

Postprint version of accepted manuscript. The paper was published as: Pettersson-Traba, Daniela. 2021. 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): 138-165. DOI <https://doi.org/10.1558/jrds.40245>

## **Measuring semantic distance across time: An analysis of the collocational profiles of a set of near-synonyms in American English**

*Daniela Pettersson Traba*

Over the last decades, several studies have analyzed the collocational preferences of particular sets of near-synonyms from a synchronic viewpoint, while their diachronic development has generally been disregarded. The aim of this paper is to partially fill this gap by examining the collocational behavior of the adjectives *fragrant*, *perfumed*, and *scented*, which denote the concept SWEET-SMELLING, over the time span 1810--2009. To this purpose, instances of the three near-synonyms and their L5-R5 collocates were extracted from the *Corpus of Historical American English (COHA)* and then submitted to statistical modelling. Results indicate that, at the beginning of the time span analyzed, the collocational preferences of *scented* and *perfumed* are very similar but, over time, *scented* becomes semantically closer to *fragrant*, while at the same time taking over some of its functions.

Keywords: near-synonymy; collocation; semantic vector spaces; collocational networks; diachrony

### **1 Introduction**

It is a well known fact that most synonyms existing in language are actually near-synonyms, that is, words or constructions which share the same core denotational meaning, but differ in peripheral aspects or in other dimensions of meaning such as connotation, style, and/or collocation (e.g. Cruse, 2000: 159--160; Liu, 2010).<sup>1</sup> In fact, languages tend to work against absolute synonymy, which leads to competition between semantically related words (e.g. Bolinger, 1977: ix--x, 9; Croft, 2000: 176). As argued by Samuels (1972: 62), '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.' This has a clear diachronic implication, since synonyms are expected to become functionally more dissimilar over time. The idea that semantically related concepts compete is recurrent in research on language change, in which we can find several examples of synonymous expressions that have eventually become semantically less similar or cases in which one item has come to substitute the other. For instance, following the Norman Conquest, many Anglo Saxon words were duplicated by French loans with the same meaning. Over time, however, they underwent semantic change; this is the case, for example, of the pairs *pig-pork* and *cow-beef* (Jackson, 1988: 66; Murphy, 2003: 161). However, recent research has argued that the competition theory could be an oversimplification since, besides differentiation, synonyms can also undergo a process of attraction in which they become semantically more similar (De Smet, D'hoedt, Fonteyn, and Goetham, 2018). In this scenario, synonymous words or constructions begin to mirror each other's behavior and

thus come to share more semantic space. In fact, a certain degree of attraction might very well be a precondition for replacement, as synonymous expressions probably need to share a great deal of their semantic features for them to be considered interchangeable by language users and for substitution to occur. Finally, yet another possibility is that of stability, to wit, when no changes in the functional profiles of the near-synonyms occur.

One way in which one can measure semantic (dis)similarity between near-synonyms is by analyzing their collocational preferences. In particular, it is possible to examine the degree of collocate overlap between two or more near-synonymous expressions by considering the number of significant collocates they share as well as the number of significant collocates one synonym exhibits but not the other/s (Gries, 2001: 83). According to this criteria, two related words would be more similar the more significant collocates they have in common, and more dissimilar the more collocates they do not have in common. The importance of collocations, first emphasized by Firth (1957) and Sinclair (1966) and summarized in the quote ‘You shall know a word by the company it keeps’ (Firth, 1957: 11), has been one of the central tenets of lexical semantics. The term ‘collocation’ has received several slightly different interpretations over the years along the following dimensions identified by Gries (2013: 138–139): (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 for an expression cooccurring with a node word in order to be considered a collocate, (iv) the distance between the items making up the collocation, i.e. whether they are directly adjacent, syntactically related, or within a context window of  $x$  words, (v) the specificity of the lexical items involved, to wit, word forms or lemmas, and (vi) the degree of compositionality and predictability of the collocation. For the purposes of the present study, we will follow Stubbs (2001: 24) and define collocation as ‘a lexical relation between two or more words which have a tendency to cooccur within a few words of each other in running text’, although we will establish in detail in Section 2 a specific frequency threshold and context window for items to be considered collocates in the present analysis.

Over the last fifty years, advancements in corpus linguistics have led to the emergence of rigorous investigations into the role of individual collocates on various linguistic phenomena, including semantic ones such as polysemy and near-synonymy. Although several studies have analyzed the collocational preferences of particular sets of near-synonyms (among other types of distributional patterns) from a synchronic viewpoint to quantify their semantic (dis)similarity (e.g. Kjellmer, 2003; Taylor, 2003; Divjak and Gries, 2006, 2008; Divjak, 2010; Liu and Espino, 2012; Liu, 2013; Desagulier, 2014), very few have paid attention to the diachronic evolution of the collocational behavior of specific groups of lexical near-synonyms (but see Primahadi-Vijaya-R. and Rajeg, 2014; Pettersson-Traba (2018)). Primahadi-Wijaya-R. and Rajeg (2014) conducted a corpus based analysis of the nominal collocational profiles of the near-synonymous adjectives *hot* and *warm* during the last one and a half centuries (i.e. 1860–2009) in American English. Their results uncovered various diachronic patterns that contributed to these two near-synonyms exhibiting different collocational behavior in Contemporary American English. For example, over time, *warm* comes to cooccur more frequently with nouns such as *heart*, *welcome*, and *smile*, therefore highlighting this this adjective’s prominence

in the metaphorical sense ‘of the heart, feelings, etc.: Full of love, gratitude, approbation, etc.; very cordial or tender’ (*Oxford English Dictionary [OED]* s.v. *warm* adj. 12a). In contrast, *hot* increasingly collocates with lexical items such as *dog* from the 1920s onwards, and with nouns referring to people (e.g. *girl*, *guy*, and *woman*) from the 2000s onwards with the meaning ‘[...] (originally a woman): sexually attractive; sexy’ (*OED* s.v. *hot* adj. 12i). Pettersson-Traba (2018) conducts a preliminary study on the diachronic development of the attributive uses of *fragrant*, *perfumed*, *scented*, and *sweet-smelling*, in the latter part of Late Modern and Present-day American English. She delineates the internal semantic structure of this set of synonyms by paying special attention to their noun collocates, which are grouped into semantic categories on the basis of the classification in the *Historical Thesaurus of the Oxford English Dictionary*. The results show that the four adjectives undergo major changes over the time span examined (1850-2009), going 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 extra-linguistic 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 have led to an ever-increasing need to allude to artificially scented lotions and candles rather than naturally fragrant plants. Moreover, *fragrant* and *perfumed*, which initially were the most frequent adjectives, are gradually replaced by *scented*, thus reflecting a change in the relation between the synonyms over time.

The present paper further contributes to the line of research initiated in Pettersson-Traba (2018) by examining the individual collocates of three of the four adjectives examined, to wit, *fragrant*, *perfumed*, and *scented*, in the *Corpus of Historical American English (COHA)*; Davies 2010--). Illustrative examples of the three lexical items are found in (1)-(3).

- (1) The kitchen was warm from the slowly burning range and *fragrant* with the beans which had been cooking since yesterday. (*COHA*, 1934, Fiction, Folks)
- (2) She too, in a sense, is a portrait of captive refinement, but she is more captive than refined. Her tiny body *perfumed* and irritable, she rubs against the bars of her cage. (*COHA*, 1993, Fiction, CityManyDays)
- (3) She worked the cloth over a small sliver of *scented* soap she had scavenged and lathered herself liberally, reveling in the pungent fragrance. (*COHA*, 1972, Fiction, FlameFlower)

The decision of analysing these three particular near-synonymous adjectives was made after careful examination of dictionaries and thesauri. While some additional adjectives are defined in the same way as the adjectives selected here (e.g. *fragranced* and *sweet-smelling*) or are listed as their synonyms (e.g. *odorous* and *redolent*) and could consequently have been included in the study, these adjectives have not been chosen for different reasons. The adjective *fragranced* is categorized as ‘rare’ in some of the sources and is not attested in *COHA*. Similarly, *sweet-smelling* displays a relatively low frequency in the corpus and thus the retrieval of significant collocates of this adjective yields few hits. The adjective *odorous* does not necessarily entail the positive connotation that the selected adjectives display and the meaning of *redolent* differs somewhat from *fragrant*, *perfumed*, and *scented* as it does not imply the trait ‘sweetness’ by definition. However,

an analysis with a wider range of near-synonyms, including the abovementioned adjectives, might prove valuable in future research.

Despite the exclusion of *sweet-smelling*, the results of the present study are based on an almost four times larger dataset than that in Pettersson-Traba (2018), given that occurrences of the three adjectives in all their syntactic environments, including also predicative and postpositive uses, are considered (cf. (1)--(3)). Furthermore, the semantic analysis of the adjectives is considerably more fine-grained, since the focus here is on specific collocates and not on more abstract semantic classifications as in the case of Pettersson-Traba (2018). By zooming in on the individual collocates it is possible to identify differences between the adjectives which may be obscured if the collocates are grouped into broader classes such as semantic categories (e.g. CLEANING AND PERSONAL CARE, FARMING AND HORTICULTURE, PEOPLE, PLANTS, and WEATHER). In sum, the present paper aims at analyzing the idiosyncratic collocational preferences of the three near-synonymous adjectives, which denote the concept SWEET-SMELLING, in 19th- and 20th-century American English.

This set of near-synonyms is particularly interesting to examine as made evident by a thorough examination of different types of reference material (e.g. dictionaries and thesauri): generally speaking, no clear and detailed information about the usage patterns and nuances of meaning of the three near-synonyms under investigation is provided. In fact, in many cases the adjectives are defined in terms of each other as the following definition of *fragrant* illustrates:

- A. Pleasant smelling, perfumed. (*Newbury House Dictionary of American English* s.v. *fragrant*)

Moreover, the definitions and examples of usage, though offering valuable information about their similarities, do not provide a comprehensive picture of their differences, which prevents users from fully comprehending how to distinguish them. In what follows, an overview of the information provided in seven reference materials is offered. These are the historical dictionary *Oxford English Dictionary* (*OED*; 2012--) and six present-day English dictionaries and thesauri, namely *Lexico* (2019), *Cambridge Dictionary* (*CD*; 2019), *Collins online Unabridged English Dictionary* (*Collins*; 2012--), *Newbury House Dictionary of American English* (*NHDAE*; 2019), *Longman Dictionary of Contemporary English* (*LDOCE*; 2015--), *Merriam-Webster Dictionary and Thesaurus* (2019), and *MacMillan Dictionary* (2009--). The main reasons for choosing these reference materials are that (i) with the exception of the *OED*, they are all PDE dictionaries, and should therefore reflect how the adjectives are currently used; (ii) most of them include an American English section, which is of great importance here, since the present analyses focus on this variety of English; and (iii) they provide definitions, examples of usage, and suggested synonyms of the adjectives at issue.

Both in the *OED* and the PDE dictionaries and thesauri, the three adjectives seem to be entirely interchangeable as they share the same basic and central meaning: 'having a sweet or pleasant smell or odor'. In fact, many of the dictionaries provide only this sense for each of the adjectives, thus making them seem monosemic and undistinguishable. Only a few of the dictionaries consulted offer information about their particular nuances of meaning and usage patterns. Some reference sources provide additional senses for

*perfumed* (*Lexico*, *Collins*, and *OED*) and *scented* (*CD*, *Collins*, and *OED*). According to these dictionaries, there seems to be a difference in nuance in the case of these two adjectives depending on whether the source of the pleasant smell or odor is natural or artificial. This difference is reflected in the examples of usage in the types of nouns that the two adjectives modify. When denoting a ‘naturally sweet or pleasant smell or odor’ the two adjectives modify nouns such as *flower*, *bower*, *cherry*, *pine*, and *breeze*. Contrariwise, when denoting an ‘artificial pleasant smell or odor’, they modify nouns which refer to manmade objects, as for instance, *SOAP*, *GLOVE*, *CANDLE*, *PAPER*, *LAMP*, and *LEATHER*. Although these apparent differences in nuances of meaning can be expected to bring about differences in usage patterns among the adjectives, especially when it comes to the types of nouns they typically modify, the information provided is too limited to know for which types of nouns the adjectives normally serve as modifiers. To illustrate this point, even though the ‘artificial’ sense is only provided for *perfumed* and *scented*, examples in which *fragrant* collocates with nouns referring to manmade objects, can easily be found, as in (4) and (5):

- (4) Inside are quirky old settees, painted chests and weathered wood hutches brimming with *fragrant soaps* and *candles*. (*LDOCE*, s.v. *fragrant*)
- (5) With soap still to be invented, the *fragrant oils* and waters were used in bathing and for perfuming hair. (*Lexico*, s.v. *fragrant*)

Distinguishing the contexts in which *fragrant*, *perfumed*, and *scented* can be used interchangeably and those in which they cannot is, therefore, not a straightforward task. It is only by means of a careful examination of their collocational preferences that we can begin to understand the complex interrelations of this particular set of near-synonyms, and this is precisely the goal of the present contribution. Section 2 deals with the data and the methodology employed to achieve this goal and, in Section 3, the results of the study are presented and discussed. Finally, some concluding remarks and suggestions for future research are put forward in Section 4.

## 2 Data and methodology

The data for the present study was extracted from *COHA*, which contains more than 400 million words from American English. It covers the time span 1810–2009, which, for the purposes of the present paper, was divided into four fifty year periods: (i) P1, from 1810 to 1859, (ii) P2, from 1860 to 1909, (iii) P3, from 1910 to 1959, and (iv) P4, from 1960 to 2009.<sup>2</sup> Since the focus here lies on the collocational behavior of the near-synonyms *fragrant*, *perfumed*, and *scented*, the lemmas of all noun collocates in a context window of 5 words to the left and 5 words to the right of the adjectives were also retrieved from *COHA* by making use of the COLLOCATE and POS tag options in the corpus. This L5-R5 context window was selected in this case since it has been shown that tighter windows such as 1 or 2 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, and Geeraerts, 2008: 40). In turn, if one is interested in typical collocates of the target, it is desirable 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*) or adverbs which modify the adjectives (e.g. *the deliciously fragrant flower*), among others. 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) and Gries (2001; 2003), nouns are more informative than other word types when it comes to the semantics of adjectives. Geeraerts (1986) study on the Dutch adjective *vers* ‘fresh’ demonstrated that the fine-grained aspects of meaning of polysemous adjectives can be discovered by examining the nouns they modify. He provides evidence of *vers* having different meanings depending on the nouns it accompanies. For instance, when modifying *wound*, *vers* means ‘fresh, recent’, whereas when modifying the noun *air*, it takes on a slightly different meaning, namely ‘fresh, pure, untainted’, or ‘optimal.’

The retrieval process resulted in 4,990 tokens of the adjectives, which collocated with a total of 10,740 tokens and 2,682 types (lemmas) of context words. These were subsequently fed into two types of statistical analyses in order to identify (dis)similarities in the near-synonyms’ collocational preferences and, therefore, their semantic structure. The methodology employed here demonstrates how techniques of a more quantitative nature can complement and enhance qualitative inquiries into linguistic data. First, the data was analyzed by means of Semantic Vector Space (SVS) modelling. SVS is a technique that enables us to measure semantic (dis)similarity between related words on the basis of their collocational profiles. An SVS model is constructed following a series of steps (e.g. Levshina, 2015: 326; Hilpert and Correia Saavedra, 2017):

1. We create a table containing the raw cooccurrence frequencies of the target words (in the columns) and the context words (in the rows). In this case, a separate column was included for each near-synonym in each period, since we are interested in the evolution of their collocational preferences over time.
2. The raw cooccurrence frequencies are transformed into collocational strength values using an association measure such as Pointwise Mutual Information (PMI; e.g. Church and Hanks, 1990: 23). Even though a variety of association measures exist to compute collocational strength such as log likelihood or  $\Delta P$  (e.g. Gries, 2013), PMI is used in the present analysis given that it is already provided in *COHA* when searching for the collocates of a specific word in the corpus.
3. We compute cosine similarity scores (Levshina, 2015: 328) between the target words (here *fragrant*, *perfumed*, and *scented* in each period) and transform them into distance values for an easier visualization of the results.

Two different visualization techniques are employed in the present study in order to explore and interpret the semantic distances between the adjectives. These are cluster analysis (Levshina, 2015: 301), a method that helps identifying groups or clusters of objects on the basis of their (dis)similarities, and multidimensional scaling (MDS; Levshina, 2015: 336--337), an approach that depicts differences between two or more objects as distances in a two- or three-dimensional plot. Cluster analysis and MDS complement one another because the former is useful to identify coherent and delimited groups in the data, while the latter can be said to be semantically more realistic in that 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 and Gries, 2017: 14).

The second type of statistical procedure to which the data was submitted was collocational network analysis which supplements the findings of the SVS analysis (Brezina, McEnery, and Wattam, 2015; Baker, 2017: 95--101).<sup>3</sup> Put simply, this method consists in selecting the most prominent collocates of the target words (in this case *fragrant*, *perfumed*, and *scented*) and visualizing them in a collocational network.<sup>4</sup> The idea is that one can observe, on the one hand, increases or decreases in the frequency of the target words and, on the other, the competition between them: if one target word is drawing away collocates from another, this means that it is taking over some of its semantic space. Four collocational networks were built, one per period, which enabled us to determine which collocates the near-synonyms share at different points in time and to identify potential variations in the internal semantic structure of the set. 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 frequency of 5 with the target words and a PMI value of 3 or higher in each of the four periods. These thresholds drastically decreased the number of collocates of the near-synonyms, if compared to the SVS analysis, thus allowing us to conduct a more qualitative and careful examination of the collocational profiles of the adjectives. Higher thresholds would lead again to data sparseness, particularly in the case of *perfumed* and *scented*, which display a lower frequency in *COHA* (cf. Table 1). Therefore, the threshold proposed by Baker (2017: 96) for words of a higher frequency, namely a minimum raw frequency of 10 and a PMI of 6 was here discarded. On the contrary, lower thresholds, that is a minimum frequency of lower than 5 and a PMI lower than 3, would include collocates which might not be particularly typical of the adjectives. Additionally, 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 and Hanks 1990; Church et al. 1991; Church et al. 1994; Liu 2010).

### 3 Results

A total of 4,990 instances of the three near-synonyms were retrieved from *COHA*. As shown in Table 1, *fragrant* is the most frequent adjective of the set, with a total frequency of 3,395 (68.04 %), followed by *perfumed* (808 instances, 16.19 %) and then *scented* (787 instances, 15.77 %). If we examine their relative frequencies in each of the four periods, an interesting trend can be observed. *Fragrant*, despite being the most common adjective in all four periods, decreases in frequency over time, from 75 % in P1 (1810--1859) to 59.33 % in P4 (1960--2009). *Scented*, on the other hand, goes from being the least frequent adjective of the set in P1, with a relative frequency of 10.17 %, to occupying the second position (22.60 %) in P2 (1860--1909). The most dramatic changes, however, take place from P2 to P3 (1910--1959), that is, during the transition from the 20<sup>th</sup> to the 21<sup>st</sup> centuries, since *fragrant* decreases more than 10 percentual points in frequency while *scented* increases by almost 10 %. Finally, *perfumed* does not change much in the time span examined but its relative frequency increases slightly, from 14.86 % in P1 to 18.06 % in P4. However, if the normalized frequencies of the adjectives per million words are considered, it is clear that all three synonyms become less frequent over time, although

this downward tendency is much more pronounced in the case of *fragrant* and *perfumed*. *Scented*, on the other hand, remains relatively stable with only a minor decrease in P4. The distribution shown in Table 1 is statistically significant according to a chi-square test of independence ( $\chi^2 = 125.27$ ,  $df = 6$ ,  $p < 0.001$ ): *fragrant* is significantly more frequent than expected in P1 and P2, but significantly less frequent than expected in P3 and P4, while the opposite tendency is true for *scented*. The distribution of *perfumed*, in turn, is not statistically significant.

Table 1: Frequency of the synonyms per period

Synonym	P1 (1810-- 1859)	P2 (1860-- 1909)	P3 (1910-- 1959)	P4 (1960-- 2009)	Total
<i>Fragrant</i>					
N	804	1291	712	588	3,395
%	75.00	72.49	62.13	59.33	68.04
NF	14.77	12.87	5.87	4.51	8.36
<i>Perfumed</i>					
N	159	281	189	179	808
%	14.83	15.78	16.49	18.06	16.19
NF	2.92	2.80	1.56	1.37	1.99
<i>Scented</i>					
N	109	209	245	224	787
%	10.17	11.73	21.38	22.60	15.77
NF	2.00	2.08	2.02	1.72	1.94
Total					
N	1072	1781	1146	991	4,990
%	100	100	100	100	100
NF	19.70	17.75	9.45	7.61	12.28

Therefore, *scented*, despite being the least common adjective of the set at beginning of the 19<sup>th</sup> century, increases in frequency at the expense of *fragrant*, which decreases over time despite still being the default choice to denote the concept SWEET-SMELLING. *Fragrant* and *scented* thus seem to have undergone a process of convergence in terms of frequency, particularly at the turn of the century. However, it is not clear whether this gradual confluence is limited to frequency alone or if it is also the case that *fragrant* and *scented* have become semantically more similar over time, with *scented* overtaking some of the functions of *fragrant*. This will be the focus of sections 3.1 and 3.2, with the former offering a more quantitative bird's eye view of the evolution of the collocational preferences of the three near-synonyms and the latter presenting the results of a more qualitative in depth analysis of their most prominent collocates per period.

### 3.1 Diachronic Semantic Vector Space analysis

Figure 1 plots the two dimensional MDS map of the collocational preferences of *fragrant*, *perfumed*, and *scented* per period. The figure is interpreted as follows: each point in the graph represents one near-synonym per period, and the closer two points are located in the graph, the more similar the collocational preferences of those adjectives are. The vertical axis in Figure 1 (Dimension 2) arranges the points in the graph according to adjective, with *perfumed* at the top, *fragrant* at the bottom, and *scented* occupying the middle ground but overall being much closer to *perfumed* (except in P4). The horizontal

axis (Dimension 1), on the other hand, arranges the points on a temporal continuum from left to right, with P1 (1810--1859) and P2 (1860--1909) exhibiting negative values and P3 (1910--1959) and P4 (1960--2009) exhibiting positive ones. Independent-samples *t*-tests on the MDS coordinates confirm the visual interpretation of Figure 1. *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 differences between the near-synonyms along the horizontal axis are not significant. In fact, on this axis the points are arranged 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 P3 and P4, on the other, are not significant. The differences among the periods on the vertical axis do not reach statistical significance. Therefore, whereas the adjectives seem to be fairly similar in P1 and P2, on the one hand, and in P3 and P4, on the other, concerning their collocational profiles, there is a considerable change from P1/P2 to P3/P4, that is, at the turn of the 19<sup>th</sup> to the 20<sup>th</sup> century.

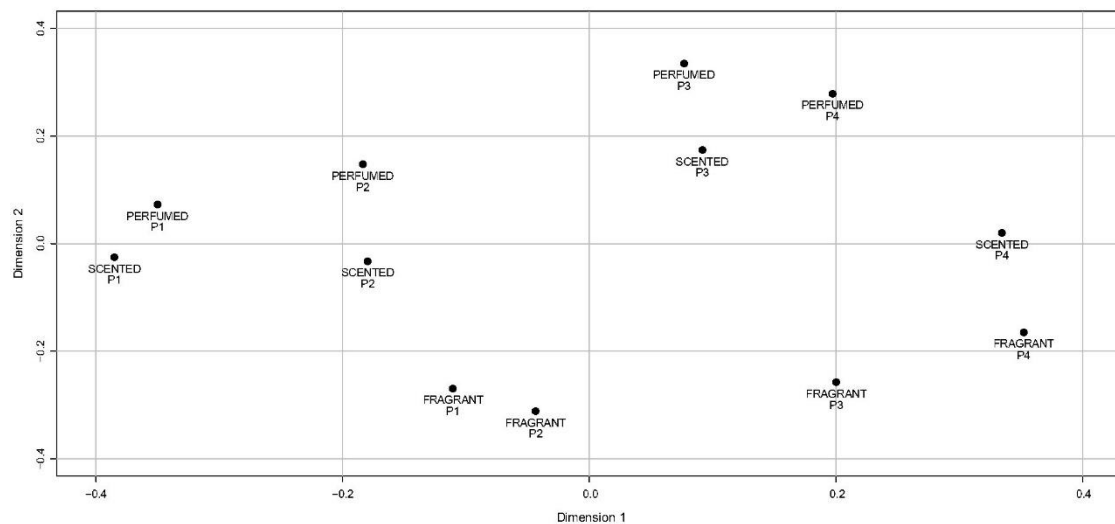


Figure 1: Two dimensional MDS map of the collocational preferences of *fragrant*, *perfumed*, and *scented* in four time periods (stress = 0.21)<sup>5</sup>

By making use of MDS we have been able 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 19<sup>th</sup> century and the 20<sup>th</sup> and 21<sup>th</sup> centuries. However important these differences may be, this analysis does still not clarify the nature of such differences. To this end, the 2,682 types of noun collocates were classified into semantic categories by employing the *UCREL Semantic Analysis System*, or *USAS* for short (Archer, Wilson, and Rayson, 2002). *USAS* contains twenty-one major semantic classes that are then further divided into more specific subclasses.<sup>6</sup> This tool allows researchers to automatically analyze strings of words according to their semantics and thus check the domains to which particular words belong. Then, a PMI score for each semantic category and sub-category was calculated for each near-synonym in each period. This was done so as to determine whether correlations existed between the points' MDS coordinates and

their PMI scores with the semantic categories and sub-categories. The following significant correlations emerged from the analysis:

- (a) Dimension 1 and USAS semantic category F4 (FARMING AND HORTICULTURE) are negatively correlated:  $r = -0.63, p < 0.05$ .
- (b) Dimension 1 and USAS semantic category W4 (WEATHER) are negatively correlated:  $r = -0.58, p < 0.05$ .
- (c) Dimension 2 and USAS semantic category B (THE BODY AND THE INDIVIDUAL) are positively correlated:  $r = 0.88, p < 0.001$ .
- (d) Dimension 2 and USAS semantic category L3 (PLANTS) are (marginally) negatively correlated:  $r = -0.54, p = 0.06$ .
- (e) Dimension 2 and USAS semantic category M3 (MOVEMENT/TRANSPORTATION: LAND) are (marginally) negatively correlated:  $r = -0.53, p = 0.07$ .
- (f) Dimension 2 and USAS semantic categories Q1 (COMMUNICATION) and Q4 (THE MEDIA) are positively correlated:  $r = 0.77, p < 0.01$  and  $r = 0.69, p < 0.05$ .
- (g) Dimension 2 and USAS semantic category S2 (PEOPLE) are positively correlated:  $r = 0.8, p < 0.01$ .

First, regarding the behavior of the three near-synonymous adjectives across periods, that is, Dimension 1 (horizontal axis of Figure 1), two semantic categories are significantly correlated with the MDS coordinates. These are F4 (FARMING AND HORTICULTURE) and W4 (WEATHER). Table 2 displays examples of types of noun collocates (lemmas) belonging to each category. The correlations in both cases are negative, which implies that *fragrant*, *perfumed*, and *scented* collocate with such nouns significantly more often the lower their values on the horizontal axis are. In other words, in P1 (1810--1859) and P2 (1860--1909) the adjectives appear with F4 and W4 nouns significantly more frequently than in P3 (1910--1959) and P4 (1960--2009). This change could perhaps be understood as a reflection of the transformation of American society, going from a preindustrial and mainly rural society in the 19<sup>th</sup> century to the world's major industrial and commercial power in the 20<sup>th</sup> and 21<sup>st</sup> centuries (Jones, 1996: chapter 16; Baker, 2017: 193--194). Consequently, the importance of farming and horticultural activities in the daily life of most American citizens have probably radically decreased over time, specifically after the First and Second Industrial Revolutions, which led to a mass migration from the countryside to the cities. In a more indirect fashion, the importance of weather phenomena can also be said to have declined with time as many of the farming and horticultural activities heavily relied or even depended on the climate.

Table 2: Examples of types of noun collocates across semantic (sub-)categories

Semantic (sub-)category	Examples of types of noun collocates
B THE BODY AND THE INDIVIDUAL	B1: <i>hand, hair, neck, shoulder, wrist</i> B4: <i>detergent, perfume, lotion, shampoo, soap</i> B5: <i>clothes, handkerchief, garment, glove, lace</i>
F4 FARMING AND HORTICULTURE	<i>crop, farm, farmhouse, field, mow pasture, plantation, vineyard</i>
L3 PLANTS	<i>cactus, garden, magnolia, pollen, pumpkin, root, vine</i>

M3 MOVEMENT/TRANSPORTATION: LAND	<i>avenue, path, pathway, road, wayside</i>
Q1 COMMUNICATION	<i>draft, leaflet, letter, message, note, notepaper, stationery</i>
Q4 THE MEDIA	<i>article, book, edition, newspaper, paper</i>
S2 PEOPLE	<i>boy, child, gentleman, madam, maiden, man, people, person, woman</i>
W4 WEATHER	<i>breeze, flood, fog, haze, mist, rain, snowfall, storm, wind</i>

Second, the MDS coordinates on the Dimension 2 (vertical axis of Figure 1) are negatively correlated with the PMI scores of the near-synonyms and the semantic categories L3 (PLANTS) and M3 (MOVEMENT/TRANSPORTATION: LAND), which means that the adjectives collocate with nouns belonging to these two categories (cf. Table 2) more often the lower their values on MDS Dimension 2. They are thus more closely associated with *fragrant* than with *perfumed* and *scented*. The nouns in L3 and M3 are most likely used in contexts where the adjectives are used to refer to natural pleasant and sweet smells. This is specifically the case of L3 nouns which refer to plants in general (e.g. *flower, shrub, tree*) particular flowers or trees (e.g. *azalea, cedar, geranium*), and parts of different plants (*bud, leaf, stem*), as illustrated in examples (6) and (7):

- (6) 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)
- (7) 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, categories or sub-categories B (THE BODY AND THE INDIVIDUAL), Q1 (COMMUNICATION), Q4 (THE MEDIA), and S2 (PEOPLE) are positively correlated with MDS Dimension 2. Noun collocates belonging to these semantic classes therefore appear more frequently with the adjectives the higher their values on the vertical axis, meaning that they are more closely connected with *perfumed* and *scented* than with *fragrant*. Most of these noun collocates with all likelihood correspond to uses of the adjectives in the artificial sense, particularly in the case of 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, that is, 49.2 %. The same is true of sub-categories Q1 and Q4 (cf. Table 2). Categories B1 (ANATOMY AND PHYSIOLOGY), i.e. the most prominent sub-class of category B (76 types, i.e. 40.6 %), and S2 are more ambiguous as it is not always clear whether the smell referred to in such contexts is natural or artificial (cf. example (8), but many occurrences with such collocates do correspond to the artificial sense of the adjectives, as illustrated in (9):

- (8) 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. (COHA, 1984, FIC, CrescentCity)
- (9) 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)

In (8), *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 point us to the correct interpretation, if there even is one. In turn, (9) *fragrant* is clearly used in the artificial sense, as evidenced by the context, where the artificial perfumes used to impregnate the locks of the women are mentioned.

It is here worth noting that two additional SVS analyses were also conducted. First, the contribution of non-nominal content words was also assessed by including adjective, adverb, and verb collocates of the three near-synonyms, as well as the already examined nominal collocates. Although there is a minimal improvement in the two-dimensional MDS solution (stress = 0.19) when all content words are considered, the resulting patterns were not as elucidating as the ones discussed in the previous paragraphs. This is in line with Geeraerts (1986) and Gries (2001; 2003) claims that nouns seem to be more informative than other parts-of-speech when the semantics of adjectives is concerned. Second, an SVS analysis was also conducted on the basis of only nouns which are modified by the adjectives, as is the case of *kitchen*, *body*, and *soap* in examples (1)-(3) (see Section 1), respectively. This represents a further refined and stricter notion of collocation, namely one which only considers as collocates those nouns which are involved in a dependency relation with the synonymous adjectives. In this case, however, the resulting MDS solution is much worse than the one visualized in Figure 1 (stress = 0.30).

The results of a hierarchical cluster analysis of the near-synonyms can help us interpret the evolution of their collocational preferences over time. Figure 2 displays the distances between the adjectives across periods; the higher two components are merged, the more dissimilar they are. The red 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 the clusters are well-formed (Levshina, 2015: 311). In this case, a solution of five clusters exhibits the highest average silhouette width (0.49): (i) *perfumed* and *scented* in P1 and P2 (cluster 1), (ii) *scented* in P4 (cluster 2), (iii) *perfumed* in P3 and P4 and *scented* in P3 (cluster 3), (iv) *fragrant* in P3 and P4 (cluster 4), and (v) *fragrant* in P1 and P2 (cluster 5). The stability of the clusters can be assessed by calculating their Approximately Unbiased (AU) *p*-values, which determine how well supported they are by the data. In this case, the closer the values are to 1, the more stable the clusters are (Levshina, 2015: 315--317). The AU values for clusters are the following: 0.89 for cluster 1, 0.9 for cluster 2, 0.93 for cluster 3, 0.88 for cluster 4, and 0.78 for cluster 5. Thus, none of the five cluster reach the traditional 0.95 significance threshold, but most of them are close to this value. The least stable cluster is cluster 5, that is, *fragrant* in P1 (1810--1859) and P2 (1860--1909), which means that if we examined a new data sample, they would probably not be grouped together. On the other hand, cluster 3, i.e. *perfumed* in P3 (1910--1959) and P4 (1960--2009), and *scented* in P3, almost reach significance and can thus be considered to be fairly stable.

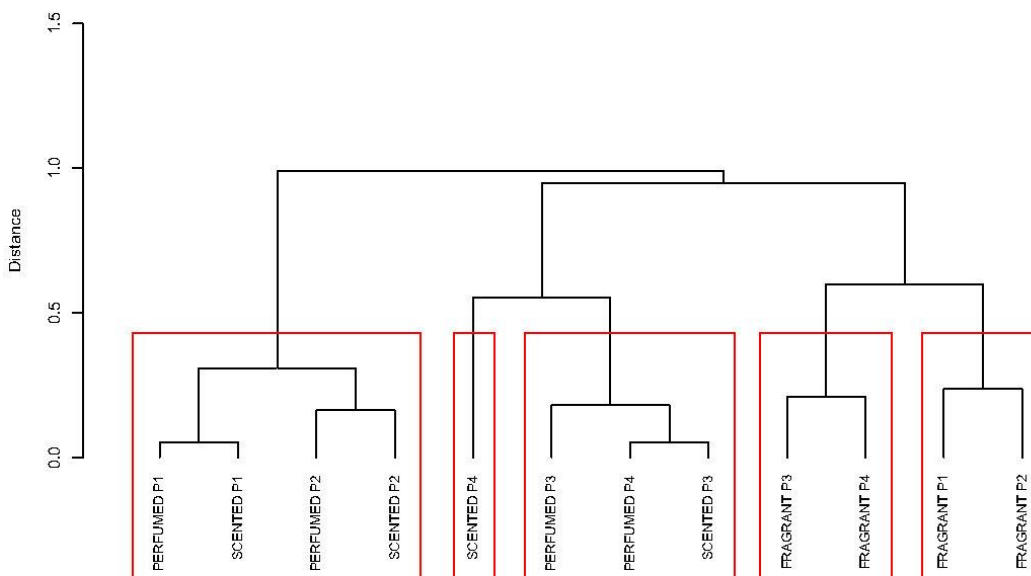


Figure 2: Dendrogram of the collocational preferences of *fragrant*, *perfumed*, and *scented* in four time periods

According to Figure 2, the collocational profiles of the adjectives are grouped according to period and near-synonym. On the one hand, the collocational profiles of the adjectives in P1 and P2 are grouped together, as well as those in P3 and P4. On the other hand, the collocational profiles of *fragrant* are kept apart from those of *perfumed* and *scented*, which are grouped together. The only exception to this pattern is the collocational behavior of *scented* in P4, which forms its own cluster, thus being separate from the cluster containing *perfumed* in P3 and P4 and *scented* in P3.

The diachronic evolution of the collocational preferences of the three near-synonyms across the four time periods, as shown by both the MDS and the cluster analyses, can thus be summarized as follows. In P1, the collocational profiles of *perfumed* and *scented* are very similar, while *fragrant* differs substantially from them, and not much seems to change from P1 to P2, particularly in the case of *fragrant*. This means that 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 collocations and, therefore, semantic space. From P2 to P3, the collocational behavior of the three adjectives seems to evolve substantially, but *perfumed* and *scented* are still closer to each other than they are to *fragrant*. Not much evolution is observed from P3 to P4 in the collocational profiles of *fragrant* and *perfumed* but, contrariwise, that of *scented* does change. In fact, in P4 *scented* moves away from *perfumed* and becomes more similar to *fragrant* in terms of collocations, as displayed by the MDS map (cf. Figure 1). Therefore, as in the case of the evolution of the adjectives as to their frequency, most changes in collocational behavior seem to take place from P2 to P3, that is, during the transition from the 19<sup>th</sup> to the 20<sup>th</sup> centuries, with *fragrant* and *scented* undergoing a process of convergence and becoming

relatively similar in P4. This indicates that the gradual attraction between *fragrant* and *scented* over time is not only in terms of their productivity (i.e. frequency), with *scented* becoming a more productive member of the set at the expense of *fragrant*, but also in terms of their semantics, with the collocational profiles of *fragrant* and *scented*, and thus the semantic space occupied by both adjectives, becoming more similar. As demonstrated by the interpretation of the vertical axis on the MDS map (i.e. Dimension 2), the downward movement of *scented* from P3 to P4 could reflect a change from denoting mostly artificial smells to becoming more neutral in this respect, not showing a clear preference for either the artificial or the natural sense. This is so because, in P4, *scented* is positioned very close to 0 on the vertical axis. On the other hand, *perfumed* and *fragrant* seem to be more specialized to the artificial and natural senses respectively, thereby being positioned on opposite ends of the continuum. Put simply, a process of attraction between *fragrant* and *scented* has therefore been identified at the end of the time span considered, both in terms of their productivity and collocational preferences. At the same time, *scented* becomes differentiated from *perfumed*, which seems to be the default choice to refer to artificial smells. This bird's eye view of the evolution of the collocational profiles of *fragrant*, *perfumed*, and *scented* has enabled us to identify potential semantic changes in the internal structure of this set of near-synonyms. However, given the large number of noun collocates (i.e. 2,682 types and 10,740 tokens) considered in the analyses, we have not been able to discuss the specific individual collocates of each near-synonym. This is the focus of the next section, which zooms in on the top noun collocates of the three adjectives.

### 3.2 Collocational networks per period

Figures 3--6 plot the four collocational networks, one per time period, of the three near-synonyms, constructed on the basis of their most significant collocates (cf. Section 2). In P1 (1810--1860), *fragrant*, with 45 collocates, dominates over *perfumed* (8) and *scented* (3) in terms of their number of collocates (Figure 3). This is the case in all four time periods, which is not surprising given that *fragrant* is the most common adjective of the set. On the other hand, *scented* is the item with the lowest number of collocates and all of them are shared with one or both of the other adjectives. With respect to the type of entities that the adjectives collocate with, *fragrant* has a clear natural orientation, that is, it is commonly used to describe natural smells. In fact, out of its 45 collocates, a great majority is clearly used to refer to entities which emit a natural smell on their own, such as *bower*, *dew*, *fruit*, *garden*, *petal*, and *plant*. To get a more detailed picture of the collocates, we can resort again to USAS and classify the nouns into semantic categories. Of the 45 collocates of *fragrant*, 14 types, almost one third, belong to L3 (PLANTS; e.g. *plant*, *garden*, *honeysuckle*), 8 to the general category F (FOOD AND FARMING; e.g. *wine*, *spice*, *fruit*), and another 4 to W4 (WEATHER; e.g. *wind*, *dew*, *gale*). A large amount of the most prominent noun collocates of *fragrant* therefore belong to categories corresponding to the natural sense which were already discussed in detail in Section 3.1. Given the low number of instances, no clear conclusions can be extracted at this stage from the collocates of *perfumed* and *scented*.

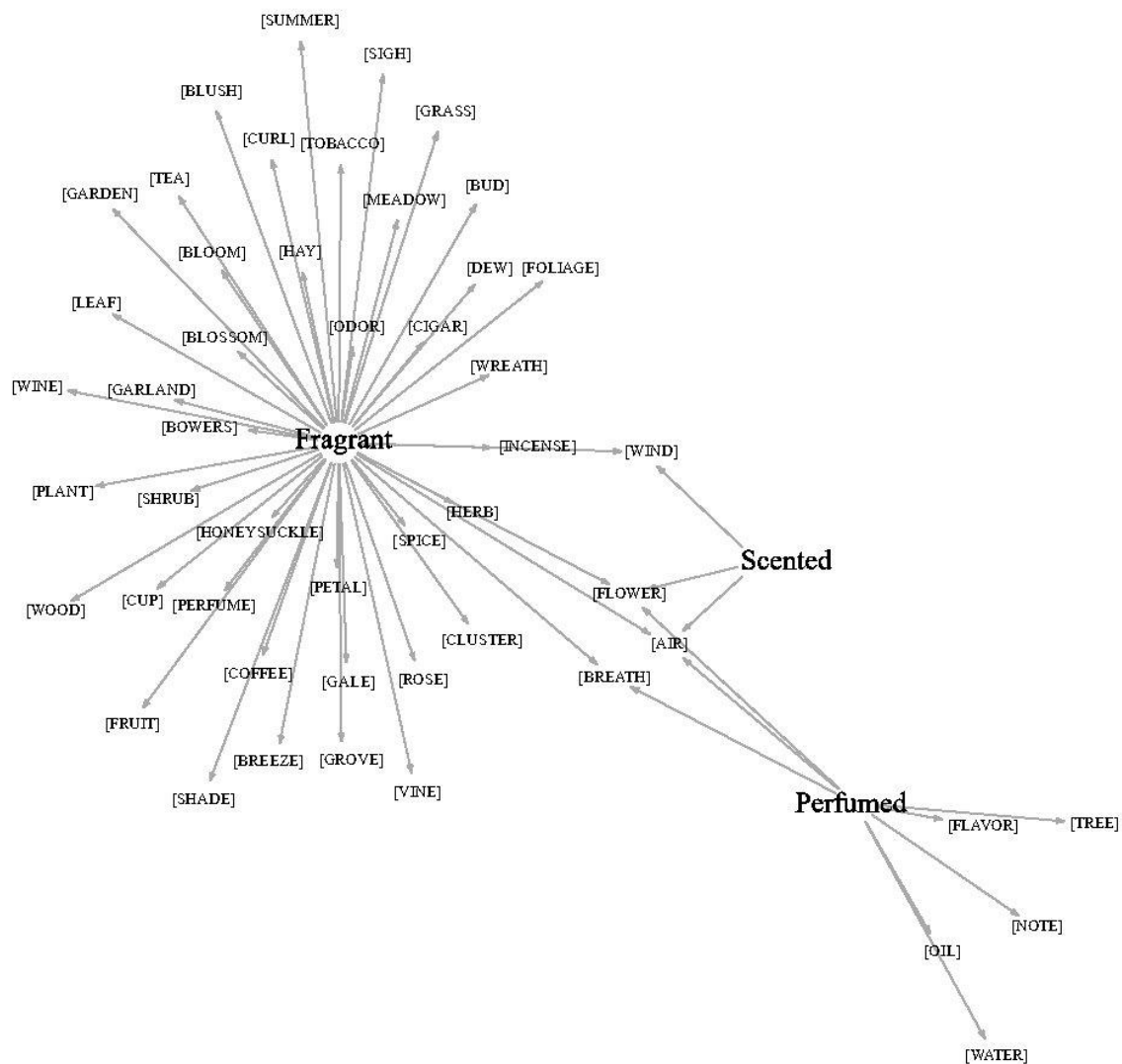


Figure 3: Collocational network of *fragrant*, *perfumed*, and *scented* in P1 (1810--1859)

In P2 (1860--1909), *fragrant* is still the dominant near-synonym of the set, with 74 collocates (Figure 4). Both *perfumed* and *scented* increase in number of collocates as compared to P1, with 16 and 10 collocates, respectively, but they are still much less productive than *fragrant*. The latter still exhibits a natural orientation, with 54 ‘natural’ collocates (e.g. *clover*, *foliage*, *herb*, *leaf*, and *vine*). Again, there is a strong correspondence between the noun collocates of *fragrant* and the categories L3 (23 types), F (8 types) and W4 (4 types). Nevertheless, the collocates of *fragrant* are now more varied, including also other types of nouns, such as several belonging to the category X3.5 (SENSORY: SMELL; e.g. *odor*, *scents*, and *smell*). Nouns belonging to the artificial sense, for instance, B4 (CLEANING AND PERSONAL CARE; 1 type) and B5 (CLOTHES AND PERSONAL BELONGINGS; 0 types), on the contrary, do not often co-occur with *fragrant*. Similarly, *scented* seems to display a preference for entities denoting a natural smell, such as *bloom*, *breeze*, *grass*, and *rose*, which are mainly nouns belonging to 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 the categories B1 (ANATOMY AND PHYSIOLOGY; 2 types) and B5 (3 types), but also to O1 (SUBSTANCES AND MATERIALS GENERALLY; 5 types).





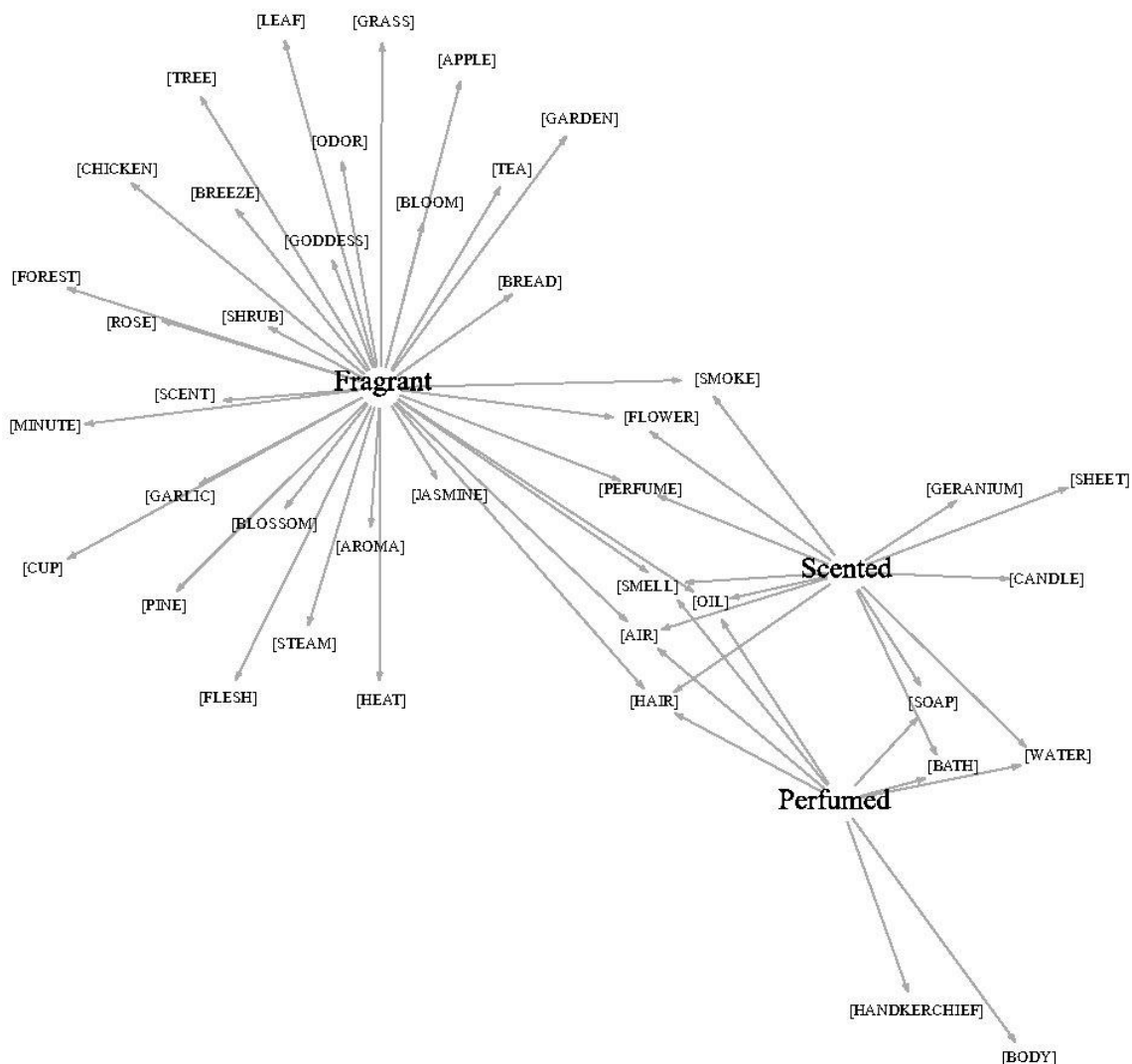


Figure 6: Collocational network of *fragrant*, *perfumed*, and *scented* in P4 (1960--2009)

The results of the collocational network analysis point to two main conclusions. First, as in the case of the frequency and collocational profiles of the near-synonyms, the most dramatic change takes place from P2 to P3, when the number of significant collocates of *fragrant* starts to decrease substantially, particularly in favor of *scented*. It is also in P3 and P4 that *scented* becomes more semantically neutral than *fragrant* and *perfumed* in the sense of not exhibiting a clear preference for either the natural or artificial senses, but being fairly common with both types of nouns. Its more neutral character is probably the reason why *scented* also occupies an intermediate position in P4 in the MDS in Figure 1. In fact, as we saw in Section 3.1, the value of *scented* on the vertical axis (Dimension 2) is almost 0 in P4, while *perfumed* and *fragrant* are located at opposite ends, the former with a positive value and the latter with a negative one.

#### 4 Conclusions and future research

The SVS analysis discussed in Section 3.1 demonstrated that, over time, *scented* goes from sharing more collocates with *perfumed* to behaving more similarly to *fragrant*. Furthermore, the findings of both the SVS and the collocational network analyses showed

that the similarities and differences between the synonyms have to do with the sense in which they are more commonly used: whereas *fragrant* and *perfumed* seem to progressively have become more specialized towards the natural and artificial senses, respectively, as shown by the nouns they typically cooccur with, *scented* is more neutral in this respect and becomes even more so at the end of the period. This explains its position in P4 (1960--2009) in the MDS plot (cf. Figure 1). In conclusion, there seems to be a scenario of competition between the near-synonyms, with *scented* gaining ground at the expense of both *perfumed* and *fragrant* over time. However, a process of attraction is also observable, whereby *scented* becomes more similar to *fragrant*, as well as simultaneously differentiating from *perfumed* by virtue of increasingly collocating with ‘natural’ collocates. The semantic specialization of *fragrant* and *perfumed*, discernible in the collocational networks, suggests that these two adjectives are also undergoing a process of differentiation, thus progressively moving towards opposite ends of the natural-artificial sense continuum.

In this contribution, we have opted for examining the collocational profiles of the three near-synonyms by averaging over all their exemplars, a so called type based SVS analysis, thus providing a description of the general semantics of the adjectives. However, the present dataset could be modified so as to be submitted to so called token based SVS modelling (e.g. Heylen, Speelman, and Geeraerts, 2012), an approach which represents the meaning of each individual occurrence of the target word(s). Type based SVS models such as the one employed here are commonly used to investigate meaning relations between words (e.g. synonymy, or antonymy), as was the original goal of the present study. Token based SVS analysis, on the other hand, is employed to explore the existence of different senses (i.e. polysemy) within the same target word/s. Creating a token based SVS model of the individual occurrences of *fragrant*, *perfumed*, and *scented* would thus be useful to further test whether the distances between the adjectives, as represented in the MDS plot of Figure 1, indeed correspond to them occupying different positions on the natural artificial sense continuum. Unfortunately, this is an issue that must be left for future research.

### **Acknowledgements**

For financial support, I am grateful to the following institutions: the Regional Government of Galicia (grant ED481A-2016/1687) and the European Regional Development Fund, the Spanish Ministry of Science, Innovation and Universities (grant FFI2017-86884-P). Thanks are also due to Iván Tamaredo for feedback on an earlier version of this paper. Lastly, I would like to express my sincere gratitude to two anonymous reviewers for their fruitful comments and thorough revision, as well as to the editors of the journal for their time and consideration.

### **References**

Archer, D., Wilson, A., and Rayson, P. (2002) Introduction to the USAS category system, *Benedict Project Report*: 1–37. <http://ucrel.lancs.ac.uk/usas/> (accessed 19 April 2020).

- Baker, P. (2017) *American and British English. Divided by a common language?* Cambridge: Cambridge University Press. Doi 10.1017/9781316105313
- Bolinger, D. (1977) *Meaning and form*. London: Longman.
- Brezina, V., McEnery, T., and Wattam, S. (2015) Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics* 20: 139--173. Doi 10.1075/ijcl.20.2.01bre
- Cambridge Dictionary*. (2019). <https://dictionary.cambridge.org/> [last accessed 23 November 2019].
- Collins online Unabridged English Dictionary*. (2012--) <https://www.collinsdictionary.com/> [last accessed 23 November 2019].
- Church, K. W., and Hanks P. (1990) Word association norms, mutual information, and lexicography. *Computational Linguistics* 16: 76--83. Doi 10.3115/981623.981633
- Church, K. W., Gale, W., Hanks, P., and Hindle, D. (1991) Using statistics in lexical analysis, in Zernik U. (ed.) *Lexical acquisition: Exploiting on-line resources to build a lexicon* 115--164. Hillsdale: Lawrence Erlbaum.
- Church, K. W., Gale, W., Hindle, D., and Rosamund M. (1994) Lexical Substitutability, in Levin, B., and Zampolli, A. (eds.) *Computational approaches to the lexicon* 153--177. Oxford and New York: Oxford University Press.
- Croft, W. (2000) *Explaining language change. An evolutionary approach*. Essex: Pearson Education.
- Cruse, A. D. (2000) *Meaning in language: An introduction to semantics and pragmatics*. Oxford: Oxford University Press.
- Csardi G., and Nepusz T. (2006) The igraph software package for complex network research. *InterJournal Complex Systems* 1695. <http://igraph.org>
- Davies, M. (2010--) *The Corpus of Historical American English (COHA): 400 million words, 1810--2009*. <https://corpus.byu.edu/coha/> [last accessed: 8 November 2019]
- De Smet. H., D'hoedt, F., Fonteyn, L., and Van Goetham, K. (2018) The changing functions of competing forms: Attraction and differentiation. *Cognitive Linguistics* 29: 197--234. Doi 10.1515/cog-2016-0025
- Desagulier, G. (2014) Visualizing distances in a set of near-synonyms: *Rather, quite, fairly, and pretty*, in Glynn, D., and Robinson, J. A. (eds.) *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* 145--178. Amsterdam and Philadelphia: John Benjamins. Doi 10.1075/hcp.43.06des
- Divjak, D. (2010) *Structuring the lexicon: A clustered model for near-synonymy*. Berlin and New York: Mouton de Gruyter. Doi 10.1515/9783110220599
- Divjak, D., and Gries, S. Th. (2006) Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2: 23--60. Doi 10.1515/CLLT.2006.002

- Divjak, D., and Gries, S. Th. (2008) Clusters in the mind? Converging evidence from near synonymy in Russian. *The Mental Lexicon* 3: 188--213. Doi 10.1075/ml.3.2.03div
- Firth, J. R. (1957) *Papers in linguistics, 1934-1951*. London and New York: Oxford University Press.
- Geeraerts, D. (1986) On necessary and sufficient conditions. *Journal of Semantics* 5: 275--291. Doi 10.1093/jos/5.4.275
- Gries, S. Th. (2001) A corpus-linguistic analysis of English *-ic* vs *-ical* adjectives. *ICAME Journal* 25: 65--108.
- Gries, S. Th. (2003) Testing the sub-test: An analysis of *-ic* and *-ical* adjectives. *International Journal of Corpus Linguistics* 8: 31--61. Doi 10.1075/ijcl.8.1.02gri
- Gries, S. Th. (2013) 50-something years of work on collocations: What is or should be next... . *International Journal of Corpus Linguistics* 18 (1): 137--165. Doi 10.1075/ijcl.18.1.09gri
- Heylen, K., Speelman, D., and Geeraerts, D. (2012) Looking at word meaning. An interactive visualization of semantic vector spaces for Dutch synsets, in Butt, M., Carpendale, S., Penn, G., Prokić, J., and Cysouw, M. (eds.) *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.
- Hilpert, M., and Correia Saavedra, D. (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. Doi 10.1515/cllt-2017-0009
- Jackson, H. (1988) *Words and their meanings*. London: Longman.
- Jansegers, M., and Gries, S. Th. (2017) Towards a dynamic Behavioral Profile: a diachronic study of polysemous *sentir* in Spanish. *Corpus Linguistics and Linguistic Theory*, Ahead of Print: 1--43. Doi 10.1515/cllt-2016-0080
- Jones, M. A. (1996) *Historia de Estados Unidos 1607–1992*. Translated by Carmen Martínez Gimeno. Madrid: Cátedra.
- Kjellmer, G. (2003) Synonymy and corpus work: On *almost* and *nearly*. *ICAME Journal* 27: 19--27.
- Levshina, N. (2015) *How to do linguistics with R*. Amsterdam and Philadelphia: John Benjamins. Doi 10.1075/z.195
- Lexico (2019). <https://www.lexico.com/en> [last accessed 23 November 2019]
- Longman Dictionary of Contemporary English. (2015--). <https://www.ldoceonline.com/> [last accessed 23 November 2019].

- Liu, D. (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: 56--87. Doi 10.1075/ijcl.15.1.03liu
- Liu, D. (2013) Salience and construal in the use of synonymy: A study of two sets of near-synonymous nouns. *Cognitive Linguistics* 24: 67--113. Doi 10.1515/cog-2013-0003
- Liu, D., and Espino, M. (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. Doi 10.1075/ijcl.17.2.03liu
- MacMillan Dictionary*. (2009--). <https://www.macmillandictionary.com/> [last accessed 23 November 2019].
- Merriam Webster Dictionary and Thesaurus*. (2019) <https://www.merriam-webster.com/> [last accessed 23 November 2019].
- Murphy, M. L. (2003). *Semantic Relations and the Lexicon. Antonymy, Synonymy, and Other Paradigms*. Cambridge: Cambridge University Press. Doi 10.1017/CBO9780511486494.002
- Newbury House Dictionary of American English*. (2019) <http://nhd.heinle.com/home.aspx> [last accessed 23 November 2019].
- Oxford English Dictionary*. 3<sup>rd</sup> edition (2012--). <http://www.oed.com/> [last accessed 23 November 2019].
- Peirsman, Y., Heylen, K. and Geeraerts D. (2008) Size matters. Tight and loose context definitions in English word space models”, in Baroni, M., Ewert, S., and Lessi A. (eds.) *Proceedings of the ESSLLI workshop on distributional lexical semantics: Bridging the gap between semantic theory and computational linguistics* 34--41. Hamburg: ESSLLI.
- Pettersson-Traba, D., (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. Doi 10.1515/cllt-2018-0025
- Primahadi-Wijaya-R., G., and Rajeg, I M. (2014) Visualising diachronic change in the collocational profiles of lexical near-synonyms, in Sudipa, I N., and Primahadi-Wijaya-R., G. (eds.) *Cahaya Bahasa: A Festschrift in honour of Prof. I Gusti Made Sutjaja* 247--258. Denpasar: Swasta Nulus.
- R Core Team. (2017) *R: A Language and Environment for Statistical Computing* (version 3.4.3). Vienna: R Foundation for Statistical Computing.
- Sahlgren, M. (2006) *The word-space model. Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. Ph.D. thesis. Stockholm: Stockholm University.
- Samuels, M. L. (1972) *Linguistic evolution with special reference to English*. London and New York: Cambridge University Press. Doi 10.1017/CBO9781139086707

Sinclair, J. (1966) Beginning the study of Lexis. in Bazell, C.E., Catford, J.C., Halliday, M. A. K., and Robins, R.H. (eds.) *In memory of J. R. Firth* 410--430. Harlow: Longman.

Stubbs, M. (2001) *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Taylor, J. R. (2003) Near Synonyms as co-extensive categories: 'High' and 'Tall' Revisited. *Language Sciences* 25: 263--284. Doi 10.1016/S0388-0001(02)00018-9

---

<sup>1</sup> In this paper, the terms 'synonym'/'synonymy' and 'near-synonym'/'near-synonymy' are used interchangeably given that most scholars agree on the fact that absolute synonyms are uncommon in language.

<sup>2</sup> The division of the corpus into four fifty year periods was undertaken due to the relatively low frequency of the near-synonyms, particularly *perfumed* and *scented*, which would have made the task of observing changes in their collocational preferences much more complicated if shorter time periods, such as decades or years, had been considered instead.

<sup>3</sup> I would like to thank Professor Stefanie Wulff for suggesting the use of collocational networks for the analysis of my data as it has indeed proved to be a useful resource.

<sup>4</sup> Brezina et al. (2015) and Baker (2017) make use of the software *GraphColl* to build collocational networks. However, in this paper, the networks were created by means of the *igraph* package (Csardi & Nepusz 2006) in *R* (R Core Team 2017). The reasons for this decision were twofold: (i) loading the corpus into the program required a lot of computation time due to its large size and (ii) *R* allows a greater flexibility in the customization of the visual parameters of the collocational networks.

<sup>5</sup> A better fit is achieved by a three dimensional MDS solution (stress = 0.16). However, for the purposes of the present paper, only a two dimensional solution will be discussed as it allows a much clearer visualization of similarities in collocational behavior between the near-synonyms.

<sup>6</sup> For more detailed information about the different classes and sub-classes distinguished, see <http://ucrel.lancs.ac.uk/usas/>.