.e. amelioration), so that already in ME the two terms were not regarded as synonymous. Nowadays, *quean* is used almost exclusively —except in Scottish English— to refer to a woman of lower status, namely a bold or impudent one, or a prostitute. *Queen*, on the other hand, though associated with positive undertones already in OE, has become further ameliorated over time, its predominant sense in PDE being that of 'a female ruler'. The case of

[40] For more information on hyponymy, cf. Section 2.2.

pejoration can be exemplified also by means of two nouns from the semantic domain of SMELL, to wit *stench* and *odor,* which were formerly synonymous with *smell*. Both nouns were originally used to denote any type of smell, be it an unpleasant or a pleasant one, but eventually they became increasingly associated with negative undertones and finally used either exclusively (in the case of *stench*) or almost exclusively (in the case of *odor*) to refer to disagreeable or even disgusting smells. The pejoration of *stench* took place already after the OE period (*OED*, s.v. *stench* n. 1 and 2). In the case of *odor* the change occurred only rather recently, but the process of pejoration can be said to be more dramatic, as the early uses of this noun in ME were typically of a positive nature, referring to sweet and pleasing aromas (*OED*, s.v. *odour* n. I. 1). These examples demonstrate that amelioration and pejoration do not exclude other types of semantic changes pertaining to the denotative domain, such as specialization, which is also present in the cases of *queen*, *quean*, *stench*, and *odor*.

The view of competition between synonyms just exemplified is recurrent in research on language change and it is believed to be an inherent trait of languages to aim at an isomorphic state. In fact, the literature is brimmed with examples of synonymous expressions —either words or constructions— that have ultimately become more differentiated, particularly via the process of specialization which, as argued by Kay & Allan (2015: 32), "[…] is one of the commonest forms of semantic change among words which start off as synonyms". During the period following the Norman Conquest of 1066, many Anglo-Saxon terms were duplicated by French loan-words with the same meaning. A well-known case is that of the pairs *pig*/*pork*, *cow*/*beef*, *deer*/*venison*, and *sheep*/*mutton*. Though originally absolute synonyms, over time the first noun (i.e. Anglo-Saxon) in each pair became specialized to denote only the living animal, while the second one (i.e. French) came to be used solely to refer to the flesh of those animals when conceptualized as food (e.g. Jackson 1988: 66; Murphy 2003: 161). In this case, the motivations behind the semantic shift are well understood. Throughout ME and (part of) eModE, French, the language of the ruling class, was more prestigious and thus associated with the more sophisticated spheres of life, such as that of "fine dining" (Murphy 2003: 161). Consequently, extralinguistic (i.e. social) factors steered the course of semantic change.

Anglo-Saxon and French pairings also serve as examples of synonyms becoming differentiated not concerning their denotative meaning, but as regards style. For instance, in the cases of Anglo-Saxon *begin*/*start*, *ask*, and *hearty* vs. French *commence*, *demand*, and *cordial*, respectively, the French loanwords came to be associated with formal and literary styles,

whereas the Anglo-Saxon counterparts are more colloquial or neutral and thus used in a wider range of stylistic contexts (Jackson 1988: 66–67; Kay & Allan 2015: 12).

Although the competition theory is widespread and has proved successful to understand many changes in language, not only at the lexical level but also at the syntactic, morphological, and phonological ones (cf., for instance, Mondorf 2010; Berg 2014), some scholars consider it an oversimplification, as it fails to account for situations in which synonyms become more similar over time, rather than more dissimilar, a process that has been labeled *attraction* (De Smet et al. 2018: 203). De Smet et al. (2018) provide evidence of several pairs of near-synonymous constructions that have come to share more semantic space across time. In one of their case studies, they show that the verb *begin* has come to be frequently complemented by *-ing*-clauses at the expense of *to*-infinitive clauses. In the 1840s *begin* was usually followed by *to*-infinitive clauses, with both agentive and non-agentive subjects, though particularly with the latter. Progressively, *-ing*-clauses became more frequent, at first only with agentive subjects, and later on also with non-agentive ones. *To*-infinitive complements simultaneously experienced a significant downward tendency in frequency, but did not retract from contexts with agentive subjects, as would be expected in a process of differentiation. On the contrary, the decrease in frequency of *to*-infinitives took place in non-agentive contexts, which led to the semantic profiles of the two constructions becoming more similar. Thus, while an instance of ongoing replacement of *to*-infinitive clauses by *-ing*-clauses, the development described here is also a clear case of attraction. The example provided by De Smet et al. (2018) suggests that a certain level of attraction is possibly a prerequisite for substitution, since synonymous expressions may need to share at least some functional features for replacement to take place.

De Smet et al. (2018: 204, 217) explain the process of attraction by means of analogical change, the process whereby one form becomes more similar to another which it already resembled, usually from a formal perspective (Trask 2007: 15–16). However, analogy is here understood in an extended sense to include also expressions which are functionally rather than formally or structurally similar (see also Nuyts & Byloo 2015: 36). De Smet et al. (2018) thus argue that expressions exhibiting functional similarity parallel each other's behavior through an interchange of characteristics. This theory of semantic parallelism is also evinced in the fact that various synonymous lexical expressions can occur in the same metaphorical mappings and thus come to share more senses over time (e.g. Stefanowitsch 2008; Turkkila 2014). For instance, as argued by Stefanowitsch (2008: 96–99), the nouns *joy* and *happiness*, which share

the literal sense 'the emotion or state of being highly pleased or delighted', have developed similar metaphorical mappings in which the emotion is described as a liquid (e.g. *river of joy/happiness*) or as a source of heat (e.g. *sparks of joy/happiness* and *burn with joy/happiness*). Another example concerns the near-synonymous adjectives from the domain of TASTE, for example, *delicious*, *delectable*, *luscious*, and *tasty*, which besides referring to 'foods of pleasant flavor', have all been metaphorically extended to 'very attractive' when used for people (e.g. *OED*, s.v. *tasty* adj. 1a and 1b). These cases suggest that with time near-synonyms can come to develop the same figurative senses, thus becoming more similar in meaning.

Additionally, De Smet et al. (2018: 205) claim that the processes of differentiation and attraction are mutually exclusive. While this seems to be the case when pairs of synonyms are considered, the argument does not necessarily hold if larger groups are taken into account. This is so since different members of a synonym set may hold diverse and sometimes opposite relations to one another. Consider a hypothetical set of near-synonyms with three members, *A*, *B*, and *C*. If we focus only on the diachronic development of *A* and *B*, over time they can become either more similar (i.e. attraction) or more dissimilar (i.e. differentiation), but not both processes at the same time.[41] However, if we take all the members of the set into account, *A* and *B* may become attracted, while simultaneously *B* and *C* may become differentiated. Therefore, by adopting a broader perspective and including larger groups of near-synonyms in the analysis, both types of changes can be found to be at work in one and the same synonym set. This is an issue that has, to my knowledge, not yet been explored, and that can be of crucial importance for lexical synonymy, where, as discussed in Section 3.2.1, near-synonyms tend to come in larger groups rather than in pairs.

3.4 SUMMARY

The present chapter has traced the semantic phenomenon of synonymy, dating back to the structuralist school, which proposes various classifications of this semantic relation. Although different types of synonymy were postulated, only the two-fold distinction between absolute and non-absolute or near-synonymy is relevant for our purposes in this dissertation. Section 3.2 presented an overview of the existing synchronic research on the internal semantic structure of pairs and sets of near-synonyms that has been conducted from a distributional corpus-based

[41] A third possibility is that of stability, that is, when no changes in the functional profiles of the near-synonyms occur.

perspective. Two waves of research were distinguished, which point to a constant evolution of the field, both concerning the number of near-synonyms and the factors included in the analyses, but also regarding other methodological aspects, such as the particular techniques used to measure semantic (dis)similarity. Then, an in-depth review of studies dealing with adjectival near-synonyms was provided, focusing on the types of factors that are typically considered and the most important findings of such studies. In the majority of cases, semantic factors emerged as the most significant ones to explain the choice between synonymous adjectives, outperforming morphosyntactic variables. Lastly, Section 3.3 was devoted to the few diachronic studies on lexical synonymy with a corpus-based distributional orientation, which demonstrate that the relation between particular near-synonyms is often unstable over time. Additionally, typical semantic changes of a semasiological nature such as specialization, generalization, amelioration, and pejoration were shown to be equally relevant in onomasiology, as these processes are also often resorted to in order to explain changes in the internal structure of synonym sets. It was seen that the prototypical view in the literature is that synonyms become differentiated with time if one is not displaced by another that takes over its functions. Nevertheless, recent research has shown that another possible outcome is for synonyms to progressively become more similar instead of more dissimilar, thus undergoing attraction.

The shortage of diachronic research on lexical synonymy shows that there is a clear need for more analyses of this nature. In this context, the present dissertation aims at partially filling this existing gap by analyzing in the recent history of AmE a set of adjectival near-synonyms from the olfactory domain, namely *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*. Chapters 5–7 present the results of corpus-based analyses of this synonym set. First, however, Chapter 4 introduces the synonyms object of study, as well as the reasons for the selection of this particular near-synonym set. Moreover, the data retrieval and annotation processes of the databases employed in the subsequent corpus-based analyses in Chapters 5–7 are also explained in detail.

4 THE CONCEPT PLEASANT SMELLING: DATA SELECTION AND ANNOTATION

As became evident in the previous chapter, although absolute synonyms are virtually nonexistent, languages abound in near-synonyms, that is, words with similar, though not identical meanings. This is particularly true for English, which, due to its long history of borrowing from other languages, such as French and Latin in ME and eModE, displays an exceptionally large number of roughly synonymous expressions (Kay & Allan 2015: 12, 31, 88). Consequently, the task of selecting one specific synonym for the analysis is not an easy one. As seen in Sections 3.3 and 3.4, where previous studies of individual pairs and sets of adjectives were reviewed, certain tendencies are observed, with some semantic domains and even specific synonyms receiving particular attention. Thus, for instance, several analyses have focused on *-ic* and *-ical* adjectives (i.e. Gries 2001; 2003; Kaunisto 2001) and others have paid attention to basic descriptive adjectives from the domains of SIZE and AMOUNT, namely Biber, Conrad & Reppen (1998: Section 2.6) on *big*, *large*, and *great*, Taylor (2003) on *high* and *tall*, and Gries & Otani (2010) on the following two sets: *big*, *large*, and *great*, on the one hand, and *little*, *small*, and *tiny*, on the other. This dissertation, however, sets out to examine a set of adjectival near-synonyms from the relatively understudied semantic domain of SMELL: *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*, which designate the concept PLEASANT SMELLING. Section 4.1 is devoted to a description of this synonym set, drawing on data from different dictionaries and thesauri, and to an explanation of the motivations for selecting this particular group of adjectives.

Different analyses conducted on this synonym set are presented in subsequent chapters (i.e. Chapters 5–7) on the basis of corpus data. Corpora are the mainstream data sources for research in basically all fields of contemporary linguistics, semantics and historical linguistics constituting no exceptions. For the purposes of the present dissertation and due to the relatively low frequency of some of the items in the synonym set object of study (cf. Section 5.1), a rather

large corpus is desirable in order to achieve a representative sample. The corpus used here is *COHA* (Davies 2010–), which contains more than 400 million words. A description of this corpus and its suitability for the analyses in this dissertation is provided in Section 4.2, together with the data retrieval process of the occurrences of the near-synonyms employed in the case studies in Chapters 5, 6, and 7. The resulting databases were annotated for a series of language-external and language-internal variables pertaining to the contextual features of each attestation of the five near-synonyms. This annotation process is described in depth in Section 4.3.

4.1 INTRODUCING THE SYNONYM SET

Five adjectival near-synonyms from the olfactory domain are the object of study of the analyses conducted in the present dissertation, namely *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*. These five items are exemplified in (11)–(15):

(11) Her feet, broad and solid from years of walking, easily passed over the tricky terrain of low shrubs, dead leaves, fallen trees, and trailing vines. It had rained a little last night, and the moist earth was **fragrant**. (COHA, 2009, FIC, WifeGodsNovel)

(12) Jake touched warm, breathing woman, inhaled her freshly bathed scent and found her primal essence beneath the **perfumed** soap and body lotion. (*COHA*, 2007, FIC, WolfTalesIII)

(13) Smoked Pork Chops with Apple-Red Chile Compote # You can think of this smoky entre from W. Park Kerr's "Burning Desires: Salsa, Smoke and Sizzle From Down by the Rio Grande," as pork chops with red-hot **scented** applesauce all grown. (*COHA*, 2007, NEWS, Denver)

(14) As his eyes adjusted ever so slowly to the gloom, he saw looming shadows, blurred shapes like enormous trees that stirred not at all in the gentle **sweet-scented** wind. (*COHA*, 2005, FIC, FantasySciFi)

(15) Tapping one slim cigarette out of the pack, she brought the **sweet-smelling** tobacco to her lips and searched blindly for her lighter, then leaned back in her chaise lounge on the patio of the villa she'd rented for the summer on the island of Saint Martin. (*COHA*, 2003, FIC, RoomService)

The main motivations for selecting these particular lexical items are explained in Section 4.1.1. Section 4.1.2, in turn, offers a review of the existing descriptions of these adjectives as provided in dictionaries and thesauri, concentrating specifically on their etymology, different senses, and examples of use.

4.1.1 Main motivations for the selection of the concept PLEASANT SMELLING

As mentioned in the introduction to this chapter, the semantic domain of SMELL has hitherto been relatively understudied in semantic research (but cf. Ibarretxe-Antuñano 1999; Digonnet 2018 for metaphor analyses on words referring to the sense of smell). A common claim in the literature is that humans, despite being able to identify and discriminate a wide range of smells, experience difficulties when having to name them (cf. Ibarretxe-Antuñano 1999: 36; Lorig 1999: 392;; Yeshurun & Sobel 2010: 216, among others). This can be one of the reasons why the odor vocabulary has not been as widely researched as other domains such as those of COLOR (e.g. Berlin & Kay 1969; Biggam 2010; 2012; Anderson & Bramwell 2014) and COOKERY (e.g. Lehrer 1969; 1974), or the lexicon used to describe other senses (i.e. hearing, vision, taste, and touch), which have been argued to be much more extensive and precise in English (e.g. Sperber 1975: 115–116; Digonnet 2018: 178–179). However, the *HTOED*, which provides a semantic classification of all the senses of words, lists a whole 904 word senses under the semantic category SMELL AND ODOUR, [42] which have been and/or are still used in English, a fact that appears to contradict previous claims about the extensiveness and precision of this semantic domain in English. Of these words, a great amount are adjectives, some of which are similar to those selected in the present study with some minor variations in definition or use, making this a particular interesting semantic field for the study of near-synonymy. Examples of these semantically related adjectives include *aromatic* (*OED*, s.v. *aromatic* adj. 1), *balmy* (*OED*, s.v. *balmy* adj. 3), *odoriferous* (*OED*, s.v. *odoriferous* adj. 1), *odorous* (*OED*, s.v. *odorous* adj.), *perfumy* (*OED*, s.v. *perfumy* adj), *redolent* (*OED*, s.v. *redolent* adj. 1 and 2), and *sweet* (*OED*, s.v. *sweet* adj. 2), among others, many of which are loanwords from French and/or Latin that entered English between the 15th and the 17th centuries (cf. also Durkin 2014: 413–414).

[42] Although AmE orthographic conventions are used in the present dissertation, the British spelling of the labels of semantic categories in the *HTOED* is retained in order to maintain the original nomenclature.

Second, as demonstrated in previous pilot studies on this synonym set (Pettersson-Traba 2015; 2018; 2019; 2020a; 2020b; forthcoming), there seem to be some intriguing diachronic developments regarding the relationship between the selected adjectives in the latter part of Late Modern and Present-Day AmE. By analyzing the noun collocates of the adjectives in attributive function, Pettersson-Traba discovers in these works a tendency for the adjectives to be increasingly used to modify artificial, as opposed to natural, smells. Additionally, *fragrant* and *perfumed*, which were initially the most frequent adjectives, are gradually replaced by *scented*, thus reflecting a change, which still seems to be ongoing, in the relation between the near-synonyms over time. The concept PLEASANT SMELLING therefore appears to be a particularly interesting and relevant object of study for the purposes of the present dissertation, which aims at uncovering the semantic processes that specific lexical synonyms undergo with time, as well as the underlying motivations for such processes.

Finally, another crucial reason for selecting this synonym set is its rather low degree of polysemy, which is advantageous for the innovative approach adopted here. Given that the main focus is on the semantic relation of synonymy, examining a set of lexical items that, besides being synonymous, are also highly polysemous would result in an extremely complex and almost unfeasible study. For example, analyzing a highly polysemous synonym set would require a great amount of manual pruning in order to ensure that the final datset contained exclusively those instances which are truly synonymous and therefore interchangeable. This would entail a great deal of "donkey work" as, to my knowledge, no semantically tagged diachronic corpora exist to date.[43] Nevertheless, the semantic structure of the selected five near-synonymous adjectives is not devoid of complexity, with at least three senses being shared by all of them, including a figurative one. This issue is further discussed in the next section.

4.1.2 Revising reference works

The information about the near-synonyms reviewed in this section is based on different English dictionaries and thesauri, both historical and present-day. The list of present-day dictionaries

[43] It is even difficult to find semantically tagged synchronic corpora, although a few do exist. Examples are the *English SemCor Corpus* created by the *WordNet* project research team (Landes, Leacock & Fellbaum 1998), which includes semantic annotations for sense, and the *BBN Pronoun Coreference and Entity Type Corpus* (Weischedel & Brunstein 2005), which provides information about the co-indexation of pronouns and their antecedents, as well as a semantic categorization into entity types (e.g. PERSON, PRODUCT, PLANT, TIME, QUANTITY, and EVENT).

includes the following: *American Heritage Dictionary of the English Language* (*AHDOE*), *Cambridge Dictionary* (*CD*), *Collins Dictionary* (*CoD*), *Lexico*, *Longman Dictionary of Contemporary English* (*LDOCE*), *MacMillan Dictionary* (*MD*), *Merriam-Webster* (*MW*), and *Newbury House Dictionary of American English* (*NHDAE*). All of them offer information about the current senses of the selected adjectives and examples of use. Moreover, several of the dictionaries contain a thesaurus section, thus providing synonyms and antonyms of words. *CoD*, which uses data from the *COBUILD Corpus* to exemplify the different meanings of words, also gives information about their usage frequency, not only in PDE but also in earlier periods of the language. The reference work therefore serves to shed light on the different senses and uses of the near-synonyms at issue in contemporary English. In addition, historical information has been drawn from the *OED*, which is considered to be the largest and most comprehensive dictionary of English (e.g. Hoffmann 2004: 18). First published in 1884 and now on its third edition, available online, the *OED* constitutes an unparalleled database when it comes to tracking the history of English words and their meanings from the time when they entered the language until the present-day, thus serving as an essential guide for research on historical English. As such, it includes earlier meanings and etymology alongside current meanings. Consequently, the *OED* offers more illustrations of word usage than most other dictionaries, as it contains quotations from different periods of English. Each entry in the *OED* provides information about the pronunciation, spelling, frequency band in PDE, etymology, POS, and senses of a word. By making use of the examples of usage one can ascertain the first known attestation of the different meanings of a word and, in many cases, it is also possible to establish the relation between senses.

The *OED* makes evident that *fragrant*, *perfumed*, and *scented* are loanwords that were borrowed from French in eModE, specifically during the first half of the 16th century (cf. examples (16)–(18) below), as is the case of many words in this particular semantic domain (Durkin 2014: 414; cf. Section 4.1.1 above). However, despite this common origin, the way they entered the English language differs. *Fragrant*, which comes from the French adjective *fragrant,* ultimately dates back to Latin *frāgrant-em,* the present participle of *frāgrāre* 'to smell sweetly.' However, although it derives from a participial form, when *fragrant* became an English word, its participial nature was no longer recognizable. Both *perfumed* and *scented*, in turn, are derivatives formed by the suffix *-ed* (*OED*, s.v. *-ed* suffix[1]) attached to the verbal base

forms *perfume* and *scent*, which were also borrowed indirectly from Latin via French.[44] Contrary to *fragrant*, *perfumed* and *scented* were already formed in English by derivation, thus retaining their participial shape, which was used to form the past participles of weak verbs. It is, therefore, probably correct to assume that *perfumed* and *scented* still retain part of their verbal semantic functions in PDE, namely that of 'denoting a process or action' (Biber et al. 1999: 63). This may also be the case of the adjective *sweet-scented*, first attested at the end of the 16th century (cf. example (19)), given that it is a compound formed by the loanword *scented* and the OE adjective *sweet*. On the other hand, *sweet-smelling* is a composite form first attested in the early 15th century that combines the adjective *sweet* and the gerund form of *smell*, also of OE origin (cf. example (20)). Therefore, as on many other occasions in the history of English, near-synonymy arises in this specific case as a result of borrowing from Latin and French in the eModE period, with loanwords —*fragrant*, *perfumed*, *scented*, and, to a certain extent, *sweet-scented*— coming to be established in the language alongside an already existing native expression with the same meaning, that is, *sweet-smelling*.

(16) The **fragraunt** odour, and oyntment of swete flour (c.1530. *OED*, s.v. *fragrant* adj.)

(17) Suffitus, **perfumed**. (1538. *OED*, s.v. *perfumed* adj. 1)

(18) Many here smell strong but none so ranke as he A stronger **sented** knaue [knave] then he was cannot bee. (?c.1562. *OED*, s.v. *scented* adj. 1)

(19) **Sweet sented** Roe (1591. *OED*, s.v. *sweet-scented* adj. a)

(20) A place..Y-set aboute with floures so **swete smellyng**. (c.1400. *OED*, s.v. *sweet-smelling* adj. 1)

Although not all the dictionaries consulted distinguish the same senses and subsenses for the five near-synonyms, they all provide the same basic meaning 'having a sweet pleasant smell' for the five adjectives. This general sense is in fact the only definition provided for *sweet-scented* and *sweet-smelling* in the *OED*, *Lexico*, and *CoD*, which include an independent entry

[44] Another possibility, according to the *OED*, is that *perfumed* and *scented* derive from the nominal forms *perfume* and *scent* plus the suffix *-ed* (*OED*, s.v. *-ed* suffix[2]).

for these two compound words.[45] Similarly, *fragrant* occurs only with this sense in all eight PDE dictionaries, while the *OED* points out that this adjective can also be used figuratively to mean simply 'sweet or pleasant', as in example (21), where *fragrant* modifies the noun *memory*, which refers to an abstract entity and can therefore not emit an odor:

(21) Their **fragrant mem'ry** will out last their tomb (1782. *OED*, s.v. *fragrant* adj.)

Whereas a few dictionaries also define *perfumed* and *scented* exclusively in the general and figurative senses just mentioned, others make a distinction between two senses, which will here be labeled 'natural' and 'artificial'.[46] This is the case of, for instance, *MD*, which lists the following two definitions for *perfumed* and thus serves to illustrate these two senses:

(i) '[P]leasant to smell because of natural qualities' (*MD,* s.v. *perfumed* adj.)
(ii) '[P]leasant to smell because perfume has been added or used' (*MD*, s.v. *perfumed* adj.)

Therefore, according to the dictionaries that draw this distinction, there is a difference in nuance of meaning depending on whether the source of the smell is natural, in which case modified nouns designate entities which can release a smell on their own, such as *wallflowers* in (22), or artificial, where the adjectives collocate with nouns referring to entities which can acquire a

[45] *MW* does not provide an entry for *sweet-smelling*, only for *sweet-scented*. *MW* and the *OED* also offer an additional subsense for *sweet-scented*, namely 'in names of species or varieties of plants having sweet-smelling flowers, leaves, etc.' (*OED*, s.v. *sweet-scented* adj. b), as in *sweet-scented pea* or *sweet-scented geranium*. Examples exhibiting this sense are excluded from the present study as the adjective in this case forms part of the vernacular name of the plants. None of the other four adjectives seem to be used in this sense according to the reference material consulted in this dissertation.

[46] Moreover, two additional senses are provided for *scented* in the *OED*:

(i) 'Of tea, tobacco, etc.: flavoured with an aromatic ingredient. Also: having a fragrant taste as if flavoured by an aromatic ingredient. […]' (*OED*, s.v. *scented* adj. 2b)
(ii) 'With modifying word. Chiefly of an animal: having a sense of smell of the specified kind. […]' (*OED*, s.v. *scented* adj. 3)

The former meaning is classified in the *OED* as a subsense of the artificial sense and, since the other adjectives have also been found to collocate with tea- and tobacco-related nouns, for instance *sweet-smelling tobacco* in example (15) above, examples featuring this sense are included in the subsequent corpus analyses. Meaning (ii), on the contrary, is exclusive to *scented* and is therefore excluded here.

pleasant smell only by being impregnated by a pleasantly smelling substance, as in the case of *bath products* in (23).

(22) I planted some wonderful, dark-red, **scented wallflowers**. (*CD*, s.v. *scented* adj.)

(23) Indulge in your favorite **scented bath products** (*CD*, s.v. *scented* adj.)

In fact, in the *OED* the primary and first attested sense of *perfumed* is '[i]mpregnated with perfume; wearing perfume; scented with a (usually pleasant) odour' (*OED*, s.v. *perfumed* adj. 1), exemplified in (17) above. Therefore, in terms of the prototypicality or onomasiological salience of the adjectives, the reference works seem to indicate that *perfumed* and *scented* are possibly more prototypical than *fragrant*, *sweet-scented*, and *sweet-smelling* to refer to artificial aromas. This state of affairs may be a result of the etymological roots of the adjectives. As previously mentioned, *perfumed* and *scented* retain their participial form and thus probably also some of their original verbal functions, even when used as adjectives, namely '[t]o impregnate with a (usually pleasant) odour; to impart a (sweet) smell to […]' (*OED*, s.v. *perfume* v. 2). However, it is important to note that *fragrant*, *sweet-scented*, and *sweet-smelling* can also be used in the artificial sense (cf. examples (24)–(26)), as proved by several instances of these three adjectives in the dictionaries.[47] Moreover, *sweet-smelling*, in contrast to *sweet-scented*, also appears in the figurative sense, although no dictionary mentions this particular sense for this adjective (cf. example (27)).

(24) Inside are quirky old settees, painted chests and weathered wood hutches brimming with **fragrant** soaps and candles. (*LDOCE*, s.v. *fragrant* adj.)

(25) A scrumptious dinner with some **sweet-scented** candles and flowers is really worth coming home to. (*Lexico*, s.v. *sweet-scented* adj.)

(26) A light, **sweet-smelling** formula to support soft, fine curls that are prone to drooping. (2009. *CoD*, s.v. *sweet-smelling* adj.)

(27) But the project's success is not that **sweet-smelling**. (2007. *CoD*, s.v. *sweet-smelling* adj.)

[47] Although *sweet-scented* also retains the participial *-ed* ending, the dictionaries only provide the basic general sense '[e]mitting a sweet or pleasant odor' for this adjective, unlike *perfumed* and *scented*.

In (24)–(26), *fragrant*, *sweet-scented*, and *sweet-smelling* collocate with nouns referring to entities that need to be impregnated by an artificial substance so as to release an aroma, namely *soaps* and *candles* (24), *candles* (25), and *formula* (26).[48] In (27) *sweet-smelling* co-occurs with the noun phrase *project's success*, where the head noun *success* refers to an abstract entity that, as in the case of *memory* in (21) above, cannot emit an odor, and thus has to be interpreted as a metaphorical use of the adjective with the meaning 'pleasant or agreeable.' Table 2 summarizes the division into senses provided by each of the dictionaries and thesauri consulted for each of the five near-synonymous adjectives.

Table 2. Sense division of the five adjectives according to nine dictionaries and thesauri

Synonyms / Senses	FRAGRANT	PERFUMED	SCENTED	SWEET-SCENTED	SWEET-SMELLING
General basic sense 'Emitting a sweet or pleasant odor' / 'Having a pleasant or sweet smell'	*AHDOE*, *CD*, *CoD*, *LDOCE*, *Lexico*, *MD*, *MW*, *NHDAE*, *OED*	*CD*	*CD*, *LDOCE*, *Lexico*, *MD*, *MW*	*CoD*, *Lexico*, *MW*, *OED*	*CoD*, *Lexico*, *OED*
Natural sense 'Naturally having or producing a sweet, pleasant smell'	—	*CoD*, *Lexico*, *MD*, *OED*	*CD*, *CoD*, *OED*	—	—
Artificial sense 'Impregnated or scented with a sweet-smelling substance'	—	*CoD*, *Lexico*, *MD*, *OED*	*CD*, *CoD*, *OED*	—	—
Figurative sense 'Sweet, pleasant; cloying; made to seem sweet, disguised'	*OED*	*OED*	*OED*	—	—
NO ENTRY	–	*AHDOE*, *LDOCE*, *MW*, *NHDAE*	*AHDOE*, *NHDAE*	*AHDOE*, *CD*, *LDOCE*, *NHDAE*, *MD*	*AHDOE*, *CD*, *LDOCE*, *MD*, *MW*, *NHDAE*

[48] Note that in example (25), *sweet-scented* also modifies the noun *flowers*, which refers to an entity that can release a smell on its own (cf. Section 4.3 below for an explanation of how coordinated nouns modified by the adjectives under analysis are treated in the data annotation process).

Concerning the thesauri sections in the reference works, five out of the eight PDE dictionaries include such a section in which synonyms are provided: *CD*, *CoD*, *Lexico*, *MD*, and *MW*. Most of them list the five adjectives as synonymous, with the exception of *sweet-smelling*, for which a thesaurus entry is not included in any of the dictionaries consulted. However, in both *CoD* and *Lexico*, *sweet-smelling* is registered as a synonym of *fragrant*, *perfumed*, *sweet-scented*, and *scented*.

As can be observed in most thesauri entries, several other adjectives are also listed as being synonymous with the five selected items, including the seven mentioned in Section 4.1.1 above, that is, *aromatic*, *balmy*, *odoriferous*, *odorous*, *perfumy*, *redolent*, and *sweet*, together with *fragranced*. Nevertheless, after careful examination of the definitions and examples of usage provided by the PDE dictionaries and the *OED*, these eight adjectives have been excluded from the analysis for different reasons:

(i) In the case of *aromatic* (*OED*, s.v. *aromatic* adj. 1), the definition differs slightly from that of the five adjectives under scrutiny, as it typically implies that the smell it refers to stems from spices or herbs ('[h]aving the fragrant smell, and warm slightly pungent, taste, of spice […]'), which is not always the case with *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*.

(ii) The polysemous word *balmy*, although included in the *HTOED* in the category of SMELL AND ODOUR in one of its senses (i.e. *OED*, s.v. *balmy* adj. 3), seems to be more often used in other senses in PDE, as this particular meaning, namely '[d]elicately and deliciously fragrant', is not often listed in the contemporary dictionaries (cf., for instance, the entries for *balmy* in *CD*, *Lexico*, *LDOCE*, *MD*, and *MW*). In fact, in most of the PDE reference material only the sense '[o]f wind, air, weather, etc. Deliciously mild, fragrant, and soothing' (*OED*, s.v. *balmy* adj. 5) is included, and most examples of usage are illustrations of this particular sense, which indicates that *balmy* is today more semantically specialized than the five selected near-synonyms.

(iii) The two adjectives *odoriferous* and *odorous*, which originally shared the same definition as *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*, have become generalized with time to denote unpleasant smells alongside pleasant ones.

Moreover, in the case of *odorous*, it seems to have undergone recently a process of pejoration, just as the related noun *odor* discussed in Section 3.3, so that many of the PDE dictionaries point out that it is now mostly used in the sense of 'malodorous' (e.g *CD*, s.v. *odorous* adj.; *Lexico*, s.v. *odorous* adj. 1). The recently acquired negative semantic prosody of *odorous* becomes evident if its noun collocates are compared in a historical corpus such as *COHA* and a contemporary one such as *COCA*: whereas the top noun collocates of *odorous* in the former include words such as *blossom*, *perfume*, *sweets*, and *breezes*, in the latter they include items such as *emissions*, *stinkweed*, and *toxin*, which clearly have negative connotations, as well as more neutral ones, for instance, *air* and *compound*. Such preliminary analyses of *odorous* suggest that it is an antonym rather than a synonym of *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling* in PDE.

(iv) A similar scenario is found for *redolent*, which originally also shared the same definition as the five selected near-synonyms (*OED*, s.v. *redolent* adj. 1a), but which over time became differentiated from them. Although its first attested sense is 'of a pleasant sweet smell', it has come to acquire a different nuance of meaning and is nowadays more often used to denote a strong smell, thus being synonymous with *pungent*. Moreover, *redolent*, as evinced in most dictionaries, including the *OED*, is most typically used in PDE in the constructions *redolent with* or *redolent of*, which suggests that it is contextually more restricted.[49] When followed by *with* or *of*, *redolent* can also be used in the figurative sense '[i]mbued with or rich in a quality, a feeling, etc.; strongly suggestive or reminiscent of a particular thing', as in *[t]his dingy street was redolent with history* […] (1989. *OED*, s.v. *redolent* adj. 2b), which does not belong to the semantic field of SMELL AND ODOUR. In fact, this particular sense seems to be one of the most common of *redolent* nowadays, given that it is the meaning that is first mentioned in many of the PDE dictionaries (e.g. *CoD*, *LDOCE*, *Lexico*, and *MD*), and therefore its most the basic sense.

(v) The adjective *sweet* has been excluded from the analysis primarily due to the fact that it is a highly polysemous word with eleven different main senses and a large

[49] Out of the 452 occurrences of the adjective *redolent* in *COHA*, 378 instances, that is, 83.7%, are followed by either *with* or *of*.

number of subsenses in the *OED*, of which only one belongs to the olfactory domain (*OED*, s.v. *sweet* adj. 2). Given the extensive polysemy of the adjective and its large number of occurrences in *COHA*, to wit 41,970, it would have been extremely time-consuming to go through them manually in order to select only those which correspond to the meaning 'of a pleasant sweet smell'. Additionally, the difficulty of teasing apart some of the other senses of *sweet*, particularly that in which the adjective is used to designate a pleasant taste or flavor (*OED*, s.v. *sweet* adj. 1), from the one of interest here would further complicate the task. Consider examples (28) and (29), extracted from *COHA*, in which *sweet* and *fragrant* modify the noun *tea*. Whereas *fragrant* in (29) is most probably used to refer to the smell of the tea, *sweet* in (28) is more ambiguous and could refer either to the taste or to the aroma of the tea, or to both at the same time.

(28) Recovering her composure in the servant's parlor, watched by Mr. Grove's kind eyes as she sipped at her second cup of **sweet tea**, Mrs. France tried once again to explain to him the swirl of feelings that were in her head. (*COHA*, 1995, FIC, FantasySciFi)

(29) Greer nodded a greeting to Blythe as the other woman sipped a **fragrant tea** from a delicate china cup. (*COHA*, 2006, FIC, MatchMaker)

(vi) Finally, the two adjectives *fragranced* and *perfumy*, while sharing the same definition as the five near-synonyms at issue, seem to be relatively infrequent in English. *Fragranced* is classified as rare or archaic in several of the dictionaries and, in fact, it is not attested in *COHA*. In turn, although *perfumy* occurs in *COHA*, there are only 15 instances of this adjective in the corpus. Such a small number of instances are not enough to extract any generalizations about the contextual behavior of *perfumy*, which has therefore been left out.

In conclusion, although *aromatic*, *balmy*, *odoriferous*, *odorous*, *redolent*, *sweet*, *fragranced*, and *perfumy* are all still semantically related to *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*, with some of their senses belonging to the domain of SMELL AND ODOUR, their descriptions in the reference material suggest that they are not as prototypical to

denote the concept PLEASANT SMELLING as the five selected adjectives. First, *aromatic* has never been used in exactly the same sense and often seems to be restricted to smells stemming from spices and herbs. Second, *balmy*, *odoriferous*, *odorous*, and *redolent*, though originally synonymous with *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*, appear to have become differentiated from them through processes of generalization, specialization, and/or pejoration. Third, *perfumy*, *fragranced*, and *sweet* are still near-synonyms of the five adjectives object of study, but the first two seem to be falling into disuse and the last one has many senses which are not related to the five selected adjectives. Therefore, the degree of synonymy holding between *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling* is probably higher, as they seem to be interchangeable in most, if not all, contexts of use. This is so because (i) they are defined in practically the same way in all dictionaries, (ii) the examples of usage provided are very similar, with some nouns appearing with all of them (e.g. *flower*, *smell*, and *water*), (iii) they are practically always listed as synonyms of one another, and (iv) they are sometimes defined in terms of each other.

Nonetheless, the dictionaries also point to some differences between the five terms at issue here concerning their frequency of use. The *OED*, which provides a frequency range between 1 (very low frequency) to 8 (very high frequency) for words in PDE, assigns a rating of 5 to *fragrant* —that is, between 1 and 10 times per million words—, a rating of 4 to *perfumed*, *scented*, and *sweet-smelling* —that is, between 0.1 and 1 times per million words—, and a rating of 3 to *sweet-scented*— that is, between 0.01 and 0.1 times per million words.[50] This suggests that *fragrant* is probably the dominant term in the set in PDE. Additionally, *CoD* includes information regarding the frequency of four out of the five adjectives, namely *fragrant*, *perfumed*, *scented,* and *sweet-scented*, across time, particularly from the year 1708 to 2008. According to this data, based on *Google Ngrams*, *fragrant* and *perfumed* have decreased in frequency over time from 2.62 and 0.48 instances per million words in 1708, respectively, to 0.54 and 0.23 in 2008.[51] Similarly, *sweet-scented* has gone from a frequency of 0.2 in 1720 to 0.01 in 2008. On the contrary, *scented* has increased in frequency throughout the same period,

[50] For further details on how frequency is calculated in the *OED* and what each of the frequency bands indicate, see https://public.oed.com/how-to-use-the-oed/key-to-frequency/.

[51] *Google Ngrams* is a tool that offers the relative frequency of lexical units over time in a corpus of books. For more information, see https://books.google.com/ngrams/info.

from 0 to 0.4 instances per million words. Interestingly, most changes seem to have taken place from approximately the 19th century onwards, precisely the time span covered in the present dissertation. Despite this valuable frequency information for PDE and earlier periods of English, neither the *OED* nor *CoD* shed any light on the distribution of the adjectives in different senses and more specific contexts of use.

The dictionaries and thesauri reviewed in the present section offer some important insights into the relation between the five near-synonymous adjectives *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling* regarding both their similarities and possible differences in terms of frequency and semantic characteristics. However, the information provided by these reference works is still insufficient in order to acquire a full understanding of the structure of the concept PLEASANT SMELLING and of the degree to which the five adjectival near-synonyms are interchangeable. As such, some questions are yet left unanswered:

(i) Are *perfumed* and *scented* the only adjectives of the set that are used in the artificial sense, i.e. 'pleasant to smell because perfume has been added or used', or do *fragrant*, *sweet-scented*, and *sweet-smelling* also appear in such semantic contexts?

(ii) If *fragrant*, *sweet-scented*, and *sweet-smelling* are also used in the artificial sense, is there any difference in their distribution, both in this and in the other senses (i.e. natural and figurative) of the adjectives?

(iii) Does the distribution of the five adjectives across senses remain stable over time or does it fluctuate? In other words, are the apparent changes in frequency of the adjectives revealed in *CoD* specific to one or several contexts of use or do they, on the contrary, occur equally across all contexts?

(iv) The reference works point to some semantic (dis)similarities among the adjectives, but do differences concerning, for instance, connotation, style, and/or morphosyntactic features also exist?

The analyses in the present dissertation are designed in order to try to answer these research questions. However, before moving on to the results of these analyses, which are provided in Chapters 5–7, Sections 4.2 and 4.3 offer a description of the corpus used, as well as the data retrieval and annotation processes.

4.2 CORPUS DESCRIPTION AND DATA RETRIEVAL

Given the relatively low frequency of the five near-synonymous adjectives object of study, a large corpus is necessary to carry out the analyses in Chapters 5–7. For this reason, the data used was drawn from *COHA*. Released in 2010, *COHA* is the largest historical corpus of English, containing about 400 million words of running text from more than 100,000 individual texts, which are divided into four different genres or text-types: fiction, popular magazines, newspapers, and non-fiction.[52] *COHA* covers the period 1810–2009 in AmE, and this time span of 200 years is split into 20 decades. Table 3 displays the number of words across the different decades and text-types. As shown in the table, *COHA* is not a balanced corpus in terms of number of words, neither across decades nor across genres, that is, it contains considerably more data for the later decades than for the earlier ones, and fiction accounts for a much higher percentage of the data than the other three genres. For instance, until the 1850s there are less than 17 million words per decade, whereas the last two decades, 1990s and 2000s, include nearly double the amount of running text, almost 28 and 30 million words, respectively. Despite the unequal distribution of the data, an effort was made by the corpus compilers to keep the distribution of the different genres even through the 20 decades. For instance, fiction totals between 48–55% of the data in each decade (see the last column in Table 3). Similarly, roughly the same percentage of popular magazines, non-fiction, and newspapers is maintained over time, with the exception of newspapers in the first six decades. This distribution ensures that possible changes are not just a byproduct of fluctuations in the number of words per genre over time, but that they reflect actual changes.

[52] Within the four genres there are further subclassifications. For more details about the main sources used for each of the four text-types, see the information available at http://www.helsinki.fi/varieng/CoRD/corpora/COHA/basic.html.

Table 3. Number of words in *COHA* across decade and genre

DECADE	FICTION	POPULAR MAGAZINES	NEWSPAPERS	NON-FICTION BOOKS	TOTAL	% FICTION
1810s	641,164	88,316	–	451,542	1,181,022	54%
1820s	3,751,204	1,714,789	–	1,461,012	6,927,005	54%
1830s	7,590,350	3,145,575	–	3,038,062	13,773,987	55%
1840s	8,850,886	3,554,534	–	3,641,434	16,046,854	55%
1850s	9,094,346	4,220,558	–	3,178,922	16,493,826	55%
1860s	9,450,562	4,437,941	262,198	2,974,401	17,125,102	55%
1870s	10,291,968	4,452,192	1,030,560	2,835,440	18,610,160	55%
1880s	11,215,065	4,481,568	1,355,456	3,820,766	20,872,855	54%
1890s	11,212,219	4,679,486	1,383,948	3,907,730	21,183,383	53%
1900s	12,029,439	5,062,650	1,433,576	4,015,567	22,541,232	53%
1910s	11,935,701	5,694,710	1,489,942	3,534,899	22,655,252	53%
1920s	12,539,681	5,841,678	3,552,699	3,698,353	25,632,411	49%
1930s	11,876,996	5,910,095	3,545,527	3,080,629	24,413,247	49%
1940s	11,946,743	5,644,216	3,497,509	3,056,010	24,144,478	49%
1950s	11,986,437	5,796,823	3,522,545	3,092,375	24,398,180	49%
1960s	11,578,880	5,803,276	3,404,244	3,141,582	23,927,982	48%
1970s	11,626,911	5,755,537	3,383,924	3,002,933	23,769,305	49%
1980s	12,152,603	5,804,320	4,113,254	3,108,775	25,178,952	48%
1990s	13,272,162	7,440,305	4,060,570	3,104,303	27,877,340	48%
2000s	14,590,078	7,678,830	4,088,704	3,121,839	29,479,451	49%
TOTAL	207,633,395	97,207,399	40,124,656	61,266,574	**406,232,024**	51%

The size and scope of *COHA* make it the perfect resource for the purposes of the present dissertation, as it allows us to examine low frequency items such as the near-synonyms under study with greater confidence than smaller corpora. Additionally, *COHA* is freely available and easily accessible via its inbuilt online interface, which provides many convenient user-friendly search tools. In particular, given that *COHA* is POS-tagged, it is possible to search only for the lexical items of interest here, and thus exclude the verbal uses of *perfumed* and *scented.* The interface also makes it possible to retrieve the collocates of a given word. By means of this option, one can select a search window of up to nine words to the left and nine words to the right of the node word (i.e. L9–R9). Moreover, the search for collocates can be restricted to specific POS and, for instance, retrieve only noun collocates. The interface also permits the researcher to establish a frequency and/or PMI threshold for the collocates, thereby limiting the search to include only those which occur relatively frequently with a node.

Two separate databases were created from the data in *COHA*, given that the aims of the case studies in Chapters 5 and 6, on the one hand, and Chapter 7, on the other, differ and thus impose distinct requirements on the data. For the first database the following queries were made by using *COHA*'s online interface: 'fragrant_j*', 'perfumed_j*', 'scented_j*', 'sweet-

scented_j*', 'sweetscented_j*', 'sweet scented', 'sweet-smelling_j*', 'sweetsmelling_j*', and 'sweet smelling', where the POS-tag '_j*' indicates that only adjectives were searched for. The strings 'sweet scented' and 'sweet smelling', written as two separate words without hyphenation, were not specified for POS because it is not possible to add one POS-tag to two separate words, as the tag would only affect one of them. In addition, the instances of *perfumed* and *scented* tagged as verbs in *COHA* were also retrieved, that is, 'perfumed_v*' and 'scented_v*', where '_v*' stands for verb. This step was taken in order to identify possible adjectival uses of these lexical items which had been erroneously tagged as verbs.

The instances of the adjectives *sweet-scented* and *sweet-smelling* were retrieved by means of three distinct search strings given that, although the two compounds are mostly spelt with a hyphen, alternative spellings exist, as in many other cases of compounds (Huddleston & Pullum et al. 2002: 451).[53] It is a well-known fact that it is often very difficult to disambiguate between compound words and syntactic constructions formed by two independent lexical items (e.g. *a blackbird* vs. *a black bird*), given that there is not a clear-cut division between the two, but rather a cline (e.g. Biber et al. 1999: 589–590; Huddleston & Pullum et al. 2002: 1644). This problem is further aggravated when texts from different historical periods are considered, as in the present dissertation, because as pointed out by Quirk et al. (1985: 1569), it is usually the case that compounds first occur as individual words and only once they become more established and lexicalized are they written as hyphenated or even as single words. Although alternative interpretations may exist for such ambiguous cases, the decision was taken to include in the analysis the examples corresponding to the search strings 'sweet scented' and 'sweet smelling' because they seem to behave exactly as those instances in which the adjectives are written with a hyphen or as one word. Consider in this connection examples (30)–(32):

(30) Before it lay a neat smooth little court, surrounded by a close hedge, of a **sweet scented** red and white **flower**, resembling the honeysuckle in shape. (*COHA*, 1828, MAG, NorthAmRev)

[53] For the sake of convenience, throughout the present dissertation, these two compounds are always referred to by means of the hyphenated alternative, that is, *sweet-scented* and *sweet-smelling*, as the vast majority of examples retrieved from *COHA* correspond to these spellings (cf. Table 4 below).

(31) It has white, **sweetscented flowers**, arranged in the same manner as the last; stem without spots. Leaves ovate lanceolate, quite smooth. (*COHA*, 1851, NF, FlowerGardenBrecks)

(32) Bladder Senna. Colutea, an ancient name of a bush with **sweet-scented flowers**. The genus includes a number of species of shrubs, with yellow or orange, pea-shaped flowers, which are succeeded by seed-vessels like bladders. (*COHA*, 1851, NF, FlowerGardenBrecks)

In all three examples, *sweet*(-)*scented* modifies the noun *flower*(*s*) without any discernible difference in meaning. It is worth mentioning here that a series of criteria, both syntactic and non-syntactic, have been used to distinguish between a true compound and a mere combination of two independent words (Biber et al. 1999: 589–590; Huddleston & Pullum et al. 2002: 448–451). Among syntactic criteria, combinations of two words can be individually coordinated with and modified by a different word, whereas this is not possible in the case of compounds. As to non-syntactic criteria, compounds are typically stressed on the first element, are spelt as one word (either without or with a hyphen), and have a non-transparent meaning. Combinations of two independent words, in turn, are typically stressed on the second element, are written as two separate words, and have a transparent meaning. These syntactic and non-syntactic criteria, however, were not easily applicable to the current dataset for various reasons. First, given that *COHA* contains only written language, it is not possible to know the stress patterns of the relevant examples. Second, orthography is not a reliable clue, as many compounds have alternative spellings and, as mentioned above, compounds are typically first written as separate words and later on as hyphenated or single words as they become more accepted and established. Moreover, hyphenation is claimed to be less common in AmE (Quirk et al. 1985: 1569). Finally, meaning is also problematic in this case, given that the semantics of the compounds *sweet-smelling* and *sweet-scented* are highly transparent in the sense that they are the sum of their components, namely ‘that smells sweet’ and ‘that has a sweet scent’, respectively. Therefore, their meaning is the same if they are spelt as two separate units. Considering the difficulty of applying these criteria to the data from *COHA*, the decision to include in the analyses all instances corresponding to the search queries ‘sweet scented’ and ‘sweet smelling’ was further reinforced.

Table 4 summarizes the data retrieval process of the first database, relevant to the analyses in Chapters 5 and 6, and the initial number of examples extracted by means of each query.

Table 4. Number of instances retrieved per query

QUERY	NUMBER OF INSTANCES RETRIEVED
fragrant_j*	3,374
perfumed_j*	792
scented_j*	792
sweet-scented	173
sweetscented_j*	5
sweet scented	37
sweet-smelling_j*	207
sweetsmelling_j*	4
sweet smelling	37
perfumed_v*	340
scented_v*	577
Total	**6,338**

The second database, relevant for the discussion in Chapter 7, does not include the instances of the adjectives, but their noun collocates. Therefore, the collocates option in *COHA* was used to retrieve the lemmas of the collocates of *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling* by means of the POS-tag _nn* in an L5–R5 context window. An L5–R5 context window was selected in this case since it has been shown that tighter windows, such as one or two words, often lead to data sparseness, especially if low frequency items are considered (Sahlgren 2006). Moreover, such tight windows are often more appropriate to retrieve semantically (dis)similar terms, such as synonyms or antonyms of the target word (Peirsman, Heylen & Geeraerts 2008: 40). On the contrary, if one is interested in typical collocates of the target, it is convenient to loosen the context window somewhat, for instance, to L5–R5. Another option would be to consider only collocates which are syntactically connected to the target, in the present case either syntactically —or semantically— modified nouns of the adjectives (e.g. *the* ***fragrant flower***; *the* ***flower*** *is* ***fragrant***). However, this would also lead to a lower number of retrieved collocates, and thus again to data sparseness, as in the case of tighter context windows. In addition, only the noun collocates of the adjectives were considered because, as argued by, for example, Geeraerts (1986), Justeson & Katz (1995), and Gries (2001; 2003), nouns are more informative than other word types when it comes to the semantics of adjectives (cf. Section 3.2.2). As mentioned earlier (cf. Section 2.3.2), Geeraerts' (1986) study on the

Dutch adjective *vers* 'fresh' demonstrates that the fine-grained aspects of meaning of polysemous adjectives can be discovered by examining the nouns they modify.

The second database did not require further annotation due to the fact that it is used only in the analyses of individual collocates in Chapter 7. On the contrary, the instances retrieved for the first database were manually pruned and subsequently annotated for a number of variables, both language-external and language-internal ones. The pruning and annotation processes of the first database are the focus of the next section.

4.3 DATA ANNOTATION

Prior to the annotation process, the instances in the first database were revised one by one to exclude false positives, as well as other problematic examples. First, in the case of duplicated occurrences, that is, when the same instance appeared more than once in the corpus, only one of them was maintained in the database. Second, three cases of *perfumed* from the same source (examples (33) and (34) below) were used in a metalinguistic context and were therefore excluded, as they are not actual instances of the adjective:

(33) A new alliterative bombardment is inaugurated by "poop" with its doubled "p's." closely followed by the doubled "p's" in "purple," and then hammered home by another "p" in "**perfumed**." (*COHA*, 1909, MAG, Harpers)

(34) Note also the magical use of the "f's" and "v's" in "**perfumed**," "love," and "silver," and of the "m's" in "**perfumed**," "them," "made," and "amorous." (*COHA*, 1909, MAG, Harpers)

Third, two instances of *perfumed* and two of *fragrant* were also removed given that they were part of the name of a brand (35) and of the title of a book (36):

(35) She burned your translation of **The Perfumed Garden**, claiming you would not have wanted to publish it unless you needed the money for it, and you didn't need it, of course, because you were now dead. "Burton was speechless for one of the few times in his life. |p109 Frigate looked out of the corner of his eyes at Burton and grinned.

He seemed to be enjoying Burton's distress." Burning **The Perfumed Garden** wasn't so bad, though bad enough. (*COHA*, 1971, FIC, YourScatteredBodies)

(36) It wasn't Sabine who had mixed up some poisonous sludge as a joke to mock at him and his obsessions. Why hadn't he believed her? She had given him a bottle of **Arkady Ferson's Fragrant Goddess**, prepared by him and described in his 1969 treatise, the veracity of which Jeremy had always rejected out of hand. (*COHA*, 2007, FIC, FantasySciFi)

Fourth, during the course of the manual revision, some false positives of 'perfumed_j*'and 'scented_j*' were identified, that is, examples which were erroneously tagged as adjectives, but in fact corresponded to verbal uses. This is the case of, for instance, examples (37) and (38):

(37) He was determined to have a luxurious bath, to be shaved and **perfumed**, to leave behind him the very dust of his past life. (*COHA*, 1893, FIC, SingerFromSea)

(38) The cops had heard that one before, and **scented** trouble. One of them began to advance. (*COHA*, 1970, MAG, Time)

In (37), *perfumed* is a past participle in a verb phrase in the passive voice, with the meaning '[t]o impregnate with a (usually pleasant) odour; to impart a (sweet) smell to; to apply perfume to' (*OED*, s.v. *perfume* v. 2). In (38), *scented* is the past participle or past simple of the verb *scent*, used figuratively to mean '[t]o perceive or discover as if by smell; to detect or discern instinctively or from subtle indications; to intuit' (*OED*, s.v. *scent* v. I. 2b).

Among the instances of *perfumed* and *scented* tagged as verbs, that is, those retrieved by means of the queries 'perfumed_v*' and 'scented_v*', some were in fact adjectival rather than verbal uses of the items in question. As is well-known, the distinction between past participles and participial adjectives is not always straightforward, and several such ambiguous instances were identified in the database. Therefore, a decision had to be made regarding which of these cases counted as adjectival uses and which, on the contrary, counted as verbal ones. A number of criteria have been proposed in the literature (e.g. Granger 1983; Andersson 1985; Quirk et al. 1985: Section 7.16; Biber et al. 1999: 505–506; Huddleston & Pullum et al. 2002: 78–79) to classify these ambiguous examples as being located closer to the adjectival end or closer to the

verbal end of the continuum. Even though participial adjectives are not considered prototypical instances of the adjectival class, a participle is regarded more adjective-like than verb-like if:[54]

(i) the focus is more on the state resulting from an action rather than on the action itself, as in example (39), where the emphasis is on the look of the man, not on the process whereby he acquired his *barbered and perfumed* appearance;
(ii) it can occur with lexical copular verbs such as *appear*, *feel*, or *seem* (cf. example (40));
(iii) it can be modified by adjectival intensifiers such as *very* (cf. example (41));
(iv) it can be inflected for degree, either comparative or superlative (cf. examples (39) and (42)); and
(v) it can be coordinated with central or true adjectives (cf. examples (40) and (43)).

(39) And, oddly enough, this **look** of premature senility was not masculine but feminine. Though no **more** barbered and **perfumed** than the next Italian man, he evoked the black mass of the dressing-table and the hand-mirror (*COHA*, 1950, FIC, CastColdEye)
(40) Our voices **seemed low** and **scented** with fragrant incense, hidden behind the static-cracked passion of the Baptists, the Pentecostal brimstone and fury, the preacher's calling from the pulpit, the congregation's response from the pews. (*COHA*, 1998, FIC, MassachRev)
(41) They did, however, keep up a suspicious intimacy with a brilliantly lighted, though not **very** fragrantly **scented**, saloon on the left. (*COHA*, 1858, FIC, LifeAdventures)
(42) The breezes that bore them were **more** exquisitely **scented** than those which had soothed my doubts, bad stimulated my imagination on her feast-days. (*COHA*, 1922, MAG, Atlantic)

[54] It is worth noting here that these criteria are postulated to disambiguate between past participles and participial adjectives in PDE. Therefore, it is not entirely clear whether they can also be safely applied to data from earlier periods in the history of English. However, these criteria are often the only way of reaching a relatively objective decision regarding such ambiguous cases.

(43) The rest of him, which was Spanish, was **indolent** and **arrogant** and **perfumed**. (*COHA*, 1933, FIC, Harpers)

Examples of *perfumed* and *scented* which satisfied any of these five criteria were consequently kept in the database and subsequently included in the analyses. All other occurrences of these two items tagged as verbs were excluded from the count.

Finally, some occurrences of the five near-synonyms were counted more than once. This is the case of examples in which the adjectives modify two or more coordinated nouns, as in (44):

(44) For years Italy led in perfuming; it supplied the rest of Europe with sweet bags, perfume cakes for throwing on fires, **fragrant candles and cosmetics, scented gloves and pomanders** (*COHA*, 1922, MAG, Mentor)

Four entries in the database were derived from example (44): two instances of *fragrant* and two of *scented*. This was so because both adjectives here modify more than one noun, which appear in coordination, to wit, *candles and cosmetics*, in the case of *fragrant*, and *gloves and pomanders*, in the case of *scented*. The reasons for taking this decision were two. First, in many such cases, the coordinated nouns belong to different semantic domains and even instantiate different senses of the adjectives, which poses problems for the annotation of the variables Sense (cf. Section 4.3.1.1 below) and Semantic category (cf. Section 4.3.1.2 below). In (44), for instance, the nouns modified by *fragrant*, namely, *candles* and *cosmetics*, were classified as belonging to the semantic categories OBJECT and COSMETICS, respectively. Similarly, the nouns modified by *scented*, *gloves* and *pomanders*, were categorized as TEXTILE AND CLOTHING and COSMETICS, respectively. Second, given that one of the variables for which the data was annotated included the specific nouns modified by the adjectives (cf. Section 4.3.2 below), coordinated nouns had to be kept separate.

As mentioned in Section 4.2, the data retrieval process resulted in a total of 6,338 instances (cf. Table 4 above). Following the manual pruning discussed in the preceding paragraphs, 5,764 occurrences of the five adjectives remained in the first database. These examples were subsequently annotated for eleven variables of different types, including language-internal

semantic (Section 4.3.1) and non-semantic factors (Section 4.3.2), as well as language-external ones (Section 4.3.3).

4.3.1 Language-internal semantic variables

As argued in Section 3.2, the variables which influence the choice between near-synonyms vary from one set to another, depending on factors such as their POS or their semantic domain (cf., for instance, Hanks 1996: 92, 96; Liu 2010: 61). Previous studies on adjectival near-synonyms reveal that semantic factors are the most significant, particularly those pertaining to the nouns that the adjectives modify, such as animacy, concreteness, countability, and more specific semantic categorizations (e.g. Gries 2001; 2003; Gries & Otani 2010; Liu 2010). Therefore, in the present dissertation, the semantic factors examined also focus on features of the nouns modified by *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*. In fact, out of the eleven variables for which the data was annotated, five are language-internal semantic factors which directly or indirectly relate to the modified nominal elements. These are Sense, Semantic category, Animacy, Concreteness, and Countability.

4.3.1.1 Sense

The basic meaning of all five adjectives is 'having a sweet pleasant smell'. However, as discussed in Section 4.1 above, some dictionaries distinguish between two related senses, namely the artificial and the natural senses, depending on the source of the smell. While this distinction is made explicit only in the entries for *perfumed* and *scented*, we saw that *fragrant, sweet-scented*, and *sweet-smelling* could also be used to denote both natural and artificial smells (cf. examples (24)–(26)). Moreover, a figurative sense, 'sweet or pleasant', was also distinguished in the *OED* for *fragrant*, *perfumed*, and *scented*, and instances of *sweet-smelling* in this sense also occurred in examples of usage offered in the dictionaries (cf. example (27)). Although a considerable amount of information is provided in the dictionaries and thesauri, one issue that is yet not clear concerns the distribution of the adjectives across these three senses (i.e. natural, artificial, and figurative). Therefore, the data was first annotated for a variable called Sense, which contained these three senses as levels. Examples of each of them are provided again for convenience under (45)–(47):

(45) Not more beautiful by early dawn and dew had been this ride, than it was through lengthening shadows, and violet glow of sunset, and silvery light of moon, the peaches ripening on sunny walls, and the odors of mint and **sweet-smelling herbs** rising through the gathering damps of evening, the birds singing their vesper songs, and in the deep forest glades the lonely nightingale pouring out his soul to the moon. (*COHA*, 1904, FIC, RoseOldStLouis)

(46) There was something about this bathroom he liked. It was done in black and gold, with decanters of **scented soap** and lovely **scented tissue** and **toilet paper**, also gold, and small exquisite guest towels that were white linen with gold embroidering. (*COHA*, 1970, FIC, WheelLove)

(47) Elsie was quite refreshing after Mignonette and Yvonne. She was British and had none of the elusive, **perfumed charm** of my other loves. She was utterly frank and good-natured and rather innocent. (*COHA*, 1945, FIC, HappyTime)

Example (45) was categorized as 'natural', given that *sweet-smelling* modifies the noun *herbs*, an entity which can release a natural smell on its own. Example (46), in turn, displays three instances of *scented* in the artificial sense, as the adjective modifies the nouns *soap*, *tissue*, and *toilet paper*, all denoting entities that need to be impregnated in order to emit a sweet pleasant aroma. Finally, (47) illustrates the use of *perfumed* in the figurative sense, since *charm* is an abstract notion which refers to the personality of Elsie and cannot release a smell. Therefore, *perfumed* has to be interpreted figuratively with the meaning 'sweet or agreeable'.

During the annotation process, it became clear that many cases could not be safely assigned to any of the three aforementioned senses, given that their interpretation was ambiguous between the natural and artificial readings.[55] Consider in this respect example (48):

(48) Miriam moved the book to make room on her lap for the little girl. For a moment she laid her cheek against Angie's warm **fragrant hair**. | She was almost choked by her surging love for the child and by the crisis of the last hour. "I always like this story because it has a happy ending, Mama." (*COHA*, 1984, FIC, CrescentCity)

[55] The high number of ambiguous occurrences may well be one of the reasons why many dictionaries do not contain separate senses for what we call here the natural and artificial senses, but include just one basic or general sense.

In (48), *fragrant* could be interpreted either as referring to the natural smell of Angie's hair or as designating an artificial aroma, such as a fragrance left by shampoo or another cleansing agent used to wash the child's hair, or even a mixture of artificial and natural smells. In fact, there is nothing in the context which can help in disambiguation. It may well be the case that the specific nature of the smell was not relevant or important for the author of this particular text at the point of writing.

Even though ambiguity introduces more complexity to the annotation process and subsequent analyses, it is also rampant in the lexicon, as discussed in Chapter 2 (cf. Sections 2.1, 2.2, and 2.4.1). As demonstrated by scholars such as Erdmann (1910: 5), Gipper (1959), and advocates of the prototype theory (e.g. Taylor 1995: Chapters 2 & 3; Geeraerts 1997: 11), vagueness is widespread in language and communication, particularly in lexical semantics, where the limits between one meaning and another are not always clear-cut. The specific type of vagueness involved in examples such as (48) is what Geeraerts (2010: 197) denominates *interpretative indeterminacy*, which refers to instances that are impossible to disambiguate contextually and can thus not be interpreted through the information available. Geeraerts gives the example of the adjective *plain* in *she is a plain girl*, where we cannot be certain whether *plain* is used in the sense of 'simple' or rather in that of 'ugly'. Therefore, it was necessary to add a further level to the variable Sense to account for instances in which it was not possible to identify the source of the smell. By making use of the notion introduced by Geeraerts, this level was called 'indeterminate'. Besides examples of the type illustrated in (48), this level also includes those in which there exists a mixture of natural and artificial aromas, as in (49).[56]

(49) She smiled in reply and, holding by the hand the child she had named Love, she stepped over the high wooden doorsteps and into the temple. The air was **fragrant with incense and lilies**. Incense burned before the gods, and lilies bloomed in the court. (*COHA*, 1946, FIC, PavilionWomen)

[56] The extremely low number of examples in the corpus where there is clearly a mixture of natural and artificial aromas made it inefficient to have an independent level for such instances.

The smell of the air described in (49) clearly comes from both artificial and natural sources, as proved by the prepositional complement of *fragrant*, namely *with incense and lilies*, where incense is artificially fragrant, and lilies naturally so.

Indeterminate cases were maintained in the count because of two main reasons. First, given that vagueness is a pervasive feature in the lexicon, excluding such cases would result in a rather incomplete and unrealistic reflection of the near-synonyms in particular, and of language in general. Second, and more importantly for our purposes, excluding indeterminate instances from the analysis would certainly skew the results for other variables, particularly Semantic category, as many of the ambiguous cases belong to the semantic categories BODY AND PEOPLE and SPACE, which comprise, respectively, more than half and almost one third of the indeterminate examples in the corpus (cf. Table 12 in Section 5.2.2 below). The semantic categorization of modified nouns is discussed next.

4.3.1.2 Semantic category

As seen in Section 2.3.2, besides individual collocates, Sinclair (1991a; 1996) identifies three additional levels of co-occurrence: colligation, semantic preference, and semantic prosody. Semantic preference concerns more schematic semantic features of the collocates of a particular node, in this case the five near-synonymous adjectives. Drawing on this notion, the data was annotated for a variable called Semantic category, which classifies the nouns modified by the adjectives according to the semantic domain to which they belong.[57] This variable thus provides a more fine-grained picture than Sense regarding the contexts of use in which the adjectives occur. In fact, this variable is the most complex of those considered in the analysis, given that it contains twelve different levels or semantic categories. The identification of semantic classes is often highly subjective and far from being straightforward. Consequently, in order to achieve a certain degree of objectivity, two well-known systems of semantic classification were used, namely the *UCREL Semantic Analysis System* or *USAS* for short (Archer, Wilson & Rayson 2002; Rayson et al. 2004) and the *HTOED*.

[57] Whenever a pronoun was modified by any of the near-synonymous adjectives at issue, its antecedent was used for the semantic classification.

USAS contains twenty-one major semantic classes, displayed in Table 5, which are then further divided into more specific subclasses.[58]

Table 5. Major semantic classes in *USAS*

A GENERAL AND ABSTRACT TERMS	**B** THE BODY AND THE INDIVIDUAL	**C** ARTS AND CRAFTS	**E** EMOTION
F FOOD AND FARMING	**G** GOVERNMENT AND PUBLIC	**H** ARCHITECTURE, HOUSING AND THE HOME	**I** MONEY AND COMMERCE IN INDUSTRY
K ENTERTAINMENT, SPORTS AND GAMES	**L** LIFE AND LIVING THINGS	**M** MOVEMENT, LOCATION, TRAVEL AND TRANSPORT	**N** NUMBERS AND MEASUREMENTS
O SUBSTANCES, MATERIALS, OBJECTS AND EQUIPMENT	**P** EDUCATION	**Q** LANGUAGE AND COMMUNICATION	**S** SOCIAL ACTIONS, STATES AND PROCESSES
T TIME	**W** WORLD AND ENVIRONMENT	**X** PSYCHOLOGICAL ACTIONS, STATES AND PROCESSES	**Y** SCIENCE AND TECHNOLOGY
Z NAMES AND GRAMMAR			

Each semantic tag groups together semantically related items which are associated with the same mental concept, including synonyms, antonyms, hypernyms, and hyponyms. *USAS* allows researchers to input up to 100,000 words of running text, which it then automatically tags according to their semantic category. Given enough context for a word, *USAS* is in some cases able to identify its correct sense. However, on many occasions this is not possible and the tagger then provides a list of tags for the same word, ranked according to their frequency order, with the most probable tag first. To illustrate this point, consider the word *bath* in *[h]e had imprudently taken a bath at too high a temperature* (1837. *OED*, s.v. *bath* n.[1] I. 1). *USAS* analyzes *bath* as belonging to two different categories, namely H5: FURNITURE AND HOUSEHOLD FITTINGS and B4: CLEANING AND PERSONAL CARE, even though in this particular instance the noun *bath* does not refer to the receptacle for bathing, which would be considered a piece of furniture. Therefore, it is often necessary to manually verify the semantic classification provided by *USAS* with the help of a more precise database such as the *HTOED*. This second source also offers a taxonomic classification of semantic classes, with the great majority of

[58] For more detailed information about the different semantic classes and subclasses, see the official *USAS* website, available at: http://ucrel.lancs.ac.uk/usas/.

word senses in the *OED* being categorized into semantic domains which are hierarchically ordered (Kay 2010; 2012). At the most schematic level, three divisions are made between THE EXTERNAL WORLD, THE MIND, and SOCIETY. These three general levels are then subdivided into specific classifications, which in turn display even more specific subclasses. Table 6 provides an overview of the major semantic classes within each of the three most schematic levels.

Table 6. Major semantic classes in the *HTOED*

THE EXTERNAL WORLD	THE MIND	SOCIETY
THE UNIVERSE	MENTAL CAPACITY	SOCIETY AND COMMUNITY
THE EARTH	ATTENTION AND JUDGEMENT	INHABITING AND DWELLING
LIFE	GOODNESS AND BADNESS	ARMED HOSTILITY
HEALTH AND DISEASE	EMOTION	AUTHORITY
PEOPLE	WILL	LAW
ANIMALS	POSSESSION	MORALITY
PLANTS	LANGUAGE	FAITH
FOOD AND DRINK		COMMUNICATION
TEXTILES AND CLOTHING		COMPUTING AND INFORMATION TECHNOLOGY
PHYSICAL SENSATION		TRAVEL
MATTER		OCCUPATION AND WORK
EXISTENCE AND CAUSATION		TRADE AND FINANCE
SPACE		LEISURE
TIME		
MOVEMENT		
ACTION OR OPERATION		
RELATIVE PROPERTIES		
THE SUPERNATURAL		

For instance, the noun *bath*, as used in the example discussed above, is classified in the *HTOED* as follows, thus providing a number of semantic categories for this word at different degrees of granularity:

THE EXTERNAL WORLD → PHYSICAL SENSATION → CLEANNESS AND DIRTINESS → CLEANING → WASHING → WASHING ONESELF OR BODY → [noun] → BATHING.

The semantic classes are ordered in chronological order, that is, words in a particular class are ordered by date of first attestation. Therefore, even obsolete or rare senses are included in the classification, which is important for a study of a diachronic nature such as the present one. Additionally, in the more specific categories, it is possible to retrieve only the words belonging to a given POS, such as noun in the case of *bath*.

Both *USAS* and the *HTOED* provide detailed hierarchies of semantic classes and domains, which in fact overlap to a large degree. For instance, both sources distinguish categories for body parts, people, plants, time, and weather. However, some important differences between them are worth highlighting, especially concerning their respective advantages and disadvantages. On the one hand, whereas *USAS* automatically tags words according to their semantic category —albeit with a certain degree of manual editing—, the *HTOED* is less automatic, as the researcher has to select manually the correct sense for the lexical items of interest in the *OED* before being able to retrieve its semantic classification.[59] On the other hand, while the *HTOED* is extremely sense specific, that is, the same word typically occurs in many different semantic classes, this is often not the case of *USAS*, which commonly provides only one or two tags for each word. In fact, many examples of metonymies and metaphors are not captured by *USAS*, while they do appear in the *HTOED*.[60] To illustrate this point, consider example (50):

(50) Down in the vale where cowslips are growing, Where violets breathe thro' sweet **scented lips**, Where brook o'er the bright pebbly bottom is flowing, And bee of the nectar of columbine sips. A monarch it stands of regnative power, In a graceful symmetrical pose; (*COHA*, 1895, FIC, OurProfessionOther)

In (50), the noun *lips* is a typical example of metaphorical sense extension, in which it is used to refer to a part of a flower, rather than to a body part, thus being used with the meaning '[s]omething resembling the lips of the mouth', specifically '[o]ne of the two divisions of a bilabiate corolla or calyx' (*OED*, s.v. *lip* n. II. 5c). The *HTOED* provides an entry for this sense of the noun, whereas *USAS* classifies this instance of *lip* as a body part (i.e. B1: ANATOMY AND

[59] One can also use the *HTOED* directly by searching for all the items included in a given semantic class (for such use of the *HTOED*, cf., for instance, Allan 2015 on EDUCATION). However, in order to do so, it is first necessary to know which specific category or categories to examine, which is not the case in the present dissertation, whose goal is to discover to which category the modified nouns belong in a particular context of use.

[60] The definitions of metaphor and metonymy followed in the present dissertation are those of conceptual approaches within cognitive semantics (cf., for instance, Lakoff 1987; 1993; Lakoff & Johnson 2003 for metaphor and Panther & Radden 1999; Kövecses 2010 for metonymy; for more information, see Section 2.4).

PHYSIOLOGY), even when access is given to the whole context in (50) and even though this is a fairly well-established metaphor.[61]

The classification of nouns modified by *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling* resulted in a total of twelve semantic categories: (i) ABSTRACT; (ii) BODY AND PEOPLE; (iii) CLEANING; (iv) COSMETICS; (v) EARTH, ATMOSPHERE, AND WEATHER; (vi) FOOD AND DRINK; (vii) OBJECT; (viii) PLANTS AND FLOWERS; (ix) SENSATION; (x) SPACE; (xi) SUBSTANCE AND MATERIAL; and (xii) TEXTILE AND CLOTHING.

The category ABSTRACT groups together under one single heading all abstract nouns, such as those denoting actions, processes, beliefs, states, and temporal terms. These are all entities which cannot emit an odor and thus the near-synonymous adjectives, when modifying such nouns, must be interpreted in the figurative sense 'sweet or agreeable'. An example is provided in (51):

(51) When she slips away in the dusk to-night I shall put a period to my thought of Maria de Guadalupe Rosalia Merced Castello. I want to keep this **fragrant memory** of her. (*COHA*, 1922, FIC, JaneJourneysOn)

The category BODY AND PEOPLE contains nouns designating human beings and human body parts, together with proper nouns and pronouns that refer to people. Both *USAS* and the *HTOED* propose separate classes for body (i.e. B1: ANATOMY AND PHYSIOLOGY and THE BODY, respectively), on the one hand, and for people (i.e. S2: PEOPLE in *USAS* and PEOPLE in the *HTOED*), on the other. However, in the present study the decision was made to keep the two together due to the similarities of the examples, which all refer to the pleasant smell people exhibit, either coming from a specific body part, as in (52), or in general, as in (53).

(52) The Spencer woman was there with us -- before us -- all around us. "I am Armand Dalberg's wife" was pounding in my brain. Then I felt a soft little hand slip into mine; a **perfumed hair tress** touched my cheek; and the sweetest voice, to me, on earth whispered in my ear. (*COHA*, 1906, FIC, ColonelRedHuzzars)

[61] In fact, body parts constitute a very common source of conventional metaphors, as they represent familiar concepts, e.g. *head/foot of a mountain* or *mouth of a cave* (Kay & Allan 2015: 159–160; cf. also Kövecses 2010: 18).

(53) Tea and toast unobserved before them, music drifting unheard about them, furred and **fragrant women** coming and going; all this was but the vague setting for their own thrilling drama of love and confidence. (*COHA*, 1921, FIC, BelovedWoman)

A few instances of occasional meanings of clothing nouns which present extensions of meaning via the metonymical pattern PIECE OF CLOTHING FOR PERSON (e.g. Waag 1908; Nyrop 1913; Paul 1920; Esnault 1925; cf. also Geeraerts 2010: 33) or OBJECT USED FOR USER (Lakoff & Johnson 2003: 38; Kövecses 2010: 172), previously mentioned in Section 2.1, were also categorized as BODY AND PEOPLE. This is the case of (54), where *bodice* does not refer to the piece of clothing, but to the person wearing it:

(54) One member of that body was not present. Well-nigh all Williamsburgh knew by now that Mr. Marmaduke Haward lay at Marot's ordinary, ill of a raging fever. Hooped petticoat and **fragrant bodice** found reason for whispering to laced coat and periwig; significant glances traveled from every quarter of the building toward the tall pew where, collected but somewhat palely smiling, sat Mistress Evelyn Byrd beside her father. (*COHA*, 1902, FIC, Audrey)

Nouns referring to cleaning and personal care products were classified into two separate categories, following the organization in the *HTOED*, but contrary to that in *USAS*, which keeps the two together under the label CLEANING AND PERSONAL CARE (i.e. class B4). These two categories are (i) CLEANING, which encompasses those products whose essential purpose is that of cleansing, either the body or areas and objects which people inhabit or use on a daily basis, and (ii) COSMETICS, which includes those products used mainly to enhance or improve physical appearance, that is, for beautification purposes. Examples of *perfumed* and *sweet-smelling* modifying a CLEANING and a COSMETICS noun are provided in (55) and (56), respectively:

(55) He watched her empty the leather pouch, daintily clean it out, fill the washbowl again and sink into it to her upper arms. The **soap** was **perfumed**. She inhaled deeply." Luxury is everywhere." (*COHA*, 1973, FIC, EagleEye)

(56) "This smoke is getting thick. Some of this toilet water might help if I sprinkled it about." One whiff of the **sweet-smelling cologne** was enough for Bragdon and he bolted up the companionway, leaving the stateroom door wide open and the prisoner free to go where he pleased. (*COHA*, 1902, FIC, BrewstersMillions)

Apart from instances like (56), CLEANING also comprises cases such as (57), where *tub* is an example of an extension of meaning through the conceptual metonymy CONTAINER FOR CONTENTS (e.g. Kay & Allan 2015: 165). Here what is fragrant is not the tub itself, but the water it contains, which has been impregnated with a cleansing agent of some type.

(57) Sherry, always an early riser, had already breakfasted in her chamber and was, Adam knew, even now sunk deep in a hot, **fragrant tub**. Washing herself clean of him, arming herself for another day of battle. (*COHA*, 2000, FIC, ComeNearMe)

Nouns referring to natural geographical terms, expressions related to the weather, and atmospheric conditions were coded as EARTH, ATMOSPHERE, AND WEATHER. A great majority of the nouns belonging to this category are grouped into the rather general class THE EARTH in the *HTOED,* particularly within the subclasses LAND, WATER, and WEATHER AND THE ATMOSPHERE. In *USAS* nouns of these types are included in the category W: THE WORLD AND OUR ENVIRONMENT, particularly in W3: GEOGRAPHICAL TERMS and W4: WEATHER. Thus, the decision of maintaining such nouns together in one individual category is well justified, as they in fact belong to the same general semantic domain in both the *HTOED* and *USAS*. An example of this category is given in (58):

(58) Turning from the mountain scenes we have described, let us back once more to Constantinople, and direct our footsteps up the **fragrant valley** where the Barbyses threads its meandering course. (*COHA*, 1851, FIC, CircassianSlave)

In this category we also find a wide range of metonymical or metaphorical extensions of meaning in which temporal expressions and states are used to refer to the surrounding environment and atmospheric conditions of an outdoor location, and sometimes even to the

characteristic actions and processes of the natural entities present in those locations. Consider (59) and (60):

(59) And how we used to drive the sleepy old cows through **the wood and meadow land**, wading knee deep in the clover that seemed to be holding up its red lips to be kissed, while the **wild flowers** like gossamer chalices filled with dew, made the rosy **morning fragrant**. (*COHA*, 1873, FIC, WhiteSlave)

(60) Accordingly, she had attired herself in a becoming negligee, and had spent the fore part of the night somewhat restlessly, occasionally emerging on the veranda and gazing down **into the perfumed gloom of the garden**. (*COHA*, 1892, FIC, GoldenFleeceRomance)

In (59), the noun *morning* does not denote '[…] the early part of the day, esp. from sunrise until noon or lunchtime' (*OED*, s.v. *morning* n. A 1a), but instead refers to '[…] the early part of a day as characterized by the particular weather, conditions, sentiments, etc., prevailing or experienced during that time' (*OED*, s.v. *morning* n. A 2). Therefore, *morning* is in this instance not just an abstract temporal expression, but it also refers to the concrete surroundings and atmospheric conditions of *the wood and meadow land* through which the narrator and other individuals drive the cows. In fact, it is evident that the smell stems from the *wild flowers* mentioned in the example. Consequently, instances of the type in (59) would correspond to the metonymical pattern TIME PERIOD FOR A CHARACTERISTIC ACTIVITY IN THAT PERIOD (Kövecses 2010: 258), in which the characteristic activity could be that of the flowers blooming again in the morning after closing at night. Note that TIME is a particularly complex concept. In fact, it is one of the most common target domains in metaphor and metonymy and thus appears in a wide range of different mappings (Kövecses 2010: 26), which makes it particularly difficult to pinpoint the exact pattern of such examples. Similarly, in (60), *gloom* does not just refer to the darkness of the garden, but also to the garden itself, which is most probably filled by flowers and plants. The conceptualization of the state as a space is reinforced by the preposition *into* (*into the perfumed gloom of the garden*), which expresses motion or direction. As in (59), difficulties arise when trying to characterize the mapping, given that SPACE is also a common target domain. However, one possibility could be that of the metaphorical pattern STATES ARE

LOCATIONS, where a more abstract aspect is understood in terms of physical location (Kövecses 2010: 163). Be that as it may, the importance here lies not in identifying the specific figurative pattern at stake, which is beyond the purposes of the present dissertation, but in the fact that when the adjectives modify such nouns, these nouns are often not used literally but figuratively.

Within the category FOOD AND DRINK we find nouns corresponding to the semantic class with the same label in the *HTOED* and to the F1: FOOD and F2: DRINKS *USAS* categories, which include edible and drinkable products, as well as nouns referring to different meals of the day (e.g. *breakfast*, *lunch*, and *dinner*). Example (61) was, therefore, grouped into this category:

(61) Every morning believers came to offer Han-shan a bowl of **fragrant soup** that steamed his face as if to make him sweat. (*COHA*, 1970, NF, BuddhistLeader)

Moreover, the category FOOD AND DRINK also comprises several metonymic examples corresponding to the conceptual metonymy CONTAINER FOR CONTENTS (e.g. Kay & Allan 2015: 165), already discussed in relation to CLEANING, in which the near-synonymous adjectives occur with nouns denoting containers (e.g. *bowl*, *cup*, and *glass*) but modify the contents within them, as in (62):

(62) They enjoy afternoon tea precisely as other women enjoy it, and over a **fragrant cup** they relax, talk and laugh and tell stories and have what they need more than anything else in their lives, a really good time. (*COHA*, 1905, NF, RadiantMotherhood)

A relatively general category is that of OBJECT, as it contains a mixture of concrete nouns referring to material things which can be seen and touched, including furniture (e.g. *bed*, *couch*, and *drawer*), paper documents (e.g. *letter*, *envelope*, and *book*), and other objects (e.g. *box*, *candle*, and *lamp*). Many of these nouns belong to the semantic classes MATTER in the *HTOED* and O2: OBJECTS GENERALLY, Q1.2: PAPER DOCUMENTS AND WRITING, and Q4: THE MEDIA in *USAS*, among others. The decision was made to group these more specific classes into a more general one due to data sparseness, as retaining the more specific categories would yield very low figures, which would make it difficult to obtain significant results. An example of OBJECT is shown in (63):

(63) We who produce the world where Heirston finds John Robshaw pillows along with Calypso's own Pom Pom cotton throws and **scented candles** ($38 each) in eleven fragrances. (*COHA*, 2009, MAG, COCA)

Categories L3: PLANTS in *USAS* and PLANTS in the *HTOED* were here classified under the label PLANTS AND FLOWERS. This category is illustrated in (64):

(64) Then the angel went down to the earth, and he came to a beautiful rose-bush upon which bloomed a **rose** lovelier and more **fragrant** than any of her kind. (*COHA*, 1896, FIC, SecondBookTales)

Again, we find several metonymical instances in this semantic category. In this case, the metonymical extension follows the pattern CHARACTERISTIC FOR CHARACTERIZED ENTITY (Waag 1908; Nyrop 1913; Paul 1920; and Esnault 1925; cf. also Geeraerts 2010: 32), in which the near-synonymous adjectives modify a noun that denotes an attribute of a plant or a flower. Such an example is illustrated in (65):

(65) The woman sprang back from the **flowers** as though a poisonous serpent, hidden in their **fragrant beauty**, had struck her. (*COHA*, 1912, FIC, TheirYesterdays)

In this example, even though *fragrant* modifies the noun *beauty*, the pleasant and sweet smell applies to the flowers, since *beauty* here is clearly a characteristic of the flowers.

The ninth category, SENSATION, includes nouns referring to any of the five physical senses, that is, smell (e.g. *aroma*, *odor*, and *scent*), taste (e.g. *flavor* and *taste*), hearing (e.g. *music*, *sound*, and *voice*), sight (e.g. *look*), and touch (e.g. *kiss*). Evidently, the great majority of the nouns in my data correspond to the sense of smell, since the adjectives under analysis belong to this particular semantic domain, as in example (66), where *fragrant* collocates with *smell*:

(66) He spread the faded army blanket, which was old but clean, on the straw of the second loft. The **smell** was **fragrant** and sweet. (*COHA*, 1979, FIC, SilverGhost)

Nevertheless, several instances of nouns belonging to one of the other four senses were also found in the data and were therefore included in this category, which corresponds to the broad semantic classes PHYSICAL SENSATION in the *HTOED* and X3: SENSORY in *USAS*. In general, when the adjectives at issue here modify any of the senses apart from smell, they are used in the figurative sense 'pleasant or agreeable', as in example (67), where *perfumed* modifies the noun *murmur*, which refers to '[a] word or sentence spoken softly or indistinctly; faint or barely audible speech […]' (*OED*, s.v. *murmur* n. 4):

(67) He picked up one of the small books. "This is The Book of Everything." There was a rustle among the girls and a **perfumed murmur** of "Everything?" "Everything." he said firmly. (*COHA*, 1958, FIC, Maggie-Now)

Examples such as (67) are instances of *synesthetic metaphors*, which describe one sensory domain on the basis of another one (e.g. Geeraerts 2010: 35). Here, an adjective from the olfactory domain is used to describe a noun from the auditory domain, and thus *perfumed* is used figuratively to mean simply 'sweet or agreeable'.

While the category EARTH, ATMOSPHERE, AND WEATHER included, as seen above, natural geographical terms, among other types of nouns, the category SPACE contains nouns referring to enclosed or indoor locations, such as *room*, *boudoir*, and *kitchen*, as well as those denoting human —as opposed to natural— geographical terms, for instance, *avenue*, *road*, and *city*. The principal reason for maintaining these two categories separate is that whereas the category EARTH, ATMOSPHERE, AND WEATHER contains instances of the adjectives solely in the natural sense, SPACE comprises also occurrences of the adjectives in the artificial sense, as well as indeterminate cases. Additionally, while the great majority of nouns in the former group are classified as THE EARTH in the *HTOED*, especially within the subclasses LAND, WATER, and WEATHER AND THE ATMOSPHERE, and in W3: GEOGRAPHICAL TERMS and W4: WEATHER in *USAS*, this is not the case of nouns in the category SPACE. Instead, these nouns appear mainly in the general category SOCIETY, especially within the more specific classes INHABITING AND DWELLING and TRAVEL, in the *HTOED*, and in H: ARCHITECTURE, BUILDINGS, HOUSES AND THE HOME and in M3: MOVEMENT/TRANSPORTATION: LAND and M7: PLACES in *USAS*. An example of the category SPACE is given in (68), where *scented* modifies the noun *room*:

(68) As he and Jock left the warm, **scented room** behind them, and faced the white, still cold of an apparently dead St. Ange, the boy turned a drawn face upon Jock, and cried tremblingly, "Say, you". (*COHA*, 1911, FIC, JoyceNorthWoods)

As in the case of EARTH, ATMOSPHERE, AND WEATHER, the category SPACE also contains instances of metaphors following the pattern STATES ARE LOCATIONS, but in which the space is an indoor location, as in example (69):

(69) Look, there are **orchids and camellias and oleanders**.' He was silent so long that she thought he hadn't heard her. She looked up inquiringly and saw that he was staring not into the **perfumed dampness of the hall** but at her. (*COHA*, 1944, FIC, Dragonwyck)

Similarly to example (60) above, *dampness* does not just refer here to the condition or state of the hall, but also to the hall itself, that is, an indoor space, where there are several flowers (i.e. orchids, camellias, and oleanders).

The category SUBSTANCE AND MATERIAL includes nouns referring to liquid, solid, or gaseous substances (e.g. *ash*, *ether*, *liquid*, and *steam*), and to materials from which objects are made (e.g. *amber* and *wood*). Moreover, this category also contains nouns referring to drugs, especially tobacco (e.g. *cigar*, *cigarette*, and *tobacco*). These drug-related nouns, very few in number, were included here as they are in fact substances of a kind. In *USAS* most of the nouns belonging to SUBSTANCE AND MATERIAL are grouped into the semantic class O1: SUBSTANCES AND MATERIALS GENERALLY, whereas they mainly correspond with the general category MATTER in the *HTOED*. Example (70) illustrates this semantic category:

(70) Most wood gives off a pleasant aroma when it's cut, but **wood dust** is more than just **fragrant** - it's also hazardous to your health. (*COHA*, 2003, MAG, MotherEarth)

Finally, the last category, TEXTILE AND CLOTHING, comprises nouns referring either to textile material (e.g. *linen*, *silk*, and *wool*) or to different types of clothing (e.g. *clothes*, *dress*,

garment, and *glove*). These nouns all belong to the categories TEXTILES AND CLOTHING in the *HTOED* and B5: CLOTHES AND PERSONAL BELONGINGS, in the case of clothing, and O1.1: SUBSTANCES AND MATERIALS GENERALLY: SOLID, in the case of textile material, in *USAS*. An example is provided in (71), where *perfumed* collocates with the noun *gloves*:

(71) […] and lay piled in heaps beneath the Sand Hill fort-many youthful gallants from Spain and Italy among them, noble volunteers recognized by their **perfumed gloves** and golden chains. (*COHA*, 1868, NF, HistoryUnitedNetherlands)

The twelve semantic categories distinguished and illustrative examples of nouns belonging to each of them are provided in Table 7.

Table 7. Classification of nouns into semantic categories

SEMANTIC CATEGORY	EXAMPLES OF NOUNS
ABSTRACT (ABS)	*charm, darkness, fear, knowledge, memory, personality*
BODY AND PEOPLE (B&P)	*arm, bodice, cheek, girl, hair, lock, woman, wrist*
CLEANING (CL)	*bath, deodorant, disinfectant, shampoo, soap, tub, water*
COSMETICS (COS)	*cologne, cosmetics, cream, gloss, oil*
EARTH, ATMOSPHERE, AND WEATHER (EAW)	*air, breeze, darkness, gloom, hill, mist, morning, rain, summer, valley, water*
FOOD AND DRINK (F&D)	*apple, beverage, bowl, chicken, coffee, cup, glass, rice*
OBJECT (OBJ)	*book, box, candle, couch, lamp, letter, notepaper, stationary*
PLANTS AND FLOWERS (P&F)	*beauty, bloom, bud, flower, leaf, loveliness, pine, rose, shrub, tree*
SENSATION (SEN)	*aroma, flavor, kiss, look, music, odor, scent, smell, taste, tone*
SPACE (SPA)	*avenue, bakery, boudoir, chamber, city, dampness, house, room*
SUBSTANCE AND MATERIAL (S&M)	*amber, dust, fume, liquid, oil, smoke, steam, vapor, water, wood*
TEXTILE AND CLOTHING (T&C)	*cambric, cloth, dress, glove, linen, pillow*

Since the classification was made on the basis of the referents of the nouns, which may vary from instance to instance, some of the nouns, especially those which are highly polysemic, such as *breath*, *darkness*, *lip*, *oil*, and *water*, were classified into more than one semantic category (cf. Table 7). First, the adjectives at issue here can function as modifiers of several senses of

the same noun. The following examples of *breath* modified by *perfumed* and *fragrant*, respectively, serve to illustrate this point:

(72) Close by her side I rested, and once when she stooped toward me in the telling of the tale, I felt her **perfumed breath** upon my cheek. (*COHA*, 1934, FIC, Witchfinders)

(73) And they passed under a shadowed grape-arbour and past a dead locust-tree, which a vine had made into a green tower of waving tendrils, and from which came the **fragrant breath** of wild grape […] (*COHA*, 1900, FIC, Crittenden)

While *breath* in (72) refers to '[t]he air exhaled from the lungs' (*OED*, s.v. *breath* n. 3a), in (73) it refers to '[t]he air exhaled from anything, or impregnated with its exhalations, and retaining its characteristic odour' (*OED*, s.v. *breath* n. 2b). Therefore, these two senses of *breath* were classified into two distinct categories, namely BODY AND PEOPLE (sense 3a) and SENSATION (sense 2b), following the categorization in the *HTOED*. The same is true of nouns which have developed figurative meanings besides their literal ones, as is the case of conventional metaphors. This happens with, for instance, the noun *lip*, which appears in the category BODY AND PEOPLE, as in (74), and in PLANTS AND FLOWERS, as in (50), repeated here for convenience as (75):

(74) […] Diane swayed toward him and his arms enfolded her. Her body met his, yielded; her face was upturned; her **fragrant,** half-opened **lips** were crushed to his in a fierce, impassioned kiss of genuine ecstasy. (*COHA*, 1917, FIC, LaughingBillHyde)

(75) = (50) Down in the vale where cowslips are growing, Where violets breathe thro' **sweet scented lips**, Where brook o'er the bright pebbly bottom is flowing, And bee of the nectar of columbine sips. A monarch it stands of regnative power, In a graceful symmetrical pose; (*COHA*, 1895, FIC, OurProfessionOther)

Another example is that of the noun *darkness*, which, similarly to its near-synonym *gloom* (cf. example (60)), can be used metonymically to denote a location along the mapping STATES ARE LOCATIONS. This is the case of example (76), where *darkness* has been grouped into the category

EARTH, ATMOSPHERE, AND WEATHER, as it does not only refer to the darkness of the place Natasha is entering, but also to the place itself:

(76) No nun could mistake her meaning. Entering steamy **perfumed darkness**, dimly lit by sunlight coming through slits in the stone, Natasha stripped before a hot foamy cascade, sponging away the sweat and grime of travel, carefully cleaning the pink flesh growing around her scab. (*COHA*, 2003, FIC, FantasySciFi)

Darkness in (77), in turn, belongs to the semantic category ABSTRACT, since it here refers to a psychological condition, namely that of '[g]loom of sorrow, trouble, or distress' (*OED*, s.v. *darkness* n. 5).

(77) She might be grave by nature, she might be sad by circumstance, she might have secret doubts and pangs, but she was essentially young and strong and fresh and able to enjoy. Her enjoyment was not especially demonstrative, but it was curiously diligent. Rowland felt that it was not amusement and sensation that she coveted, but knowledge -- facts that she might noiselessly lay away, piece by piece, in the **perfumed darkness** of her serious mind […] (*COHA*, 1876, FIC, RoderickHudson)

Second, occasional non-conventional meanings of nouns emerged in specific contexts, which led to the classification of some of them into semantic categories to which their usual conventional meanings do not belong. A case in point is (78), where the noun *sister* refers to a flower (i.e. *sassafras*) and is thus coded as PLANTS AND FLOWERS:

(78) It was a swell of land, on the summit very rocky, covered with beech and maple trees, and with an undergrowth of spice-wood and its **fragrant sister**, the **sassafras**. (*COHA*, 1835, FIC, HawksHawk-Hollow)

Similarly, many of the types of metonymical extensions of meaning mentioned throughout this section (e.g. CONTAINER FOR CONTENTS, CHARACTERISTIC FOR CHARACTERIZED ENTITY) also result in a deviation from the conventional semantic categorization of nouns. By way of

illustration, *tub* was categorized as CLEANING rather than as OBJECT in example (57) above. An even more complex deviation from the conventional meaning of a noun is displayed in (79), where *fragrant* modifies the noun *womb*:

(79) "I have brought down all of the old homestead that I could," said Martha, who had her **willow basket** at the bedside; "and it is here." She unclasped it; and as Fob glanced down into its **fragrant womb**, his eyes shone with a new light. He saw whole tracts and acres there. "These, you know," continued Martha, producing a handful of **green cresses**, "I plucked them from the Mower's Nook in the wood, so calm and shady in the summer time. You remember it?" (*COHA*, 1842, FIC, CareerPufferHopkins)

Here, *womb* is used metaphorically to denote a specific part of the *willow basket* mentioned in the previous sentence, namely its interior. This metaphorical extension of *womb* corresponds to the meaning '[s]omething likened to the uterus (†or belly) in being hollow or enveloping; a hollow space or cavity' (*OED*, s.v. *womb* n. 4). However, in this instance a metonymy following the pattern CONTAINER FOR CONTENTS is also present, with *womb* referring not to the actual *willow basket*, which is here a container, but to its contents, namely the *green cresses* that Martha plucked from the wood and which she now has in her willow basket.[62] Therefore, *womb* in this instance was classified as PLANTS AND FLOWERS. This example demonstrates that the semantic categorization of nouns is sometimes not straightforward and that the extended context must in some cases necessarily be analyzed in depth to reach an adequate classification.

A further complication in the categorization process that needs to be acknowledged is the lack of a clear-cut distinction between some of the twelve semantic categories, with some nouns being located at the periphery of more than one category. This is in line with the non-discrete nature of semantic categories demonstrated by prototype theory (cf. Section 2.4.1). For instance, nouns such as *berry*, *orange*, and *herb* can be said to lie at the border between the categories FOOD AND DRINK, on the one hand, and PLANTS AND FLOWERS, on the other. Such

[62] The interaction of the two cognitive processes of metaphor and metonymy has been labeled *metaphtonymy* (Goossens 1990). The interplay between the two processes is discussed in, for instance, Geeraerts (2002), Ruíz de Mendoza & Díez (2002), and Ruíz de Mendoza & Galera-Masegosa (2011).

nouns can therefore be conceptualized either as food or as parts of plants (most commonly the former in my data), and again the context of use must therefore be carefully examined to reach a decision. Similarly, the distinction between the categories CLEANING and COSMETICS is not devoid of problems either. Consider in this connection example (80):

(80) When it is a question of cleansing the hair of the head, washing should only be employed with the greatest caution... Instead, rub with bran stewed in a pan, changing it frequently; or spread over and between the hair some **drying and cleansing powder** before going to bed, and remove it in the morning with a comb.17 **Powder** had, in fact, already enjoyed a lasting success at the court of Henri III. **Sweet-smelling**, it was no longer only a **means of washing the hair**, but had become a **hair cosmetic**. (*COHA*, 1988, NF, ConceptsCleanliness)

The noun *powder* is almost always conceptualized as a cleansing agent in my data, thus belonging to the category CLEANING. However, (80) is somewhat ambiguous, showing that the noun could also be regarded as a hair cosmetic, thus maybe belonging to the category COSMETICS as well. Given that in all other cases analyzed, *powder* is clearly a CLEANING noun and that the context in (80) does not help to disambiguate, this example has been categorized as such.

Finally, it is worth mentioning that a correspondence can be established between the variables Semantic category, discussed in this section, and Sense (cf. Section 4.3.1.1). Of the twelve categories, three comprise only nouns which refer to entities that can release a smell on their own, namely EARTH, ATMOSPHERE, AND WEATHER, FOOD AND DRINK, and PLANTS AND FLOWERS. In other words, these three categories correspond with the natural sense. On the other hand, CLEANING, COSMETICS, and TEXTILE AND CLOTHING include only nouns which exhibit artificial smells, thus corresponding with the artificial sense. Moreover, the category ABSTRACT displays a one-to-one correspondence with the figurative sense. The five remaining categories, that is, BODY AND PEOPLE, OBJECT, SENSATION, SPACE, and SUBSTANCE AND MATERIAL, include nouns which can denote either an artificial or a natural smell depending on the context in which they are used, as well as indeterminate instances (cf. example (48)). For instance, the noun

smoke, which belongs to SUBSTANCE AND MATERIAL, can refer either to a natural or to an artificial smell:

(81) The trees were being vigorously pruned, and burning heaps of **branches**, a few leaves still adhering, left trailing scarves of **fragrant smoke** through the tattered orchards. (*COHA*, 1930, FIC, Seed)

(82) Strange white boys looked at him from a battered car, laughing and smoking **cigarettes**. "Man," one of them said enviously, blowing acrid, **scented smoke**. (*COHA*, 1995, FIC, SouthernRev)

In (81), *fragrant* is clearly used to refer to a natural smell, namely that of the smoke from burning branches. Contrarily, in (82), *scented* modifies the noun *smoke* referring to an artificial odor, as it denotes the smoke of a cigarette. This correspondence between the two variables Semantic Category and Sense poses a problem for the type of analysis carried out in Chapter 6, where this issue is further discussed (cf. Section 6.1 below).

4.3.1.3 Other semantic variables: Animacy, Concreteness, and Countability

The nouns modified by the adjectives *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling* were also annotated for three additional semantic factors, to wit, Animacy, Concreteness, and Countability, which provide relatively schematic semantic classifications.

Following, for instance, Gries (2001) on adjectival *-ic* and *-ical* pairs (cf. Section 3.2.2), the variable Animacy contains two levels: animate and inanimate nouns.[63] The value animate refers to nouns denoting living entities, an overwhelming majority of which are human beings in my data (e.g. *women* in example (83)).

[63] I am well aware that conceptualizing animacy as a binary choice between animate and inanimate referents is probably too limiting in contrast to animacy scales or hierarchies such as human, animal nouns, collective nouns, temporal nouns, locative nouns, and common inanimate nouns (e.g. Rosenbach 2008: 153) or human, organizations, animals, intelligent machines, vehicles, concrete inanimate, non-concrete inanimate, place, and time (e.g. Zaenen et al. 2004: 121). For a particularly detailed division of animacy, cf. Yamamoto (1999). However, for the purposes of the analyses in the present dissertation, the data was annotated only for the binary distinction animate vs. inanimate in order to avoid data sparsity. Remember that a fine-grained classification of nouns according to their semantic category with twelve different levels is used (cf. Section 4.3.1.2), which makes it inadvisable to consider further variables with a high number of levels (cf. Section 6.1).

(83) On the crowded courtroom benches the red tarbooshes bobbed up &; down. The whispers of the **perfumed** mascaraed **women** rose to an excited buzz. (*COHA*, 1945, MAG, TIME)

However, following scholars such as Yamamoto (1999: 23), some typically inanimate nouns, such as those referring to weather or plants and flowers, were also sometimes classified as animate when they were personified, that is, given human or animate characteristics, especially in poetic or highly figurative texts. This is the case in example (84), where all flowers and plants mentioned (e.g. *Blue-bells*, *Marigolds*, and *Mr. Thistle*), including the *fragrant Lily*, are described as people. In all other cases, PLANTS AND FLOWERS nouns were classified as inanimate, as in animacy hierarchies they tend to be given a fairly low animacy score (cf., for instance, Corbett 1981: 59; Yamamoto 1999: 16, 38).

(84) And the **Blue-bells** and **Hare-bells**, in simple array, With all their Scotch cousins from highland and brae. **Ragged Ladies** and **Marigolds** clustered together, And gossiped of scandal, the news, and the weather -- What dresses were worn at the wedding so fine Of sharp **Mr. Thistle** and sweet **Columbine**; Of the loves of Sweet **William** and **Lily** the prude, Till the clamours of Babel again seemed renewed. In a snug little nook sate the **Jessamine** pale, And that pure **fragrant Lily**, the gem of the vale. (*COHA*, 1834, FIC, Poems)

On the contrary, the value inanimate was given to those nouns denoting non-living entities. An example is provided in (85), where *sweet-smelling* collocates with *jar*:

(85) Katya's body cried for the fluids that had spilled out the hole in her side. Blue-Eyes propped her up, which hurt terribly, and held a **sweet-smelling** clay **jar** to her lips. (*COHA*, 2003, FIC, FantasySciFi)

As regards the degree of concreteness of a noun modified by the adjectives at issue, this was coded for in two different variables, Concreteness rating and Concreteness binary.

Brysbaert, Warriner & Kuperman (2014) provide a concreteness rating of about 40,000 English lexical items —37,058 words and 2,896 two-word expressions. By means of an online survey, they asked 4,000 participants from different groups to rate a series of lexical items according to how concrete they thought the items were, ranging from 1 (abstract) to 5 (concrete). Explicit instructions were given to the participants to consider all five physical senses when giving a rating for each item, rather than basing their decisions on only one or two of them (e.g. sight and touch). The average rating for each word was included in a database.[64] The concreteness ratings of the nouns modified by the near-synonymous adjectives analyzed in this dissertation were extracted from this database. For instance, a concrete noun like *handkerchief* was given an average rating of 5, whereas an abstract noun such as *gloom* was given a rating of 1.86. Even though the vast majority of modified nouns in my data were found in Brysbaert, Warriner & Kuperman's (2014) database, some were not included. In those cases, the rating of a semantically related word, most typically a hyperonym, was used instead to retrieve the rating. By way of illustration, on one occasion *fragrant* modified the noun *wildrose*, which was not included in Brysbaert, Warriner & Kuperman's (2014) database. Therefore, the rating for the hyperonym of *wildrose*, that is, *rose* (i.e. 4.9), was used instead.

Despite the fact that the participants were asked to bear all five physical senses in mind when providing a rating, the authors conclude that participants attach more importance to sight and touch than to the other three senses. Given that the present analyses focus on adjectives from the olfactory domain and that context of use is of crucial importance here, the data was annotated for an additional variable, to wit, Concreteness binary. This variable offers a more qualitative assessment of the concreteness of the modified nouns, since it takes into account the sense in which each noun is used as well as the referent it denotes in each particular context. As the label itself suggests, Concreteness binary comprises two levels, namely concrete, that is, entities which can be perceived by any of the five senses, and abstract, that is, those which cannot be experienced directly through the senses. Examples (86) and (87) show an instance of a concrete (*apple*) and of an abstract noun (*malice*), respectively:

[64] The database with the average concreteness rating for the almost 40,000 English lexical items is available at: http://crr.ugent.be/archives/1330.

(86) Their return from these visits was still more grotesque, for their family-carriage generally trundled into town garnished with baskets of fresh, **sweet-scented apples**, and a pair or two of tender poultry […] (*COHA*, 1843, FIC, VariousWritings)

(87) Finally, the late Richard Burton as O'Brien, the couple's betrayer and interrogator, gives a last performance that is all silky corruption, **perfumed malice** in every beautifully measured phrase. (*COHA*, 1985, MAG, Time)

The last variable pertaining to the semantics of the nouns modified by the adjectives at issue here is Countability. Following, for instance, Gries & Otani (2010) on the adjectival near-synonym sets *big*, *great*, and *large*, on the one hand, and *little*, *small*, and *tiny*, on the other, two types of common nouns were distinguished, namely count nouns and non-count nouns (cf. Section 3.2.2). As the labels of the values indicate, the former group contains common nouns referring to entities that can be counted, whereas the latter comprises common nouns denoting entities which cannot be counted (Huddleston & Pullum et al. 2002: Section 5.3.1). A count noun modified by *scented* is seen in (88), while a non-count noun co-occurring with *sweet-smelling* is found in (89):

(88) They both made a fuss over Johnny and petted him like a little dog or a precocious child; she taught him to dance the "Boston," and the Frenchman, who it turned out was a duke or a baron or something, kept offering him drinks and cigars and **scented cigarettes**. (*COHA*, 1930, FIC, 42ndParallel)

(89) A breeze blew in through the opened hay door, and the **hay** was clean and fresh and **sweet-smelling**. (*COHA*, 1961, FIC, JourneyMatecumb)

As in the case of the other variables discussed so far, it is crucial to consider the context of use in order to reach a decision regarding the countability of some nouns. This is specially so with nouns which are countable in some of their senses, and uncountable in others. A case in point is *air*, which is modified by *scented* in (90) and (91):

(90) At the town florist's he rapped a timid signal to the driver to stop, and, glowing with anticipation, spryly shuffled into the warm, **scented air** of the little shop. (*COHA*, 1913, FIC, UncleNoahsChristmas)

(91) The wide-open windows looked out upon the slopes of that lovely hill on whose summit is perched Fiesole, the poor little old mother of Florence, who still holds watch over her beautiful daughter stretched at her feet. **Scented airs** which had swept all the way from distant blue hills over countless orange, olive, and mulberry groves filled the room, and fluttered the paper upon which the girls were writing; it was their weekly letter budget. (*COHA*, 1899, FIC, BarbarasHeritage)

In (90), *air* is used in the sense '[a] special state, condition, or quality of the atmosphere, as affected by temperature, moisture, etc., or as modified by time or place' (*OED*, s.v. *air* n[1]. I. 4), i.e. a non-count noun. In (91), in turn, *air* refers to the '[a]ir in motion; a current or draught of air; (esp. in the context of sailing) a breeze, a light wind' (*OED*, s.v. *air* n[1]. I. 6), i.e. a count noun. Therefore, although we speak of count and non-count nouns, we in fact mean count and non-count uses of nouns.[65]

Given that the five near-synonymous adjectives also co-occur with proper nouns (cf. example (92)) and pronouns (cf. example (93)), it was necessary to add a third value to the variable Countability, labelled 'other'.[66] Although not applicable to these two types of nominal elements, the decision was taken to maintain the semantic variable Countability because proper nouns and pronouns account only for a relatively low amount of the modified entities in the data, namely 5.7%, of which about 0.9% are proper nouns and about 4.8% are pronouns. This means that over 94% of the modified nominal elements are common nouns which can be characterized as either count or non-count.

[65] For more examples, cf. Huddleston & Pullum et al. 2002: 335–337.

[66] The three reference grammars consulted make a distinction between common and proper nouns. Whereas the former can be further divided into count and non-count nouns, the latter cannot (Quirk et al. 1985: 245–247; Biber et al. 1999: 241; Huddleston & Pullum et al. 2002: 328). It must be noted, however, that Huddleston & Pullum et al. (2002: 520) also claim that the primary uses of proper nouns can be categorized as non-count. Additionally, some proper nouns can acquire uses as common nouns, as in *I got a Bentley, two Cadillacs, a Chrysler station wagon, and an MG for my boy* (Biber et al. 1999: 247; cf. also Quirk et al. 1985: 288–289). However, such uses of proper nouns were not attested in my data.

(92) Tweet, his wife, "a fair thich being"; Mama Crumb, passive housewife; Pearl, "a lovely, listless sister, a too mellow fruit"; **Richmiel**, sleek and **perfumed**, "whose body had seemed nine-tenths of her being." (*COHA*, 1923, MAG, TIME)

(93) Another variety, with white flowers, has yellow berries. This shrub is in bloom early in April. **It** is **sweet-scented**; and, where there are many together, they will perfume the air to a considerable distance. (*COHA*, 1851, NF, FlowerGardenBrecks)

4.3.2 Language-internal non-semantic variables

Besides the semantic features pertaining to modified nouns, the data was also annotated for morphosyntactic characteristics of the five near-synonymous adjectives under analysis. As mentioned in Section 3.2.2, previous studies on adjectival synonym sets have shown that morphological and/or syntactic factors in some cases determine the choice between the members of a synonym set (e.g. Biber, Conrad & Reppen 1998: Section 4.2; Gries 2003; Gries & Otani 2010; Liu 2010). The most common syntactic feature examined for adjectives is their syntactic function, but others, such as degree, have also been considered. Therefore, these two factors were also included in the present analyses under the labels 'Syntactic function' and 'Degree'.

Most reference grammars agree as to the number and types of syntactic functions of adjectives. The three basic functions of adjectives are attributive, predicative complement, and postpositive (e.g. Huddleston & Pullum et al. 2002: 528–529).[67] While we here speak of adjectives as being attributive, postpositive, and predicative, it is not the adjectives alone that perform these functions, but rather the adjective phrases in which they serve as heads (Quirk et al. 1985: 416–417; Biber et al. 1999: 101–102; Huddleston & Pullum et al. 2002: 529).

Attributive adjectives are those that serve as premodifiers of the head of a noun phrase (Quirk et al. 1985: 417; Biber et al. 1999: 510; Huddleston & Pullum et al. 2002: 528). In (94), for instance, *sweet-smelling* occurs in attributive position as a premodifier of the noun *wind*:

(94) The **sweet-smelling** warm **wind** moved gently in Francie's hair. (*COHA*, 1943, FIC, TreesGrowInBrooklyn)

[67] Although different labels are in some cases used in the reference grammars to refer to the different syntactic functions of adjectives, the nomenclature employed in Huddleston & Pullum et al. (2002) is followed here.

As to predicative adjectives, two types can be distinguished. Subject predicatives function as complements of copular verbs, such as *appear*, *be*, and *seem*, among others, and qualify the nominal expression in subject position (cf. example (95)). Moreover, non-copular verbs which fulfil similar functions to copular verbs in particular contexts, such as *grow* and *come*, can also take predicative complements (cf. example (96)). Object predicatives, in turn, typically occur with verbs such as *declare*, *find*, and *make*, and characterize the direct object of the clause, as in (97) (Quirk et al. 1985: 417; Biber et al. 1999: 515–516; Huddleston & Pullum et al. 2002: 528; 530).

(95) Luis drank the tea, sitting on the side of the bed, while he opened the letters. Two of them were the kind he got in every town; **the paper of one** was pink and **scented**, the other pale violet. (*COHA*, 1949, FIC, BraveBulls)

(96) I could smell the tang of woodsmoke from the chimney. **It** came strong and **fragrant** to meet me, and I guessed that someone had just thrown a fresh log on the living-room fire. (*COHA*, 1942, FIC, AlongStreet)

(97) The apple-trees **made the air fragrant**, and some of the delicate pink of their blossoms was in Mousie's cheeks. (*COHA*, 1885, FIC, DrivenBackEden)

Besides the prototypical cases of predicative complements just exemplified, some examples featured adjectives in predicative complement function in constructions in which the subject and the verb of the clause were omitted. Such examples occurred mostly in recipes, a genre which exhibits a high incidence of ellipsis (e.g. Brown & Yule 1983: 174–176; Massam & Roberge 1989; Culy 1996; Todolí & Dolón 2006). In particular, ellipsis is highly frequent in adverbial clauses with *until* (Klenová 2010: 90–91), as in (98), where we find two instances of *fragrant* in a context of this type:

(98) Add shallots and curry paste; stir **until fragrant**, about 1 minute. # Skim off top creamy part of coconut milk (about 1/2 cup) and add to pan. Let mixture bubble and sizzle **until fragrant**, another 2 minutes (*COHA*, 1997, NEWS, SanFran)

Both occurrences of *fragrant* in (98) were classified as predicative complements in a construction which shows the omission of both the subject, *shallots and curry paste* and *mixture*, respectively, and the verb of the clause, probably *be*. Compare (98) to (99), where *fragrant* occurs in a non-elliptical adverbial clause introduced by *until* and functions as a prototypical predicative subject complement:

(99) Add onion paste, paprika and pepper flakes and cook, stirring, **until mixture is fragrant** and begins to brown, about 5 minutes. (*COHA*, 2000, MAG, VegTimes)

Postpositive adjectives immediately follow the nominal element they modify, thus serving as postmodifiers (Quirk et al. 1985: 418–419; Biber et al. 1999: 519; Huddleston & Pullum et al. 2002: 528–529). Even though they are not as common as attributive and predicative adjectives, they are frequently found following indefinite pronouns such as *something*, *somebody*, and *no one* (cf. example (100)).

(100) Sky Eyes took hot water from a kettle near the fire and mixed it with cold in a large, shallow bowl, then dripped **something** light green and **fragrant** into it. (*COHA*, 1991, FIC, FantasySciFi)

Apart from these three basic syntactic roles, adjectives can also occur in a number of minor functions, including predicative adjunct, fused modifier-head, and predeterminer (Quirk et al. 1985: 421–428; Biber et al. 1999: 519–521; Huddleston & Pullum et al. 2002: 529). Predicative adjuncts qualify a nominal element, but do so from a syntactically freer position than predicative complements (Quirk et al. 1985: 424–426; Biber et al. 1999: 520–521; Huddleston & Pullum et al. 2002: 529). Consider in this connection (101), where the predicative adjunct *sweet-smelling or odoursome* is separated by means of commas from the rest of the clause and is thus detached from it:

(101) **Breezes**, **sweet-smelling or odoursome**, would come off the water towards him. (*COHA*, 1988, FIC, "E"IsEvidence)

Predicative adjuncts can occur in different positions in the sentence and are therefore mobile. For instance, in example (101) it is possible to move the adjunct to clause-initial or to clause-final position.

Adjectives serving as fused modifier-heads function simultaneously as modifiers and as heads of the noun phrase. Fused modifier-head adjectives behave almost like nouns but do not exhibit the typical morphological characteristics of nouns, such as inflection for number (Quirk et al. 1985: 421–424; Biber et al. 1999: 519–520; Huddleston & Pullum et al. 2002: 415–418). An example is provided in (102):

(102) Of natural **odors** the most agreeable are **the aromatic and fragrant** that emanate from plants. (*COHA*, 1922, MAG, Century)

In this example, *the aromatic and fragrant* refers to the noun *odors*, which is mentioned in the preceding context. However, in order to avoid repetition of the noun, and since it is clear from the context which noun the two adjectives qualify, *odors* is omitted. Consequently, *aromatic* and *fragrant* combine simultaneously the role of premodifier and that of head of the noun phrase.

Finally, predeterminer adjectives function as external modifiers of noun phrases. According to Huddleston & Pullum et al. (2002: 550–551), adjectives in this function have to fulfil two criteria: first, they must precede the indefinite article *a*/*an* and, second, the adjective phrase in which they appear must either begin with any of the modifiers *how*, *as*, *so*, *too*, *this*, or *that*, or contain *such* or exclamatory *what* as head. In (103), *fragrant* is a predeterminer of the noun phrase *an odor*, and is preceded by the modifier *so*:

(103) She could not understand it, but concluded that it must be some great delicacy among the vessel's stores lying on the bank, which had **so** very **fragrant an** /z/ **odor**. (*COHA*, 1869, NF, BookBoys)

The variable Syntactic function therefore comprises the following values: attributive, predicative complement, postpositive, and a level labelled 'other', with all the minor functions

explained in the preceding paragraphs grouped together, given their low frequency of occurrence (less than 5% of the data).[68]

The second syntactic variable coded for was the degree of the adjective, with three levels: positive, comparative, and superlative, exemplified in (104)–(106), respectively:

(104) Sympathizing friends from his own and distant cities sent fresh, **fragrant flowers** he loved so well; and these were placed upon his breast and pillow. (*COHA*, 1855, MAG, NorthAmRev)

(105) Dr. Grey, it is more than either or all of the three you mention; for **Persian mythology**, like Persian wines and Persian roses, is richer, more subtle, **more fragrant**, more glowing than any other. (*COHA*, 1869, FIC, VashtiQuotUntil)

(106) "Say, Barker, what do you think of that new flower which Mrs. Marston has put into our garden?" "I think," said Barker, "she is the prettiest and **most fragrant bud** I have seen; a very rare specimen." (*COHA*, 1902, FIC, CaliforniaGirl)

Even though many adjectives can be morphologically marked for the comparative and superlative degrees through the suffixes *-er* and *-est*, respectively, all five near-synonymous adjectives examined in the present dissertation are always marked via analytical means, that is, with the adverbs *more* and *most*.[69]

In addition to the six semantic and two syntactic factors discussed, the data was also annotated for a language-internal lexical variable termed Collocate. This variable contains the specific semantically modified nouns of the adjectives in each instance and allows us to control for lexical effects. Previous studies on near-synonyms have shown the importance of including the specific collocates of nodes, as members of synonym sets often display their own idiosyncratic lexical behavior (e.g. Gries 2003; Kjellmer 2003; Liu 2010; 2013; Desagulier 2014; cf. Section 3.2.2 above). In the case of adjectives, noun collocates seem to be more informative than other POS. In particular, many studies on attributive adjectives consider their

[68] Even though predicative adjuncts occur much more frequently than the other two minor functions, this was not enough to warrant an independent level for this syntactic function

[69] In English, morphological marking occurs only in cases of comparison of superiority, whereas comparison of inferiority or equality is always analytically signaled through the adverbs *less* and *as*/*so*, among others, respectively (Huddleston & Pullum et al. 2002: 1123–1124).

R1 collocates, that is, the head nouns they modify (e.g. Church et al. 1991; Biber, Conrad & Reppen 1998: Section 2.6; Gries 2001). In the present analyses, modified nouns include head nouns modified by attributive adjectives, which often occur directly to their right (but cf. *wind* in (94) above), but also subjects and objects qualified by predicative complements, as *paper* in (95) and *air* in (97), and nominal elements modified by postpositive adjectives (cf. example (100)), among others. In the majority of cases, the task simply entailed retrieving the specific noun collocate from each example. However, in a few cases this was not possible, as in example (107), where the nominal element modified by *fragrant* is the pronoun *something*.

(107) Everyone in the house was sleeping when I crept downstairs. It was my day off. In the kitchen, Coxie was waiting. She handed me a napkin wrapped around **something warm and fragrant**. And a mug of tea. "Breakfast. You'd best be goin' afore that hellcat up there wakes up an' thinks of some errand fer you." (*COHA*, 1993, FIC, SplitHeirs)

Whenever a pronoun was modified by any of the adjectives at issue, its antecedent was retrieved from the context. By way of illustration, in example (108), the pronoun *it* is used as an anaphor for (*lavender-colored*) *letter*, and consequently *letter* is considered to be the noun collocate of *scented* in this particular instance.

(108) "Who was your **lavender-colored letter** from?" Eleanor said, yawning; "I forgot to ask you. **It** was awfully **scented**!" (*COHA*, 1992, FIC, VehementFlame)

In some cases, it was not possible to locate the precise referent of anaphoric or cataphoric pronouns. For instance, in (107), even when the extended context in *COHA* is examined, it is impossible to identify the exact referent of *something* in *something warm and fragrant*. We know that it is used to denote an edible item of some kind, which Coxie gives to the narrator as breakfast, but we cannot know for sure exactly what type of food it is. Consequently, the collocate of *fragrant* in example (107) has been taken to be the generic noun *food*. Similarly, in contexts such as instructional discourse segments (e.g. recipes), where ellipsis is very frequent, it is difficult to pinpoint the exact source of the smell, as the near-synonyms usually

qualify a mixture of ingredients of different types, for instance, spices, meat, and vegetables. This is the case in example (109), in which *fragrant* does not refer to one specific ingredient, but to the resulting mixture of all of them. Therefore, the generic noun *food* was again selected as the collocate of *fragrant* in (109).

(109) Meanwhile, in large skillet, heat oil over medium-high heat. Add **onion** and cook, stirring often, until softened, about 5 minutes. Stir in **garlic, curry powder, cumin** and **cloves** and cook, stirring, until **fragrant**, about I minute. (*COHA*, 2002, MAG, VegTimes)

Even though most studies on near-synonyms focus on language-internal variables such as those considered so far, language-external factors have also been shown to exert a strong influence on lexico-semantic variation and change (cf. Section 2.4.2). Therefore, variables of an extralinguistic nature were also included in the analyses of the present dissertation. These are explained in the next section.

4.3.3 Language-external variables

Since the main aim of the present dissertation is to analyze the diachronic development of the five near-synonymous adjectives *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*, a variable that must necessarily be included in the analyses is Period. As mentioned in Section 4.2, *COHA* covers the time span 1810–2009 in AmE and is divided into 20 decades. However, given the relatively low frequency of the adjectives examined, especially *sweet-scented* and *sweet-smelling* (cf. Table 4 above and Table 9 in Section 5.1), it would have been difficult to identify changes in their distributional behavior if decades or years had been considered. For this reason, the temporal span covered by the corpus was divided into four fifty-year periods, which are the values of the variable Period: 1810–1859 (period 1 or P1), 1860–1909 (P2), 1910–1959 (P3), and 1960–2009 (P4).

The second language-external factor is Text-type, which refers to the genre in *COHA* to which each of the examples belong. As seen in Section 4.2 above, *COHA* distinguishes four major text-types, to wit, fiction, non-fiction, magazines, and newspapers. In the analyses in this dissertation, however, magazines and newspapers were merged into one single category, here

labelled 'periodicals', due to the low number of instances of the adjectives at issue in newspapers and to the fact that the corpus does not contain texts from this genre until the 1860s. Therefore, Text-type is here a categorical variable with three levels: fiction, non-fiction, and periodicals. It is worth noting at this point that within fiction *COHA* includes a considerable amount of poetic texts which, as is well-known, are subject to esthetic considerations that do not often apply to other genres, such as a higher degree of figurative language. Therefore, the inclusion in the analysis of examples of the near-synonyms from such texts may lead to an overrepresentation of figurative uses in the sense 'pleasant or agreeable'. Despite this, the decision was made to maintain examples from poetry in the database, given that poetic discourse is also a manifestation of language, and removing this genre would result in an incomplete picture of the contextual patterns of the near-synonyms and of the language of the time span examined, particularly from the 1810s to the 1860s.

4.4 SUMMARY

The present chapter has dealt with the data used for the analyses in the present dissertation. First, the synonym set object of study was described (cf. Section 4.1). The main motivations for selecting the five adjectives *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling* were explained: on the one hand, that fact that the domain of SMELL has been relatively understudied in semantic research and, on the other, that the results of pilot studies pointed to interesting historical developments of this particular synonym set. A semantic description of the members of the set was provided on the basis of both PDE and historical reference works, which were deemed insufficient to reach an adequate understanding of the semantic relation holding between them.

Second, the corpus from which the data was extracted, namely *COHA*, was described (cf. Section 4.2). It was argued that *COHA* constitutes an optimal data source for the analyses in this dissertation, mainly due to its large size and its diachronic nature. Next, the data extraction process was clarified, focusing on the search strings used to retrieve the examples of the near-synonymous adjectives from *COHA*. Two databases were created, one containing all the occurrences of the adjectives, and a second containing their noun collocates in an L5–R5 context window.

Finally, Section 4.3 discussed the annotation of the first database, starting with the examples that had to be excluded from the count for different reasons, before moving on to the description of the variables considered in the analysis, both intralinguistic semantic and non-semantic ones, as well as extralinguistic factors. Table 8 summarizes the eleven variables included in the study, together with their values.

Table 8. Variables in the first dataset and their levels

Variable types	Variable	Variable levels
Language-internal semantic variables	Sense	Natural Artificial Figurative Indeterminate
	Semantic category	ABSTRACT BODY AND PEOPLE CLEANING COSMETICS EARTH, ATMOSPHERE, AND WEATHER FOOD AND DRINK OBJECT PLANTS AND FLOWERS SENSATION SPACE SUBSTANCE AND MATERIAL TEXTILE AND CLOTHING
	Animacy	Animate Inanimate
	Concreteness rating	Average rating of concreteness from 1 to 5
	Concreteness binary	Concrete Abstract
	Countability	Count Non-count Other
Language-internal non-semantic variables	Syntactic function	Attributive Predicative complement Postpositive Other
	Degree	Positive Comparative Superlative
	Collocate	Specific noun collocate (lemma)
Language-external variables	Period	Period 1 (1810–1859) Period 2 (1860–1909) Period 3 (1910–1959) Period 4 (1960–2009)
	Text-type	Fiction Non-fiction Periodicals

The remaining chapters of the present dissertation provide the results of three separate analyses. The findings in Chapters 5 and 6 are based on the first dataset (cf. Section 4.2), and thus focus on the effects of the variables in Table 8 on the choice between the five near-synonyms. Chapter 5 is of a more descriptive and univariate nature, including both semasiological and onomasiological perspectives of the adjectives, while Chapter 6 presents a confirmatory multivariate analysis in which the emphasis lies on the differences between the members of the synonym set over time. Chapter 7, in turn, offers a more detailed account of the idiosyncratic collocational patterns of *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*, by zooming in on their specific noun collocates, thus employing the second database described in Section 4.2 above.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5 SEMASIOLOGICAL AND ONOMASIOLOGICAL ANALYSES OF THE SYNONYM SET

Before the emergence of cognitive linguistics, different theoretical frameworks in lexical semantics tended to focus almost exclusively on either semasiological or onomasiological structures: whereas, for instance, the historical-philological school was mainly devoted to the former, structuralist scholars primarily paid attention to the latter (cf. Sections 2.1 and 2.2). As seen earlier on, semasiology refers to the study of particular words and the senses or concepts that they designate, thus being closely associated with polysemy. A typical semasiological study would therefore consider a polysemic word such as *water* in isolation and its different senses. Conversely, onomasiology takes a particular concept or category as a starting point and examines the various expressions which are used to designate it, hence being closely related to synonymy. A typical onomasiological study would therefore select a concept such as PLEASANT TO THE TASTE and analyze the different lexical items that can be used to refer to it, for instance, *delicious*, *luscious*, *delectable*, and *tasty*. As argued by Levshina (2011: 11–12), analyses which take into account both perspectives are rather scarce. However, both semasiological and onomasiological descriptions are necessary in order to reach a full understanding of the semantics of individual lexical items, the concepts they denote, and the relationship between neighboring categories. This is so because, depending on the perspective one adopts, prototypicality and salience effects may vary. In other words, what is semasiologically salient or prototypical is not necessarily onomasiologically so.

Salience is in itself a very complex concept, which refers to the degree of typicality of the members of a given category. This category can be an individual word, in which case the emphasis is on the typicality of its different senses, or a larger linguistic category, in which case the focus is on the typicality of the words used to denote it. In this line, semasiological salience, when applied to the phenomenon of polysemy, refers to how central a specific sense or reading of a concrete lexical item is. Central or salient senses are usually those from which other senses

derive and those that are more frequent (e.g. Levshina 2011: 13; Kay & Allan 2015: 41). This is the reason why metaphorically extended senses are often less salient, as they tend to derive from literal ones and are, at the same time, often less common. The noun *bird* serves to illustrate this point: besides its literal meaning in PDE '[a]ny feathered vertebrate animal […]' (*OED*, s.v. *bird* n. 2), *bird* can also be used metaphorically to denote '[a]n aeroplane' (*OED*, s.v. *bird* n. 4d). The latter reading is in fact a metaphorical extension of the former and is less salient, both logically and diachronically, as it originates from the 'biological species' sense, but also statistically, as it is much less frequent (Lewandowska-Tomaszcyk 2007: 147). Onomasiological salience, on the other hand, refers to the degree of entrenchment of a lexical item used to designate a particular semantic concept or category in comparison with other alternatives, be they synonyms or hyponyms/hypernyms. As mentioned in Section 2.4.1, onomasiological salience can be operationalized as "the ratio between (a) the frequency with which the members of a category are named with an item that is a unique name for that category, (b) and the total frequency with which the category occurs in a corpus" (Geeraerts 2010: 202; cf. also Geeraerts, Grondelaers & Bakema 1994). Salience is not a static but a dynamic phenomenon and, as such, can vary over time, since the experiences of language users often change due to, for instance, transformations in culture and society (cf. Section 2.4.1). This applies both to semasiological salience, and thus senses can become more or less prototypical over time, and to onomasiological salience, where the competing alternatives may differ as regards entrenchment from one period to another.

The previous chapter offered an overview of the semantics of the selected adjectival near-synonyms *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*, as described in existing reference works. This overview provided some preliminary clues as to the frequency and prototypicality of the five adjectives, as well as to their possible usage patterns in terms of the senses that they share (cf. Section 4.1.2). This information led us to formulate some research questions, some of which are explored in the present chapter. These are here repeated for the sake of convenience:

(i) Are *perfumed* and *scented* the only adjectives of the set that are used in the artificial sense, i.e. 'pleasant to smell because perfume has been added or used', or do *fragrant*, *sweet-scented*, and *sweet-smelling* also appear in such semantic contexts?

(ii) If *fragrant*, *sweet-scented*, and *sweet-smelling* are also used in the artificial sense, is there any difference in their distribution, both in this and in the other senses (i.e. natural and figurative) of the adjectives?

(iii) Does the distribution of the five adjectives across senses remain stable over time or does it fluctuate? In other words, are the apparent changes in frequency of the adjectives revealed in *CoD* specific to one or several contexts of use or do they, on the contrary, occur equally across all contexts?

Therefore, the goal of this chapter is to analyze the first database described in Section 4.2, focusing on the language-internal semantic variables Sense and Semantic category, the latter offering a more fine-grained semantic division of categories closely associated with a particular sense. This allows us to contrast the information provided by the reference material against actual corpus data. To this purpose, the absolute and relative frequencies of the five adjectives at issue in different periods are examined, first generally, and then in particular contexts of use, that is, across senses and semantic categories. Only minor confirmatory testing—to wit, chi-square tests of independence— is carried out in this chapter, as the main objective here is to provide a more descriptive analysis of the data. Chi-square tests are conducted to determine if there is a significant association between two categorical variables. The test compares the observed frequencies in a contingency table with the expected frequencies that would be found if the variables were independent. If the observed frequencies differ significantly from the expected frequencies, this means that the variables are correlated (cf., for instance, Cantos Gómez 2013: 75–80; Levshina 2015: 210–213).[70] Multivariate analyses of the data are, in turn, carried out in Chapter 6, which thus offers a more sophisticated quantitative approach.

Chapter 5 is structured as follows. Section 5.1 presents the overall frequency in *COHA* of the concept PLEASANT SMELLING and of the five adjectival near-synonyms, as well as their frequency over the time span 1810–2009. Then, Section 5.2 provides an account of the semantics of the concept PLEASANT SMELLING as a whole, that is, as designated by the five near-synonyms, both generally and diachronically throughout the four historical periods distinguished. A hypothesis accounting for the existent variation of the concept over time is postulated on the basis of the social and technological transformations experienced by American society after the First and Second Industrial Revolutions. Section 5.3 focuses on the

[70] All statistical procedures conducted throughout the present dissertation were computed in *R* (R Core Team 2017).

semasiological structure of each of the five adjectives individually in order to identify their prototypical uses. Section 5.4, in turn, deals with the onomasiological structure of the concept PLEASANT SMELLING, exploring the competition between the synonyms in different contexts of use. Finally, Section 5.5 offers a summary and some concluding remarks.

5.1 OVERALL FREQUENCY AND FREQUENCY OVER TIME

Out of the initial 6,338 hits, a total number of 5,764 tokens of the five adjectival near-synonyms remained in the database following the manual pruning, and were subsequently analyzed (cf. Section 4.3). The frequency distribution is provided in Table 9, which shows the absolute and relative frequencies of the concept PLEASANT SMELLING and of the individual adjectives in general and over time. The frequencies are shown by means of both percentages and normalized frequencies (NFs) per million words. The percentages display the relative frequencies of the adjectives with respect to one another, thus comparing the frequency of the five synonyms in each period. Therefore, the percentages have to be interpreted vertically in Table 9, that is, they total 100% in each period. The NFs, in turn, display the frequency of each adjective individually over time.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 9. Absolute and relative frequencies of the five near-synonyms

Period / Synonym	P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
Fragrant					
N	804	1291	712	588	3,395
%	66.83	64.87	52.88	48.32	58.90
NF	14.77	12.87	5.87	4.51	8.36
Perfumed					
N	193	333	227	224	977
%	16.04	16.73	16.77	18.41	16.95
NF	3.55	3.32	1.87	1.72	2.41
Scented					
N	109	235	296	301	941
%	9.06	11.81	21.86	24.73	16.33
NF	2,00	2.34	2.44	2.31	2.32
Sweet-scented					
N	78	78	29	13	198
%	6.48	3.92	2.14	1.07	3.44
NF	1.43	0.78	0.24	0.10	0.49
Sweet-smelling					
N	19	53	90	91	253
%	1.58	2.66	6.65	7.48	4.39
NF	0.35	0.53	0.74	0.70	0.62
Total					
N	1,203	1,990	1,354	1,217	5,764
%	100	100	100	100	100
NF	22.10	19.83	11.17	9.34	14.19

Starting with the overall frequency, the findings evince major differences among the synonyms, with *fragrant* being considerably more frequent than the other four adjectives, followed by *perfumed* and *scented*, which display a similar frequency, and finally *sweet-smelling* and *sweet-scented*, which together account for less than 8% of the data. In fact, out of the total 5,764 occurrences, more than half, namely 3,395 (58.90%), correspond to instances of *fragrant*, while 977 (16.95%), 941 (16.33%), 198 (3.44%), and 253 (4.39%) are cases of *perfumed*, *scented, sweet-scented*, and *sweet-smelling*, respectively. This considerable variation in frequency among synonyms is not uncommon, since synonym sets often contain one or more dominant expressions (cf. Section 2.4.1).

In addition, the frequency distribution of the five adjectives varies greatly over time, as shown in Figure 3.

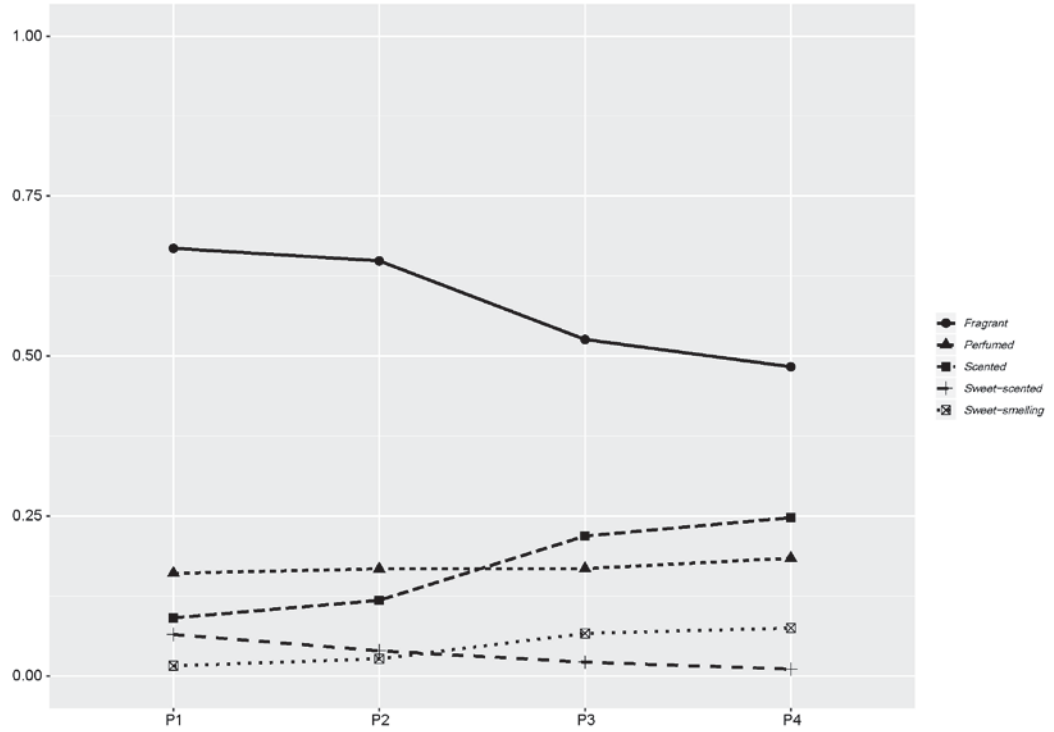


Figure 3. Proportions of the five near-synonyms across periods

While *fragrant* clearly decreases diachronically, from 66.83% in P1 to 48.32% in P4, *scented* increases substantially, from 9.06% of the total occurrences in P1 to 24.73% in P4. Similarly, *sweet-scented* and *sweet-smelling*, despite their low overall frequency, also fluctuate considerably over time. The former displays an important downward tendency, from 6.48% in P1 to 1.07% in P4, and the latter an important upward tendency, from 1.58% in P1 to 7.48% in P4. Conversely, *perfumed* remains rather stable diachronically, with only a minor increase of approximately 2.5 percentual points. The distribution of the five near-synonyms across the four periods is statistically significant, as proved by a chi-square test of independence ($\chi^2 = 338.94$, df = 12 $p < 0.001$). By examining the residuals calculated by the test, it is possible to identify which specific changes contribute the most to the obtained chi-square value and its significance. The residuals constitute the differences between the observed and expected frequencies, with positive and negative values indicating a higher and a lower observed frequency than expected, respectively (Levshina 2015: 218; Sharpe 2015: 2; Desagulier 2017: 183–185). Residuals are to be interpreted as follows: the degree of contribution of a particular cell in the contingency table to the obtained chi-square value depends on the absolute size of its residual, so that the larger the residual, the higher the contribution. As a rule of thumb, residual values that are larger than +2 or lower than –2 are considered to contribute substantially to the distribution (Levshina

2015: 220–221; Sharpe 2015: 3; Desagulier 2017: 185). Here, the residuals demonstrate that both *fragrant* and *sweet-scented* are considerably more frequent than expected in P1, and *fragrant* also in P2, but less frequent than expected in P3 and P4, while the opposite tendency is true for *scented* and *sweet-smelling.* The distribution of *perfumed*, in turn, does not seem to change much over time. These chi-square residuals are provided in Table A.1 in Appendix A.

If the NFs of the individual adjectives per million words are considered, however, it is clear from Table 9 that three of the five items decrease over time, namely *fragrant*, *perfumed*, and *sweet-scented*, while the remaining two, *scented* and *sweet-smelling* increase. In the case of *fragrant* and *sweet-scented*, the downward tendencies are particularly sharp, from 14.77 and 1.43 in P1 to 4.51 and 0.10 in P4, respectively, while the decrease of *perfumed* is more moderate. The upward tendencies of *scented* and *sweet-smelling*, in turn, are only minor, and both adjectives decline somewhat from P3 to P4. The frequency fluctuations over time evince a competition between the five synonyms, with some of them converging in frequency as time passes. In fact, in a few cases one adjective comes to overthrow another. For instance, *scented*, which in P1 is less frequent than *perfumed*, becomes more frequent over time. Similarly, *sweet-smelling* starts out as being approximately four times less common than *sweet-scented*, but in P4 it is seven times more common. The increase of *scented* and the decrease of *fragrant* also lead to a gradual convergence in the frequency of the two adjectives: while *fragrant* is seven times more frequent than *scented* in P1, in P4 it is only twice as frequent. This convergence between the two synonyms is particularly pronounced from P2 to P3, that is, in the transition from the 19th to the 20th century.

The frequency clines in Table 10 summarize the data in Table 9 and Figure 3, where '>' stands for 'more frequent than':

Table 10. Frequency clines of the five adjectives in the period 1810–2009

P1 (1810–1859):	*fragrant* > *perfumed* > *scented* > *sweet-scented* > *sweet-smelling*
P2 (1860–1909):	*fragrant* > *perfumed* > *scented* > *sweet-scented* > *sweet-smelling*
P3 (1910–1959):	*fragrant* > *scented* > *perfumed* > *sweet-smelling* > *sweet-scented*
P4 (1960–2009):	*fragrant* > *scented* > *perfumed* > *sweet-smelling* > *sweet-scented*

Therefore, although *fragrant* maintains its dominance over time, the other four adjectives change positions in the frequency cline throughout the period 1810–2009. In particular, *scented*

displaces *perfumed*, and *sweet-smelling* displaces *sweet-scented*. These changes take place especially between P2 and P3.

The findings obtained regarding the frequency of the five near-synonyms are mostly in line with the information provided in the *OED* and *CoD* discussed in Section 4.1.2. First, *fragrant* is the most common adjective in PDE, followed by *scented*, *perfumed*, and *sweet-smelling*, and lastly *sweet-scented*, thus coinciding with the frequency ranges provided in the *OED*. Additionally, the specific ratings offered in the *OED* for *fragrant*, *sweet-scented*, and *sweet-smelling* coincide with the results obtained in *COHA*. *Fragrant* occurs between 1 and 10 times per million words (i.e. 4.51 in P4), *sweet-smelling* between 0.1 and 1 (i.e. 0.70 in P4), and *sweet-scented* between 0.01 and 0.1 (i.e. 0.10 in P4). Nonetheless, the frequencies of both *perfumed* and *scented* are somewhat higher in *COHA* than those suggested in the *OED*, ranging between 1 and 10 times per million words (i.e. 1.72 and 2.31 in P4, respectively), rather than between 0.1 and 1. Second, the diachronic tendencies in *CoD*, which are based on *Google Ngrams*, correspond with the frequency fluctuations obtained here: *fragrant*, *perfumed*, and *sweet-scented* all decrease according to their NFs, while *scented* increases somewhat.[71] However, the NFs provided in *CoD* for the different periods do not correspond with those obtained here. This is to be expected given that *Google Ngrams* is only based on data from books, while *COHA* is a more heterogenous corpus, including different text-types.

Finally, considering the frequency of the concept PLEASANT SMELLING as a whole, which is displayed in the last row in Table 9, a general downward tendency is observed, from 22.10 instances per million words in P1 to just 9.34 in P4. The decrease of the concept seems to have taken place particularly between P2 and P3, where the decline is most pronounced, falling from almost 20 occurrences per million words in P2 to approximately 11 in P3, thus coinciding with the periods in which most changes of the individual adjectives occur. This global decrease seems to suggest that the concept PLEASANT SMELLING is not as frequent in PDE as in the latter part of the Late Modern English (LModE) period, although in order to confirm this claim it is necessary to investigate a wider range of items belonging to this semantic domain. Interestingly, despite the decline of the concept PLEASANT SMELLING, as we have seen in this section, some adjectives (i.e. *scented* and *sweet-smelling*) do not follow a downward trend, and even increase somewhat in frequency in the course of the 20th century, which calls for further explanation.

[71] Nothing can be said about *sweet-smelling* in this respect, since *CoD* does not include information regarding the frequency of this adjective over time (cf. Section 4.1.2).

In sum, this section has demonstrated that the concept PLEASANT SMELLING and the five adjectives *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*, which are used to designate it, experience changes over time. However, it is not clear whether these fluctuations in frequency occur across the board, that is, in all contexts of use, or whether different patterns emerge depending on the senses in which the adjectives are used or on the semantic categories of the nouns which they modify. Moreover, it is not clear either why the concept as a whole experiences such a dramatic decrease in frequency in the course of the 20th century (i.e. P3 and P4). It might well be the case that the downward tendency observed is a result of smells being less common and prominent in present-day American culture and society than before, either generally or only some specific types of aromas. In addition, the decrease might also be a direct consequence of the fluctuations of specific adjectives, either in terms of their overall frequency or due to changes in their prototypical structure. The remainder of this chapter delves into these issues and tries to uncover the underlying motivations for the changes identified in the present section, starting with the concept PLEASANT SMELLING as a whole.

5.2 THE CONCEPT pleasant smelling

This section considers the diachronic development of the concept PLEASANT SMELLING across senses and semantic categories. Given that Sense represents a rather coarse-grained distinction, while Semantic category represents a more fine-grained classification of the data, the former is first discussed in Section 5.2.1. Then, to further localize the possible changes of the concept across senses, the different semantic categories are analyzed in Section 5.2.2. Chi-square tests are again conducted, in this case so as to determine if there is a significant association between the variables Sense and Period, on the one hand, and between Semantic category and Period, on the other. As the focus here lies on the concept in its entirety, the five near-synonyms are considered together at this point. Finally, Section 5.2.3 explores the possible motivations for the changes uncovered in Sections 5.2.1 and 5.2.2 on the basis of external factors pertaining to the modernization of American society.

5.2.1 Sense

Figure 4 displays the distribution of the concept PLEASANT SMELLING across senses and periods: each bar represents one period, and within each period the senses are signaled by different shades of grey. Table 11 provides the specific figures per period, as well as the overall

frequency of the concept in each sense. Again, the percentages are to be interpreted vertically. Overall, the concept is much more commonly used to denote natural smells (Na), with a general frequency of 3,844 instances, equaling 66.69% of the total data. The other three senses (i.e. artificial, figurative, and indeterminate) are much less common. The concept is used in the artificial sense (Art) in a total of 1,211 instances, representing 21.01% of the cases. The figurative (Fig) and indeterminate (Indet) senses together account for less than 15% of the data, with the former amounting to 235 instances (4.08%) and the latter to 474 instances (8.22%).

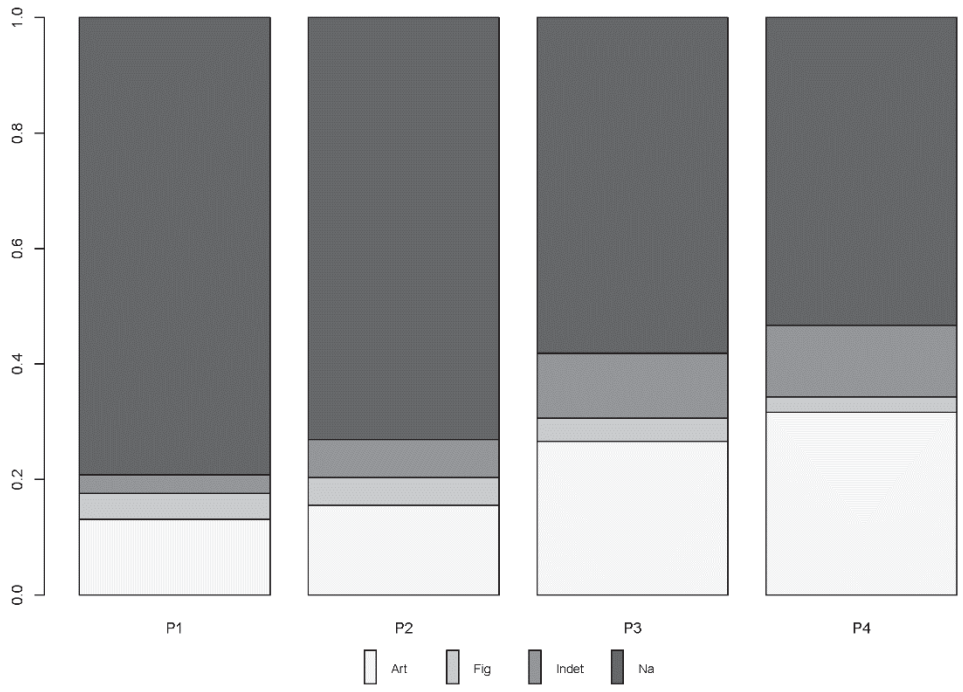


Figure 4. Proportions of the concept PLEASANT SMELLING in each sense across periods

Table 11. Distribution of the concept PLEASANT SMELLING across senses

Period / Sense	P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
Artificial					
N	157	309	360	385	1,211
%	13.05	15.53	26.59	31.64	21.01
NF	2.88	3.08	2.97	2.96	2.98
Figurative					
N	54	95	54	32	235
%	4.49	4.77	3.99	2.63	4.08
NF	0.99	0.95	0.45	0.25	0.58
Indeterminate					
N	39	131	153	151	474
%	3.24	6.58	11.30	12.41	8.22
NF	0.72	1.31	1.26	1.16	1.17
Natural					
N	953	1455	787	649	3,844
%	79.22	73.12	58.12	53.33	66.69
NF	17.51	14.51	6.49	4.98	9.46
Total					
N	1,203	1,990	1,354	1,217	5,764
%	100	100	100	100	100
NF	22.10	19.83	11.17	9.34	14.19

As mentioned in Section 4.3.1.1 above, the value 'indeterminate' groups together those instances which could not be safely assigned to either the natural or artificial senses. An instance is provided in (110):

(110) Off came her earrings, then her clothes, without embarrassment. Breasts or none, her woman's wealth satisfied him. When he touched her nipples the effect was electric. She could be a little wild in bed. "Your fingers are fires." He enjoyed her long body, the cool flesh, sex smell, **fragrant hair**. He enjoyed the possession and adventure, her intensity with detachment. (*COHA*, 1861, FIC, ChairmanBored)

Similarly to example (48) in Section 4.3.1.1, *fragrant* can be interpreted in (110) as referring either to a natural or to an artificial smell. This is so because we cannot be sure whether the man finds the hair of the woman naturally fragrant or whether, on the contrary, its aroma is due to a previous impregnation of the hair by some cleaning or personal care product. In occurrences such as these, it may also be the case that the agreeable fragrance stems from a mixture of natural and artificial pleasantly smelling substances. Note that indeterminate cases such as the

one illustrated in (105) are more frequent in the data than occurrences illustrating the figurative sense.

If the distribution of the concept across senses is analyzed over time, considerable changes are identified. This is so because, even though the natural sense remains the most common one throughout the time span 1810–2009, this sense displays a major downward tendency as time passes, from about four fifths of the data in P1 (79.22%) to only about half in P4 (53.33%). The figurative sense, despite its overall low frequency, also decreases in use from P1 to P4, from 4.49% to 2.63%. The artificial sense and indeterminate examples, in turn, show the opposite trend, becoming more and more frequent with each subsequent period. Concerning the artificial sense, it totals only 13.05% of the data in P1, while it rises up to 31.64% in P4. In the case of the value 'indeterminate', we see an upward tendency from 3.24% in P1 to 12.41% in P4. These changes lead to a convergence of the senses in terms of frequency, especially in the case of the natural and artificial senses. In other words, whereas in P1, the concept PLEASANT SMELLING is mainly used to refer to natural aromas, in P4 this predominance is not as strong. In fact, whereas in P1 the natural sense is about six times more common than the artificial one, in P4 it is only around two times more common. Crucially, both the decrease of the natural sense and the increase of both the artificial sense and indeterminate uses are particularly pronounced from P2 to P3, that is, around the transition from the 19th to the 20th century. Interestingly, it is precisely at this point in time that most fluctuations in the frequency of individual adjectives, as well as of the concept as a whole (cf. Section 5.1), take place, thus suggesting the existence of some underlying motivation(s) for change.

A chi-square test of independence confirms that the distribution of the concept PLEASANT SMELLING across senses and periods shown in Table 11 and Figure 4 is statistically significant ($\chi^2 = 331.76$, df = 9, $p < 0.001$). Again, by considering the residuals of the test, the specific changes that contribute the most to the chi-square value and its significance can be identified. In this case, the residuals indicate that the natural sense is considerably more frequent than expected in both P1 and P2, but less frequent than expected in P3 and P4. The opposite trend emerges in the case of the artificial sense and of indeterminate uses, both of which are substantially less frequent than expected in P1 and P2, but more so in P3 and P4. Some of the changes undergone by the figurative sense are also quite notable, as this sense is considerably less frequent than expected in P4.[72]

[72] The residuals of this statistical analysis are provided in Table A.2 in Appendix A.

If we consider the NFs per million words, similar trends emerge, with some differences. The natural and figurative senses still decrease substantially over time, particularly from P2 to P3, from 14.51 and 0.95 to 6.49 and 0.45, respectively. In fact, these decreases are even more pronounced than those resulting from the examination of the percentages of use. On the other hand, even though the artificial sense and indeterminate uses still experience a growth across the period 1810–2009, these increases are more moderate if the NFs are taken into consideration, from 2.88 and 0.72 in P1 to 2.96 and 1.16 in P4, respectively. However, since the concept as a whole clearly declines over time, from 22.10 in P1 to 9.34 in P4, that two of the senses still increase in frequency is particularly noteworthy and further points to some reorganization over time of the sense structure of the concept PLEASANT SMELLING.

The findings presented in this section have uncovered rather general diachronic tendencies regarding the distribution of the concept PLEASANT SMELLING as a whole across senses. In the next section we turn to the distribution of the concept across semantic categories in order to identify whether these changes occur with specific types of nouns or whether, on the contrary, they are equally distributed across the natural, artificial, figurative, and indeterminate uses.

5.2.2 Semantic category

As mentioned in Section 4.3.1.2, there is a one-to-one correspondence between some of the four senses discussed in the previous section and the twelve semantic categories. First, instances in which the concept PLEASANT SMELLING co-occurs with nouns in the semantic categories EARTH, ATMOSPHERE, AND WEATHER (e.g. *breeze*, *hill*, and *rain*), FOOD AND DRINK (e.g. *apple*, *chicken*, and *cup*), and PLANTS AND FLOWERS (e.g. *bud*, *rose*, and *tree*) are invariably used in the natural sense.[73] Second, all occurrences with nouns from the category CLEANING (e.g. *bath*, *disinfectant*, and *soap*), COSMETICS (e.g. *cologne*, *gloss*, and *oil*), and TEXTILE AND CLOTHING (e.g. *dress*, *glove*, and *linen*) refer to smells exhibited by synthetic substances, thus corresponding with the artificial sense. Lastly, with noun collocates belonging to the category ABSTRACT (e.g. *charm*, *knowledge*, and *memory*), the concept is always used in the figurative sense ‘sweet or pleasant’. Therefore, in seven of the twelve semantic categories distinguished here, we witness a complete match between the two variables. On the contrary, in none of the five remaining categories, namely BODY AND PEOPLE (e.g. *hair*, *woman*, and *wrist*), OBJECT (e.g. *candle*, *couch*, and *letter*), SENSATION (e.g. *scent*, *taste*, and *tone*), SPACE (e.g. *chamber*, *city*,

[73] For more examples of nouns belonging to each semantic category, see Table 7 in Section 4.3.1.2.

and *house*), and SUBSTANCE AND MATERIAL (e.g. *liquid*, *smoke*, and *wood*), is there an exact parallelism between the variables, although some clear tendencies can be observed. The distribution of these five categories with regard to the four senses is shown in Table 12.

Table 12. Sense distribution of the semantic categories BODY AND PEOPLE, OBJECT, SENSATION, SPACE, and SUBSTANCE AND MATERIAL

Semantic category / Sense	B&P	OBJ	SEN	SPA	S&M
Artificial					
N	204	179	17	49	232
%	38.93	69.92	6.51	14.41	61.38
Figurative					
N	5	7	15	5	–
%	0.95	2.73	5.75	1.47	
Indeterminate					
N	270	29	21	102	39
%	51.53	11.33	8.05	30.00	10.32
Natural					
N	45	41	208	184	107
%	8.59	16.02	79.69	54.12	28.31
Total					
N	524	256	261	340	378
%	100	100	100	100	100

As seen here, SENSATION and SPACE are overall 'natural' categories, as more than half of the instances occur in this sense, 79.69% in the former case and 54.12% in the latter. On the contrary, OBJECT and SUBSTANCE AND MATERIAL are mainly 'artificial,' with 69.92% and 61.38% of cases, respectively, corresponding with this sense. Finally, BODY AND PEOPLE is a more heterogenous category, exhibiting a great deal of ambiguity (51.53% of indeterminate instances). Similarly, SPACE, although chiefly a 'natural' category, also displays a high percentage of indeterminate cases (30%).

Figure 5 shows the distribution of the concept PLEASANT SMELLING across semantic categories per period: each bar represents a separate period, and within each of them the semantic categories are signaled by different colors. Table 13 provides the specific figures for each period and the overall frequency of the concept with respect to the twelve semantic categories, both in absolute numbers and in percentages. As in previous cases, the percentages should be interpreted vertically.

Starting with the overall frequencies, which are shown in the rightmost column, it is evident that some semantic categories of nouns are modified by the concept under analysis much more

frequently than others. For instance, PLANTS AND FLOWERS and EARTH, ATMOSPHERE, AND WEATHER represent 28.57% and 19.52%, respectively, while nouns belonging to COSMETICS and CLEANING correspond to just 2.31% and 3.09%. Overall, the prototypically 'natural' categories are much more common than the prototypically 'artificial' ones.

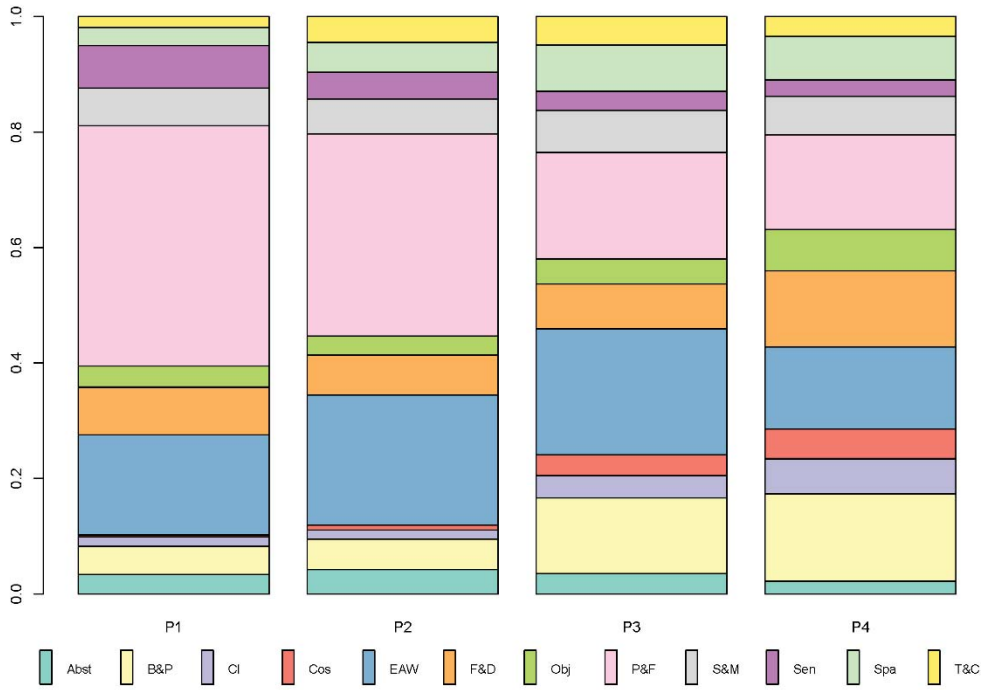


Figure 5. Proportions of the concept PLEASANT SMELLING in each semantic category across periods

Table 13. Distribution of the concept PLEASANT SMELLING across semantic categories

Period / Sem. category		P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
ABS						
	N	40	84	48	27	199
	%	3.33	4.22	3.55	2.22	3.45
	NF	0.73	0.84	0.40	0.21	0.49
B&P						
	N	59	104	177	184	524
	%	4.90	5.23	13.07	15.12	9.09
	NF	1.08	1.04	1.46	1.41	1.29
CL						
	N	20	32	52	74	178
	%	1.66	1.61	3.84	6.08	3.09
	NF	0.37	0.32	0.43	0.57	0.44
COS						
	N	4	17	49	63	133
	%	0.33	0.85	3.62	5.18	2.31
	NF	0.07	0.17	0.40	0.48	0.33
EAW						
	N	208	448	296	173	1,125
	%	17.29	22.51	21.86	14.22	19.52
	NF	3.82	4.47	2.44	1.33	2.77
F&D						
	N	100	138	105	160	503
	%	8.31	6.93	7.75	13.15	8.73
	NF	1.84	1.14	0.87	1.23	1.24
OBJ						
	N	44	66	59	87	256
	%	3.66	3.32	4.36	7.15	4.44
	NF	0.81	0.66	0.49	0.67	0.63
P&F						
	N	501	697	249	200	1,647
	%	41.65	35.03	18.39	16.43	28.57
	NF	9.21	6.95	2.05	1.54	4.05
S&M						
	N	78	120	99	81	378
	%	6.48	6.03	7.31	6.66	6.56
	NF	1.43	1.20	0.82	0.62	0.93
SEN						
	N	89	92	45	35	261
	%	7.40	4.62	3.32	2.88	4.53
	NF	1.64	0.92	0.37	0.27	0.64
SPA						
	N	38	103	108	91	340
	%	3.16	5.18	7.98	7.48	5.90
	NF	0.70	1.03	0.89	0.70	0.84
T&C						
	N	22	89	67	42	220
	%	1.83	4.47	4.95	3.45	3.82
	NF	0.40	0.89	0.55	0.32	0.54
Total						
	N	1,203	1,990	1,354	1,217	5,764
	%	100	100	100	100	100
	NF	22.10	19.83	11.17	9.34	14.19

If the diachronic evolution of the concept is considered, some clear trends can be observed. On the one hand, the frequency of the concept decreases in four categories, namely ABSTRACT, EARTH, ATMOSPHERE, AND WEATHER, PLANTS AND FLOWERS, and SENSATION. The downward tendency seems to be particularly pronounced in the latter two categories, from 41.65% and 7.40% in P1 to 16.43% and 2.88% in P4, respectively. Interestingly, both PLANTS AND FLOWERS and SENSATION are mainly 'natural' categories: PLANTS AND FLOWERS corresponds completely with the natural sense, while in approximately 80% of the instances of SENSATION the concept is used to denote natural aromas. Therefore, the considerable decrease of the natural sense uncovered in the previous section is mainly located in these two semantic categories. On the other hand, seven semantic categories exhibit an upward tendency over time; these are BODY AND PEOPLE, CLEANING, COSMETICS, FOOD AND DRINK, OBJECT, SPACE, and TEXTILE AND CLOTHING. Note that three of these, to wit CLEANING, COSMETICS, and TEXTILE AND CLOTHING, all display a one-to-one correspondence with the artificial sense. Additionally, OBJECT is also a prototypically 'artificial' category, given that almost 70% of the instances of the concept with nouns belonging to this class are used in this sense. In other words, the substantial increase of the artificial sense discussed earlier (cf. Section 5.2.1) seems to occur in most prototypically 'artificial' semantic categories. The only exception to this trend is SUBSTANCE AND MATERIAL (61.38% in the artificial sense), which remains rather stable over the period 1810–2009. In turn, BODY AND PEOPLE, the semantic category which displays the highest degree of ambiguity, also increases considerably, from 4.90% in P1 to 15.12% P4. Although this semantic category contains a large amount of indeterminate cases, the percentage of uses in the artificial sense (i.e. 38.93%) is still substantially greater than that in the natural one (8.59%). In the case of BODY AND PEOPLE, the upward tendency may be related to the growing use of lotions, perfumes, cleansing agents, and other artificial products that are often applied to the body. This would also go a long way towards explaining the increase in both the CLEANING and COSMETICS categories.

These results are mostly in line with the distribution of the concept PLEASANT SMELLING across senses discussed in Section 5.2.1. However, despite the decrease over time in the frequency of the concept in the natural sense, two semantic categories that are placed close to the natural end of the continuum do in fact increase. First, FOOD AND DRINK, a category which comprises only natural uses of the concept, rises from 8.31% in P1 to 13.15% in P4. Second, SPACE, a category used in almost 55% of instances in the natural sense, also increases, from

3.16% in P1 to 7.48% in P4. These findings, however, do not necessarily contradict the ones obtained in the previous section, namely that the concept decreases in the natural sense and rises in the artificial sense and in indeterminate uses. This is so because, although FOOD AND DRINK is prototypically 'natural', it could be argued that food and drinks have progressively become less 'natural' over time with the growing addition of food coloring, sweeteners, and preservatives. In the case of SPACE, the increase is attested mainly in the indeterminate uses (cf. Table A.3 in Appendix A), but not in the natural sense.

Many of diachronic tendencies discussed in the previous paragraphs are statistically significant according to a chi-square test of independence ($\chi^2 = 686.41$, df = 33, $p < 0.001$). The residuals of the test confirm the following important tendencies over the period 1810–2009 (cf. Table A.4 in Appendix A):[74]

(i) Substantial downward tendency over time: ABSTRACT; PLANTS AND FLOWERS; EARTH, ATMOSPHERE, AND WEATHER; and SENSATION.

(ii) Substantial upward tendency over time: BODY AND PEOPLE, CLEANING, COSMETICS, FOOD AND DRINK, OBJECT, SPACE, and TEXTILE AND CLOTHING.

As in the case of the distribution across senses shown in Table 11 and Figure 4 (cf. Section 5.2.1), the respective decreases and increases in frequency of the concept across semantic categories are particularly pronounced from P2 to P3, i.e. in the transition from the 19th to the 20th century, especially in the following categories: BODY AND PEOPLE, CLEANING, COSMETICS, PLANTS AND FLOWERS, SENSATION, and SPACE.

Again, the NFs per million words suggest the existence of similar diachronic tendencies, but here the differences are more pronounced than in Section 5.2.1. The four semantic categories for which percentages show a decrease in frequency, to wit, ABSTRACT; EARTH, ATMOSPHERE, AND WEATHER; PLANTS AND FLOWERS; and SENSATION, also decline and even more noticeably, particularly between P2 and P3. Contrariwise, only four of the seven semantic categories that increased according to the percentages also do so if NFs are considered; these are BODY AND PEOPLE, CLEANING, COSMETICS, and TEXTILE AND CLOTHING.[75] Interestingly, the

[74] Some of these semantic categories do not exhibit statistically significant residuals in all four periods, but they do show at least one significant residual, which points to a potentially significant upward or downward tendency.

[75] The category TEXTILE AND CLOTHING decreases only from P3 to P4.

categories that increase according to both measures are prototypically 'artificial', except BODY AND PEOPLE which, despite containing a high percentage of indeterminate uses, is still placed towards the artificial end of the semantic continuum (cf. Table 12). On the other hand, both FOOD AND DRINK and OBJECT decrease in terms of their NFs, while SPACE remains relatively stable. Lastly, SUBSTANCE AND MATERIAL now undergoes a substantial decline, in contrast to the stability it showed in terms of percentages.

In conclusion, the distributional changes experienced by the concept PLEASANT SMELLING throughout the time span 1810–2009 uncovered in the previous section and in the preceding paragraphs indicate that the concept undergoes a change from being used mainly to denote natural smells to becoming more and more frequent to refer to artificial ones. In other words, we witness a change as regards the kinds of nouns typically modified by the concept. Between 1810 and 1909 (i.e. in P1 and P2), the concept chiefly appears with nouns referring to entities with a natural pleasant smell, especially those belonging to the categories PLANTS AND FLOWERS and EARTH, ATMOSPHERE, AND WEATHER. However, from 1910 onwards (i.e. in P3 and P4), the concept increasingly collocates with nouns referring to entities with an agreeable aroma by virtue of having been impregnated by a pleasantly smelling synthetic substance. In the next section, this change is hypothesized to have been caused principally by language-external motivations, to wit, the social and technological transformations experienced by American society after the First and Second Industrial Revolutions. This hypothesis is further tested by examining the diachronic distribution of a series of lexical indicators from three semantic domains: INDUSTRY AND TECHNOLOGY, ADVERTISING AND MEDIA, and COSMETICS AND HYGIENE.

5.2.3 Ongoing semantic change in a modernizing society

During the period examined in the present dissertation (1810–2009), important social and technological changes took place, both around the globe and, in particular, in the United States of America. From a primarily pre-industrial economy in the 19th century, American society became throughout the 20th century the world's major industrial and commercial power (Jones 1996: Chapter 16; Baker 2017: 193–194). In such a context, the social and technological changes experienced as a result of the First and Second Industrial Revolutions may well constitute the underlying motivations accounting for the rise in the use of the concept PLEASANT SMELLING in the artificial sense at the expense of the natural one. First, there has been an

important upsurge in the production of goods with artificial aromas, such as cosmetics and cleansing agents, which most certainly also led these products to be more readily available to the population. In the second half of the 19th century, advances in industrial manufacturing and, particularly, chemistry enabled the production of cosmetics so that by the 1920s the mass production of these commodities became much cheaper, which, together with the fall of traditional Victorian standards of beauty and marketing campaigns, allowed more and more people to access cosmetics and wear them in public (Jones 1996: 308–309; Johnson 1997: 393, 476). Additionally, the growth of the advertising industry and its increasing influence on people's choices probably also aided in making these kinds of goods more important and visible in people's daily lives. Throughout the time span 1880–1914, the number and importance of periodicals, both newspapers and magazines, increased considerably in the US and, as a consequence, the percentage of the American population having access to this type of publications –and the advertisements they featured– increased as well (Johnson 1997: 397): in 1850, there were merely around 260 newspapers in circulation in the country, while only 60 years later, by 1910, approximately 2,600 different titles (i.e. ten times more) were published on a daily basis. The diachronic development of the concept PLEASANT SMELLING throughout the period 1810–2009 could hence constitute a reflection of language-external factors, to wit, socio-economic changes such as industrialization and mass production, that were at play during the same period and led to an ever-increasing need to allude to *artificially* scented lotions and candles rather than to *naturally* fragrant plants and flowers.

The effect of the aforementioned socio-economic changes on the usage patterns of the concept PLEASANT SMELLING in AmE can be tested by examining the distribution over time of a series of lexical indicators related to processes of modernization. This idea is in line with Baker's (2017: 177) claim that the frequencies of words in particular semantic domains reflect what "people actually write about at the conceptual level", thus uncovering aspects connected to "matters of culture and national identity" (cf. also Leech & Fallon 1992; Oakes 2003; Potts & Baker 2012). By drawing on data from corpora belonging to the Brown family, Leech & Fallon (1992), Oakes (2003), Potts & Baker (2012), and Baker (2017: Chapter 7) identify cultural and social differences between the US and Great Britain from the 1930s to the 2000s. The frequencies of words belonging to various semantic domains are examined in the British and American components of the Brown family of corpora in order to determine social, political, and cultural aspects which are particularly salient in each society in different

subperiods. These studies focus on two general domains which are particularly relevant for our purposes here, namely MASS MEDIA (e.g. *advertising*, *article*, *bulletin*, *editor*, *journal*, and *press*) and SCIENCE AND TECHNOLOGY (e.g. *electronics*, *machine*, *radiation*, *scientific*, and *technical*). In both cases, all four studies find a considerable increase in frequency over time in both varieties. However, the overall frequency of use of words in these two domains is higher in AmE than in BrE. These findings seem to be in line with the hypothesis postulated above, although the great majority of changes in the concept PLEASANT SMELLING took place at the turn of the 19th to the 20th century, that is, before the time span examined in the studies by Leech & Fallon (1992), Oakes (2003), Potts & Baker (2012), and Baker (2017). Be that as it may, these analyses do demonstrate that corpus data can be used as an indicator, if not as conclusive evidence, of cultural and social changes.

In order to test the hypothesis that changes in the concept PLEASANT SMELLING are influenced by extralinguistic factors, in what follows the evolution of lexical items belonging to three semantic domains is analyzed. Two of these correspond closely with some of the domains studied by Leech & Fallon (1992), Oakes (2003), Potts & Baker (2012), and Baker (2017), namely INDUSTRY AND TECHNOLOGY and ADVERTISING AND MEDIA. In addition, a field which was not included in these four studies is also examined, that is, COSMETICS AND HYGIENE, as it is especially relevant for the purposes of the present dissertation. Following the aforementioned studies, if an increase in the frequency of words in the semantic domains of INDUSTRY AND TECHNOLOGY, ADVERTISING AND MEDIA, and COSMETICS AND HYGIENE takes place at the same time as the rise in the artificial uses of the concept PLEASANT SMELLING, this would suggest that the transformation of American society from a pre-industrial stage to a fully industrialized nation is indeed one of the underlying causes for the changes discovered in Sections 5.1 and 5.2. To this purpose, searches were conducted in *COHA* of words pertaining to the three semantic domains and their NFs per million words were examined over time. In order to get a more fine-grained picture of the evolution of these words, the results were visualized across the 20 decades in *COHA* instead of across the four fifty-year periods distinguished previously. Twenty words (lemmas) were selected from each of the three domains:[76]

[76] Potts & Baker (2012) and Baker (2017: Chapter 7) analyze the changes in frequency of entire semantic domains across the two varieties of English and different subperiods from the 1930s to the 2000s. They do so by first semantically tagging all words in the corpora they employ, a process which was done automatically using the software *Wmatrix* (Rayson 2008) and then comparing the frequencies of specific semantic domains across the different corpora. Given that the study of lexical indicators and the socio-cultural changes they may reflect is not the main purpose of the present dissertation, the analysis is

(i) INDUSTRY AND TECHNOLOGY: *chemical*, *chemistry*, *coal*, *electric*, *electrical*, *electricity*, *engine*, *engineer*, *engineering*, *factory*, *fuel*, *gasoline*, *industrial*, *industry*, *manufacture*, *mechanic*, *mechanical*, *science*, *scientific*, and *workshop*.

(ii) ADVERTISING AND MEDIA: *advertisement*, *advertiser*, *advertising*, *audience*, *billboard*, *brand*, *broadcast*, *broadcaster*, *catalog(ue)*, *commercial*, *journalism*, *journalist*, *magazine*, *marketing*, *media*, *newspaper*, *press*, *publicity*, *radio*, and *reporter*.

(iii) COSMETICS AND HYGIENE: *barber*, *broom*, *clean*, *comb*, *cosmetic*, *detergent*, *haircut*, *hygiene*, *laundry*, *lipstick*, *lotion*, *make-up*, *manicure*, *mascara*, *mop*, *perfume*, *razor*, *shampoo*, *soap*, and *sponge*.

The selection of words was based on the semantic categorizations of the *HTOED* and *USAS* (cf. Section 4.3.1.2). Highly polysemous words in which the various senses belong to different semantic domains were avoided, as it would have been necessary to manually disambiguate their different senses and consider their frequencies separately (Leech & Fallon 1992: 38; Potts & Baker 2012: 298). This is the case of nouns such as *oil*, which could be included in both the domains of INDUSTRY AND TECHNOLOGY and of COSMETICS AND HYGIENE with the meanings 'fuel or petroleum' and 'ointment or lotion applied on the body as a cosmetic', respectively.[77] In a similar vein, lexical items with a very low frequency were not selected either, as their contribution to the average NFs of the semantic domains would have been almost negligible. A case in point is the word *deodorant*, which exhibits a maximum NF of 1.66.

restricted here to twenty words per domain. Moreover, the large size of *COHA* would have made it extremely time-consuming to add semantic annotations to it.

[77] Cf. Table 7 in Section 4.3.1.2, where *oil* is classified into two different semantic categories depending on its sense, namely SUBSTANCE AND MATERIAL or COSMETICS.

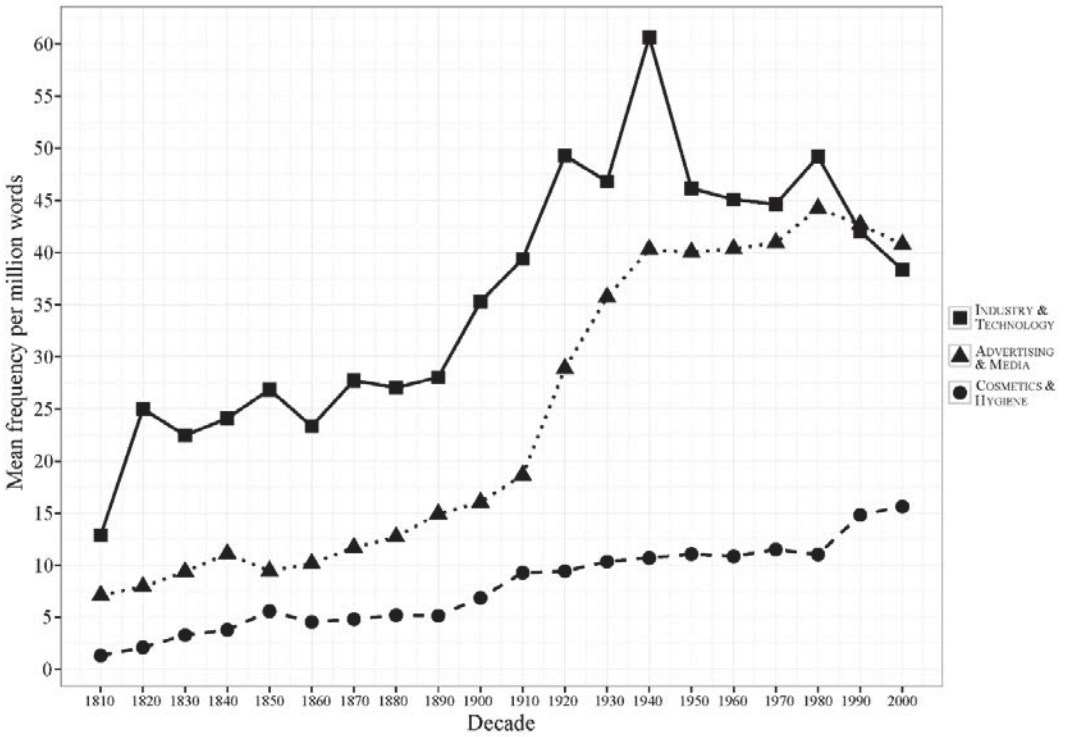


Figure 6. Average frequency per million words in *COHA* of the semantic domains INDUSTRY AND TECHNOLOGY, ADVERTISING AND MEDIA, and COSMETICS AND HYGIENE in the period 1810–2009

The NFs per decade of the 60 selected lexical items were retrieved from *COHA*, and an average NF was computed for each semantic domain. Figure 6 shows the average NFs of the three selected domains across the twenty decades distinguished in *COHA*.[78] All three domains increase steadily in frequency over time, but more dramatically so at the end of the 19th century and the beginning of the 20th century, especially INDUSTRY AND TECHNOLOGY and ADVERTISING AND MEDIA. The domain INDUSTRY AND TECHNOLOGY, the most frequent of the three, exhibits a low NF in the 1810s, before a substantial upsurge takes place from the 1810s to the 1820s and a period of relative stability until the 1890s. From then onwards, however, its NF increases noticeably and steadily, particularly before the 1940s, when it starts losing some ground, although the domain is still much more commonly used than in the 19th century. Similarly, ADVERTISING AND MEDIA increases steadily in frequency throughout the 19th century. However, from the 1910s onwards its NF rises dramatically: between 1910 and 1940 its frequency more than doubles, from 18.63 to 40.31. After the 1940s, the frequency of this domain increases only slightly until it reaches its peak in the 1980s. In turn, COSMETICS AND HYGIENE exhibits the

[78] For the exact figures, cf. Table A.5 in Appendix A.

lowest frequency of the three domains examined here, but its NF still shows the same trend as the other two: it rises steadily over time, but the increase is somewhat more pronounced in the transition from the 19th to the 20th century and onwards. Therefore, the findings obtained regarding these lexical indicators point to the existence of a change at the conceptual level in AmE, which might well reflect the aforementioned socio-economic transformations in American society. It is likely that technological and scientific advancements, which increasingly affected and permeated people's daily lives, implied that these three semantic domains became more frequent topics of conversation. Similarly, it is probably safe to assume that these social, economic, and technological advancements influenced the way in which the concept PLEASANT SMELLING developed over time, from primarily denoting natural smells to also referring frequently to artificial ones. In the next section, a more detailed analysis of each of the five selected adjectival near-synonyms is offered separately to identify how the changes in the concept uncovered here are distributed across the different members of the synonym set. To this purpose, a semasiological perspective is adopted, thus accounting for the prototypical uses of each adjective across senses and semantic categories, both generally and over time.

5.3 SEMASIOLOGICAL ANALYSIS

In Section 4.1.2, an overview of the sense division of the near-synonymous adjectives *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling* was provided on the basis of a series of dictionaries and thesauri (cf. Table 2). It was shown there that while the reference works provided the same basic meaning 'having a sweet pleasant smell' for the five adjectives, a few also explicitly distinguished between the artificial and the natural senses in the case of *perfumed* and *scented*. Therefore, it was hypothesized that *perfumed* and *scented* could be more prototypical than the three remaining adjectives in order to refer to artificial, as opposed to natural, aromas. In this section, the distribution of the five adjectives across the different senses and semantic categories is examined from a semasiological perspective, that is, considering each adjective separately to identify their prototypical uses. In addition, the adjectives are analyzed over time to uncover possible changes in their prototypical structure and to determine whether the changes in the concept PLEASANT SMELLING as a whole may be due to only one or a few of the adjectives or whether, on the contrary, all five near-synonyms undergo similar diachronic semantic developments. The structure of this section mirrors that of the previous one

in that the distribution of the adjectives across senses is first analyzed in Section 5.3.1 and then across semantic categories in 5.3.2. The section closes with a discussion of the findings in 5.3.3.

5.3.1 Sense

Figure 7 displays the overall distribution of each adjective, represented by a separate bar, across the four senses distinguished.[79] Each shade of grey stands for a different sense.

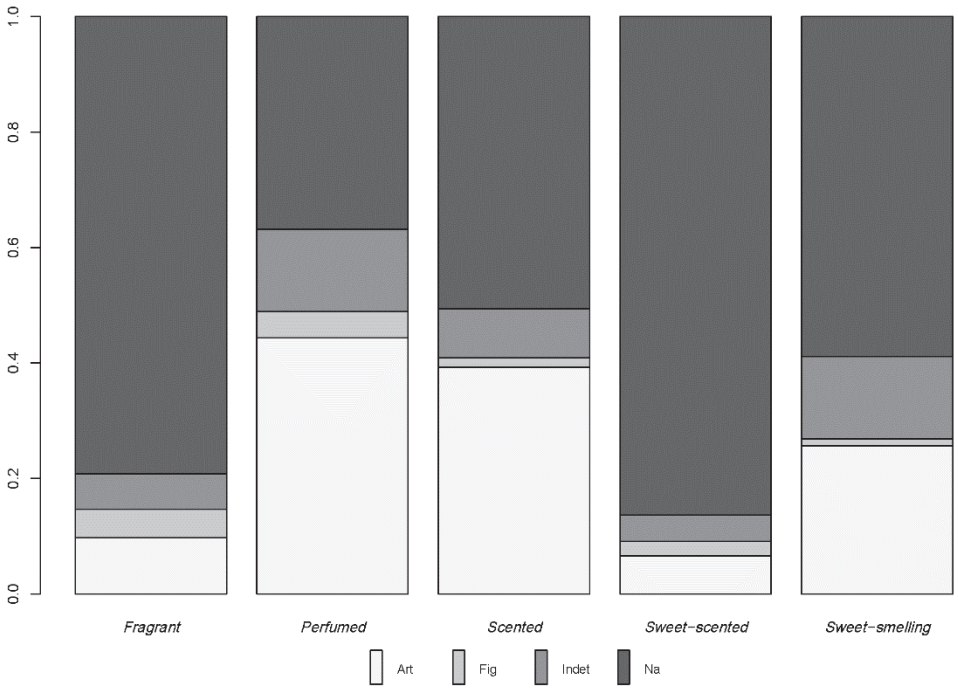


Figure 7. Overall sense distribution (in proportions) per near-synonym

Starting with *fragrant*, it is used to refer to natural smells in almost 80% of the cases, while artificial smells amount to only 9.75% of the data.[80] Similarly, *sweet-scented* is also predominantly 'natural', with more than 85% of the instances of this adjective occurring in this particular sense. On the contrary, both *perfumed* and *scented* are common in both the artificial

[79] No chi-square tests are conducted in this section because (i) the overall distributions of the adjectives across senses and semantic categories are not compared to one another, as a semasiological perspective is adopted here; and (ii) on some occasions, the frequency of some of the adjectives over time across senses and semantic categories is lower than 5. The chi-square test requires that at least 80% of cells in tables larger than 2 by 2 exhibit a frequency of 5 or higher and no cells with zero (Cantos Gómez 2013: 79; Levshina 2015: 213–214). As can be observed in Tables A.6-A.15 in Appendix A, this requirement is not always met, particularly in the case of *sweet-scented* and *sweet-smelling*.

[80] The relative frequencies visualized in the figures presented in this section are provided in tabular form, together with absolute frequencies, in Appendix A (cf. Tables A.6-A.10).

and natural senses. *Perfumed* is more frequent in the artificial sense (45%) vs. natural (37%), while *scented* is used somewhat more frequently in the natural (50.58%) than in the artificial sense (39.21%). Lastly, *sweet-smelling*, although more prototypically 'natural' (58.89 %), also displays a considerable proportion of artificial uses (25.69%), at least if compared to *fragrant* and *sweet-scented*. The figurative and indeterminate senses are not particularly frequent for any of the adjectives. Nevertheless, the value 'indeterminate' is generally more common with the three adjectives that occur relatively frequently in both the natural and artificial senses, that is, *perfumed*, *scented*, and *sweet-smelling*. The low overall frequency of the figurative sense is to be expected, given that metaphorical extensions tend to be less salient than the literal senses from which they originate.

If the distribution of each adjective across senses in the four periods is examined, some interesting developments seem to be at play, as shown in Figure 8 below, which includes five independent plots, one per adjective, with the bars representing the four periods and a different shade of grey for each sense.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

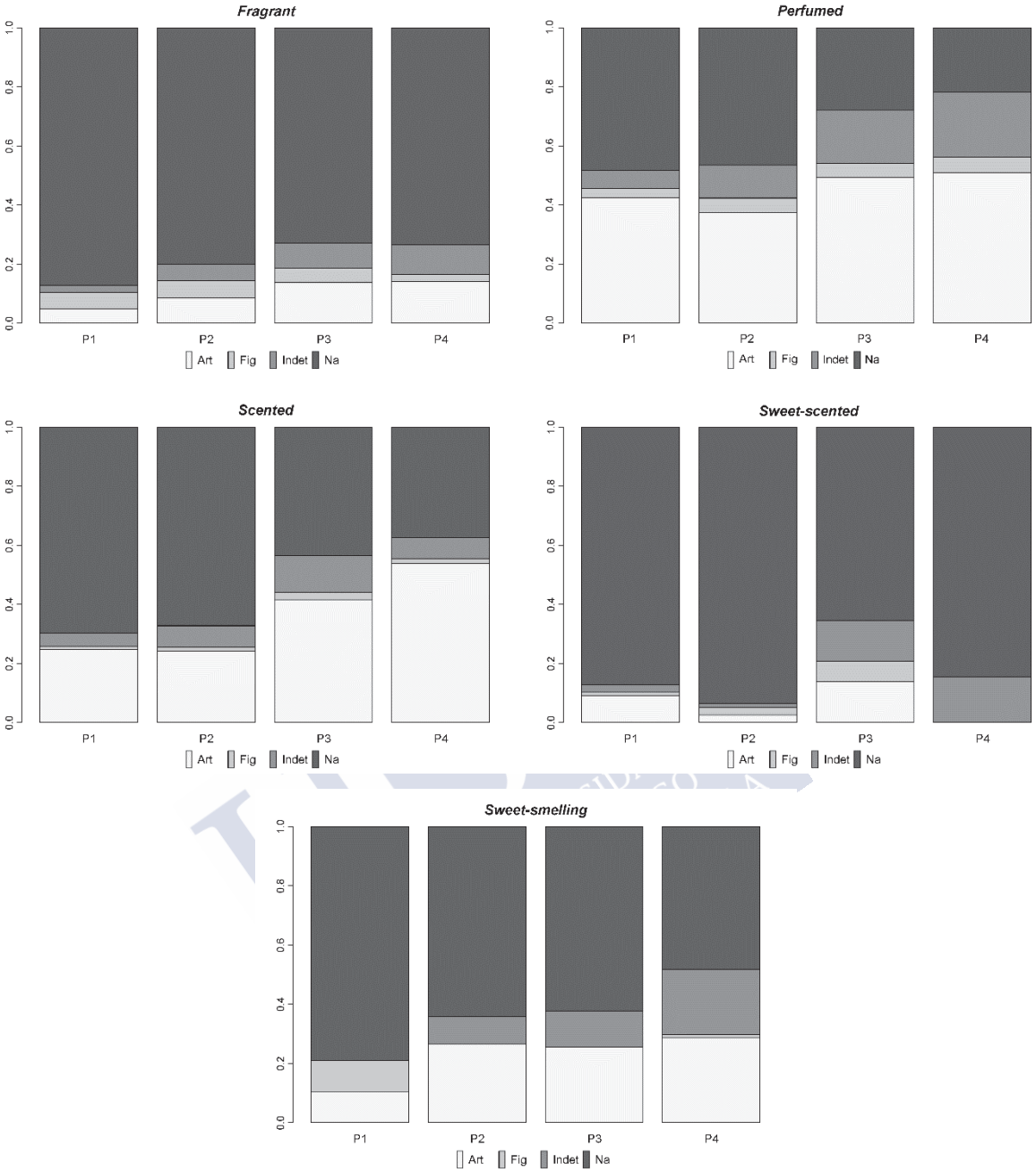


Figure 8. Sense distribution (in proportions) per adjectival near-synonym across periods

As regards *fragrant*, although it remains prototypical in the natural sense, its use in such contexts decreases over time, from nearly 90% in P1 to 73.47% in P4. Similarly, we also witness a decrease of this adjective in the figurative sense across periods, while it becomes more common in both the artificial and indeterminate senses throughout the period 1810–2009, from 4.85% and 2.49% in P1 to 14.12% and 10.03% in P4, respectively. In the case of *perfumed*, we observe an increase in three of the four senses, namely artificial, figurative, and indeterminate,

although this upward tendency is much more pronounced in indeterminate cases, from 6.22% in P1 to 21.88% in P4. The use of this adjective to denote natural aromas, in turn, declines substantially, as it is used in the natural sense more than twice as frequently in P1 (48.19%) than in P4 (21.88%). Similar trends emerge concerning the development of *scented*: it increases in the artificial, figurative, and indeterminate senses and decreases substantially in the natural one over time. Nevertheless, in this case, the rise in use is most noticeable in the artificial sense, from 24.77% in P1 to 53.82% in P4. In fact, the prototypical structure of *scented* changes throughout the time span considered, from being used mostly to denote natural smells in P1 to becoming more prototypically 'artificial' in P4. Concerning the adjective *sweet-scented*, it experiences less fluctuations over time as regards its distribution across senses, as it remains predominantly 'natural' throughout the period 1810–2009, with only an increase in indeterminate uses. However, due to its low absolute frequencies in all senses except the natural one, specifically in P3 and P4, no reliable conclusions can be drawn from the data. Finally, *sweet-smelling* becomes more frequent over time in the artificial sense and in indeterminate uses, while its occurrence in the natural sense drastically declines. However, despite this substantial downward tendency, *sweet-smelling* is still more prototypically used in the natural sense in P4.

5.3.2 Semantic category

The present section analyzes whether the five near-synonymous adjectives occur as modifiers of some specific types of nouns by considering their distribution across the twelve semantic categories. Figure 9 provides the overall tendencies of the adjectives individually. Each bar represents one of the adjectives and, within the bars, the colors indicate the proportion of use of the adjectives with particular semantic categories.

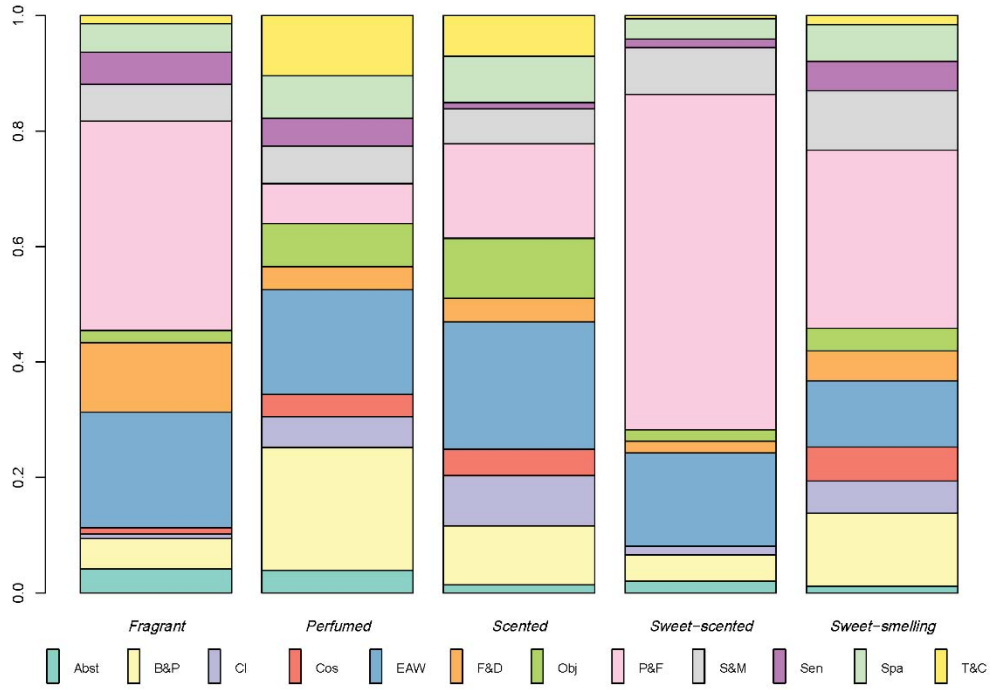


Figure 9. Overall semantic category distribution (in proportions) per near-synonym

The overall trends shown in Figure 9 demonstrate that *fragrant*, as expected, is most often used with the three prototypically natural semantic categories PLANTS AND FLOWERS (36.29%), EARTH, ATMOSPHERE, AND WEATHER (20.00%), and FOOD AND DRINK (12.05%), while it almost never occurs with nouns belonging to CLEANING (0.80%), COSMETICS (1.09%), and TEXTILE AND CLOTHING (1.38%), that is, the prototypically 'artificial' categories.[81] Similarly, *sweet-scented* is also very frequently used with PLANTS AND FLOWERS and EARTH, ATMOSPHERE, AND WEATHER nouns, which together account for more than 70% of its total occurrences. In the remaining semantic categories this adjective is relatively uncommon, with less than 10 instances in each of them, except in SUBSTANCE AND MATERIAL (16 examples; 8.08%). In turn, the three categories of nouns with which *perfumed* collocates mostly are BODY AND PEOPLE (21.29%), EARTH, ATMOSPHERE, AND WEATHER (18.12%), and TEXTILE AND CLOTHING (10.44%). In the other nine categories *perfumed* occurs with a lower frequency and with a rather balanced distribution. In the case of *scented*, the order of semantic categories according to frequency is as follows: EARTH, ATMOSPHERE AND WEATHER (22.10%), PLANTS AND FLOWERS

[81] The relative frequencies which are discussed in the text and visualized in the figures in this section are given in tabular form, together with absolute frequencies, in Tables A.11-A.15 in Appendix A.

(16.37%), OBJECT (10.41%), and BODY AND PEOPLE (10.20%). In ABSTRACT and SENSATION, the use of *scented* is, on the contrary, only residual, with 13 (1.38%) and 10 (1.06%) instances, respectively. Lastly, *sweet-smelling* is most commonly used with nouns denoting PLANTS AND FLOWERS (30.83%). Additionally, this adjective is also relatively frequent in the semantic categories BODY AND PEOPLE (12.65%), EARTH, ATMOSPHERE, AND WEATHER (11.46%), and SUBSTANCE AND MATERIAL (10.28%).

Figure 10 shows the distribution of each adjective across semantic categories in the four periods. Again, five separate plots, one per adjective, are included, with the bars representing the four periods and, within each period, one color standing for each semantic category.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

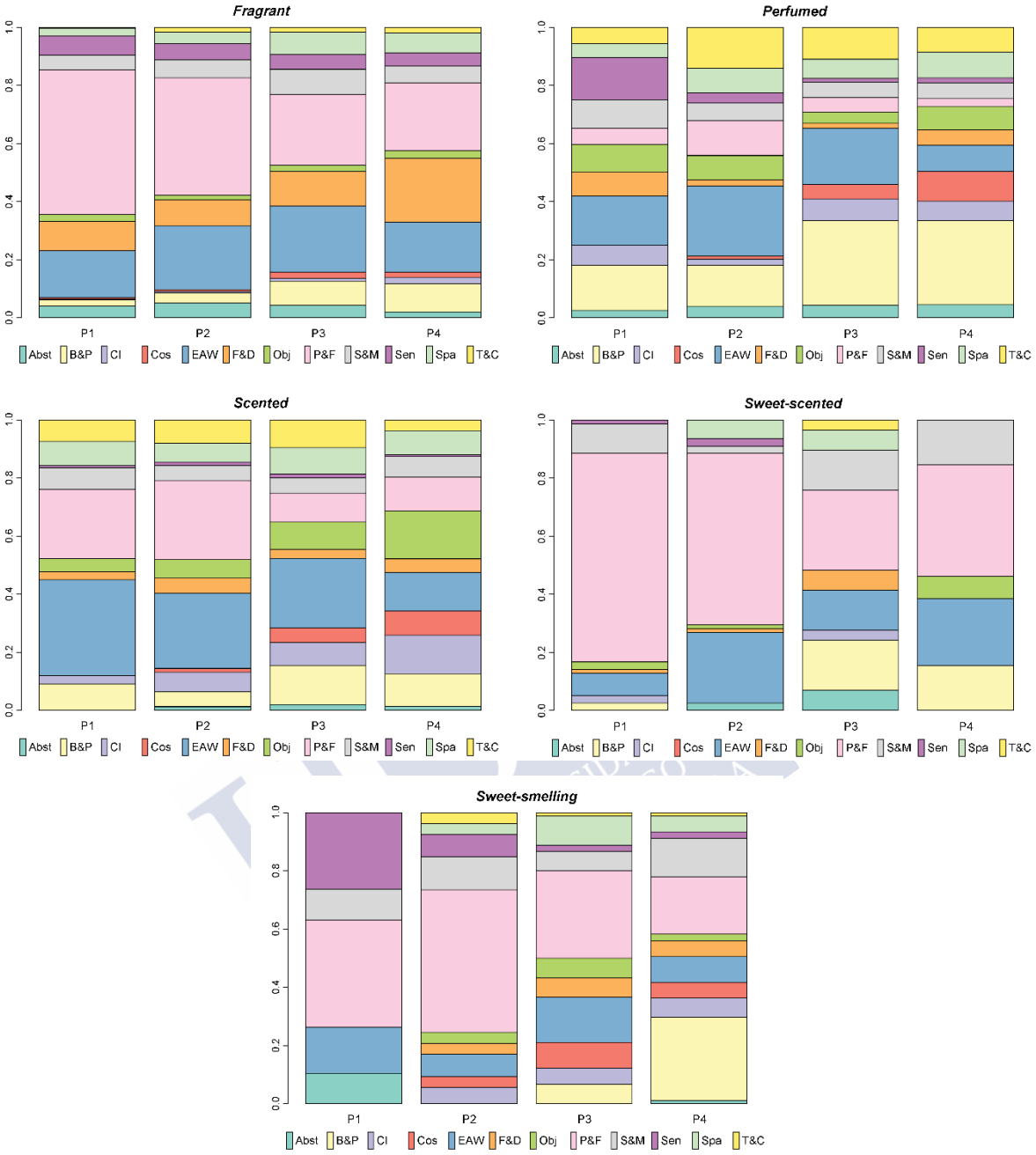


Figure 10. Semantic category distribution (in proportions) per near-synonym across periods

In the case of *fragrant*, the most prototypical semantic category in P1 is clearly PLANTS AND FLOWERS, with a relative frequency of 49.88%, followed by EARTH, ATMOSPHERE, AND WEATHER (16.17%) and FOOD AND DRINK (9.95%). However, this state of affairs changes substantially over time: *fragrant* decreases considerably with PLANTS AND FLOWERS nouns, and increases considerably with FOOD AND DRINK nouns, to the point that both categories are almost equally frequently modified by this adjective in P4 (23.13% and 21.94%, respectively). The

category EARTH, ATMOSPHERE, AND WEATHER, on the other hand, remains rather stable. Interestingly, two categories of nouns which are very infrequent with *fragrant* in P1 become more common by the end of the time span considered. These are BODY AND PEOPLE and SPACE, which go from 2.11% and 2.49% in P1 to 9.69% and 6.97% in P4, respectively.

In the case of *perfumed*, the following semantic categories experience noticeable fluctuations diachronically. On the one hand, BODY AND PEOPLE, COSMETICS, SPACE, and TEXTILE AND CLOTHING increase greatly in frequency over the period 1810–2009. The increase is particularly pronounced in the case of BODY AND PEOPLE and COSMETICS, which go from 15.54% and 0% in P1 to 29.02% and 10.27% in P4, respectively. On the other hand, four categories exhibit a downward tendency with this adjective, namely EARTH, ATMOSPHERE, AND WEATHER, FOOD AND DRINK, PLANTS AND FLOWERS, SUBSTANCE AND MATERIAL, and SENSATION. In the case of SENSATION, there is a sharp decrease from P1 to P2 (from 14.51% to 3.60%. However, the high frequency of this category in P1 is due to just three particular texts in *COHA* in which SENSATION nouns are commonly modified by *perfumed*.[82]

Several categories of nouns also become either more or less frequently modified by *scented* over time. The most noticeable changes in the case of this adjective occur in the categories CLEANING, COSMETICS, and OBJECT, on the one hand, which increase substantially from P1 to P4, and EARTH, ATMOSPHERE, AND WEATHER and PLANTS AND FLOWERS, on the other, which decrease considerably.

Concerning the evolution of *sweet-scented*, although the scarcity of data did not allow to draw reliable conclusions, one interesting trend is observed: this adjective becomes a much less common modifier of nouns belonging to the category PLANTS AND FLOWERS, from 71.79% in P1 to 38.46% in P4. Nevertheless, PLANTS AND FLOWERS is still the category which is most often modified by *sweet-scented* in P4.

[82] In these texts a total of 14 instances of *perfumed* occur as modifier of the SENSATION noun *flavor*. The three texts are the following non-fictional sources: *He fruits and fruit trees of America: the culture, propagation, and management, in the garden and orchard, of fruit trees generally: With descriptions of all the finest varieties of fruit of all the finest varieties of fruit, native and foreign, cultivated in the country* (HeFruitsFruitTree) from the year 1845; *The American fruit culturist: Containing directions for the propagation and culture of fruit trees, in the nursery, orchard, and garden, with descriptions of the principal American variety and foreign varieties, cultivated in the United States* (AmericanFruitCulturalist) from the year 1850; and *Pear culture. A manual for the propagation, planting, cultivation, and management of the pear tree* (PearCultureAManual) from the year 1858.

Similarly to *sweet-scented*, *sweet-smelling* also exhibits a low frequency in some semantic categories and thus the reliability of the results is at least questionable. However, some types of nouns do undergo changes over time, although only in relative, rather than absolute, terms. A considerable increase is observed in the semantic category BODY AND PEOPLE, from 0% in P1 to 28.57% in P4, while an important decrease takes place with PLANTS AND FLOWERS (from 36.84% to 19.78%) and SENSATION (from 26.32% to 2.20%).

5.3.3 Interim summary and discussion

This section has focused on the semasiological patterns of the five near-synonymous adjectives individually, both in general and throughout the time span 1810–2009. Their distribution across the four senses and twelve semantic categories uncovers some interesting differences, but also some similarities in their semasiological structures. First, we have seen that, overall, two of the adjectives, namely *fragrant* and *sweet-scented*, are prototypically used in the natural sense, whereas they are not very frequent in the artificial one. *Sweet-smelling* is also predominantly used in the natural sense, but, contrary to *fragrant* and *sweet-scented*, it is not so uncommon in the artificial sense. Finally, *perfumed* and *scented* are more evenly distributed across the natural and artificial senses, with *perfumed* coming closer to the artificial end of the semantic spectrum and *scented* to the natural one. Figurative and indeterminate uses display a much lower frequency, and none of the adjectives are particularly salient in any of these two uses. This is not surprising, given that figurative senses are often less salient than more literal ones, and since the value 'indeterminate' does not, as explained earlier (cf. Section 4.3.1.1), constitute a separate sense, but merely groups together ambiguous examples.

A more fine-grained picture of the semasiological structures of the adjectives was obtained by examining their distribution across the semantic categories of nouns, though similar tendencies were observed: *fragrant* and *sweet-scented* occur mostly in the prototypically 'natural' categories, particularly in PLANTS AND FLOWERS, EARTH, ATMOSPHERE, AND WEATHER, and, in the case of *fragrant*, also in FOOD AND DRINK. Two of these semantic categories of nouns, to wit, PLANTS AND FLOWERS and EARTH, ATMOSPHERE, AND WEATHER, are in fact also among the most common contexts of use with the remaining three adjectives, *perfumed*, *scented*, and *sweet-smelling*, with the exception of *perfumed* in the former. The fact that these two semantic categories of nouns are frequently modified by all five adjectives explains the very high proportion of these particular categories with the concept PLEASANT SMELLING as a whole (cf.

Figure 5 in Section 5.2.2). In addition, *perfumed*, *scented*, and *sweet-smelling* are also relatively salient in the categories BODY AND PEOPLE, TEXTILE AND CLOTHING (in the case of *perfumed*), OBJECT (in the case of *scented*), and SUBSTANCE AND MATERIAL (in the case of *sweet-smelling*). The remaining five categories of nouns, i.e. ABSTRACT, CLEANING, COSMETICS, SENSATION, and SPACE, are not particularly typical with any of the five adjectives and therefore display a much lower overall frequency.

These results seem to be mostly in line with the definitions provided by the reference material (cf. Section 4.1.2), as *fragrant* and *sweet-scented* are both infrequent to refer to artificial aromas. Nevertheless, even though an artificial sense is not provided for *sweet-smelling* in any of the dictionaries consulted, this adjective does occur in this sense relatively frequently in the data. As hypothesized in Section 4.1.2, *perfumed* and *scented* are the two adjectives with the highest proportion of uses in the artificial sense, which may explain why some of the dictionaries distinguish between a natural and an artificial sense for these two items. It is plausible that this state of affairs, i.e. *perfumed* and *scented* being the most 'artificial' adjectives of the synonym set, derives from their etymology. As mentioned in Section 4.1.2, both *perfumed* and *scented* are derivatives formed by the suffix *-ed* (*OED*, s.v. *-ed* suffix[1]), which were coined after their base form *perfume* and *scent* had already been borrowed into English. Therefore, *perfumed* and *scented* still retain their participial shape and also possibly part of their verbal semantic functions, namely, denoting a process or action whereby something is impregnated with a pleasant smell. Consequently, it is not at all surprising that *perfumed* and *scented* are more salient in the artificial sense than the three remaining adjectives.

Regarding the diachronic evolution of the semasiological structure of the adjectives, similar trends can be observed for several of the five near-synonyms. First, four of the five adjectives, namely *fragrant*, *perfumed*, *scented*, and *sweet-smelling*, become more frequent in the artificial sense and, to a certain extent, in indeterminate uses, while they undergo a considerable decrease in the natural sense. Such tendencies seem to be more pronounced in the case of *perfumed* (particularly in the natural sense and indeterminate cases) and of *scented* (particularly in the natural and artificial senses). On the other hand, *sweet-scented* constitutes an exception, since it remains rather stable over time and is mainly used to refer to natural aromas, with more than 80% of its occurrences corresponding to this specific sense in three of the four periods and only a slight dip in P3. Consequently, the diachronic trend across senses uncovered in Section 5.2.1 regarding the concept PLEASANT SMELLING as a whole seems to be relatively general, occurring

across the board, with the exception of *sweet-scented*.[83] Second, in Section 5.2.2 two main diachronic trends were identified with respect to the distribution of the concept as a whole across semantic categories: (i) a substantial decrease with nouns belonging to PLANTS AND FLOWERS, and (ii) an increase with BODY AND PEOPLE, CLEANING, and COSMETICS. The findings of the semasiological analysis demonstrate that whereas the decrease in PLANTS AND FLOWERS affects all five adjectives, the increase in BODY AND PEOPLE is restricted mainly to the adjectives *fragrant*, *perfumed*, and *sweet-smelling*. Similarly, the upward tendency in COSMETICS is witnessed mostly with *perfumed* and *scented*, and that of CLEANING only with *scented*.[84]

In conclusion, the findings in Section 5.3 suggest the existence of a continuum of senses with *sweet-scented* and *fragrant* being located at or towards the natural end of the cline and *perfumed* at the artificial end. *Sweet-smelling* and *scented*, in turn, occupy intermediate positions on the cline, since although both exhibit the same preference for the natural sense overall, they are also quite salient in the artificial sense, especially in the case of *scented*. Figure 11 visualizes this cline:

+ natural *sweet-scented – fragrant – sweet-smelling – scented – perfumed* **+ artificial**

Figure 11. Continuum of senses of the five adjectives

Interestingly, there seems to exist a correlation between the position occupied by the adjectives on the continuum of senses and the diachronic frequency fluctuations identified in Section 5.1 (cf. Figure 3 and Table 9): *sweet-scented* and *fragrant*, the prototypically 'natural' adjectives, are the only near-synonyms in the set that undergo a significant decrease in frequency from P1 to P4. On the contrary, the adjectives occupying intermediate positions on the continuum, i.e. *scented* and *sweet-smelling*, are the only ones that increase significantly in frequency over time, while *perfumed*, the prototypically 'artificial' member of the set, remains rather stable in terms of frequency. It may be the case that the relative neutrality of *scented* and *sweet-smelling*, not being restricted to particular semantic categories, enables them to occur relatively often in a wider range of contexts than the other adjectives. This, in turn, may explain why they are the

[83] However, it has to be borne in mind that the low frequency of *sweet-scented* in many of its senses does not allow to reach a definitive conclusion about its semantic development.

[84] Although Figure 10 suggests a noticeable increase of CLEANING and COSMETICS also with *sweet-smelling*, as mentioned earlier, the scarcity of instances of the adjective in these two semantic categories (cf. Table A.15 in Appendix A) makes it impossible to prove the strength of such tendencies.

only adjectives that increase diachronically, given that the natural and artificial senses seem to even out progressively in terms of frequency over time. However, it is surprising that *perfumed*, the prototypically 'artificial' adjective, remains stable even though the artificial sense becomes diachronically more and more salient at the expense of the natural one.

So far, attention has been paid to the distribution of the concept PLEASANT SMELLING as a whole (cf. Section 5.2), as well as to the semasiological structure of each adjective individually (cf. Section 5.3). Nonetheless, given that the five lexical items are semantically related and occur in very similar contexts of use, the analysis of the competition between them in such contexts may help to reach a fuller understanding of their relation. To this purpose, an onomasiological perspective must also be adopted in which the five near-synonymous adjectives are compared to one another across senses and semantic categories. As mentioned in the introductory section to this chapter, semasiological and onomasiological analyses offer only a partial picture, since prototypicality and salience effects may vary if either one or the other perspective is adopted. In other words, what is semasiologically salient or prototypical may not necessarily be onomasiologically so. Therefore, leaving out one of the two perspectives could lead to an incomplete interpretation of the data. In this context, the next section deals with the onomasiological structure of the synonym set.

5.4 ONOMASIOLOGICAL ANALYSIS

To analyze the onomasiological structure of the synonyms in different contexts of use, the different senses and semantic categories constitute the point of departure in this section. Within each sense and semantic category, the proportions of use of the five adjectives are compared, first generally and then across time. As argued in Section 2.4.1, the degree of onomasiological salience —or entrenchment— of a lexical item with respect to other competing forms depends on its frequency in a particular sense or function rather than on its overall frequency (Schmid 2007: 119). Thus, despite the general dominance of *fragrant* in terms of frequency, equaling 58.90% of the total data (cf. Table 9 in Section 5.1), it may well be the case that *fragrant* is outperformed by some of the other adjectives in specific contexts of use. In particular, both *perfumed* and *scented* are relatively semasiologically salient in the artificial sense, as opposed to *fragrant*, which is more semasiologically salient in the natural sense (cf. Section 5.3.1). Therefore, if an onomasiological perspective is adopted, *perfumed* and/or *scented* may surpass *fragrant* in 'artificial' contexts of use, including semantic categories which are prototypically

'artificial' (e.g. CLEANING, COSMETICS, and TEXTILE AND CLOTHING). As in the case of the two preceding sections, the adjectives are first examined across senses in Section 5.4.1, and then across semantic categories in 5.4.2. Finally, the results are discussed in Section 5.4.3.

5.4.1 Sense

Figure 12 provides the overall proportion of the five adjectives across each of the four senses with respect to one another.[85] In this figure, therefore, each bar represents one of the four senses, and within each bar, a different shade of grey stands for one of the near-synonymous adjectives.

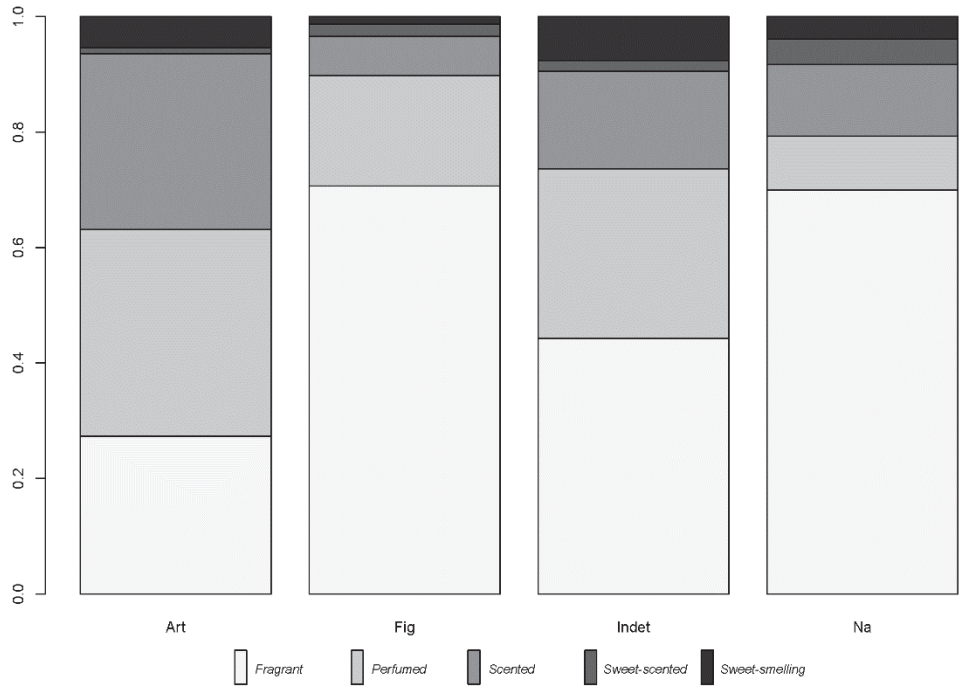


Figure 12. Overall proportions of the five near-synonyms in each sense

As was to be expected due its overall much higher frequency, *fragrant* is the most entrenched variant in three of the four senses, namely figurative, indeterminate, and natural. Its dominance

[85] As in the previous section, no chi-square analyses are conducted here. This is so because (i) in several cells in Tables A.16-A.25 in Appendix A the frequency is zero or lower than 5 and (ii) the diachronic analyses in this section involve three variables, namely Synonym, Period, and Sense/Semantic category, while chi-square tests are used to test possible correlations only between two categorical variables (Desagulier 2017: 178). In order to apply chi-square tests to the analyses in this section, they would have had to be implemented to the data from each period separately, which would not allow to test for significant differences between the periods. To determine the significance of the diachronic tendencies identified in this section, more sophisticated statistical tests which allow to account for the simultaneous effects of more than two independent variables on a dependent one are conducted in Chapter 6.

is particularly pronounced in the figurative and natural senses, where it accounts for 70.64% and 69.93% of the cases, respectively.[86] In the indeterminate sense, *fragrant* also dominates (44.30%), but *perfumed* is relatively common as well (29.32%). However, despite its overall predominance in terms of frequency, *fragrant* is outperformed by both *perfumed* and *scented* in the artificial sense (35.76% and 30.47%, respectively vs. 27.33%). Nevertheless, the difference between these three adjectives in the artificial sense is not particularly large. Finally, neither *sweet-scented* nor *sweet-smelling* seem to be especially entrenched in any of the four senses, as is to be expected due to their overall low frequency (cf. Table 9 in Section 5.1).

Figure 13 offers the proportions of the five adjectives with respect to the different senses in the four periods so as to identify potential differences concerning their distribution over time. In this case, four separate plots, one per period, are provided. Again, the bars stand for the different senses and within them, the various shades of grey represent the near-synonymous adjectives.

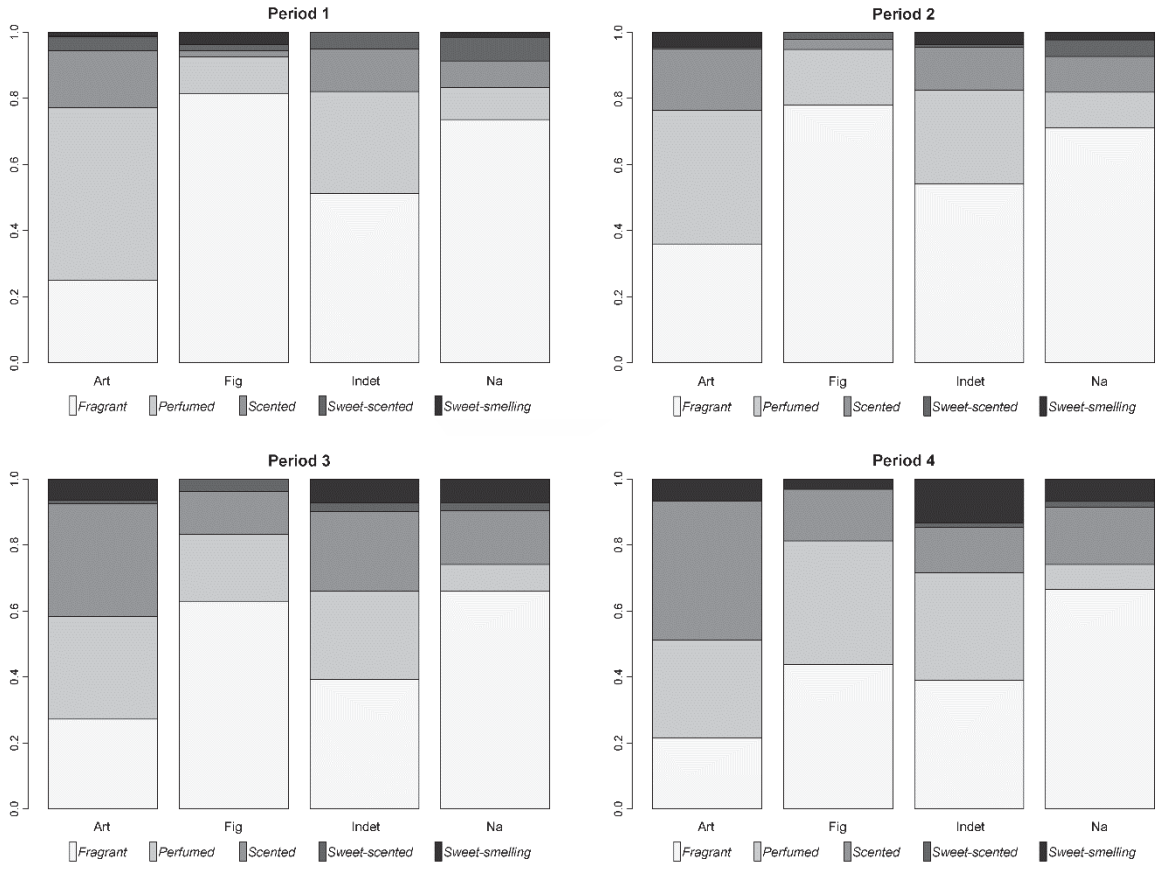


Figure 13. Proportions of the five near-synonyms in each sense across periods

[86] The relative frequencies visualized in the figures provided throughout Section 5.4.1 are given in tabular form, together with absolute frequencies, in Tables A.16–A.20 in Appendix A.

Several changes seem to have taken place over the period 1810–2009 as concerns the proportions of the five adjectives across the different senses. First, in the artificial sense we observe a change in the frequency order of the adjectives over time: in P1 *perfumed* is clearly dominant as it accounts for 52.23%. However, from P2 onwards, its proportion gradually decreases, first in favor of *fragrant* in P2, and then mainly in favor of *scented* in P3 and P4. In fact, in P4, *scented* totals 42.08% of the data in the artificial sense, while *perfumed* amounts to 29.61%. As regards the figurative sense, *fragrant* continues to be the most frequent member of the set throughout the time span analyzed but, with each subsequent period, *perfumed* takes up more and more ground. In P4, for instance, *perfumed* is almost as frequent as *fragrant* in this sense (37.50% vs. 43.75%). Similarly, in the value 'indeterminate' *fragrant* decreases in favor of *sweet-smelling*, which is the only adjective that increases considerably with respect to the others in this use, from 0% in P1 to 13.25% in P4.[87] Lastly, in the natural sense, no remarkable changes can be observed, except for a moderate decrease of *fragrant* in favor of *scented*, with the latter increasing from 7.97% in P1 to 17.41% in P4. Nevertheless, *fragrant* is still clearly predominant in the natural sense even in P4, totaling 66.56% of the data.

In sum, the data suggests an overall reorganization in the onomasiological structure of the set over time, which results in (i) *fragrant* no longer being markedly dominant in the figurative sense and in indeterminate uses by P4 and (ii) *scented* surpassing *perfumed* in the artificial sense at the end of the time span considered here. In what follows, the proportions of use of the near-synonyms across semantic categories are examined. This is done with the aim of identifying in which particular contexts of use the changes uncovered regarding the senses are most pronounced.

5.4.2 Semantic category

Figure 14 presents the overall proportions of the five near-synonyms across the twelve semantic categories distinguished.[88] In this case, each bar represents one semantic category, and within each of them, a different shade of grey stands for the proportion of use of each adjective.

[87] The proportions of *perfumed* and *scented* also increase marginally in the value 'indeterminate' over time, from 30.77% and 12.82% in P1 to 32.45% and 13.91% in P4, respectively.

[88] The relative frequencies visualized in Figures 14 and 15 are given in Tables A.21–A.25 in Appendix A, together with the absolute frequencies of the five adjectives.

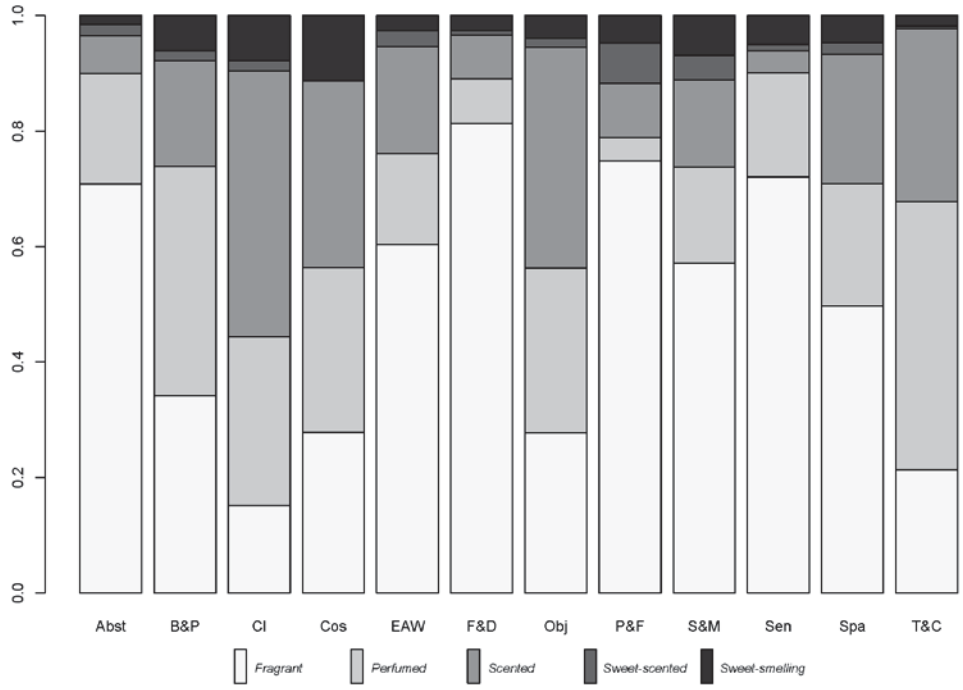


Figure 14. Overall proportions of the five near-synonyms in each semantic category

As can be seen, *fragrant* is overall the default choice in seven out of the twelve categories, namely ABSTRACT (70.85%); EARTH, ATMOSPHERE, AND WEATHER (60.36%); FOOD AND DRINK (81.31%), PLANTS AND FLOWERS (74.80%); SUBSTANCE AND MATERIAL (57.14 %); SENSATION (72.03%); and SPACE (49.71%). It is also relatively frequent, although surpassed by *perfumed* and/or *scented*, in BODY AND PEOPLE (34.16%), COSMETICS (27.82%), OBJECT (27.73%), and TEXTILE AND CLOTHING (21.36%). Finally, in CLEANING *fragrant* exhibits its lowest proportion of use, accounting for only 15.17% of the total instances in this semantic category. In turn, *perfumed* is the dominant adjective with two semantic types of nouns, to wit, BODY AND PEOPLE (39.69%), where it is closely followed by *fragrant* (34.16%), and TEXTILE AND CLOTHING (46.36%). In addition, it is also relatively common in CLEANING (29.21%), COSMETICS (28.57%), OBJECT (28.52%), and SPACE (21.18%). *Scented* surpasses the other adjectives in the set in the categories CLEANING (46.07%), COSMETICS (32.33%), and OBJECT (38.28%), and is reasonably frequent in SPACE (22.35%) and TEXTILE AND CLOTHING (30.00%). Finally, *sweet-scented* and *sweet-smelling* do not stand out in any of the twelve semantic categories. Nevertheless, the former reaches its peak in PLANTS AND FLOWERS (6.98%), where it is more frequent than both *perfumed* and *sweet-smelling* (4.13% and 4.74%, respectively), and the latter

in COSMETICS (11.28%), followed by CLEANING (7.87%), SUBSTANCE AND MATERIAL (6.88%), and BODY AND PEOPLE (6.11%).

Therefore, from an onomasiological perspective, *fragrant* is most salient in all five prototypically 'natural' semantic categories (i.e. EARTH, ATMOSPHERE, AND WEATHER, FOOD AND DRINK, PLANTS AND FLOWERS, SENSATION, and SPACE), as well as in the 'figurative' category ABSTRACT and in the relatively 'artificial' category SUBSTANCE AND MATERIAL. On the contrary, both *perfumed* and *scented* display a clear preference for 'artificial' categories, as the former dominates in TEXTILE AND CLOTHING and the latter in CLEANING, COSMETICS, and OBJECT. In addition, *perfumed* is the most salient adjective in BODY AND PEOPLE, a mixed category with an extremely high proportion of indeterminate cases (51.53%; see Table 12 in Section 5.2.2). As in the case of the distribution of the adjectives across senses (cf. Section 5.4.1), neither *sweet-scented* nor *sweet-smelling* are particularly entrenched in any of the twelve semantic categories, a fact which is undoubtedly related to their low overall frequency in the corpus.

Figure 15 presents the diachronic evolution of the five adjectives across semantic categories. Once again, a separate graph is plotted for each of the four periods.

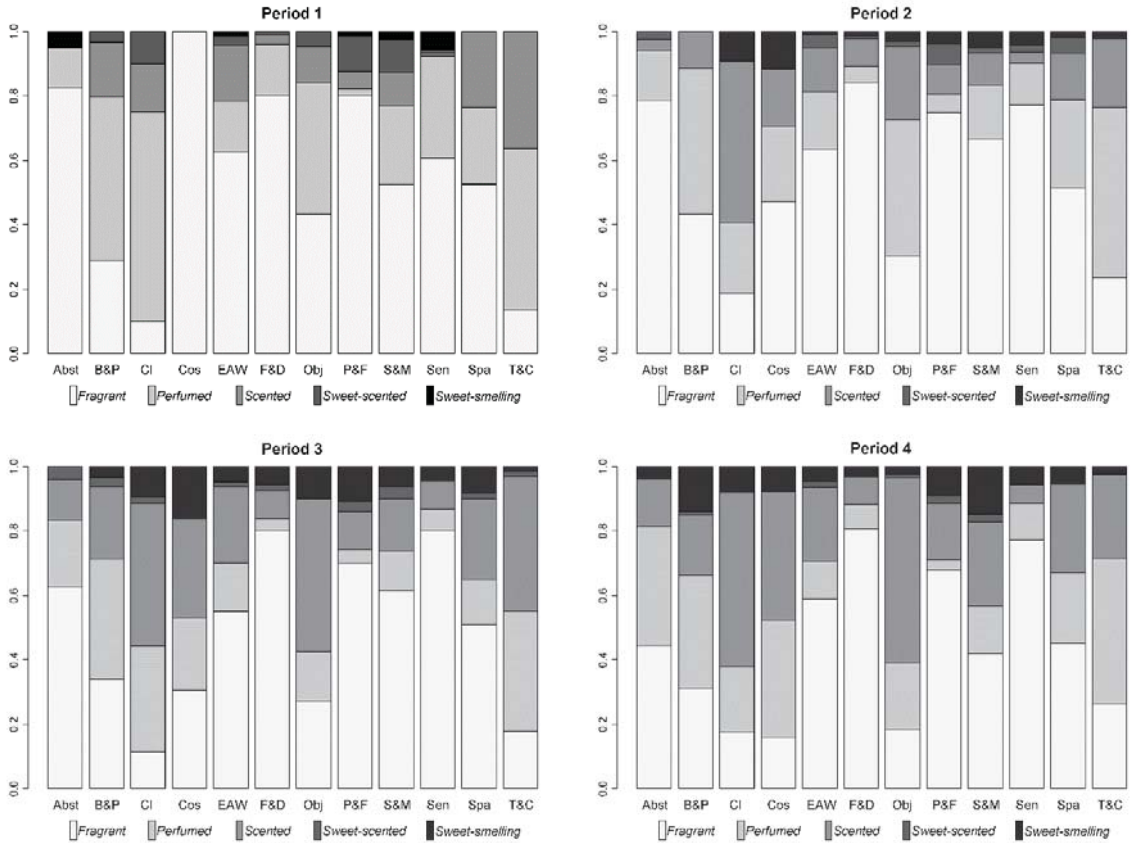


Figure 15. Proportions of the five near-synonyms in each semantic category across periods

Two clear tendencies can be observed. On the one hand, as shown in the figure, in half of the semantic categories, the situation remains rather stable from P1 to P4. This is the case of the categories EARTH, ATMOSPHERE, AND WEATHER, FOOD AND DRINK, PLANTS AND FLOWERS, and SENSATION, where *fragrant* maintains a clear dominance over time with only a minimal decrease in some cases, mostly in favor of *scented*. Additionally, stability also applies to the distribution of the near-synonyms in the categories SPACE and TEXTILE AND CLOTHING. In SPACE, *fragrant* is the most common adjective throughout the period 1810–2009, but its dominance is not as clear as in the four semantic categories mentioned above, where the frequency of both *perfumed* and *scented* is much lower. In TEXTILE AND CLOTHING, *perfumed* and, to some extent, *scented* dominate the scene in all four periods, with only a slight increase of *fragrant* in P4.

On the other hand, diachronic tendencies can be observed as regards the remaining semantic categories:

(i) ABSTRACT shows a decrease of *fragrant* in favor of *perfumed*, so that in P4 the former is only marginally more frequent than the latter (44.44% vs. 37.04%).

(ii) Concerning BODY AND PEOPLE, *perfumed* loses its initial dominance in P1 (50.85%), as the frequencies of both *scented* and *sweet-smelling* increase over time. By P4, there exists a more balanced distribution of the adjectives in this category, with *perfumed* and *fragrant* outperforming the others (35.33% and 30.98%, respectively) but being rather closely followed by *scented* and *sweet-smelling* (18.48% and 14.13%, respectively).

(iii) In CLEANING, *scented* increases dramatically over time at the expense of *perfumed* and surpasses it already from P2 onwards.

(iv) As regards COSMETICS, from P2 onwards *perfumed* and *scented* gain more and more ground at the expense of *fragrant*, and in P4 they are the most frequent adjectives with nouns denoting cosmetic products, with 36.51% and 39.68%, respectively (vs. the 15.87% of *fragrant*).[89]

(v) With respect to the category OBJECT, in P1 *fragrant* and *perfumed* are clearly predominant (43.18% and 40.91%, respectively). However, over time, particularly

[89] Little can be said about the distribution of COSMETICS in P1, as there are only 4 cases in total, which correspond to instances of *fragrant*.

from P2 to P3, there is a drastic increase of *scented*, which comes to represent 57.47% of the total data in P4, that is, more than double the amount of the second most frequent adjective (i.e. *perfumed*; 20.69%).

(vi) Lastly, in SUBSTANCE AND MATERIAL, although *fragrant* remains the most frequent adjective throughout the whole time span considered, *perfumed* and *scented* invert positions over time: whereas *perfumed* is more frequent than *scented* in P1 and P2, *scented* is more common than *perfumed* in P3 and P4. In fact, by P4 the frequency of *perfumed* equals that of *sweet-smelling* (i.e. 14.81%), as the latter also increases substantially in the period 1810–2009 in this category.[90]

5.4.3 Interim summary and discussion

The results presented in this section regarding the onomasiological structure of the synonym set therefore demonstrate that *fragrant* is the most salient adjective in the natural sense across all five prototypically 'natural' categories, that is, EARTH, ATMOSPHERE, AND WEATHER, FOOD AND DRINK, PLANTS AND FLOWERS, SENSATION, and SPACE. Moreover, it maintains its pronounced dominance in all these contexts throughout the four periods, with only a slight decrease in frequency, mostly in favor of *scented*. On the contrary, all prototypically 'artificial' categories, except TEXTILE AND CLOTHING, do exhibit distributional changes over time. In particular, *scented* gains ground at the expense of *perfumed* and, to a certain extent, *fragrant*, and becomes the most salient adjective in the artificial sense by P4, more specifically in the semantic categories CLEANING, COSMETICS, and OBJECT. Additionally, changes are also identified in the figurative sense and, consequently, in the category ABSTRACT, to which a great majority of instances in this sense belong. In this case, *perfumed* increases considerably and becomes almost as salient as *fragrant* in the last period. Finally, in the value 'indeterminate', *fragrant* decreases somewhat over time, mainly in favor of *sweet-smelling*, which increases considerably from P1 to P4. In particular, this increase of *sweet-smelling* seems to have taken place with nouns belonging to the category BODY AND PEOPLE, which, as has been argued before, comprises a very high number of ambiguous examples classified here as indeterminate.

[90] Nothing can be said about the diachronic evolution of *sweet-scented* across semantic categories when compared to the other four adjectives, as its frequency is almost negligible in all four periods, even in the category in which it is most frequent, i.e. PLANTS AND FLOWERS.

In sum, as was the case in the previous sections of this chapter,[91] several diachronic changes also emerge if an onomasiological perspective is adopted, as proved by the rearrangement in the proportions of use of the different adjectives in several of the senses and semantic categories. In Section 5.1, it was shown that both *fragrant* and *sweet-scented* decreased in frequency over time, while *scented* and *sweet-smelling* increased, and *perfumed* remained relatively stable (cf. Table 9 and Figure 3). The results in Section 5.4 demonstrate that the decrease of *fragrant* takes place mainly in the figurative sense and in indeterminate uses, and semantic categories associated with these two values, but that this adjective also undergoes a slight decline in 'natural' contexts of use. The increase of *scented*, in turn, affects mostly the artificial sense (i.e. CLEANING, COSMETICS, and OBJECT), with a minimal increase also in the natural one. In the case of *sweet-smelling*, it increases primarily in the indeterminate cases, particularly with nouns denoting BODY AND PEOPLE. Finally, regarding *perfumed* and *sweet-scented*, the former becomes more salient over time in the figurative sense, although it does not exhibit an overall increase in frequency, while the latter decreases mainly in the natural sense, that is, the only sense for which there is enough data to trace diachronic fluctuations for this adjective. Therefore, the changes in the frequency of the adjectives uncovered in Section 5.1 do not occur across the board, that is, in all contexts of use. On the contrary, they seem to be restricted to particular semantic functions.

5.5 CHAPTER SUMMARY AND DISCUSSION

This chapter has been devoted to the analysis of the synonym set object of study from both semasiological and onomasiological perspectives. Both approaches are important to account for the semantics of the individual adjectives, the relation among them, and the concept they designate, given that they provide different viewpoints of the data. Despite the fact that the semantic relation studied in the present dissertation is that of synonymy and, therefore, the main focus is on the onomasiological structure of the set, considering the semasiological structure of each adjective enabled us to conduct a more holistic analysis and thus explain some of the diachronic tendencies uncovered.

Section 5.1 proved the existence of several changes over time in the frequency of the five adjectives under analysis. Put briefly, *fragrant* and *sweet-scented* decreased considerably

[91] Cf. Section 5.1 for diachronic changes in the frequency of the adjectives, Section 5.2 for those affecting the concept PLEASANT SMELLING as a whole, and Section 5.3 for those related to the semasiological structure of the five near-synonyms.

throughout the period 1810–2009, whereas *scented* and *sweet-smelling* increased. Interestingly, the increase of the latter two adjectives took place despite the overall downward tendency in the frequency of the concept PLEASANT SMELLING as a whole. Some potential explanations for these respective increases and decreases were identified in the semasiological study conducted in Section 5.3. A continuum of senses emerged from this analysis, with *fragrant* and *sweet-scented* being prototypically 'natural', thus occupying the natural end of the continuum, and *perfumed* being prototypically 'artificial'. *Scented* and *sweet-smelling*, on the other hand, occupied intermediate positions on this cline, as they occurred relatively frequently in both the natural and artificial senses. It therefore became evident that there exists a clear correspondence between the fluctuations in frequency over time and the continuum of senses postulated: the two adjectives which decreased in frequency are both prototypically 'natural', whereas the two that increased are relatively salient in both the artificial and the natural senses. On the contrary, the frequency of the prototypically 'artificial' adjective, namely *perfumed*, remained relatively stable diachronically.

Concerning changes in the semasiological structure of the near-synonyms, a common pattern was identified for all the adjectives except *sweet-scented*. Interestingly, despite their different degree of prototypicality in the various senses and semantic categories, *fragrant*, *perfumed*, *scented*, and *sweet-smelling* all became increasingly more frequent to denote artificial aromas across periods, while they became less and less common in the natural sense, although to different degrees. This shared trend seems to point to a gradual convergence of these four adjectives concerning their semantic functions, whereby they become more similar rather than dissimilar over time.

Additionally, the shared diachronic development of the four near-synonyms in this respect also becomes evident when examining the concept PLEASANT SMELLING as a whole (cf. Section 5.2). In fact, it was postulated that this general tendency to denote more and more artificial aromas may have been motivated by external socio-cultural and technological forces, namely the transformation of American society as a result of the First and Second Industrial Revolutions. It was argued in Section 5.2.3 that the growing manufacturing of goods with synthetic aromas due to technological advances and the increased availability of such products, together with the changing habits and a higher degree of exposure to these products through media and advertising, led to changes at the conceptual level. In particular, goods with artificial, as opposed to natural, smells became more central in people's daily lives and so the concept

PLEASANT SMELLING, as well as the adjectives denoting it, came to be more commonly used to describe such aromas. This hypothesis found preliminary support in a study of a series of lexical indicators from three semantic domains connected to processes of modernization. However, the conclusions are yet tentative and further research is necessary to confirm or refute this hypothesis. One way in which this could be achieved would be to examine the diachronic development of the counterparts of the terms making up the concept PLEASANT SMELLING in other geographical varieties of English (e.g. BrE or Australian English) or in languages spoken in societies which have undergone similar socio-cultural and technological transformations. In fact, similar synonym sets can be found in other languages, for instance, Spanish *perfumado*, *fragante*, *bienoliente*, and *odorífero*. If such synonym sets were to follow the same or similar developmental patterns uncovered in this chapter for AmE *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*, this would yield further support to the hypothesis here.

Section 5.4 focused on the onomasiological structure of the synonyms, thus comparing their frequency in different contexts of use. This allowed us to zoom in on the competition between the five adjectives and to identify the exact semantic functions in which the frequency fluctuations took place. It was found that the increase of *scented* affected mostly the artificial sense and the semantic categories associated with it, whereas that of *sweet-smelling* took place mainly in the value 'indeterminate', particularly in the category BODY AND PEOPLE. On the contrary, *fragrant* decreased especially in the figurative sense and in indeterminate uses, but also to a certain extent in the natural sense. Therefore, not all contexts of use are equally affected by the general fluctuations in frequency uncovered in Section 5.1, as they seem to be limited to particular senses and to particular semantic categories. In these contexts of use in which changes are identified, a rearrangement is often found in the onomasiological salience of the adjectives, so that in some of the senses and semantic categories the originally predominant adjective is ousted by another member of the set. This is, for instance, the case of the artificial sense and of some of the prototypically 'artificial' categories, to wit, CLEANING, COSMETICS, and OBJECT. On other occasions, the end result is a more equal distribution of the adjectives by P4, in such a way that none of the adjectives are clearly dominant, as happens in the figurative sense and in the value 'indeterminate' and the semantic categories ABSTRACT, BODY AND PEOPLE, and TEXTILE AND CLOTHING.

By taking into account the results of the different types of analyses conducted in the present chapter, we can characterize the semantic development of the synonyms object of study as one

of attraction and ongoing substitution, at least in some contexts of use. As mentioned in Section 3.3, most studies on near-synonymy have resorted to a simplified theory of competition in which languages aim at an isomorphic state, mainly through the processes of differentiation and substitution. However, De Smet et al. (2018) argue that situations can be found in which near-synonymous expressions become over time more similar, rather than dissimilar, thus coming to share more semantic ground, a process they call *attraction*. The results presented throughout this chapter seem to point to the conclusion that at least four of the adjectives have undergone, and may still be undergoing, such a process of attraction. This was made evident in Section 5.3, in which it was shown that *fragrant*, *perfumed*, *scented*, and *sweet-smelling* follow a similar developmental path in semasiological terms, becoming more 'artificial' and less 'natural' across time. Moreover, this process of attraction seems to be accompanied by an ongoing process of substitution, whereby *scented* increases at the expense of both *perfumed* and *fragrant*, having already surpassed the former in terms of its overall frequency by P4 (cf. Section 5.1), and both *perfumed* and *fragrant* in particular semantic functions (cf. Section 5.4). It is not entirely surprising to find the processes of attraction and substitution operating at the same time, as they are not mutually exclusive: while the former is concerned with the functions of a set of competing expressions, the latter relates to their relative frequencies (De Smet et al. 2018: 205). In fact, a certain degree of attraction is probably a prerequisite for substitution, given that synonyms may need to be sufficiently functionally similar for one to be replaced by another. The interrelation between these two processes is further explored in the next chapter, where the onomasiological tendencies identified here are statistically tested by means of multivariate techniques and where further contexts of use are examined beyond the semantic ones considered up to now.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

6 IN-DEPTH ONOMASIOLOGICAL ANALYSIS OF THE SYNONYM SET: A MULTIVARIATE APPROACH

Even though the five near-synonymous adjectives *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling* seem to be mostly interchangeable, their distribution, as shown in Chapter 5, depends largely on the sense in which they occur and the type of nouns with which they collocate. However, the evidence provided in the previous two chapters can be further refined by examining the competition between the adjectives in greater depth and with more sophisticated statistical techniques. Therefore, the focus of the present chapter is, contrary to Chapter 5, only on the onomasiological structure of the synonym set in order to discover the contexts of use in which speakers favor each adjective at the expense of the others and whether these preferences have changed over time. In particular, the preliminary findings of the onomasiological analysis in Section 5.4 are here tested for significance, and a larger number of variables, including further semantic but also morphosyntactic and extralinguistic ones, are examined. It was shown in Section 3.2.2 that, although semantic variables, specifically those pertaining to noun collocates, are crucial to uncover fine-grained distinctions in meaning between adjectival near-synonyms, variables of a morphosyntactic nature often play a role as well. Moreover, the stylistic dimension is also considered here by taking into account the text-types in which the adjectives occur. Special attention is paid in this chapter to the diachronic tendencies of the adjectives by ascertaining whether their probability of occurrence in particular functions and contexts of use, be they semantic, morphosyntactic, or stylistic, varies over time. The results will allow us to explore the processes at work in the diachronic evolution of the near-synonym set and thus confirm or refute the preliminary conclusion drawn in Chapter 5 that the adjectives are undergoing processes of attraction and substitution.

The present chapter is divided into four sections.[92] Section 6.1 briefly summarizes the structure of the dataset employed and provides an explanation of the statistical methods applied

[92] The results presented in this chapter are partly based on those in Pettersson-Traba (2020a; 2020b).

to the data. Then, in Section 6.2, the results of the statistical analyses are presented: while Section 6.2.1 focuses on the effects of the variables Concreteness, Degree, Countability, Syntactic function, Text-type, Sense, Semantic category, and Period, Section 6.2.2. zooms in on the idiosyncratic collocational patterns of the adjectives. The results of the different analyses are jointly discussed in Section 6.3, and a summary of the entire chapter is offered in Section 6.4.

6.1 DATASET AND METHODOLOGY

As in the case of Chapter 5, the dataset used for the analyses in the present chapter is the first one explained in Section 4.2. Apart from the two semantic variables Sense and Semantic category, the instances of the adjectives in this dataset were also annotated for a series of additional language-internal variables, both semantic and non-semantic, and two language-external factors. These are Animacy, Concreteness (rating and binary), Countability, Syntactic function, Degree, Collocate, Period, and Text-type.[93] For reasons further explained below, only three of the adjectives were included in the analyses, namely *fragrant*, *perfumed*, and *scented*. On the other hand, *sweet-scented* and *sweet-smelling* had to be excluded due to their low frequency of occurrence, which makes it difficult to extract conclusions regarding their diachronic tendencies.[94]

In order to account for the independent effects of each variable on the probabilities of the three adjectives, while at the same time controlling for the effects of the other variables, multivariate statistical tests were applied. First, the data was analyzed by means of multinomial logistic regression and, second, by means of binary mixed-effects logistic regression. Logistic regression analysis is a statistical technique that aims at predicting the probability of occurrence of the levels of a dependent or response variable on the basis of one or more independent variables, also called predictors (Speelman 2014: 488–489; Levshina 2015: 253–254). In turn, multinomial regression serves to model situations in which the response variable comprises more than two levels (Levshina 2015: 277), here the three near-synonyms *fragrant*, *perfumed*, and *scented*. Although several approaches exist to conduct this test, the `polytomous` package (Arppe 2013) in *R* was selected for the present analyses, given that it provides several goodness-

[93] Cf. Table 8 in Section 4.4 for a summary of these variables and their values.

[94] The distribution of these two adjectives in comparison with the rest as regards the variables Animacy, Concreteness rating, Countability, Syntactic function, Degree, and Text-type is provided in Tables B.1–B.6 in Appendix B.

of-fit statistics of the resulting model and easily interpretable coefficients. The heuristic used in this package is the so-called *one vs. rest* approach, in which the probabilities of each level of the dependent variable are compared with those of the other levels together (Levshina 2015: 277). Here, this implies comparing first the probability of *fragrant* vs. the joint probability of *perfumed* and *scented*, then that of *perfumed* vs. *fragrant* and *scented*, and finally that of *scented* vs. *fragrant* and *perfumed*. Two separate models were computed, with Synonym functioning as dependent variable and Sense/Semantic category, Countability, Concreteness rating, Syntactic function, Degree, Period, and Text-type as predictors, as well as the interactions between Period and the other variables, given that the main focus is on the diachronic development of the adjectives. An interaction between two variables entails that the effect of one of them on the response variable depends on the other (Levshina 2015: 162). For instance, a significant interaction between Period and Degree would mean that the influence of the degree of the adjectives (i.e. positive, comparative, or superlative) on their probability of occurrence would vary across the four periods distinguished. Model A included Sense, not Semantic category, as a predictor of Synonym, while Model B included Semantic category, not Sense. Two separate models were computed because, as previously mentioned in Section 4.3.1.2, Sense and Semantic category are strongly correlated, since there is often a one-to-one correspondence between their values (e.g. natural and PLANTS AND FLOWERS). Correlated predictors are problematic if added to the same model and, consequently, these predictors had to be kept apart (Levshina 2015: 272). Moreover, Animacy was not included in any of the models, since it is also too strongly correlated with Semantic category.[95] As shown in Table 14, the vast majority of animate referents occur in the semantic category BODY AND PEOPLE (88.54%), with only a few occurring in EARTH, ATMOSPHERE, AND WEATHER (9.48%) and PLANTS AND FLOWERS (1.58%), which correspond to highly poetic or figurative texts where natural phenomena such as weather terms and nouns denoting plants and flowers are personified (cf. example (84) in Section 4.3.1.3 above).

[95] Nevertheless, to shed some light on the distribution of the adjectival near-synonyms across animate and inanimate referents, the proportions of each adjective with respect to this variable are given in Table B.1 in Appendix B.

Table 14: Distribution of animate and inanimate referents per Semantic category

Animacy / Sem. category		Animate	Inanimate
ABS	**N**	1	191
	%	0.40	3.78
B&P	**N**	224	259
	%	88.54	5.12
CL	**N**	–	161
	%		3.18
COS	**N**	–	118
	%		2.33
EAW	**N**	24	1040
	%	9.48	20.55
F&D	**N**	–	486
	%		9.60
OBJ	**N**	–	242
	%		4.78
P&F	**N**	4	1450
	%	1.58	28.66
S&M	**N**	–	336
	%		6.64
SEN	**N**	–	245
	%		4.85
SPA	**N**	–	317
	%		6.26
T&C	**N**	–	215
	%		4.25
Total	**N**	253	5,060
	%	100	100

Given that concreteness was coded by means of two separate variables, one binary and one numerical (i.e. Concreteness binary and Concreteness rating, respectively; cf. Section 4.3.1.3), only one of them was finally included in the analyses, as they measure the same semantic feature in two different ways. By computing two separate models, each one with a different variable

measuring concreteness, it was possible to identify which one was a better predictor of the alternation between the near-synonyms. A model containing Concreteness rating and the other predictors performed significantly better ($p < 0.001$) than a model with Concreteness binary. Therefore, the former variable was selected for inclusion in the final models.

To determine which predictors made a significant contribution to the models, a procedure known as *stepwise forward selection* was performed. This process consists in first computing a model without any predictors and then adding variables, one by one, until the addition of new variables does not lead to a significant improvement of the model (Levshina 2015: 149). First, the significance of each individual variable was ascertained and then the interactions between Period and the other predictors were tested for.

In order to interpret the results of the multinomial logistic regression models, the effects of the predictors were visualized by means of partial effect plots, which were carried out using the `ggplot2` package (Wickham 2016). Partial effect plots show the effects of each individual variable when the other variables remain constant.

The effect of the variable Collocate was measured by means of a different type of regression analysis, namely binary mixed-effects logistic regression. Mixed-effects models permit the inclusion of both fixed and random effects (Baayen 2008: 241–242; Tagliamonte & Baayen 2012: 143). The difference between these two kinds of effects lies in their repeatability. In the case of fixed effects, the levels of a predictor are repeatable, as happens with variables such as Sense: if we extracted a new sample of data of the adjectives under analysis from a different corpus, we would still annotate the resulting occurrences according to the values 'artificial', 'figurative', 'indeterminate', or 'natural'. On the contrary, random effects are open-ended, since there is not a limited number of values for such variables. This happens with variables such as Collocate, which only includes a subset of nouns out of all the potential nouns that could be modified by *fragrant*, *perfumed*, and/or *scented*. Therefore, if we extracted a new sample of the near-synonyms from a corpus other than *COHA*, it would inevitably contain individual noun collocates which do not appear in the current dataset. By including Collocate as a random effect, it is possible to identify particular lexical items that show a preference for one of the near-synonymous adjectives, thus accounting for their idiosyncratic collocational preferences. The same heuristic as in the multinomial models was used here, namely comparing the probabilities of each adjective against those of the other two combined (e.g. *fragrant* vs.

perfumed and *scented*). Mixed-effects models were computed by using the `glmer` function in the `lme4` package (Bates et al. 2015).

All the resulting models are relatively complex, given that they include a large number of regressors. In logistic regression, each of the levels of a predictor counts as one regressor, excluding the reference value (Speelman 2014: 504).[96] To illustrate this point, consider the predictor Countability. This variable contains three levels, to wit, count, non-count, and other, which translates into two regressors. The most complex model, that is multinomial Model B, i.e. the one including Semantic category, contains 64 regressors. As a common rule of thumb, the number of instances of the least frequent level of the dependent variable should be at least ten to fifteen times larger than the total number of regressors (Baayen 2008: 195; Levshina 2015: 143–144). In the present dataset, the total number of instances of *sweet-scented* and *sweet-smelling*, the two least frequent near-synonyms, is 198 and 253, respectively (cf. Table 9 in Section 5.1). This implies that the maximum number of regressors allowed in a model with these two adjectives ranges between 13 and 19 in the case of *sweet-scented*, and between 16 and 25 in the case of *sweet-smelling*. These figures are much lower than the actual number of regressors in the models, which is the main reason why *sweet-scented* and *sweet-smelling* were excluded from the present analyses. Following the exclusion of these two near-synonyms, *scented* is the least frequent adjective, amounting to 941 total occurrences in *COHA*, which permits the inclusion of 62 to 94 regressors. The 64 regressors of the most complex model are thus within the limits.

Table 15 provides the overall frequency of the reduced dataset with the three near-synonyms *fragrant*, *perfumed*, and *scented*, as well as their distribution in absolute and relative frequencies, both generally and over time.[97] The percentages present the relative frequencies of

[96] Logistic regression modeling calculates the effects of the predictors on the dependent variable by determining whether there is an increase or a decrease in the probability of each level of the dependent variable when a predictor changes from one variable, called the reference level, to another one. For instance, in order to compute the effect of Degree, the test identifies whether there is an increase or a decrease in the likelihood of *fragrant* (in comparison with *perfumed* and *scented*) if we switch from the positive degree, i.e. the reference level, to the comparative or the superlative degrees. Reference levels are set by the analyst, but can be changed to allow for different contrasts between the levels of predictor variables. In a model, reference levels are bundled together into only one regressor, which is called the intercept.

[97] The absolute frequencies of the three adjectives are identical to those given in Table 9 in Section 5.1, but they are provided here again for the sake of convenience.

the near-synonyms in comparison to one another and are hence to be interpreted vertically, since they total 100% in each period.

Table 15. Absolute and relative frequencies of *fragrant*, *perfumed*, and *scented*

Period / Synonym	P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
Fragrant					
N	804	1291	712	588	3,395
%	72.69	69.45	57.65	52.83	63.90
Perfumed					
N	193	333	227	224	977
%	17.45	17.91	18.38	20.13	18.39
Scented					
N	109	235	296	301	941
%	9.86	12.64	23.97	27.04	17.71
Total					
N	1,106	1,859	1,235	1,113	5,313
%	100	100	100	100	100

Finally, although logistic regression models can point to the relative importance of predictors to account for the choice between the levels of a dependent variable, here *fragrant*, *perfumed*, and *scented*, another statistical test is more commonly used for this purpose, namely random forest analysis (Tagliamonte & Baayen 2012: 158–161; Levshina 2015: 297–299). This statistical test functions by computing many so-called conditional inference trees. Conditional inference trees, in turn, predict the probability of the values of a dependent variable by means of a series of binary splits in the data. The algorithm takes the predictor that is more strongly correlated with the dependent variable and considers whether dividing the data in two creates two new subsets of cases in which one of the values of the dependent variable is more probable than the rest. This process is then repeated until no further significant splits are identified (Tagliamonte & Baayen 2012: 159; Levshina 2015: 291–297). The individual trees in the forest are constructed on the basis of random samples of both predictors and instances of the dataset. Next, random forest analysis calculates the relative importance of predictors by arbitrarily changing the values of a given predictor. In this way, any discriminatory capacity that a particular predictor may have disappears. The test then compares the explanatory power of the original predictor with its reordered counterpart: if their explanatory power is the same, this means that the predictor does not have a significant effect on the choice between the levels of the dependent variable. On the contrary, if the explanatory power of the reordered predictor is

worse than that of the original one, this implies that the latter has indeed a significant effect. The relative importance of a predictor is then calculated as the extent to which the model performs worse with the reordered version as opposed to the original one. Put simply, the worse the model becomes, the more important the effect of the predictor. Table 16 below exemplifies this process with a hypothetical scenario with the dependent variable Synonym and the predictor Degree.

Table 16. Hypothetical scenario of reordering of levels in random forest analysis

Synonym	Degree (original)	Degree (reordered)
fragrant	positive	positive
scented	comparative	superlative
perfumed	superlative	positive
perfumed	superlative	comparative
scented	comparative	positive
fragrant	positive	superlative
fragrant	positive	comparative
scented	comparative	comparative
fragrant	positive	comparative

In this hypothetical scenario, *fragrant* always occurs in the positive degree, *perfumed* in the superlative degree, and *scented* in the comparative degree (cf. second column in Table 16). However, when we arbitrarily reorder the levels of the predictor, the correlation between Synonym and Degree disappears, as none of the three adjectives occur exclusively or even mainly in a particular degree (cf. third column in Table 16). In the made-up example shown in the table, Degree would, consequently, exhibit a fairly high relative importance. The output of random forest analysis consists in a set of variable importance values which are usually interpreted as a ranking, where the predictor with the highest value is the most important and so on. Random forests were computed in *R* using the function `cforest()` in the `party` package (Hothorn et al. 2006; Strobl et al. 2007; Strobl et al. 2008).

6.2 RESULTS

This section deals with the results obtained in the different regression models. While Section 6.2.1 presents the findings of the multinomial regression analyses, focusing on the fixed effect

structures, Section 6.2.2 zooms in on the random effects of the variable Collocate, thus accounting for the outcome of the mixed-effects regression analyses.

6.2.1 Multinomial models: Fixed effect structure

As mentioned in the previous section, two multinomial regression models had to be computed separately in order to take the two variables Sense and Semantic category into consideration, as it is not possible to include both of them in the same model. The results of the two models, named 'Model A' and 'Model B', are here presented separately. However, given that besides the two highly correlated predictors Sense and Semantic category, the two models contain exactly the same significant predictors and predictor interactions, their effects are reported only once, as they are practically identical in Model A and Model B.

Starting with Model A, the following predictors and predictor interactions emerged as significant: Concreteness rating, Countability, Syntactic function, Degree, and the interactions between Period and Sense and between Period and Text-type. Table 17 displays the summary statistics of Model A, which provides several goodness-of-fit statistics, namely accuracy, McFadden R^2, Nagelkerke R^2, and likelihood ratio test.

Table 17. Model summary of multinomial logistic regression Model A (Sense)

Model summary	
accuracy	67.34% (baseline = 63.90%; $p < 0.001$)
McFadden R^2	0.1341
Nagelkerke R^2	0.2575
likelihood ratio test	1300.9 ($p < 0.001$)

The first row shows the accuracy, a reflection of the percentage of correct predictions of the model. As explained above, logistic regression analysis works as follows: the algorithm calculates the probability of each of the values of the response variable, in this case *fragrant*, *perfumed*, and *scented*, on the basis of the independent variables. Consequently, the test predicts for each individual instance in the dataset whether it is most likely that one of the three adjectives occurs by using the information provided by the independent variables. In some cases, these predictions are correct, while in others they are not. For instance, the model may estimate on the basis of the predictors (i.e. their specific values) that a concrete example in the data should be realized as *perfumed*, when the actual observed near-synonym is either *fragrant* or *scented*. Therefore, it is possible to identify the amount of times that the predictions of the model are in line with the actual observed data or, in other words, the predictive capacity of the

model. In Model A, the accuracy value is 67.34%. The baseline accuracy is 63.90%, which is equal to the total amount of the most frequent value of the dependent variable, namely *fragrant* (cf. Table 15 in Section 6.1). The accuracy value of Model A is therefore significantly better than the baseline value ($p < 0.001$), which means that the predictors included entail an improvement over a null model (i.e. a model with no predictors).

The following two rows in Table 17 presents the McFadden's R^2 and Nagelkerke's R^2, both of which are indications of how well the model fits the actual observed data. R^2 is an additional index of predictive power which, put simply, translates into the proportion of variance that the fitted model explains.[98] As concerns McFadden R^2, values between 0.2 and 0.4 indicate a good fit of a model (Levshina, 2015: 280); therefore, the value obtained in Model A (i.e. 0.1341) signals that there is still some unexplained variance. The same holds for the obtained Nagelkerke R^2 index, the values of which range from zero, if a model has no predictive power, to 1, if a model achieves a perfect fit. In this case, the value 0.2575 points to a similar conclusion as the McFadden R^2. Finally, the last row displays the findings of the likelihood ratio test, which provide the overall significance of the model. As we can observe, this model is statistically significant ($p < 0.001$).

Continuing with Model B, the same predictors and predictor interactions are significant. The only difference concerns a significant interaction between Semantic category —as opposed to Sense— and Period. Table 18 displays the summary statistics of Model B, which provides the same goodness-of-fit statistics used in Table 17. As shown in the table, the indexes are similar to those of Model A. There is, however, an improvement in the predictive power of the model, given that the accuracy, McFadden R^2, and Nagelkerke R^2 values are all higher than in Model A. In fact, Model B performs significantly better than Model A ($p < 0.001$). This was to be expected, given that Semantic category provides a more fine-grained division than Sense and therefore identifies differences between the preferences of the near-synonyms which may be obscured if only the sense in which they are used is considered. A likelihood ratio test indicates that Model B is also statistically significant ($p < 0.001$).

[98] It is noteworthy to point out that this value is, when logistic regression is concerned, not as straightforward as in linear regression, where this index stems from. This is so because the value obtained in logistic regression is often lower than an equivalent model in linear regression (Speelman 2014: 515; Levshina 2015: 259).

Table 18. Model summary of multinomial logistic regression Model B (Semantic category)

Model summary	
accuracy	69.13% (baseline = 63.90%; $p < 0.001$)
McFadden R^2	0.1754
Nagelkerke R^2	0.3250
likelihood ratio test	1702.5 ($p < 0.001$)

As regards the importance of the predictors in order to discriminate between the three adjectival near-synonyms, the variables of a semantic nature (i.e. Sense, Semantic category, Concreteness rating, and Countability) generally have a stronger explanatory power than the language-internal and language-external factors. This is confirmed by computing regression models with only the semantic variables and their significant interactions with Period, i.e. between Sense and Period and between Semantic category and Period, respectively, and comparing the resulting indexes with those obtained for Models A and B. The models with semantic variables only are displayed in Tables 19 and 20:

Table 19. Model summary of multinomial logistic regression with only semantic predictors (Sense)

Model summary	
Accuracy	67.04 % (baseline = 63.90 %; $p < 0.001$)
McFadden R^2	0.1187
Nagelkerke R^2	0.2310
Likelihood Ratio test	1146.1 ($p < 0.001$)

Table 20. Model summary of multinomial logistic regression with only semantic predictors (Semantic category)

Model summary	
Accuracy	67.83 % (baseline = 63.90 %; $p < 0.001$)
McFadden R^2	0.1567
Nagelkerke R^2	0.2952
Likelihood Ratio test	1514.6 ($p < 0.001$)

As can be seen here, the accuracy, McFadden R^2, and Nagelkerke R^2 in the models including only semantic variables are not much lower than their counterparts in Models A and B, respectively. For instance, while the accuracies of Models A and B are 67.34% and 69.13% (cf. Tables 17 and 18 above), respectively, that of their counterparts with only semantic variables are 67.04% (Table 19) and 67.83% (Table 20).

The greater importance of semantic variables is further supported by the findings of a random forest analysis, with all the independent variables in Models A and B as predictors.[99] Figure 16 displays the variable importance values of the eight predictors included in the model: the predictors are located on the vertical axis and their importance values are displayed on the horizontal axis.[100] Therefore, the longer the line of a predictor, the greater its importance in the model. As can be observed, all the semantic variables exhibit a higher importance value than the other predictors, with the exception of Period, which is ranked third, following Semantic category and Sense. On the other hand, the three predictors that obtain the lowest rankings are the non-semantic variables, both language-internal (i.e. Syntactic function and Degree) and language-external (i.e. Text-type). This implies that the choice between the adjectives *fragrant*, *perfumed*, and *scented* is mostly determined by semantic features, especially the two highly correlated predictors Semantic category and Sense, but also by the extralinguistic factor Period. The fact that Period achieves such a high ranking already indicates that the synonym set object of study undergoes important changes over the time span analyzed.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[99] The fact that the results of the random forest analysis are computed on the basis of random samples of data points and predictors makes it relatively immune to problems caused by highly correlated predictors such as Sense and Semantic category. This is the reason why these two variables can be included in the same random forest model (Tagliamonte & Baayen 2012: 161).

[100] Although not the focus of the present discussion, it is worth mentioning here that the random forest model achieves a fairly high prediction accuracy of 73.52%, which is again statistically significantly higher than the baseline accuracy of 63.90% ($p < 0.001$).

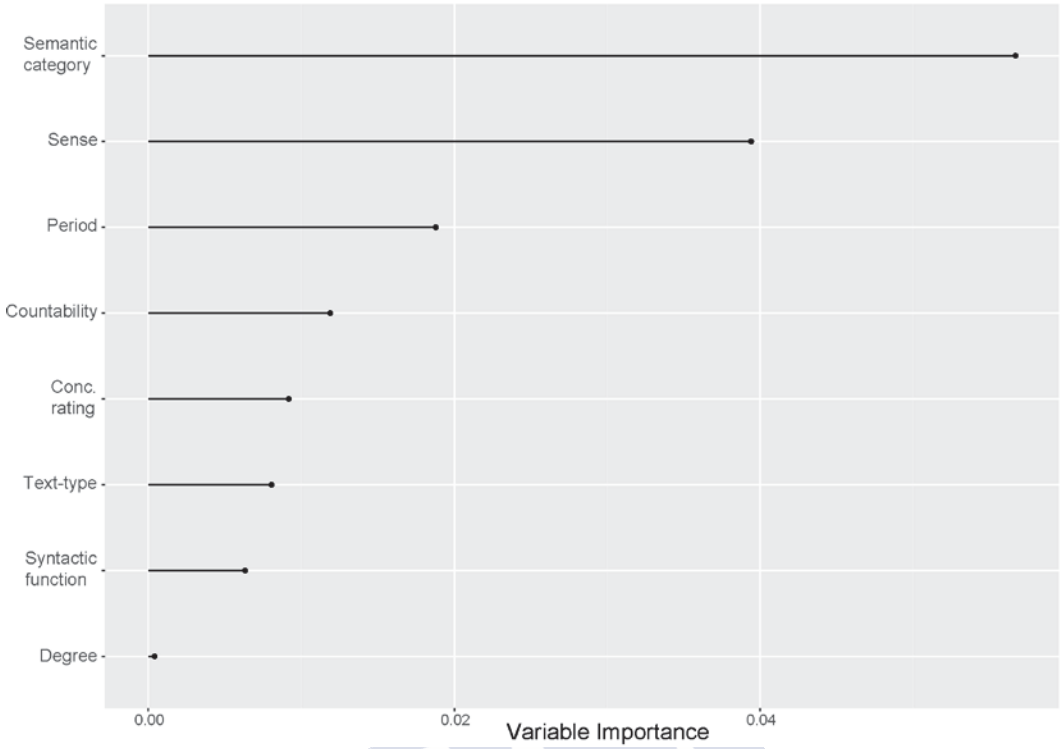


Figure 16. Variable importance of predictors according to random forest analysis

Turning now to the predictions of multinomial regression Models A and B, these are displayed in tabular form in Tables 21 and 22.[101] In both tables the first column displays the regressors of the model. The first row provides the values for the intercept which, as explained previously (cf. Section 6.1), represents the probability of each of the adjectives in the reference levels of each of the eight predictors, namely fiction (Text-type), positive (Degree), count (Countability), P2 (Period), natural (Sense), EARTH, ATMOSPHERE, AND WEATHER (Semantic category), and attributive (Syntactic function), as well as the mean value for the numerical variable Concreteness rating, which is 4.365. The rest of the regressors signal the probability of the adjectives when the level of a specific predictor changes, while the rest are kept at the reference level. To illustrate this point, the regressor called 'Degree = Comparative' (i.e. row five in Table 21) presents the probabilities of *fragrant*, *perfumed*, and *scented* when they occur in comparative, instead of positive degree and the other predictors are maintained at their reference level. The regressors which include more than one predictor (e.g. 'Period = P1: Sense = Artificial' in row twelve in Table 21) provide the values of interactions. However, given the

[101] Table 22 contains only the values of the predictors Semantic category and Period, and their interaction, given that the remaining independent variables behave very similarly in Models A and B.

difficulty of interpreting interactions in this format, they are visualized below by means of partial effect plots. The remaining columns in Tables 21 and 22 offer the coefficients for each adjective. The first column presents the estimates, that is, the increases or decreases in probability of the adjectives as compared to the reference level (i.e. intercept). The estimates are measured in log odds: whereas positive values point to an increase in the probability of the adjectives, the reverse holds in the case of negative values, and the higher the absolute number, the larger the respective increase/decrease. Finally, the columns entitled '*p*-value' show whether the decreases or increases in probability are statistically significant or not. As is customary in most statistical analyses, *p*-values that are lower than 0.05 are considered to be significant.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 21. Model A coefficients (Sense)

Predictor	*Fragrant*		*Perfumed*		*Scented*	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.2100	< 0.001	-1.3180	< 0.001	-2.7950	< 0.001
Concreteness rating	-0.0007	0.989	-0.1541	0.005	0.1676	0.007
Text-type = Non-fiction	0.3321	0.053	-0.3221	0.127	-0.1967	0.383
Text-type = Periodicals	0.8170	< 0.001	-0.6296	0.020	-0.7541	0.020
Degree = Comparative	1.5910	< 0.001	-1.1300	0.001	-1.4980	0.001
Degree = Superlative	1.8500	0.020	-1.6270	0.120	-1.3700	0.1870
Countability = Non-count	-0.5127	< 0.001	0.1224	0.158	0.5617	< 0.001
Countability = Other	-0.1891	0.224	0.3368	0.051	-0.1025	0.587
Period = P1	0.4424	< 0.001	-0.3775	0.022	-0.3662	0.026
Period = P1: Text-type = Non-fiction	-1.002	< 0.001	1.213	< 0.001	0.2020	0.573
Period = P1: Text-type = Periodicals	-1.1160	0.006	1.1580	0.015	0.5480	0.385
Period = P1: Sense = Artificial	-0.8259	0.001	0.7340	0.004	0.2148	0.480
Period = P1: Sense = Figurative	0.3195	0.530	-0.4040	0.466	-0.2460	0.835
Period = P1: Sense = Indeterminate	-0.6112	0.134	0.5325	0.225	0.4072	0.478
Period = P3	-0.0855	0.470	-0.3866	0.022	0.4261	0.003
Period = P3: Text-type = Non-fiction	-0.4307	0.158	0.2767	0.480	0.3283	0.348
Period = P3: Text-type = Periodicals	-0.8875	0.002	0.7093	0.038	0.7639	0.041
Period = P3: Sense = Artificial	-0.1585	0.439	-0.1217	0.598	0.3882	0.091
Period = P3: Sense = Figurative	-0.3738	0.367	0.3794	0.422	0.9582	0.190
Period = P3: Sense = Indeterminate	-0.4491	0.104	0.3481	0.271	0.4532	0.200
Period = P4	-0.0075	0.955	-0.3995	0.035	0.3218	0.041
Period = P4: Text-type = Non-fiction	-1.425	<0.001	0.5223	0.212	1.156	0.002
Period = P4: Text-type = Periodicals	-0.8974	0.001	0.3521	0.305	1.0610	0.004
Period = P4: Sense = Artificial	-0.5440	0.011	-0.1306	0.589	0.7349	0.002
Period = P4: Sense = Figurative	-1.300	0.006	1.472	0.003	1.0750	0.169
Period = P4: Sense = Indeterminate	-0.5292	0.065	0.8371	0.010	-0.0587	0.879
Sense = Artificial	-1.7010	<0.001	1.7770	<0.001	0.5268	0.002
Sense = Figurative	0.0030	0.991	0.4098	0.172	-1.0550	0.078
Sense = Indeterminate	-0.8125	<0.001	1.094	<0.001	0.06071	0.827
Syntactic function = Other	-0.5313	0.001	0.8028	<0.001	-0.1603	0.431
Syntactic function = Postpositive	-0.2697	0.018	0.0711	0.603	0.3016	0.020
Syntactic function = Predicative	0.5007	<0.001	-0.4043	0.003	-0.3108	0.020

Table 22. Model B coefficients (Semantic category)

Predictors	*Fragrant*		*Perfumed*		*Scented*	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.3220	<0.001	-1.3550	<0.001	-2.8840	<0.001
Semantic category = ABST	0.3649	0.240	-0.0772	0.819	-1.0540	0.085
Semantic category = B&P	-0.9288	<0.001	1.2160	<0.001	-0.3052	0.371
Semantic category = CL	-1.9090	<0.001	0.2655	0.561	1.7720	<0.001
Semantic category = COS	-0.6148	0.262	0.5143	0.395	0.3477	0.607
Semantic category = F&D	1.167	<0.001	-1.415	0.001	-0.6243	0.065
Semantic category = OBJ	-1.562	<0.001	1.2500	<0.001	0.6869	0.040
Semantic category = P&F	0.9180	<0.001	-1.2380	<0.001	-0.3555	0.074
Semantic category = S&M	0.4714	0.049	-0.1651	0.556	-0.6100	0.075
Semantic category = SEN	0.8692	0.005	-0.4004	0.237	-1.4830	0.015
Semantic category = SPA	-0.5804	0.014	0.6255	0.016	0.1585	0.617
Semantic category = T&C	-1.8420	<0.001	1.613	<0.001	0.4606	0.127
Period = P1	-0.0084	0.964	-0.2352	0.314	0.2636	0.267
Period = P1: Semantic category = ABST	0.7757	0.196	-0.2968	0.636	-12.6400	0.957
Period = P1: Semantic category = B&P	-0.5832	0.154	0.4720	0.249	0.2063	0.695
Period = P1: Semantic category = CL	-0.5052	0.580	2.189	0.003	-2.2600	0.004
Period = P1: Semantic category = COS	12.5600	0.938	-12.4700	0.963	-14.4100	0.984
Period = P1: Semantic category = F&D	0.1407	0.735	1.0270	0.057	-1.709	0.016
Period = P1: Semantic category = OBJ	0.7663	0.099	0.0528	0.911	-1.2360	0.044
Period = P1: Semantic category = P&F	1.2140	<0.001	-1.2140	0.005	-1.0230	0.004
Period = P1: Semantic category = S&M	-0.4355	0.252	0.7135	0.102	-0.1737	0.749
Period = P1: Semantic category = SEN	-0.3858	0.366	0.8269	0.078	-1.6260	0.174
Period = P1: Semantic category = SPA	-0.2325	0.603	-0.1122	0.827	0.4285	0.426
Period = P1: Semantic category = T&C	-0.5138	0.462	-0.0584	0.913	0.3359	0.555
Period = P3	-0.295	0.079	-0.2503	0.240	0.6639	0.001
Period = P3: Semantic category = ABST	-0.1811	0.693	0.3652	0.482	0.5014	0.513
Period = P3: Semantic category = B&P	-0.0255	0.934	0.0186	0.955	0.2588	0.530
Period = P3: Semantic category = CL	-0.0326	0.961	0.7334	0.200	-1.0500	0.047
Period = P3: Semantic category = COS	-0.1401	0.831	0.1206	0.867	-0.0368	0.962
Period = P3: Semantic category = F&D	0.5336	0.209	-0.1604	0.812	-0.7309	0.149
Period = P3: Semantic category = OBJ	0.2959	0.505	-1.1420	0.021	0.5481	0.228
Period = P3: Semantic category = P&F	(0.2027)	0.444	-0.0274	0.947	-0.3799	0.225
Period = P3: Semantic category = S&M	0.1860	0.599	-0.1066	0.812	-0.0751	0.871
Period = P3: Semantic category = SEN	0.3468	0.514	-0.5778	0.415	0.4744	0.561
Period = P3: Semantic category = SPA	0.3739	0.270	-0.5692	0.173	0.0445	0.915
Period = P3: Semantic category = T&C	-0.0473	0.915	-0.3842	0.331	0.3312	0.424

Table 22. Continued

Predictors	*Fragrant*		*Perfumed*		*Scented*	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Period = P4	-0.0309	0.878	-0.5064	0.066	0.4868	0.041
Period = P4: Semantic category = ABST	-1.1710	0.030	1.588	0.006	0.6324	0.455
Period = P4: Semantic category = B&P	-0.3808	0.253	0.4568	0.225	0.3217	0.462
Period = P4: Semantic category = CL	0.2573	0.669	0.3127	0.598	-0.6455	0.212
Period = P4: Semantic category = COS	-1.252	0.063	0.9308	0.187	0.2124	0.778
Period = P4: Semantic category = F&D	0.1191	0.763	0.9094	0.109	-0.8048	0.096
Period = P4: Semantic category = OBJ	-0.5303	0.232	-0.5626	0.221	0.9173	0.038
Period = P4: Semantic category = P&F	-0.0022	0.994	-0.2598	0.622	-0.0855	0.799
Period = P4: Semantic category = S&M	-0.8290	0.031	0.5555	0.255	0.7626	0.106
Period = P4: Semantic category = SEN	0.0541	0.926	0.3629	0.593	0.0370	0.970
Period = P4: Semantic category = SPA	-0.2695	0.462	0.2501	0.569	0.2234	0.612
Period = P4: Semantic category = T&C	0.4187	0.386	0.0982	0.834	-0.4794	0.340

The remainder of this section focuses on the effects of each of the predictors in Models A and B, starting with the main effects and then moving to the interactions of Period and Text-type, Period and Sense, and Period and Semantic category. As mentioned previously, the effects are visualized by means of partial effect plots, which are interpreted as follows: the levels of the predictors are shown on the horizontal axis and the probability of the adjectives in each level on the vertical axis. Each adjective is represented with a different shape, or line type in the case of the numerical variable Concreteness rating, as indicated in the legend to the right of the plots. Within the plots, whiskers, or a shaded grey area in the case of Concreteness rating, show the 95% confidence intervals. Confidence intervals constitute the range of values within which we can be 95% sure that the true values of the regressors are located (Levshina 2015: 98).

Figure 17 displays the effect of Concreteness rating on the choice between the near-synonyms. As is clearly observed in the figure, for *fragrant* no significant effect emerges, which means that it is more or less equally probable regardless of the degree of concreteness of the nominal elements it modifies, with all values ranging between 66% and 67%. Despite the lack of effect, *fragrant* is by far the most prominent adjective in all cases, which is not surprising, given its high overall frequency (63.90% of the total number of data points). On the contrary, both *perfumed* and *scented* are significantly influenced by the feature Concreteness. The former adjective is less likely as the rating increases, that is, the more concrete the modified nominal

element, the lower the likelihood of encountering *perfumed*. On the other hand, *scented* exhibits the opposite trend, becoming more likely as the concreteness rating increases. However, for none of the two adjectives is the effect extremely pronounced: the probabilities of *perfumed* and *scented* are 23.46% and 10.03%, respectively, when they modify nouns with a rating of 1, and 15.21% and 17.52%, when they modify nouns with a rating of 5. This entails a decrease in probability of 8.25% in the case of *perfumed* and an increase of 7.49% in the case of *scented*. Moreover, the probabilities of the two adjectives are very similar when they collocate with highly concrete nouns, that is, those with a rating ranging between 4 and 5 (e.g. *lady*, *meadow*, and *tea*). This means that the main difference between the adjectives *perfumed* and *scented* lies in their frequency with more abstract nouns, especially those with a rating between 1 and 2 (e.g. *knowledge*, *peace*, and *regret*), in which case *perfumed* is more likely.

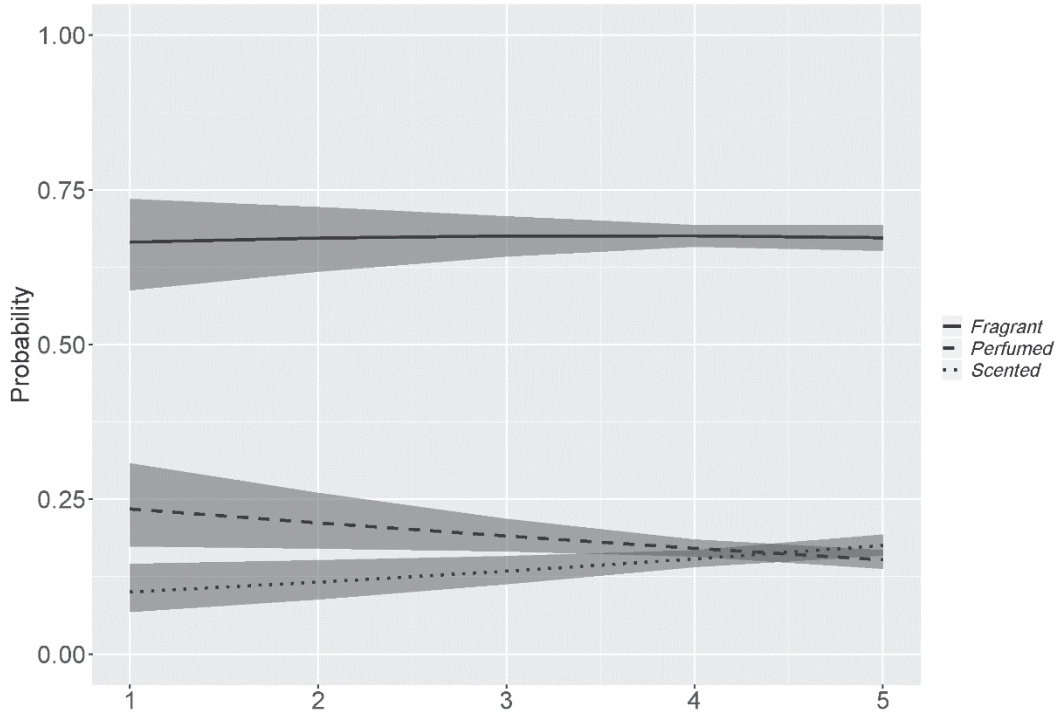


Figure 17. Effect of Concreteness rating on the probabilities of the adjectives *fragrant*, *perfumed*, and *scented*

Continuing with the effect of Degree, the probabilities of the three adjectives are plotted in Figure 18. With respect to this variable, *fragrant* is again the dominant variant in all three levels (positive, comparative, and superlative), but its dominance is more pronounced in the comparative and superlative degrees, with a probability of 90.84% and 92.37%, respectively.

In fact, the high probability of *fragrant* in these two values, as compared to that for the positive degree, is statistically significant. In turn, both *perfumed* and *scented* are significantly less likely in the comparative and the superlative degrees, and more probable in the positive degree. As can be deduced from the large confidence intervals —represented by whiskers— in this graph for the superlative degree, the number of instances corresponding to this level is very low, with a total of 22 instances. Therefore, the results for the superlative degree should be taken with caution.

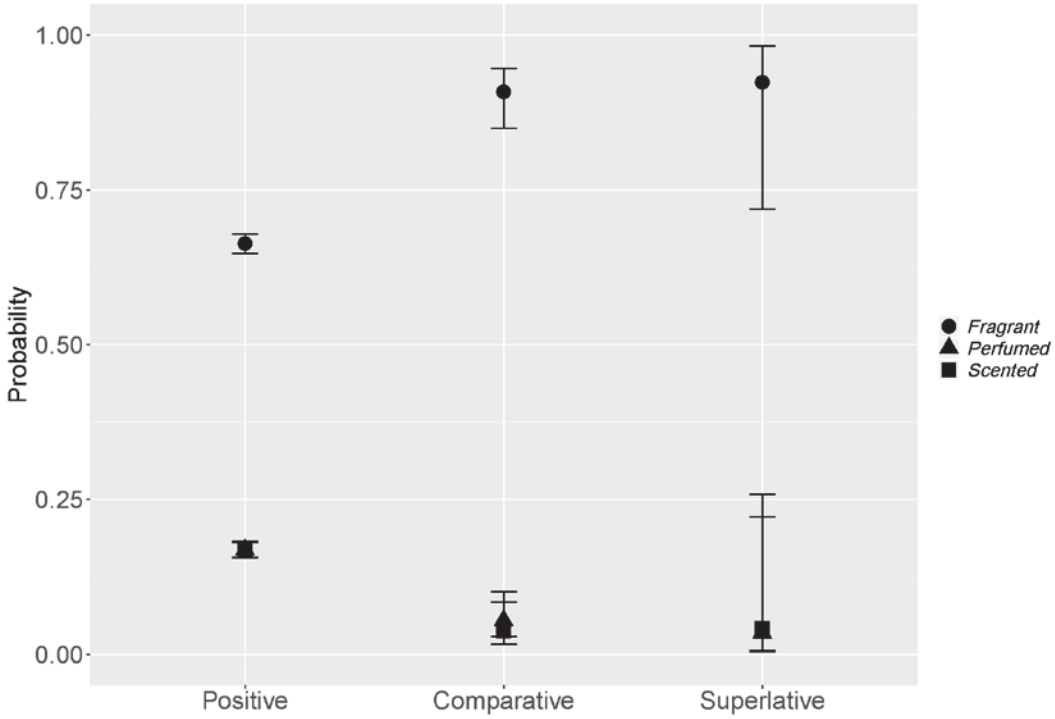


Figure 18. Effect of Degree on the probabilities of the adjectives *fragrant*, *perfumed*, and *scented*

In the case of Countability (cf. Figure 19), *fragrant* is once more the most likely choice with the three values (count, non-count, and other). However, it is significantly less likely with non-count nouns (58.79%) than with count nouns (70.46%). On the other hand, *scented* exhibits the opposite effect, being significantly more likely with non-count (23.07%) than with count nouns (14.18%). Finally, the variable Countability does not have a significant effect on the probability of *perfumed*, but as can be seen in Table 21 (cf. row 8), this adjective is almost significantly more likely in the value 'other' (i.e. with pronouns and proper nouns), with a likelihood of 20.01%, than with count nouns, with which it shows a likelihood of only 15.36% ($p = 0.051$).

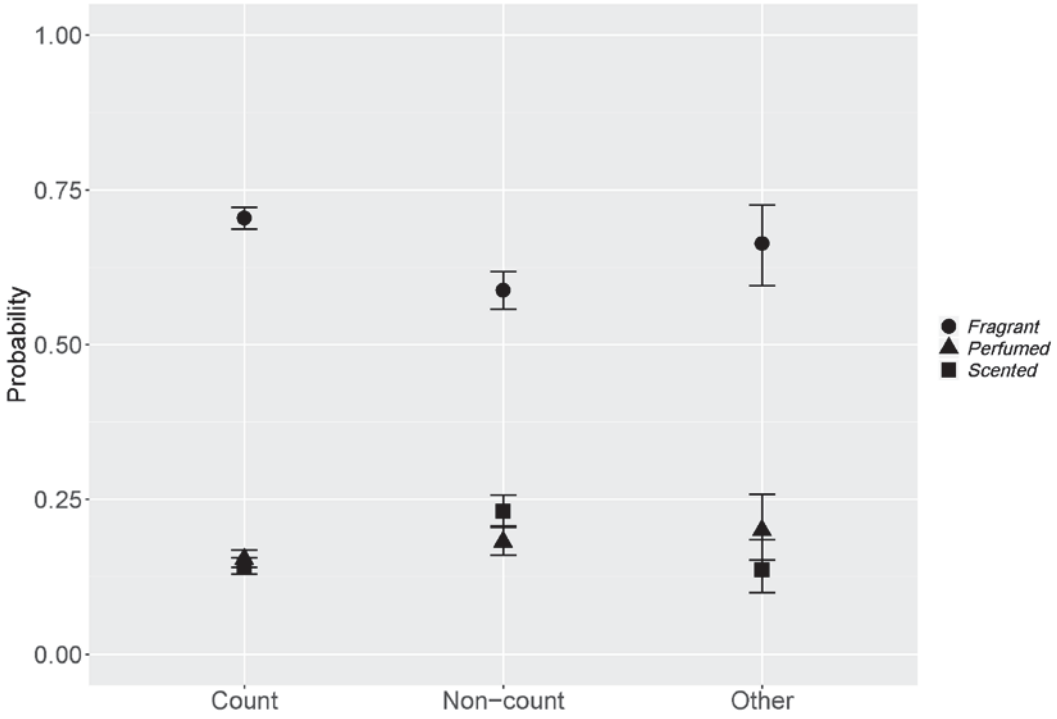


Figure 19. Effect of Countability on the probabilities of the adjectives *fragrant*, *perfumed*, and *scented*

The last main effect is that of Syntactic function, which is visualized in Figure 20. The probabilities of the adjectives with respect to this variable follow the general tendency, with *fragrant* representing again the prevailing option in all levels. Nonetheless, some significant differences emerge. First, *fragrant* is less likely in postpositive and other minor functions (i.e. predicative adjunct, fused modifier-head, and predeterminer; cf. examples (101)–(103) in Section 4.3.2) than in attributive function, while it occurs more often as a predicative complement. Second, *perfumed* is more likely in the category 'other' if compared to attributive function, but less likely as a predicative complement. Similarly, *scented* is also less likely in predicative function, whereas it is more probable as a postpositive modifier. The fairly large confidence intervals of the adjectives in the value 'other' are due to the relatively low number of examples corresponding to the minor functions.[102]

[102] As mentioned in Section 4.3.2, this was the reason why these different syntactic functions where bundled together in the same level labeled 'other'.

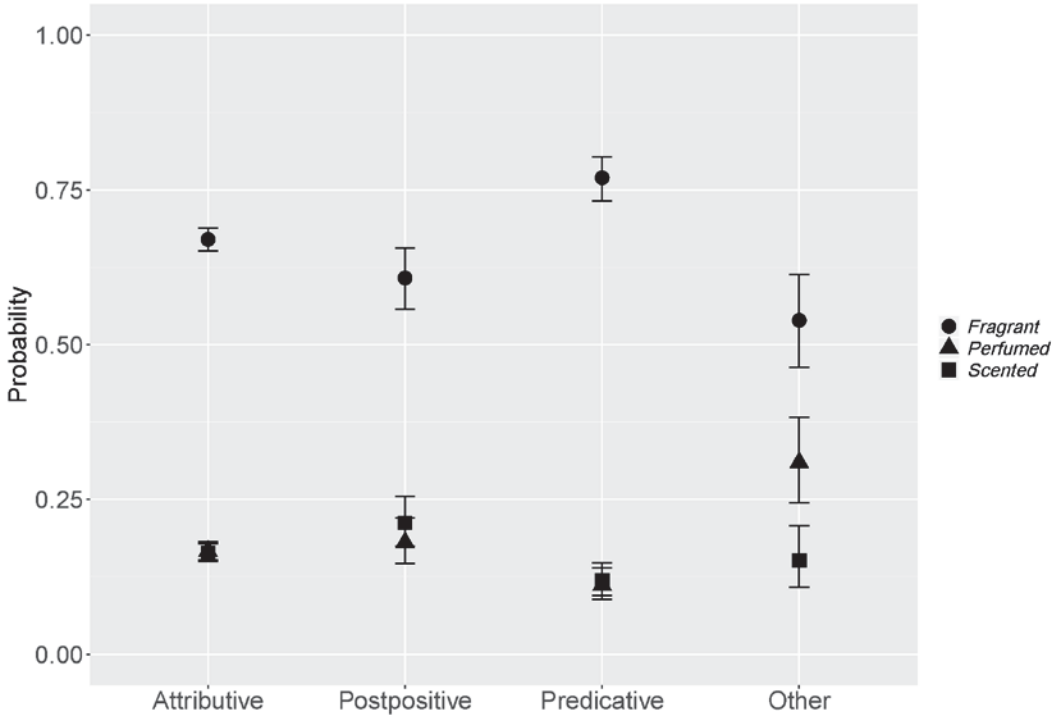


Figure 20. Effect of Syntactic function on the probabilities of the adjectives *fragrant*, *perfumed*, and *scented*

Turning now to the interactions, Figure 21 plots the effects of Text-type and Period in three different graphs, one per level of the variable Text-type. In the case of interactions between predictors, the focus is on the competition between the adjectives over time within each of the levels of the variables interacting with Period, to wit, Text-type (Figure 21), Sense (Figure 22), and Semantic category (Figure 23). To examine some of these additional contrasts, the reference levels of Period and the variables with which it interacts have been altered (cf. Tables B.7–B.25 in Appendix B for the relevant coefficients of these supplementary models).

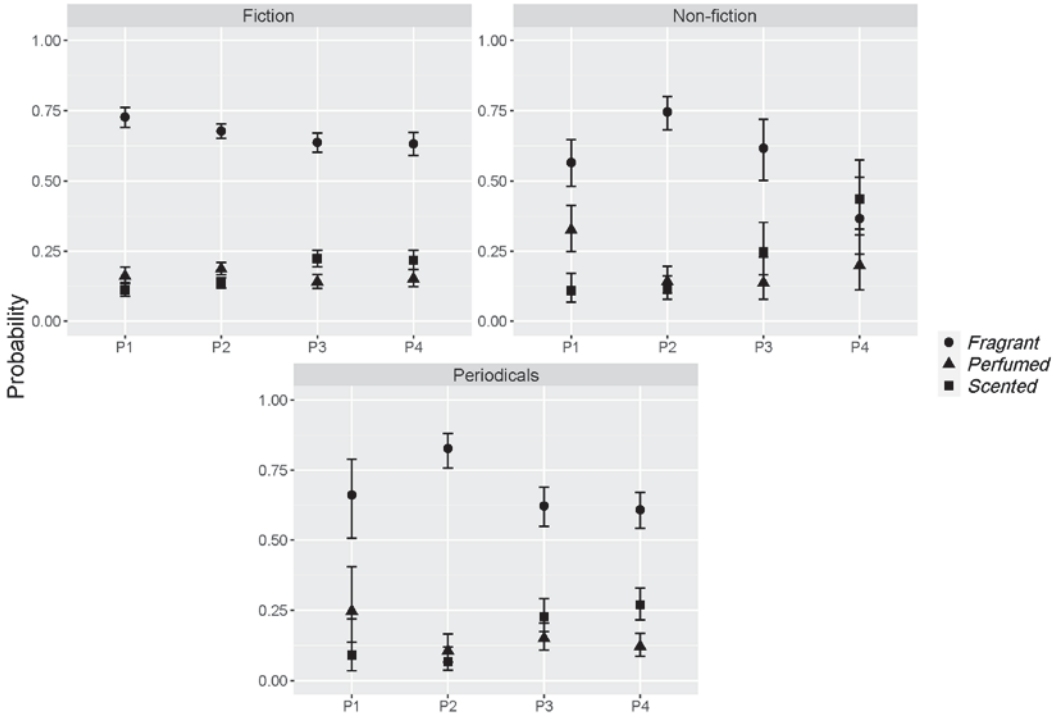


Figure 21. Effect of Text-type on the probabilities of the adjectives *fragrant*, *perfumed*, and *scented* across periods

As shown in the figure, in fictional texts *fragrant* maintains its dominance through the period 1810–2009, but we witness a significant decrease of this adjective from P1 (72.77%) to P2 (67.79%). *Perfumed* also experiences a significant downward tendency in fiction over time from P2 (18.58%) to P3 (14.03%). Contrary to *fragrant* and *perfumed*, the likelihood of *scented* in fiction increases diachronically: it is more likely in P2 (13.62%) than in P1 (11.14%), and also in P3 (22.24%) and P4 (21.71%) than in P2.

As regards non-fictional texts, the probability ranking of the adjectives changes over time. *Fragrant* is the most likely adjective in P1 (56.54%), P2 (74.58%), and P3 (61.65%), but then in P4 it is overthrown by *scented* (36.58% vs 43.56%, respectively). Although there is a significant increase of *fragrant* from P1 to P2 its probability drastically and significantly declines throughout the 20th century (P3 and P4). Similarly, the likelihood of *perfumed* displays a significant downward tendency, particularly from P1 (32.53%) to P2 (14.11%). Since then, *perfumed* remains rather stable, with only a slight increase again in P4 (19.86%). The respective decreases of *fragrant* and *perfumed* in non-fictional texts seem to be in favor of *scented*, which increases significantly from P2 onwards: while it is the least probable adjective in this text-type

in P1 (10.92%), it becomes the most likely choice one in P4 by first outperforming *perfumed* in P3 and then *fragrant* in P4.

Finally, in periodicals, the results for P1 are not entirely reliable due to the low number of instances of the adjectives in this text-type.[103] From P2 onwards, a similar trend to that identified in non-fiction is followed by *fragrant* and *scented*, whereby the latter increases at the expense of the former. Nevertheless, in this case, although the probabilities of the two adjectives become more similar, *scented* does not surpass *fragrant* in P4, as the latter still clearly dominates (26.97% vs 60.85%, respectively). Concerning *perfumed*, no significant differences emerge from P2 onwards in periodicals, which implies that this adjective remains stable over time in this particular text-type.

Figure 22 plots the interaction between Period and Sense (i.e. Model A). Again, as in the previous case, a separate graph for each Sense is provided.

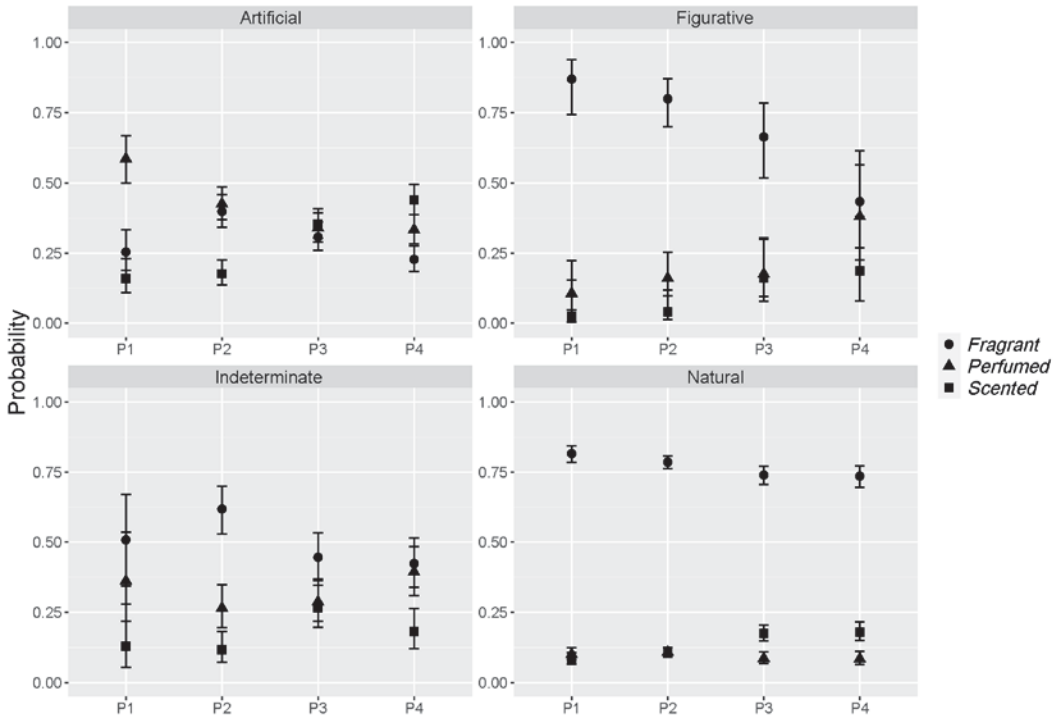


Figure 22. Effect of Sense on the probabilities of the adjectives *fragrant*, *perfumed*, and *scented* across periods

[103] Remember that *COHA* does not contain any data from newspapers in the first five decades, i.e. 1810s to 1850s (see Table 3 in Section 4.2).

As regards the artificial sense, various significant tendencies can be observed. The adjective *fragrant* displays a gradual downward tendency from P2 onwards, although the only significant contrast is that between P2 (39.80%) and P4 (22.74%). Similarly, although more drastically, *perfumed* also declines over time, particularly from P1 (58.61%) and P2 (42.55%) to P3 (33.93%) and P4 (33.30%). In fact, this adjective loses its initial clear dominance in this sense in favor of *scented* from P3 onwards: in P3 the probabilities of these two adjectives are very similar, 33.93% and 35.32%, respectively, but then *scented* becomes much more likely in P4. In fact, in this last period *scented* displays a likelihood of 43.96%, while that of *perfumed* is 33.30%, that is, a difference of more than 10 percentual points. The increase in probability of *scented* just mentioned is particularly noticeable between P2 (17.74%) and P3 (35.32%), which coincides with the turn of the 19th to the 20th century. In this specific sense, *scented*, the least likely adjective in P1 (16.07%), becomes the most probable choice in P4.

In the figurative sense, *fragrant* experiences a decrease from P2 onwards. In particular, the adjective occurs significantly less often in P3 (66.41%) and P4 (43.35%) than in P1 (86.98%), on the one hand, and in P4 than in P2 (79.94%), on the other. *Perfumed* shows the opposite trend, becoming more likely with time. In fact, it is significantly more likely in P4 (37.97%) than in P1 (10.60%), P2 (16.04%), and P3 (17.50%). As can be observed in Figure 22, the probabilities of *fragrant* (43.35%) and *perfumed* (37.97%) in the figurative sense are rather similar in P4. However, as the confidence intervals reflect, the number of instances of the adjectives in this sense at the end of the time span under analysis is considerably lower. Finally, as regards *scented*, no significant changes are observed for this adjective in the figurative sense.

In the value 'indeterminate', *fragrant* exhibits a downward tendency from P2 onwards, as it is significantly less likely in P3 (44.63%) and P4 (42.42%) than in P2 (61.82%). In turn, *scented* increases significantly from P2 (11.64%) to P3 (26.60%), but then decreases from P3 to P4 (18.17%) and *perfumed* does not display any significant changes. However, due to the decline of *fragrant*, the probabilities of *perfumed* and *fragrant* converge over time, and in P4 the two adjectives are almost equally likely (39.41% vs 42.42%, respectively).

Finally, in the natural sense, *fragrant* once again decreases significantly, particularly from P1 onwards, but it still remains by far the most likely variant in all four periods. *Perfumed* also decreases over time, but only from P2 onwards. In the case of *scented*, we witness an increase in this sense with each subsequent period, but although this upward trend is statistically

significant, the probability of *scented* continues to be much lower than that of *fragrant*, even in P4 (18.02% vs 73.55%, respectively).

The last interaction, namely that between Period and Semantic category (i.e. Model B), is shown in Figure 23: twelve separate plots are given, one for each semantic category, and within each plot the probabilities of the three adjectives per period are provided.

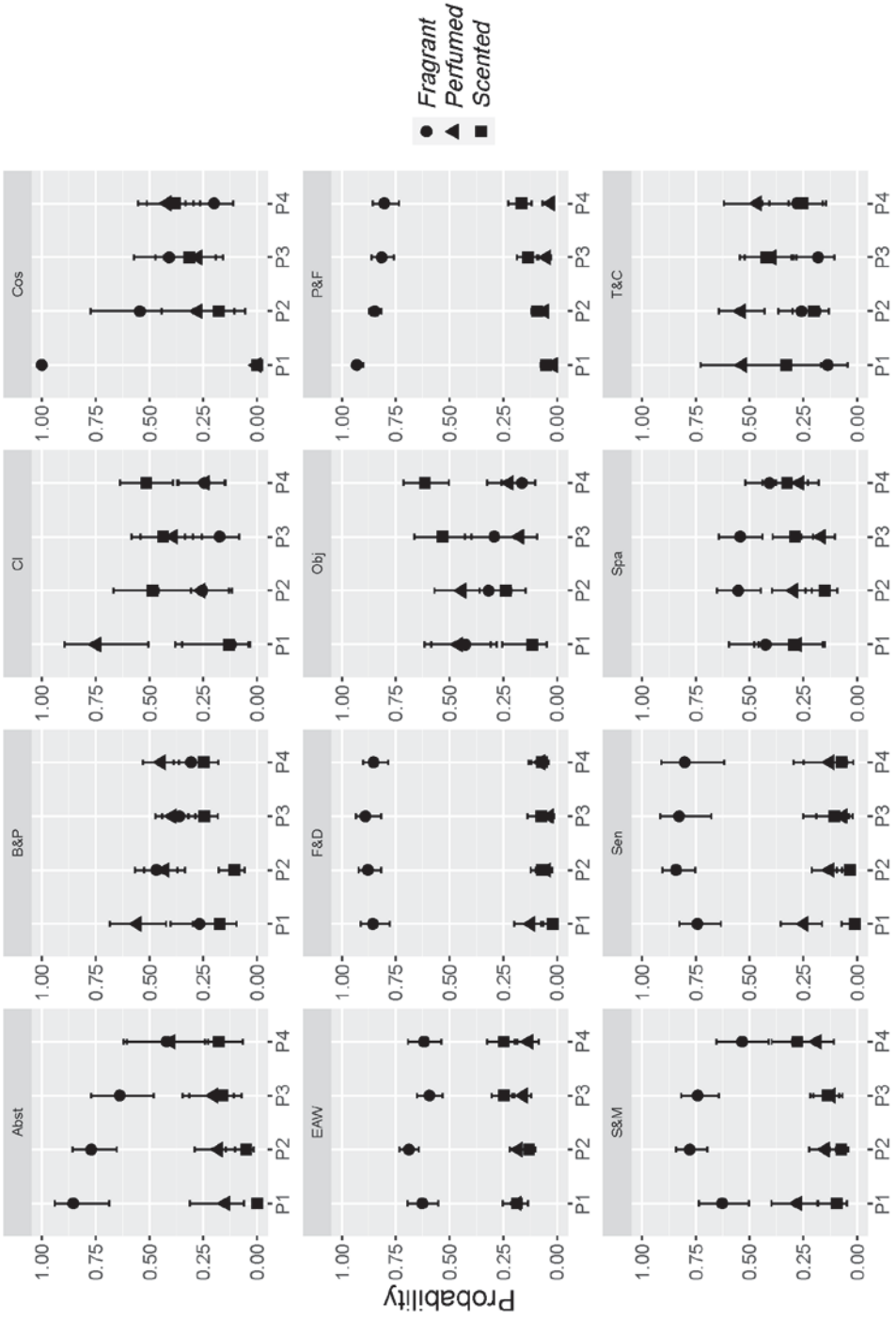


Figure 23. Effect of Semantic category on the probabilities of the adjectives *fragrant*, *perfumed*, and *scented* across periods

Starting with the level ABSTRACT, Figure 22 shows a significant decrease of *fragrant* and an increase of *perfumed*. In both cases the diachronic trends appear gradually over the time span 1810–2009: *fragrant* is less likely in P3 (63.91%) and P4 (42.04%) than in P1 (85.41%), and in P4 than in P2 (77.05%); *perfumed*, in turn, is more likely in P4 (40.13%) than in P1 (14.59%) and P2 (17.91%). By P4, *perfumed* is almost as likely as *fragrant*, although the latter was clearly predominant at the beginning of the time span examined, being 70.82 percentual points more likely than *perfumed* in P1. No significant effect emerges for *scented* in this semantic category.

In BODY AND PEOPLE, on the other hand, only *scented* exhibits a significant change diachronically, increasing in probability from P2 onwards: it is more likely in P3 (24.45%) and P4 (24.74%) than in P2 (10.43%).

With nouns belonging to the category CLEANING, *scented* increases significantly, while *perfumed* declines. For both adjectives, the significant change occurs from P1 onwards, and already in P2 *scented* becomes more likely than *perfumed* (48.56% vs. 26.09%), which was initially the most likely choice.

As concerns COSMETICS, the probabilities of the adjectives in P1 are meaningless, given that, as mentioned in Section 5.4.2, just 4 instances are attested in this period, all of which correspond to *fragrant*. From P2 onwards, only *fragrant* exhibits significant differences in probability as it is less likely in P4 (19.95%) than in P2 (54.54%): in P4 it loses its dominance and is surpassed by both *scented* (38.17%) and *perfumed* (41.88%).

In the category EARTH, ATMOSPHERE, AND WEATHER, the plot shows a significant increase of *scented*, which undergoes an upward tendency from P2 onwards: this adjective is significantly more likely in both P3 (24.79%) and P4 (24.92%) than in P2 (13.05%). Despite this increase of *scented*, *fragrant* maintains its dominance throughout the four periods.

Another semantic category in which *fragrant* remains the dominant variant over time is FOOD AND DRINK, and no significant contrasts are identified for this adjective. However, both *scented* and *perfumed* exhibit significant diachronic differences with this category of nouns: *scented* increases slightly, especially between P1 (2.11%) and P3 (7.41%), while *perfumed* decreases from 12.21% in P1 to 3.54% in P3. Nevertheless, neither *perfumed* nor *scented* are particularly likely in this category. This is so because these adjectives do not exceed an 8% likelihood in any period and *fragrant* always exhibits a probability of 85%–89%.

OBJECT is a semantic category in which all three adjectives undergo significant changes across time. While both *fragrant* and *perfumed* decrease in probability, the opposite tendency

applies to *scented*. In the case of *fragrant*, the only contrast that reaches significance is that between P4 and P1, with the adjective being significantly less likely in the former (16.41%) than in the latter period (42.60%). As regards *perfumed* and *scented*, their respective decrease and increase occur from P2 onwards: in both cases a significant change is observed between P3/P4 and P2, on the one hand, and between P3/P4 and P1, on the other. In fact, in this category, *scented* starts out as the least probable adjective in P1 (11.61%) and P2 (23.77%), but already in P3 (53.30%) it is the most likely one. Finally, in P4 the distance between *scented* (61.35%) and the other two adjectives becomes more pronounced.

The category PLANTS AND FLOWERS displays a similar development to that of FOOD AND DRINK as concerns the clear dominance of *fragrant*, which is maintained throughout the period 1810–2009. However, *fragrant* does exhibit significant changes in this category, as its probability decreases from P1 (93.14%) to P4 (80.31%). Conversely, both *perfumed* and *scented* display a significant upward trend with these types of nouns, also from P1 onwards.

In SUBSTANCE AND MATERIAL, although *fragrant* is the dominant variant in all four periods, this predominance is not as clear in P4 as in the previous stages. In fact, the probability of this adjective in P4 (53.48%) is significantly lower than in P2 (77.78%) and P3 (74.07%). *Perfumed*, in turn, remains rather stable in this category, although it shows a considerable dip in probability from P1 (27.73%) to P3 (12.10%). Finally, *scented* increases significantly if we compare its likelihood in P4 (27.83%) to that of P1 (9.53%) and P2 (7.34%).

Turning now to SENSATION, once again, *fragrant* is the default choice in all four periods, and does not exhibit any significant diachronic trends. In the case of *perfumed*, a significant decrease emerges between P1 (24.74%) and P3 (6.68%), but the high probability of this adjective in P1 is, as mentioned earlier (cf. Section 5.3.2), a result of its very frequent use with the noun *flavor* in only three particular texts. *Scented*, despite its low likelihood in all four periods with SENSATION nouns, exhibits a slight increase from P1 (1.06%) to P3 (10.58%).

In the category SPACE, the only significant change concerns *perfumed*, which decreases from P2 (29.51%) to P3 (16.65%). The probabilities of *fragrant* and *scented*, on the other hand, seem to remain stable over the time span analyzed.

Lastly, in TEXTILE AND CLOTHING, only the probability of *scented* experiences significant differences over time: it first increases from P2 (20.25%) to P3 (42.24%) and then decreases between P3 and P4 (25.51%). In turn, no significant changes in the probabilities of *fragrant* and *perfumed* emerge for this semantic category of nouns.

As mentioned in Section 6.1, the effect of the last predictor, to wit, Collocate, had to be measured by computing a different type of regression analysis, due to its nature. Therefore, before turning to the discussion of the findings obtained, the analyses comprising the variable Collocate are provided in the next section in order to determine whether the inclusion of this predictor significantly improves the models including only fixed effects that have been presented so far.

6.2.2 Mixed-effects models: The effects of individual noun collocates

As explained previously (cf. Section 6.1), the mixed-effects regression models were computed using the same heuristic employed in the multinomial models, namely one vs. rest. Three separate models were calculated, each one focusing on the collocational preferences of one of the three adjectives. The summary statistics of the models zooming in on the behavior of *fragrant*, *perfumed*, and *scented*, respectively, are presented in Tables 23–25.

Table 23: Model summary of the binary mixed-effects logistic regression (*fragrant*)

Model summary	
C-index of concordance	0.8417
accuracy	78.11% (baseline = 63.90%; $p < 0.001$)
Nagelkerke R^2 (marginal)	0.2106
Nagelkerke R^2 (conditional)	0.4053

Table 24: Model summary of the binary mixed-effects logistic regression (*perfumed*)

Model summary	
C-index of concordance	0.8349
accuracy	84.49% (baseline = 81.61%; $p < 0.001$)
Nagelkerke R^2 (marginal)	0.1538
Nagelkerke R^2 (conditional)	0.3683

Table 25: Model summary of the binary mixed-effects logistic regression (*scented*)

Model summary	
C-index of concordance	0.7903
accuracy	83.74% (baseline = 82.29%; $p < 0.01$)
Nagelkerke R^2 (marginal)	0.1508
Nagelkerke R^2 (conditional)	0.2775

In this case, the indexes differ somewhat from those presented for the multinomial regression analyses. First, in the upper row of each table we have the *C*-index of concordance, which is an indicator of how well the model discriminates between *fragrant* and the other two adjectives (i.e. *perfumed*/*scented*) in Table 23, between *perfumed* and the other two (i.e. *fragrant/scented*) in Table 24, and between *scented* and the other two (i.e. *fragrant/perfumed*) in Table 25. The

values of this index range between 0.5 and 1, and the higher the value, the better the discrimination of the model. The following thresholds are commonly distinguished for the *C*-index: (i) $C = 0.5$ no discrimination, (ii) $C = 0.7–0.8$ acceptable discrimination, (iii) $C = 0.8–0.9$ excellent discrimination, and (iv) $C \geq 0.9$ outstanding discrimination (Speelman 2014: 514–515; Levshina 2015: 259). Therefore, the values obtained in the models for *fragrant* and *perfumed* (i.e. 0.8417 and 0.8349, respectively) entail that these models exhibit an excellent discrimination capacity. However, the model for *scented* exhibits only an acceptable discrimination (i.e. 0.7903), as it does not reach the 0.8 threshold, although it is very close to it.

In the second row of Tables 23–25, the accuracies of the models are provided. As can be observed, the values for all three models point to an improvement with respect to the multinomial models A and B with only fixed effects presented in the previous section (i.e. 67.34% and 69.13%, respectively; cf. Tables 17 and 18). The *p*-values also confirm that these three accuracy values are significantly better than their respective baselines.

In the last two rows of the tables, two Nagelkerke R^2 values are shown, termed marginal and conditional.[104] These values were retrieved by using the `MuMIn` package and, in particular, the `r.squaredGLMM()` function (Barton 2020). The marginal Nagelkerke R^2 indexes indicate the predictive power of the models accounting solely for the fixed effects. The conditional Nagelkerke R^2 indexes, in turn, signal the predictive power of the models considering both the fixed and the random effects. As can be observed in Tables 23–25, there is an important difference between the two values, as the latter is almost twice as large as the former in the models for *fragrant* and *scented* (cf. Tables 23 and 25, respectively) and more than twice as large in the model for *perfumed* (cf. Table 24). This means that the random effects of the variable Collocate are extremely important to discriminate between the near-synonymous adjectives, since by adding them to the models, the amount of explained variance in the data increases dramatically.

The models achieved by means of mixed-effects regression analyses constitute a substantial improvement over the models including only fixed effects, given that all indexes reach values indicating a good fit, which was not the case in the models with fixed effects only (cf. Section 6.2.1 for more details). This highlights the importance of taking the individual collocates of the near-synonyms into consideration, as this predictor accounts for a considerable

[104] McFadden R^2 values are not shown in Tables 23–25, since the function used, namely `glmer`, does not provide this particular index. Only the Nagelkerke R^2 values are therefore discussed here.

amount of the variance in the data. In the remainder of this section, the idiosyncratic collocational preferences of the three adjectives are discussed in more depth by focusing on their most salient noun collocates.

Figures 24–26 plot the collocational behavior of the adjectives according to mixed-effects regression analysis. In these figures, each individual point in the plot represents one specific noun collocate. Vertical axes represent the adjustments to the intercept for each noun collocate, with positive values indicating collocates which are more strongly associated with the near-synonym of interest (*fragrant* in Figure 24, *perfumed* in Figure 25, and *scented* in Figure 26) and negative values indicating collocates which are more strongly associated with the other two adjectives combined (*perfumed* and *scented* in Figure 24, *fragrant* and *scented* in Figure 25, and *fragrant* and *perfumed* in Figure 26). Finally, horizontal axes plot the absolute adjustments to the intercept for each noun collocate, that is, the same values shown on the vertical axes but without taking into account whether they are positive or negative. In more conceptual terms, horizontal axes signal the strength of the association between the noun collocates and the adjectives: the higher the value, the stronger the association. For each adjective, the most prominent collocates have been included in the figures by selecting an arbitrary threshold value of 1.25. Given that the goal here is to offer only a bird's-eye-view of the general collocational tendencies of the near-synonymous items in order to show the importance of the idiosyncratic preferences of each adjective, the plots represent the tendencies in all four periods together. A more fine-grained diachronic analysis of the near-synonyms' collocational behavior is provided in Chapter 7, where more adequate methods to undertake this task are employed.

Starting with *fragrant*, Figure 24 displays the nouns that are more strongly associated with this variant. In this particular case, positive values on the vertical axis indicate those collocates that are associated with *fragrant*, as opposed to *perfumed*/*scented*, whereas the values on the horizontal axis signal the strength of this association.

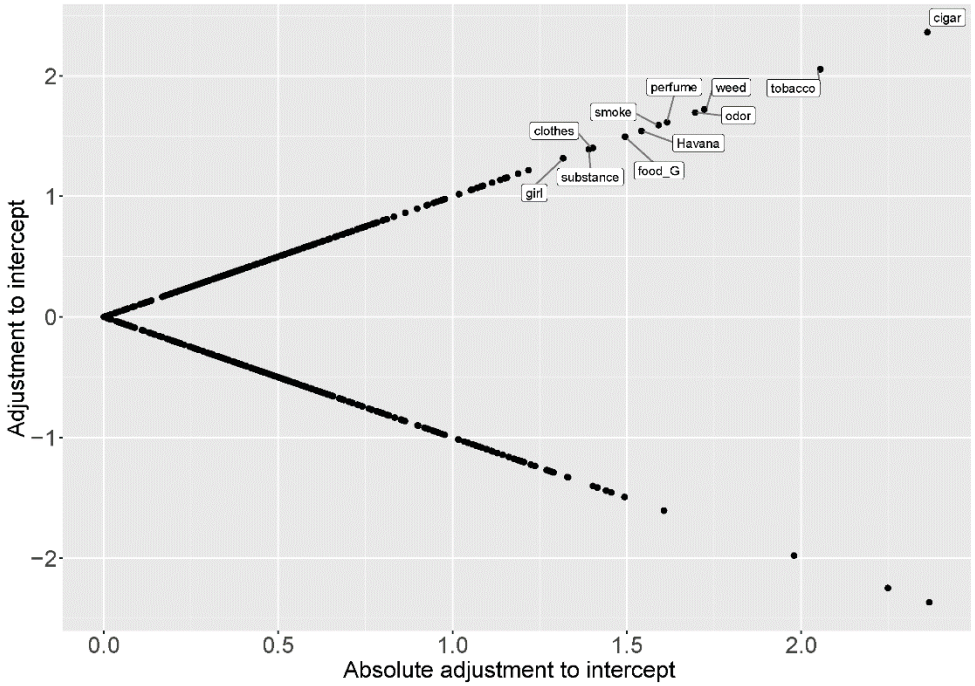


Figure 24. Collocational preferences of *fragrant*

As can be observed, in the case of *fragrant*, 11 collocates emerge as prominent and some interesting trends can be extracted from this figure. The specific noun collocates which are more strongly connected with *fragrant* are those located in the top right corner of the plot (e.g. *cigar*, *tobacco*, *weed*, and *odor*). Of the 11 collocates included in Figure 24, three are clearly related to the semantic domain of TOBACCO AND SMOKING, namely *cigar*, *tobacco*, and *Havana*. Moreover, two additional nouns, though polysemous, are also often used in this domain: *smoke* and *weed*. In fact, of the 42 total instances in *COHA* in which *fragrant* occurs as a modifier of *smoke*, 25 (59.53%) are of the type illustrated in example (111), and hence belong to the same semantic domain:

(111) He kicked the newspaper from his feet, sat back, and felt in his pocket for a cigar. It was a short thick one, and he lit it. and puffed out a cloud of blue and **fragrant smoke**. (*COHA*, 1951, FIC, GodsMen)

Similarly, of the 12 times *fragrant* co-occurs with *weed*, 8 (66.67%) are used to refer to a substance that is used for smoking, as in example (112):

(112) My father always smoked a pipe. I like it. “He took the pipe down and packed it with tobacco, held a taper to the fire and lighted the **fragrant weed**. Then he drew the other rocker up to the fire beside her.” (*COHA*, 1955, FIC, QuicksilverPool)

The rest of the collocates of *fragrant* represent a varied selection of domains. The most prominent are the two nouns *odor* and *perfume*, which are in nearly all cases used to denote types of scents, therefore belonging to the semantic category SENSATION.

Continuing with *perfumed*, its collocates are visualized in Figure 25.

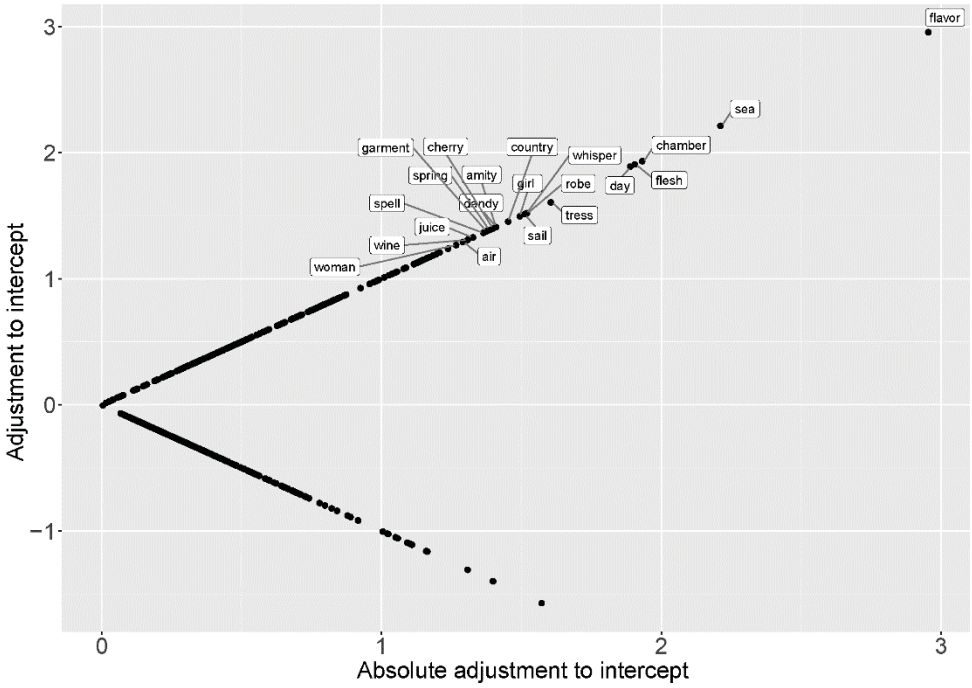


Figure 25. Collocational preferences of *perfumed*

Again, some clear tendencies can be observed in the plot:

(i) Several of the most prominent nouns pertain to the semantic category BODY AND PEOPLE, a category in which *perfumed* is the preferred variant in most periods (cf. Figure 23 in Section 6.2.1); these are *flesh* (of a human body; cf. example (113)), *tress* (of hair), *girl*, *dandy*, and *woman*.

(113) Astonishingly, Cheryl began to laugh. Why didn't you let go, you idiot? Now look at the mess we're in. But Alvin wasn't looking at anything. His nose was buried in the soft, **perfumed flesh** just below her left ear. (*COHA*, 1972, FIC, Clone)

(ii) More surprisingly, several collocates belonging to FOOD AND DRINK, to wit, *flesh* (of fruit; cf. example (114)), *cherry*, *juice*, and *wine*, are strongly associated with *perfumed*, although the probability of finding this particular adjective in this semantic category is relatively low (cf. Figure 23 above). In the case of *flesh* (of fruit), most instances come from the same text (i.e. *The American fruit culturist, containing directions for the propagation and culture of fruit trees, in the nursery, orchard, and garden, with descript* (AmericanFruitCulturist). The noun *flavor*, belonging to SENSATION, is the most strongly associated collocate of *perfumed* according to the mixed-effects model. However, similarly to *flesh* (of fruit), this is due its high co-occurrence frequency with *perfumed* in three specific texts (cf. Section 5.3.2). Therefore, these two associations may be due to the idiosyncratic preferences of the authors of these texts.

(114) Large Pears. BEURRE BOSC. (Syn. Calebasse Bosc.) Large, very distinct pyriform, neck rather long and very narrow, acute; body large oblate; surface nearly smooth, deep yellow, russeted in patches; stalk an inch and a half long, slender, curved; basin very shallow; **flesh** juicy, buttery, rich, perceptibly **perfumed**, sweet, excellent. Mid-autumn. Growth moderate, a regular, even bearer. (*COHA*, 1850, NF, AmericanFruitCulturist)

(iii) Three of the collocates that emerge as being highly associated with *perfumed* represent instantiations of this adjective being used in the figurative sense: *whisper*, *amity*, and *spell*. For an illustrative example, see (115). As in the case of the nouns belonging to BODY AND PEOPLE, this is not at all surprising, given that *perfumed* is relatively likely in the figurative sense, especially at the end of the time span examined (cf. Figure 22 above). The rest of the collocates mostly belong to the categories EARTH, ATMOSPHERE, AND WEATHER (e.g. *sea*, *spring*, and *country*) and

TEXTILE AND CLOTHING (i.e. *garment* and *robe*), two categories in which *perfumed* is relatively productive.

(115) I hate the prostitution of the name of friendship to signify modish and worldly alliances. I much prefer the company of ploughboys and tin-peddlers to the silken and **perfumed amity** which celebrates its days of encounter by a frivolous display, by rides in a curricle and dinners at the best taverns. (*COHA*, 1841, FIC, Essays–FirstSeries)

Finally, Figure 26 plots the collocational preferences of *scented*.

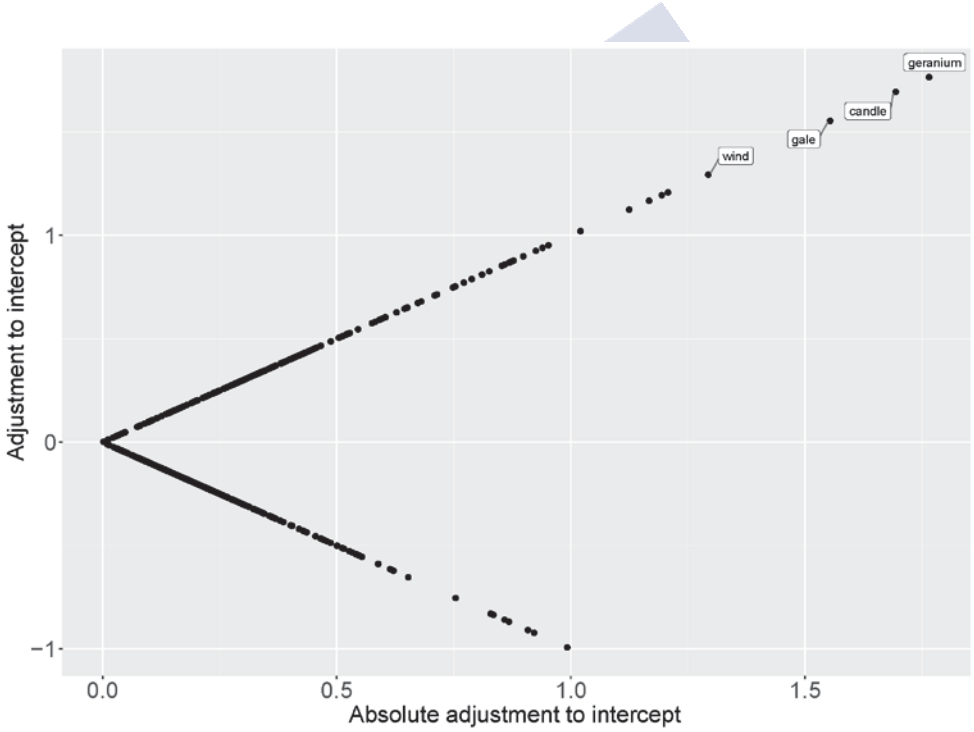


Figure 26. Collocational preferences of *scented*

In this case, only 4 nouns are included in the plot, namely *geranium*, *candle*, *gale*, and *wind*. That a considerably lower number of collocates emerge for this particular adjective is probably due to the fact that the mixed-effects regression model for *scented* have a lower discriminatory power than those for *fragrant* and *perfumed* (cf. compare the first row in Tables 23–25 above). This, in turn, may be the result of the collocational preferences of *scented* being more similar to the collocational behavior of *fragrant* and *perfumed* than these two are with respect to each other. In Section 5.3, it was postulated that *fragrant* and *perfumed* are located on opposite ends

of a continuum of senses, with the former closer to the natural end of the cline and the latter placed at the artificial end. *Scented*, on the other hand, occupied a more intermediate position, given that it seemed to be relatively salient in both the artificial and natural senses (cf. Figures 7 and 8 in Section 5.3.1). Therefore, it may well be the case that *scented* shares more semantic ground with *fragrant* and *perfumed*, than either of them does with one another. The fact that the model for *scented* performs worse than those of the other two adjectives may thus be a consequence of *scented* occupying a middle position between *fragrant* and *perfumed*. This is an issue which is further explored in Chapter 7.

Concerning the 4 noun collocates which are strongly associated with *scented*, the most prominent one is *geranium*. However, as in the case of *flesh* (of fruit) and *flavor* discussed above in relation to *perfumed*, this is probably also a corollary of the preferences of one particular author, as out of the 13 total instances in which *scented* modifies *geranium*, 11 appear in the same text in *COHA* (i.e. *Scan*: *Homesteading*). The second noun collocate that is more strongly associated with *scented* according to the mixed-effects regression analysis is *candle*, which represents a particularly interesting case. This is so because *candle* is modified by *scented* a total of 25 times in the corpus, but only from the 1950s onwards, and of these, 20 instances are dated from the 1990s onwards. Therefore, most occurrences of the collocation *scented candle* appear in P4. Moreover, the collocate *candle* is almost exclusively used with the adjective *scented*, as only two cases of this noun are found with *fragrant* and one with *perfumed* in the whole corpus. These findings raise the question of whether *scented candle* should be considered either a noun phrase formed by an attributive adjective and a head noun or whether, by the end of the time span analyzed here, it could already be considered a compound or a multiword unit (Liu 2010: 64–66) used to refer to a specific concept, namely that of a candle used for aromatic purposes. This hypothesis seems to be strengthened by the fact that *scented* occurs immediately preceding *candle*, without any kind of intervening material, in 20 of the 25 cases in *COHA*, as in example (116):

(116) The agricultural economy is in no better shape than the stock market at this time. We who produce the world where Heirston finds John Robshaw pillows along with Calypso's own Pom Pom cotton throws and **scented candles** ($38 each) in eleven fragrances. (*COHA*, 2009, MAG, Coca (*Town and Country*)).

In fact, it could even be argued that the combination *scented candle* is in the early stages of a *lexicalization* process, understood as

> the change whereby in certain linguistic contexts, speakers use a syntactic construction or word formation as a new contentful form with formal and semantic properties that are not completely derivable or predictable from the constituents of the construction or word formation pattern (Brinton & Traugott 2005: 96).

This is so because, while a candle is employed primarily as a source of artificial light (*OED*, s.v. *candle* n. 1a), the main use of a scented candle nowadays is that of refreshing or perfuming the air of a place, much like air fresheners. This can be deduced from examples such as (117), where *scented candle* is referred to as a type of air freshener delivery method:

(117) Many other **air freshener delivery methods** have become popular since, including **scented candles**, reed diffusers, potpourri, and heat release products. (*CD*, s.v. *candle* n. collocations)

The remaining two prominent collocates of *scented* are *gale* and *wind*, two related words, as the former is a hyponym of the latter, which belong to the semantic category EARTH, ATMOSPHERE, AND WEATHER.

6.3 DISCUSSION

The analyses conducted in this chapter proved the adequacy of the set of variables selected for examination since they all emerged as significant predictors of the choice between the near-synonymous adjectives *fragrant*, *perfumed*, and *scented*. This indicates that variables of a varied nature should be considered in order to reach a better understanding of the variation between lexical near-synonyms. However, not all predictors play an equally important role according to the statistical methods employed here. In Section 6.2.1, both the multinomial regression models and the random forest analysis pointed to the conclusion that the intralinguistic semantic variables were stronger determinants for the particular synonym set under analysis, especially the two related predictors Semantic category and Sense, but also Countability and Concreteness. Similarly, the mixed-effects analyses computed in Section 6.2.2 indicated that the individual noun collocates considerably improved the performance of the

regression models, as idiosyncratic collocational patterns explained a great amount of the variation existing between the near-synonyms. This is in line with previous research (cf. Section 3.2.2), which demonstrates that zooming in on the nouns that adjectives modify is a particularly effective approach to uncover their internal semantic structure (Geeraerts 1986: 282–284). On the other hand, the morphosyntactic variables Syntactic function and Degree, although significant, played a relatively minor role, particularly the latter (cf. Figure 16). This finding highlights the importance of considering the POS of the near-synonyms object of study when selecting the variables to be included in the analysis: while morphosyntactic variables are indeed crucial for near-synonymous verbs (e.g. Divjak 2010: 183–193), their effect size is rather negligible in the case of adjectives, at least in the set examined here. This reminds us of Hank's (1996: 92; 96) and Liu's (2010: 61) arguments that, despite the similar goals of usage-based corpus research on near-synonymy, the micro-procedures that need to be employed in each study vary, since the ways in and the dimensions on which near-synonyms differ also vary from one set to another. Additionally, one of the two extralinguistic variables, to wit, Period, also exerted a great influence on the distribution of the near-synonyms at stake, being ranked third in order of importance according to the random forest analysis. The fact that Period is a more powerful predictor than most of the other variables, both semantic and morphosyntactic ones, suggests that the diachronic dimension of lexical variation cannot be disregarded. As shown in Section 3.3, this issue has not received the attention it clearly deserves, as only a handful of diachronic distributional corpus-based studies on lexical near-synonyms exist to date, especially if compared to the relatively large amount of historical research on other linguistic phenomena, including polysemy and constructional synonymy.

Concerning the specific effects of the predictors, most of the diachronic onomasiological tendencies of Sense and Semantic category uncovered in Section 5.4 are confirmed by the regression analyses, as many of these trends reach statistical significance. Once again, *fragrant* emerges as the default choice in the natural sense and all five prototypically 'natural' semantic categories (i.e. EARTH, ATMOSPHERE, AND WEATHER, FOOD AND DRINK, PLANTS AND FLOWERS, SENSATION, and SPACE). We witness, however, a slight decrease in the probability of this adjective in most of these contexts, mainly in favor of *scented*, which increases significantly over time, although it does not threaten the prominence of *fragrant* with these types of nouns. On the other hand, some substantial changes do occur in the artificial sense and with most of the related semantic categories, to wit, CLEANING, COSMETICS, OBJECT, and SUBSTANCE AND

MATERIAL. In fact, we observe a reorganization of the near-synonyms in many of these contexts of use.

(i) The probability of *fragrant* decreases somewhat in the artificial sense, and that of *perfumed* does so more drastically. *Scented*, in turn, gains ground at the expense of the other two adjectives, particularly from P2 to P3, and becomes the default choice in this sense in P4.

(ii) With respect to specific semantic categories, the probability of *scented* increases particularly in CLEANING, COSMETICS, and OBJECT, but also somewhat in SUBSTANCE AND MATERIAL. In fact, in CLEANING and in OBJECT *scented* becomes the default choice, ousting *perfumed* in the former and both *fragrant* and *perfumed* in the latter. In COSMETICS and SUBSTANCE AND MATERIAL, *scented* converges in probabilities with *perfumed* and *fragrant*, respectively, although it does not surpass them.

Similarly, the internal structure of the set is also reorganized in the figurative sense and in the category ABSTRACT, where *fragrant* loses its dominance over time; by P4, *perfumed* and *fragrant* are almost equally likely. Finally, in the value 'indeterminate', *fragrant* also loses ground in favor of *scented*, while *perfumed* remains stable. This change seems to be restricted to BODY AND PEOPLE, where *scented* also displays a substantial upward tendency.

As in Chapter 5, the results of this chapter suggest the existence of a probably still ongoing process of substitution within the synonym set object of study, whereby *scented* gains ground at the expense of *fragrant* and *perfumed*. Figure 27 depicts this process in a more visual manner, where both the overall frequencies of each adjective, as well as those within the artificial, indeterminate, and natural senses, are displayed.[105] Here, the vertical axis represents the percentages of use of the adjectives in each sense, whereas the horizontal axis shows the four historical periods. Within the plot, a different pattern is used to signal the frequencies of each adjective: squares are used to indicate those of *fragrant* and black diamonds those of *scented*, while the absence of a pattern in the middle area of the graph corresponds to *perfumed*. Additionally, a different color signals the frequencies of each sense: purple stands for the artificial sense, white for indeterminate uses, and green for the natural sense.

[105] The figurative sense is not included in this graph as the frequencies of the three adjectives in this sense are almost negligible.

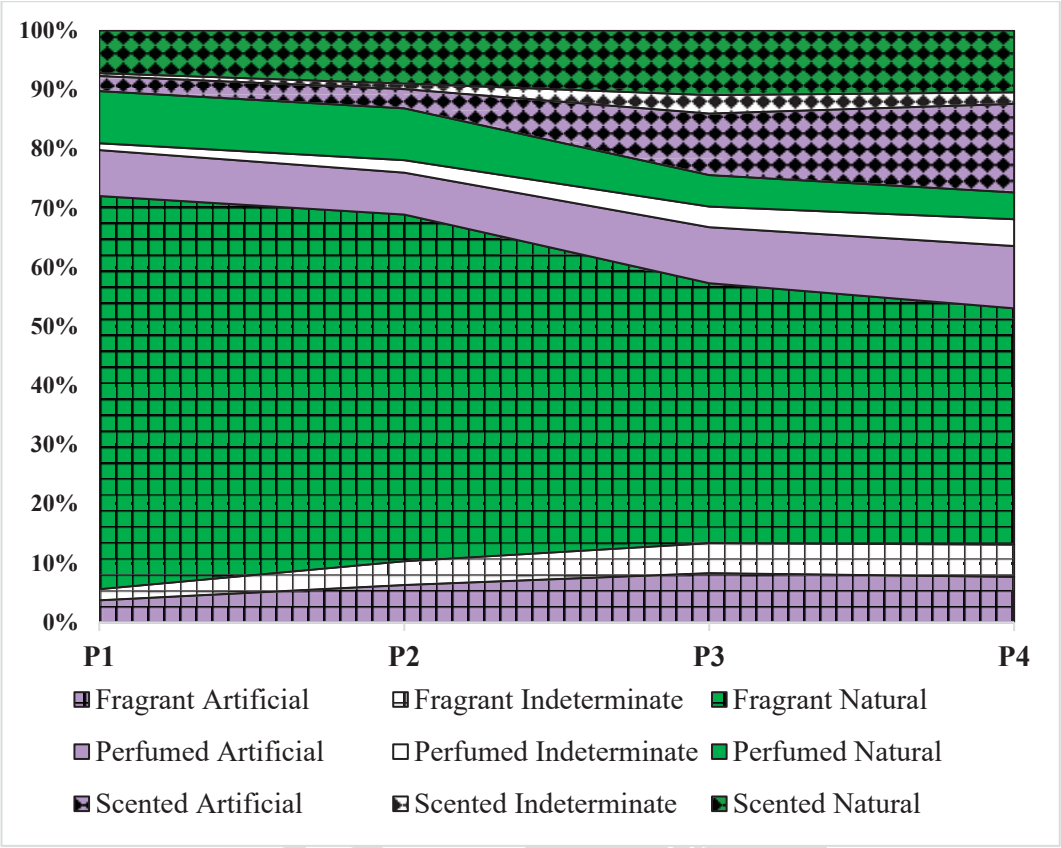


Figure 27. The use of *fragrant*, *perfumed*, and *scented* in the artificial, indeterminate, and natural senses over time

Figure 27 combines the patterns and colors to portray the percentages of use of the three adjectives in each sense over time. By focusing exclusively on the patterns, the ongoing process of replacement can clearly be observed. Overall, *fragrant* —located at the bottom of the figure— decreases considerably in frequency throughout the period 1810–2009, while the opposite tendency is true for *scented* —located at the top of the figure. However, while *fragrant* becomes less and less commonly selected over time in favor of *scented*, it still remains the dominant choice overall in P4. *Perfumed*, in turn, does neither increase nor decrease substantially. In P4, the situation seems to be one of a greater degree of competition between the near-synonyms, with all three adjectives being more equally distributed than in P1. Nevertheless, this greater degree of competition between the variants does not appear to result in an increasingly higher degree of differentiation between the synonyms as time passes. In fact, the contrary tendency seems to hold in this particular case, namely attraction. This can be observed by focusing on the colors in Figure 27, that is, by zooming in on the frequencies of each adjective in the different senses. By P4, all three adjectives are much more frequently used in the artificial sense

—the purple areas— than in earlier periods. On the contrary, none of the adjectives exhibit a substantial upward tendency in the natural sense —the green areas—: *fragrant* and *perfumed* both clearly decrease in this sense over time, and *scented* shows only a minor increase. Similarly to the development of the adjectives in the artificial sense, all of them also increase in the value 'indeterminate' —the white areas— from P1 to P4. Interestingly, the diachronic development of the adjectives in so-called indeterminate contexts is very similar to their evolution in the artificial sense. This could be taken as an indication that a relatively large amount of such ambiguous cases is, in fact, 'artificial', although this cannot be discerned in the actual instances in *COHA*, not even if their extended contexts are analyzed. Moreover, more than half of the examples (56.96%) which are classified as indeterminate belong to the semantic category BODY AND PEOPLE. The reader may recall that a considerable increase over time of this particular semantic category emerged from the analysis (cf. Section 5.2.2), and that this upward tendency may have been the result of the growing availability and use of cosmetics and other personal care products applied to the body (cf. Section 5.2.3). Therefore, it may well be the case that the general rise of the indeterminate sense is, at least in part, due to the increase of the category BODY AND PEOPLE and that of potential artificial uses of the adjectives which had to be categorized as 'indeterminate' due to a lack of sufficient evidence.

The end result of the frequency fluctuations described in the previous paragraph is a more even distribution of the adjectives in P4, both generally and across the different senses. In general terms, the overall frequencies of the adjectives are more similar in P4, although *fragrant* remains the dominant variant of the set. Additionally, their individual distribution across senses is also more similar at the end of the time span analyzed, with all three adjectives being more commonly used than before in both the artificial sense and in indeterminate cases, and less so in the natural sense, except in the case of *scented*, which does not recede from natural contexts of use. Therefore, the findings point to the conclusion reached in Chapter 5 that both processes of substitution and attraction are at play in the history of this synonym set. De Smet et al. (2018: 217) postulate that attraction results from analogical change, whereby functionally similar words or structures parallel each other's behavior through an interchange of characteristics, an explanation that would account, for instance, for near-synonymous expressions developing the same metaphorical mappings. However, a more likely explanation in this particular case, as argued in Section 5.2.3, is that the tendency of all three adjectives to denote more and more artificial aromas derives from extralinguistic forces. In particular, the underlying motivations

for change may be related here to the socio-cultural and technological advances of American society over the time span 1810–2009, whereby these specific types of aromas have become increasingly prominent in the day-to-day life of the American population.

Turning now to the significant predictors which do not interact with Period, namely the semantic factors Concreteness rating and Countability and the morphosyntatic variables Degree and Syntactic function, *fragrant* is the default choice in all levels of the four predictors throughout the entire time span analyzed. This is not surprising, given the much higher overall frequency of *fragrant* in the corpus as compared to the other two adjectives. However, there are some contexts of use in which both *perfumed* and *scented* are more likely than expected, whereas the reverse is often true for *fragrant*:

(i) *Perfumed* is more likely with highly abstract nouns, i.e. those with a low concreteness rating, than with more concrete nouns, while the opposite tendency holds for *scented*.

(ii) As concerns Countability, *fragrant* is dominant in all three levels (count, non-count, and other), but less so with non-count nouns, where *scented* is more probable than expected, and with pronouns and proper nouns ('other'), where *perfumed* is marginally more likely than expected.

(iii) In the case of the variable Degree, *fragrant* is again salient in all levels, but much more so in the comparative and superlative degrees, because *perfumed* and *scented* are hardly ever used in either comparative or superlative form in *COHA*.

(iv) Finally, as regards Syntactic function, *fragrant* outperforms the other two adjectives in all functions, especially as a predicative complement. However, in the postpositive use and in other minor functions its probability is not as pronounced: in the former, *scented* exhibits a higher probability than expected, whereas the same is true for *perfumed* when serving as a predicative adjunct, fused-modifier head, or predeterminer.

These results entail that the overall prominence of *fragrant* is not threatened in any of these morphosyntactic or semantic contexts, and that this situation is maintained diachronically. However, some interesting differences are identified between *perfumed* and *scented*, especially

as regards Concreteness rating, Countability, and Syntactic function, which suggests that a division of labor may exist between these two adjectives in some of these contexts.

Some stylistic differences between the three near-synonyms also emerged from the regression analyses. Crucially, some diachronic changes can be said to exist as regards the probability of the adjectives in some text-types in *COHA*. An overall increase of *scented* is attested in all three text-types, but it is in non-fictional texts where this upward tendency is most pronounced. In fact, *scented*, the least probable adjective in non-fiction in P1, becomes the most likely choice in P4. In fiction and periodicals *scented* also increases and outperforms *perfumed* in P3 and P4, but it does not reach the probability of *fragrant* in any of these two text-types. Therefore, the overall increase of *scented* in frequency uncovered in Section 5.1 seems to occur across all text-types, though the increase is sharper in non-fictional texts. Thus, as in previous studies on the distribution of near-synonyms, stylistic differences also play a role on the choice between the three adjectives examined here (e.g. Biber, Conrad & Reppen 1998: Section 2.6; Kjellmer 2003; Liu & Espino 2012). However, it must be noted here that the categorization of text-types used in the present dissertation (fiction, non-fiction, and periodicals) is relatively coarse-grained and could be further refined in future research to pinpoint exactly in which particular subgenres each adjective is preferred (e.g. poetry, novels, movie and play scripts, editorials, advertisements, academic writing from different sciences).

Lastly, the goodness-of-fit indexes of the multinomial regression models provided in Section 6.2.1 signal that there is still room for improvement. As previously explained, the accuracy and the R^2 values of both Models A and B indicate that the set of predictors included in the analyses and their interactions are insufficient to explain all the variation in the data. Three possible reasons why these models are not optimal are postulated here. First, although a rather large number of predictors of different types has been examined, other variables may also play a role in the choice between the near-synonyms. Two factors which have been shown to influence linguistic variation in different dimensions of language are the opposing forces of *priming* (Gries 2005; Szmrecsanyi 2005; 2006) and the so-called *horror aequi* principle (e.g. Rohdenburg 2003; Vosberg 2003). The former consists in the (explicit) repetition or reiteration of linguistic elements precisely due to the fact that they have been employed in a recently produced utterance. The latter, in turn, implies the avoidance of the repeated use of "formally (near-) identical and (near-) adjacent […]" linguistic elements (Rohdenburg 2003: 236). For instance, in example (118) *scented*, as opposed to *fragrant* or *perfumed*, may have been chosen

as a modifier of the noun *fruit* precisely because it has already been used earlier in the same sentence (i.e. *scented bloom*), thus potentially constituting a case of priming. On the contrary, in (119) *fragrant* may have been selected as a modifier of the COSMETICS noun *oils* because *scented*, which is a more likely choice with this particular semantic category, has already been employed in the previous noun phrase (i.e. *scented soap*), hence a possible instance of *horror aequi*. Similarly, the avoidance of *perfumed* occurring in the neighboring contexts of the noun *perfume* or of the verb *perfume* may illustrate cases of the *horror aequi* principle.[106]

(118) About me were miles on miles of apple orchards, and in our own I worked and played from the time of pink and **scented** bloom to that of **scented** and yellow fruit. (*COHA*, 1930, MAG, Atlantic)

(119) "Everybody is waiting for it to happen down here, hoping every- thing will jell," said Sarah Laight, an ex-pat Brit who sells **scented** soaps, **fragrant** oils and summer-of-love dresses in her Life boutique. (*COHA*, 1994, NEWS, SanFran)

Moreover, the non-optimal values of the multinomial models may result from interactions between variables which were not tested for in the present analyses. The reader may recall that only the interactions between the extralinguistic variable Period and the other predictors were included here, as the main aim of this dissertation is to examine the diachronic development of the near-synonymous adjectives. However, interesting findings may be revealed if a larger number of predictors, for instance, Text-type and Sense or Semantic category, are cross-cut so as to see whether functional differences exist between the near-synonyms in different genres. Nevertheless, the inclusion of additional variables and variable interactions would require a larger dataset than the one used in the present dissertation.

The second reason why the models with only fixed effects are not optimal could be that some of the variation in the data may be relatively random. After all, we are here dealing with the phenomenon of lexical synonymy and semantically related terms need to exhibit a sufficiently high degree of interchangeability without a change in meaning in order to be considered near-synonyms. If this was not the case, language users would not be able to resort to near-synonyms for the purposes of avoidance of repetition and coherence in text production, among others (Murphy 2013: 289–290).

[106] The same holds for the combinations *scented scent* or *fragrant fragrance*.

Finally, and more importantly, the mixed-effects regression analyses presented in Section 6.2.2 highlighted the great importance of taking into consideration the idiosyncratic behavior of the near-synonyms. By including the specific noun collocates of the adjectives in the model, a substantial improvement in model fit was achieved. The resulting models with both fixed and random effects (cf. Section 6.2.2) explained almost double the amount of variation than a model with only fixed effects (cf. Section 6.2.1). This means that, in this particular case, what really distinguishes between the adjectives are the specific nouns that they modify. Thus, Firth's (1957a: 11) famous quote "you shall know a word by the company it keeps" (cf. Section 2.3.3), by means of which he emphasizes the importance of collocational patterns in lexical semantics, seems to be an adequate description of the synonym set at issue. However, as shown by the models with fixed effects only presented in Section 6.2.1, collocations do not tell the whole story, since the fixed effects are also significant determinants of the variation between *fragrant*, *perfumed*, and *scented*.

6.4 SUMMARY

This chapter has focused on the onomasiological structure of the synonym set under analysis, thus emphasizing the competition over time between the three adjectives *fragrant*, *perfumed*, and *scented* in different contexts of use. A wider range of variables than in Chapter 5 was examined by means of more powerful statistical tools, which enabled us to confirm several of the findings obtained previously and also to reach additional conclusions. In particular, it was shown that variables of a semantic nature, specifically those pertaining to the noun collocates of the adjectives, had a much stronger explanatory power than morphosyntactic and stylistic ones, although the latter two variable types also contributed significantly to the models. Moreover, the effect of the extralinguistic variable Period emerged as highly significant in this particular synonym set, thus highlighting the importance of adding a diachronic dimension to the study of near-synonyms. However, models containing only fixed effects proved to be insufficient to explain the variation in the data and had to be complemented by random effects capable of capturing the idiosyncratic collocational behavior of the lexical items object of study. Given the great importance of the individual noun collocates of the near-synonymous adjectives, the next chapter zooms in on their idiosyncratic collocational behavior throughout the period 1810–2009 by using different methodological approaches that are specifically geared towards analyses of this kind.

7 IDIOSYNCRATIC COLLOCATIONAL PREFERENCES OF THE NEAR-SYNONYMS

As seen in Section 2.3.2, The notion of collocation constitutes one of the central tenets of the distributional approach from its earliest stages, being emphasized already in several of Firth's most important works (e.g. 1935; 1957a). With the advent of corpora, the notion became even more essential in lexico-semantic analysis and, therefore, research on collocational behavior flourished in this field of study. Similarly, advancements in the methods and statistical techniques used both to measure the degree of association between nodes and their collocates and to represent such associations led to an increasing interest in this phenomenon. In particular, the notion of collocation has figured prominently in research on near-synonymy as one of the types of contextual patterns used to quantify the degree of (dis)similarity between potentially near-synonymous expressions. In fact, many scholars argue that the more contextual ground two lexemes share, including specific collocates, the more similar they are (e.g. Rubenstein & Goodenough 1965; Miller & Charles 1991; Gries 2001; 2003; Gries & Otani 2010; Liu 2010). For instance, Gries (2001; 2003) examines the degree of collocate overlap between pairs of near-synonymous *-ic*/*-ical* adjectives by accounting for the amount of significant collocates they share, as well as the amount of significant collocates one near-synonym exhibits, but not the other.

Due to the broad application of the notion of collocation and the boost it received from the distributional corpus-based approach, it is not surprising that this term has been interpreted in various ways. For instance, Gries (2013: 138–139), in his overview of collocational research over the last 50 years, identifies a series of parameters on which the notion has varied. Among these, he includes the nature of the linguistic items examined (e.g. POS), the frequency threshold for a co-occurring item to be considered a collocate, the distance between the items making up the collocation (i.e. context window), and the relation between the node and the collocate, that is, whether mere adjacency between the items in question suffices or whether the

two items need to be syntactically or semantically related.[107] Such variations in the use of the concept have logically resulted in the proposal of several definitions of the term in the specialized literature, namely the textual, the statistical, and the associative definitions. These definitions are not mutually exclusive and, in fact, in the present chapter both the textual and the statistical definitions are employed, as the latter can be seen as a refinement of the former with the addition of some frequency or associative thresholds.

Most research on near-synonyms reviewed in Chapter 3 analyzed collocational patterns in one way or another, either by considering specific collocates or more schematic classifications, that is, semantic preferences and/or semantic prosody, among others (e.g. Persson 1989; Gries 2001; 2003; Kjellmer 2003; Arppe & Järvikivi 2007; Primahadi-Wijaya-R & Rajeg 2014). The great majority of these studies are of a synchronic nature, thus measuring the contextual overlap of near-synonyms in, most commonly, PDE (e.g. Persson 1989; Kjellmer 2003). However, a few studies examine how the collocational preferences of such related words have changed over time, thus offering a fine-grained perspective of how the relation between near-synonyms can vary from one period to another. Two cases in point are Primahadi-Vijaya-R. & Rajeg (2014) on the adjectives *hot* and *warm* and Baker (2017: 95–101) on the prepositional pairs *round*/*around* and *on*/*upon* (cf. Section 3.3 above). In the present dissertation, the synonym set under examination exhibits important diachronic tendencies, with Period emerging as a particularly strong predictor which is involved in significant interactions with Sense and Semantic category, among others (cf. Section 6.2.1). Additionally, the idiosyncratic collocational patterns of the adjectives also play a crucial role in their choice, as confirmed by the mixed-effects regression analyses in Section 6.2.2. Nevertheless, changes over time in the collocational profiles of the adjectives were not analyzed in depth in Chapter 6, where only their general collocational tendencies were visualized (cf. Figures 24–26). The main aim of this chapter is therefore to determine whether these tendencies vary diachronically, much as in the case of the semantic preferences of the near-synonymous adjectives uncovered in the previous two chapters.

The structure of this chapter mirrors that of the preceding one.[108] The dataset and the statistical techniques employed are first described in Section 7.1. Then, Section 7.2 focuses on the results of the different collocational analyses, with 7.2.1 providing a bird's-eye view of all

[107] For more information on these parameters, cf. Section 2.3.2.

[108] The results presented in this chapter are partly based on those in Pettersson-Traba (2019; forthcoming).

noun collocates of the near-synonyms and 7.2.2 zooming in on just the most prominent ones. Finally, Section 7.3 offers a summary and a discussion of the main findings.

7.1 DATASET AND METHODOLOGY

Contrary to the first dataset which served as the basis for the analyses in Chapters 5 and 6, the second dataset does not comprise the actual instances of the near-synonymous adjectives, but instead their noun collocates in an L5–R5 context window (cf. Section 4.2). The collocates of the adjectives were extracted automatically from *COHA* by making use of the Collocates and POS-tag options available in the corpus. Additionally, in order to be able to provide the diachronic development of the collocational behavior of the adjectives, their collocates were retrieved per period. This allows us to compare (i) the collocational patterns of each adjective across periods (e.g. the collocates of *fragrant* in P1 with those of *fragrant* in P2) and (ii) the collocational patterns of the different adjectives (e.g. the collocates of *fragrant* in P1 with those of *scented* in P1). Table 26 displays six of the rows in the database to provide an overview of its structure. For instance, the lemma *afternoon* occurs in *COHA* as a collocate of *fragrant* zero times in P1, 3 in P2, and so on and so forth. Similarly, this particular noun only collocates with *perfumed* once, in P3. As further explained below, the adjectives *sweet-scented* and *sweet-smelling* were excluded also from the present analyses for similar reasons to those put forward in Section 6.1 above. The retrieval process resulted in a total of 10,740 tokens and 2,682 types of noun collocates, which were then submitted to two types of analyses so as to measure the (dis)similarities of the collocational preferences of the near-synonymous adjectives and, consequently, identify their semantic structure.

Table 26. Structure of the dataset

Node / Collocate	*Fragrant*				*Perfumed*				*Scented*			
	P1	**P2**	**P3**	**P4**	**P1**	**P2**	**P3**	**P4**	**P1**	**P2**	**P3**	**P4**
afternoon	0	3	3	0	0	0	1	0	0	0	1	0
age	0	1	0	0	0	0	0	0	0	0	0	0
air	33	74	44	34	11	28	9	7	8	17	15	21
aisle	0	3	0	1	0	0	0	0	0	0	0	0
alabaster	2	0	0	0	0	0	0	0	0	0	0	0
alcohol	0	0	0	0	0	1	0	0	0	0	0	0

First, the data was analyzed by means SVS modeling.[109] This technique, which was first used in computational linguistics, is specifically geared towards the comparison of the contexts of words or word senses. It is therefore a particularly suitable method to measure the semantic (dis)similarity between related words, such as near-synonyms, on the basis of their collocational profiles. Different types of SVS analyses exist depending on how context and the notion of collocation are conceived. If one adopts a more textual definition where all neighboring words in a specific context window (e.g. L3–R3 or L5–R5) are considered collocates, a so-called *bag-of-words* SVS model is applied (Peirsman, Heylen & Geeraerts 2008: 34–35; Levshina 2015: 323–324). If, on the contrary, only those context words which exhibit a specific syntactic relation with the node are accounted for, a so-called *dependency-based* SVS model is conducted (Padó & Lapata 2007; Heylen, Speelman & Geeraerts 2012: 18).

Additionally, another distinction is made between *type*- and *token-based* SVS. The former averages "collocate frequencies over many occurrences of the same word […]", whereas the latter "operates at the level of word tokens, thus capturing meaning differences between individual occurrences of the same word" (Hilpert & Correia Saavedra 2017: 2; cf. also Heylen, Speelman & Geeraerts 2012). The dataset analyzed here and illustrated in Table 26 is an example of a data sample used for type-based SVS. This is so because the main units of analysis are the word types or lemmas *fragrant*, *perfumed*, and *scented* and the frequencies of occurrence of their collocates are averaged over all the instances of these lemmas in the corpus. If a token-based SVS were to be computed, then the columns in the database would not be the lemmas *fragrant*, *perfumed*, and *scented*, but rather each individual example of these words in *COHA* and the frequency of the collocates in each specific instance would have been considered.

These various types of SVS, as well as the context windows selected, have different advantages and disadvantages, some of them being more appropriate for the study of specific semantic phenomena (for more information, cf., for instance, Padó & Lapata 2007; Peirsman 2008; Peirsman, Heylen & Geeraerts 2008; 2010; Kiela & Clark 2014). Thus, for instance, bag-of-words models are especially useful to uncover issues connected to particular discourse contexts and to semantic prosody (Peirsman, Heylen & Geeraerts 2010). On the contrary, dependency-based models are claimed to perform better at measuring semantic relations between words or word senses (Padó & Lapata 2007; Peirsman 2008). Moreover, as mentioned

[109] SVS is sometimes also referred to by the term *word space models* (e.g. Peirsman 2008; Peirsman, Heylen & Geeraerts 2010; Heylen et al. 2015).

in Section 4.2, the size of the context window employed in bag-of-words models also affects the results of SVS modeling: while tighter windows of one or two words to the left and to the right typically retrieve near-synonyms or antonyms of the target word, somewhat looser windows (e.g. L5–R5) are more suitable to retrieve the most typical collocates of the node (Peirsman, Heylen & Geeraerts 2008: 40). In the analyses conducted in the present chapter, the results of a restricted bag-of-words SVS with an L5–R5 context window are provided, focusing exclusively on the noun collocates of the adjectives. The decision to include only noun collocates was justified in Section 4.2 on the basis of findings of previous studies, which suggest that nouns are more informative than other POS when the semantic structure of adjectives is concerned (Geeraerts 1986; Justeson & Katz 1995; Gries 2001; 2003). Therefore, other content words (e.g. adjectives and verbs) and function words (e.g. prepositions and pronouns) were excluded. By means of illustration, in example (115) the square brackets delimit the L5–R5 collocation window and, within this span, just *air* and *breath* are considered collocates of *fragrant* as they are the only nouns that appear within the limits of the square brackets.

(115) He drew it forth, and read the blessed name of Hermione, and the city beneath him vanished **[**away, and the **air** grew **fragrant** as with the **breath** of**]** May-flowers, and a light streamed through the shadowy forest […]. (*COHA*, 1839, FIC, Hyperion)

Nevertheless, it must be noted that two additional SVS analyses were also conducted to assess their suitability for the present data. First, the contribution of non-nominal content words was evaluated by including adjective, adverb, and verb, as well as noun, collocates of the adjectives. Despite the fact that a minimal improvement in the quality of the analysis was achieved when all content words were considered, the resulting patterns were not as elucidating as the ones obtained on the basis of the restricted SVS with only noun collocates. Second, a dependency-based SVS model was also computed by considering only those nouns that are semantically modified by the adjectives (e.g. *flower* in *the fragrant flower* or *the flower is fragrant*). In this case, however, the resulting SVS model performed much worse than the one with all noun collocates.[110] Furthermore, as explained above, a type-based, as opposed to a token-based, SVS was carried out here. Token-based SVS typically results in data sparseness, given that the database contains many more cells with zero than that for a type-based SVS. To

[110] For a comparison of the goodness-of-fit statistics of these three models, cf. Table C.1 in Appendix C.

solve this problem, token-based SVS models resort to *second-order collocates*, that is, the collocates of the collocates of the node. This means that one would first search for the collocates of *fragrant* in each instance of this adjective in *COHA* (e.g. *flower* in *the flower is fragrant*) and then for the collocates of *flower*. This is indeed an interesting approach, but retrieving second-order collocates in a time-sensible manner requires access to the whole corpus. In addition, token-based SVS analysis is more typically employed to trace the existence of different senses (i.e. polysemy) within the same target word(s) in a data-driven fashion, whereas type-based SVS is more adequate when working with near-synonyms or other semantically related words (Heylen et al. 2015: 155). Given that the focus here is on the degree of (dis)similarity of the adjectives rather than on the existence of different senses (e.g. natural, artificial, figurative), which was already established for this synonym set on the basis of reference materials in Section 4.1, type-based SVS modeling was chosen.

Regardless of the type of SVS one wants to conduct, a similar series of steps is typically followed to construe the models (cf., for instance, Levshina 2015: 326; Hilpert & Correia Saavedra 2017: 4–9):

(i) A table containing the raw co-occurrence frequencies of the nodes (in the columns) and the context words (in the rows) is created. Here, a separate column was included for each near-synonym in each period (cf. Table 26), since we are interested in the evolution of their collocational preferences over time.

(ii) The raw co-occurrence frequencies are transformed into collocational strength values using an association measure such as PMI (e.g. Church & Hanks 1990: 23). Although a variety of association measures exist to compute collocational strength, such as log-likelihood or ΔP (e.g. Gries 2013), PMI is used in the present analysis. The main reason for choosing this association measure is that it is the standard one used in SVS analysis, since it works well with low co-occurrence frequencies (Levshina 2015: 327).

(iii) Finally, cosine similarity scores are computed (Levshina 2015: 328) between the collocational behavior of the node words (i.e. *fragrant*, *perfumed*, and *scented* in each period). Cosine values range between 0 and 1: the higher the value between two nodes (e.g. ‘fragrant P1’ and ‘fragrant P2’), the higher their degree of similarity. The final result is a similarity matrix, shown in Table 27:

Table 27. Similarity matrix based on cosine similarity scores for the data

		Fragrant				Perfumed				Scented			
		P1	**P2**	**P3**	**P4**	**P1**	**P2**	**P3**	**P4**	**P1**	**P2**	**P3**	**P4**
Fragrant	**P1**	1	0.76	0.59	0.53	0.83	0.47	0.34	0.37	0.71	0.58	0.53	0.47
	P2	0.76	1	0.71	0.42	0.43	0.57	0.34	0.36	0.48	0.48	0.39	0.36
	P3	0.59	0.71	1	0.79	0.32	0.45	0.39	0.40	0.43	0.42	0.70	0.49
	P4	0.53	0.42	0.79	1	0.25	0.36	0.43	0.56	0.20	0.40	0.42	0.45
Perfumed	**P1**	0.83	0.43	0.32	0.25	1	0.87	0.50	0.51	0.95	0.65	0.49	0.23
	P2	0.47	0.57	0.45	0.36	0.87	1	0.76	0.60	0.83	0.84	0.66	0.41
	P3	0.34	0.34	0.39	0.43	0.50	0.76	1	0.85	0.40	0.30	0.83	0.33
	P4	0.37	0.36	0.40	0.56	0.51	0.60	0.85	1	0.35	0.54	0.95	0.73
Scented	**P1**	0.71	0.48	0.43	0.20	0.95	0.83	0.40	0.35	1	0.78	0.61	0.22
	P2	0.58	0.48	0.42	0.40	0.65	0.84	0.30	0.54	0.78	1	0.59	0.42
	P3	0.53	0.39	0.70	0.42	0.49	0.66	0.83	0.95	0.61	0.59	1	0.66
	P4	0.47	0.36	0.49	0.45	0.23	0.41	0.33	0.73	0.22	0.42	0.66	1

The value in each cell in Table 27 represents the degree of similarity between a node in the columns and another one in the rows. For instance, the degree of similarity between 'fragrant P1' and 'fragrant P2' is 0.76. The series of cells with a value of 1 signal the degree of similarity of each node with itself. As mentioned above, a value of 1 represents complete collocate overlap between two nodes, which only occurs when a node is compared with itself. The diagonal line formed by the cells with a value of 1 separates the matrix into two identical halves, with the top right half reduplicating the values of the bottom left half.

The similarity scores obtained in this third and last step of the process were subsequently fed into cluster analysis and Multidimensional Scaling (MDS), two different dimensionality reduction techniques used to explore and interpret the (dis)similarities between the adjectives in the different periods. For easier interpretability of the resulting cluster and MDS analyses, the similarity matrix was first transformed into a distance matrix by subtracting each of the similarity scores in Table 27 from 1. The distance matrix is shown in Table 28: in this case, distance values go from 0 to 1, and the higher the value between two nodes, the larger the semantic distance —in collocational terms— between them. In this case, the diagonal line of zero crossing Table 28, which stands for absolute similarity in terms of collocational behavior, divides the matrix into two identical halves.

Table 28. Distance matrix for the data

		Fragrant				Perfumed				Scented			
		P1	**P2**	**P3**	**P4**	**P1**	**P2**	**P3**	**P4**	**P1**	**P2**	**P3**	**P4**
Fragrant	**P1**	0	0.24	0.41	0.47	0.17	0.53	0.66	0.63	0.29	0.42	0.47	0.53
	P2	0.24	0	0.29	0.58	0.57	0.43	0.66	0.64	0.52	0.52	0.61	0.64
	P3	0.41	0.29	0	0.21	0.68	0.55	0.61	0.60	0.57	0.58	0.30	0.51
	P4	0.47	0.58	0.21	0	0.75	0.65	0.57	0.44	0.80	0.60	0.58	0.55
Perfumed	**P1**	0.17	0.57	0.68	0.75	0	0.13	0.50	0.49	0.05	0.36	0.51	0.77
	P2	0.53	0.43	0.55	0.65	0.13	0	0.24	0.40	0.17	0.16	0.34	0.59
	P3	0.66	0.66	0.61	0.57	0.50	0.24	0	0.15	0.60	0.70	0.17	0.67
	P4	0.63	0.64	0.60	0.44	0.49	0.40	0.15	0	0.65	0.46	0.05	0.27
Scented	**P1**	0.29	0.52	0.57	0.80	0.05	0.17	0.60	0.65	0	0.22	0.39	0.78
	P2	0.42	0.52	0.58	0.60	0.35	0.16	0.70	0.46	0.22	0	0.41	0.58
	P3	0.47	0.61	0.30	0.58	0.51	0.34	0.17	0.05	0.39	0.41	0	0.34
	P4	0.53	0.64	0.51	0.55	0.77	0.59	0.67	0.27	0.78	0.58	0.34	0

Cluster analysis is a group of methods that helps to identify groups or clusters of objects on the basis of their (dis)similarities (Levshina 2015: 301). The objective of such methods is therefore to create groups comprising objects that are maximally similar to one another and maximally dissimilar to those objects in other groups, all this on the basis of the data at hand, in this case the distance matrix displayed in Table 28. Here, Hierarchical Agglomerative Clustering (HAC) analysis is used, which organizes the data into clusters and subclusters in a tree-like structure by first considering each individual node (here, 'fragrant P1', 'fragrant P2', etc.) as one separate cluster, then continually merging those two which are most similar until there is only one macro cluster containing all nodes (Steinbach, Tan & Kumar 2005: Chapter 8). MDS, in turn, is an approach that depicts differences between two or more objects as distances in a two- or three-dimensional plot (Levshina 2015: 336–337). The input for this test is again the distance matrix shown in Table 28 above and the output consists in a spatial representation of the nodes so that the more similar two nodes are, the closer they are located to each other in this spatial configuration. The goal of MDS is to shed light on the structure of the data in an easily understandable and visual manner (Kruskal & Wish 1978: 7; Wickelmaier 2003: 4). The two methods, HAC and MDS, are used here since they complement each other: HAC is useful to identify coherent and delimited groups in the data, while MDS can be said to be semantically more realistic in the sense that it does not imply a categorical either/or split

between individual clusters, but instead displays a more precise and continuous cline of semantic (dis)similarity (Jansegers & Gries 2017: 14).

The second type of analysis to which the data was submitted was collocational network analysis, which supplements the findings of SVS modeling. This is so because, as argued by Heylen et al. (2015: 154–155), "the technique [SVS] is too much of a black box to be suitable for in-depth lexical analysis […]". In other words, SVS results in a series of distance values, but it averages over a great amount of individual collocates so that it is not always easy to determine exactly where the (dis)similarities in the data reside. The idea of visualizing the associations between nodes and their collocates by means of networks was first proposed by Philips (1985; 1989) and has since then been put into practice and further refined by numerous scholars (e.g. Baker 2005; 2017: 95–101; McEnery 2006a; 2006b; Brezina, McEnery & Wattam 2015). Put simply, collocational networks are built by automatically selecting the collocates of the node words, usually the most prominent ones according to some statistical measure of collocational strength (e.g. PMI, ΔP, or log-likelihood), and then visualizing them in the form of a network. In this network, collocates are connected to their nodes by means of arrows, with the length of the arrow indicating the strength of the association between them: the shorter the arrow, the stronger the association.

Collocational networks have been used for various purposes in the specialized literature, ranging from discourse analysis (e.g. McEnery 2006a; 2006b; Brezina, McEnery & Wattam 2015) to historical linguistics (Baker 2017: 95–101). In particular, Baker's study on the competing prepositional forms *on*/*upon* and *round*/*around* described in Section 3.3 is of special relevance for the purposes of the present dissertation, as it constitutes the first application of this method to analyze the development of the collocational behavior of near-synonyms over time. Baker demonstrates that, through collocational networks, one can observe increases or decreases in the frequency of near-synonyms, as well as the competition between them: if one near-synonym draws away collocates from another, this most probably means that it is taking over some of its semantic space. Here, four separate collocational networks were built, one per period, hence enabling us to determine which collocates the near-synonymous adjectives under analysis share at different points in time, but also to identify potential variations in the internal semantic structure of the synonym set. The association measure employed was again PMI and, following the methodology in Baker (2017: 98–100) for low-frequency words, two thresholds were established for a collocate to be considered significant and thus be included in the

networks: a minimum raw frequency of 5 with the node words and a PMI value of 3 or higher in each of the four periods.[111] These thresholds drastically decreased the number of collocates of the near-synonyms in the dataset if compared to the SVS analysis, thus allowing us to conduct a more qualitative and careful examination of the collocational profiles of the adjectives at issue. At this point, it became evident that *sweet-scented* and *sweet-smelling* could not be included in the collocational network analysis, as almost none of their collocates emerged as prominent after applying these two thresholds: only the noun *flower* complied with these criteria as a collocate of *sweet-scented* in P1 and *sweet-smelling* in P2. For this reason, the two adjectives were excluded from the collocational networks and, for the sake of comparability, also from the SVS analysis.[112] Establishing lower thresholds, that is a raw frequency lower than 5 and a PMI lower than 3, would not be a particularly effective solution, as such networks would include collocates which may not be especially prominent. Additionally, as mentioned earlier (cf. Section 2.3.2), a PMI of 3 or higher is often considered to indicate that two items co-occur significantly more often than expected by chance (e.g. Church & Hanks 1990; Church et al. 1991; Church et al. 1994; Liu 2010), so that lowering this threshold would result in non-significant collocates entering the analyses. On the contrary, higher thresholds would lead to even greater data sparseness, particularly in the case of *perfumed* and *scented*, which are, as seen in Section 5.1, not as frequent in *COHA* as *fragrant* (cf. Table 9). Therefore, the thresholds proposed by Baker (2017: 96) for words of a higher frequency, namely a minimum raw frequency of 10 and a PMI of 6 were here discarded. Table 29 shows the settings used to identify collocates, namely the PMI and frequency cut-off values, the context window, and some additional filters. The additional filter column makes it explicit that only noun collocates were considered, thus discarding other content (e.g. adjectives and verbs) and function words (e.g. prepositions and pronouns).

[111] Brezina, McEnery & Wattam (2015: 159) argue that it is useful to apply a frequency threshold when selecting collocates on the basis of PMI scores, given that these scores have "a propensity to highlight unusual combinations […] that occur only once or twice in the corpus" and which are thus not particularly informative.

[112] The most frequent noun collocates of *sweet-scented* and *sweet-smelling* in each period are provided in Tables C.2 and C.3 in Appendix C.

Table 29. Settings for the construction of collocational networks

PMI threshold	Frequency threshold	Context window	Additional filters
3	5	L5–R5	Function words and content words other than nouns removed

Nowadays different tools are available in order to build collocational networks, including *CONE* (Gullick et al. 2010) and *GraphColl* (Brezina, McEnery & Wattam 2015). However, in this dissertation, the networks were created in *R* by means of the `igraph` package (Csardi & Nepusz 2006). The reasons for this decision were twofold: (i) loading *COHA* into a program like *GraphColl* required a lot of computation time due to its large size and (ii) *R* allows greater flexibility in the customization of the visual parameters of collocational networks.

7.2 Results

As mentioned in the introductory section to this chapter, the results of the two methods employed to examine the idiosyncratic collocational behavior of the near-synonyms *fragrant*, *perfumed*, and *scented*, to wit, SVS modeling and collocational network analysis, are presented separately in what follows. First, in Section 7.2.1 the findings of SVS are provided through the two different visualizations techniques selected here, i.e. HAC and MDS. Then, Section 7.2.2 presents the most prominent collocates of each adjective by means of collocational networks.

7.2.1 Semantic Vector Space modeling

Figure 28 displays the distance between the near-synonymous adjectives across periods according to HAC in the form of a dendrogram: the higher two nodes are merged, the more dissimilar they are. The grey rectangles indicate the optimal number of clusters on the basis of the average silhouette width of this solution, an index which serves as an indicator of the internal coherence of the clusters. This index ranges from 0 to 1, with values higher than 0.2 signaling that clusters are well-formed (Levshina 2015: 311).



Figure 28. Dendrogram of the collocational preferences of *fragrant*, *perfumed*, and *scented* throughout the period 1810–2009

A HAC solution with five clusters obtains the highest average silhouette width, namely 0.49. The five clusters are the following (from left to right in Figure 28):

(i) Cluster 1: *perfumed* and *scented* in P1 and P2.
(ii) Cluster 2: *scented* in P4.
(iii) Cluster 3: *perfumed* in P3 and P4 and *scented* in P3.
(iv) Cluster 4*: fragrant* in P3 and P4.
(v) Cluster 5: *fragrant* in P1 and P2.

The stability of the clusters obtained can be estimated by computing their Approximately Unbiased (AU) *p*-values, which determine how well supported they are by the data and thus also the chances of achieving the same clusters if a new data sample was to be analyzed. AU *p*-values range from 0 to 1 and the closer they are to 1, the more stable the clusters (Levshina 2015: 315–317). The AU *p*-values for the five clusters in Figure 28 are shown in Table 30:

Table 30. AU *p*-values for the five-cluster solution in Figure 28

Cluster	AU *p*-value
1 (*perfumed* and *scented* in P1 and P2)	0.89
2 (*scented* in P4)	0.90
3 (*perfumed* in P3 and P4 and *scented* in P3)	0.93
4 (*fragrant* in P3 and P4)	0.88
5 (*fragrant* in P1 and P2)	0.78

Of the five clusters obtained, none reaches the traditional 0.95 significance threshold, but most of them are close to this value. The least stable cluster is that containing *fragrant* in P1 and P2 (i.e. cluster 5), which implies that if a new data sample was to be examined, these two nodes would possibly not be part of the same cluster. On the other hand, the cluster containing *perfumed* in P3 and P4 and that of *scented* in P3 (i.e. cluster 3) is almost statistically significant and therefore represents the most stable cluster. The remaining three clusters exhibit values that range from those obtained for cluster 5 to those for cluster 3, but they are closer to the latter.

Figure 28 above shows that the collocational profiles of the three adjectives seem to pattern according to period and near-synonym. First, the collocational profiles of the near-synonyms in P1 and P2 are grouped together (i.e. clusters 1 and 5), as well as those in P3 and P4 (i.e. clusters 3 and 4). Second, the collocational profiles of *fragrant* (i.e. clusters 4 and 5) differ from those of *perfumed* and *scented*, which are part of the same clusters (i.e. clusters 1 and 3). The only exception to this pattern is the collocational behavior of *scented* in P4, which forms its own cluster (i.e. cluster 2) and is thus kept apart from the nodes 'perfumed P3', 'perfumed P4', and 'scented P3' (i.e. cluster 3).

A semantically more realistic picture with no categorical either/or splits is provided by the MDS analysis of the data. Figure 29 plots a two-dimensional MDS solution of the collocational preferences of *fragrant*, *perfumed*, and *scented* across periods. This figure should be interpreted in the following way: each separate node in the graph represents the collocational behavior of one of the near-synonyms in a specific period, and the closer two nodes are located on the MDS map, the more similar their collocational preferences. The quality of the MDS solution can be examined by computing the so-called stress value (Levshina 2015: 341): the smaller the stress, the better the performance of the model. As a rule of thumb, values higher than 0.2 indicate that the model can be further improved. The stress of the MDS solution visualized in Figure 29 is 0.21, indicating an almost acceptable fit. A slightly better fit is obtained by a three-dimensional

MDS solution (i.e. stress = 0.16). However, only a two-dimensional solution is discussed here, since it allows for a much clearer and easily interpretable visualization of (dis)similarities in the collocational behavior of the three adjectives (cf. Szmrecsanyi & Kortmann 2009: 1650; Szmrecsanyi 2017: 351 for a similar approach).[113]

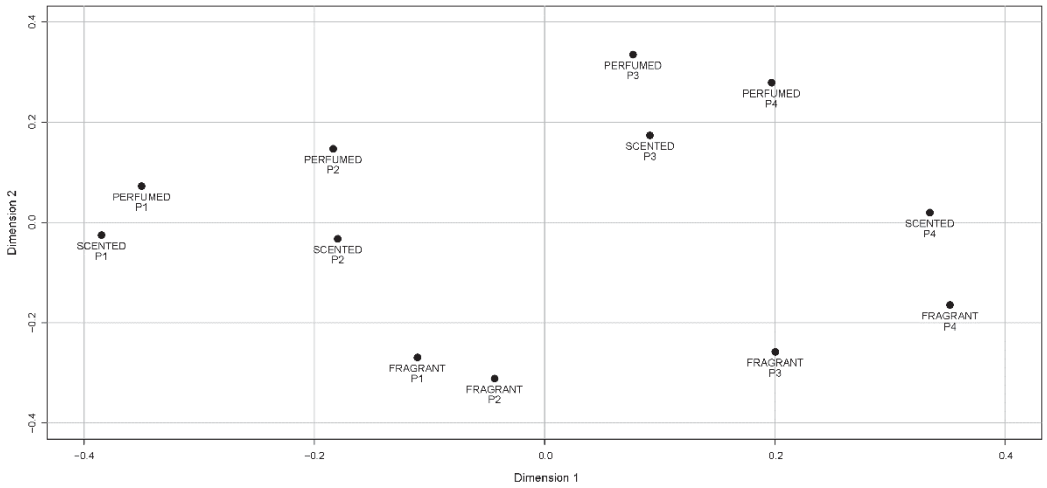


Figure 29. Two-dimensional MDS map of the collocational preferences of *fragrant*, *perfumed*, and *scented* throughout the period 1810–2009

The vertical axis (Dimension 2) in Figure 29 organizes the nodes according to near-synonym, with *perfumed* being located at the top of the figure and *fragrant* at the bottom. *Scented*, in turn, occupies the middle ground but, overall, it is much closer to *perfumed* than to *fragrant*, with the exception of P4. The horizontal axis (Dimension 1), on the contrary, organizes the nodes on a temporal continuum from left to right, with P1 and P2 exhibiting negative values and P3 and P4 exhibiting positive ones. Independent-samples *t*-tests on the MDS coordinates, that is, the values of the nodes on Dimensions 1 and 2, confirm the visual interpretation of Figure 29.[114] This is so because *fragrant* differs significantly from both *perfumed* and *scented* with respect to their positions on the vertical axis (t = -6.8, df = 4.49, p < 0.01 and t = -4.98, df = 5.12, p < 0.01, respectively), but the difference between *perfumed* and *scented* is only marginally significant (t = 2.27, df = 5.72, p = 0.06). However, the

[113] Although only a slightly better fit is achieved with a three-dimensional MDS, more variation is of course captured by this solution. For the three-dimensional MDS map, cf. Figure C.1 in Appendix C.

[114] This test is used to compare the means of two groups of datapoints, here the mean MDS coordinates of the near-synonyms across periods (e.g. *fragrant* vs. *perfumed*), on the one hand, and the mean MDS coordinates of the periods across near-synonym (e.g. P1 vs. P2), on the other (Levshina 2015: 87).

differences between the near-synonyms along the horizontal axis are not significant. In fact, on this axis, where the nodes are organized according to period, P1 diverges significantly from P3 and P4 ($t = 4.27$, df = 2.78, $p < 0.05$ and $t = 5.81$, df = 3.17, $p < 0.01$, respectively), P2 diverges significantly from P3 and P4 ($t = 4.28$, df = 3.89, $p < 0.05$ and $t = 6.4$, df = 3.98, $p < 0.01$, respectively), but the differences between P1 and P2, on the one hand, and between P3 and P4, on the other, are not significant ($t = 1.50$, df = 3.06, $p = 0.23$ and $t = 2.74$, df = 3.81, $p = 0.055$, respectively). In turn, the differences among the periods on the vertical axis do not reach statistical significance. This translates into the following: concerning their collocational profiles, whereas the near-synonyms seem to be relatively similar in P1 and P2, on the one hand, and in P3 and P4, on the other, there is a substantial change from P1/P2 to P3/P4, that is, at the turn of the 19th to the 20th century. This finding is in line with those obtained in the previous two chapters, where important changes were also discovered at this particular point in time.

The two-dimensional MDS solution of the SVS analysis displayed in Figure 29 allows us to identify differences in collocational behavior between, on the one hand, *fragrant* and the other two adjectives and, on the other, all three near-synonyms in the 19th century versus the 20th and 21st centuries. However, as mentioned above, Heylen et al. (2015: 154–155) argue that SVS is "too much of a black box" to clarify the nature of the (dis)similarities uncovered. Therefore, in order to shed some light on such (dis)similarities, further analyses are required. Before turning to the more qualitative analysis carried out in Section 7.2.2 by means of collocational networks, in the remainder of the present section, quantitative techniques are applied to the results of the SVS to identify what the Dimensions (i.e. axes) in the MDS plot in Figure 29 stand for. To this end, the 2,682 types of noun collocates identified in Section 7.1 were classified into semantic categories by resorting to *USAS* (cf. Section 4.3).[115] Following *USAS* semantic classification for each near-synonym in each period (e.g. 'fragrant P1', 'fragrant P2'), a PMI score signaling their collocational strength with each semantic category and sub-category in *USAS* was calculated. This was done with the aim of determining whether correlations exist between the MDS coordinates of the nodes and their PMI scores with the

[115] Given the data-driven nature of SVS analysis, the semantic classification of the noun collocates in this chapter was carried out automatically using *USAS*, instead of the *HTOED*. As mentioned in Section 4.3, despite being less precise, *USAS* allows researchers to automatically analyze strings of words according to their semantics and thus check the domains to which particular words belong

semantic categories and sub-categories. The significance of correlations was assessed by means of Pearson′s product-moment correlation tests (Levshina 2015: 116–130). Besides a significance value, this test provides a correlation coefficient, represented by the letter *r*, which ranges from 0 to 1 and indicates the strength of the correlation: the higher the value, the stronger the correlation. Positive coefficients signal a positive correlation, which means that the values of the tested variables, here the MDS coordinates of the nodes and their PMI scores with the *USAS* semantic categories, jointly decrease and increase. In other words, if the values of one variable increase, so do the values of the other variable, and vice versa. On the contrary, negative coefficients signal a negative or inverse correlation, implying that the values go in opposite directions: if the values of one variable increase, the values of the other decrease, and vice versa. The eight correlations shown in Table 31 emerged as significant:

Table 31. Significant correlations between the dimensions in Figure 29 and semantic (sub-)categories in *USAS*

MDS dimension	*USAS* semantic (sub-)category	Direction of correlation	Correlation coefficient (*r*)	*p*-value
Dimension 1	F4: FARMING AND HORTICULTURE	negative	-0.63	< 0.05
Dimension 1	W4: WEATHER	negative	-0.58	< 0.05
Dimension 2	B: THE BODY AND THE INDIVIDUAL	positive	0.88	< 0.001
Dimension 2	L3: PLANTS	negative	-0.54	0.06
Dimension 2	M3: MOVEMENT/TRANSPORTATION (LAND)	negative	-0.53	0.07
Dimension 2	Q1: COMMUNICATION	positive	0.77	< 0.01
Dimension 2	Q4: THE MEDIA	positive	0.69	< 0.05
Dimension 2	S2: PEOPLE	positive	0.80	< 0.01

Given that some of the semantic categories in Table 31 differ from those used in previous chapters, examples of nouns belonging to these categories that exhibit a significant correlation with any of the MDS dimensions are shown in Table 32:

Table 32. Examples of types of noun collocates in the semantic (sub-)categories in *USAS*

Semantic (sub-)category	Examples of types of noun collocates
B: THE BODY AND THE INDIVIDUAL	B1: *hand*, *hair*, *neck*, *shoulder*, *wrist* B4: *detergent*, *perfume*, *lotion*, *shampoo*, *soap* B5: *clothes*, *handkerchief*, *garment*, *glove*, *lace*
F4: FARMING AND HORTICULTURE	*crop*, *farm*, *farmhouse*, *field*, *mow*, *pasture*, *plantation*, *vineyard*
L3: PLANTS	*cactus*, *garden*, *magnolia*, *pollen*, *pumpkin*, *root*, *vine*
M3: MOVEMENT/TRANSPORTATION (LAND)	*avenue*, *path*, *pathway*, *road*, *wayside*
Q1: COMMUNICATION	*draft*, *leaflet*, *letter*, *message*, *note*, *notepaper*, *stationery*
Q4: THE MEDIA	*article*, *book*, *edition*, *newspaper*, *paper*
S2: PEOPLE	*boy*, *child*, *gentleman*, *madam*, *maiden*, *man*, *people*, *person*, *woman*
W4: WEATHER	*breeze*, *flood*, *fog*, *haze*, *mist*, *rain*, *snowfall*, *storm*, *wind*

First, regarding the behavior of the three near-synonymous adjectives across periods, that is, Dimension 1 (horizontal axis in Figure 29), two semantic categories exhibit significant negative correlations with the MDS coordinates, namely F4: FARMING AND HORTICULTURE and W4: WEATHER. This means that the higher the value of a node on Dimension 1, the lower its PMI score with these categories. Therefore, in P3 and P4 the adjectives occur with F4 and W4 nouns significantly less frequently than in P1 and P2. This change could again perhaps constitute a reflection of the transformation of American society, evolving from a preindustrial and mainly rural society in the 19th century to the world's major industrial and commercial power in the 20th and 21st centuries (cf. Section 5.2.3). As a result of this process of modernization, the importance of farming and horticultural activities in the daily life of most American citizens has surely drastically declined over time, particularly after the First and Second Industrial Revolutions of the 19th century, which resulted in mass migration from the countryside to the cities. In a more indirect way, the importance of weather phenomena may also have decreased with time, as farming and horticultural activities rely greatly on the climate.

Second, the MDS coordinates on Dimension 2 (vertical axis in Figure 29) are negatively correlated with the PMI scores of the near-synonyms and the semantic categories L3: PLANTS and M3: MOVEMENT/TRANSPORTATION (LAND). This implies that the adjectives collocate with nouns belonging to these two categories more often when their values on MDS Dimension 2

are lower. Such nouns are therefore more strongly associated with *fragrant* than with *perfumed* and *scented*. In all likelihood, nouns belonging to the categories L3 and M3 appear mostly in contexts where the near-synonyms are used in the natural sense. This is particularly true of L3 nouns, which refer to plants in general (e.g. *flower*, *shrub*, and *tree*), particular species of flowers or trees (e.g. *azalea*, *cedar*, and *geranium*), and different parts of plants (e.g. *bud*, *leaf*, and *stem*), as illustrated in the following two examples:

(116) The night was clamorous with frogs and katydids, warm, **fragrant** with **cedar**, clear like day, the moon powdering the trees in light as starkly white as the dust of bone. (*COHA*, 1967, FIC, WhenSheWasGood)

(117) At 10 in the morning, the air is still so moist that the hairy **leaves** of the **scented geraniums** are silvered and shining. (*COHA*, 1991, NEWS, CSMonitor)

On the contrary, (sub-)categories B: THE BODY AND THE INDIVIDUAL, Q1: COMMUNICATION, Q4: THE MEDIA, and S2: PEOPLE are positively correlated with MDS Dimension 2. Therefore, nouns belonging to these semantic classes co-occur more frequently with the adjectives when their values on the vertical axis are higher. This means that B, Q1, Q4, and S2 nouns are more strongly associated with *perfumed* and *scented* than with *fragrant*. A large proportion of the occurrences of these nouns most probably correspond to uses of the near-synonyms in the artificial sense, particularly in the case nouns in the sub-categories B4: CLEANING AND PERSONAL CARE and B5: CLOTHES AND PERSONAL BELONGINGS, which amount to 26 and 66 of the types in category B (49.2% of the total). The same is true of nouns in the sub-categories Q1 and Q4. On the contrary, nouns belonging to category B1: ANATOMY AND PHYSIOLOGY, the most recurrent sub-class in category B (76 types; 40.6%), and those in S2: PEOPLE are more ambiguous, as it is not always clear whether the smell referred to by the adjectives *fragrant*, *perfumed*, and *scented* in such contexts is either natural or artificial. As a matter of fact, as shown in Section 4.3.1.2 above, a relatively large amount of such nouns was classified as indeterminate cases. Despite this high degree of ambiguity, many occurrences of the adjectives with B1 and S2 noun collocates do correspond to the artificial sense of the adjectives, as is the case in example (118), where the smell comes from the artificial perfumes used to impregnate the locks of the fair women mentioned in the extended context:

(118) This is a great improvement on the Persian poets who go into raptures over the **fragrant locks** of fair women, not for their inherent sweetness, however, but for the artificial perfumes used by them, including the disgusting musk! (*COHA*, 1899, NF, PrimitiveLoveLove)

Figure 30 visualizes the significant correlations between the MDS dimensions and semantic (sub-)categories in *USAS* just discussed:

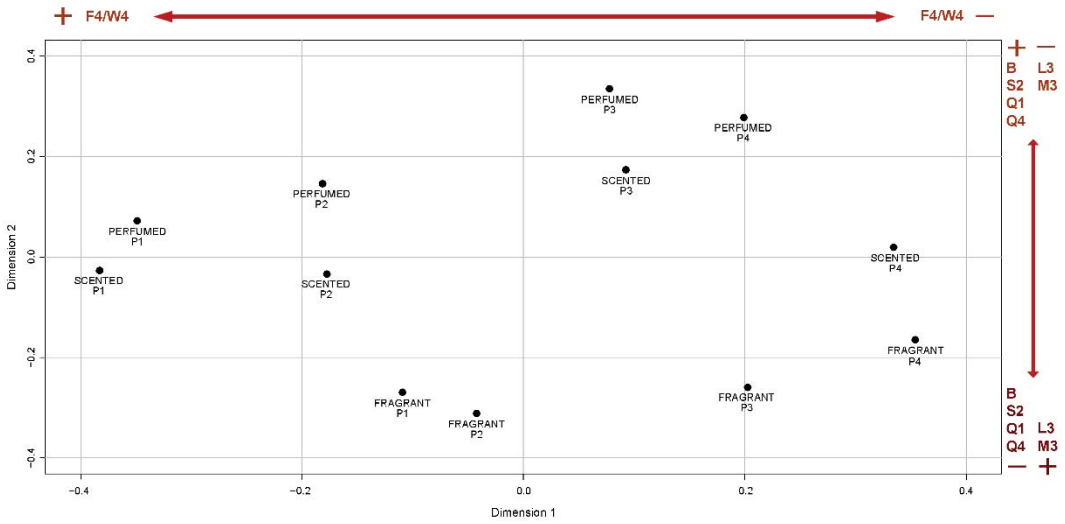


Figure 30. Significant correlations between MDS dimensions and semantic (sub-)categories in *USAS*

The diachronic development of the collocational behavior of the three near-synonymous adjectives across the four time periods, as shown by both the HAC solution and the MDS plot in Figures 28 and 29, respectively, can be summarized as follows.

In P1, the collocational preferences of *perfumed* and *scented* are very similar, while *fragrant* differs considerably from them and does not change much from P1 to P2. Therefore, in P1 and P2, *perfumed* and *scented* are semantically closer to one another than either of them is to *fragrant*, as they seem to share a great deal of collocates and, therefore, considerable semantic space. This can be observed in the HAC solution and MDS map since, *perfumed* and *scented* are clustered together according to the former and are located in close proximity of one another on the latter. On the contrary, *fragrant* is part of a different cluster and is further removed from the other two adjectives in the MDS solution.

However, a substantial development is identified from P2 to P3 in the collocational behavior of the adjectives, as they all move considerably towards the right-hand side of the MDS map.

This change is also reflected in the dendrogram in Figure 28, given that the P3 components of the adjectives are kept apart from their P1 and P2 components. Despite these changes, *perfumed* and *scented* are still closer to each other in P3 than they are to *fragrant*.

The collocational profiles of *fragrant* and *perfumed* remain rather stable from P3 to P4 but, contrariwise, that of *scented* does indeed change. In fact, in P4 *scented* moves away from *perfumed* and occupies an intermediate position on the MDS map between *perfumed* and *fragrant*, which are located at opposite ends of Dimension 2 (i.e. the vertical axis). This change is again reflected in the dendrogram, given that, contrary to previous periods, *scented* in P4 does not form part of the same cluster as *perfumed*.

As in the analyses of the previous two chapters, the results of SVS again indicate that most changes seem to take place from P2 to P3, that is, in the transition from the 19th to the 20th centuries. The changes identified here at the level of individual collocates seem to replicate the tendencies uncovered in Chapters 5 and 6. This is so because, as shown by the interpretation of the MDS dimensions and their correlation with specific *USAS* semantic (sub-)categories, changes along the vertical axis seem to correspond to changes in the frequency with which the adjectives are used in different senses. Again, *perfumed* and *fragrant* seem to be more closely associated with the artificial and natural senses, respectively, as proved by their positions on opposite ends of MDS Dimension 2, with *perfumed* at the top and *fragrant* at the bottom. In turn, *scented* occupies an intermediate position, with values close to 0 on this dimension in all periods except P3. Therefore, as already suggested in the previous two chapters, *scented* does not show a particularly clear preference for neither the artificial nor the natural sense. Finally, changes along the horizontal axis (i.e. Dimension 1) appear to correspond with some of the transformations of American society explained in more detail in Section 5.2.3, thus pointing to similar conclusions such as those reached in the analyses conducted in Chapter 5.

This bird's-eye view of the evolution of the collocational preferences of *fragrant*, *perfumed*, and *scented* has enabled us to identify potential semantic changes concerning the idiosyncratic collocational patterns of the three near-synonyms. However, given the large number of noun collocates present in the dataset used for the SVS analysis (i.e. 2,682 types and 10,740 tokens), it has not yet been possible to identify the most prominent collocates of each near-synonym in a more qualitative fashion. This is the focus of Section 7.2.2, which zooms in on the top noun collocates of each of the three adjectives.

7.2.2 Collocational networks

The most prominent collocates of *fragrant*, *perfumed*, and *scented*, that is, those occurring at least 5 times and displaying a PMI of 3 or higher in a given period, are visualized in the form of collocational networks in Figures 31–34 below. In these figures, the adjectives are linked to their collocates by means of arrows, and the shorter the arrow, the stronger the degree of association.[116] Additionally, the collocates are shown in tabular form (cf. Tables 34–36) to clarify which ones occur in the different periods with each adjective, with a plus sign (+) indicating the occurrence of a particular collocate. These tables facilitate the task of identifying which collocates occur with the adjectives in all periods, and which ones are lost or gained at some point in time. This, in turn, makes it possible to spot shifts in word meaning or in the productivity of particular senses.

To discuss the findings in this section, the terminology proposed by McEnery & Baker (2017: 25–30) is employed. These authors distinguish between four types of collocates, to wit, *consistent* (cf. also Gabrielatos & Baker 2008), *terminating*, *initiating*, and *transient*. The first type refers to those collocates which are relatively stable and thus appear with a node regularly over a period of time. Here, a collocate is considered consistent when it appears in at least three of the four periods. Initiating collocates, in turn, are words which occur for the first time in an intermediate period and then occur consistently throughout the time span analyzed. Terminating collocates show the opposite trend, i.e. occurring at the beginning of the time span, and disappearing in later periods. Here, those collocates which occur either only in the first two or only in the latter two periods (that is, P1/P2 or P3/P4) are considered terminating and initiating, respectively. Similarly, those which occur either only in P1 or only in P4 are considered potentially terminating and potentially initiating collocates, respectively. Finally, transient collocates are those that occur with a node only for a short period of time without gaining consistency. In this case, transient collocates are those which do not meet the criteria for any of the other three types, occurring in one or two intermediate periods or in a maximum of two non-continuous periods (e.g. in P2 and P4).

Figure 31 plots the collocational network of the three adjectives in P1.

[116] For ease of visualization of the networks, in the case of collocates shared by two or more adjectives (e.g. *flower* in Figure 31), the length of the arrows does not reflect the association strength between the collocates and the adjectives. Tables C.4–C.6 in Appendix C provide the PMI values of the collocates present in Figures 31–34.

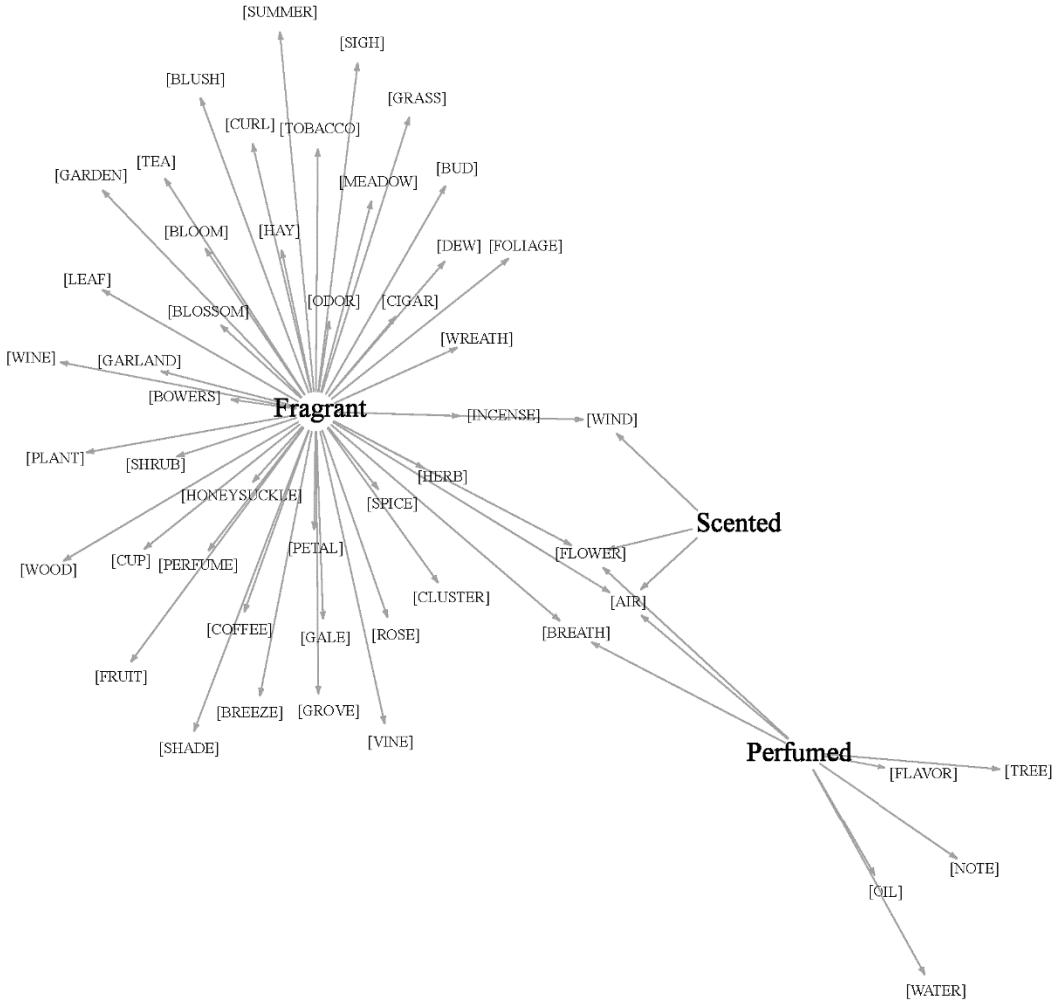


Figure 31. Collocational network of *fragrant*, *perfumed*, and *scented* in P1 (1810–1859)

As shown here, in terms of the number of collocates, *fragrant* clearly dominates (45 collocates) over *perfumed* (8 collocates) and *scented* (3 collocates). As we will see, *fragrant* is the adjective with the largest amount of collocates in all four time periods, a fact which is unsurprising, given the overall much higher frequency of this adjective in *COHA*. On the other hand, *scented* is the item with the lowest number of collocates in P1 and all three prominent collocates of *scented* are shared with either *fragrant* (i.e. *wind*), or both *fragrant* and *perfumed* (i.e. *flower* and *air*). With respect to the semantics of the noun collocates of the near-synonyms in this period, *fragrant* shows a strong natural orientation, being commonly used to describe natural smells. Out of its 45 collocates, a great majority is clearly used to refer to entities which emit a natural smell, such as *bower*, *dew*, *fruit*, *garden*, *petal*, and *plant*. A more detailed picture of the semantics of the nouns with which the adjectives co-occur can be obtained by resorting again to *USAS* and automatically grouping them into semantic (sub-)categories. By doing so, it

becomes evident that out of the 45 collocates of *fragrant* in P1, 14 collocates (almost one third) belong to the semantic (sub)-category L3: PLANTS (e.g. *plant*, *garden*, and *honeysuckle*), 8 to the general category F: FOOD AND FARMING (e.g. *wine*, *spice*, and *fruit*), and another 4 to category W4: WEATHER (e.g. *wind*, *dew*, and *gale*). This means that more than half of the prominent noun collocates of *fragrant* belong to categories corresponding to the natural sense. In fact, of the five collocates with the highest PMI score with *fragrant* in P1 (cf. Table C.4 in Appendix C), four clearly correspond to natural entities (i.e. *honeysuckle*, *herb*, *petal*, and *blossom*), the only exception being the noun *odor*, which can be used in both 'natural' and 'artificial' contexts. Given the low number of instances, no generalizations can be made at this stage from the collocates of *perfumed* and *scented*. The most prominent collocates of these two adjectives are *flavor* and *flower*, respectively. The case of *flavor* was discussed in previous chapters (cf. Sections 5.3.2 and 6.2.2 above), where attention was drawn to the fact that the use of *perfumed* with this noun was restricted to only a few particular texts.

In P2, *fragrant* is still the dominant near-synonym of the set, with 73 collocates (cf. Figure 32).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

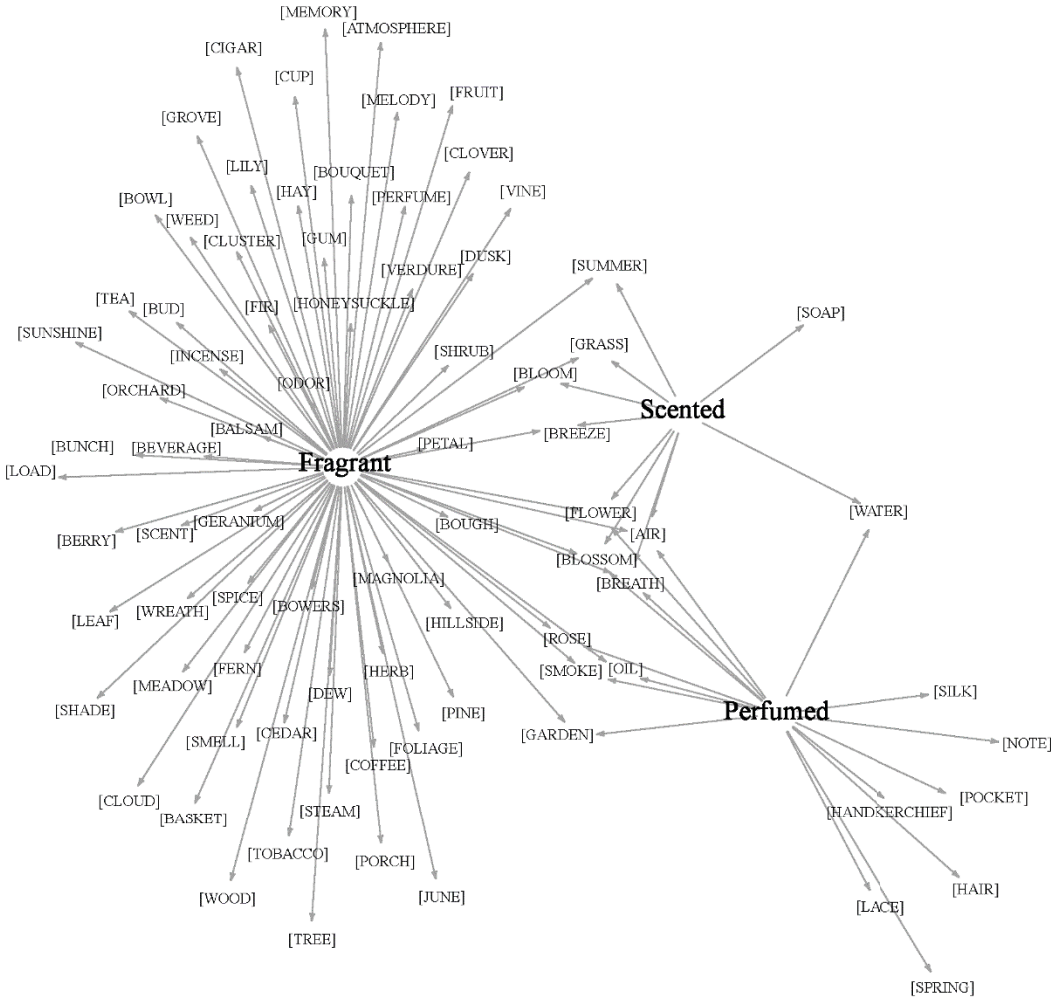


Figure 32. Collocational network of *fragrant*, *perfumed*, and *scented* in P2 (1860–1909)

Both *perfumed* and *scented* increase in number of collocates with respect to P1, with 16 and 10 collocates, respectively, but they are still much less productive than *fragrant*. As in P1, *fragrant* again exhibits a clear natural orientation, with 54 'natural' noun collocates (e.g. *clover*, *foliage*, *herb*, *leaf*, and *vine*). As in the previous period, there is a strong correspondence between the noun collocates of this adjective and the (sub-)categories L3 (23 types), F (8 types), and W4 (4 types). In fact, we witness an increase in nouns belonging to L3 in this period, with collocates such as *pine*, *lily*, and *clover* emerging also as prominent collocates of *fragrant*. Additionally, its collocates now seem to be somewhat more varied, including also nouns belonging to other (sub-)categories, such as X3.5: SENSORY (SMELL) (e.g. *scent* and *smell*) and O1: SUBSTANCES AND MATERIALS GENERALLY (e.g. *steam* and *smoke*). Nouns belonging to the artificial sense, for instance, B4: CLEANING AND PERSONAL CARE and B5: CLOTHES AND PERSONAL BELONGINGS, on the contrary, do not often co-occur with *fragrant*. Similarly, *scented* seems to display a

preference for entities denoting a natural smell, collocating with nouns such as *bloom*, *blossom*, *breeze*, and *grass*, which mainly belong to (sub-)categories L3 and W4. Contrariwise, *perfumed* seems to be the most common adjective to modify artificial smells, as suggested by collocates such as *hair*, *handkerchief*, *lace*, *pocket*, and *silk*, that is, nouns in (sub-)categories B1: ANATOMY AND PHYSIOLOGY (1 collocate) and B5 (4 collocates), but also in O1: SUBSTANCES AND MATERIALS GENERALLY (3 collocates, namely *oil*, *water*, and *smoke*). The five most prominent collocates of each adjective in this period —in decreasing order according to their PMI scores (cf. Tables C.4, C.5, and C.6 in Appendix C)— are the following:

(i) *fragrant*: *odor*, *balsam*, *honeysuckle*, *geranium*, and *petal*.
(ii) *perfumed*: *handkerchief*, *lace*, *blossom*, *oil*, and *silk*.
(iii) *scented*: *soap*, *grass*, *blossom*, *bloom*, and *breeze*.

This list confirms the general tendencies described above, with *fragrant* and *scented* occurring mostly with 'natural' collocates, and *perfumed* with 'artificial' ones. Nevertheless, in the case of *scented*, the most prominent collocate is *soap*, which belongs to the category B4: CLEANING AND PERSONAL CARE, and is thus a noun that occurs with the adjectives at issue when they are used in the artificial sense.

Figure 33 displays the collocational network for P3.

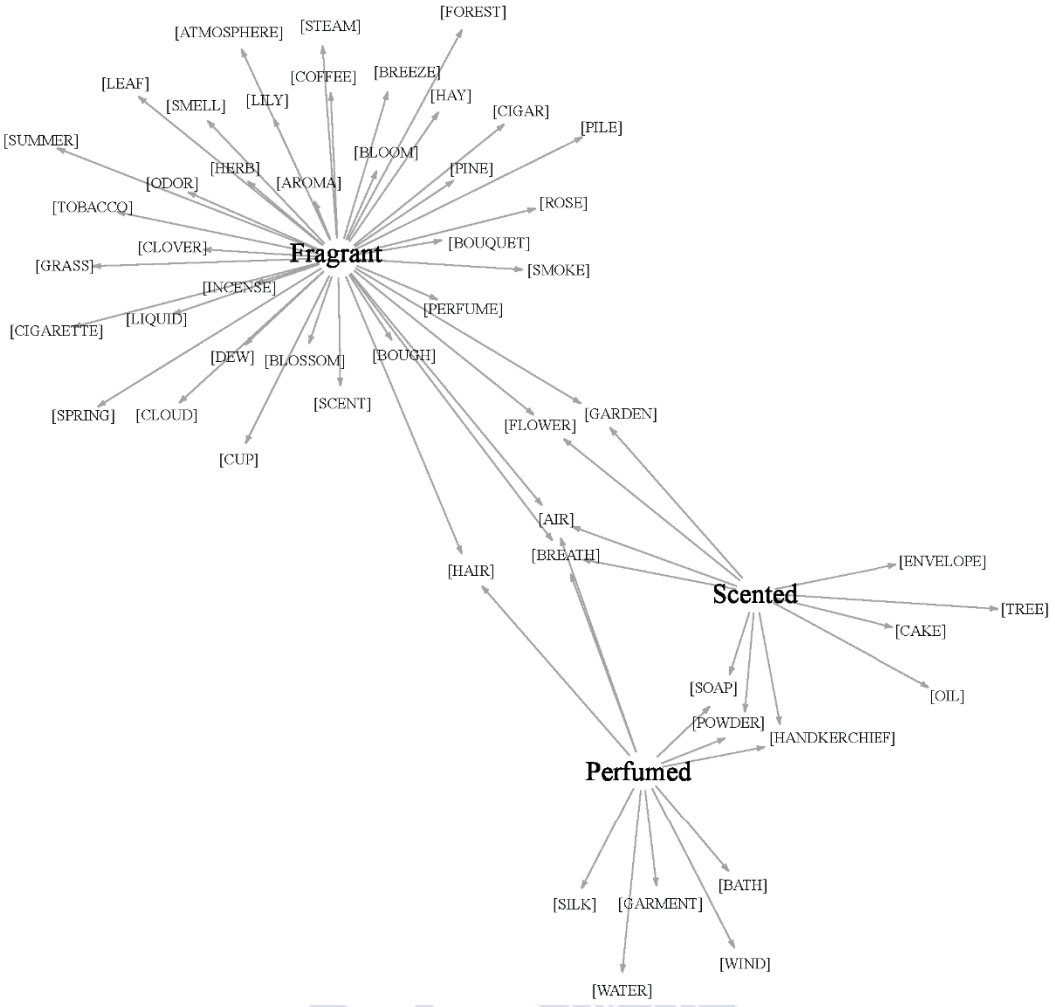


Figure 33. Collocational network of *fragrant*, *perfumed*, and *scented* in P3 (1910–1959)

With 39 collocates, *fragrant* is still the most productive adjective of the synonym set in this period, but it now seems to start losing some ground if compared with previous stages. Similarly, *perfumed* decreases somewhat in comparison to P2, with 11 nouns, i.e. five less than in P2. *Scented*, in turn, appears with one more collocate (11 in all) than in the preceding period. As regards the semantics of the collocates of the near-synonyms, the results point to a certain degree of specialization of the adjectives: *fragrant* still dominates in the natural sense, with 25 collocates clearly depicting natural smells (e.g. *blossom*, *forest*, *lily*, *pine*, and *rose*), while *perfumed* favors the artificial sense (e.g. *garment*, *handkerchief*, *powder*, and *soap*). As becomes evident from the collocational network in Figure 33, the noun collocates of *fragrant* and *perfumed* mainly correspond with the same semantic categories as those discussed for P2, namely L3, F, W4, X3.5 for *fragrant*, and O1 and B1, B5, and O1 for *perfumed*. On the other hand, *scented* does not seem to exhibit a clear preference for any of the two senses, occurring both in

the natural sense, especially with nouns in category L3 (e.g. *flower*, *garden*, and *tree*), and in the artificial sense, with nouns from different categories, especially O1 (e.g. *powder* and *oil*), but also B4 (e.g. *soap* and *handkerchief*). Again, the five most prominent collocates of each adjective reveal that *fragrant* is mostly used to denote natural smells and *perfumed* artificial ones:

(i) *fragrant*: *aroma*, *incense*, *blossom*, *perfume*, and *clover*.
(ii) *perfumed*: *soap*, *handkerchief*, *garment*, *bath*, and *powder*.
(iii) *scented*: *soap*, *powder*, *envelope*, *handkerchief*, and *cake*.

In the case of *scented*, all top five nouns clearly belong to the artificial domain.[117] Interestingly, in the case of *fragrant*, L3 nouns seem to become less dominant, with only 2 items (*blossom* and *clover*) in its top five, as opposed to 4 in previous periods. Also worthy of notice is the fact that *perfumed* and *scented* share several of their most prominent collocates in this period, namely *soap*, *handkerchief*, and *powder*, which may indicate that their prototypical uses are more similar than to those of *fragrant*.

Figure 34 shows the collocational network for P4.

[117] The noun *cake* can be used either in the semantic category F: FOOD AND FARMING, as in example (i), thus referring to a natural aroma, or in B4: CLEANING AND PERSONAL CARE, as in (ii), thus referring to an artificial one.

(i) Half the bread compartment was filled with dainty sandwiches of bread and butter sprinkled with the yolk of egg and the remainder with three large slices of the most **fragrant** spice **cake** imaginable. (*COHA*, 1909, FIC, GirlLimberlost)
(ii) "I put a **cake** of **scented** soap among your handkerchiefs," she said, rather breathlessly. (*COHA*, 1922, FIC, BreakingPoint)

In the data, *scented* always collocates with *cake* in the latter category (B4: CLEANING AND PERSONAL CARE), as opposed to *fragrant*, which, in all cases but one, co-occurs with *cake* in the former category (F: FOOD AND FARMING).

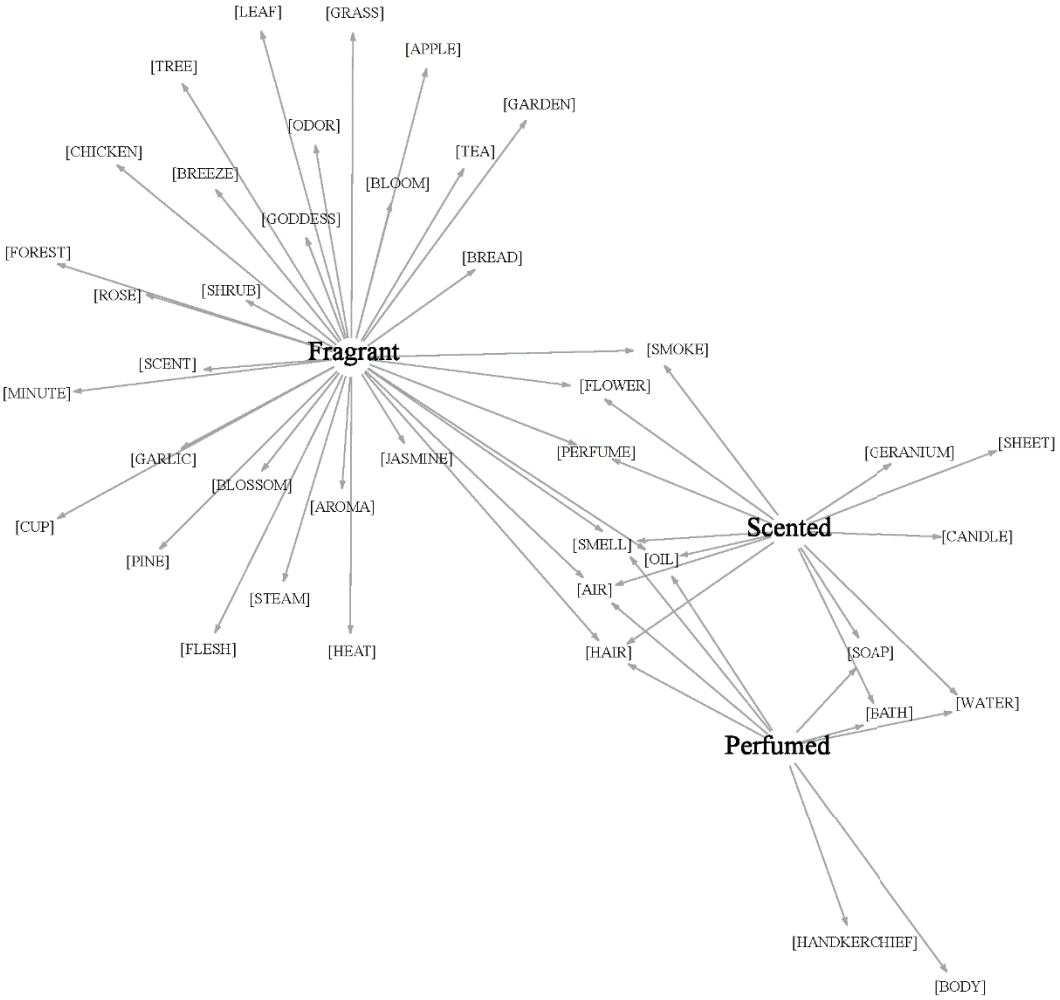


Figure 34. Collocational network of *fragrant*, *perfumed*, and *scented* in P4 (1960–2009)

As shown here, *fragrant* continues losing ground, occurring now with only 33 collocates. The number of collocates of *perfumed* also decreases slightly in contrast with P3, from 11 to 9. Contrariwise, the number of collocates of *scented* continues rising (13 collocates). As in the previous periods, many of the collocates of *scented* are shared with one or both of the other adjectives, but many of them are now more strongly associated with *scented*. In order words, the PMI scores of these nouns are now higher with *scented* than with *fragrant* or *perfumed*. This is the case of *hair*, *perfume*, and *smoke* for *fragrant*, and of *bath*, *oil*, and *water* for *perfumed* (cf. Tables C.4–C.6 in Appendix C). This finding seems to be in line with those of previous chapters: *scented* comes to occupy some of the semantic space previously belonging to *fragrant* and/or *perfumed* (cf. Sections 5.4 and 6.2.1). As concerns the semantics of the nouns, those related to natural smells still collocate more commonly with *fragrant* (21 natural collocates, e.g. *bloom*, *flesh*, *jasmine*, and *shrub*), while *perfumed* remains as the preferred

choice with artificial smells (e.g. *bath*, *handkerchief*, *oil*, and *soap*). Again, the semantic categories of nouns with which the adjectives collocate are basically the same: mainly categories F, L3, W4, X3.5, and O1 in the case of *fragrant*, and B1, B4, B5, and O1 in the case of *perfumed*. As in the previous period, *scented* seems to occupy an intermediate position in this respect, being common with both 'natural' (mainly category L3; e.g. *geranium* and *flower*) and 'artificial' collocates (mainly categories B1, B4, B5, and O1; e.g. *hair*, *perfume*, *soap*, *sheet*, and *oil*).

The top five noun collocates of the three adjectives in P4 are the following:

(i) *fragrant*: *aroma*, *jasmine*, *shrub*, *goddess*, and *blossom*.
(ii) *perfumed*: *soap*, *handkerchief*, *bath*, *smell*, and *oil*.
(iii) *scented*: *geranium*, *soap*, *candle*, *perfume*, and *bath*.

The five most prominent collocates of *scented* are, as in P3, mostly 'artificial', with the exception of *geranium*, which as we saw in Section 6.2.2, occurs with *scented* only in one specific text in *COHA*. All top noun collocates of *perfumed* except *smell* are also clearly 'artificial', while the opposite is true of those of *fragrant*. Interestingly, two of the artificial noun collocates of *scented* are, once more, shared with *perfumed*, namely *soap* and *bath*, again indicating a higher degree of similarity between these two adjectives.

A different but complementary perspective of the collocational preferences of the adjectives can be achieved by classifying their most prominent collocates visualized in the collocational networks in Figures 31–34 into the types proposed by McEnery & Baker (2017: 25–30), to wit, consistent, initiating, terminating, and transient. A summary of the absolute number and percentages of the different types of collocates of the three near-synonyms is provided in Table 33. The detailed classification of noun collocates into types is shown in Tables 34–36.

Table 33. Number and percentage of consistent, initiating, terminating, and transient collocates of *fragrant*, *perfumed*, and *scented*

	Consistent	Initiating	Terminating	Transient	TOTAL
Fragrant					
N	27	12	22	35	96
%	28.13	12.51	22.91	36.46	100
Perfumed					
N	6	4	4	11	25
%	24.00	16.00	16.00	44.00	100
Scented					
N	3	9	1	13	26
%	11.54	34.62	3.85	50.00	100

Here, the focus is mainly on consistent, initiating, and terminating noun collocates, particularly on the last two types, as they are the most informative ones regarding possible diachronic fluctuations in collocational preferences. Transient collocates are disregarded, as they are basically indicative of transient or punctual changes (McEnery & Baker 2017: 28), while the focus here is on more stable diachronic tendencies or the lack of such trends.

As displayed in Table 33, *fragrant* and *perfumed* seem to be the most stable adjectives, with a much higher amount of consistent collocates, 28.13% and 24%, respectively. *Scented*, in turn, exhibits a lower rate of such collocates, namely 11.54%. If we consider the share of initiating and terminating collocates of the adjectives, an interesting picture emerges. *Fragrant* displays the largest amount of terminating collocates (22.91%) and the lowest amount of initiating ones (12.51%), which means that it loses more collocates over time than it gains. This finding replicates those of Section 5.1 (cf. Table 9 and Figure 3), where an overall decrease in frequency of this adjective was identified during the time span analyzed. Interestingly, some patterns can be deduced from the semantics of the noun collocates that *fragrant* gains and loses over time (cf. Table 34).[118]

[118] In Tables 34–36, the following abbreviations are used to refer to consistent, initiating, terminating, and transient collocates, respectively: cons., init., ter., and tran.

Table 34. Consistent, initiating, terminating, and transient noun collocates of *fragrant*

COLLOCATE	P1	P2	P3	P4	TYPE	COLLOCATE	P1	P2	P3	P4	TYPE
air	+	+	+	+	Cons.	*hay*	+	+	+		Cons.
apple				+	Init.	*heat*				+	Init.
aroma			+	+	Init.	*herb*	+	+	+		Cons.
atmosphere		+	+		Tran.	*hillside*		+			Tran.
balsam		+			Tran.	*honeysuckle*	+	+			Ter.
basket		+			Tran.	*incense*	+	+	+		Cons.
berry		+			Tran.	*jasmine*				+	Init.
beverage		+			Tran.	*June*		+			Tran.
bloom	+	+	+	+	Cons.	*leaf*	+	+	+	+	Cons.
blossom	+	+	+	+	Cons.	*lily*		+	+		Tran.
blush	+				Ter.	*liquid*			+		Tran.
bough		+	+		Tran.	*load*		+			Tran.
bouquet		+	+		Tran.	*magnolia*		+			Tran.
bower	+	+			Ter.	*meadow*	+	+			Ter.
bowl		+			Tran.	*melody*		+			Tran.
bread				+	Init	*memory*		+			Tran.
breath	+	+	+		Cons.	*minute*				+	Init.
breeze	+	+	+	+	Cons.	*odor*	+	+	+	+	Cons.
bud	+	+			Ter.	*oil*		+		+	Tran.
bunch		+			Tran.	*orchard*		+			Tran.
cedar		+			Tran.	*perfume*	+	+	+	+	Cons.
chicken				+	Init.	*petal*	+	+			Ter.
cigar	+	+	+		Cons.	*pile*			+		Tran.
cigarette			+		Tran.	*pine*		+	+	+	Cons.
cloud		+	+		Tran.	*plant*	+				Ter.
clover		+	+		Tran.	*porch*		+			Tran.
cluster	+	+			Ter.	*rose*	+	+	+	+	Cons.
coffee	+	+	+		Cons.	*scent*		+	+	+	Cons.
cup	+	+	+	+	Cons.	*shade*	+	+			Ter.
curl	+				Ter.	*shrub*	+	+		+	Cons.
dew	+	+	+		Cons.	*sight*	+				Ter.
dusk		+			Tran.	*smell*		+	+	+	Cons.
fern		+			Tran.	*smoke*		+	+	+	Cons.
fir		+			Tran.	*spice*	+	+			Ter.
flesh				+	Init.	*spring*			+		Tran.

Table 34. Continued

COLLOCATE	P1	P2	P3	P4	TYPE	COLLOCATE	P1	P2	P3	P4	TYPE
flower	+	+	+	+	Cons.	*steam*		+	+	+	Cons.
foliage	+	+			Ter.	*summer*	+	+	+		Cons.
forest			+	+	Init.	*sunshine*		+			Tran.
fruit	+	+			Ter.	*tea*	+	+		+	Cons.
gale	+				Ter.	*tobacco*	+	+	+		Cons.
garden	+	+	+	+	Cons.	*tree*		+		+	Tran.
garland	+				Ter.	*verdure*		+			Tran.
garlic				+	Init.	*vine*	+	+			Ter.
geranium		+			Tran.	*weed*		+			Tran.
goddess				+	Init.	*wind*	+				Ter.
grass	+	+	+	+	Cons.	*wine*	+				Ter.
grove	+	+			Ter.	*wood*	+	+			Ter.
gum		+			Tran.	*wreath*	+	+			Ter.
hair			+	+	Init.						

Terminating collocates correspond mostly with nouns belonging to the semantic categories L3: PLANTS (e.g. *bower*, *foliage*, *honeysuckle*, *petal*, and *plant*) and W4: WEATHER (e.g. *gale*, *meadow*, and *wind*), whereas initiating ones are concentrated mainly in categories F: FOOD AND FARMING (e.g. *apple*, *bread*, *chicken*, and *garlic*) and B1: ANATOMY AND PHYSIOLOGY (e.g. *flesh* [of body] and *hair*). As seen in section 5.3.2, *fragrant* increased in P4 with nouns belonging to the semantic category FOOD AND DRINK; the findings obtained here point to a similar conclusion. However, it is worth mentioning that *fragrant* also gains some new L3 collocates (e.g. *forest* and *jasmine*) and loses some F collocates (e.g. *fruit*, *spice*, and *wine*), which suggests that its prototypical uses remain largely the same over time. This can also be observed by zooming in on the consistent collocates of this adjective, most of which are L3 nouns, especially basic level terms such as *bloom*, *blossom*, *flower*, *garden*, *grass*, and *leaf* and W4 nouns, such as *breeze*, *dew*, and *summer*. Many of the consistent collocates of *fragrant* correspond with those that exhibited a high degree of association with this adjective in the mixed-effects regression analysis carried out in Section 6.2.2. These are the three tobacco-related nouns *cigar*, *smoke*, and *tobacco*. In turn, the noun *weed*, which emerged as a prominent collocate of *fragrant* in the mixed-effects regression analyses, is here categorized as a transient collocate, occurring only in P2. This shows the importance of conducting a fine-grained qualitative analysis of the individual collocates in different periods, since some collocates are more informative than

others, occurring continually over time. Finally, two other consistent collocates of *fragrant* are also in line with the results of the mixed-effects regression model. These are the two nouns *odor* and *perfume*, which belong to the sensory domain. While both lemmas appear as prominent collocates in all four periods, *odor* displays a downward tendency with *fragrant* in P4: out of its 6 occurrences in this period, 5 are dated before the 1980s, with only one example of this noun co-occurring with *fragrant* in the last 30 years in *COHA*. The decrease of *odor* as a collocate of *fragrant* may be related to the fact that this noun has undergone a process of pejoration in recent years, as explained in Section 3.3, being nowadays used mainly with a negative connotation to depict 'unpleasant smells', much as many of its related forms (e.g. *odorous*; cf. Section 4.1.2). On the contrary, *fragrant* and the other adjectives analyzed in the present dissertation are explicitly used to denote pleasant smells, and no indication has been found in the corpus of their use to designate disagreeable ones. Therefore, it is not at all surprising that *odor*, which progressively becomes more strongly associated with negative smells, almost vanishes in the last decades in *COHA* as a collocate of *fragrant*. Additionally, in the examination of the last decade in *COCA* (2010s), *odor* appears only once in the surrounding context of *fragrant* (i.e. L5–35 window), a fact which reinforces this hypothesis.

As regards *perfumed*, its relative stability in terms of its overall frequency (cf. Section 5.1) seems to be reflected also in its proportion of initiating and of terminating noun collocates. As can be seen in Table 33 above, these two types of collocates are evenly distributed, namely 16%. This indicates that, over time, *perfumed* has gained as many new prominent collocates as it has lost. As concerns the semantics of these collocates (cf. Table 35), 3 out of 4 initiating ones correspond to uses in the artificial sense and in indeterminate uses, namely *bath* and *soap*, i.e. category B4: CLEANING AND PERSONAL CARE, and *body*. i.e. category B1: ANATOMY AND PHYSIOLOGY. On the contrary, two of its terminating collocates belong to L3: PLANTS (*flower* and *tree*) and another one is the sensation noun *flavor*, which has already been commented on at several points in this dissertation, as the use of *perfumed* with this noun is restricted to just a few particular texts in *COHA* in P1 (cf. Sections 5.3.2 and 6.2.2).

Table 35. Consistent, initiating, terminating, and transient noun collocates of *perfumed*

COLLOCATE	P1	P2	P3	P4	TYPE	COLLOCATE	P1	P2	P3	P4	TYPE
air	+	+	+	+	Cons.	*oil*	+	+		+	Cons.
bath			+	+	Init.	*pocket*		+			Tran.
blossom		+			Tran.	*powder*			+		Tran.
body				+	Init.	*rose*		+			Tran.
breath	+	+	+		Cons.	*silk*		+	+		Tran.
flavor	+				Ter.	*smell*				+	Init.
flower	+	+			Ter.	*smoke*		+			Tran.
garden		+			Tran.	*soap*			+	+	Init.
garment			+		Tran.	*spring*		+			Tran.
hair		+	+	+	Cons.	*tree*	+				Ter.
handkerchief		+	+	+	Cons.	*water*	+	+	+	+	Cons.
lace		+			Tran.	*wind*			+		Tran.
note	+	+			Ter.						

The consistent noun collocates of *perfumed* also speak to its prototypical use being that of denoting artificial aromas or those labeled here 'indeterminate': out of 6 consistent collocates, 4 are related to uses in the artificial sense and/or indeterminate uses, namely *hair*, *handkerchief*, *oil*, and *water*. Although the latter two nouns are highly polysemous, most of their instances co-occurring with *perfumed* in *COHA* clearly fall into the domains of PERSONAL CARE (*oil*) and CLEANING (*water*), as illustrated in examples (119) and (120):

(119) He was a dandy who garbed himself in colorful silk shirts and a broad belt with gold buckles; he stopped at the barber in the straw market every morning before coming to work to be shaved and have his hair combed with a **perfumed oil**. (*COHA*, 1961, FIC, AgonyEcstasy)

(120) This morning, he spent more time than usual, being sure that no vestige of beard was left on his face, and that he was perfectly clean. He completed his bath by dashing **perfumed water** over his entire body. He opened his traveling chest, picking out clothing he had worn but few times, and those in private. (*COHA*, 1955, FIC, Millenium)

The other two consistent noun collocates of *perfumed* are *air* and *breath*. The former also emerged as a prominent collocate of this adjective in the mixed-effects regression analysis (cf. Figure 25 in Section 6.2.2), together with *flavor*, while most other prominent collocates identified here are absent from the regression model, except the two transient collocates *garment* (in P3) and *spring* (in P2). This lack of consistency of the results between the present analysis and those of the mixed-effects regression modeling (cf. Section 6.2.2) again emphasize the need for careful in-depth qualitative analysis in order to distinguish the collocates which are truly informative from those that are only transient.

The significant upward tendency of *scented* in terms of its overall frequency (cf. Table 9 in Section 5.1) can also be observed in the results of the present analysis. In fact, *scented* is the adjective that exhibits the largest share of initiating collocates (34.62%) and the lowest amount of terminating ones (3.85%) (cf. Table 33 above). Hence, over time, *scented* gains more collocates than it loses, thus showing the opposite tendency to that of *fragrant*. Table 36 shows the consistent, initiating, terminating, and transient collocates of *scented* in *COHA*.

Table 36. Consistent, initiating, terminating, and transient noun collocates of *scented*

COLLOCATE	P1	P2	P3	P4	TYPE	COLLOCATE	P1	P2	P3	P4	TYPE
air	+	+	+	+	Cons.	*hair*				+	Init.
bath				+	Init.	*handkerchief*			+		Tran.
bloom		+			Tran.	*oil*			+	+	Init.
blossom		+			Tran.	*perfume*				+	Init.
breath		+	+		Tran.	*powder*			+		Tran.
breeze		+			Tran.	*sheet*				+	Init.
cake			+		Tran.	*smell*				+	Init.
candle				+	Init.	*smoke*				+	Init.
envelope			+		Tran.	*soap*		+	+	+	Cons.
flower	+	+	+	+	Cons.	*summer*		+			Tran.
garden			+		Tran.	*tree*			+		Tran.
geranium				+	Init.	*water*		+		+	Tran.
grass		+			Tran.	*wind*	+				Ter.

As seen here, many of the initiating collocates of *scented* suggest that it has come to be used more and more often in the artificial sense over time. In fact, out of its nine initiating collocates, more than half correspond to artificial smells, namely *bath*, *candle*, *oil*, *perfume*, and *sheet*. In

Section 6.2.2, *scented candle* was already discussed as a plausible instance of a newly-formed compound or a multiword unit —as opposed to a noun phrase formed by an attributive adjective and a head noun—, or even as a case of lexicalization. By resorting again to a contemporary corpus such as *COCA*, it can be ascertained that the degree of association between *scented* and *candle* keeps increasing in American English after the time span covered in *COHA*: in the 2010s, there are 72 occurrences of *candle* in the surrounding context of *scented* (i.e. L5–R5 window) and, out of these, 63 (87.5%) are instances of *scented* immediately preceding *candle* (i.e. *scented candle*(*s*)). Furthermore, the PMI score between *scented* and *candle* rises from 8.54 in P4 (1960–2009; cf. Table C.6 in Appendix C) to 11.04 in the 2010s, as calculated in *COCA*, which also provides the PMI of the collocates of the word(s) searched for in the corpus. The two initiating collocates *oil* and *perfume* behave in a similar way to *oil* and *water* as collocates of *perfumed*: although polysemous, most cases of these two nouns collocating with *scented* are used to denote artificial aromas, especially in the more recent periods. The remaining 4 initiating collocates of *scented*, to wit, *hair*, *geranium*, *smell*, and *smoke*, represent a more varied picture. The nouns *hair*, *smell*, and *smoke* can be used together with *scented* in the natural or artificial senses and, on some occasions, it is not possible to identify the source of the aroma (i.e. indeterminate cases). On the contrary, all instances of *scented geranium* refer to natural uses of the adjective, but almost all these examples come from a single text in the corpus (cf. Section 6.2.2).

The only terminating noun of *scented* is *wind*, belonging to category W4: WEATHER and thus to uses of the adjective in the natural sense. *Scented* exhibits 3 consistent collocates, namely *air*, *flower*, and *soap*, which point to the relative neutrality of this adjective in comparison to *perfumed* and *fragrant*. Note that one collocate is clearly 'natural' (*flower*), one is clearly 'artificial' (*soap*), and another one (*air*) can be used in both 'natural' and 'artificial' contexts, although a closer look at the individual occurrences of this noun in *COHA* reveals that in most cases *scented* qualifies natural smells when it collocates with *air*.

Finally, out of the 4 noun collocates that emerged as being closely associated with *scented* in the mixed-effects regression analysis (cf. Section 6.2.2), three have also been discussed here, namely *candle*, *geranium*, and *wind*. Interestingly, none of these collocates is classified as consistent in the present analysis: whereas *candle* and *geranium* are initiating collocates, occurring only in P4, *wind* is a terminating collocate which occurs only in P1.

The more in-depth qualitative analysis carried out in this section by means of collocational networks and of the classification of the most prominent collocates into consistent, initiating, terminating, and transient ones has served to confirm many of the tendencies identified in earlier sections and chapters. Among them, note the preference of *fragrant* for the natural sense, that of *perfumed* for the artificial one, and the relative neutrality of *scented*, although its top collocates in several periods, especially in P3 and P4, seem to correspond mostly to artificial rather than to natural uses of the adjective. This slight inclination of *scented* towards 'artificial' contexts may go a long way towards explaining its closer similarity to *perfumed* than to *fragrant* observed in previous analyses, namely the regression analyses conducted in Chapter 6 and the SVS modeling discussed in Section 7.2.1. Similarly, the frequency fluctuations uncovered in Section 5.1 are also reflected in the qualitative analysis provided here: *fragrant* loses more collocates than it gains over time, *scented* gains more than it loses, and *perfumed* remains relatively stable, losing and gaining the same amount of collocates over time, although a slight decrease is noticeable in the collocational networks. However, the main findings of the qualitative analysis conducted in this section are the highly idiosyncratic collocational preferences of the adjectives, preferences which cannot be captured by methods which average over a very large amount of collocates (that is, SVS) or those which take into account more coarse-grained semantic categorizations (that is, the analyses in Chapters 5 and 6). Two idiosyncratic trends with important implications are the cases of *fragrant* and *odor*, on the one hand, and of *scented* and *candle*, on the other. The former suggests that the positive connotation of the adjective has probably not changed diachronically, as the noun *odor*, which has recently acquired a negative connotation, becomes a less and less likely collocate of *fragrant* over time. In turn, the combination *scented candle* may constitute a case of compounding or even of ongoing lexicalization, as it has become established in the language as a multiword unit over a relatively short period of time.

7.3 Summary

The findings of the present chapter have demonstrated that, as in the case of the semantic preferences of the adjectives *fragrant*, *perfumed*, and *scented*, their idiosyncratic collocational preferences also vary over time. First, when averaging over all the noun collocates of the near-synonyms, that is, the SVS analysis provided in Section 7.2.1, similar tendencies as those found in previous chapters emerge. A semantic division of labor is evident between *fragrant* and

perfumed, with the former being the preferred choice to designate natural aromas and the latter to designate artificial ones. As in the previous analyses, *scented* remains a more neutral adjective, not being particularly associated with either natural or artificial smells. Additionally, the results of the SVS analysis also seem to point to external forces playing an important role on the semantic development of the three adjectives, with most changes being located particularly in the transition from the 19th to the 20th century (cf. Dimension 1 in Figure 29 above).

Second, the more in-depth qualitative analysis of collocational networks carried out in Section 7.2.2 allows us to delve deeper into the prototypical uses of the adjectives. Again, some tendencies observed previously are also reflected in the findings of this section. To begin with, the specialization of *fragrant* and *perfumed* to the natural and artificial ends of the semantic continuum, respectively, as well as the relative neutrality of *scented* in this respect, are ascertained by their most prominent collocates. Moreover, the similarity of *perfumed* and *scented*, if compared to *fragrant*, is also evident in their top prominent collocates, with *perfumed* and *scented* sharing a number of them in several periods. In fact, *scented* shares a large amount of its prominent collocates with the other two adjectives, which may explain why the mixed-effects regression model of this adjective in Section 6.2.2 performed worse than those of *fragrant* and *perfumed* (cf. Table 25). This is so because the collocational preferences of *scented* seem to be more similar to those of the other two adjectives, and thus less distinctive, which implies that that the individual collocates of *scented* introduced as random effects in the mixed-effects regression model are not particularly useful to distinguish it from the other two adjectives. The lower degree of distinctiveness of *scented* in terms of its collocational behavior was also reflected in the intermediate position of this adjective on the MDS map in all periods (cf. Figure 29).

The process of substitution whereby *scented* becomes progressively more frequent at the expense of the other two adjectives identified in Sections 5.1, 5.4, and 6.2.1 can also be observed in the collocational networks discussed in this chapter. By P4, *scented* shows a larger amount of collocates than *perfumed*, and the dominance of *fragrant* in this respect is not as evident as in previous periods; in fact, *scented* is the only adjective that gains collocates over time, some of which are drawn away from the other two near-synonyms. Although the process of attraction undergone by the synonym set is also reflected in the present chapter, it is more

subtle here: as mentioned in Section 7.2.2, *scented* shares many collocates with *fragrant* and *perfumed* from the earliest period but comes to share even more collocates with them over time.

Nevertheless, the most significant finding of the analyses carried out in the present chapter is that, regardless of any shared tendencies and abstract semantic preferences, the three adjectives display rather idiosyncratic collocational preferences, as the specific nouns with which they tend to co-occur vary from one adjective to another and from one period to the next. This is not entirely surprising, given that the mixed-effects regressions provided in Section 6.2.2 already pointed to a similar conclusion due to the large amount of variance explained by the predictor Collocate when included in the random effect structure of the models. Therefore, examining the individual collocates of lexical near-synonyms, as done in the present chapter, is an indispensable task to further uncover differences between such semantically related words.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

8 CONCLUDING REMARKS AND SUGGESTIONS FOR FUTURE RESEARCH

The semantic relation of synonymy is a topic with a fairly long research tradition in lexical semantics, especially since the structuralist movement brought to the fore a focus on onomasiological structures in the language (cf. Section 2.2). Synonymy, as well as other lexico-semantic relations, was considered by structuralists to be of crucial importance for the delineation of meaning and the organization of lexical knowledge. The perspective adopted by structuralists was paradigmatic in nature, as priority was given to the language system —as opposed to language use— and therefore to the identification of different types and subtypes of semantic relations existing in language (Section 3.1). Neostructuralist approaches, particularly the distributional corpus-based one, moved away from this excessive emphasis on the system to a focus on language use by adopting a syntagmatic perspective (cf. Section 2.3). In this way, research on synonymy widened in scope and the interest shifted from the mere classification of semantic relations in general, and types of synonyms in particular, to an in-depth analysis of the factors that determine the choice between near-synonymous expressions. One of the main tenets of distributional semantics was the great importance of collocations in language knowledge and language use. Collocation, in its different forms (e.g. semantic preference and semantic prosody), has been repeatedly found to influence the use of alternative expressions such as near-synonyms and the notion is still widely used in other contemporary movements (cf. Section 3.2). In particular, cognitive semantics continued along the lines set out by distributional semantics and largely converged with many of the principal foundations of this neostructuralist approach, mainly due to its strong usage-based orientation (cf. Section 2.4). Additionally, research on synonymy within cognitive semantics benefitted from the theoretical apparatus of cognitive linguistics, thus improving the distributional method.

Despite the prominence of synonymy in previous research in the field, both within the (neo)structuralist schools and, more recently, within cognitive semantics, a dimension which has been somewhat overlooked is that of the diachronic development of particular sets of lexical near-synonyms (but cf. Section 3.3 for some diachronic approximations). The overarching aim of the present dissertation was therefore to partially fill this gap in the specialized literature by analyzing a set of near-synonymous adjectives from the relatively understudied semantic domain of SMELL in 19th and 20th century AmE, to wit, *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*, which designate the concept PLEASANT SMELLING. An in-depth description of the semantics of the near-synonyms was provided in Section 4.1 on the basis of both present-day and historical dictionaries and thesauri, followed by an explanation of the data retrieval process (Section 4.2). Drawing on historical data from *COHA*, the main focus was on the contextual behavior of the five selected adjectives, i.e. their degree of (dis)similarity was quantified in terms of their distributional patterns operationalized through a series of variables from different linguistic domains —semantic, lexical, morphosyntactic, and stylistic— and levels of granularity —i.e. from more abstract semantic preferences to individual collocates (cf. Section 4.3). To this purpose, state-of-the-art methods used in synchronic corpus-based studies on synonymy were applied to the data in subsequent chapters (cf. Chapters 5–7) and their usefulness for diachronic research on this same phenomenon was tested.

Chapter 5 offered descriptive univariate analyses concerning both the semasiological and the onomasiological structures of the near-synonymous expressions at issue, as well as their overall frequency and frequency development over the time span examined. The results suggested the existence of a large difference between the items in the set in terms of their frequency of use, with one clearly dominant adjective, namely *fragrant*, which accounted for more than half of the total data (cf. Section 5.1). With a frequency of approximately 17% each, *perfumed* and *scented* were ranked second and third in the frequency cline, respectively. Lastly, *sweet-smelling* and *sweet-scented* were far less common in the corpus, each amounting to less than 5% of the data.

Interesting developments over time regarding the frequency of the near-synonyms emerged from the data, given that whereas *fragrant* and *sweet-scented* both decreased significantly from the 1810s to the 2000s, *scented* and *sweet-smelling* increased. In turn, *perfumed* remained relatively stable diachronically. These frequency fluctuations led to a

reorganization of the members of the synonym set: although *fragrant* remained the most common adjective throughout the 19th and 20th centuries, *scented* displaced *perfumed* and *sweet-smelling* outperformed *sweet-scented*. Interestingly, the concept PLEASANT SMELLING as a whole, that is, as designated by all five adjectives, decreased considerably in frequency over time, which makes the increase of *scented* and *sweet-smelling* even more striking. The semantics of the concept also exhibited diachronic trends (cf. Section 5.2). Although overall it was used mostly to qualify natural aromas —and the associated semantic categories of nouns—, a shift towards artificial smells was observed, particularly at the turn of the 19th to the 20th century. A hypothesis was put forward to explain this diachronic development on the basis of extralinguistic forces, namely the sociocultural and technological transformations experienced by American society following the First and Second Industrial revolutions (cf. Section 5.2.3). In a nutshell, it was argued that the increasing need to refer to artificial —as opposed to natural— smells sprang from the growing availability of synthetic products and goods, as well as from their presence in the daily life of American people. This hypothesis found preliminary support in a pilot study on lexical indicators from three semantic domains related to processes of modernization and industrialization (i.e. INDUSTRY AND TECHNOLOGY, ADVERTISING AND MEDIA, and COSMETICS AND HYGIENE). The items belonging to these domains were also observed to occur substantially more frequently in the corpus after the 1900s, precisely at the same point in time when the concept experienced a significant upward tendency to denote artificial aromas.

However, one of the main goals of Chapter 5 was to confirm or refute the information provided by the reference material introduced in Section 4.1.2 concerning the semantics of the five near-synonyms. This was done by examining their semasiological structures (cf. Section 5.3). The distribution of each of them across senses and semantic categories of modified nouns confirmed most of the information available in the reference material, with *perfumed* and *scented* being used more often to qualify artificial aromas, as opposed to the other three adjectives. However, the actual state of affairs was somewhat more nuanced than that represented in the dictionaries consulted. First, although for *fragrant* and *sweet-scented* most dictionaries provided only the general basic sense ‘having a sweet pleasant smell’, both adjectives were found to be prototypically ‘natural’, as the great majority of the occurrences of these two adjectives in the corpus corresponded to the natural sense and its associated semantic categories. Second, *sweet-smelling*, although more common also in the natural

sense, exhibited a fairly large amount of 'artificial' uses. In addition, although not specified in the dictionaries, important differences regarding the use of the adjectives across semantic categories of modified nouns emerged from the corpus. This finding highlighted valuable usage differences which might have been overlooked if only the sense in which the adjectives are used were considered. For instance, although both *perfumed* and *scented* were common to qualify artificial aromas, the former was more prominent with nouns belonging to the categories BODY AND PEOPLE and TEXTILE AND CLOTHING, while the latter was more prominent with those in the category OBJECT. As concerns diachronic patterns in semasiological terms, similar trends arose for all adjectives except *sweet-scented*, namely an increase over time in their frequency of use in the artificial sense and most of its corresponding semantic categories. The importance of this finding resided in the fact that, irrespective of their prototypicality, be it natural or artificial, most near-synonyms in the set ended up being more 'artificial' at the end of the time span examined, hence suggesting that these changes were not coincidental, but rather the result of some underlying motivation, which, as suggested in Chapter 5, could be related to language-external forces, in particular the modernization of American society from a preindustrial to a fully industrialized nation.

Finally, a first glimpse into the onomasiological structure of the synonym set was provided by zooming in on the competition between the five semantically related adjectives across senses and across semantic categories (cf. Section 5.4). This analysis allowed us to determine which adjectives were the most likely in different contexts of use in comparison with the other members of the set. It was shown that, despite the much higher overall frequency of *fragrant*, this adjective was not the most likely one in the artificial sense and some of its associated semantic categories, where it was outperformed by *perfumed* and/or *scented*. In addition, a change over time in the proportions of use of the near-synonyms in some senses and in some semantic categories was identified. First, *scented* displaced *perfumed* and, to a certain extent, *fragrant* in several of the 'artificial' semantic categories and, second, in the figurative sense and in indeterminate uses, *fragrant* lost its initial clear dominance over time. All this brought about a more balanced distribution of the adjectives at the end of the time span analyzed.

The separate analyses carried out in Chapter 5 offered different but complementary interpretations of the data in as much as the results of one type of analysis aided in the interpretation of other findings. This resulted in a more holistic approach to the study of the

five near-synonyms. First, a correlation was observed between the frequency fluctuations of the adjectives and their semasiological salience: the adjectives which decreased in frequency over time were those which were prototypically 'natural' (i.e. *fragrant* and *sweet-scented*), while those that increased occupied an intermediate position on the continuum of senses postulated in Section 5.3.3, to wit, *scented* and *sweet-smelling*. In turn, *perfumed*, that is, the prototypically 'artificial' adjective, remained relatively stable. Second, the onomasiological analyses uncovered the specific contexts of use in which each of the five adjectives increased or decreased in frequency. For instance, it became clear that *fragrant* decreased primarily in the figurative sense and in indeterminate uses, but also somewhat in the natural sense. On the contrary, the increase of *scented* occurred mainly in the artificial sense and its associated semantic categories, while that of *sweet-smelling* was located mainly in indeterminate uses, particularly with BODY AND PEOPLE nouns. Lastly, the analysis of the concept PLEASANT SMELLING as a whole suggested the existence of some external motivations underlying most of the diachronic trends identified in Chapter 5, namely the increasing tendency to qualify more and more artificial aromas as time passed, a trend which was common to four out of the five adjectives object of study, with the exception of *sweet-scented*.

Chapter 6 delved deeper into the onomasiological organization of the synonym set by analyzing the competition between the three most common adjectives (i.e. *fragrant*, *perfumed*, and *scented*) across a wider range of internal and external variables. To this purpose, a series of multivariate statistical techniques were applied to the data in order to quantitatively measure the degree of (dis)similarity between the adjectives. The variables included semantic (Sense, Semantic category, Concreteness, and Countability) and non-semantic determinants (Collocate, Degree, and Syntactic function), as well as extralinguistic factors (Period and Text-type). The main goals of this chapter were to determine which variables exerted a stronger influence on the choice between the near-synonyms and to identify in which contexts of use the relation between the three adjectives changed over time. The findings confirmed most of the tendencies uncovered in Chapter 5 regarding the competition between the adjectives across senses and semantic categories, but it also shed new and valuable light on other patterns. All the other semantic and non-semantic linguistic factors played a significant role on the choice between the alternative expressions, but the adjectives did not display diachronic tendencies in any of these contexts of use. The extralinguistic predictor Text-type also affected the alternation between the near-synonymous expressions and, in fact,

significant changes over time emerged with respect to this variable, which indicated that the increase of *scented* took place in all text-types, but that it was particularly pronounced in non-fiction.

Although all the variables included in the analyses in Chapter 6 played a significant role on the choice between the adjectives, some of them were found to have a stronger impact. In particular, the semantic variables outperformed all the non-semantic ones except Period, the latter being of crucial importance in the present data. This specific finding was in line with previous research in the field (e.g. Gries 2001; Liu 2010; Liu & Espino 2012), which indicated that semantic factors, particularly those relating to modified nouns, are stronger determinants of the alternation between adjectival near-synonyms. Despite the fact that all variables included in the analyses turned out to be significant, the goodness-of-fit statistics of the multinomial logistic regression models conducted in Section 6.2.1 showed that these predictors were not enough to explain all the variance in the data. The effects of the individual noun collocates of the adjectives were assessed in Section 6.2.2, and the results of these analyses demonstrated that the variable Collocate significantly improved the quality of the models: almost double the amount of variance in the data was explained when this particular predictor was considered, hence proving the great importance of idiosyncratic collocational preferences in lexical choice.

Given that the individual collocates of the adjectives were found to exert such a strong influence on the alternation between *fragrant*, *perfumed*, and *scented*, Chapter 7 set out to examine the collocational preferences of the three adjectives in more detail by means of methods specifically geared towards this type of analysis. First, Section 7.2.1 offered a bird's-eye view of the collocational behavior of the near-synonyms by averaging over all individual collocates. The findings again confirmed many of the tendencies identified on the basis of more schematic semantic preferences (that is, senses and semantic categories): *fragrant* collocated more frequently with nouns located at the natural end of the semantic spectrum, whereas the opposite tendency was observed for *perfumed*. In turn, *scented* occupied an intermediate position, showing no strong association with any of the two senses and thus being more neutral in this respect. The results of Section 7.2.1 also pointed to a similar conclusion regarding the influence of extralinguistic forces on the diachronic evolution of the near-synonymous adjectives as that reached in Chapter 5. The three adjectives were found to co-occur significantly more often with nouns in 'natural' semantic categories in the 19th

century (i.e. P1 and P2) than in the 20th century (i.e. P3 and P4). This again seemed to indicate a process of convergence regarding the diachronic patterns of the adjectives, all of them becoming less and less common to denote natural smells as time passed. Second, Section 7.2.2 provided a more qualitative examination of the collocational preferences of the near-synonyms by zooming in on their most prominent noun collocates. The tendencies identified in previous chapters received further support by many of the patterns uncovered by means of this fine-grained qualitative analysis. Nevertheless, by far the most important finding of Section 7.2.2 was that, regardless of any diachronic abstract semantic trends shared by the adjectives, the three of them exhibited rather idiosyncratic collocational behavior when their most prominent individual collocates were accounted for.

The results of Chapters 5–7 suggested that two diachronic processes were at work in this particular synonym set, namely those of attraction and substitution. As mentioned in Section 3.3, a common assumption in the specialized literature is that languages work against absolute synonymy. In other words, if two or more synonyms come to exist in the language, either they become increasingly differentiated over time, or one or more of them eventually drop out of use. The present study revealed, however, that only the process of substitution seemed to be partially at work in the synonym set object of study, with *scented* increasing in frequency over time, particularly at the expense of *fragrant*, the predominant member of the set. The outcomes of the increases and decreases in frequency of these two adjectives were twofold. First, a much more balanced picture was identified in many of the contexts of use in present-day AmE, which no longer shows a clearly dominant option. Second, a rearrangement took place concerning the likelihood of the adjectives in some contexts, with *scented* ousting *fragrant* and/or *perfumed* and hence becoming the default option in such cases. Therefore, the development of the near-synonyms cannot be characterized as one of differentiation, since they do not become more dissimilar throughout the period 1810–2009. Rather, the reverse process, namely that of attraction, is observed in the present data, given that all adjectives except *sweet-scented* came to be used more and more frequently in the artificial sense and its associated semantic categories, irrespective of whether they were semasiologically more salient in the natural or in the artificial sense. This finding is in line with recent theories of language change, which claim that attraction is a far more common development than previously assumed (De Smet et al. 2018: 201–202). According to this theory, the underlying motivation for attraction is that of analogy. Nevertheless, a more likely motivation in this

specific synonym set can be identified in language-external factors which steer the course of semantic change: the sociocultural and technological developments explained in detail in Section 5.2.3 constitute the most likely cause for the shift of the concept PLEASANT SMELLING and the terms used to denote it from more natural to more artificial aromas over the time span examined, particularly at the turn of the 19th to the 20th century.

From a theoretical perspective, therefore, the findings of the present dissertation provide further support to the claim that the process of attraction figures more prominently in language change than has been assumed in most previous work on synonymy. This, in turn, implies that the forces behind the diachronic evolution of semantically related words and expressions such as near-synonyms is more complex than originally believed, given that, besides differentiation and substitution, other processes such as attraction may be at work. Additionally, the fact that both attraction and substitution seem to be at play in the case of the selected synonym set again corroborate De Smet et al.'s (2018) hypothesis that attraction may well be a prerequisite for substitution to occur, since synonyms need to share enough semantic ground for one to replace another.

From a methodological point of view, the analyses conducted in this dissertation show that the methods used in synchronic research on synonymy can be successfully applied to diachronic data. The techniques employed for the analyses of the concept PLEASANT SMELLING uncover insightful diachronic patterns and demonstrate that such methods are not only useful to measure the (dis)similarities between the members of a synonym set at a particular point in time, but also the potential diachronic changes in the use of a given near-synonym in successive periods. Therefore, assuming a historical viewpoint allows for the visualization of the changing relation of an entire set of related words or expressions, hence making it possible to observe potential diachronic rearrangements concerning their frequency and saliency in particular contexts of use.

The contributions discussed in the previous paragraphs serve to fill an important gap in the specialized literature on lexical near-synonymy. However, it is necessary to mention some limitations of the studies conducted in this dissertation, as well as possible ways of remedying them and further improving and expanding the present piece of research. First of all, the hypothesis put forward regarding the influence of extralinguistic motivations on the diachronic changes observed was only preliminary tested through a series of lexical indicators connected to processes of modernization and industrialization. Though suggestive of the

hypothesis being on the right track, these findings are not conclusive and additional research is desirable to either verify or refute it. One way in which this hypothesis could be further examined is by undertaking a cross-linguistic study of the equivalent terms of *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling* in other languages spoken in societies which experienced similar sociocultural and technological developments as American society (e.g. Spanish *bienoliente*, *fragante*, *perfumado*, and *odorífero*). If the concept PLEASANT SMELLING and the terms denoting it in other languages follow identical or similar diachronic patterns as those uncovered here and the changes go hand in hand with language-external developments in the technological sphere, this could provide further validation to the hypothesis postulated in the present dissertation. Moreover, the levels of the variable Period used here (i.e. four 50-year periods) could be rearranged to make the time divisions coincide with changes in the extralinguistic reality that may have affected the evolution of the use of the near-synonyms at issue. For instance, the variable could be modified in such a way that its different levels reflect, instead of arbitrary 50-year periods, times of economic growth vs. times of economic recession or other important moments in the history of industrialization and other modernization processes in American society.

Another shortcoming of the studies carried out in this dissertation concerns the insufficient number of examples of *sweet-scented* and *sweet-smelling*, as well as of other low-frequency near-synonymous expressions (e.g. *fragranced* and *perfumy*; cf. Section 4.1.2), to include them in the analyses provided in Chapters 6 and 7. To solve this problem, it is necessary to resort to additional corpora. For instance, a fairly compatible corpus is *COCA*, which covers the time span 1990–2019 and contains data from the same variety of English as *COHA*, as well as from many of the genres featured in the latter corpus (e.g. fiction and magazines). This would allow us to trace the development of the near-synonyms in more contemporary sources and extract a larger amount of data from the 1990s and 2000s. However, *COCA* also contains spoken material and texts from genres not found in *COHA* (e.g. blogs and TV/movies), which would make it difficult to identify whether the (dis)similarities in the use of the adjectival near-synonyms between *COHA* and *COCA* are actual diachronic changes or, on the contrary, respond to stylistic patterns. Besides *COCA*, other diachronic corpora from earlier periods in the history of English could be used. For instance, *Early English Books Online Corpus 1.0* (*EEBOCorp 1.0*) would enable the analysis of the five near-synonyms examined here in previous stages, thus making it possible to extract

a larger number of attestations of these adjectives, but also of other near-synonyms which were not discussed in the present dissertation. As we saw in Section 4.1.2, some of the semantically related adjectives from the same semantic domain (e.g. *balmy*, *odorous*, and *redolent*) were discarded here due to the fact that in PDE they are no longer used predominantly in the sense 'having a sweet pleasant smell', as they have undergone processes of specialization and pejoration and came to be differentiated from *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*. If one goes back to earlier historical periods, it might be worth including a larger set of adjectives from the domain of SMELL to pinpoint when and why they became differentiated from the five lexical items at issue here. An additional advantage of employing a corpus such as *EEBOCorp 1.0* is the fact that several of the adjectives of the synonym set examined in this dissertation, to wit, *fragrant*, *perfumed*, and *scented*, were borrowed into English in the first half of the 16th century. Given that this particular corpus goes back to the 15th century, it would give us the opportunity to trace the history of this synonym set from the time that it originated in the English language.

Finally, the multinomial regression models carried out in Section 6.2.1 made it evident that if the individual noun collocates of the near-synonyms were not taken into account, the set of language-internal and language-external variables examined in Chapter 6 were not enough to explain the alternation between the selected adjectives. It was there argued that enlarging the set of variables to include factors such as priming and *horror aequi*, among others, might alleviate this problem. In addition, a level of co-occurrence which was only briefly touched upon in the analysis of one of the prominent noun collocates of *fragrant*, namely *odor*, is that of semantic prosody and, more generally, connotation. Given that neither the reference material discussed in Section 4.1.2 nor impressionistic observations of the data seem to point to a loss of the positive connotation of the adjectives in the synonym set, semantic prosody was not considered in depth in the present dissertation. However, it may be worth examining this issue in more detail, particularly if other adjectives belonging to the same semantic domain are included (e.g. *odorous*). This would allow us to identify whether differences also exist between the near-synonyms as regards this dimension of meaning, which, as has been shown in the specialized literature, often influences lexical choice (cf. Section 3.1). One relatively straightforward way in which the potential effects of semantic prosody could be uncovered is to submit the collocates in a determinate context window of the near-synonyms in the corpus to automatic sentiment analysis using, for instance, software

such as *Linguistic Inquiry and Word Count 2015* (Pennebaker et al. 2015). This tool automatically assigns words to one or more grammatical, semantic, and punctuation categories from among 90 different ones (e.g. negative and positive emotion). Therefore, if one near-synonym co-occurs significantly more often with words with a negative connotation than another near-synonym, this may suggest the existence of a difference in semantic prosody between them. Nonetheless, this particular issue and those discussed in the previous paragraphs must be left for future research.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

LIST OF REFERENCES AND SOURCES

References

Adamska-Salaciak, Arleta. 2013. Equivalence, synonymy, and sameness of meaning in a bilingual dictionary. *International Journal of Lexicography* 26(3). 329–345.

Allan, Kathryn. 2015. Education in the *Historical Thesaurus of the Oxford English Dictionary*. In Jocelyne Daems, Eline Zenner, Kris Heylen, Dirk Speelman & Hubert Cuyckens (eds.), *Change of paradigms - new paradoxes: Recontextualizing language and linguistics*, 81–96. Berlin & Boston: De Gruyter Mouton.

Anderson, Wendy & Ellen Bramwell. 2014. A metaphorical spectrum: Surveying colour terms in English. In Wendy Anderson, Carol P. Biggam, Carole Hough & Christian Kay (eds.), *Colour studies: A broad spectrum*, 140–152. Amsterdam & Philadelphia: Benjamins.

Andersson, Evert. 1985. *On verb complementation in written English* (Lund Studies in English). Vol. 71. Lund: CWK Gleerup.

Arppe, Antti. 2002. The usage patterns and selectional preferences of synonyms in a morphologically rich language. In Annie Morin & Pascale Sébillot (eds.), *JADT-2002. 6th international conference on textual data statistical analysis*. Vol. 1, 21–32. Rennes: INRIA.

Arppe, Antti. 2008. *Univariate, bivariate, and multivariate methods in corpus-based lexicography: A study of synonymy*. Helsinki: University of Helsinki dissertation.

Arppe, Antti. 2013. *Polytomous: Polytomous logistic regression for mixed and fixed effects.*

Arppe, Antti & Juhani Järvikivi. 2007. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Linguistic Theory* 3(2). 131–159.

Atkins, Beryl T.S. 1987. Semantic ID tags: Corpus evidence for dictionary senses. In *Proceedings of the third annual conference of the UW centre for the New Oxford English Dictionary*, 17–36. Waterloo: University of Waterloo.

Atkins, Beryl T.S & Beth Levin. 1995. Building on a corpus: A linguistic and lexicographical look at some near-synonyms. *International Journal of Lexicography* 8(2). 85–114.

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge & New York: Cambridge University Press.

Baker, Paul. 2005. *Public discourses of gay men*. Abington: Routledge.

Baker, Paul. 2009. The BE06 corpus of British English and recent language change. *International Journal of Corpus Linguistics* 14(3). 312–337.

Baker, Paul. 2017. *American and British English: Divided by a common language?* Cambridge: Cambridge University Press.

Baldinger, Kurt. 1980. *Semantic theory*. Oxford: Basil Blackwell.

Barton, Kamil. 2020. *MuMIn: Multi-model inference*. https://CRAN.R-project.org/package=MuMIN.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Berg, Thomas. 2014. Competition as a unifying concept for the study of language. *The Mental Lexicon* 9(2). 338–370.

Bergenholtz, Henning & Rufus Gouws. 2012. Synonymy and synonyms in Lexicography. *Lexicographica: International Annual for Lexicography* 28(1). 309–336.

Berlin, Brent & Paul Kay. 1969. *Basic color terms: Their universality and evolution.* Berkeley & Los Angeles: University of California Press.

Biber, Douglas. 1993. Co-occurrence patterns among collocations: A tool for corpus-based lexical knowledge acquisition. *Computational Linguistics* 19. 531–538.

Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use* (Cambridge Approaches to Linguistics). Cambridge: Cambridge University Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Bierwisch, Manfred. 1983a. Formal and lexical semantics. *Linguistische Studien* 114. 56–79.

Bierwisch, Manfred. 1983b. Major aspects of the psychology of language. *Linguistische Studien* 114. 1–38.

Bierwisch, Manfred & Ewald Lang (eds.). 1989. *Dimensional adjectives: Grammatical structure and conceptual interpretation*. Berlin: Springer.

Biggam, Carol P. 2010. The development of the basic colour terms of English. In Alaric Hall, Olga Timofeeva, Ágnes Kiricsi & Bethany Fox (eds.), *Interfaces between language and culture in medieval England: A festschrift for Matti Kilpiö*, 231–266. Leiden & Boston: Brill.

Biggam, Carol P. 2012. *The semantics of colour. A historical approach*. Cambridge: Cambridge University Press.

Blank, Andreas. 1999. Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change. In Andreas Blank & Peter Koch (eds.), *Historical semantics and cognition*, 61–89. Berlin & New York: Mouton de Gruyter.

Bolinger, Dwight. 1977. *Meaning and form*. London: Longman.

Bréal, Michel. 1897. *Essai de sémantique: science des significations*. Paris: Hachette.

Breban, Tine & Henrik De Smet. 2019. How do grammatical patterns emerge? The origins and development of the English proper noun modifier construction. *English Language and Linguistics* 23(4). 879–899.

Brezina, Vaclav, Tony McEnery & Stephen Wattam. 2015. Collocations in context. A new perspective on collocation networks. *International Journal of Corpus Linguistics* 20(2). 139–173.

Brinton, Laurel J. & Elizabeth Closs Traugott. 2005. *Lexicalization and language change*. Cambridge & New York: Cambridge University Press.

Brown, Gillian & George Yule. 1983. *Discourse analysis*. Cambridge & New York: Cambridge University Press.

Brysbaert, Marc, Amy Beth Warriner & Victor Kuperman. 2014. Concreteness rating for 40 thousand generally known English word lemmas. *Behavior Research Methods* 46(3). 904–911.

Cantos Gómez, Pascual. 2013. *Statistical methods in language and linguistic research*. Sheffield and Bristol: Equinox.

Carnoy, Albert. 1927. *La science des mots: Traité de sémantique*. Leuven: Editions Universitas.

Chadwyck Healey Literature Online. 1996-2020. Cambridge: Chadwyck Healey Ltd. (Bell & Howell Information and Learning Company). http://collections.chadwyck.com/marketing/list_of_all.jsp

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: Massachusetts Institute of Technology Press.

Chung, Siaw-Fong. 2011. A corpus-based analysis of "create" and "produce." *Chang Gung Journal of Humanities and Social Sciences* 4(2). 399–425.

Church, Kenneth W., William Gale, Patrick Hanks & Donald Hindle. 1991. Using statistics in lexical analysis. In Uri Zernik (ed.), *Lexical Exploiting On-line resources to build a lexicon*, 115–164. New Jersey, Hove, and London: Lawrence Erlbaum.

Church, Kenneth W., William Gale, Donald Hindle & Rosamund Moon. 1994. Lexical Substitutability. In Beth Levin & Antonio Zampolli (eds.), *Computational approaches to the lexicon*, 153–177. Oxford & New York: Oxford University Press.

Church, Kenneth W. & Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16. 22–29.

Clear, Jeremy. 1993. From Firth principles: computational tools for the study of collocation. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.), *Text and Technology: In honour of John Sinclair*, 271–292. Amsterdam: John Benjamins.

Cole, Peter & Jerry Morgan (eds.). 1975. *Syntax and semantics*, *Vol 3: Speech acts*. New York: Academic Press.

Corbett, Greville G. 1981. Syntactic features. *Journal of Linguistics* 17. 55–76.

Coseriu, Eugenio. 1962. *Teoría del lenguaje y lingüística general: cinco estudios*. Madrid: Gredos.

Coseriu, Eugenio. 1964. Pour une sémantique diachronique structurale. *Travaux de linguistique et de littérature* 2. 139–186.

Croft, William. 2000. *Explaining language change. An evolutionary approach.* Essex: Pearson Education.

Cruse, D. Alan. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Cruse, D. Alan. 2000. *Meaning in language: An introduction to semantics and pragmatics*. Oxford & New York: Oxford University Press.

Cruse, D. Alan. 2002. Paradigmatic relations of inclusion and identity III: Synonymy. In D. Alan Cruse, Franz Hundsnurcher, Michael Job & Peter Wolf Lutzeier (eds.), *Lexicology. An international handbook on the nature and structure of words and vocabularies*, 485–497. Berlin: Mouton.

Crystal, David. 1985. *A dictionary of linguistics and phonetics*. Oxford & New York: Blackwell.

Csardi, Gavor & Tamas Nepusz. 2006. The igraph software package for complex network research. *InterJournal Complex Systems* (1695). 1–9.

Culy, Christopher. 1996. Null objects in English recipies. *Language Variation and Change* 8(1). 91–124.

D'hoedt, Frauke, Henrik De Smet & Hubert Cuyckens. 2019. Constructions waxing and waning: A diachronic overview of the zero-Secondary Predicate Construction. *Journal of English Linguistics* 47(1). 3–28.

De Smet, Henrik, Frauke D'hoedt, Lauren Fonteyn & Kristel Van Goethem. 2018. The changing functions of competing forms: Attraction and differentiation. *Cognitive Linguistics* 29(2). 197–234.

Desagulier, Guillaume. 2014. Visualizing distances in a set of near-synonyms: *rather*, *quite*, *fairly*, and *pretty*. In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus methods for semantics. Quantitative studies in polysemy and synonymy*, 145–178. Amsterdam & Philadelphia: John Benjamins.

Desagulier, Guillaume. 2017. *Corpus linguistics and statistics with R: Introduction to quantitative methods in linguistics*. Cham: Springer.

Digonnet, Rémi. 2018. The linguistic expression of smells: from lack to abundance. In Annalisa Baicchi, Rémi Digonnet & Jodi. L Sandford (eds.), *Sensory perceptions in language, embodiment and epistemology*, 177–192. Cham: Springer.

Divjak, Dagmar. 2006. Ways of intending: Delineating and structuring near synonyms. In Stefan Th. Gries & Anatol Stefanowitsch (eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*, 19–56. Berlin & New York: Mouton de Gruyter.

Divjak, Dagmar. 2010. *Structuring the lexicon: A clustered model for near-synonymy*. Berlin & New York: Mouton de Gruyter.

Divjak, Dagmar & Stefan Th. Gries. 2006. Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2(1). 23–60.

Divjak, Dagmar & Stefan Th. Gries. 2008. Clusters in the mind? Converging evidence from near synonymy in Russian. *The Mental Lexicon* 3(2). 188–213.

Durkin, Philip. 2009. *The Oxford guide to etymology*. Oxford: Oxford University Press.

Durkin, Philip. 2014. *Borrowed words: A history of loanwords in English*. Oxford: Oxford University Press.

Early English Books Online. Chadwyck Healey. 2003-2020. Available at: http://eebo.chadwyck.com/home.

Edmonds, Philip. 1999. *Semantic representations of near-synonyms for automatic lexical choice*. Toronto: University of Toronto dissertation.

Edmonds, Philip & Graeme Hirst. 2002. Near-Synonymy and Lexical Choice. *Computational Linguistics* 28(2). 105–144.

Ellis, Nick C., Rita Simpson-Vlach & Carson Maynard. 2008. Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly* 42. 375–396.

Erdmann, Karl-Otto. 1910. *Die Bedeutung des Wortes: Aufsätze aus dem Grenzgebeit der Sprachpsychologie und Logik*. Leipzig: Avenarius.

Esnault, Gaston. 1925. *Métaphores occidentales: essai sur les valeurs imaginatives concrètes du français parlé en Basse-Bretagne comparé avec les patois, parlers techniques et argots français*. Paris: Presses Universitaires de France.

Evert, Stefan. 2005. *The statistics of word cooccurrences: Word pairs and collocations*. Stuttgart: University of Stuttgart dissertation.

Fellbaum, Christiane (ed.). 1998. *WordNet: An electronic lexical database*. Cambridge, MA: Massachusetts Institute of Technology Press.

Fellbaum, Christiane. 2005. WordNet and wordnets. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics,* 665–670. 2nd edn. Oxford: Elsevier.

Fillmore, Charles J. 1968. The case for case. In Emmon Bach & Robert Thomas Harms (eds.), *Universals in Linguistic Theory*, 1–88. New York: Holt Rinehart & Winston.

Fillmore, Charles J. 1977. Scenes-and-frames semantics. In Antonio Zampolli (ed.), *Linguistic structures processing*, 55–81. Amsterdam: New Holland.

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6. 222–254.

Fillmore, Charles J. & B. T. Sue Atkins 1992. Toward a frame-based lexicon: The semantics of *risk* and its neighbors. In Adrienne Lehrer & Eva Feder Kittay (eds.), *Frames, fields and contrasts: New essays in semantic and lexical organization*, 75–102. Hillsdale, NJ: Erlbaum.

Firth, John R. 1935. The technique of semantics. *Transactions of the Philological Society* 36–72.

Firth, John R. 1957a. A synopsis of linguistic theory 1930-1955. In John R. Firth (ed.), *Studies in linguistic analysis*, 1–32. Oxford: Philological Society.

Firth, John R. 1957b. *Papers in linguistics*. Oxford: Oxford University Press.

Gablasova, Dana, Vaclav Brezina & Tony McEnery. 2017. Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning* 67. 155–179.

Gabrielatos, Costas & Paul Baker. 2008. Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK press 1996-2005. *Journal of English Linguistics* 36(1). 5–38.

Geeraerts, Dirk. 1986. On necessary and sufficient conditions. *Journal of Semantics* 5(4). 275–291.

Geeraerts, Dirk. 1988. Where does prototypicality come from? In Brygida Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics*, 207–229. Amsterdam & Philadelphia: John Benjamins.

Geeraerts, Dirk. 1997. *Diachronic prototype semantics*. Oxford: Clarendon Press.

Geeraerts, Dirk. 2002. The interaction of metaphor and metonymy in composite expressions. In René Dirven & Ralf Pörings (eds.), *Metaphor and metonymy in comparison and contrast*, 435–468. Berlin & New York: Mouton de Gruyter.

Geeraerts, Dirk. 2006a. Prototype theory. In Dirk Geeraerts (ed.), *Cognitive Linguistics: Basic readings*, 141–165. Berlin: Mouton de Gruyter.

Geeraerts, Dirk. 2006b. Introduction: A rough guide to Cognitive Linguistics. In Dirk Geeraerts (ed.), *Cognitive Linguistics: Basic readings*, 1–28. Berlin: Mouton de Gruyter.

Geeraerts, Dirk. 2010. *Theories of lexical semantics*. Oxford: Oxford University Press.

Geeraerts, Dirk & Hubert Cuyckens. 2007. Introducing Cognitive Linguistics. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford Handbook of Cognitive Linguistics*, 1–21. Oxford & New York: Oxford University Press.

Geeraerts, Dirk, Stefan Grondelaers & Peter Bakema. 1994. *The structure of lexical variation: Meaning, naming, and context*. Berlin & New York: Mouton de Gruyter.

Gipper, Helmut. 1959. *Sessel* oder *Stuhl*? Ein Beitrag zur Bestimmung von Wortinhalten im Bereich der Sachkultur. In Helmut Gipper (ed.), *Sprache, Schlüssel zur Welt: Festschrift für Leo Weisgerber*, 271–292. Düsseldorf: Schwann.

Glynn, Dylan. 2010. Corpus-driven cognitive semantics. Introduction to the field. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: corpus-driven approaches*, 1–41. Berlin: Mouton de Gruyter.

Goddard, Cliff & Anna Wierzbicka (eds.). 2002. *Meaning and universal grammar: Theory and empirical findings*. Amsterdam: Benjamins.

Goodenough, Ward. H. 1956. Componential analysis and the study of meaning. *Language* 32. 195–216.

Goossens, Louis. 1990. Metaphtonymy: The interaction of metaphor and metonymy in expressions for linguistic action. *Cognitive Linguistics* 1(3). 323–340.

Granger, Sylviane. 1983. *The be + past participle construction in spoken English with special emphasis on the passive*. Amsterdam: North-Holland.

Grice, Paul. 1989 [1975]. Logic and conversation. In Paul Grice (ed.), *Studies in the way of words*, 22–40. Cambridge, MA: Harvard University Press. (Originally published in Peter Cole & Jerry Morgan 1975, 41–58.)

Gries, Stefan Th. 2001. A corpus-linguistic analysis of English *-ic* vs *-ical* adjectives. *ICAME Journal* 25. 65–108.

Gries, Stefan Th. 2003. Testing the sub-test: An analysis of *-ic* and *-ical* adjectives. *International Journal of Corpus Linguistics* 8(1). 31–60.

Gries, Stefan Th. 2005. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research* 34(4). 365–399.

Gries, Stefan Th. 2009. *Statistics for linguistics with R*. Berlin: Mouton de Gruyter.

Gries, Stefan Th. 2010. Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon* 5(3). 323–346.

Gries, Stefan Th. 2013. 50-something years of work on collocations: what is or should be next. *International Journal of Corpus Linguistics* 18(1). 137–166.

Gries, Stefan Th. & Dagmar Divjak. 2009. Behavioral profiles: A corpus-based approach to cognitive semantic analysis. In Vyvyan Evans & Stéphanie Pourcel (eds.), *New directions in cognitive linguistics*. Amsterdam & Philadelphia: John Benjamins.

Gries, Stefan Th. & Naoki Otani. 2010. Behavioral profiles: corpus-based perspective on synonymy and antonymy. *ICAME Journal* 34. 121–150.

Grondelaers, Stefan & Dirk Geeraerts. 2003. Towards a pragmatic model of cognitive onomasiology. In Hubert Cuyckens, René Dirven & John R. Taylor (eds.), *Cognitive approaches to lexical semantics*, 66–92. Berlin: Mouton de Gruyter.

Grondelaers, Stefan, Dirk Speelman & Dirk Geeraerts. 2007. Lexical variation and change. *The Oxford Handbook of Cognitive Linguistics*, 988–1011. Oxford & New York: Oxford University Press.

Gullick, David, Paul Rayson, John Mariani, Scott Piao & François Taiani. 2010. *CONE: COllocational Network Explorer*. Lancaster University. http://ucrel.lancaster.ac.uk/cone/.

Halliday, M.A.K. 1991. Corpus studies and probabilistic grammar. In Karin Aijmer & Bengt Altenberg (eds.), *English corpus linguistics*, 30–43. London: Longman.

Halliday, M.A.K. 1992. Language as system and language as instance. In Jan Svartvik (ed.), *Directions in corpus linguistics: Proceedings of Nobel symposium 81 Stockholm, 4-8 August 1991*, 61–77. Berlin: Mouton de Gruyter.

Hanks, Patrick. 1996. Contextual dependency and lexical sets. *International Journal of Corpus Linguistics* 1. 75–98.

Hanks, Patrick. 2013. *Lexical analysis: Norms and exploitations*. Cambridge, MA: The MIT Press.

Harris, Randy. 1993. *The linguistics war*. Oxford: Oxford University Press.

Harris, Zellig. 1954. Distributional structure. *Word* 10. 146–162.

Heylen, Kris, Yves Peirsman, Dirk Geeraerts & Dirk Speelman. 2008. Modelling word similarity: An evaluation of automatic synonymy extraction algorithms. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis & Daniel Tapias (eds.), *Proceedings of the sixth International conference on language resources and evaluation*, 3243–3249. Marrakech: European Language Resource Association.

Heylen, Kris, Dirk Speelman & Dirk Geeraerts. 2012. Looking at word meaning. An interactive visualization of semantic vector spaces for dutch synsets. In *Proceedings of the EACL-2012 joint workshop of LINGVIS & UNCLH: Visualization of language patterns and uncovering language history from multilingual resources*, 16–24. Stroudsburg: Association for Computational Linguistics.

Heylen, Kris, Thomas Wielfaert, Dirk Speelman & Dirk Geeraerts. 2015. Monitoring polysemy: Word space models as a tool for large-scale lexical semantic analysis. *Lingua* 157. 153–172.

Hilpert, Martin. 2011. Dynamic visualizations of language change: Motion charts on the basis of bivariate and multivariate data from diachronic corpora. *International Journal of Corpus Linguistics* 16(4). 435–461.

Hilpert, Martin. 2013. *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press.

Hilpert, Martin. 2016. Change in modal meanings: Another look at the shifting collocates of *may*. *Constructions and Frames* 8(1). 66-85.

Hilpert, Martin & David Correia Saavedra. 2017. Using token-based semantic vector spaces for corpus-linguistic analyses: From practical applications to tests of theoretical claims. *Corpus Linguistics and Linguistic Theory* Ahead of print. 1–32.

Hirst, Graeme. 1995. Near-synonymy and the structure of lexical knowledge. In Judith Klavans (ed.), *AAAI Symposium on representation and acquisition of lexical knowledge: Polysemy, ambiguity, and generativity*, 51–56. Cambridge, MA: AAAI Press.

Hoey, Michael. 1991. *Patterns of lexis in text*. Oxford: Oxford University Press.

Hoffmann, Sebastian. 2004. Using the *OED* quotations database as a corpus - a linguistic appraisal. *ICAME Journal* 28. 17–30.

Hopper, Paul J. & Elizabeth Traugott. 2003. *Grammaticalization* (Cambridge Textbooks in Linguistics). 2nd edn. Cambridge: Cambridge University Press.

Hothorn, Torsten, Peter Buehlmann, Sandrine Dudoit, Annette Molinaro & Mark Van Der Laan. 2006. Survival ensembles. *Biostatistics* 7(3). 355–373.

Huddleston, Rodney, Geoffrey Pullum, Laurie Bauer, Betty Birner, Ted Briscoe, Peter Collins, David Denison, et al. 2002. *The Cambridge grammar of the English language*. Cambridge & New York: Cambridge University Press.

Ibarretxe-Antuñano, Iraide. 1999. Metaphorical mappings in the sense of smell. In Raymond W. Gibbs Jr. & Gerard J. Steen (eds.), *Metaphor in cognitive linguistics*, 29–45. Amsterdam & Philadelphia: John Benjamins.

Inkpen, Diana. 2007. A statistical model for near-synonym choice. *ACM Transactions on Speech and Language Processing* 4(1). 1–17.

Inkpen, Diana & Graeme Hirst. 2006. Building and using a lexical knowledge base of near-synonym differences. *Computational Linguistics* 32(2). 223–262.

Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: Massachusetts Institute of Technology Press.

Jackendoff, Ray. 1996. Conceptual semantics and cognitive linguistics. *Cognitive Linguistics* 7. 93–129.

Jackendoff, Ray. 2006. on conceptual semantics. *Intercultural Pragmatics* 3(3). 353–358.

Jackson, Howard. 1988. *Words and their meanings*. London: Longman.

Jansegers, Marlies & Stefan Th. Gries. 2020. Towards a dynamic behavioral profile: A diachronic study of polysemous *sentir* in Spanish. *Corpus Linguistics and Linguistic Theory* 16 (1). 145–187.

Johnson, Paul. 1997. *A history of the American people*. New York: Harper Collins.

Jones, Maldwyn A. 1996. *Historia de Estados Unidos 1607-1992* (Carmen Martínez Gimeno, Trans.). Madrid: Cátedra.

Justeson, John S. & Slava M. Katz. 1995. Principles disambiguation: Discriminating adjective senses with modified nouns. *Computational Linguistics* 21(1). 1–27.

Katz, Jerrold J. 1972. *Semantic theory*. New York: Harper & Row.

Katz, Jerrold J. & Jerry A. Fodor. 1963. The structure of semantic theory. *Language* 39. 170–210.

Kaunisto, Mark. 2001. Nobility in the history of adjectives ending in *-ic* and *-ical*. In Ruth Brend, Alan K. Melby & Arle R. Lommel (eds.), *LACUS Forum XXVII: Speaking and comprehending*, 35–46. Fullerton: LACUS.

Kay, Christian. 2010. What is the *Historical Thesaurus of the OED*? In *Oxford English Dictionary*. Oxford: Oxford University Press. https://www.oed.com/public/whatishtoed/what-is-the-historical-thesaurus/loginpage (24 April, 2020).

Kay, Christian. 2012. Developing *The Historical Thesaurus of the OED*. In Kathryn Allan & Justyna A. Robinson (eds.), *Current methods in historical semantics*, 41–58. Berlin & Boston: De Gruyter Mouton.

Kay, Christian & Kathryn Allan. 2015. *English historical semantics*. Edinburgh: Edinburgh University Press.

Kennedy, Graeme. 1991. *Between* and *through*: The company they keep and the function they serve. In Karin Aijmer & Bengt Altenberg (eds.), *English corpus linguistics: Studies in honour of Jan Svartvik*, 95–110. London: Longman.

Kiela, Douwe & Stephen Clark. 2014. A systematic study of semantic vector space model parameters. In *Proceedings of the second workshop on continuous vector space models and their compositionality (CVSC)*, 21–30. Gothenburg: Association for Computational Linguistics. https://www.aclweb.org/anthology/W14-1503/.

Kilgarriff, Adam. 2006. Word senses. In Eneko Agirre & Philip Edmonds (eds.), *Word sense disambiguation: Algorithms and applications* (Text, Speech and Language Technology), 29–46. Berlin: Springer.

Kilgarriff, Adam & Pavel Rychlý. 2010. Semi-automatic dictionary drafting download. In Gilles-Maurice de Schryver (ed.), *A way with words: Recent advances in lexical theory and analysis. A Festschrift for Patrick Hanks*, 299–312. Kampala: Menha Publishers.

Kjellmer, Göran. 2003. Synonymy and corpus work: On *almost* and *nearly*. *ICAME Journal* 27. 19–27.

Klenová, Dominika. 2010. *The language of cookbooks and recipes*. Brno: Masaryk University MA thesis.

König, Ekkerhard & Elizabeth Traugott. 1988. Pragmatic strengthening and semantic change: The conventionalizing of conversational implicature. In Werner Hüllen & Rainer Schulze (eds.), *Understanding the lexicon: Meaning, sense and world knowledge in lexical semantics*, 110–124. Tübingen: Niemeyer.

Kövecses, Zoltan. 2010. *Metaphor: A practical introduction*. Oxford: Oxford University Press.

Krawczak, Karolina. 2014. *Shame* and its near-synonyms in English: A multivariate corpus-driven approach to social emotions. In Iva Novakova, Peter Blumenthal & Dirk Siepmann (eds.), *Les émotions dans le discourse / Emotions in discourse*, 83–94. Frankfurt am Main: Peter lang.

Krawczak, Karolina. 2018. Reconstructing social emotions across languages and cultures: A multifactorial account of the adjectival profiling of SHAME in English, French, and Polish. *Review of Cognitive Linguistics* 16(2). 455–493.

Kruskal, Joseph B. & Myron Wish. 1978. *Quantitative applications in the social sciences: Multidimensional scaling*. Thousand Oaks, CA: SAGE Publications.

Lakoff, George. 1970. Linguistics and natural logic. *Synthese* 22. 151–271.

Lakoff, George. 1971. On generative semantics. In Danny D. Steinberg & Leon A. Jakobovits (eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology*, 232–296. Cambridge: Cambridge University Press.

Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago & London: The University of Chicago Press.

Lakoff, George. 1993. The contemporary theory of metaphor. In Andrew Ortony (ed.), *metaphor and thought*, 202–251. 2nd edn. Cambridge: Cambridge University Press.

Lakoff, George & Mark Johnson. 2003. *Metaphors we live by*. Chicago & London: The University of Chicago Press.

Landes, Shari, Claudia Leacock & Christiane Fellbaum. 1998. Building semantic concordances. In Christiane Fellbaum (ed.), *WordNet: An electronic lexical database*, 199–216. Cambridge, MA: Massachusetts Institute of Technology Press.

Lang, Ewald. 1993. The meaning of German projective prepositions: A two-level approach. In Cornelia Zelinsky-Wibbelt (ed.), *The semantics of propositions: From mental processing to natural language processing*, 249–91. Berlin: Mouton de Gruyter.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar. Theoretical prerequisites*. Vol. 1. Stanford CA: Stanford University Press.

Leech, Geoffrey. 1974. *Semantics: The study of meaning*. 2nd edn. London: Penguin Books.

Leech, Geoffrey & Roger Fallon. 1992. Computer corpora – what do they tell us about culture? *ICAME Journal* 16. 29–50.

Leech, Geoffrey & Nicholas Smith. 2005. Extending the possibilities of corpus-based research on English in the twentieth century: A prequel to LOB and FLOB. *ICAME Journal* 29. 83–98.

Lehmann, Winfred P. 1988. *Prototypes in language and cognition*. Ann Arbor: Karoma.

Lehrer, Adrienne. 1969. Semantic cuisine. *Journal of Linguistics* 5. 39–55.

Lehrer, Adrienne. 1974. *Semantic fields and lexical structure*. Amsterdam & London: North Holland.

Leith, Dick. 1983. *A social history of English*. London & Boston: Routledge & Kegan Paul.

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.

Levshina, Natalia. 2011. Doe wat je niet laten kan*: A usage-based analysis of Dutch causative constructions*. Leuven: KU Leuven dissertation.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam & Philadelphia: John Benjamins.

Lewandowska-Tomaszcyk, Barbara. 2007. Polysemy, prototypes, and radial categories. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford Handbook of Cognitive Linguistics*, 139–169. Oxford & New York: Oxford University Press.

Liu, Dilin. 2010. Is it a *chief*, *main*, *major*, *primary* or *principal concern*? A corpus-based behavioral profile study of the near-synonyms. *International Journal of Corpus Linguistics* 15(1). 56–87.

Liu, Dilin. 2013a. Using corpora to help teach difficult-to-distinguish English words. *English Teaching* 68(3). 27–50.

Liu, Dilin. 2013b. Salience and construal in the use of synonymy: A study of two sets of near-synonymous nouns. *Cognitive Linguistics* 24(1). 67–113.

Liu, Dilin & Maggie Espino. 2012. *Actually, genuinely*, *really*, and *truly*. A corpus-based behavioral profile study of the near-synonymous adverbs. *International Journal of Corpus Linguistics* 17. 198–228.

Liu, Dilin & Shouman Zong. 2016. L2 vs L1 use of synonymy: An empirical study of synonym use/acquisition. *Applied Linguistics* 37(2). 239–261.

López-Couso, María José. 2017. Inferential-based approaches. In Laurel J. Brinton (ed.), *English historical linguistics: Approaches and perspectives*, 158–184. Cambridge: Cambridge University Press.

Lorig, Tyler S. 1999. On the similarity of odor and language perception. *Neuroscience and Biobehavioral Reviews* 23. 391–398.

Lyons, John. 1963. *Structural semantics*. Oxford: Blackwell.

Lyons, John. 1968. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.

Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.

Lyons, John. 1981a. *Language and linguistics*. Cambridge: Cambridge University Press.

Lyons, John. 1981b. *language, meaning, and context*. Bungay: Fontana.

Lyons, John. 1995. *Linguistic semantics*. Cambridge: Cambridge University Press.

MacLaury, Robert E. 1987. Coextensive semantic ranges: Different names for distinct vantages of one category. In *Papers from the Twenty-third Regional Meeting of the Chicago Linguistic Society*, 268–282. Chicago: Chicago Linguistic Society.

MacLaury, Robert E. 1995. Vantage theory. In John R. Taylor & Robert E. MacLaury (eds.), *Language and the cognitive construal of the world*, 231–276. Berlin: Mouton de Gruyter.

MacLaury, Robert E. 1997. *Color and cognition in Mesoamerica: Constructing categories as vantages*. Austin: Texas University Press.

Martin, Marylin. 1984. Advanced vocabulary teaching: The problem of synonyms. *The Modern Language Journal* 68(2). 130–137.

Massam, Diane & Yves Roberge. 1989. Recipe context null objects in English. *Linguistic Inquiry* 20. 134–139.

McEnery, Tony. 2006a. *Swearing in English: Bad language, purity and power from 1586 to the present*. Abington: Routledge.

McEnery, Tony. 2006b. The moral panic about bad language in England, 1691-1745. *Journal of Historical Pragmatics* 7(1). 89–113.

McEnery, Tony & Helen Baker. 2017. *Corpus linguistics and 17th-century prostitution: Computational linguistics and history*. London: Bloomsbury Academic.

McEnery, Tony, Richard Xiao & Yukio Tono. 2006. *Corpus-based language studies: An advanced resource book*. Abingdon & New York: Routledge.

Mel'čuk, Igor A. 1988. Semantic description of lexical units in an Explanatory Combinatorial Dictionary. *International Journal of Lexicography* 1. 165–188.

Mel'čuk, Igor A. 1996. Lexical functions: a tool for the description of lexical relations in a lexicon. In Leo Wanner (ed.), *Lexical functions in lexicography and natural language processing*, 37–102. Amsterdam: Benjamins.

Mel'čuk, Igor A. 1998. Collocations and lexical functions. In Anthony P. Cowie (ed.), *Phraseology. Theory, analysis, and applications*, 23–63. Oxford: Clarendon Press.

Mel'čuk, Igor A., André Clas & Nadia Arbatchewsky-Jumarie. 1984. *Dictionnaire explicatif et combinatoire du français contemporain: recherses lexico-sémantiques I-IV*. Montréal: Presses de l'Université de Montréal.

Miller, George A. 1990. *WordNet*: an on-line lexical database. *International Journal of Lexicography* 3. 235–312.

Miller, George A. & Walter G. Charles. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6(1). 1–28.

Mondorf, Britta. 2010. Variation and change in English resultative constructions. *Language Variation and Change* 22(3). 397–421.

Moon, Rosamund. 2013. Braving synonymy: From data to dictionary. *International Journal of Lexicography* 26(3). 260–278.

Murphy, M. Lynne. 2003. *Semantic relations and the lexicon: Antonymy, synonymy, and other paradigms*. Cambridge: Cambridge University Press.

Murphy, M. Lynne. 2013. What we talk about when we talk about synonyms (and what it can tell us about thesauruses). *International Journal of Lexicography* 26(3). 279–304.

Nattinger, James R. & Jeanette S. DeCarrico. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Nesselhauf, Nadja. 2005. *Collocations in a learner corpus* (Studies in Corpus Linguistics 14). Amsterdam: John Benjamins.

Nida, Eugene A. 1951. A system for the description of semantic elements. *Word* 7. 1–14.

Nuyts, Jan & Pieter Byloo. 2015. Competing modals: Beyond (inter)subjectification. *Diachronica* 32(1). 34–68.

Nyrop, Kristoffer. 1913. *Grammaire historique de la langue française IV: Sémantique*. Copenhagen: Gyldendalske Boghandel Nordisk Forlag.

Oakes, Michael P. 2003. Contrasts between US and British English in the 1991. In Elzbieta H. Oleksy & Barbara Lewandowska-Tomaszcyk (eds.), *Research and scholarship in integration processes*, 213–222. Lódz: Lódz University Press.

Padó, Sebastian & Mirella Lapata. 2007. Dependency-based construction of Semantic Space Models. *Computational Linguistics* 33(2). 161–199.

Panther, Klaus-Uwe & Günther Radden. 1999. *Metonymy in language and thought*. Amsterdam: John Benjamins.

Paquot, Magali & Sylviane Granger. 2012. Formulaic language in learner corpora. *Annual Review of Applied Linguistics* 32. 130–149.

Partington, Alan. 1998. *Patterns and meanings: Using corpora for English language research and teaching* (Studies in Corpus Linguistics). Amsterdam & Philadelphia: John Benjamins.

Paul, Hermann. 1920. *Prinzipien der Sprachgeschichte*. Halle: Niemeyer.

Peirsman, Yves. 2008. Word Space Models of semantic similarity and relatedness. In *Proceedings of the ESSLLI-2008 student session*, 1–10. Hamburg: ESSLII.

Peirsman, Yves, Kris Heylen & Dirk Geeraerts. 2008. Size matters. Tight and loose context definitions in English word space models. In Marco Baroni, Stefan Evert & Alessandro Lenci (eds.), *Proceedings of the ESSLLI workshop on distributional lexical semantics: Bridging the gap between semantic theory and computational linguistics*, 34–41. Hamburg.

Peirsman, Yves, Kris Heylen & Dirk Geeraerts. 2010. Applying word space models to sociolinguistics. Religion names before and after 9/11. In Dirk Geeraerts, Gitte Kristiansen & Yves Peirsman (eds.), *Recent advances in cognitive sociolinguistics*, 111–137. Berlin & New York: Mouton de Gruyter.

Pennebaker, James W., Roger J. Booth, Ryan L. Boyd & Martha E. Francis. 2015. *Linguistic Inquiry and Word Count: LIWC2015*. Austin: Pennebaker Conglomerates. https://liwc.wpengine.com/.

Persson, Gunnar. 1989. Deep *and* profound*: A study in so-called synonymy*. Umeå: Printing office of Umeå University.

Pettersson-Traba, Daniela. 2015. *A preliminary corpus-based diachronic analysis of the behavioral profile of a set of near-synonyms in American English*. Santiago de Compostela: University of Santiago de Compostela MA thesis.

Pettersson-Traba, Daniela. 2018. A diachronic perspective on near-synonymy: The concept of SWEET-SMELLING in American English. *Corpus Linguistics and Linguistic Theory* Ahead of Print. 1–31.

Pettersson-Traba, Daniela. 2019. Measuring semantic distance across time: An analysis of the collocational profiles of a set of near synonyms in American English. *Journal of Research Design and Statistics in Linguistics and Communication Science* 6(2). 1–28.

Pettersson-Traba, Daniela. 2020a. Modeling the choice of near-synonyms in the recent history of American English: The case of *fragrant*, *perfumed* and *scented*". In Pilar Guerrero-Medina, Macarena Palma Gutiérrez & María Valero Redondo (eds.), *Advances in English and American Studies: Current developments, future trends*, 149–162. Córdoba: UCOPress / Editorial Universidad de Córdoba.

Pettersson-Traba, Daniela. 2020b. Lexical variation in the history of American English: Attraction and differentiation in a synonym set. In Lourdes López Ropero, Sara Prieto García-Cañedo & José Antonio Sánchez Fajardo (eds.), *Thresholds and ways forward in English studies*, 235–248. Alicante: Publicaciones de la Universidad de Alicante.

Pettersson-Traba, Daniela. Forthcoming. Distributional changes in synonym sets: The case of *fragrant*, *scented*, and *perfumed* in 19th- and 20th-century American English. In Chris Cummins, Lisa Gotthard, Alpo Honkapohja & Benjamin Molineaux (eds.), *Historical English in contact* (Studies in Language Companion Series). Amsterdam: John Benjamins.

Phillips, Martin. 1985. *Aspects of text structure: An investigation of the lexical organisation of text*. Amsterdam: North-Holland.

Phillips, Martin. 1989. *Lexical structure of text* (Discourse Analysis Monograph). Vol. 12. Birmingham: University of Birmingham.

Porzig, Walter. 1934. Wesenhafte Bedeutungsbeziehungen. *Beiträge zur Geschichte der deutschen Sparche und Literatur* 58. 70–97.

Pottier, Bernard. 1964. Vers une sémantique moderne. *Travaux de linguistique et de littérature* 2. 107–137.

Pottier, Bernard. 1965. La définition sémantique dans les dictionnaires. *Travaux de linguistique et de littérature* 3. 33–39.

Potts, Amanda & Paul Baker. 2012. Does semantic tagging identify cultural change in British and American English? *International Journal of Corpus Linguistics* 17(3). 295–324.

Primahadi-Wijaya-R, Gede & I Made Rajeg. 2014. Visualising diachronic change in the collocational profiles of lexical near-synonyms. In I Nengah Sudipa & Gede Primahadi-Wijaya-R (eds.), *Cahaya bahasa: A Festschrift in honour of Prof. I Gusti Made Sutjaja*, 247–258. Denpasar: Swasta Nulus.

Pustojevsky, James. 1995. *The Generative Lexicon*. Cambridge, MA: Massachusetts Institute of Technology Press.

Quine, Willard Van Orman. 1951. Main trends in recent philosophy: Two dogmas of empiricism. *The Philosophical Review* 60. 20–43.

Quirk, Randolph, Sydney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London & New York: Longman.

R Core Team. 2017. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Ravin, Yael. 1992. Synonymy from a computational point of view. In Adrienne Lehrer & Eva Feder Kittay (eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization*, 397–419. New York & Abington: Routledge.

Rayson, Paul. 2008. From key words to key semantic domains. *International Journal of Corpus Linguistics* 13(4). 519–549.

Rohdenburg, Günter. 2003. Cognitive complexity and horror aequi as factors determining the use of interrogative clause linkers in English. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English* (Topics in English Linguistics 43), 205–249. Berlin & New York: Mouton de Gruyter.

Rosch, Eleanor. 1973a. Natural categories. *Cognitive Psychology* 4. 328–350.

Rosch, Eleanor. 1973b. On the internal structure of perceptual and semantic categories. In Timothy E. Moore (ed.), *Cognitive development and the aquisition of language*, 111–144. New York: Academic Press.

Rosenbach, Anette. 2008. Animacy and grammatical variation – Findings from English dative variation. *Lingua* 118. 151–171.

Rubenstein, Herbert & John B. Goodenough. 1965. Contextual correlates of synonymy. *Communications of the ACM* 8(10). 627–633.

Ruíz de Mendoza, Francisco José & Olga Isabel Díez. 2002. Patterns of conceptual interaction. In René Dirven & Ralf Pörings (eds.), *Metaphor and metonymy in comparison and contrast*, 489–532. Berlin & New York: Mouton de Gruyter.

Ruíz de Mendoza, Francisco José & Alicia Galera-Masegosa. 2011. Going beyond metaphtonymy: metaphoric and metonymic complexes in phrasal verb interpretation. *Language Value* 3(1). 1–29.

Sahlgren, Magnus. 2006. *The word-space model. Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. Stockholm: Stockholm University dissertation.

Samuels, Michael L. 1972. *Linguistic evolution with special reference to English*. London & New York: Cambridge University Press.

Saussure, Ferdinand de. 1916. *Course in general linguistics*. (Roy Harris, Trans). London: Duckworth.

Schmid, Hans-Jörg. 1993. *Cottage und. co., idea, start vs. begin. Die Kategorisierung als Grundprinzip einer differenzierten Bedeutungsbeschreibung*. Tübinger: Max Niemeyer.

Schmid, Hans-Jörg. 2007. Entrenchment, salience, and basic levels. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford Handbook of Cognitive Linguistics*, 117–138. Oxford & New York: Oxford University Press.

Schulz, Philip & Wilker Aziz. 2016. Fast collocation-based Bayesian HMM word alignment. In Yuji Matsumoto & Rashmi Prasad (eds.), *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 3146–3155. Osaka: COLING 2016 Organizing Committee.

Scott, Mike. 2020. *WordSmith Tools*. Stroud: Lexical Analysis Software Ltd.

Shank, Christopher, Koen Plevoets & Hubert Cuyckens. 2014. A diachronic corpus-based multivariate analysis of "I think *that*" vs. "I think zero". In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 279-303. Amsterdam & Philadelphia: John Benjamins.

Sharpe, Donald. 2015. Chi-square test is statistically significant: Now what? *Practical Assessment, Research, and Evaluation* 20(8). 1–10.

Sinclair, John M. 1966. Beginning the study of lexis. In Charles E. Bazell (ed.), *In memory of J.R. Firth*, 410–429. Harlow: Longman.

Sinclair, John M. 1987. Collocation: A progress report. In Ross Steele & Terry Threadgold (eds.), *Language topics: Essays in honor of Michael Halliday*. Vol. 2, 319–332. Amsterdam & Philadelphia: John Benjamins.

Sinclair, John M. 1991a. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, John M. 1991b. Shared knowledge. In James Atlatis (ed.), *Linguistics and language pedagogy. The state of the art* (Georgetown University Round Table on Languages and Linguistics), 489–500. Washington D.C: Georgetown University Press.

Sinclair, John M. 1996. The search for units of meaning. *Textus* 9. 75–106.

Sinclair, John M. 2004. *Trust the text: Language, corpus and discourse*. London: Routledge.

Soares da Silva, Augusto. 2013. Comparing objective and subjective linguistic distances between European and Brazilian Portuguese. In Monika Reif, Justyna A. Robinson & Martin Pütz (eds.), *Variation in language and language use: Linguistic, socio-cultural and cognitive perspectives*, 241–269. Frankfurt, Berlin & New York: Peter lang.

Soares da Silva, Augusto. 2014. The pluricentricity of Portuguese: A sociolectometrical approach to divergence between European and Brazilian Portuguese. In Augusto Soares da Silva (ed.), *Pluricentricity: Language variation and sociocognitive dimensions*, 143–188. Berlin & Boston: De Gruyter Mouton.

Soares da Silva, Augusto. 2015. Competition of synonyms through time: Conceptual and social salience factors and their interrelations. *Catalan Journal of Linguistics* 14. 199–2018.

Speelman, Dirk. 2014. Logistic regression: A confirmatory technique for comparisons in corpus linguistics. In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*, 487–533. Amsterdam & Philadelphia: John Benjamins.

Speelman, Dirk, Stefan Grondelaers & Dirk Geeraerts. 2003. Profile-based linguistic uniformity as a generic method for comparing language varieties. *Computers and the Humanities* 37. 317–337.

Sperber. 1975. *Rethinking symbolism*. Cambridge: Cambridge University Press.

Stefanowitsch, Anatol. 2008. Words and their metaphors: A corpus-based approach. In Anatol Stefanowitsch & Stefan Th. Gries (eds.), *Corpus-based approach to metaphor and metonymy*. Berlin & New York: Mouton de Gruyter.

Steinbach, Michael, Pang-Ning Tan & Vipin Kumar. 2005. *Introduction to data mining*. Boston: Longman.

Stern, Gustaf. 1931. *Meaning and change of meaning, with special reference to the English language*. Gothenburg: Elanders Bocktryckeri Aktiebolag.

Storjohann, Petra. 2009. Plesionymy: A case of synonymy or contrast? *Journal of Pragmatics* 41. 2140–2158.

Strobl, Carolin, Ann-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9(307). http://www.biomedcentral.com/1471-2105/9/307.

Strobl, Carolin, Ann-Laure Boulesteix, Achim Zeileis & Torsten Hothorn. 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* 8(25). http://www.biomedcentral.com/1471-2105/8/25

Stubbs, Michael. 1993. British traditions in text analysis: from Firth to Sinclair. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.), *Text and technology: In honour of John Sinclair*, 1–36. Amsterdam: John Benjamins.

Stubbs, Michael. 2002a. *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Stubbs, Michael. 2002b. Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics* 7(2). 215–244.

Szmrecsanyi, Benedikt. 2005. Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1(1). 113–149.

Szmrecsanyi, Benedikt. 2006. *Morphosyntactic persistence in spoken English. A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin & New York: Mouton de Gruyter.

Szmrecsanyi, Benedikt & Bernd Kortmann. 2009. The morphosyntax of varieties of English worldwide: A quantitative perspective. *Lingua* 119. 1643–1663.

Szmrecsanyi, Benedikt. 2017. Featurometry. In Martijn Wieling, Martin Kroon, Gertjan van Noord & Gosse Bouma (eds.), *From semantics to dialectometry. Festschrift in honor John Nerbonne* (Tributes 32), 345–353. UK: College Publications

Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: *was*/*were* variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178.

Taylor, John R. 1995. *Linguistic categorization*. 2nd edn. Oxford & New York: Oxford University Press.

Taylor, John R. 2003. Near synonyms as co-extensive categories: 'high' and 'tall' revisited. *Language Sciences* 25(3). 263–284. https://doi.org10.1016/S0388-0001(02)00018-9.

Todolí, Julia & Rosanna Dolón. 2006. Translating cooking recipes in English, Spanish and Catalan. In Cristina Mourón-Figueroa & Teresa Iciar Moralejo-Gárate (eds.), *Studies in contrastive linguistics: Proceedings of the 4th International Contrastive Linguistic Conference*, 1051–1060. Santiago de Compostela: Servicio de Publicacións da Universidade de Santiago de Compostela.

Trask, Robert L. 1993. *A dictionary of grammatical terms in linguistics*. London: Routledge.

Trask, Robert L. 2007. *Language and linguistics: The key concepts*. 2nd edn. Abingdon & New York: Routledge.

Traugott, Elizabeth & Richard Dasher. 2002. *Regularity in semantic change*. Cambridge & New York: Cambridge University Press.

Trier, Jost. 1931. *Der deutsche Wortschatz im Sinnbezirk des Verstandes: Die Geschichte eines sprachlichen Feldes I. Von den Anfängen bis zum Beginn des 13. Jhdts.* Heidelberg: Winter.

Trier, Jost. 1934. Das sprachliche Feld: eine Auseinandersetzung. *Neue Jahrbücher für Wissenschaft und Jugendbildung* 10. 428–449.

Turkkila, Kaisa. 2014. Do near-synonyms occur with the same metaphors: A comparison of anger terms in American English. *Metaphorik* 25. 129–154.

Ullmann, Stephen. 1957. *The principles of semantics*. 2nd edn. Oxford: Basil Blackwell.

Ullmann, Stephen. 1962. *Semantics: An introduction to the science of meaning*. Oxford: Blackwell.

Verhagen, Arie. 2007. Construal and perspectivization. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford Handbook of Cognitive Linguistics*, 48–81. Oxford & New York: Oxford University Press.

Vosberg, Uwe. 2003. The role of extractions and horror aequi in the evolution of *-ing*-complements in Modern English. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English* (Topics in English Linguistics 43), 305–327. Berlin & New York: Mouton de Gruyter.

Waag, Albert. 1908. *Bedeutungsentwicklung unseres Wortschatzes: Ein Blicke in das Seelenleben der Wörter*. 2nd edn. Lahr: Schauenburg.

Weischedel, Ralph & Ada Brunstein. 2005. *BBN Pronoun Coreference and Entity Type Corpus LDC2005T33*. Philadelphia: Linguistic Data Consortium.

Wickelmaier, Florian. 2003. *An introduction to MDS*. Aalborg: Aalborg Universitetsforlag.

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

Wiegand, Herbert Ernst. 1999. Synonymy and its role in monolingual lexicography. In Antje Immken & Werner Wolski (eds.), *Semantics and lexicography: Selected studies (1976-1996)* (Lexicographica Series Maior), 11–53. Tübingen & Niemeyer: Max Niemeyer.

Wierzbicka, Anna. 1972. *Semantic primitives*. Frankfurt: Athenaeum.

Wierzbicka, Anna. 1988. *The semantics of grammar*. Amsterdam: John Benjamins.

Wierzbicka, Anna. 1992. Defining emotion concept. *Cognitive Science* 16(4). 539–581.

Wierzbicka, Anna. 1996. *Semantics. Primes and universals*. Oxford: Oxford University Press.

Wierzbicka, Anna. 1997. *Understanding cultures through their key words: English, Russian, Polish, German, and Japanese*. New York: Oxford University Press.

Wierzbicka, Anna. 1999. Emotional universals. *Language Design* 2. 23–69.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Oxford: Blackwell.

Xiao, Richard & Tony McEnery. 2006. Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics* 27(1). 103–129.

Yamamoto, Mutsumi. 1999. *Animacy and reference: A cognitive approach to corpus linguistics*. Amsterdam & Philadelphia: John Benjamins.

Yatandu Uba, Sani. 2015. A corpus-based behavioural profile study of near-synonyms: *Important*, *essential*, *vital*, *necessary* and *crucial*. *International Journal of English Language and Linguistics Research* 3(5). 9–17.

Yeh, Yuli, Hsien-Chin Liou & Yi-Hsin Li. 2007. Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning* 20. 131–152.

Yeshurun, Yara & Noam Sobel. 2010. An odor is not worth a thousand words: From multidimensional odors to unidimensional odor objects. *Annual Review of Psychology* 61. 219–241.

Zaenen, Annie, Jan Carlette, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, M.Catherine O'Connor & Tom Wasow. 2004. Animacy in coding in English: Why and how. In Donna Byron & Bonnie Webber (eds.), *Proceedings of the 2004 ACL workshop on discourse*, 118–125. Barcelona: Association for Computational Linguistics.

Ziff, Paul. 1964. *Semantic analysis*. Ithaca, New York: Cornell University Press.

Sources

AHDOE = The American Heritage Dictionary of the English Language. 2016–. Houghton Mifflin Harcourt. https://ahdictionary.com/ (December 16, 2020)

BNC = The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. http://www.natcorp.ox.ac.uk/. (December 16, 2020)

CD = Cambridge Dictionary. 2016–. Cambridge University Press. Available at: https://dictionary.cambridge.org/ (December 16, 2020)

COCA = Davies, Mark. 2008–. *The Corpus of Contemporary American English (COCA): 520 million words, 1990-present*. Available at: http://corpus.byu.edu/coca/.

CoD = Collins COBUILD Dictionary. 2012–. Collins. Available at: https://www.collinsdictionary.com/ (December 16, 2020)

COHA = Davies, Mark. 2010–. *The Corpus of Historical American English (COHA): 400 Million Words, 1810–2009*. Available at: https://corpus.byu.edu/coha/.

LDOCE = Longman Dictionary of Contemporary English. 1996–. Pearson. Available at: https://www.ldoceonline.com/ (December 16, 2020)

Lexico. 2019. Oxford University Press. Available at: https://www.lexico.com// (December 16, 2020)

MD = Macmillan Dictionary. 2009–2020. Macmillan Education Limited. Available at: https://www.macmillandictionary.com/ (December 16, 2020)

MW = Merriam-Webster Dictionary and Thesaurus. 2017–. Merriam-Webster. Available at: https://www.merriam-webster.com/ (December 16, 2020)

NHDAE = Newbury House Dictionary of American English. 2018–. Heinle. Available at: http://nhd.heinle.com/home.aspx (December 16, 2020)

OED = Oxford English Dictionary Online. 3rd edn. 2000–. Oxford University Press. Available at: <http://www.oed.com/> (December 16, 2020)

USAS = Archer, Dawn, Andrew Wilson & Paul Rayson. 2002. Introduction to the USAS category system. *Benedict Project Report* 1–37.

USAS = Rayson, Paul, Dawn Archer, Scott Piao & Tony McEnery. 2004. The UCREL semantic analysis system. In Louise Guthrie, Roberto Basili, Eva Hajicova & Frederick Jelinek (eds.), *Proceedings of the workshop on beyond named entity recognition semantic labelling for NLP tasks*, 7–12. Lisbon: LREC.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

APPENDIX A:
UNIVARIATE ANALYSES

Table A.1. Chi-square residuals of the frequency of the five adjectives in the period 1810–2009

Period / Synonym	P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)
Fragrant	3.59	3.47	-3.03	-4.81
Perfumed	-0.76	-0.23	-0.17	1.23
Scented	-6.24	-4.99	5.04	7.26
Sweet-scented	5.71	1.17	-2.57	-4.46
Sweet-smelling	-4.65	-3.68	3.97	5.14

Table A.2 Chi-square residuals of the concept PLEASANT SMELLING across senses in the period 1810–2009

Period / Sense	P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)
Artificial	-6.02	-5.34	4.48	8.09
Figurative	0.71	1.54	-0.16	-2.50
Indeterminate	-6.03	-2.55	3.95	5.09
Natural	5.32	3.51	-3.86	-5.71

Table A.3. Diachronic development of the semantic category SPACE according to sense

Sense / Period	P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
Artificial					
N	6	9	16	18	49
%	15.79	8.74	14.81	19.78	14.41
Figurative					
N	2	2	–	1	5
%	5.26	1.94		1.10	1.47
Indeterminate					
N	6	36	30	30	102
%	15.79	34.95	27.78	32.97	30.00
Natural					
N	24	56	62	42	184
%	63.16	54.37	57.41	46.15	54.12
Total					
N	38	103	108	91	340
%	100	100	100	100	100

Table A.4. Chi-square residuals of the concept PLEASANT SMELLING across semantic categories in the period 1810–2009

Period / Sem. Category	P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)
ABS	-0.24	1.85	0.18	-2.32
B&P	-4.82	-5.72	4.86	6.97
CL	-2.81	-3.76	1.58	5.94
COS	-4.51	-4.27	3.18	6.59
EAW	-1.75	3.02	1.95	-4.19
F&D	-0.49	-2.71	-1.21	5.22
OBJ	-1.29	-2.38	-0.15	4.48
P&F	8.48	5.38	-7.01	-7.92
S&M	-0.10	-0.92	1.08	0.13
SEN	4.68	0.20	-2.08	-2.71
SPA	-3.91	-1.33	3.15	2.27
T&C	-3.53	1.50	2.13	-0.65

Table A.5. Average frequency per million words in *COHA* of the semantic domains INDUSTRY AND TECHNOLOGY, ADVERTISING AND MEDIA, and COSMETICS AND HYGIENE in the period 1810–2009

Domain / Decade	INDUSTRY AND TECHNOLOGY	ADVERTISING AND MEDIA	COSMETICS AND HYGIENE
1810	12.87	7.11	1.31
1820	24.99	7.93	2.09
1830	22.46	9.36	3.29
1840	24.09	11.09	3.78
1850	26.84	9.43	5.58
1860	23.33	10.15	4.55
1870	27.71	11.69	4.81
1880	27.03	12.70	5.20
1890	28.03	14.92	5.15
1900	35.29	16.03	6.87
1910	39.39	18.63	9.26
1920	49.30	28.86	9.42
1930	46.83	35.73	10.34
1940	60.67	40.31	10.69
1950	46.16	40.03	11.06
1960	45.10	40.35	10.84
1970	44.64	40.94	11.50
1980	49.21	44.24	11.01
1990	42.02	42.70	14.82
2000	38.36	40.77	15.62

Table A.6. Sense distribution of *fragrant*

Period / Sense	P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
Artificial					
N	39	111	98	83	331
%	4.85	8.60	13.76	14.12	9.75
Figurative					
N	44	74	34	14	166
%	5.47	5.73	4.78	2.38	4.89
Indeterminate					
N	20	71	60	59	210
%	2.49	5.50	8.43	10.03	6.19
Natural					
N	701	1,035	520	432	2,688
%	87.19	80.17	73.03	73.47	79.18
Total					
N	804	1,291	712	588	3,395
%	100	100	100	100	100

Table A.7. Sense distribution of *perfumed*

Period / Sense	P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
Artificial					
N	82	125	112	114	433
%	42.49	37.54	49.34	50.89	44.32
Figurative					
N	6	16	11	12	45
%	3.11	4.80	4.85	5.36	4.61
Indeterminate					
N	12	37	41	49	139
%	6.22	11.11	18.06	21.88	14.23
Natural					
N	93	155	63	49	360
%	48.19	46.55	27.75	21.88	36.85
Total					
N	193	333	227	224	977
%	100	100	100	100	100

Table A.8. Sense distribution of *scented*

Period / Sense	P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
Artificial					
N	27	57	123	162	369
%	24.77	24.26	41.55	53.82	39.21
Figurative					
N	1	3	7	5	16
%	0.92	1.28	2.36	1.66	1.70
Indeterminate					
N	5	17	37	21	80
%	4.59	7.23	12.50	6.98	8.50
Natural					
N	76	158	129	113	476
%	69.72	67.23	43.58	37.54	50.58
Total					
N	109	235	296	301	941
%	100	100	100	100	100

Table A.9. Sense distribution of *sweet-scented*

Period / Sense	P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
Artificial					
N	7	2	4	–	13
%	8.97	2.56	13.79		6.56
Figurative					
N	1	2	2	–	5
%	1.28	2.56	6.90		2.53
Indeterminate					
N	2	1	4	2	9
%	2.56	1.28	13.79	15.38	4.55
Natural					
N	68	73	19	11	171
%	87.18	93.59	65.52	84.62	86.36
Total					
N	78	78	29	13	198
%	100	100	100	100	100

Table A.10. Sense distribution of *sweet-smelling*

Period / Sense	P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
Artificial					
N	2	14	23	26	65
%	10.53	26.42	25.56	28.57	25.69
Figurative					
N	2	–	–	1	3
%	10.53			1.10	1.19
Indeterminate					
N	–	5	11	20	36
%		9.43	12.22	21.98	14.23
Natural					
N	15	34	56	44	149
%	78.95	64.15	62.22	48.35	58.89
Total					
N	19	53	90	91	253
%	100	100	100	100	100

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table A.11. Semantic category distribution of *fragrant*

Period / Sem. category		P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
ABS						
	N	33	66	30	12	141
	%	4.10	5.11	4.21	2.04	4.15
B&P						
	N	17	45	60	57	179
	%	2.11	3.49	8.43	9.69	5.27
CL						
	N	2	6	6	13	27
	%	0.25	0.46	0.84	2.21	0.80
COS						
	N	4	8	15	10	37
	%	0.50	0.62	2.11	1.70	1.09
EAW						
	N	130	284	163	102	679
	%	16.17	22.00	22.89	17.35	20.00
F&D						
	N	80	116	84	129	409
	%	9.95	8.99	11.80	21.94	12.05
OBJ						
	N	19	20	16	16	71
	%	2.36	1.55	2.25	2.72	2.09
P&F						
	N	401	521	174	136	1,232
	%	49.88	40.36	24.44	23.13	36.29
S&M						
	N	41	80	61	32	216
	%	5.10	6.20	8.57	5.78	6.56
SEN						
	N	54	71	36	27	188
	%	6.72	5.50	5.06	4.59	5.54
SPA						
	N	20	53	55	41	169
	%	2.49	4.11	7.72	6.97	4.98
T&C						
	N	3	21	12	11	47
	%	0.37	1.63	1.69	1.87	1.38
Total						
	N	804	1,291	712	588	3,395
	%	100	100	100	100	100

Table A.12. Semantic category distribution of *perfumed*

Period / Sem. category		P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
ABS						
	N	5	13	10	10	38
	%	2.59	3.90	4.41	4.46	3.89
B&P						
	N	30	47	66	65	208
	%	15.54	14.11	29.07	29.02	21.29
CL						
	N	13	7	17	15	52
	%	6.74	2.10	7.49	6.70	5.32
COS						
	N	–	4	11	23	38
	%		1.20	4.85	10.27	3.89
EAW						
	N	33	80	44	20	177
	%	17.10	24.02	19.38	8.93	18.12
F&D						
	N	16	7	4	12	39
	%	8.29	2.10	1.76	5.36	3.99
OBJ						
	N	18	28	9	18	73
	%	9.33	8.41	3.96	8.04	7.47
P&F						
	N	11	40	11	6	68
	%	5.70	12.01	4.85	2.68	6.96
S&M						
	N	19	20	12	12	63
	%	9.84	6.01	5.29	5.36	6.45
SEN						
	N	28	12	3	4	47
	%	14.51	3.60	1.32	1.79	4.81
SPA						
	N	9	28	15	20	72
	%	4.66	8.41	6.61	8.93	7.73
T&C						
	N	11	47	25	19	102
	%	5.70	14.11	11.01	8.48	10.44
Total						
	N	193	333	227	224	977
	%	100	100	100	100	100

Table A.13. Semantic category distribution of *scented*

Period / Sem. category		P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
ABS	N	–	3	6	4	13
	%		1.28	2.03	1.33	1.38
B&P	N	10	12	40	34	96
	%	9.17	5.11	13.51	11.30	10.20
CL	N	3	16	23	40	82
	%	2.75	6.81	7.77	13.29	8.71
COS	N	–	3	15	25	43
	%		1.28	5.07	8.31	3.89
EAW	N	36	61	71	40	208
	%	33.03	25.96	23.99	13.29	22.10
F&D	N	3	12	9	14	38
	%	2.75	5.11	3.04	4.65	4.04
OBJ	N	5	15	28	50	98
	%	4.59	6.38	9.46	16.61	10.41
P&F	N	26	64	29	35	154
	%	23.85	27.23	9.80	11.63	16.37
S&M	N	8	12	16	21	57
	%	7.34	5.11	5.41	6.98	6.06
SEN	N	1	3	4	2	10
	%	0.92	1.28	1.35	0.66	1.06
SPA	N	9	15	27	25	76
	%	8.26	6.38	9.12	8.31	8.08
T&C	N	8	19	28	11	66
	%	7.34	8.09	9.46	3.65	10.44
Total	N	109	235	296	301	941
	%	100	100	100	100	100

Table A.14. Semantic category distribution of *sweet-scented*

Period / Sem. category		P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
ABS	N	–	2	2	–	4
	%		2.56	6.90		2.02
B&P	N	2	–	5	2	9
	%	2.56		17.24	15.38	4.55
CL	N	2	–	1	–	3
	%	2.56		3.45		1.52
COS	N	–	–	–	–	–
	%					
EAW	N	6	19	4	3	32
	%	7.69	24.36	13.79	23.08	16.16
F&D	N	1	1	2	–	4
	%	1.28	1.28	6.90		2.02
OBJ	N	2	1	–	1	4
	%	2.56	1.28		7.69	2.02
P&F	N	56	46	8	5	115
	%	71.79	58.97	27.59	38.46	58.08
S&M	N	8	2	4	2	16
	%	10.26	2.56	13.79	15.38	8.08
SEN	N	1	2	–	–	3
	%	1.28	2.56			1.52
SPA	N	–	5	2	–	7
	%		6.41	6.9		3.54
T&C	N	–	–	1	–	1
	%			3.45		0.58
Total	N	78	78	29	13	198
	%	100	100	100	100	100

Table A.15. Semantic category distribution of *sweet-smelling*

Period / Sem. category	P1 (1810–1859)	P2 (1860–1909)	P3 (1910–1959)	P4 (1960–2009)	Total
ABS					
N	2	–	–	1	3
%	10.53			1.10	1.19
B&P					
N	–	–	6	26	32
%			6.67	28.57	12.65
CL					
N	–	3	5	6	14
%		5.66	5.56	6.59	5.53
COS					
N	–	2	8	5	15
%		3.77	8.89	5.49	4.93
EAW					
N	3	4	14	8	29
%	15.79	7.55	15.56	8.79	11.46
F&D					
N	–	2	6	5	13
%		3.77	6.67	5.49	5.14
OBJ					
N	–	2	6	2	10
%		3.77	6.67	2.20	3.95
P&F					
N	7	26	27	18	78
%	36.84	49.06	30.00	19.78	30.83
S&M					
N	2	6	6	12	26
%	10.53	11.32	6.67	13.19	10.28
SEN					
N	5	4	2	2	13
%	26.32	7.55	2.22	2.20	5.14
SPA					
N	–	2	9	5	16
%		3.77	10.00	5.49	6.32
T&C					
N	–	2	1	1	4
%		3.77	1.11	1.10	1.58
Total					
N	19	53	90	91	253
%	100	100	100	100	100

Table A.16. Overall absolute and relative frequencies of the five near-synonyms in each sense

Sense / Synonym	Artificial	Figurative	Indeterminate	Natural	Total
Fragrant					
N	331	166	210	2,688	3,395
%	27.33	70.64	44.30	69.93	58.90
Perfumed					
N	433	45	139	360	977
%	35.76	19.15	29.32	9.37	16.95
Scented					
N	369	16	80	476	941
%	30.47	6.81	16.88	12.38	16.33
Sweet-scented					
N	13	5	9	171	198
%	1.07	2.13	1.90	4.45	3.44
Sweet-smelling					
N	65	3	36	149	253
%	5.37	1.28	7.59	3.88	4.39
Total					
N	1,211	235	474	3,844	5,764
%	100.00	100.00	100.00	100.00	100.00

Table A.17. Absolute and relative frequencies of the five near-synonyms in each sense in P1 (1810–1859)

Sense / Synonym	Artificial	Figurative	Indeterminate	Natural	Total
Fragrant					
N	39	44	20	701	804
%	24.84	81.48	51.28	73.56	66.83
Perfumed					
N	82	6	12	93	193
%	52.23	11.11	30.77	9.76	16.04
Scented					
N	27	1	5	76	109
%	17.20	1.85	12.82	7.97	9.06
Sweet-scented					
N	7	1	2	68	78
%	4.46	1.85	5.13	7.14	6.48
Sweet-smelling					
N	2	2	–	15	19
%	1.27	3.70		1.57	1.58
Total					
N	157	54	39	953	1,203
%	100.00	100.00	100.00	100.00	100.00

Table A.18. Absolute and relative frequencies of the five near-synonyms in each sense in P2 (1860–1909)

Sense / Synonym	Artificial	Figurative	Indeterminate	Natural	Total
Fragrant					
N	111	74	71	1,035	1,291
%	35.92	77.89	54.20	71.13	64.87
Perfumed					
N	125	16	37	155	333
%	40.45	16.84	28.24	10.65	16.73
Scented					
N	57	3	17	158	235
%	18.45	3.16	12.98	10.86	11.81
Sweet-scented					
N	2	2	1	73	78
%	0.65	2.11	0.76	5.02	3.92
Sweet-smelling					
N	14	–	5	34	53
%	4.53		3.82	2.34	2.66
Total					
N	309	95	131	1,455	1,990
%	100.00	100.00	100.00	100.00	100.00

Table A.19. Absolute and relative frequencies of the five near-synonyms in each sense in P3 (1910–1959)

Sense / Synonym	Artificial	Figurative	Indeterminate	Natural	Total
Fragrant					
N	98	34	60	520	712
%	27.22	62.96	39.22	66.07	52.58
Perfumed					
N	112	11	41	63	227
%	31.11	20.37	26.80	8.01	16.77
Scented					
N	123	7	37	129	296
%	34.17	12.96	24.18	16.39	21.86
Sweet-scented					
N	4	2	4	19	29
%	1.11	3.70	2.61	2.41	2.14
Sweet-smelling					
N	23	–	11	56	90
%	6.39		7.19	7.12	6.65
Total					
N	360	54	153	787	1,354
%	100.00	100.00	100.00	100.00	100.00

Table A.20. Absolute and relative frequencies of the five near-synonyms in each sense in P4 (1960–2009)

Sense / Synonym	Artificial	Figurative	Indeterminate	Natural	Total
Fragrant					
N	83	14	59	432	588
%	21.56	43.75	39.07	66.56	48.32
Perfumed					
N	114	12	49	49	224
%	29.61	37.50	32.45	7.55	18.41
Scented					
N	162	5	21	113	301
%	42.08	15.63	13.91	17.41	24.73
Sweet-scented					
N	–	–	2	11	13
%			1.32	1.69	1.07
Sweet-smelling					
N	26	1	20	44	91
%	6.75	3.13	13.25	6.78	7.48
Total					
N	385	32	151	649	1,217
%	100.00	100.00	100.00	100.00	100.00

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table A.21. Overall absolute and relative frequencies of the five near-synonyms in each semantic category

Sem. category / Synonym		ABS	B&P	CL	COS	EAW	F&D	OBJ	P&F	S&M	SEN	SPA	T&C	Total
Fragrant														
	N	141	179	27	37	679	409	71	1,232	216	188	169	47	3,395
	%	70.85	34.16	15.17	27.82	60.36	81.31	27.73	74.80	57.14	72.03	49.71	21.36	58.90
Perfumed														
	N	38	208	52	38	177	39	73	68	63	47	72	102	977
	%	19.10	39.69	29.21	28.57	15.73	7.75	28.52	4.13	16.67	18.01	21.18	46.36	16.95
Scented														
	N	13	96	82	43	208	38	98	154	57	10	76	66	941
	%	6.53	18.32	46.07	32.33	18.49	7.55	38.28	9.35	15.08	3.83	22.35	30.00	16.33
Sweet-scented														
	N	4	9	3	–	32	4	4	115	16	3	7	1	198
	%	2.01	1.72	1.69		2.84	0.80	1.56	6.98	4.23	1.15	2.06	0.45	3.44
Sweet-smelling														
	N	3	32	14	15	29	13	10	78	26	13	16	4	253
	%	1.51	6.11	7.87	11.28	2.58	2.58	3.91	4.74	6.88	4.98	4.71	1.82	4.39
Total														
	N	199	524	178	133	1,125	503	256	1,647	378	261	340	220	5,764
	%	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table A.22. Absolute and relative frequencies of the five near-synonyms in each semantic category in P1 (1810–1859)

Sem. category / Synonym		ABS	B&P	CL	COS	EAW	F&D	OBJ	P&F	S&M	SEN	SPA	T&C	Total
Fragrant														
	N	33	17	2	4	130	80	19	401	41	54	20	3	804
	%	80.50	28.81	10.00	100.00	62.50	80.00	43.18	80.04	52.56	60.67	52.63	13.64	66.83
Perfumed														
	N	5	30	13	–	33	16	18	11	19	28	9	11	193
	%	12.50	50.85	65.00		15.87	16.00	40.91	2.20	24.36	31.46	23.68	50.00	16.04
Scented														
	N	–	10	3	–	36	3	5	26	8	1	9	8	109
	%		16.95	15.00		17.31	3.00	11.36	5.19	10.26	1.12	23.68	36.36	9.06
Sweet-scented														
	N	–	2	2	–	6	1	2	56	8	1	–	–	78
	%		3.39	10.00		2.88	1.00	4.55	11.18	10.26	1.12			6.48
Sweet-smelling														
	N	2	–	–	–	3	–	–	7	2	5	–	–	19
	%	5.00				1.44			1.40	2.56	5.62			1.58
Total														
	N	40	59	20	4	208	100	44	501	78	89	38	22	1,203
	%	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table A.23. Absolute and relative frequencies of the five near-synonyms in each semantic category in P2 (1860–1909)

Sem. category / Synonym		ABS	B&P	CL	COS	EAW	F&D	OBJ	P&F	S&M	SEN	SPA	T&C	Total
Fragrant														
	N	66	45	6	8	284	116	20	521	80	71	53	21	1,291
	%	78.57	43.27	18.75	47.06	63.39	84.06	30.30	74.75	66.67	77.17	51.46	23.60	64.87
Perfumed														
	N	13	47	7	4	80	7	28	40	20	12	28	47	333
	%	15.48	45.19	21.88	23.53	17.86	5.07	42.42	5.74	16.67	13.04	27.18	52.81	16.73
Scented														
	N	3	12	16	3	61	12	15	64	12	3	15	19	235
	%	3.57	11.54	50.00	17.65	13.62	8.70	22.73	9.18	10.00	3.26	14.56	21.35	11.81
Sweet-scented														
	N	2	–	–	–	19	1	1	46	2	2	5	–	78
	%	2.38				4.24	0.72	1.52	6.60	1.67	2.17	4.85		3.92
Sweet-smelling														
	N	–	–	3	2	4	2	2	26	6	4	2	2	53
	%			9.38	11.76	0.89	1.45	3.03	3.73	5.00	4.35	1.94	2.25	2.66
Total														
	N	84	104	32	17	448	138	66	697	120	92	103	89	1,990
	%	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table A.24. Absolute and relative frequencies of the five near-synonyms in each semantic category in P3 (1910–1959)

Sem. category / Synonym		ABS	B&P	CL	COS	EAW	F&D	OBJ	P&F	S&M	SEN	SPA	T&C	Total
Fragrant														
	N	30	60	6	15	163	84	16	174	61	36	55	12	712
	%	62.50	33.90	11.54	30.61	55.07	80.00	27.12	69.88	61.62	80.00	50.93	17.91	52.58
Perfumed														
	N	10	66	17	11	44	4	9	11	12	3	15	25	227
	%	20.83	37.29	32.69	22.45	14.86	3.81	15.25	4.42	12.12	6.67	13.89	37.31	16.77
Scented														
	N	6	40	23	15	71	9	28	29	16	4	27	28	296
	%	12.50	22.60	44.23	30.61	23.99	8.57	47.46	11.65	16.16	8.89	25.00	41.79	21.86
Sweet-scented														
	N	2	5	1	–	4	2	–	8	4	–	2	1	29
	%	4.17	2.82	1.92		1.35	1.90		3.21	4.04		1.85	1.49	2.14
Sweet-smelling														
	N	–	6	5	8	14	6	6	27	6	2	9	1	90
	%		3.39	9.62	16.33	4.73	5.71	10.17	10.84	6.06	4.44	8.33	1.49	6.65
Total														
	N	48	177	52	49	296	105	59	249	99	45	108	67	1,354
	%	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table A.25. Absolute and relative frequencies of the five near-synonyms in each semantic category in P4 (1960–2009)

Sem. category / Synonym		ABS	B&P	CL	COS	EAW	F&D	OBJ	P&F	S&M	SEN	SPA	T&C	Total
Fragrant														
	N	12	57	13	10	102	129	16	136	34	27	41	11	588
	%	44.44	30.98	17.57	15.87	58.96	80.63	18.39	68.00	41.98	77.14	45.05	26.19	48.32
Perfumed														
	N	10	65	15	23	20	12	18	6	12	4	20	19	224
	%	37.04	35.33	20.27	36.51	11.56	7.50	20.69	3.00	14.81	11.43	21.98	45.24	18.41
Scented														
	N	4	34	40	25	40	14	50	35	21	2	25	11	301
	%	14.81	18.48	54.05	39.68	23.12	8.75	57.47	17.50	25.93	5.71	27.47	26.19	24.73
Sweet-scented														
	N	–	2	–	–	3	–	1	5	2	–	–	–	13
	%		1.09			1.73		1.15	2.50	2.47				1.07
Sweet-smelling														
	N	1	26	6	5	8	5	2	18	12	2	5	1	91
	%	3.70	14.13	8.11	7.94	4.62	3.13	2.30	9.00	14.81	5.71	5.49	2.38	7.48
Total														
	N	27	184	74	63	173	160	87	200	81	35	91	42	1,217
	%	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

APPENDIX B:
MULTIVARIATE ANALYSES

Table B.1. Distribution of the near-synonyms according to Animacy

Animacy / Synonym		Animate referents	Inanimate referents	Total
Fragrant				
	N	77	3,318	3,395
	%	27.40	60.51	58.90
Perfumed				
	N	127	850	977
	%	45.20	15.50	16.95
Scented				
	N	50	891	941
	%	17.79	16.25	16.33
Sweet-scented				
	N	7	191	198
	%	2.49	3.48	3.44
Sweet-smelling				
	N	20	233	253
	%	7.12	4.25	4.39
Total				
	N	281	5,483	5,764
	%	100.00	100.00	100.00

Table B.2. Mean and median Concreteness rating

Concreteness / Synonym	Mean	Median
Fragrant	4.34	4.59
Perfumed	4.34	4.61
Scented	4.47	4.76
Sweet-scented	4.57	4.76
Sweet-smelling	4.45	4.72

Table B.3. Distribution of the near-synonyms according to Countability

Countability / Synonym	Count	Non-count	Other	Total
Fragrant				
N	2,473	738	184	3,395
%	62.26	50.44	55.93	58.90
Perfumed				
N	630	277	70	977
%	15.86	18.93	21.28	16.95
Scented				
N	553	340	48	941
%	13.92	23.24	14.59	16.33
Sweet-scented				
N	148	40	10	198
%	3.73	2.73	3.04	3.44
Sweet-smelling				
N	168	68	17	253
%	4.23	4.65	5.17	4.39
Total				
N	3,972	1,463	329	5,764
%	100.00	100.00	100.00	100.00

Table B.4. Distribution of the near-synonyms according to Syntactic function

Syntactic function / Synonym	Attributive	Postpositive	Predicative	Other	Total
Fragrant					
N	2,434	304	537	120	3,395
%	57.07	60.44	71.31	49.38	58.90
Perfumed					
N	729	90	92	66	977
%	17.09	17.89	12.22	27.16	16.95
Scented					
N	712	100	95	34	941
%	16.69	19.88	12.62	13.99	16.33
Sweet-scented					
N	180	3	7	8	198
%	4.22	0.60	0.93	3.29	3.44
Sweet-smelling					
N	210	6	22	15	253
%	4.92	1.19	2.92	6.17	4.39
Total					
N	4,265	503	753	243	5,764
%	100.00	100.00	100.00	100.00	100.00

Table B.5. Distribution of the near-synonyms according to Degree

Synonym \ Degree		Positive	Comparative	Superlative	Total
Fragrant					
	N	3,245	130	20	3,395
	%	58.04	86.09	90.91	58.90
Perfumed					
	N	965	11	1	977
	%	17.26	7.28	4.55	16.95
Scented					
	N	934	6	1	941
	%	16.71	3.97	4.55	16.33
Sweet-scented					
	N	197	1	0	198
	%	3.52	0.66	0.00	3.44
Sweet-smelling					
	N	250	3	0	253
	%	4.47	1.99	0.00	4.39
Total					
	N	5,591	151	22	5,764
	%	100.00	100.00	100.00	100.00

Table B.6. Distribution of the near-synonyms according to Text-type

Synonym \ Text-type		Fiction	Non-fiction	Periodicals	Total
Fragrant					
	N	2,507	393	495	3,395
	%	57.78	61.21	63.22	58.90
Perfumed					
	N	772	99	106	977
	%	17.79	15.42	13.54	16.95
Scented					
	N	714	90	137	941
	%	16.46	14.02	17.50	16.33
Sweet-scented					
	N	149	33	16	198
	%	3.43	5.14	2.04	3.44
Sweet-smelling					
	N	197	27	29	253
	%	4.54	4.21	3.70	4.39
Total					
	N	4,339	642	783	5,764
	%	100.00	100.00	100.00	100.00

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table B.7. Period contrasts in fiction

Reference values: Fiction and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.2430	< 0.001	-2.8250	< 0.001	-4.6590	< 0.001
Period = P2	-0.4424	< 0.001	0.3775	0.022	0.3662	0.026
Period = P3	-0.5278	< 0.001	-0.0091	0.963	0.7923	< 0.001
Period = P4	-0.4498	0.003	-0.0220	0.919	0.6880	< 0.001
Reference values: Fiction and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.2100	< 0.001	-1.3180	< 0.001	-2.7950	< 0.001
Period = P1	0.4424	< 0.001	-0.3775	0.022	-0.3662	0.026
Period = P3	-0.0855	0.470	-0.3866	0.022	0.4261	0.003
Period = P4	-0.0075	0.955	-0.3995	0.035	0.3218	0.041
Reference values: Fiction and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.7150	< 0.001	-2.8340	< 0.001	-3.8670	< 0.001
Period = P1	0.5278	< 0.001	0.0091	0.963	-0.7923	< 0.001
Period = P2	0.0855	0.470	0.3866	0.022	-0.4261	0.003
Period = P4	0.0780	0.597	-0.0129	0.953	-0.1043	0.531
Reference values: Fiction and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.7930	< 0.001	-2.8470	< 0.001	-3.9710	< 0.001
Period = P1	0.4498	0.003	0.0220	0.919	-0.6880	< 0.001
Period = P2	0.0075	0.955	0.3995	0.035	-0.3218	0.041
Period = P3	-0.0780	0.597	0.0129	0.953	0.1043	0.531

Table B.8. Period contrasts in non-fiction

Reference values: Non-fiction and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.5730	< 0.001	-1.9340	< 0.001	-4.6540	< 0.001
Period = P2	0.5594	0.016	-0.8356	0.003	0.1643	0.620
Period = P3	0.0432	0.883	-0.9456	0.013	0.9187	0.011
Period = P4	-0.8726	0.013	-0.7128	0.082	1.6420	< 0.001
Reference values: Non-fiction and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.1330	< 0.001	-2.7700	< 0.001	-4.4890	< 0.001
Period = P1	-0.5594	0.016	0.8356	0.003	-0.1643	0.620
Period = P3	-0.5162	0.078	-0.1099	0.777	0.7544	0.025
Period = P4	-1.4320	< 0.001	0.1228	0.769	1.4770	< 0.001
Reference values: Non-fiction and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.6170	< 0.001	-2.8800	< 0.001	-3.7350	< 0.001
Period = P1	-0.0432	0.883	0.9456	0.013	-0.9187	0.011
Period = P2	0.5162	0.078	0.1099	0.777	-0.7544	0.025
Period = P4	-0.9158	0.020	0.2327	0.636	0.7230	0.064
Reference values: Non-fiction and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.7010	< 0.001	-2.6470	< 0.001	-3.0120	< 0.001
Period = P1	0.8726	0.013	0.7128	0.082	-1.6420	< 0.001
Period = P2	1.4320	< 0.001	-0.1228	0.769	-1.4770	< 0.001
Period = P3	0.9158	0.020	-0.2327	0.636	-0.7230	0.064

Table B.9. Period contrasts in periodicals

Reference values: Periodicals and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.9440	< 0.001	-2.2970	< 0.001	-4.8650	< 0.001
Period = P2	0.6740	0.088	-0.7803	0.092	-0.1818	0.768
Period = P3	-0.2989	0.414	-0.4576	0.291	1.0080	0.071
Period = P4	-0.2308	0.520	-0.8277	0.056	1.2010	0.029
Reference values: Periodicals and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.6180	< 0.001	-3.0770	< 0.001	-5.0470	< 0.001
Period = P1	-0.6740	0.088	0.7803	0.092	0.1818	0.768
Period = P3	-0.9730	< 0.001	0.3227	0.346	1.1900	0.001
Period = P4	-0.9049	0.001	-0.0475	0.890	1.3830	< 0.001
Reference values: Periodicals and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.6450	< 0.001	-2.7550	< 0.001	-3.8570	< 0.001
Period = P1	0.2989	0.414	0.4576	0.291	-1.0080	0.071
Period = P2	0.9730	< 0.001	-0.3227	0.346	-1.1900	0.001
Period = P4	0.0681	0.753	-0.3701	0.218	0.1931	0.415
Reference values: Periodicals and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.7130	< 0.001	-3.1250	< 0.001	-3.6640	< 0.001
Period = P1	0.2308	0.520	0.8277	0.056	-1.2010	0.029
Period = P2	0.9049	0.001	0.0475	0.890	-1.3830	< 0.001
Period = P3	-0.0681	0.753	0.3701	0.218	-0.1931	0.415

Table B.10. Period contrasts in the artificial sense

Reference values: Artificial and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	-0.8748	0.003	0.8153	0.008	-2.4200	< 0.001
Period = P2	0.3836	0.092	-0.3565	0.087	0.1515	0.566
Period = P3	0.1396	0.541	-0.8649	< 0.001	0.9657	< 0.001
Period = P4	-0.1680	0.468	-0.8866	< 0.001	1.2080	< 0.001
Reference values: Artificial and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	-0.4912	0.055	0.4587	0.102	-2.2680	< 0.001
Period = P1	-0.3836	0.092	0.3565	0.087	-0.1515	0.566
Period = P3	-0.2439	0.172	-0.5083	0.004	0.8143	< 0.001
Period = P4	-0.5515	0.003	-0.5301	0.002	1.0570	< 0.001
Reference values: Artificial and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	-0.7351	0.004	-0.0496	0.860	-1.4540	< 0.001
Period = P1	0.1396	0.541	0.8649	< 0.001	-0.9657	< 0.001
Period = P2	0.2439	0.172	0.5083	0.004	-0.8143	< 0.001
Period = P4	-0.3076	0.092	-0.0218	0.900	0.2424	0.145
Reference values: Artificial and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	-1.0430	< 0.001	-0.0714	0.803	-1.212	< 0.001
Period = P1	0.1680	0.468	0.8866	< 0.001	-1.208	< 0.001
Period = P2	0.5515	0.003	0.5301	0.002	-1.057	< 0.001
Period = P3	0.3076	0.092	0.0218	0.900	-0.2424	0.145

Table B.11. Period contrasts in the figurative sense

Reference values: Figurative and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.9750	< 0.001	-1.6900	0.001	-4.4620	< 0.001
Period = P2	-0.7619	0.128	0.7815	0.148	0.6122	0.603
Period = P3	-1.2210	0.019	0.7742	0.181	1.9960	0.069
Period = P4	-2.0690	< 0.001	1.8540	0.002	2.0090	0.077
Reference values: Figurative and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.2130	< 0.001	-0.9084	0.006	-3.8500	< 0.001
Period = P1	0.7619	0.128	-0.7815	0.148	-0.6122	0.603
Period = P3	-0.4593	0.255	-0.0072	0.987	1.3840	0.055
Period = P4	-1.3070	0.004	1.073	0.024	1.3960	0.071
Reference values: Figurative and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	0.7539	0.025	-0.9157	0.018	-2.4660	< 0.001
Period = P1	1.2210	0.019	-0.7742	0.181	-1.9960	0.069
Period = P2	0.4593	0.255	0.0072	0.987	-1.3840	0.055
Period = P4	-0.8477	0.079	1.0800	0.037	0.0121	0.985
Reference values: Figurative and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	-0.0938	0.817	0.1643	0.696	-2.4540	< 0.001
Period = P1	2.0690	< 0.001	-1.8540	0.002	2.0090	0.077
Period = P2	1.3070	0.004	-1.0730	0.024	-1.3960	0.071
Period = P3	0.8477	0.079	-1.0800	0.037	-0.0121	0.985

Table B.12. Period contrasts in the indeterminate value

Reference values: Indeterminate and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	0.2288	0.576	-0.0688	0.875	-2.6930	< 0.001
Period = P2	0.1689	0.667	-0.1550	0.708	-0.0410	0.941
Period = P3	-0.3658	0.346	-0.1936	0.637	0.8383	0.111
Period = P4	-0.3678	0.348	0.2826	0.489	0.2221	0.684
Reference values: Indeterminate and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	0.3977	0.152	-0.2238	0.464	-2.7340	< 0.001
Period = P1	-0.1689	0.667	0.1550	0.708	0.0410	0.941
Period = P3	-0.5346	0.036	-0.0386	0.889	0.8793	0.008
Period = P4	-0.5366	0.039	0.4376	0.108	0.2631	0.464
Reference values: Indeterminate and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	-0.1369	0.616	-0.2624	0.385	-1.8550	< 0.001
Period = P1	0.3658	0.346	0.1936	0.637	-0.8383	0.111
Period = P2	0.5346	0.036	0.0386	0.889	-0.8793	0.008
Period = P4	-0.0020	0.994	0.4761	0.074	-0.6162	0.048
Reference values: Indeterminate and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	-0.1389	0.614	0.2138	0.474	-2.4710	< 0.001
Period = P1	0.3678	0.348	-0.2826	0.489	-0.2221	0.684
Period = P2	0.5366	0.039	-0.4376	0.108	-0.2631	0.464
Period = P3	0.0020	0.994	-0.4761	0.074	0.6162	0.048

Table B.13. Period contrasts in the natural sense

Reference values: Natural and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.6530	< 0.001	-1.6960	< 0.001	-3.1610	< 0.001
Period = P2	-0.4424	< 0.001	0.3775	0.022	0.3662	0.026
Period = P3	-0.5278	< 0.001	-0.0091	0.963	0.7923	< 0.001
Period = P4	-0.4498	0.003	-0.0220	0.919	0.6880	< 0.001
Reference values: Natural and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.2100	< 0.001	-1.3180	< 0.001	-2.7950	< 0.001
Period = P1	0.4424	< 0.001	-0.3775	0.022	-0.3662	0.026
Period = P3	-0.0855	0.470	-0.3866	0.022	0.4261	0.003
Period = P4	-0.0075	0.955	-0.3995	0.035	0.3218	0.041
Reference values: Natural and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.1250	< 0.001	-1.7050	< 0.001	-2.3690	< 0.001
Period = P1	0.5278	< 0.001	0.0091	0.963	-0.7923	< 0.001
Period = P2	0.0855	0.470	0.3866	0.022	-0.4261	0.003
Period = P4	0.0780	0.597	-0.0129	0.953	-0.1043	0.531
Reference values: Natural and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.2030	< 0.001	-1.718	< 0.001	-2.4730	< 0.001
Period = P1	0.4498	0.003	0.0220	0.919	-0.6880	< 0.001
Period = P2	0.0075	0.955	0.3995	0.035	-0.3218	0.041
Period = P3	-0.0780	0.597	0.0129	0.953	0.1043	0.531

Table B.14. Period contrasts in the semantic category ABSTRACT

Reference values: ABSTRACT and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.9710	< 0.001	-2.9840	< 0.001	-17.8300	0.939
Period = P2	-0.7673	0.182	0.5320	0.368	12.3800	0.958
Period = P3	-1.2430	0.036	0.6469	0.299	13.5400	0.954
Period = P4	-1.9690	0.002	1.6130	0.014	13.4900	0.954
Reference values: ABSTRACT and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.2040	< 0.001	-2.4520	< 0.001	-5.4530	< 0.001
Period = P1	0.7673	0.182	-0.5320	0.368	-12.3800	0.958
Period = P3	-0.4761	0.270	0.1149	0.812	1.1650	0.117
Period = P4	-1.2020	0.017	1.0810	0.038	1.1190	0.171
Reference values: ABSTRACT and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.7270	< 0.001	-2.3370	< 0.001	-4.2880	< 0.001
Period = P1	1.2430	0.036	-0.6469	0.299	-13.5400	0.954
Period = P2	0.4761	0.270	-0.1149	0.812	-1.1650	0.117
Period = P4	-0.7255	0.167	0.9665	0.083	-0.0460	0.949
Reference values: ABSTRACT and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.0020	< 0.001	-1.3700	0.016	-4.3340	< 0.001
Period = P1	1.9690	0.002	-1.6130	0.014	-13.4900	0.954
Period = P2	1.2020	0.017	-1.0810	0.038	-1.1190	0.171
Period = P3	0.7255	0.167	-0.9665	0.083	0.0461	0.949

Table B.15. Period contrasts in the semantic category BODY AND PEOPLE

Reference values: BODY AND PEOPLE and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.3180	0.006	-0.9218	0.072	-4.2350	< 0.001
Period = P2	0.5917	0.106	-0.2369	0.487	-0.4698	0.322
Period = P3	0.2712	0.435	-0.4686	0.145	0.4528	0.262
Period = P4	0.1799	0.608	-0.2864	0.375	0.3387	0.408
Reference values: BODY AND PEOPLE and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.9100	< 0.001	-1.1590	0.015	-4.7050	< 0.001
Period = P1	-0.5917	0.106	0.2369	0.487	0.4698	0.322
Period = P3	-0.3204	0.225	-0.2317	0.373	0.9227	0.011
Period = P4	-0.4117	0.126	-0.0495	0.850	0.8085	0.029
Reference values: BODY AND PEOPLE and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.5890	< 0.001	-1.3900	0.003	-3.7820	< 0.001
Period = P1	-0.2712	0.435	0.4686	0.145	-0.4528	0.262
Period = P2	0.3204	0.225	0.2317	0.373	-0.9227	0.011
Period = P4	-0.0913	0.707	0.1822	0.438	-0.1142	0.676
Reference values: BODY AND PEOPLE and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.4980	< 0.001	-1.2080	0.010	-3.8960	< 0.001
Period = P1	-0.1799	0.608	0.2864	0.375	-0.3387	0.408
Period = P2	0.4117	0.126	0.0495	0.850	-0.8085	0.029
Period = P3	0.0913	0.707	-0.1822	0.438	0.1142	0.676

Table B.16. Period contrasts in the semantic category CLEANING

Reference values: CLEANING and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	0.4160	0.626	-0.1547	0.826	-4.6240	< 0.001
Period = P2	0.5136	0.565	-1.9540	0.005	1.9970	0.008
Period = P3	0.1861	0.832	-1.4710	0.017	1.6110	0.023
Period = P4	0.7400	0.366	-2.1480	< 0.001	1.8380	0.007
Reference values: CLEANING and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	0.9297	0.131	-2.1090	0.001	-2.6270	< 0.001
Period = P1	-0.5136	0.565	1.9540	0.005	-1.9970	0.008
Period = P3	-0.3276	0.615	0.4831	0.368	-0.3859	0.434
Period = P4	0.2264	0.692	-0.1937	0.715	-0.1587	0.733
Reference values: CLEANING and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	0.6021	0.309	-1.6260	0.003	-3.0130	< 0.001
Period = P1	-0.1861	0.832	1.4710	0.017	-1.6110	0.023
Period = P2	0.3276	0.615	-0.4831	0.368	0.3859	0.434
Period = P4	0.5539	0.310	-0.6768	0.118	0.2272	0.567
Reference values: CLEANING and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.1560	0.023	-2.3030	< 0.001	-2.7860	< 0.001
Period = P1	-0.7400	0.366	2.1480	< 0.001	-1.8380	0.007
Period = P2	-0.2264	0.692	0.1937	0.715	0.1587	0.733
Period = P3	-0.5539	0.310	0.6768	0.118	-0.2272	0.567

Table B.17. Period contrasts in the semantic category COSMETICS

Reference values: COSMETICS and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	14.7700	0.927	-14.5600	0.957	-18.2000	0.980
Period = P2	-12.5500	0.938	12.7000	0.962	14.1500	0.984
Period = P3	-12.9800	0.936	12.5700	0.963	14.7800	0.984
Period = P4	-13.8300	0.932	13.1200	0.961	14.8500	0.984
Reference values: COSMETICS and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.2240	0.001	-1.8600	0.012	-4.0520	< 0.001
Period = P1	12.5500	0.938	-12.700	0.962	-14.1500	0.984
Period = P3	-0.4351	0.495	-0.1297	0.852	0.6271	0.400
Period = P4	-1.2830	0.047	0.4244	0.518	0.6992	0.331
Reference values: COSMETICS and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.7890	0.001	-1.9900	0.001	-3.4250	< 0.001
Period = P1	12.9800	0.936	-12.5700	0.963	-14.7800	0.984
Period = P2	0.4351	0.495	0.1297	0.852	-0.6271	0.400
Period = P4	-0.8477	0.086	0.5540	0.227	0.0722	0.870
Reference values: COSMETICS and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	0.9410	0.074	-1.4360	0.007	-3.3530	< 0.001
Period = P1	13.8300	0.932	-13.1200	0.961	-14.8500	0.984
Period = P2	1.2830	0.047	-0.4244	0.518	-0.6992	0.331
Period = P3	0.8477	0.086	-0.5540	0.227	-0.0722	0.870

Table B.18. Period contrasts in the semantic category EARTH, ATMOSPHERE, AND WEATHER

Reference values: EARTH, ATMOSPHERE, AND WEATHER and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.8300	< 0.001	-2.6100	< 0.001	-4.1360	< 0.001
Period = P2	0.0084	0.964	0.2352	0.314	-0.2636	0.267
Period = P3	-0.2866	0.155	-0.0151	0.954	0.4003	0.094
Period = P4	-0.0225	0.922	-0.2712	0.386	0.2233	0.408
Reference values: EARTH, ATMOSPHERE, AND WEATHER and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.3220	<0.001	-1.3550	<0.001	-2.8840	<0.001
Period = P1	-0.0084	0.964	-0.2352	0.314	0.2636	0.267
Period = P3	-0.295	0.079	-0.2503	0.240	0.6639	0.001
Period = P4	-0.0309	0.878	-0.5064	0.066	0.4868	0.041
Reference values: EARTH, ATMOSPHERE, AND WEATHER and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.5440	< 0.001	-2.6250	< 0.001	-3.7360	< 0.001
Period = P1	0.2866	0.155	0.0151	0.954	-0.4003	0.094
Period = P2	0.2950	0.079	0.2503	0.240	-0.6639	0.001
Period = P4	0.2641	0.217	-0.2561	0.389	-0.1771	0.459
Reference values: EARTH, ATMOSPHERE, AND WEATHER and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.8080	< 0.001	-2.881	< 0.001	-3.9130	< 0.001
Period = P1	0.0225	0.922	0.2712	0.386	-0.2233	0.408
Period = P2	0.0309	0.878	0.5064	0.066	-0.4868	0.041
Period = P3	-0.2641	0.217	0.2561	0.389	0.1771	0.459

Table B.19. Period contrasts in the semantic category FOOD AND DRINK

Reference values: FOOD AND DRINK and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	4.1380	< 0.001	-2.9980	< 0.001	-6.4690	< 0.001
Period = P2	-0.1323	0.726	-0.7915	0.109	1.4460	0.033
Period = P3	0.1063	0.798	-1.2020	0.044	1.3790	0.049
Period = P4	-0.0441	0.906	-0.3885	0.379	1.1280	0.095
Reference values: FOOD AND DRINK and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	4.0060	< 0.001	-3.7890	< 0.001	-5.0240	< 0.001
Period = P1	0.1323	0.726	0.7915	0.109	-1.4460	0.033
Period = P3	0.2386	0.547	-0.4107	0.526	-0.0670	0.887
Period = P4	0.0882	0.802	0.4030	0.427	-0.3179	0.464
Reference values: FOOD AND DRINK and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	4.2450	< 0.001	-4.2000	< 0.001	-5.0910	< 0.001
Period = P1	-0.1063	0.798	1.2020	0.044	-1.3790	0.049
Period = P2	-0.2386	0.547	0.4107	0.526	0.0670	0.887
Period = P4	-0.1504	0.701	0.8137	0.181	-0.2509	0.593
Reference values: FOOD AND DRINK and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	4.0940	< 0.001	-3.3860	< 0.001	-5.3420	< 0.001
Period = P1	0.0441	0.906	0.3885	0.379	-1.1280	0.095
Period = P2	-0.0882	0.802	-0.4030	0.427	0.3179	0.464
Period = P3	0.1504	0.701	-0.8137	0.181	0.2509	0.593

Table B.20. Period contrasts in the semantic category OBJECT

Reference values: OBJECT and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.0350	< 0.001	-1.3070	0.018	-4.6850	< 0.001
Period = P2	-0.7578	0.075	0.1823	0.657	0.9720	0.086
Period = P3	-0.7569	0.088	-1.2100	0.013	2.1840	< 0.001
Period = P4	-1.3190	0.002	-0.8866	0.035	2.3760	< 0.001
Reference values: OBJECT and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.2770	0.007	-1.1250	0.029	-3.7130	< 0.001
Period = P1	0.7578	0.075	-0.1823	0.657	-0.9720	0.086
Period = P3	0.0009	0.998	-1.3920	0.002	1.2120	0.003
Period = P4	-0.5613	0.161	-1.0690	0.004	1.4040	< 0.001
Reference values: OBJECT and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.2780	0.010	-2.5170	< 0.001	-2.5010	< 0.001
Period = P1	0.7569	0.088	1.2100	0.013	-2.1840	< 0.001
Period = P2	-0.0009	0.998	1.3920	0.002	-1.2120	0.003
Period = P4	-0.5622	0.181	0.3230	0.484	0.1922	0.599
Reference values: OBJECT and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	0.7156	0.133	-2.1940	< 0.001	-2.3080	< 0.001
Period = P1	1.3190	0.002	0.8866	0.035	-2.3760	< 0.001
Period = P2	0.5613	0.161	1.0690	0.004	-1.4040	< 0.001
Period = P3	0.5622	0.181	-0.3230	0.484	-0.1922	0.599

Table B.21. Period contrasts in the semantic category PLANTS AND FLOWERS

Reference values: PLANTS AND FLOWERS and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	4.9620	< 0.001	-5.0610	< 0.001	-5.5150	< 0.001
Period = P2	-1.2050	< 0.001	1.4490	< 0.001	0.7598	0.006
Period = P3	-1.2970	< 0.001	1.1710	0.011	1.0440	0.001
Period = P4	-1.2380	< 0.001	0.6825	0.211	1.1610	< 0.001
Reference values: PLANTS AND FLOWERS and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.7570	< 0.001	-3.6120	< 0.001	-4.7550	< 0.001
Period = P1	1.2050	< 0.001	-1.4490	< 0.001	-0.7598	0.006
Period = P3	-0.0923	0.671	-0.2777	0.439	0.2840	0.260
Period = P4	-0.0331	0.888	-0.7662	0.097	0.4013	0.119
Reference values: PLANTS AND FLOWERS and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.6640	< 0.001	-3.8900	< 0.001	-4.4710	< 0.001
Period = P1	1.2970	< 0.001	-1.1710	0.011	-1.0440	0.001
Period = P2	0.0923	0.671	0.2777	0.439	-0.2840	0.260
Period = P4	0.0592	0.831	-0.4885	0.359	0.1173	0.696
Reference values: PLANTS AND FLOWERS and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.7240	< 0.001	-4.3790	< 0.001	-4.3540	< 0.001
Period = P1	1.2380	< 0.001	-0.6825	0.211	-1.1610	< 0.001
Period = P2	0.0331	0.888	0.7662	0.097	-0.4013	0.119
Period = P3	-0.0592	0.831	0.4885	0.359	-0.1173	0.696

Table B.22. Period contrasts in the semantic category SUBSTANCE AND MATERIAL

Reference values: SUBSTANCE AND MATERIAL and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.8660	< 0.001	-2.0610	< 0.001	-4.9200	< 0.001
Period = P2	0.4440	0.182	-0.4783	0.199	-0.0899	0.855
Period = P3	0.3350	0.332	-0.8351	0.045	0.4990	0.292
Period = P4	-0.4159	0.247	-0.4292	0.312	1.1600	0.013
Reference values: SUBSTANCE AND MATERIAL and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.3100	< 0.001	-2.5400	< 0.001	-5.0090	< 0.001
Period = P1	-0.4440	0.182	0.4783	0.199	0.0899	0.855
Period = P3	-0.1089	0.731	-0.3568	0.374	0.5888	0.160
Period = P4	-0.8599	0.009	0.0491	0.904	1.2490	0.002
Reference values: SUBSTANCE AND MATERIAL and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.2010	< 0.001	-2.8970	< 0.001	-4.4210	< 0.001
Period = P1	-0.3350	0.332	0.8351	0.045	-0.4990	0.292
Period = P2	0.1089	0.731	0.3568	0.374	-0.5888	0.160
Period = P4	-0.7509	0.029	0.4059	0.367	0.6606	0.091
Reference values: SUBSTANCE AND MATERIAL and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.4500	< 0.001	-2.4910	< 0.001	-3.7600	< 0.001
Period = P1	0.4159	0.247	0.4292	0.312	-1.1600	0.013
Period = P2	0.8599	0.009	-0.0491	0.904	-1.2490	0.002
Period = P3	0.7509	0.029	-0.4059	0.367	-0.6606	0.091

Table B.23. Period contrasts in the semantic category SENSATION

Reference values: SENSATION and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.3140	< 0.001	-2.1830	< 0.001	-7.2450	< 0.001
Period = P2	0.3943	0.310	0.5917	0.154	1.362	0.246
Period = P3	0.4461	0.364	-1.4200	0.032	2.5000	0.029
Period = P4	0.4175	0.439	-0.7352	0.224	1.8860	0.134
Reference values: SENSATION and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.7080	< 0.001	-2.7750	< 0.001	-5.8820	< 0.001
Period = P1	-0.3943	0.310	0.5917	0.154	-1.3620	0.246
Period = P3	0.0518	0.919	-0.8281	0.224	1.1380	0.152
Period = P4	0.0232	0.967	-0.1435	0.819	0.5238	0.582
Reference values: SENSATION and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.7600	< 0.001	-3.6030	< 0.001	-4.7440	< 0.001
Period = P1	-0.4461	0.364	1.4200	0.032	-2.5000	0.029
Period = P2	-0.0518	0.919	0.8281	0.224	-1.1380	0.152
Period = P4	-0.0286	0.964	0.6846	0.398	-0.6145	0.501
Reference values: SENSATION and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	3.7310	< 0.001	-2.9190	< 0.001	-5.3590	< 0.001
Period = P1	-0.4175	0.439	0.7352	0.224	-1.8860	0.134
Period = P2	-0.0232	0.967	0.1435	0.819	-0.5238	0.582
Period = P3	0.0286	0.964	-0.6846	0.398	0.6145	0.501

Table B.24. Period contrasts in the semantic category SPACE

Reference values: SPACE and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.0170	< 0.001	-2.0970	< 0.001	-3.5490	< 0.001
Period = P2	0.2409	0.555	0.3474	0.453	-0.6921	0.156
Period = P3	0.3199	0.435	-0.4721	0.340	0.0164	0.972
Period = P4	-0.0594	0.887	0.0912	0.850	0.0181	0.969
Reference values: SPACE and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.2580	< 0.001	-1.7490	< 0.001	-4.2410	< 0.001
Period = P1	-0.2409	0.555	-0.3474	0.453	0.6921	0.156
Period = P3	0.0789	0.793	-0.8195	0.025	0.7085	0.054
Period = P4	-0.3004	0.335	-0.2563	0.463	0.7102	0.060
Reference values: SPACE and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	2.3370	< 0.001	-2.5690	< 0.001	-3.5320	< 0.001
Period = P1	-0.3199	0.435	0.4721	0.340	-0.0164	0.972
Period = P2	-0.0789	0.793	0.8195	0.025	-0.7085	0.054
Period = P4	-0.3793	0.226	0.5633	0.148	0.0017	0.996
Reference values: SPACE and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	1.9580	< 0.001	-2.0050	< 0.001	-3.5310	< 0.001
Period = P1	0.0593	0.887	-0.0912	0.850	-0.0181	0.969
Period = P2	0.3004	0.335	0.2563	0.463	-0.7102	0.060
Period = P3	0.3793	0.226	-0.5633	0.148	-0.0017	0.996

Table B.25. Period contrasts in the semantic category TEXTILE AND CLOTHING

Reference values: TEXTILE AND CLOTHING and P1						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	0.4743	0.521	-1.055	0.093	-3.3390	< 0.001
Period = P2	0.5223	0.438	0.2935	0.542	-0.5995	0.248
Period = P3	0.1800	0.798	-0.3410	0.496	0.3957	0.442
Period = P4	0.9101	0.206	-0.1147	0.830	-0.5920	0.303
Reference values: TEXTILE AND CLOTHING and P2						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	0.9966	0.034	-0.7612	0.131	-3.939	< 0.001
Period = P1	-0.5223	0.438	-0.2935	0.542	0.5995	0.248
Period = P3	-0.3423	0.405	-0.6345	0.060	0.9951	0.006
Period = P4	0.3878	0.378	(0.4082	0.288	0.0075	0.987
Reference values: TEXTILE AND CLOTHING and P3						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	0.6543	0.200	-1.3960	0.008	-2.9440	< 0.001
Period = P1	-0.1800	0.798	0.3410	0.496	-0.3957	0.442
Period = P2	0.3423	0.405	0.6345	0.060	-0.9951	0.006
Period = P4	0.7301	0.131	0.2263	0.580	-0.9877	0.025
Reference values: TEXTILE AND CLOTHING and P4						
Predictor	***Fragrant***		***Perfumed***		***Scented***	
	Estimate	***p*-value**	**Estimate**	***p*-value**	**Estimate**	***p*-value**
Intercept	-0.1327	0.765	-0.1497	0.731	-2.4160	< 0.001
Period = P1	-0.9101	0.206	0.1147	0.830	0.5920	0.303
Period = P2	-0.3878	0.378	0.4082	0.288	-0.0075	0.987
Period = P3	-0.7301	0.131	-0.2263	0.580	0.9877	0.025

APPENDIX C:
COLLOCATIONAL ANALYSES

Table C.1. Goodness-of-fit statistics comparison of three SVS analyses

SVS analysis	Stress value
Bag-of-words SVS with all content words	0.19
Bag-of-words SVS with only nouns	0.21
Dependency-based SVS with modified nouns	0.30

Table C.2. Most frequent noun collocates of *sweet-scented* in the four periods

COLLOCATE	P1	P2	P3	P4	TOTAL
flower	12	6	1	1	20
grass	4	5	2	1	12
rose	2	3	3	–	8
blossom	1	3	2	–	6
child	–	4	2	–	6
clover	2	3	–	–	5
herb	5	–	–	–	5
meadow	4	–	1	–	5
air	–	4	–	1	5
honeysuckle	2	2	–	–	4

Table C.3. Most frequent noun collocates of *sweet-smelling* in the four periods

COLLOCATE	P1	P2	P3	P4	TOTAL
flower	2	6	4	4	16
incense	1	4	1	1	7
herb	–	4	1	2	7
pine	–	3	3	–	6
shrub	1	1	3	–	5
blossom	–	–	3	2	5
wood	1	1	–	2	5
tree	–	3	–	–	5
grass	–	–	3	1	4
rose	–	2	2	–	4

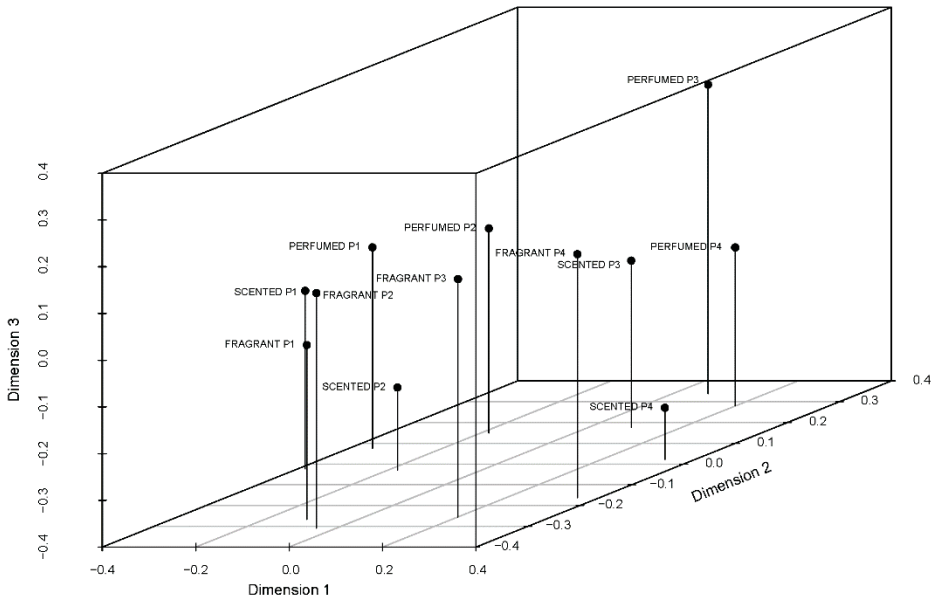


Figure C.1. Three-dimensional MDS map of the collocational preferences of *fragrant*, *perfumed*, and *scented* throughout the period 1810–2009

Table C.4. PMI score of the noun collocates of *fragrant* in the four periods distinguished

COLLOCATE	P1	P2	P3	P4	COLLOCATE	P1	P2	P3	P4
air	3.74	4.3	4.23	4.1	*hay*	6.33	6.05	5.18	0
apple	0	0	0	4.35	*heat*	0	0	0	4.51
aroma	0	0	8.67	8.17	*herb*	7.04	6.06	6.95	0
atmosphere	0	3.78	4.31	0	*hillside*	0	5.37	0	0
balsam	0	7.83	0	0	*honeysuckle*	8.1	7.69	0	0
basket	0	3.73	0	0	*incense*	5.67	6.4	7.23	0
berry	0	4.39	0	0	*jasmine*	0	0	0	7.89
beverage	0	6.41	0	0	*June*	0	3.23	0	0
bloom	5.43	6.38	6.43	5.87	*leaf*	3.88	3.91	3.81	3.67
blossom	6.89	6.68	7.16	6.58	*lily*	0	5.54	5.57	0
blush	3.8	0	0	0	*liquid*	0	0	5.63	0
bough	0	5.88	6.54	0	*load*	0	3.26	0	0
bouquet	0	6.18	6.83	0	*magnolia*	0	7.17	0	0
bower	6.69	6.88	0	0	*meadow*	5.05	5.08	0	0
bowl	0	4.78	0	0	*melody*	0	4.35	0	0
bread	0	0	0	5.69	*memory*	0	3.59	0	0
breath	5.66	4.71	3.72	0	*minute*	0	0	0	3.82
breeze	4.75	5.39	5.49	5.29	*odor*	9.9	9.9	6.58	5.89
bud	4.99	5.46	0	0	*oil*	0	4.47	0	3.94
bunch	0	5.19	0	0	*orchard*	0	5.33	0	0
cedar	0	5.31	0	0	*perfume*	6.14	5.5	7.15	6.07
chicken	0	0	0	3.82	*petal*	7	7.17	0	0
cigar	6.08	3.74	4.57	0	*pile*	0	0	3.39	0
cigarette	0	0	3.5	0	*pine*	0	5.07	5.78	4.43
cloud	0	3.22	3.99	0	*plant*	3.79	0	0	0
clover	0	4.42	6.98	0	*porch*	0	4.14	0	0
cluster	5.12	5.23	0	0	*rose*	5.17	4.67	4.91	4.92
coffee	5.42	5.42	4.67	0	*scent*	0	5.99	6.53	6.38
cup	4.61	4.27	3.78	3.02	*shade*	3.48	3.24	0	0
curl	4.36	0	0	0	*shrub*	6.1	6.28	0	7.15
dew	5.36	6.07	6.85	0	*sight*	3.77	0	0	0
dusk	0	5.08	0	0	*smell*	0	5.28	5.42	4.96
fern	0	5.87	0	0	*smoke*	0	4.54	5.19	3.93
fir	0	6.68	0	0	*spice*	6.66	6.03	0	0
flesh	0	0	0	4.17	*spring*	0	0	3.01	0

Table C.4. Continued

COLLOCATE	P1	P2	P3	P4	COLLOCATE	P1	P2	P3	P4
flower	6.26	5.96	5.87	5.82	*steam*	0	4.6	4.84	5.14
foliage	4.33	5.45	0	0	*summer*	3.25	3.11	3.03	0
forest	0	0	3.3	3.47	*sunshine*	0	3.43	0	0
fruit	3.46	4.21	0	0	*tea*	4.07	4.8	0	5.02
gale	5.4	0	0	0	*tobacco*	4.69	4.07	4.76	0
garden	3.42	4.22	4.82	4.07	*tree*	0	3.28	0	3.8
garland	5.85	0	0	0	*verdure*	0	6.64	0	0
garlic	0	0	0	5.91	*vine*	4.03	4.35	0	0
geranium	0	7.37	0	0	*weed*	0	4.86	0	0
goddess	0	0	0	7.04	*wind*	3.06	0	0	0
grass	4.26	3.94	4.18	3.95	*wine*	3.39	0	0	0
grove	5.14	4.3	0	0	*wood*	3.02	3.33	0	0
gum	0	6.54	0	0	*wreath*	6.09	5.37	0	0
hair	0	0	3.17	3.11					

Table C.5. PMI score of the noun collocates of *perfumed* in the four periods distinguished

COLLOCATE	P1	P2	P3	P4	COLLOCATE	P1	P2	P3	P4
air	4.6	5.08	3.88	3.88	*oil*	7.36	5.91	0	4.72
bath	0	0	6.68	6.68	*pocket*	0	4.46	0	0
blossom	0	6.22	0	0	*powder*	0	0	6.6	0
body	0	0	0	0	*rose*	0	3.82	0	0
breath	5.9	5.26	4.99	4.99	*silk*	0	5.45	5.99	0
flavor	11.63	0	0	0	*smell*	0	0	0	5.35
flower	4.85	4.86	0	0	*smoke*	0	4.79	0	0
garden	0	4.03	0	0	*soap*	0	0	7.45	7.96
garment	0	0	6.86	6.86	*spring*	0	3.45	0	0
hair	0	3.96	4.11	4.11	*tree*	3.41	0	0	0
handkerchief	0	7.14	7.23	7.23	*water*	3.56	3.31	3.43	3.24
lace	0	6.52	0	0	*wind*	0	0	3.86	0
note	4.44	3.54	0	0					

Table C.6. PMI score of the noun collocates of *scented* in the four periods distinguished

COLLOCATE	P1	P2	P3	P4	COLLOCATE	P1	P2	P3	P4
air	4.57	4.78	4.19	4.81	*hair*	0	0	0	3.51
bath	0	0	0	6.76	*handkerchief*	0	0	6.11	0
bloom	0	6.26	0	0	*oil*	0	0	4.37	5.34
blossom	0	6.41	0	0	*perfume*	0	0	0	7.47
breath	0	4.68	4.55	0	*powder*	0	0	6.43	0
breeze	0	6.23	0	0	*sheet*	0	0	0	5.22
cake	0	0	6.04	0	*smell*	0	0	0	4.29
candle	0	0	0	8.54	*smoke*	0	0	0	4.48
envelope	0	0	6.41	0	*soap*	0	9.28	8.6	8.67
flower	5.45	5.5	4.75	5.47	*summer*	0	4.62	0	0
garden	0	0	4.66	0	*tree*	0	0	3.25	0
geranium	0	0	0	11.43	*water*	0	3.27	0	3.86
grass	0	6.64	0	0	*wind*	4.62	0	0	0

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

RELACIÓN DE PUBLICACIÓNS DA DOUTORANDA MENCIONADAS NA TESE DE DOUTORAMENTO

Relación de publicacións	Pettersson-Traba, Daniela. 2018. A diachronic perspective on near-synonymy: The concept of SWEET-SMELLING in American English. *Corpus Linguistics and Linguistic Theory*. 1–31. Pettersson-Traba, Daniela. 2019. Measuring semantic distance across time: An analysis of the collocational profiles of a set of near synonyms in American English. *Journal of Research Design and Statistics in Linguistics and Communication Science* 6(2). 1–28. Pettersson-Traba, Daniela. 2020a. Modeling the choice of near-synonyms in the recent history of American English: The case of *fragrant*, *perfumed* and *scented*. En Pilar Guerrero-Medina, Macarena Palma Gutiérrez e María Valero Redondo (eds.), *Advances in English and American Studies: Current developments, future trends*, 149–162. Córdoba: UCOPress / Editorial Universidad de Córdoba. Pettersson-Traba, Daniela. 2020b. Lexical variation in the history of American English: Attraction and differentiation in a synonym set. En Lourdes López Ropero, Sara Prieto García-Cañedo e José Antonio Sánchez Fajardo (eds.), *Thresholds and ways forward in English studies*, 235–248. Alicante: Publicaciones de la Universidad de Alicante. Pettersson-Traba, Daniela. En prensa. Distributional changes in synonym sets: The case of *fragrant*, *scented*, and *perfumed* in 19th- and 20th-century American English. En *Historical English in contact* (Studies in Language Companion Series). Amsterdam: John Benjamins.
Título	A diachronic perspective on near-synonymy: The concept of SWEET-SMELLING in American English
Ano	2018
Revista ou Libro	Corpus Linguistics and Linguistic Theory
Volume e páxinas	Publicado en liña; 1–31 (DOI: 10.1515/cllt-2018-0025)

Contribución da doutoranda	Autora única da publicación
Índice/s de calidade	Indexada en WoS, Scopus, ERIH Plus Índices de impacto: - JCR: 0.960 (Q2; 86/184 na categoría Linguistics) - SJR: 0.388 (Q1; 154/844 na categoría Language and Linguistics) - CIRC: Categoría Excelencia A+ en Ciencias Humanas - CARHUS Plus+ 2018: Categoría A (ámbito: Filología, Lingüística e Sociolingüística) - MIAR ICDS 2018: 10.6 sobre 11 Citas: 1 (Google Académico)
Título	Measuring semantic distance across time: An analysis of the collocational profiles of a set of near synonyms in American English
Ano	2019
Revista ou Libro	*Journal of Research Design and Statistics in Linguistics and Communication Science*
Volume e páxinas	6 (2); 1–28 (https://doi.org/10.1558/jrds.40245)
Contribución da doutoranda	Autora única da publicación
Índice/s de calidade	Indexada en Linguistics Abstracts Online, ERIH Plus
Título	Modeling the choice of near-synonyms in the recent history of American English: The case of *fragrant*, *perfumed* and *scented*
Ano	2020
Revista ou Libro	*Advances in English and American Studies: Current developments, future trends*
Volume e páxinas	149–162
Contribución da doutoranda	Autora única da publicación
Índice/s de calidade	SPI (2018): posición 78 de 104 en editoriais españolas no ranking xeral

Título	Lexical variation in the history of American English: Attraction and differentiation in a synonym set
Ano	2020
Revista ou Libro	*Thresholds and ways forward in English studies*
Volume e páxinas	235–248
Contribución da doutoranda	Autora única da publicación
Índice/s de calidade	SPI (2018): posición 26 de 55 en editoriais españolas no ranking da disciplina "Lingüística, Literatura y Filología"
Título	Distributional changes in synonym sets: The case of *fragrant*, *scented*, and *perfumed* in 19th- and 20th-century American English
Ano	En prensa
Revista ou Libro	*Historical English in contact*
Volume e páxinas	28 páxinas
Contribución da doutoranda	Autora única da publicación
Índice/s de calidade	SPI (2018): posición 5 de 48 en editoriais estranxeiras no ranking da disciplina "Lingüística, Literatura y Filología"

Journal of Research Design and Statistics in Linguistics and Communication Science
ISSN 2052-417X (print)
ISSN 2052-4188 (online)
equinox

7 December 2020

AUTHORIZATION LETTER

Daniela Pettersson-Traba, author of 'Measuring semantic distance across time: An analysis of the collocational profiles of a set of near-synonyms in American English,' published in *Journal of Research Design and Statistics in Linguistics and Communication Science* (ISSN 2052-417X and e-ISSN 2052-4188), is allowed to use all or part of this article in future works in her name as long as all bibliographic details of the original publication and Equinox Publishing Ltd. as the original publisher and copyright holder are cited.



Prof. Pascual Cantos-Gómez
JRDS co-editor

Editors
Pascual Cantos Gómez, University of Murcia. Email: pcantos@um.es

Stefan Th. Gries, University of California, Santa Barbara. Email: stgries@linguistics.ucsb.edu

Please send books for review to:
Pascual Cantos Gómez, Dpto. Filología Inglesa, Universidad de Murcia, Facultad de Letras, Campus de la Merced, 30071 Murcia, Spain

For more information about *Journal of Research Design and Statistics in Linguistics and Communication Science*, visit **www.equinoxpub.com/JRDS**

UNIVERSIDAD D CÓRDOBA

UCOPress
Editorial Universidad de Córdoba

El abajo firmante, Juan Pedro Monferrer Sala, en calidad de Director de *UCOPress Editorial Universidad de Córdoba* autoriza a Dña. Daniela Pettersson a que pueda hacer uso en su tesis doctoral de los datos obtenidos en su artículo "Modeling the choice of near-synonyms in the recent history of American English: The case of fragrant, perfumed and scented", incluido en el libro *Advances in English and American Studies current developments, future trends* (Córdoba: UCOPress, 2020; ISBN: 978-84-9927-546-8), rogándole haga constar debidamente la referencia del libro que acoge la publicación.

Lo que firmo a petición de la interesada, a los efectos oportunos.

En Córdoba, a nueve de diciembre de dos mil veinte.

MONFERRER SALA JUAN PEDRO - 19095870M

Juan Pedro Monferrer Sala

Campus de Rabanales. Ctra. Nacional IV, Km. 396 • Teléfono: 957 21 8126 • E-mail: ucopress@uco.es • 14071 Córdoba (ESPAÑA)

UA | UNIVERSITAT D'ALACANT
UNIVERSIDAD DE ALICANTE
Departament de Filologia Anglesa
Departamento de Filología Inglesa

December 7, 2020

Daniela Pettersson-Traba, author of 'Lexical Variation in the History of American English: Attraction and Differentiation in a Synonym Set,' published in the ebook entitled *Thresholds and Ways Forward in English Studies* (ISBN 978-84-1302-079-2), is allowed to use all or part of this article in future works in her name as long as all bibliographic details of the original publication and Publicaciones de la Universidad de Alicante as the original publisher and copyright holder are cited.

MARIA LOURDES| LOPEZ|ROPERO

Firmado digitalmente por MARIA LOURDES|LOPEZ|ROPERO
Fecha: 2020.12.07 09:51:39 +01'00'

Prof. Lourdes López Ropero
Thresholds and Ways Forward in English Studies co-editor

RESUMO EN GALEGO

Que significa exactamente que dúas palabras sexan sinónimas? Esta é unha pregunta que esperta un grande interese tanto entre lingüistas como entre falantes das linguas. Ao longo dos anos propuxéronse respostas lixeiramente diferentes a esta cuestión na bibliografía especializada. A pesar de que esta relación semántica sempre se caracteriza en termos de semellanza de significado, existen dous grupos diferentes de definicións da palabra *sinónimo*. Mentres que no primeiro grupo o termo se define como unha relación de identidade semántica, no segundo considerase que un grao parcial de semellanza é suficiente para que unha palabra ou expresión constitúa un sinónimo doutra. A sinonimia absoluta, é dicir, a identidade completa de significado, é un fenómeno extremadamente raro ou incluso inexistente na linguaxe, mentres que a case sinonimia é extremadamente común (véxase, por exemplo, Cruse 2000: 157–158; Taylor 2003: 264; Divjak e Gries 2006: 24; Liu 2010: 56–57). Isto pódese comprobar facilmente considerando por un momento o uso de calquera par ou serie de sinónimos nunha lingua determinada. Independentemente dos sinónimos que se seleccione, o escenario máis probable é que existan algúns contextos de uso onde un dos termos sexa a única opción posible ou, polo menos, a opción máis adecuada.

Este sinxelo experimento mostra que a gran maioría dos sinónimos, con toda probabilidade, difiren dun ou doutro xeito, o que fai que non sexan completamente intercambiables en todos os contextos. Polo tanto, como xa argumentou o filósofo Thomas Reid no século XVIII, "[m]ost synonimes [sic] have some minute distinction that deserves notice" (1785: 14, citado no *Oxford English Dictionary*). Así pois, talvez a pregunta máis interesante á hora de analizar o uso de sinónimos non é que os fai semellantes senón, pola contra, onde residen exactamente as súas diferenzas. Esta é unha cuestión que non só fascina os lingüistas, senón tamén os falantes das linguas, dado que, como demostra Murphy (2013: Sección 3), uns dos usos máis frecuentes dos dicionarios é precisamente o de buscar sinónimos, especialmente na produción da linguaxe para evitar a repetición, conseguir unha escritura máis sofisticada ou axustar o estilo empregado, entre outros. De feito, seleccionar a palabra máis axeitada entre un conxunto de expresións

sinónimas é unha tarefa notoriamente difícil, tanto para os usuarios non nativos como para os nativos. Isto débese a que os sinónimos poden variar nun gran número de dimensións: ata 35 diferentes, segundo Edmonds (1999: Capítulo 3). Ademais, as diferenzas son a miúdo de natureza multidimensional (véxase, por exemplo, Cruse 1986: 284–289; 2000: 159–160; Murphy 2003: 154–157). Tendo todo isto en conta, non resulta estraño que as investigacións sobre sinonimia cambiasen recentemente de enfoque, deixando atrás unha excesiva énfase na clasificación de sinónimos en tipos e subtipos, e centrándose, no seu lugar, na identificación dos factores que exercen unha influencia na selección de sinónimos. Para este fin, nas últimas décadas empregáronse métodos e técnicas innovadoras en estudos semánticos en correntes neoestruturalistas e cognitivas para identificar con precisión as dimensións nas cales determinados pares ou grupos de sinónimos varían (véxase, por exemplo, Persson 1989; Arppe 2002; 2008; Taylor 2003; Divjak 2006; 2010; Divjak e Gries 2006; 2008; Gries e Otani 2010; Liu 2010; 2013; Liu e Espino 2012; Desagulier 2014; Krawczak 2014; 2018). Estes estudos demostran a idoneidade destes enfoques para identificar diferenzas entre sinónimos mediante o uso de datos de corpus e análises pormenorizadas dos contextos nos que se empregan este tipo de palabras, incluíndo contextos semánticos, léxicos, morfosintácticos e estilísticos, entre outros.

Malia a utilidade de tales enfoques para o estudo da sinonimia, a maioría das investigacións deste tipo adoptan unha perspectiva sincrónica, examinando o conxunto de factores que poderían influír de xeito significativo na alternancia entre sinónimos nun período determinado, nomeadamente no presente. Pola contra, poucos estudos analizan como cambia gradualmente a relación entre os membros dunha serie de sinónimos co paso do tempo, adoptando así unha perspectiva diacrónica. De feito, só se levaron a cabo un número moi limitado de investigacións deste tipo nos últimos anos, entre as que destacan Kaunisto (2001), Primahadi-Vijaya-R e Rajeg (2014) e Baker (2017: 95–101). Igualmente, as motivacións subxacentes aos cambios na estrutura semántica interna de sinónimos foron reducidas tradicionalmente á chamada *no synonymy rule* (véxase, por exemplo, Samuels 1972: 65; Bolinger 1977: ix–x, 9; Wierzbicka 1988: 13–14; Croft 2000: 176; Nuyts e Byloo 2015: 62–63), que define o comportamento das linguas como contrario á sinonimia absoluta. En concreto, esta teoría propón que as palabras sinónimas sempre se van diferenciando co paso do tempo ou, de non ser así, un ou varios dos termos deixan de ser empregados para deste xeito asegurar que as linguas continúen sendo sistemas de comunicación económicos e eficientes. Non obstante, recentemente esta teoría de

competición entre sinónimos é vista como insuficiente para explicar todos os cambios experimentados por este tipo de expresións, xa que existen outros procesos diacrónicos que tamén poderían xogar un papel nos devanditos cambios (De Smet et al. 2018). Alén dos procesos de diferenciación e substitución, os sinónimos tamén poden converxer en termos funcionais co paso do tempo, un proceso coñecido como atracción. Porén, este proceso non ten recibido tanta atención coma os fenómenos de competición.

Neste contexto, esta tese de doutoramento pretende contribuír ao coñecemento sobre que factores inflúen na selección léxica en inglés, como cambian os seus efectos co paso do tempo e cales son as principais motivacións deste tipo de mudanzas. Isto lévase a cabo mediante un estudo de corpus dunha serie de adxectivos case sinónimos do dominio semántico do olfacto no inglés americano dos séculos XIX e XX, en concreto, *fragrant*, *perfumed*, *scented*, *sweet-scented* e *sweet-smelling*, que designan o concepto PLEASANT SMELLING. Os exemplos (1)–(5) demostran o uso destes cinco adxectivos:

(1) They made campfires with top-quality hardwood, which released **fragrant** smoke, to cook for their families. (*COHA*, 2009, FIC, FarawayWarNovel)
(2) Dellius had seen enough easterners to know that the curls were not dirty greasy; easterners pomaded their locks with **perfumed** creams. (*COHA*, 2007, FIC, ColorsInsulting)
(3) Avoid chemical irritants in bubble baths and heavily **scented** soaps, and don't let her sit in a wet bathing suit for too long. (*COHA*, 2009, MAG, Coca)
(4) As his eyes adjusted ever so slowly to the gloom, he saw looming shadows, blurred shapes like enormous trees that stirred not at all in the gentle **sweet-scented** wind. (*COHA*, 2005, FIC, FantasySciFi)
(5) Route 214 turned out to be a winding lane through **sweet-smelling** fields long abandoned to goldenrod and blackberry brambles. (*COHA*, 2008, FIC, BookOldHouses)

A decisión de analizar esta serie de sinónimos susténtase en dous motivos principais. En primeiro lugar, o *Historical Thesaurus of the Oxford English Dictionary* (*HTOED*) enumera unha gran cantidade de palabras dentro do dominio semántico do olfacto, moitas das cales son precisamente adxectivos. Moitos destes son definidos de xeito similar nos dicionarios, incluíndo soamente algunhas pequenas variacións. Polo tanto, este parece ser un campo

semántico especialmente interesante para o estudo da case sinonimia. En segundo lugar, o relativamente baixo grao de polisemia destes adxectivos case sinónimos constitúe unha vantaxe para un estudo centrado nesta relación semántica. Se o grupo de sinónimos seleccionado fose altamente polisémico, o resultado sería un estudo demasiado ecléctico para unha tese doutoral, xa que non estaría enfocado soamente na relación semántica da sinonimia.

Os datos empregados para as análises incluídas na tese foron extraídos do *Corpus of Historical American English* (*COHA*; Davies 2010–), que abrangue o período comprendido entre 1810 e 2009. A atención céntrase especialmente nos patróns distribucionais dos adxectivos case sinónimos, que serven para cuantificar o seu grao de diferenza funcional. Por iso, seguindo enfoques recentes aplicados ao estudo sincrónico da sinonimia, esta relación semántica mídese na tese en termos cuantitativos por medio da análise pormenorizada dunha serie de factores que desempeñan un papel importante na selección léxica. Os factores analizados representan contextos de uso semánticos e non semánticos, así como variables de natureza extralingüística. En particular, a influencia destes factores examínase ao longo do período de 200 anos cuberto polo corpus, dividido aquí en catro subperíodos de 50 anos: 1810–1859, 1860–1909, 1910–1959 e 1960–2009. O principal obxectivo é o de descubrir como evoluciona co paso do tempo cada elemento léxico, así como a relación existente entre eles. Con este propósito, os datos foron analizados mediante técnicas estatísticas univariantes e multivariantes que xa foron empregadas en estudos previos para medir as semellanzas e as diferenzas entre case sinónimos dende unha perspectiva sincrónica. Polo tanto, un obxectivo secundario da tese é o de avaliar a utilidade destes métodos cuantitativos, principalmente sincrónicos, na investigación diacrónica sobre sinonimia.

Dende unha perspectiva teórica, esta tese enmárcase nos paradigmas da semántica neoestruturalista e cognitiva, que destacan a vital importancia do uso na estrutura da linguaxe. Ademais, o estudo está baseado en teorías de cambio semántico, xa que afonda nos distintos procesos subxacentes no desenvolvemento da estrutura semántica interna de grupos de sinónimos, particularmente os de diferenciación e substitución, pero tamén o proceso de atracción, e a súa conexión coa realidade extralingüística. Xa que logo, outro obxectivo secundario desta tese é o de determinar se a teoría tradicional de competición entre sinónimos é suficiente para dar conta dos cambios dos adxectivos obxecto de estudo ou se, pola contra, é preciso recorrer tamén a outros procesos de cambio semántico como o de atracción.

A tese está dividida en oito capítulos, organizados do seguinte xeito. Despois do capítulo introdutorio, os capítulos 2 e 3 conteñen unha revisión da bibliografía relevante sobre nocións e enfoques no campo da semántica léxica, en xeral, e da sinonimia, en particular. O capítulo 4 explica os datos empregados nas análises realizadas nos seguintes capítulos, centrándose na descrición dos case sinónimos obxecto de estudo e do corpus empregado, así como dos procesos de extracción e anotación de datos. Os capítulos 5–7 presentan os resultados de tres estudos baseados en corpus sobre o concepto PLEASANT SMELLING e os adxectivos que o designan. Finalmente, o capítulo 8 ofrece as conclusións extraídas dos resultados das diferentes análises realizadas, así como algunhas suxestións para futuras investigacións.

O propósito do capítulo 2 é o de proporcionar unha visión xeral das escolas lingüísticas máis importantes na investigación sobre semántica léxica nos últimos 200 anos, centrándose especialmente na orixe e desenvolvemento das ideas e nocións subxacentes no marco teórico dos estudos realizados. A sección 2.1 resume os principios máis importantes da semántica histórico-filolóxica, o marco que dominou o campo dende a década de 1830 ata a década de 1930. Unha gran parte desta sección está dedicada a explicar os principios centrais desta escola, especialmente o que concirne á súa perspectiva diacrónica e ás clasificacións de cambios semánticos, as cales foron adaptadas por escolas posteriores e que aínda hoxe son fundamentais en calquera estudo diacrónico neste ámbito. A sección 2.2 céntrase nas escolas estruturalistas e xenerativistas, prestando especial atención ao estudo das relacións semánticas dentro da tradición estruturalista, que eran consideradas como a base para a delimitación de significados no vocabulario. A sección 2.3 trata sobre os movementos neoestruturalistas, en particular sobre o método distribucional baseado en corpus iniciado por Firth (1957) e Sinclair (1966) nos anos cincuenta e sesenta. Este enfoque e a súa énfase no comportamento colocacional das palabras forman a base dos supostos metodolóxicos e teóricos desta tese, así como da maioría de traballos recentes sobre semántica léxica. Finalmente, a sección 2.4 céntrase na semántica cognitiva, un enfoque maximalista do estudo da linguaxe, en xeral, e do significado, en particular. Dúas das contribucións máis importantes desta escola destacan nesta sección, nomeadamente a teoría de prototipos e as teorías sobre cambio semántico baseadas no uso, dado que estas teorías se empregan á hora de interpretar os resultados das análises realizadas nos capítulos 5–7.

O capítulo 3 afonda na relación semántica da sinonimia. En primeiro lugar, as clasificacións de sinónimos máis relevantes trátanse na sección 3.1. As conclusións extraídas

destas clasificacións son que a distinción entre sinonimia absoluta e case sinonimia é a máis adecuada para investigacións empíricas sobre este fenómeno semántico. A continuación, a sección 3.2 ofrece unha revisión en profundidade dos estudos sincrónicos de corpus sobre pares ou grupos específicos de case sinónimos que se realizaron nos últimos 30 anos. Na sección 3.2.1, identifícanse e explícanse dúas correntes de investigación que suxiren que recentemente se produciu unha evolución considerable nesta área de investigación. A sección 3.2.2 céntrase exclusivamente en estudos sobre adxectivos case sinónimos, por seren estes o obxecto de análise principal da presente tese. Finalmente, a sección 3.3 ofrece unha visión xeral das investigacións diacrónicas sobre sinonimia léxica, considerando primeiro os poucos estudos de corpus existentes sobre o desenvolvemento histórico de pares ou grupos particulares de case sinónimos. A sección remata cun resumo dos procesos xerais de cambio semántico relevantes para o estudo diacrónico da sinonimia.

O capítulo 4 comeza cunha descrición, na sección 4.1, do grupo de sinónimos obxecto de estudo, das principais motivacións que levan a seleccionar esta serie específica de case sinónimos e da información sobre estes elementos léxicos existente en dicionarios e *thesaurus* do inglés actual e histórico. Seguidamente, a sección 4.2 describe o corpus do que se extraeron os datos, así como o proceso de obtención dos mesmos. Por último, a sección 4.3 céntrase no proceso de anotación dos datos. Gran parte desta sección está dedicada á explicación das variables internas e externas e os seus niveis, que se resumen a continuación:

- Sentido do uso dos adxectivos: natural, artificial, figurado e indeterminado.
- Categoría semántica dos substantivos modificados polos adxectivos: (i) ABSTRACT, (ii) BODY AND PEOPLE, (iii) CLEANING, (iv) COSMETICS, (v) EARTH, ATMOSPHERE, AND WEATHER, (vi) FOOD AND DRINK, (vii), OBJECT, (viii) PLANTS AND FLOWERS, (ix) SENSATION, (x) SPACE, (xi) SUBSTANCE AND MATERIAL e (xii) TEXTILE AND CLOTHING.
- Carácter animado ou inanimado dos referentes dos substantivos modificados polos adxectivos.
- Carácter contable ou non contable dos referentes dos substantivos modificados polos adxectivos.
- Concreción dos referentes dos substantivos modificados polos adxectivos: concreto e abstracto.

- Función sintáctica dos adxectivos: atributivo, complemento predicativo, pospositivo e outras funcións menores.
- Grao: positivo, comparativo e superlativo.
- Colocación substantiva: cada un dos substantivos modificados polos adxectivos.
- Período: período 1 (1810–1859), período 2 (1860–1909), período 3 (1910–1959) e período 4 (1960–2009).
- Tipo de texto: ficción, non ficción e publicacións periódicas (xornais e revistas).

O capítulo 5 presenta o primeiro estudo empírico da tese, que constitúe unha análise descritiva e univariante dos datos. O capítulo está dividido en catro seccións, cada unha centrada nun aspecto diferente da serie de case sinónimos. A sección 5.1 ofrece a frecuencia xeral dos cinco adxectivos, así como a súa evolución no período 1810–2009. Os resultados apuntan á existencia dunha gran diferenza entre os termos no que concirne á súa frecuencia de uso, con *fragrant* como elemento claramente dominante, xa que representa máis da metade da totalidade dos datos (58,90%), seguido, en orde decrecente de frecuencia, por *perfumed* (16,95%), *scented* (16,33%), *sweet-smelling* (4,39%) e *sweet-scented* (3,44 %). Tamén se atopan cambios interesantes en termos da frecuencia dos case sinónimos, xa que mentres que *fragrant* e *sweet-scented* diminúen significativamente ao longo do tempo, *scented* e *sweet-smelling* aumentan. Á súa vez, *perfumed* permanece relativamente estable ao longo do período analizado. Os devanditos cambios de frecuencia dan lugar a unha reorganización dos case sinónimos: se ben *fragrant* é o adxectivo máis frecuente en todos os períodos, *scented* despraza a *perfumed* e *sweet-smelling* a *sweet-scented*.

A sección 5.2 versa sobre o concepto PLEASANT SMELLING, tendo en conta os cinco adxectivos de forma conxunta e a súa distribución en sentidos de significado e categorías semánticas de substantivos modificados. O uso do concepto presenta tendencias diacrónicas: aínda que, en xeral, se emprega principalmente para cualificar aromas naturais —e as categorías semánticas de substantivos asociadas a eles (por exemplo, PLANTS AND FLOWERS e EARTH, ATMOSPHERE, AND WEATHER)—, obsérvase un cambio en favor de aromas artificiais, particularmente a finais do século XIX e principios do XX. Nesta sección proponse unha hipótese para dar conta deste desenvolvemento diacrónico con base en forzas extralingüísticas, concretamente as transformacións socioculturais e tecnolóxicas que experimentou a sociedade estadounidense tras a primeira e a segunda revolución industrial. Arguméntase aquí que a

crecente necesidade de referirse a aromas artificiais —no detrimento de aromas naturais— deriva da crecente dispoñibilidade de produtos que emiten olores sintéticos agradables, así como da súa presenza na vida cotiá do pobo norteamericano. Esta hipótese vese apoiada por un estudo preliminar de indicadores léxicos do ámbito da modernización e da industrialización.

A sección 5.3 analiza o comportamento de cada adxectivo por separado, ofrecendo así unha perspectiva semasiolóxica dos datos. A distribución de cada un deles nos diferentes sentidos de significado e categorías semánticas de substantivos modificados confirman a maior parte da información dispoñible nos dicionarios e *thesaurus*. Non obstante, cómpre matizarmos os patróns de uso dos case sinónimos, xa que son máis complexos e específicos do que se reflicte nos dicionarios. Respecto dos patróns diacrónicos en termos semasiolóxicos, existen tendencias similares para todos os adxectivos agás *sweet-scented*, dado que a súa frecuencia para referirse a aromas artificiais aumenta co tempo. A importancia deste achado reside en que, independentemente do uso prototípico de cada adxectivo, ben sexa 'natural' ou 'artificial', a maioría dos case sinónimos acaban sendo empregados máis a miúdo en contextos 'artificiais' ao final do período analizado, o que suxire a existencia dalgunha motivación subxacente que guía este cambio semántico.

Finalmente, a sección 5.4 trata sobre a competición entre os case sinónimos en diferentes contextos semánticos, adoptando deste xeito unha orientación onomasiolóxica. Nesta sección demóstrase que, a pesar de que *fragrant* ten unha frecuencia moito máis alta que o resto de case sinónimos, este adxectivo non é o máis común para referirse a aromas artificiais, onde se ve superado por *scented* e/ou *perfumed*. Ademais, identifícase un cambio no que se refire ás proporcións de uso dos case sinónimos nalgúns sentidos de significado e nalgunhas categorías semánticas. En primeiro lugar, *scented* despraza a *perfumed* e, ata certo punto, tamén a *fragrant*, en varias das categorías semánticas 'artificiais' e, segundo, nos sentidos de significado figurativos e indeterminados *fragant* perde o seu dominio inicial segundo pasa o tempo, o que resulta nunha distribución máis equilibrada dos adxectivos ao final do período analizado.

O capítulo 6 afonda na estrutura onomasiolóxica da serie de case sinónimos, medindo a competición entre *fragrant*, *perfumed* e *scented* en relación a un gran número de factores intralingüísticos e extralingüísticos. Neste caso, a perspectiva adoptada é multivariante e, polo tanto, empréganse unha serie de técnicas estatísticas máis sofisticadas, as cales se explican na sección 6.1. A sección 6.2, á súa vez, describe os resultados obtidos neste estudo, centrándose primeiro nos efectos das variables incluídas na análise na sección 6.2.1 e ofrecendo unha

investigación preliminar sobre o comportamento colocacional dos adxectivos na sección 6.2.2. Os resultados confirman a maioría das tendencias identificadas no capítulo 5 sobre a competición entre os adxectivos nos diferentes sentidos de significado e categorías semánticas, mais tamén evidencian outros patróns de uso. As restantes variables lingüísticas semánticas e non semánticas tamén xogan un papel importante na selección entre os case sinónimos, mais estes non amosan tendencias diacrónicas en ningún destes contextos de uso. A variable extralingüística 'Tipo de texto' tamén afecta á alternancia entre os adxectivos e, de feito, existen cambios significativos ao longo do tempo con respecto a esta variable. A pesar de que todas as variables incluídas nas análises presentadas no capítulo 6 inflúen de xeito significativo na selección dos adxectivos, unhas teñen un impacto máis forte ca outras. En particular, as variables semánticas superan a todas as non semánticas, agás 'Período', tendo esta última unha importancia crucial neste estudo. A variable 'Colocación' mellora significativamente a calidade dos resultados, o que indica, por tanto, a gran importancia das preferencias colocacionais e idiosincráticas na selección léxica.

Dado que as colocacións individuais dos adxectivos exercen unha grande influencia na alternancia entre eles, o capítulo 7 examina con maior detalle as preferencias colocacionais dos case sinónimos *fragrant*, *perfumed* e *scented*, empregando métodos especificamente orientados a este tipo de análises. O capítulo ábrese cunha descrición da base de datos e dos métodos empregados nas análises colocacionais na sección 7.1, mentres que a sección 7.2 describe os resultados destas análises. En primeiro lugar, a sección 7.2.1 ofrece unha visión xeral do comportamento colocacional dos adxectivos case sinónimos. Os resultados confirman de novo moitas das tendencias identificadas anteriormente con base en preferencias semánticas máis esquemáticas (é dicir, sentidos de significado e categorías semánticas): *fragrant* co-ocorre con máis frecuencia con substantivos situados no extremo natural do espectro semántico, mentres que para *perfumed* se observa a tendencia contraria. Á súa vez, *scented* ocupa unha posición intermedia, dado que non está asociado a ningún dos dous sentidos de significado. Os resultados da sección 7.2.1 tamén apuntan a conclusións semellantes ás alcanzadas no capítulo 5 sobre a influencia de forzas extralingüísticas na evolución diacrónica dos tres adxectivos case sinónimos examinados, xa que todos eles co-ocorren con máis frecuencia con substantivos en categorías semánticas 'naturais' no século XIX que no século XX. En segundo lugar, na sección 7.2.2 preséntase unha análise máis cualitativa dos substantivos máis importantes que aparecen cos tres adxectivos. As tendencias recoñecidas nos capítulos anteriores reciben, unha vez máis,

apoio dos patróns identificados mediante esta análise cualitativa máis detallada. Non obstante, o achado máis importante da sección 7.2.2 é que, independentemente da existencia de tendencias diacrónicas de carácter máis abstracto compartidas polos adxectivos, todos presentan comportamentos colocacionais considerablemente idiosincráticos cando se examinan as súas colocacións individuais máis importantes.

A tese conclúe co capítulo 8, no que se resumen os resultados máis relevantes que se poden extraer dos estudos realizados, así como algunhas áreas que poden ser melloradas en futuras investigacións sobre o tema.

En conclusión, a contribución principal desta tese de doutoramento é a de engadir a dimensión diacrónica ao estudo cuantitativo da relación de case sinonimia léxica, xa que constitúe unha das primeiras aplicacións de métodos principalmente sincrónicos á investigación histórica desta relación semántica. Ademais, a tese ten en conta un número máis amplo de case sinónimos dos que normalmente se examinan na bibliografía especializada. Ao mesmo tempo, estes sinónimos son analizados nunha variedade de contextos de uso máis numerosa que na maioría de estudos anteriores. Por iso, a pesar da relativamente longa tradición investigadora no campo da sinonimia, as análises desta tese proporcionan un enfoque innovador ao seu estudo, así como achados novos e interesantes sobre o desenvolvemento diacrónico de palabras relacionadas semanticamente.

OBRAS CITADAS:

Arppe, Antti. 2002. The usage patterns and selectional preferences of synonyms in a morphologically rich language. En Annie Morin e Pascale Sébillot (eds.), *JADT-2002. 6th international conference on textual data statistical analysis*, vol. 1, 21–32. Rennes: INRIA.

Arppe, Antti. 2008. *Univariate, bivariate, and multivariate methods in corpus-based lexicography: A study of synonymy*. Helsinki: Tese de doutoramento da Universidade de Helsinki.

Baker, Paul. 2017. *American and British English: Divided by a common language?* Cambridge: Cambridge University Press.

Bolinger, Dwight. 1977. *Meaning and form*. Londres: Longman.

Croft, William. 2000. *Explaining language change. An evolutionary approach.* Essex: Pearson Education.

Cruse, D. Alan. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Cruse, D. Alan. 2000. *Meaning in language: An introduction to semantics and pragmatics*. Oxford e Nova York: Oxford University Press.

Davies, Mark. 2010–. *The Corpus of Historical American English (COHA): 400 Million Words, 1810–2009*. Dispoñible en: https://corpus.byu.edu/coha/.

De Smet, Henrik, Frauke D'hoedt, Lauren Fonteyn e Kristel Van Goethem. 2018. The changing functions of competing forms: Attraction and differentiation. *Cognitive Linguistics* 29(2). 197–234.

Desagulier, Guillaume. 2014. Visualizing distances in a set of near-synonyms: *rather*, *quite*, *fairly*, and *pretty*. En Dylan Glynn e Justyna A. Robinson (eds.), *Corpus methods for semantics. Quantitative studies in polysemy and synonymy*, 145–178. Ámsterdam e Filadelfia: John Benjamins.

Divjak, Dagmar. 2006. Ways of intending: Delineating and structuring near synonyms. En Stefan Th. Gries e Anatol Stefanowitsch (eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*, 19–56. Berlín e Nova York: Mouton de Gruyter.

Divjak, Dagmar. 2010. *Structuring the lexicon: A clustered model for near-synonymy*. Berlín e Nova York: Mouton de Gruyter.

Divjak, Dagmar e Stefan Th. Gries. 2006. Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2(1). 23–60.

Divjak, Dagmar e Stefan Th. Gries. 2008. Clusters in the mind? Converging evidence from near synonymy in Russian. *The Mental Lexicon* 3(2). 188–213.

Edmonds, Philip. 1999. *Semantic representations of near-synonyms for automatic lexical choice*. Toronto: Tese de doutoramento da Universidade de Toronto.

Firth, John R. 1957. A synopsis of linguistic theory 1930-1955. En John R. Firth (ed.), *Studies in linguistic analysis*, 1–32. Oxford: Philological Society.

Gries, Stefan Th. e Naoki Otani. 2010. Behavioral profiles: Corpus-based perspective on synonymy and antonymy. *ICAME Journal* 34. 121–150.

Kaunisto, Mark. 2001. Nobility in the history of adjectives ending in *-ic* and *-ical*. En Ruth Brend, Alan K. Melby e Arle R. Lommel (eds.), *LACUS Forum XXVII: Speaking and comprehending*, 35–46. Fullerton: LACUS.

Krawczak, Karolina. 2014. *Shame* and its near-synonyms in English: A multivariate corpus-driven approach to social emotions. En Iva Novakova, Peter Blumenthal e Dirk Siepmann (eds.), *Les émotions dans le discourse / Emotions in discourse*, 83–94. Frankfurt: Peter lang.

Krawczak, Karolina. 2018. Reconstructing social emotions across languages and cultures: A multifactorial account of the adjectival profiling of SHAME in English, French, and Polish. *Review of Cognitive Linguistics* 16(2). 455–493.

Liu, Dilin. 2010. Is it a *chief*, *main*, *major*, *primary* or *principal concern*? A corpus-based behavioral profile study of the near-synonyms. *International Journal of Corpus Linguistics* 15(1). 56–87.

Liu, Dilin. 2013. Salience and construal in the use of synonymy: A study of two sets of near-synonymous nouns. *Cognitive Linguistics* 24(1). 67–113.

Liu, Dilin e Maggie Espino. 2012. *Actually*, *genuinely*, *really*, and *truly*. A corpus-based behavioral profile study of the near-synonymous adverbs. *International Journal of Corpus Linguistics* 17. 198–228.

Murphy, M. Lynne. 2003. *Semantic relations and the lexicon: Antonymy, synonymy, and other paradigms*. Cambridge: Cambridge University Press.

Murphy, M. Lynne. 2013. What we talk about when we talk about synonyms (and what it can tell us about thesauruses). *International Journal of Lexicography* 26(3). 279–304.

Nuyts, Jan e Pieter Byloo. 2015. Competing modals: Beyond (inter)subjectification. *Diachronica* 32(1). 34–68.

Persson, Gunnar. 1989. Deep *and* profound*: A study in so-called synonymy*. Umeå: Printing office of Umeå University.

Primahadi-Wijaya-R, Gede e I Made Rajeg. 2014. Visualising diachronic change in the collocational profiles of lexical near-synonyms. En I Nengah Sudipa e Gede Primahadi-Wijaya-R (eds.), *Cahaya bahasa: A Festschrift in honour of Prof. I Gusti Made Sutjaja*, 247–258. Denpasar: Swasta Nulus.

Reid, Thomas. 1785. *Essays on the intellectual powers of man*. 1ª edición. Volume 1. Edimburgo e Londres: John Bell and G. G. J e J. Robinson.

Samuels, Michael L. 1972. *Linguistic evolution with special reference to English*. Londres e Nova York: Cambridge University Press.

Sinclair, John M. 1966. Beginning the study of lexis. En Charles E. Bazell (ed.), *In memory of J.R. Firth*, 410–429. Harlow: Longman.

Taylor, John R. 2003. Near synonyms as co-extensive categories: 'high' and 'tall' revisited. *Language Sciences* 25(3). 263–284.

Wierzbicka, Anna. 1988. *The semantics of grammar*. Ámsterdam: John Benjamins.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA