Postprint version of accepted manuscript. The paper was published as: Pettersson-Traba, Daniela. 2022. Ongoing semantic change in a modernizing society: A look at some adjectives from the olfactory domain in the Corpus of Historical American English. *Corpora* 17(3): 389-421. DOI: 10.3366/cor.2022.0264.
Available at: https://www.euppublishing.com/doi/abs/10.3366/cor.2022.0264?role=tab

Ongoing semantic change in a modernizing society: A look at some adjectives from the olfactory domain in the *Corpus of Historical American English*

Daniela Pettersson-Traba
Complutense University of Madrid

As a diachronic corpus-based investigation into onomasiological variation, this study has two main objectives. First, the paper analyses the evolution of the concept SWEET-SMELLING as a whole, that is, as instantiated by the three near-synonymous adjectives, *fragrant*, *perfumed*, and *scented*, with a focus on language-external pressures for distributional changes. There seems to exist variation over time in the nouns that the concept typically collocates with, going from nouns referring to entities with a natural pleasant smell to entities with an artificial agreeable aroma. It is here argued that this change is motivated by the social and technological transformations experienced by American society after the First and Second Industrial Revolutions, a claim that finds preliminary empirical support in the distribution from 1820 to 2009 of a series of lexical indicators from the semantic domains CLEANING, COSMETICS, and TEXTILE & CLOTHING. Second, the distribution over time of the three adjectives is examined. The data point to a reorganization concerning the internal semantic structure of the synonym set, with *scented* gaining ground at the expense of *fragrant* and *perfumed* in several contexts of use. Additionally, the adjectives exhibit highly idiosyncratic collocational preferences, which go a long way towards explaining the alternation between them.

Keywords: Cognitive Semantics; distributional approach; near-synonymy; onomasiological variation; semantic change; extra-linguistic factors

1. Introduction

Synonymy is often defined as a relation of referential equivalence between words. However, this is a rather reductive definition since it equates synonymy with referential identity – two words that denote the same entity or event – but completely disregards

other semantic potential differences of a more sociolinguistic, stylistic, or contextual nature. Whereas many words could be classified as synonyms under the former (reductive) definition, finding examples of synonymy which share not only their core denotational meaning but are also identical in all other dimensions is extremely difficult, if not impossible (e.g., Croft 2000: 176). Otherwise synonymous words which differ in peripheral aspects, or in other dimensions of meaning such as connotation, style, and/or collocation are usually called near-synonyms and they are essential to the structure of lexical knowledge of any language (e.g., Cruse, 2000: 159–160; Liu, 2010).

Synonymy and near-synonymy have figured prominently in theories of lexical semantics, dating back to the structuralist linguistic school of the twentieth century. The structuralist approach to synonymy and other semantic relations (e.g., antonymy) focuses primarily on paradigmatic relations between words at the level of the system, disregarding issues connected to language use. In this framework, meanings of words are thought to be derived from the relations that they have with other words in the system (Lyons, 1963, 1977). Synonymous words, therefore, are those which share a connection of semantic equivalence in terms of referential identify. Later neo-structuralist approaches to semantic relations in general, and synonymy in particular, however, diverge somewhat from this emphasis on paradigms to focus on syntagmatic associations between linguistic units (e.g., Sinclair, 1966, 1991). The idea is that looking at the syntagmatic contexts in which words occur helps to identify their semantic properties. This distributional approach to lexical semantics differs substantially from structuralist frameworks in three fundamental ways. First, due to its focus on syntagmatic relations, it employs a usage-based corpus methodology instead of the system-based perspective of structuralist methods. Second, it highlights the importance of collocations to identify the meanings of words. Third, empirical and statistical analyses lie at the core of the distributional method instead of the more introspective technique of previous approaches. This is largely due to the emphasis on collocations, as identifying which words are collocates of another requires the use of sophisticated statistical procedures.

From a methodological viewpoint, therefore, the neo-structuralist distributional approach undoubtedly meant an improvement over structuralist theories. However, this approach is ultimately a method rather than a theory, and it is not always clear how the distributional patterns uncovered relate to theoretical issues in lexical semantics

(Geeraerts, 2010: 178). Recently, distributional methods have been adopted in the study of synonymy under the theoretical framework of Cognitive Semantics, which embeds the study of meaning in the study of human cognition in general. One area of research within this framework focuses on onomasiological variation from a usage-based perspective, that is, on modelling the choice between near-synonyms on the basis of both language-internal and language-external factors. To this purpose, this line of research employs a distributional corpus-based methodology, but it makes use of more advanced statistical techniques such as logistic regression or correspondence analysis, among others (Divjak and Gries, 2008; Speelman and Geeraerts, 2009; Levshina, Geeraerts, and Speelman, 2013; Liu, 2015). These studies, which have successfully uncovered subtle distinctions in meaning between near-synonyms, have approached the study of synonymy mainly from a synchronic perspective, thus somewhat overlooking the diachronic dimension (but cf. Primahadi-Wijaya-R and Rajeg, 2014; Pettersson-Traba, 2021).

Concepts with a long tradition in Cognitive Semantics, such as prototypicality and entrenchment, are often used to explain the findings of usage-based onomasiological studies, thus relating them to a theoretical framework on language and human cognition, in general, and lexical semantics, in particular (cf. Geeraerts, 1988; Geeraerts, Grondelaers, and Bakema, 1994; Soares da Silva, 2015). Nevertheless, finding the motivations underlying lexico-sematic changes is often not a straightforward endeavor. Semantic developments are in many cases subject to a wider range of factors than other types of changes, including both intra- and extralinguistic ones. Moreover, the latter type of factors play a major role in many cases (Durkin, 2009: 222–223). As such, non-linguistic history, particularly technological, socio-political, and cultural transformations, must often be resorted to in order to explain the occurrence of semantic changes. In fact, language-external pressures have been found in numerous occasions to lie at the root of change. Cases in point are the reorganization of kinship terms in English after the Norman Conquest (Kay and Allan, 2015: 140–141) and the restructuring of French names for the meals of the day after the sixteenth century due to changes in lifestyle (Blank, 1999: 73).

Recently, it has been shown that such extralinguistic changes at the cultural and social level can also be observed through the use of corpora. In particular, Baker (2017: 177) claims that the frequencies of words in specific 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, 2013; 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 periods. Therefore, these analyses do demonstrate that corpus data can be used as an indicator, if not as conclusive evidence, of cultural and social changes.

Against this background, this paper examines the diachronic evolution of three near-synonymous adjectives which instantiate the concept SWEET-SMELLING, namely *fragrant*, *perfumed*, and *scented* (cf. (1)–(3) below), in nineteenth- and twentieth-century American English.

(1) Amid the smell of love and squashed, **fragrant fruit**, Mersault realized then that the season was ending. (*COHA*, 1972)

(2) She swept to the door in her brilliant dress, her **perfumed laces**, her shining jewels (*COHA*, 1874)

(3) When Kilo buys **scented soap** she likes to have it really scented. (*COHA*, 1907)

These three near-synonyms were selected on the basis of a thorough examination of dictionaries and other reference materials. Other adjectives which share their definition with *fragrant*, *perfumed*, and *scented*, for example, *fragranced* and *sweet-smelling*, were excluded from the present study because of their low frequency in *COHA*, which made it impossible to include them in the statistical analyses conducted in Section 3.2.

The core denotational meaning of *fragrant*, *perfumed*, and *scented* is identical, as the definition in existing reference material in the three cases is that of 'having a pleasant and sweet smell' (e.g., *Merriam Webster*, *OED*, *Lexico*).[1] However, not all reference material

[1] The information regarding the senses of the near-synonymous adjectives provided here is based on the following dictionaries and thesauri: *American Heritage Dictionary of the English*

distinguish the same senses and subsenses for the three near-synonyms, which seems to point to some subtle distinctions of meaning between them: whereas, for instance, the *OED* only provides one sense for *fragrant*, namely '[e]mitting a sweet or pleasant odour' (s.v. *fragrant* adj.), as in the case of *fruit* in example (1), an additional sense is given for *perfumed* and *scented*. There seems to be a distinction in these two adjectives between entities which produce a natural pleasant smell on their own and those which have acquired an artificial agreeable aroma by being infused with a sweet-smelling substance, such as *laces* and *soap* in examples (2) and (3), respectively. This is reflected, for instance, in the definitions of *perfumed* provided by *MacMillan Dictionary*:

(i) 'pleasant to smell because of natural qualities' (*MacMillan Dictionary*, s.v. *perfumed* adj.)

(ii) 'pleasant to smell because perfume has been added or used' (*MacMillan Dictionary*, s.v. *perfumed* adj.)

In fact, in the *OED* the primary sense of *perfumed* is '[i]mpregnated with perfume; wearing perfume; scented with a (usually pleasant) odour' (s.v. *perfumed* adj. 1). In terms of their prototypicality or onomasiological salience, therefore, *perfumed* and *scented* are possibly more prototypical than *fragrant* to refer to artificial aromas, as several of the dictionaries list the 'artificial' sense in the entries of these two adjectives. *Fragrant*, in turn, seems to be more general, as the dictionaries do not specify whether it is used to denote natural or artificial smells. The following definition illustrates this point:

(iii) 'having a pleasant or sweet smell' (*Collins Dictionary*, s.v. *fragrant* adj.)

Moreover, none of the dictionaries consulted provide the 'artificial' sense for *fragrant*, although some do include examples of *fragrant* being used in this sense. Consider example (4) extracted from *Longman Dictionary of Contemporary English*, where *fragrant* modifies the noun phrase *soaps and candles*:

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

Language, *Cambridge Dictionary*, *Collins Dictionary*, *Lexico*, *Longman Dictionary of Contemporary English*, *MacMillan Dictionary*, *Merriam-Webster*, *Newbury House Dictionary of American English*, and the *OED*.

The fact that *perfumed* and *scented* are probably more prototypical to modify entities which need to be impregnated with a synthetic substance to exhibit a smell while *fragrant* is more general may be a result of the origins of the adjectives, as shown in the *OED*. *Fragrant*, which entered English indirectly via the French adjective *fragrant* in the 16th century, dates back to Latin *frāgrant-em,* the present participle of *frāgrāre* 'to smell sweetly'. However, despite ultimately deriving from a participial form, when *fragrant* became an English word, its participial nature had already been lost. On the other hand, both *perfumed* and *scented* are derivatives formed by the suffix *-ed* and their respective nominal and verbal forms *perfume* and *scent*, which were also borrowed indirectly from Latin via French. Contrary to *fragrant*, *perfumed*, and *scented* were already formed in English by derivation, thus retaining their participial shape, particularly due to the presence of the English suffix *-ed*. It is, therefore, probably correct to assume that the *perfumed* and *scented* still retain part of their verbal semantic functions, namely, denoting a process or action (Biber et al., 1999: 63), in this case, '[t]o impregnate with a (usually pleasant) odour; to impart a (sweet) smell to […]' (*OED* s.v. *perfume* v. 2). Consequently, they should be more salient in the artificial sense. *Fragrant*, on the other hand, could be expected to be more prototypical in natural contexts given that the artificial end of the spectrum seems to be occupied by *perfumed* and *scented*. Be that as it may, it should be noted that the preferences of the adjectives for one sense or the other are probably not of a categorical but of a probabilistic nature.

Additionally, all three adjectives exhibit a figurative sense, as in example (5), where *fragrant* modifies a noun referring to an abstract entity (i.e. *memory*) which cannot emit a smell and is thus used to mean just sweet or pleasant:

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

This paper is a continuation of the line of research initiated in Pettersson-Traba (2021), where the distribution over the time-span 1850–2009 of the near-synonymous adjectives *fragrant*, *perfumed*, *scented*, and *sweet-smelling* in attributive function was examined in American English, paying special attention to their noun collocates, which were classified into nine semantic categories on the basis of the categorization of the *Historical Thesaurus of the Oxford English Dictionary* (*HTOED*): BODY, CLEANING, EARTH,

AESTHETICS, FOOD & DRINK, MATTER, PLANTS & FLOWERS, SENSATION, and TEXTILE & CLOTHING. The results showed that the concept designated by the four adjectives (i.e. SWEET-SMELLING) underwent significant semantic developments throughout the period analyzed, going from being used mainly to denote entities that emit a natural pleasant smell (e.g. blossoms and spices) to modifying objects that are impregnated with synthetic substances with an artificial pleasant aroma (e.g. soaps and candles). Moreover, the results demonstrated that *fragrant* and *perfumed*, which were the most common synonyms at the beginning of the period analyzed, progressively lost ground in favor of *scented*. However, the analysis in Pettersson-Traba (2021) was still rather preliminary, which is the reason why this work is here revised and improved. First, in Pettersson-Traba (2021), only a subset of texts in *COHA* was analyzed due to the unbalanced nature of this corpus and the statistical test employed, namely Hierarchical Configural Frequency Analysis (HCFA). The later decades in *COHA* contain significantly more words of running text than the earlier ones and, since HFCA works by comparing observed and expected frequencies across cells in a contingency table, it was necessary to divide the time span into comparable periods in terms of number of words. For this reason, the analysis in Pettersson-Traba (2021) was based on approximately 16.5 million words from each decade, amounting to a total 263,901,899 words.[2] This subcorpus is only about half the size of *COHA*, which contains slightly over 400 million words. As a consequence, the analysis in Pettersson-Traba (2021) was based on a limited dataset of just 1,666 instances of the four near-synonymous adjectives. On the contrary, the present study, despite the exclusion of *sweet-smelling*, considers a much larger sample of 5,313 instances of the adjectives. This increase results from the fact that the whole corpus has been here examined as the statistical method employed does not impose the same restrictions on the data as HCFA and that all instances of the adjectives, not only those in attributive function, have been analyzed, including predicative and postpositive uses.

Second, a more sophisticated statistical analysis (i.e., mixed-effects logistic regression) is here conducted. While HCFA is an extension of the chi-square test that makes it possible to analyze more than two categorical variables, logistic regression is specifically geared towards assessing and isolating the independent effect of each of a set

[2] The decades 1810–1840 in *COHA* were excluded given that they contain fewer than 16.5 million words.

of predictors, as well as their possible interactions, thus making it more suitable for an analysis as the present one which examines the influence of five variables on the choice between the near-synonyms. In fact, while Pettersson-Traba (2021) considered only the variables 'semantic category' and 'period', two additional variables are analyzed here, namely 'sense' and 'text-type' (cf. Section 2.1). Furthermore, the type of statistical test used (i.e. mixed-effects logistic regression) allows to control for random effects such as lexical items, in this case, the particular collocates of the adjectives, a variable which has been shown to exert a great influence on the choice between near-synonyms (e.g. Gries 2001; Liu, 2010; Desagulier, 2014).

In Pettersson-Traba (2021), it was suggested that the changes in the distribution of the concept SWEET-SMELLING were the outcome of extralinguistic factors, namely the socio-technological transformations experienced by American society, especially at the turn of the nineteenth to the twentieth century, which led to an ever-increasing need to allude to artificially scented soaps and candles rather than naturally fragrant blossoms and spices. However, Pettersson-Traba (2021) offered this tentative explanation of the results obtained merely as an avenue for future research, without further delving into the matter. The present study preliminarily explores this hypothesis to see whether it could in fact constitute a possible motivation underlying the semantic changes identified.

In light of the issues discussed in this section, this study addresses the following research questions. Focusing first on the concept SWEET-SMELLING as a whole, that is, as designated by the three near-synonyms, it is possible for it to exhibit diachronic fluctuations regarding the type of nouns it modifies. In other words, the type of entities that are commonly qualified as having a pleasant sweet smell may vary diachronically. If so, the changes may be due to the aforementioned socio-technological transformations in American society, a hypothesis which is preliminarily tested by ascertaining whether such transformations can also be observed in *COHA* following the methodology in Baker (2017: Chapter 7), among others. Second, dictionaries and the results in Pettersson-Traba (2021) seem to suggest the existence of a division of semantic labor between the three near-synonyms in terms of a natural-artificial division, with *perfumed* and *scented* being prototypically artificial and *fragrant* being more general but possibly more prototypically natural. If such a semantic division of labor between the adjectives holds, it is certainly

possible for it not to remain stable over time, but rather to change as a result of diachronic reorganizations within the semantic space occupied by this synonym set.

The rest of the paper is structured as follows. Section 2 explains the data retrieval process and the methodology. Section 3 presents and discusses the results obtained: Section 3.1 focuses on the first research question, namely the diachronic evolution of the concept SWEET-SMELLING, while Section 3.2 concentrates on the second question, that is, on the division of semantic labor between the near-synonyms (i.e., natural vs. artificial) and the possible changes to the internal structure of the synonym set. Finally, Section 4 concludes the paper with some final remarks and suggestions for future research.

2. Data and method

The data were retrieved from *COHA* (Davies, 2010–), which contains approximately 400 million words of American English from the 1810s to the 2000s.[3] Four different text-types are represented in *COHA*, namely fictional texts, popular magazines, newspapers, and non-fiction books. Instances of *fragrant*, *perfumed*, and *scented* were retrieved from *COHA* by means of the following queries: fragrant_j*, perfumed_j* scented_j*, where the POS-tag j* indicates that only adjectives are searched for. 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, given that some such instances are in fact adjectival rather than verbal uses of the items, as in the case of example (6):

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

As is well-known, the distinction between past participles and participial adjectives is not always straightforward. Therefore, a decision had to be made regarding which cases counted as adjectival uses and which counted as verbal ones. A number of criteria have been proposed in the literature to classify these ambiguous examples as being located closer to the adjectival end or closer to the verbal end of the continuum (e.g. Quirk et al. 1985: Section 7.16). Instances retrieved with the strings perfumed_v* and scented_v* where included in the database if they satisfied any of the following criteria:

[3] This information is based on the pre-2021 version of the corpus, not the updated version, released in 2021.

(i) the focus is more on the state resulting from an action rather than on the action itself,
(ii) the items can occur with lexical copular verbs such as *appear*, *feel*, or *seem*,
(iii) the items can be modified by adjectival intensifiers such as *very*,
(iv) the items can be inflected for degree, either comparative or superlative, and
(v) the items can be coordinated with true adjectives.

The resulting database was annotated for the variables 'synonym', 'collocate', 'semantic category', 'sense', 'period', and 'text-type' (cf. Section 2.1), and subsequently submitted to different statistical analyses (cf. Section 2.2) to answer the research questions raised previously.

2.1. Variables

The first variable for which the instances were annotated was the synonym that occurred in each of them. This variable contains three levels, namely *fragrant*, *perfumed*, and *scented*. The rest of the intra-linguistic variables, 'collocate', 'semantic category', and 'sense', pertain to the nouns which the adjectives modify. As shown by previous studies on adjectives (e.g., Geeraerts, 1986; Gries, 2001; Liu, 2010), the fine-grained aspects of their meaning can be uncovered by analyzing the noun types they modify.

The variable 'collocate' contains the specific semantically modified nouns of the adjectives and allows us to control for lexical effects. These nouns were then grouped into two more schematic semantic classes, namely 'semantic category' and 'sense'. Nouns were sorted into semantic categories on the basis of the semantic classifications provided by the *HTOED* and the *UCREL Semantic Analysis System* (*USAS*; Archer, Wilson & Rayson 2002; Rayson et. al 2004). *USAS* contains twenty-one major sematic classes, which are then further divided into more specific subclasses. 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. However, this automatic process is not devoid of problems, as the sense of a given word is not always correctly identified, and therefore it

is often necessary to manually verify the semantic classification 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 word senses in the *OED* being categorized into semantic domains which are hierarchically ordered (Kay, 2012). At the most schematic level, three divisions are made between THE EXTERNAL WORLD, THE MIND, and SOCIETY. These three general levels are then sub-divided into specific classifications, which in turn display even more specific subclasses. The classes are ordered chronologically, that is, words in a particular class are organized by date of first attestation. Therefore, even obsolete or rare senses are included in the *HTOED*, which is important for a study of a diachronic nature such as the present one. Both sources, despite their differences, most often categorize words in similar semantic classes, albeit sometimes using different labels. For instance, both *USAS* and the *HTOED* include semantic categories for body parts, people, plants, time, and weather phenomena, among others.

The classification process resulted in twelve categories, which are displayed in Table 1 together with some prototypical examples of nouns included in each of them to provide a general picture of what types of nouns each category contains.

Table 1: Semantic categorization of head nouns

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

It is worth noting here that the classification provided is not the only possible one, and it might not be the most appropriate one if other concepts or lexical items are examined. However, this is also the case of previous classifications, such as that present in the *HTOED* (Allan 2015: 85). Additionally, a complication of categorization processes such

as this one that needs to be acknowledged is the fact that the limits between categories are not clear-cut, with some items being located at the periphery of more than one category and whose category membership is therefore not straightforward. This is in line with the non-discrete nature of semantic categories demonstrated by prototype-theory (e.g. Geeraerts, 1997: 11). For instance, in this dataset, nouns such as *berry* and *herb* can be said to lie at the border between the categories FOOD & DRINK and PLANTS & FLOWERS. Depending on the context of use, such nouns can be conceptualized either as parts of plants or as food. Similarly, the distinction between CLEANING and COSMETICS is not without problems. Consider example (7):

(7) **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 cosmeti**c. (*COHA*, 1988)

The noun *powder* is always conceptualized as a cleansing agent in the present dataset, thus belonging to the category CLEANING. However, (7) 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 cases analyzed *powder* is clearly a CLEANING noun and that the context in examples such as (7) does not help to disambiguate, *powder* has always been classified as a CLEANING noun. Something similar occurs with the noun *deodorant*, which can also be said to lie at the periphery of these two classes.

Besides being classified according to their semantic category, the instances were also annotated for whether the modified nouns referred to an entity that emits a natural pleasant smell on its own or to an entity which has a sweet smell by virtue of having been impregnated by an artificial substance. The levels of this variable, labelled ‘sense’, thus include the natural and artificial senses, as well as the figurative one mentioned in Section 1. These levels partly coincide with the semantic categorization of modified nouns: all nouns in the categories EARTH, ATMOSPHERE, AND WEATHER, FOOD & DRINK, and PLANTS & FLOWERS denote entities with a natural smell, while nouns in the categories COSMETICS, CLEANING, and TEXTILE & CLOTHING refer to artificial aromas. Similarly, the adjectives, when modifying ABSTRACT nouns, are always used in the figurative sense. The five

remaining categories, that is, BODY & PEOPLE, OBJECT, SENSATION, SPACE, and SUBSTANCE & MATERIAL, include nouns which can denote either artificial or a natural smells depending on the context in which they are used, as well as so-called indeterminate instances, that is, those that could not be assigned to any of the aforementioned senses, given that their interpretation was ambiguous between the natural and artificial readings.[4] Consider in this respect example (8):

(8) He turned and looked at her: the luminous eyes between their thick black lashes, the straight little nose, the voluptuous red mouth above the cleft chin, the transparent rose of her cheeks, the tumbled bronze of her **fragrant hair**, and the blue veins and white curves of her firm full breasts. (*COHA*, 1954)

Here, 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 hair product. Cases such as (7) were therefore categorized as indeterminate.

Finally, the database was annotated for two extra-linguistic variables. The first signals the time period from which the instances were extracted. The temporal span covered by *COHA* was divided into four fifty-year periods: 1810–1859 (Period 1 or P1), 1860–1909 (P2), 1910–1959 (P3), and 1960–2009 (P4). The reason for dividing the corpus in this manner was that the relatively low frequency of the near-synonyms examined made it difficult to notice changes in their distributional behavior in shorter time periods such as decades or years. The second variable of this type is 'text-type', which refers to the genre in *COHA* to which each example belong, namely fiction, non-fiction, magazine, and newspapers. However, due to the low number of instances of the adjectives in newspapers and since *COHA* does not contain texts from this genre until the 1860s, magazines and newspapers were merged into one single level called periodicals.

2.2. Statistical analysis

[4] These uses of the adjectives, although ambiguous, have not been excluded because omitting them would skew the results of the semantic categories as many of the instances of the adjectives modifying nouns pertaining to BODY & PEOPLE, OBJECT, SENSATION, SPACE, and SUBSTANCE & MATERIAL would have to be removed from the study.

Two different sets of statistical analyses were conducted. In order to address the first research question, namely ascertaining whether the concept SWEET-SMELLING changes throughout the time-span considered, the frequencies of the three near-synonyms taken together were examined, in the four periods, across senses and semantic categories. The statistical significance of any diachronic changes was subsequently assessed by means of chi-squared tests of independence (Levshina, 2015: 210–213).[5] Chi-squared tests are conducted using the function `chisq.test` to determine if there is an association between two 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.

The second research question concentrates on the natural-artificial divide and the possible variation in the relationship between the adjectives over time. These are more complex issues that require more complex analyses so, in this case, the data were examined by means of logistic regressions. Logistic regression analysis is a statistical test that aims at predicting the probability of occurrence of the levels of a dependent variable on the basis of one or more independent variables, also called predictors (Levshina, 2015: 277–289). Two separate models were computed using the function `polytomous` in the `polytomous` package, in both cases with 'synonym' functioning as dependent variable: Model A included 'sense', 'period', and 'text-type' as predictors of 'synonym', while Model B included 'semantic category', 'period', and 'text-type'. The reason why two independent analyses were conducted was that, as mentioned in Section 2.1, the predictors 'sense' and 'semantic category' are strongly correlated, since there is a correspondence between many of their levels. Correlated predictors are problematic if added to the same model (Levshina, 2015: 272) so, for the purposes of this paper, 'sense' and 'semantic category' were kept apart.

To control for lexical effects, mixed-effects regression models were also conducted, as they permit the inclusion of both fixed and random effects (Baayen 2008: 241–242; Tagliamonte & Baayen 2012: 143). By including 'collocate' as a random effect, it was possible to identify particular lexical items that show a preference for one of the near-

[5] All statistical procedures conducted were computed in *R* (R Core Team, 2017).

synonymous adjectives, thus accounting for their idiosyncratic collocational preferences. The same heuristic was used in all regression models, 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).

3. Results and discussion

A total of 5,313 instances of *fragrant*, *perfumed*, and *scented* were retrieved from *COHA*. Out of these, 3,395 (63.90 percent), 977 (18.39 percent), and 941 (17.71 percent) are cases of *fragrant*, *perfumed*, and *scented*, respectively, which indicates that *fragrant* clearly surpasses the other two adjectives in terms of their overall frequency. The distribution of the near-synonyms across the four periods distinguished, which is statistically significant as ascertained by a chi-squared test of independence (χ^2 = 201.36, d.f. = 6, p > 0.001), is shown in Figure 1.[6] Whereas *fragrant* clearly decreases in frequency over time – it accounts for 72.69 percent of the total number of instances in P1 but its percentage drops to 52.83 percent in P4 – *scented* increases substantially, going from 9.86 percent in P1 to 27.04 percent in P4. In turn, *perfumed* remains rather stable. Therefore, although *fragrant* is the most frequent synonym in all four periods, it progressively loses ground in favor of *scented*.

[6] All figures in the paper were generated using the `ggplot` function in the `ggplot2` package.

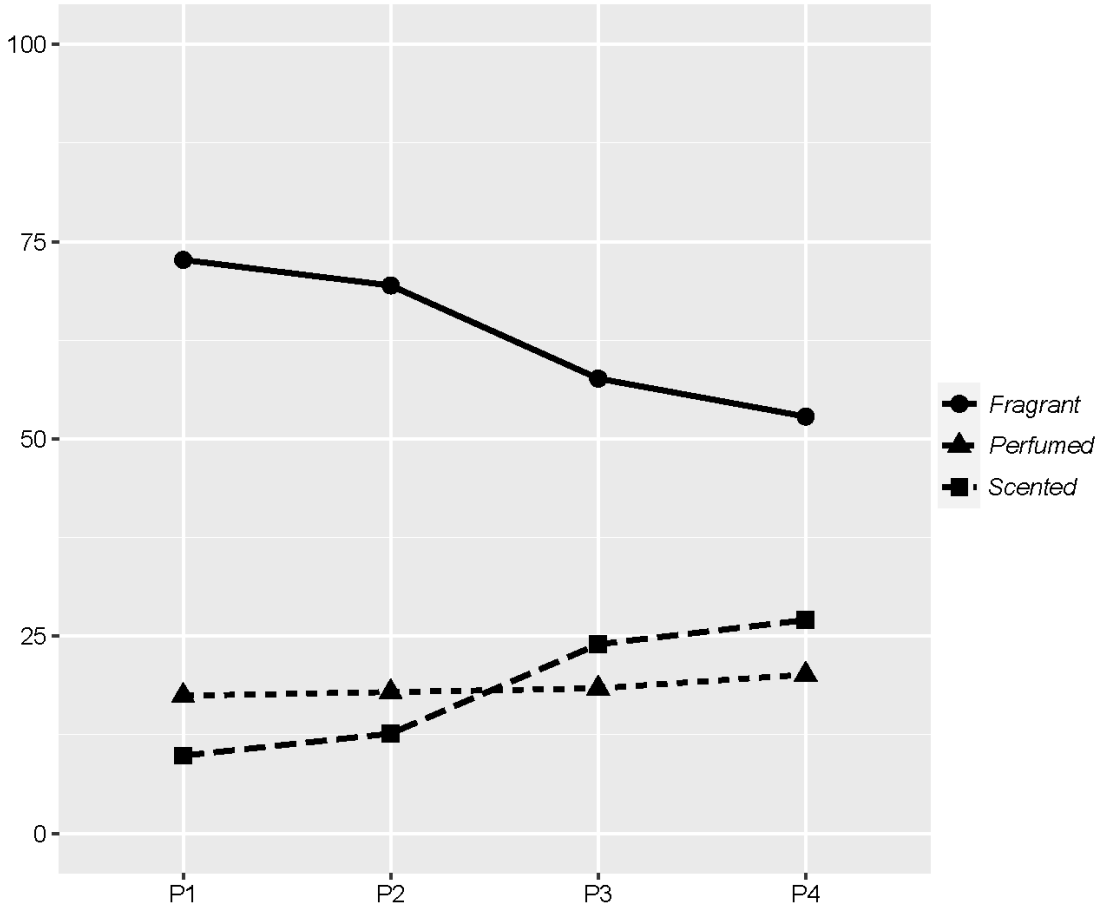


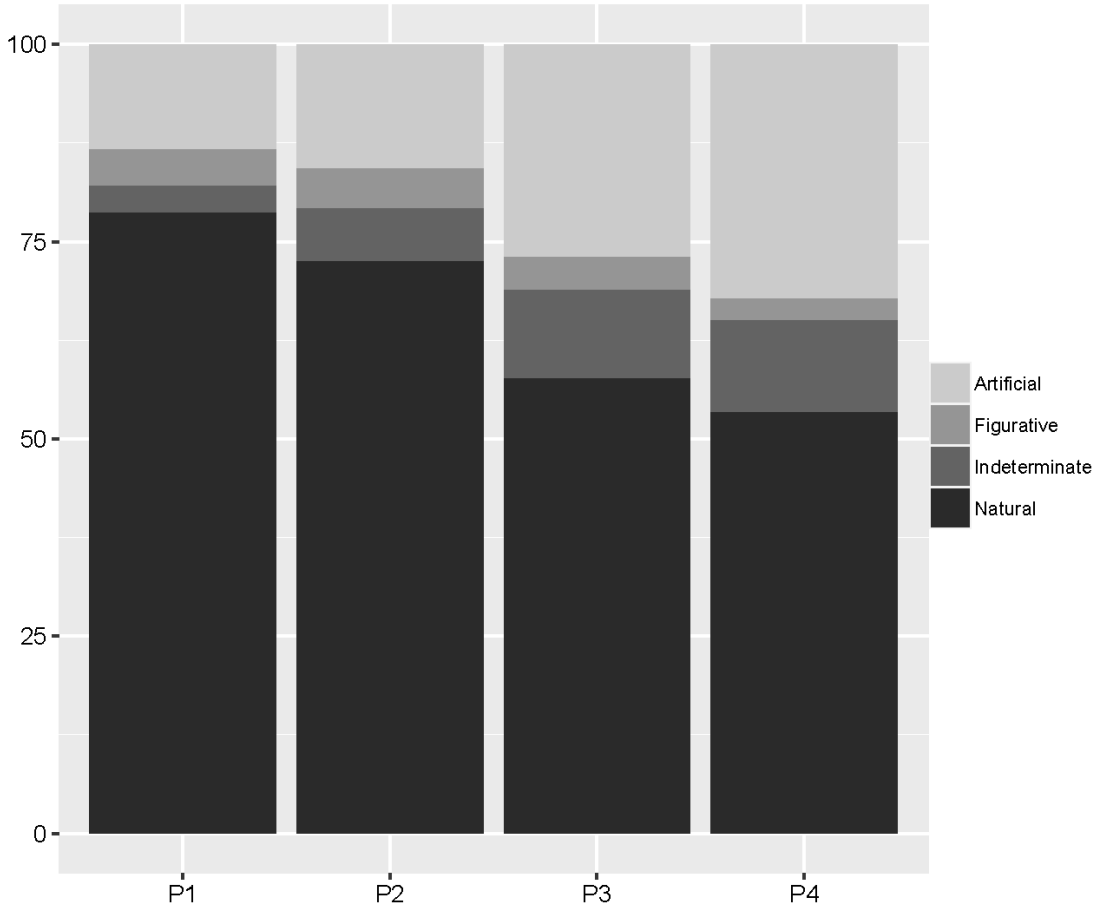
Figure 1: Distribution of the adjectives across time in percentages

In the rest of this section, two sets of results are presented. First, Section 3.1 describes the distribution of the concept SWEET-SMELLING across senses and semantic categories over time. Moreover, a hypothesis about the influence of social change in the evolution of the concept is put forward and preliminarily tested. Second, Section 3.2 focuses on the internal structure of the synonym set, that is, on changes in the co-occurrence patterns of the three adjectives from P1 to P4.

3.1. Evolution of the concept SWEET-SMELLING

Figure 2 displays the distribution of the concept across senses over the time-span analyzed. In P1, the near-synonyms are much more commonly used to modify nouns denoting entities which emanate a natural smell, with a relative frequency of occurrence of 78.66 percent, that is, 65.28 percent higher than in the artificial sense (13.38 percent). However, over time, the percentage of natural uses progressively decreases in favor of

the artificial sense, which increases to 32.26 percent in P4, only 21.11 percent lower than the relative frequency of the near-synonyms in the natural sense. Crucially, both the decrease of natural uses and the increase of artificial ones are particularly pronounced from P2 to P3, that is, around the transition from the nineteenth to the twentieth century. Similarly to the artificial sense, indeterminate uses of the synonyms also increase from P1 (3.35 percent) to P4 (11.59 percent).[7] Finally, the figurative sense is not very common and the uses of the concept in this sense remain rather stable from P1 to P3, ranging between 4 and 5 percent, with a decrease only in P4 (2.79 percent) A chi-squared test of independence indicates that the distribution of the concept SWEET-SMELLING across senses and periods shown in Figure 2 is statistically significant ($\chi^2 = 293.71$, d.f. = 9, $p >$ 0.001).



[7] It should be noted that indeterminate is not really a sense but a category including those instances which could not be classified as natural or artificial due to their ambiguity.

Figure 2: Evolution of the concept SWEET-SMELLING across senses

A more fine-grained picture of the evolution of the concept can be obtained by examining the changes in its distribution across the twelve semantic categories of modified nouns (cf. Figure 3), a distribution which is statistically significant (χ^2 = 610.28, d.f. = 33, p > 0.001). On the one hand, the frequency of the concept decreases in four semantic categories, namely ABSTRACT, EARTH, ATMOSPHERE, AND WEATHER, PLANTS & FLOWERS, and SENSATION. All examples in the category ABSTRACT coincide with the figurative sense. EARTH, ATMOSPHERE, AND WEATHER and PLANTS & FLOWERS nouns always denote entities which emanate a natural smell, so the decrease in the frequency of the concept SWEET-SMELLING in the natural sense is specifically located in these two semantic categories, especially in PLANTS & FLOWERS, which loses 23.70 percentual points from P1 to P4. Moreover, most instances of the concept SENSATION nouns also correspond to the natural sense (almost 80 percent), as they are of the type illustrated in example (9), where the aroma refers to the scent exhaled by the *bowl of gardenias*:

(9) A large bowl of gardenias sits on the table in the center of the foyer, their **perfumed aroma** trying unsuccessfully to mask the musty older-building smell. (*COHA*, 2010)

On the other hand, seven semantic categories exhibit an upward tendency in that, over time, they become more frequently modified by either *fragrant*, *perfumed*, or *scented*; these are BODY & PEOPLE, CLEANING, COSMETICS, FOOD & DRINK, OBJECT, SPACE, and TEXTILE & CLOTHING. Note that it is precisely with COSMETICS, CLEANING, and TEXTILE & CLOTHING nouns that the adjectives are always used to denote entities with an artificial smell, so these results are in line with the distribution of the concept across senses in Figure 2. Despite the decrease in the frequency of SWEET-SMELLING in the natural sense, the concept increases in four semantic categories that include 'natural' nouns, namely BODY & PEOPLE, FOOD & DRINK, OBJECT, and SPACE. However, OBJECT could also be considered a prototypically 'artificial' category, given that more than 70 percent of the instances of the concept with nouns belonging to this class are used in this sense. In the case of BODY & PEOPLE, it is the semantic category which displays the highest degree of ambiguity, as it contains an extremely large amount of indeterminate cases (50.1 percent; cf. example (8) above), but the percentage of uses in the artificial sense is still

substantially greater than in the natural one (40.37 vs. 8.48 percent). It is conceivable that the upward tendency of BODY & PEOPLE is related to the growing use of lotions, perfumes, cleansing agents, and other artificial products that are often applied to the body. In fact, this would also go a long way towards explaining the increase in both CLEANING and COSMETICS.

The two remaining categories that increase over time are FOOD & DRINK and SPACE. FOOD & DRINK is, as mentioned above (cf. Section 2.1), a prototypically natural category, as it solely includes instances of the near-synonyms used in the natural sense. Nevertheless, it could be argued that food and drinks have progressively become less natural over time with the growing addition of artificial products such as food coloring, sweeteners, and preservatives. Finally, SPACE contains a rather large number of instances in both the natural (54.25 percent) and indeterminate (29.02 percent) senses. However, the increase in the use of the concept in this category is attested mainly in the indeterminate uses, but not in the natural ones.

As in the case of the distribution across senses shown in Figure 2, the changes in the frequency of the concept across semantic categories seem to be particularly pronounced from P2 to P3, that is, in the transition from the nineteenth to the twentieth century, especially in BODY & PEOPLE, CLEANING, COSMETICS, PLANTS & FLOWERS, and SPACE.

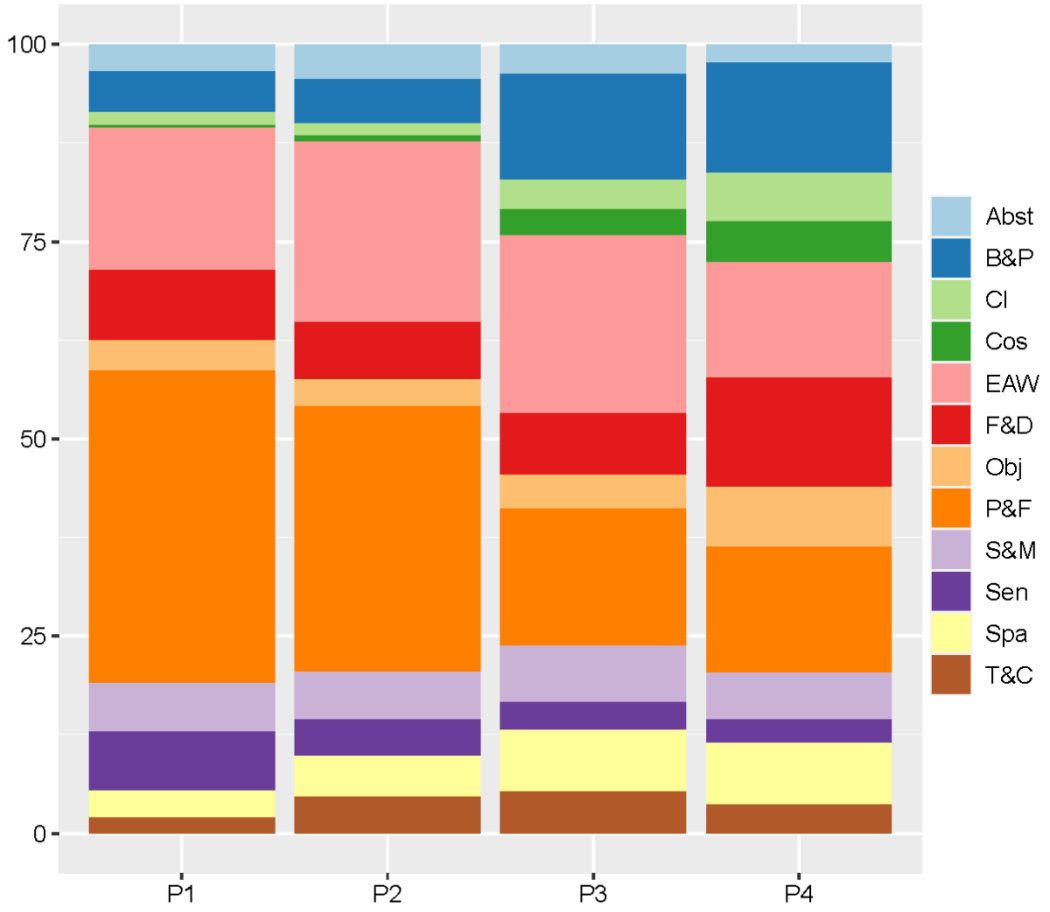


Figure 3: Evolution of the concept SWEET-SMELLING across semantic categories

A possible explanation for the rise in the use of the concept SWEET-SMELLING in the artificial sense at the expense of the natural one may lie in the transformation that American society experienced throughout the nineteenth and twentieth centuries as a result of the First and Second Industrial Revolutions, going from a primarily pre-industrial economy to the world's major industrial and commercial power (Jones, 1996: Chapter 16; Baker, 2017: 193–4). First, there was an increase in the production of goods with artificial smells, such as cosmetics and cleansing agents, which inevitably led to a growing availability of these products. In the second part of the nineteenth century, advances in industrial manufacturing and, particularly, chemistry facilitated the production of cosmetics and, by the 1920s, the mass production of these goods became much cheaper, which, together with the fall of traditional Victorian standards of beauty and marketing campaigns, enabled more and more people to access cosmetics and wear them in public (Jones, 1996: 308–9; 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 types of products more important in people's daily lives. In the period 1880–1914, the number of newspapers and magazines increased substantially in the US and, consequently, the percentage of the American population having access to this type of publications – and the advertisements they featured – also grew (Johnson, 1997: 397). The evolution of the concept SWEET-SMELLING from 1810 to 2009 could thus be a reflection of extra-linguistic factors, namely socio-economic changes such as industrialization and mass production that took place in the same period and led to an ever-increasing need to allude to *artificially* scented lotions and candles rather than *naturally* fragrant plants.

The influence of socio-economic changes on lexical choice in American English mentioned here can be ascertained by analyzing the distribution over time of lexical items related to the prototypically 'artificial' semantic domains CLEANING, COSMETICS, and TEXTILE & CLOTHING. Many such items refer to products which have probably been largely affected by the aforementioned modernization and industrialization processes. Moreover, a large number of nouns in these three domains are commonly modified by the synonymous adjectives under analysis. If such lexical items exhibit an overall increase in frequency over time, and not just when modified by the concept SWEET-SMELLING this could suggest that the transformation of American society from a pre-industrial to a fully industrialized nation could indeed be one of the underlying causes of the diachronic developments of the concept SWEET-SMELLING discovered here, that is, its rise in the artificial sense at the expense of the natural one. This idea is in line with Baker's (2017: 177) aforementioned claim (cf. Section 1) that the frequencies of words in particular semantic domains can be used to identify conceptual changes in the language, which are often directly associated to changes in the experiences and extra-linguistic reality of language users. To this purpose, searches were made in *COHA* of the lemmas of the most relevant modified nouns of the adjectives in the semantic domains CLEANING, COSMETICS, and TEXTILE & CLOTHING. Their normalized frequencies (NFs) per million words were examined across decades instead of the four fifty-year periods distinguished before to get a more detailed picture of the evolution of these words.

Fifteen nouns were chosen pertaining to each domain, as follows:

- CLEANING: *bath, bathtub, cleaning, cleanliness, deodorant, detergent, disinfectant, gel, handkerchief, laundry, shampoo, soap, sponge, suds, tissue.*
- COSMETICS: *body oil, brush, cologne, cosmetics, face cream, hairspray, haircut, lipstick, lotion, make-up, manicure, mascara, perfume, razor, shimmer.*
- TEXTILE & CLOTHING: *blouse, cambric, clothes, cushion, dress, flannel, garment, glove, lace, linen, sheet, shirt, silk, skirt, textile.*

The NFs per decade of these lexical items were extracted from *COHA*, and a mean NF per semantic domain was calculated. An effort was made to avoid highly polysemous words in which the various senses belong to different semantic domains. This is the case, for instance, of the noun *oil*, which could be included in the domain of COSMETICS with the meaning 'ointment or lotion' (as in *body oil*), but also in that of SUBSTANCE & MATERIAL (*crude oil*) or FOOD & DRINK (*olive* or *sunflower oil*). Baker (2017: Chapter 7) examines the changes in frequency of whole semantic domains across varieties of English and periods. This, however, was not feasible in the present study due to *COHA* not being semantically tagged and the fact that its large size would make it extremely difficult and time consuming to add semantic annotations to it, which are the reasons why we restricted our analysis to fifteen words per domain. In addition, for the purposes of the present paper it is probably more enlightening to focus exclusively on lexical items which in fact co-occur with the adjectives.

Figure 4 shows the mean NFs of the three domains across the decades distinguished in *COHA*.[8] All three domains increase steadily in frequency over time, even though TEXTILE & CLOTHING is much more frequent overall than the other two domains. Interestingly, this increase seems to be somewhat more pronounced at the turn of the nineteenth to twentieth century, that is, precisely the moment in time when the concept SWEET-SMELLING exhibits a more dramatic rise in the artificial sense. The domain CLEANING increases substantially from the 1890s to the 1910, after which there is sharp but short-lived decline. Then, from the 1930s onward it rises again and remains more or less stable until the present-day. The situation with respect to COSMETICS is similar, although the increase is more gradual overall. However, there is a more noticeable upward

[8] This analysis is based on the newly updated version, released in 2021. As the decade 1810s has been eliminated in the new version, it is not featured in Figure 4.

tendency from the 1910s to the 1990s. In the last decade in *COHA*, there seems to be a large decline of this domain, but the overall tendency is clearly one of an increase over time. Finally, TEXTILE & CLOTHING shows three pronounced increases: one from the 1820s until the 1870s, another from the 1890s to the 1950s, and a third from the 1980s onwards. The fact that this particular domain exhibits a more salient increase already from the beginning of the period analyzed is most probably due to the First Industrial Revolution, centering primarily on the manufacturing of textile products such as those derived from cotton (Jones, 1996: 117-123). and not on petroleum and chemicals, which were the basis of the Second Industrial Revolution.

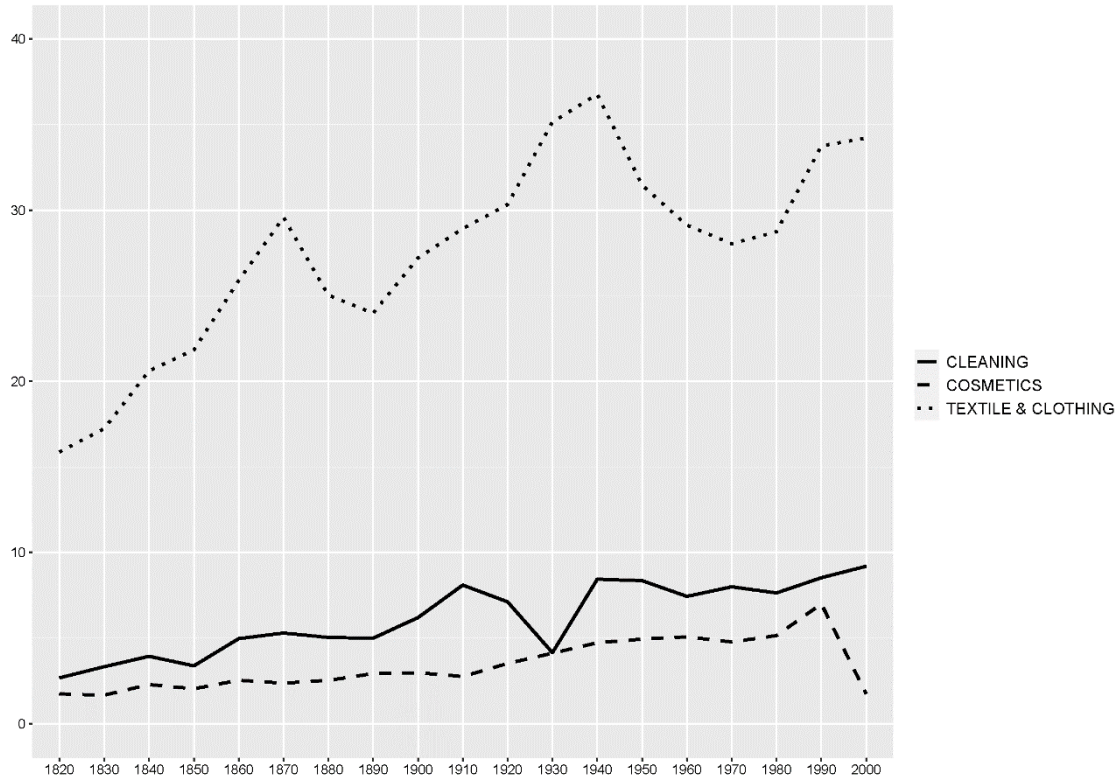


Figure 4: Mean frequency per million words of CLEANING, COSMETICS, and TEXTILE & CLOTHING (1820s–2000s)

These results, therefore, suggest that there was indeed a change at the conceptual level in American English, which might be a reflection of the well-known socio-economic transformations that took place in American society at the end of the nineteenth and the beginning of the twentieth century. With technological and scientific advancements

increasingly influencing people's daily lives, these semantic domains also became recurrent topics in conversation, as well as in written texts.

3.2. Changes in the internal structure of the synonym set

As mentioned in Section 2.2, two multinomial logistic regression models were computed to analyze the semantic division of labor between the three adjectives and to uncover potential changes in the internal structure of the synonym set: Model A included the variables 'sense', 'period', and 'text-type' as predictors of 'synonym', while Model B included 'semantic category', 'period', and 'text-type'. Table 2 displays the summary statistics of Model A, which is significant. It contains two goodness-of-fit statistics, namely Nagelkerke R^2 and an accuracy value which reflects the percentage of correct predictions. In the case of Nagelkerke R^2, values range between 0 and 1, so the value for Model A, 0.2275, signals that there is still some unexplained variance. However, this was to be expected due to the number of predictors considered and the fact that some variation is most likely random as the near-synonyms are interchangeable, which is not surprising given that they share their core denotational meaning. The accuracy of the model, 67.01 percent, points to a similar conclusion. All three predictors have significant effects on the choice of near-synonym, as well as the interactions between 'period' and 'sense' and 'period' and 'text-type'; in other words, the probabilities of the adjectives in each of the senses and text-types change from P1 to P4 (for more details, cf. Table 7 in the Appendix).

Table 2: Summary statistics of Model A

Likelihood ratio test	$\chi^2 = 1119.7$ ($p > 0.001$)
Nagelkerke R^2	0.2275
Accuracy	67.01 (baseline = 63.90)

Table 3 displays the summary statistics of Model B. This model is also statistically significant and the Nagelkerke R^2 and accuracy values, 0.2956 and 68.23 percent, respectively, again indicate that there is still some unaccounted variance in the model. However, both values are slightly better than those of Model A, which was to be expected since 'semantic category' provides a more fine-grained division than 'sense' and, therefore, it might identify differences between the preferences of the near-synonyms that

may have been obscured in Model A. Again, the interactions between 'period' and 'semantic category' and 'period' and 'text-type' are significant (for more details, cf. Table 8 in the Appendix).

Table 3: Summary statistics of Model B

Likelihood ratio test	$\chi^2 = 1509.7$ ($p > 0.001$)
Nagelkerke R^2	0.2956
Accuracy	68.23 (baseline = 63.90)

Figures 5-7 plot the probabilities of *fragrant*, *perfumed*, and *scented* across senses, semantic categories, and text-types, respectively, over time. In these figures, a separate graph is provided for each sense, semantic category, and text-type. The interaction between 'period' and 'sense' is shown in Figure 5.

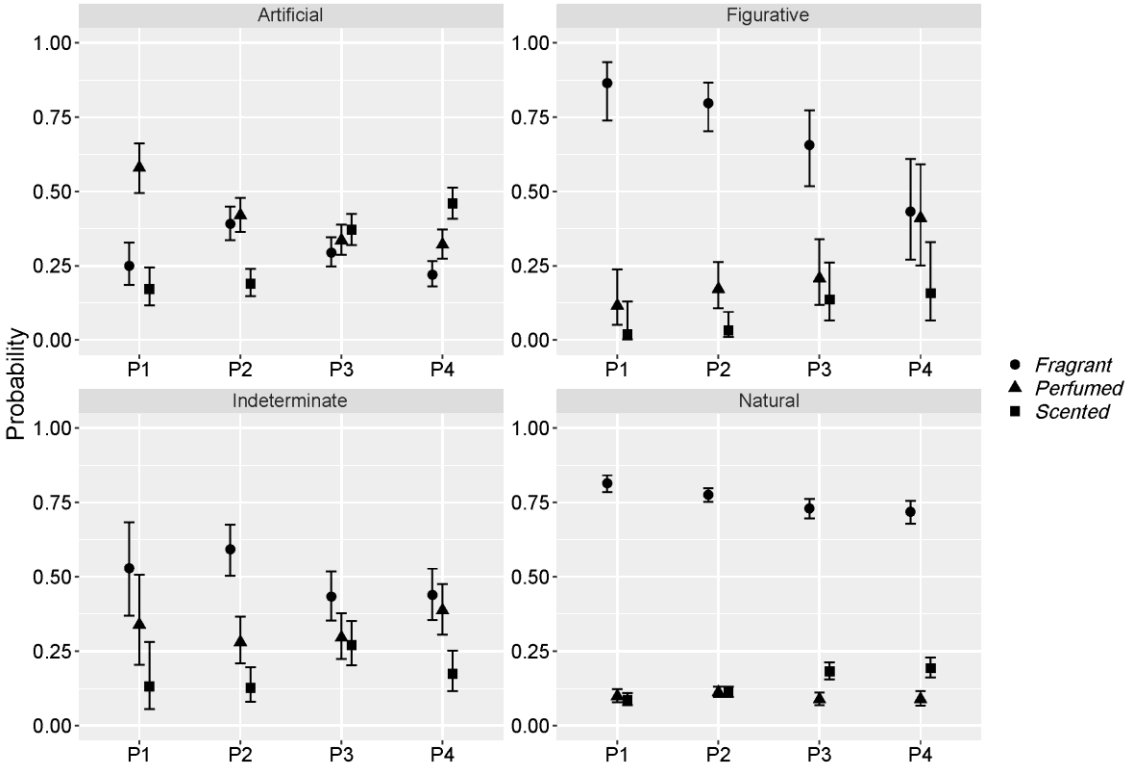


Figure 5: Interaction between 'period' and 'sense' (Model A)

Concerning the level artificial, several significant tendencies emerge from the data. *Fragrant* exhibits a gradual downward tendency from P2 onwards, but the only

significant contrast is that between P2 (39.09 percent) and P4 (21.99 percent). *Perfumed* also declines in this sense diachronically, in particular from P1 (57.94 percent) and P2 (41.96 percent) to P3 (33.53 percent) and P4 (32.07 percent). In fact, *perfumed* loses its original dominance in favor of *scented* from P3 onwards: the probabilities of these two adjectives are very similar in P3 (33.53 percent and 37.07 percent, respectively) but then in P4, *scented* becomes more probable than *perfumed* (45.95 percent vs. 41.96 percent). The rise in probability of *scented* is especially pronounced between P2 (18.05 percent) and P3 (37.07 percent), thus corresponding to the turn of the nineteenth to the twentieth century, just as in the case of many of the developments discussed up to now. In the natural sense, *fragrant* also shows a significant downward trend, especially from P1 onwards, but this adjective continues to be by far the most probable variant in all periods. Similarly, *Perfumed* declines over time, but only from P2 onwards. As regards *scented*, an increase in this sense with each subsequent period is observable, but although it is statistically significant, the likelihood of *scented* is still much lower than that of *fragrant*, even in P4 (19.28 percent vs 71.87 percent, respectively).

The confidence intervals in Figure 5 for the levels indeterminate and figurative reflect the relatively low frequency of the adjectives in these two uses. Due to this and the fact that indeterminate does not constitute a sense in its own right, the developments of the adjectives in these two levels are not further discussed.

Figure 6 displays the interaction between 'period' and the more fine-grained semantic variable 'semantic category'. A great number of significant tendencies emerged from Model B, but as the main focus here is on the natural-artificial divide, only the results of the prototypically natural and artificial semantic categories are here discussed, thus leaving out the categories ABSTRACT, BODY & PEOPLE, SENSATION, SUBSTANCE & MATERIAL, and SPACE.

Starting with the prototypically artificial categories, namely CLEANING, COSMETICS, OBJECT, and TEXTILE & CLOTHING, both *fragrant* and *perfumed* decrease significantly in OBJECT, and *perfumed* also in CLEANING, and they do so in favor of *scented*, which increases significantly in both categories. In CLEANING, *scented* overthrows *perfumed*, the initially most probable choice in this category. In OBJECT, *scented* overthrows both *fragrant* and *perfumed* and goes from the least likely option in P1 to the most likely one in P4. In the case of COSMETICS, the probabilities of the adjectives in P1 are meaningless,

given that only four examples occur in this period and all correspond to *fragrant*. *Fragrant* then decreases from P2 onwards and loses its dominance as it is surpassed by the other two adjectives in P3 and P4. Finally, *scented* is the only adjective that undergoes significant changes in TEXTILE & CLOTHING, increasing between P2 and P3.

As regards the prototypically natural categories EARTH, ATMOSPHERE, AND WEATHER, FOOD & DRINK, and PLANTS & FLOWERS, the probability of *fragrant*, the dominant option in all these domains, only decreases in PLANTS & FLOWERS. However, its dominance in this category is never threatened by the other adjectives, although *scented* increases from P1 onwards. In turn, *perfumed* decreases with FOOD & DRINK and PLANTS & FLOWER nouns. On the contrary, the probability of *scented* is significantly higher in P2 than in P1 in FOOD & DRINK. Lastly, in EARTH, ATMOSPHERE, AND WEATHER, only the probabilities of *scented* change significantly over time as it rises from P2 to P3/P4.

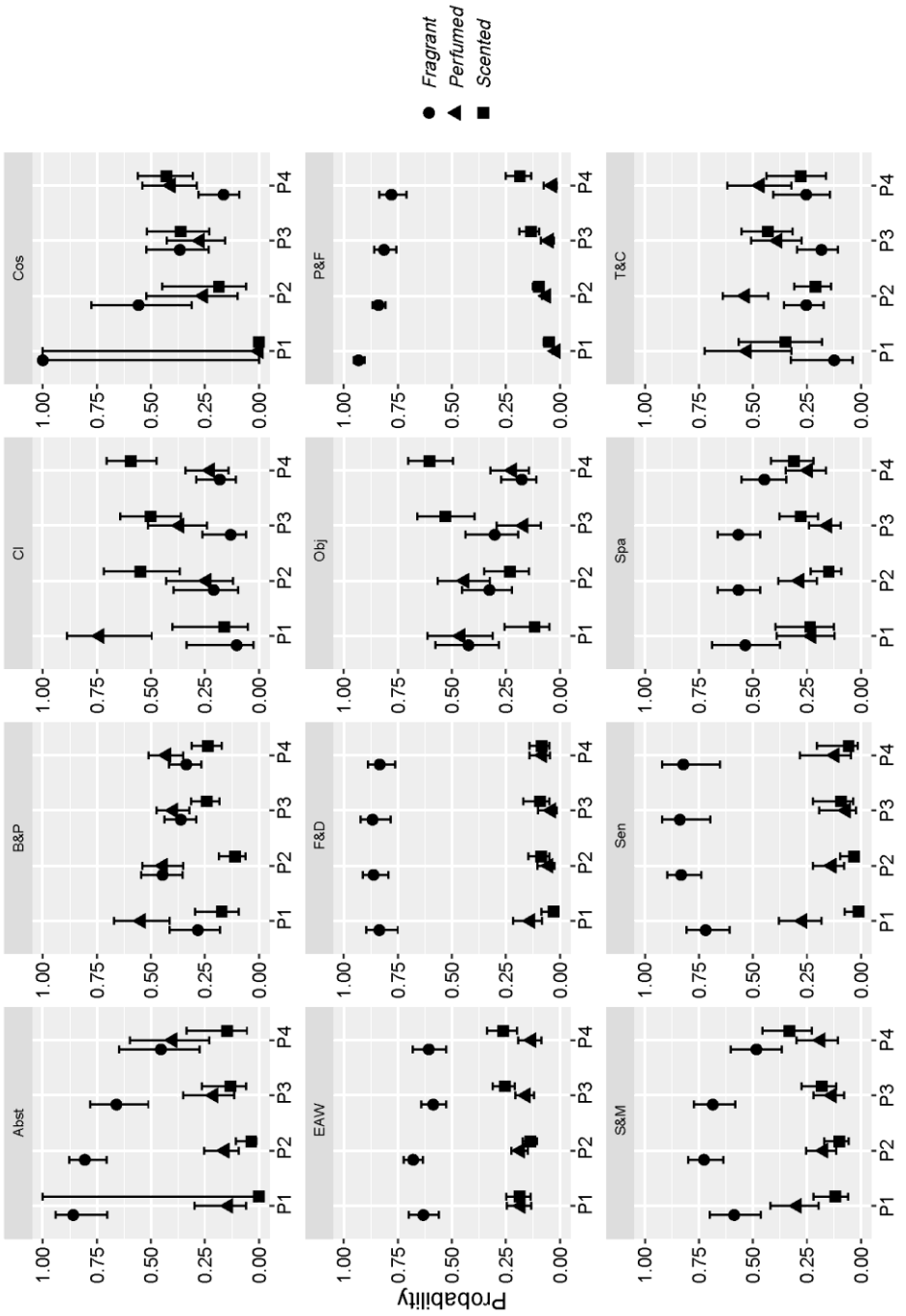


Figure 6: Interaction between ‘period’ and ‘semantic category’ (Model B)

Finally, the interaction between ‘period’ and ‘text-type’ is visualized in Figure 7. *Fragrant* maintains its dominance in both fiction and periodicals in all periods, although it decreases in both text-types, always in favor of *scented*, which increases significantly. Similarly to *fragrant*, *perfumed* also undergoes a significant decline in fiction over time, especially from P2 to P3. In non-fiction, both *fragrant* and *perfumed* also decrease significantly, while *scented* shows the opposite trend. In fact, *scented* surpasses *perfumed*

already in P3 and *fragrant* in P4, thus becoming the most likely choice in this text-type at the end of the time-span analyzed.

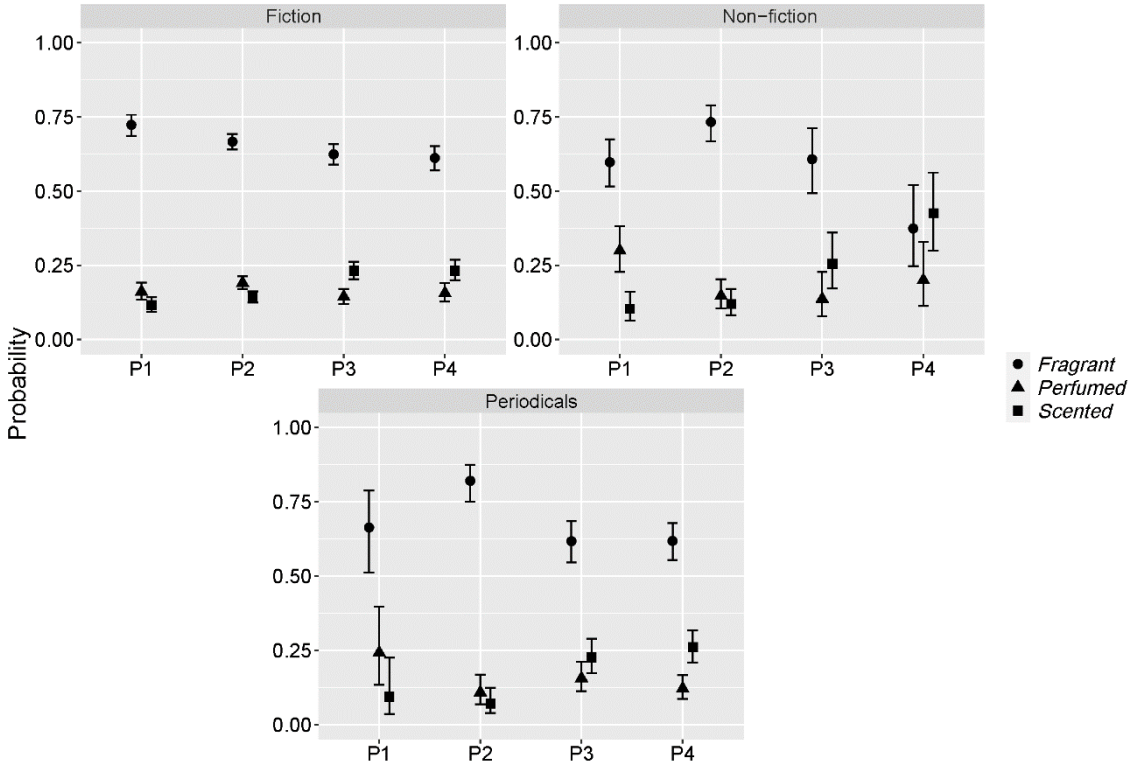


Figure 7: Interaction between ‘period’ and ‘text-type’ (Model A)

As mentioned in Section 2.2, mixed-effects regression models were also computed to account for the possible effects of the individual noun collocates of the adjectives. The lexical variable ‘collocate’ was included as a random effect to identify whether there exist idiosyncratic collocational preferences of the synonyms that explain the choice between them. Six separate models were calculated, two for the collocational preferences of each adjective, one with ‘sense’ and another with ‘semantic category’. However, only the goodness-of-fit statistics for the models comparing the probabilities of *fragrant* with the other two adjectives combined (cf. Tables 4 and 5) are here provided in order to show the considerable improvement of these models over those with only fixed effects.

Table 4: Summary statistics of Mixed Model A (*fragrant* vs. rest)

Nagelkerke R^2 (marginal)	0.1678

Nagelkerke R^2 (conditional)	0.3737
Accuracy	77.55 (baseline = 63.90%)

Table 5: Summary statistics of Mixed Model B (*fragrant* vs. rest)

Nagelkerke R^2 (marginal)	0.2217
Nagelkerke R^2 (conditional)	0.4112
Accuracy	75.48 (baseline = 63.90%)

In this case, two Nagelkerke R^2 values are provided, termed marginal and conditional. These values were retrieved with the `r.squaredGLMM` function in the `MuMIn` package (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, there is an important difference between the two values, as the latter is almost twice as large as the former in both models. This finding implies that the individual noun collocates of the adjectives are extremely important to discriminate between them, since by adding the collocates to the models, the amount of explained variance in the data increases substantially. This improvement is also reflected in the accuracy of the models, which are 10.54 and 7.25 higher than those in the fixed effects only models (cf. Tables 2 and 3), respectively.

By examining the particular noun collocates that are strongly associated with each adjective, it is possible to discern idiosyncratic preferences that help explain the behavior of the synonyms over time. In what follows, two cases of highly prominent collocates of *scented* according to the mixed models are discussed, as this is the adjective of the set that undergoes a significant increase in frequency and even comes to surpass the other two in some contexts of use. The first example concerns the noun *soap* belonging to CLEANING. Whereas in P1 there is only one example of *scented* collocating with *soap*, in P3 there are 16 occurrences out of a total of 23 instances of *scented* in this category (69.57 percent). Similarly, in P4 there are 16 occurrences of *scented soap* out of a total of 40 with CLEANING nouns (40 percent). In fact, Mixed Model B for *scented* shows that the significant increase of this adjective with the semantic category CLEANING is weaker in P3 and P4 than in Model B.

A similar case is that of the noun *candle*, which might go a long way to explain the pronounced increase of *scented* in the category OBJECT. This is so because *candle* is modified by *scented* a total of 25 times in *COHA*, with 24 of them occurring in P4. Similarly to the case of *scented soap*, *candle* represents 24 out of 50 cases of *scented* occurring in OBJECT in P4 (almost 50 percent). This is again reflected in Mixed Model B for *scented*, as the significant increase over time of this adjective in OBJECT is weaker when 'collocate' is included as a random effect than in Model B.

Moreover, *candle* represents a particularly fascinating case for a diachronic study such as the present one because this noun is almost exclusively used with *scented*. 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, it could already be considered a compound or 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 intervening material, in 20 of the 25 cases, as in example (9):

(10) 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).

In fact, it could even be argued that the combination *scented candle* is in the early stages of a *lexicalization* process (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 is that of refreshing or perfuming the air of a place, much like air fresheners.

In sum, the findings of this section demonstrate that there exists a division of semantic labor between *fragrant*, on the one hand, and *perfumed* and *scented*, on the other, in line with the conclusions extrapolated from dictionary definitions (cf. Section 1): *fragrant* is the preferred adjective with nouns denoting entities emitting natural smells, and thus it is also favored with the natural categories EARTH, ATMOSPHERE, AND WEATHER, FOOD & DRINK, and PLANTS & FLOWERS, and not much changes over time in these contexts

of use. Contrariwise, *perfumed* and *scented* lie at the opposite end of the semantic continuum, being the most common choices with nouns referring to entities with artificial aromas. In this sense, there are significant diachronic tendencies because, with time, *scented* surpasses *perfumed*, the most common choice at the beginning of the time-span. In particular, *scented* ousts *perfumed*, but also *fragrant*, in the categories CLEANING and OBJECT, where it becomes the default option in P4, and it also increases significantly in COSMETICS and TEXTILE & CLOTHING. *Perfumed* remains rather stable in two of the prototypically artificial categories, namely COSMETICS and TEXTILE & CLOTHING, although it enters into competition with *scented* in the former in P3 and P4. However, not only abstract semantic co-occurrence patterns such as these play a role on the choice between these three near-synonyms, since idiosyncratic collocational preferences explain a great deal of the existing variation between them. In fact, when adding the variable 'collocate' to the models, their goodness-of-fit increases substantially and some of the effects of the semantic categories become weaker. Finally, the general increase in frequency of *scented* is not only particularly sharp in specific senses and semantic categories, but also in concrete text-types. This is so because, although it rises significantly in all three levels of this predictor, the upward tendency of this adjective is much more pronounced in non-fictional texts.

4. Conclusions and further research

By examining the diachronic behavior of the near-synonymous adjectives *fragrant*, *perfumed*, and *scented* in American English, this paper sought to address three main research questions (cf. Section 1). First, the study focused on the evolution of the concept SWEET-SMELLING as a whole, that is, as denoted by the three near-synonyms, throughout the period 1810–2009 (cf. Section 3.1). The results pointed to a rise in the use of the concept to describe artificial aromas at the expense of natural smells, which was located in particular in BODY & PEOPLE, CLEANING, COSMETICS, FOOD & DRINK, OBJECT, and TEXTILE & CLOTHING. A hypothesis to explain the increase in frequency of SWEET-SMELLING in the artificial sense on the basis of the transformation experienced by American society as a result of processes of industrialization was examined. Preliminary results focusing on the frequency over time of a series of noun collocates of the adjectives pertaining to the prototypically artificial domains CLEANING, COSMETICS, and TEXTILE &

CLOTHING seemed to support this hypothesis, thus pointing to an influence of extra-linguistic factors on the distribution of the concept SWEET-SMELLING. However, the results presented can still be improved as, for instance, a wider range of lexical indicators and semantic domains can be investigated in order to verify or refute the claim.

Second, the study moved on to the division of semantic labor between the adjectives and the possible changes in the internal structure of the synonym set, that is, whether the relationship between *fragrant*, *perfumed*, and *scented* remained stable or fluctuated throughout the time span 1810–2009 (cf. Section 3.2). Even though a mixture of schematic co-occurrence patterns such as sense and semantic category and idiosyncratic collocational preferences emerged from the analysis, by far the most important finding in the present analyses is the extremely valuable contribution of the individual noun collocates of the adjectives to shed light on the alternation between them. In fact, the two semantic variables, although significant, are not enough to explain all the variation in the data, as the rather low accuracy and R^2 values indicated. Therefore, in the future it might be worth employing analysis which are specifically geared towards examining the collocational behavior of words, both of a quantitative (e.g. Semantic Vector Space Modelling) and qualitative nature (e.g. collocational networks). Such methods might be valuable to clarify whether the increase of *scented* at the expense of *fragrant* and *perfumed* uncovered here is (i) mostly due to *scented* modifying nouns which were previously more strongly associated with them as is the case of *soap*, previously a common collocate of *perfumed* or (ii) mostly due to *scented* collocating with nouns which did not typically co-occur with the adjectives until later periods, as is the case of *candle*. This, however, is an issue which lies outside the scope of the present paper and must thus be left for future research.

Appendix

Table 6. Model A coefficients (parentheses indicate non-significant estimates)

Coefficient	***Fragrant***	***Perfumed***	***Scented***
Intercept	1.0910	-1.9530	-1.9380
Period = P1	0.4724	-0.4174	-0.3807
Period = P1: Sense = Artificial	-0.8719	0.7974	(0.2439)
Period = P1: Sense = Figurative	(0.2402)	(-0.3242)	(-0.1904)
Period = P1: Sense = Indeterminate	(-0.4863)	(0.4180)	(0.3619)
Period = P1: Text-type = Periodicals	-1.0960	1.1410	(0.5564)

Period = P1: Text-type = Non-fiction	-0.8311	1.1040	(0.1088)
Period = P3	(-0.0835)	-0.3912	0.4253
Period = P3: Sense = Artificial	(-0.1894)	(-0.0766)	(0.3923)
Period = P3: Sense = Figurative	(-0.4833)	(0.4953)	(1.0090)
Period = P3: Sense = Indeterminate	(-0.4005)	(0.3428)	(0.3924)
Period = P3: Text-type = Periodicals	-0.8585	0.7317	(0.7222)
Period = P3: Text-type = Non-fiction	(-0.3849)	(0.2335	(0.3220)
Period = P4	(-0.0583)	-0.3684	0.3735
Period = P4: Sense = Artificial	-0.5173	(-0.1329)	0.7132
Period = P4: Sense = Figurative	-1.3080	1.4670	(1.1300)
Period = P4: Sense = Indeterminate	(-0.3285)	0.7314	(-0.2282)
Period = P4: Text-type = Periodicals	-0.7973	(0.3257)	0.9327
Period = P4: Text-type = Non-fiction	-1.2840	(0.4578)	1.0270
Sense = Artificial	-1.6840	1.7320	0.5821
Sense = Figurative	(0.1243)	(0.5028)	-1.3490
Sense = Indeterminate	-0.8653	1.1280	(0.1204)
Text-type = Periodicals	0.8290	-0.6367	-0.7620
Text-type = Non-fiction	0.3136	(-0.3046)	(-0.1933)

Table 7. Model B coefficients ((parentheses indicate non-significant estimates)

Coefficients	***Fragrant***	***Perfumed***	***Scented***
Intercept	0.6446	-1.4230	-1.7380
Semantic category = ABST	0.6731	(-0.1741)	-1.4430
Semantic category = B&P	-0.9624	1.2620	(-0.2654)
Semantic category = CL	-2.0790	(0.3229)	2.0230
Semantic category = COS	(-0.5205)	(0.4227)	(0.3536)
Semantic category = F&D	1.0820	-1.4230	(-0.5121)
Semantic category = OBJ	-1.4750	1.2390	(0.6234)
Semantic category = P&F	0.8995	-1.2100	(-0.3721)
Semantic category = S&M	(0.2272)	(-0.0698)	(-0.3457)
Semantic category = SEN	0.8558	(-0.3543)	-1.5360
Semantic category = SPA	-0.4802	0.5656	(0.0881)
Semantic category = T&C	-1.8300	1.6080	(0.4905)
Period = P1	(0.0325)	(-0.2590)	(0.2232)
Period = P1: Semantic category = ABST	(0.6044)	(-0.1468)	(-12.5700)
Period = P1: Semantic category = B&P	(-0.4893)	(0.4202)	(0.1918)
Period = P1: Semantic category = CL	(-0.5963)	2.2160	-2.1750
Period = P1: Semantic category = COS	(12.4100)	(-12.3100)	(-14.4100)
Period = P1: Semantic category = F&D	(-0.0144)	1.0940	-1.5390
Period = P1: Semantic category = OBJ	(0.6519)	(0.1011)	(-1.1360)
Period = P1: Semantic category = P&F	1.2330	-1.2710	-1.0240
Period = P1: Semantic category = S&M	(-0.4026)	(0.6979)	(-0.1917)
Period = P1: Semantic category = SEN	(-0.4942)	(0.8918)	(-1.5320)

Period = P1: Semantic category = SPA	(0.0838)	(-0.3004)	(0.1848)
Period = P1: Semantic category = T&C	(-0.6594)	(0.0253)	(0.4428)
Period = P3	(-0.2765)	(-0.2682)	0.6603
Period = P3: Semantic category = ABST	(-0.3692)	(0.5250)	(0.6183)
Period = P3: Semantic category = B&P	(0.04264)	(-0.0098)	(0.1926)
Period = P3: Semantic category = CL	(-0.1621)	(0.8081)	(-0.9542)
Period = P3: Semantic category = COS	(-0.3777)	(0.2668)	(0.1502)
Period = P3: Semantic category = F&D	(0.4471)	(-0.0691)	(-0.6994)
Period = P3: Semantic category = OBJ	(0.2929)	-1.1670	(0.5631)
Period = P3: Semantic category = P&F	(0.2302)	(-0.0442)	(-0.4133)
Period = P3: Semantic category = S&M	(0.2033)	(-0.1253)	(-0.0974)
Period = P3: Semantic category = SEN	(0.4386)	(-0.5704)	(0.3268)
Period = P3: Semantic category = SPA	(0.4022)	(-0.5936)	(0.0282)
Period = P3: Semantic category = T&C	(-0.0148)	(-0.3939)	(0.3036)
Period = P4	(-0.0453)	(-0.5274)	0.5405
Period = P4: Semantic category = ABST	-1.2720	1.6840	(0.7237)
Period = P4: Semantic category = B&P	(-0.1682)	(0.3542)	(0.1546)
Period = P4: Semantic category = CL	(0.1361)	(0.3662)	(-0.5904)
Period = P4: Semantic category = COS	-1.5420	(1.1230)	(0.4202)
Period = P4: Semantic category = F&D	(0.1083)	(0.9421)	(-0.8190)
Period = P4: Semantic category = OBJ	(-0.4980)	(-0.5646)	0.8600
Period = P4: Semantic category = P&F	(-0.0865)	(-0.1746)	(-0.0513)
Period = P4: Semantic category = S&M	(-0.7272)	(0.4941)	(0.6798)
Period = P4: Semantic category = SEN	(0.2537)	(0.3228)	(-0.2150)
Period = P4: Semantic category = SPA	(-0.1700)	(0.2117)	(0.1550)
Period = P4: Semantic category = T&C	(0.3039)	(0.1761)	(-0.3988)

References

Allan, K. 2015. ‘Education in the Historical Thesaurus of the Oxford English Dictionary’, in J. Daems et al. (eds.) *Change of paradigms – new paradoxes: Recontextualizing language and linguistics*, pp. 81–96. Berlin/Boston: De Gruyter Mouton.

Archer, D., A. Wilson, and P. Rayson. 2002. ‘Introduction to the *USAS* category system’, *Benedict Project Report*, pp. 1–37. http://ucrel.lancs.ac.uk/usas/ (accessed 7 January 2020).

Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to R*. Cambridge/New York: CUP.

Baker, P. 2017. *American and British English: Divided by a Common Language?* Cambridge: CUP.

Barton, K. 2020. *MuMIN: Multi-model inference*. https://CRAN.R-project.org/package=MuMIN.

Bates, D. et al. 2015. 'Fitting linear mixed-effects models using lme4', *Journal of Statistical Software* 67 (1), pp. 1–48.

Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Blank, A. 1999. 'Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change', in A. Blank and P. Koch (eds.) *Historical Semantics and Cognition*, pp. 61–89. Berlin/New York: Mouton de Gruyter.

Brinton, L. J., and E. C. Traugott. 2005. *Lexicalization and Language Change*. Cambridge/New York. CUP.

Church, K. W., and P. Hanks. 1990. 'Word association norms, mutual information, and lexicography', *Computational Linguistics* 16 (1), pp. 76–83.

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: OUP.

Davies, M. 2010–. *The Corpus of Historical American English*. https://www.english-corpora.org/coha/ (accessed 7 January 2020).

Desagulier, G. 2014. 'Visualizing distances in a set of near-synonyms: *rather*, *quite*, *fairly*, and *pretty*', in D. Glynn and J.A Robinson (eds.) *Corpus methods for semantics. Quantitative studies in polysemy and synonymy*, pp. 145–178. Amsterdam/Philadelphia: John Benjamins.

Divjak, D., and S. Th. Gries. 2008. 'Clusters in the mind? Converging evidence from near synonymy in Russian', *The Mental Lexicon* 3 (2), pp. 188–213.

Durkin, P. 2009. *The Oxford Guide to Etymology*. Oxford: OUP.

Lexico. 2019–. Oxford: OUP. https://www.lexico.com/ (accessed 18 May 2021).

Geeraerts, D. 1986. 'On necessary and sufficient conditions', *Journal of Semantics* 5 (4), pp. 275–291.

Geeraerts, D. 1988. 'Where does prototypicality come from?', in B. Rudzka-Ostyn (ed.) *Topics in Cognitive Linguistics*, pp. 207–229. Amsterdam: John Benjamins.

Geeraerts, D. 1997. *Diachronic prototype semantics*. Oxford: Clarendon Press.

Geeraerts, D. 2010. *Theories of Lexical Semantics*. Oxford: OUP.

Geeraerts, D., S. Grondelaers, and P. Bakema. 1994. *The Structure of Lexical Variation: Meaning, Naming, and Context*. Berlin/New York: Mouton de Gruyter.

Gries, S. Th. 2001. 'A corpus-linguistic analysis of English *-ic* vs *-ical* adjectives', *ICAME Journal* 25, pp. 65–108.

Historical Thesaurus of the Oxford English Dictionary. Oxford: OUP. https://www.oed.com/thesaurus/ (accessed 7 January 2020).

Johnson, P. 1997. *A History of the American People*. New York: Harper Collins.

Jones, M. A. 1996. *Historia de Estados Unidos 1607–1992*. Translated by Carmen Martínez Gimeno. Madrid: Cátedra.

Kay, C. 2012. 'Developing *The Historical Thesaurus of the OED'* in K. Allan and J.A Robinson (eds.) *Current methods in historical semantics*, pp. 41–58. Berlin/Boston: De Gruyter Mouton.

Kay, C., and K. L. Allan. 2015. *English Historical Semantics*. Edinburgh: Edinburgh University Press

Leech, G., and R. Fallon. 1992. 'Computer corpora – What do they tell us about culture?', *ICAME Journal* 16, pp. 29–50.

Levshina, N., D. Geeraerts, and D. Speelman. 2013. 'Mapping construal spaces: A contrastive analysis of English and Dutch analytic causatives', *Linguistics* 51, pp. 825–854.

Levshina, N. 2015. *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam/Philadelphia: John Benjamins.

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 (1), pp. 56–87.

Liu, D. 2013. 'Salience and construal in the use of synonymy: A study of two sets of near-synonymous nouns', *Cognitive Linguistics* 24 (1), pp. 67–113.

Longman Dictionary of Contemporary English. 2015–. Pearson. https://www.ldoceonline.com/ (accessed 18 May 2021).

Lyons, J. 1963. *Structural Semantics*. Oxford: Basil Blackwell.

Lyons, J. 1977. *Semantics*. Cambridge: CUP.

Merriam Webster dictionary and thesaurus. 2017–. Merriam Webster. https://www.merriam-webster.com/ (accessed 18 May 2021).

Oaks, M. P. 2003. 'Contrasts between US and British English in the 1990s', in E. H. Oleksy and B. Lewandowska-Tomaszczyk (eds.) *Research and Scholarship in Integration Processes*, 213–222. Lodz: Lodz University Press.

Oxford English Dictionary. 3rd edn. 2000–. Oxford: OUP. https://www.oed.com/ (accessed 18 May 2021).

Pettersson-Traba, D. 2021. 'A diachronic perspective on near-synonymy: The concept of SWEET-SMELLING in American English', *Corpus Linguistics and Linguistic Theory* 17 (2), pp. 319-349.

Potts, A., and P. Baker. 2012. 'Does semantic tagging identify cultural change in British and American English?', *International Journal of Corpus Linguistics* 17 (3), pp. 295–324.

Primahadi-Wijaya-R., G., and I M. Rajeg. 2014. 'Visualising diachronic change in the collocational profiles of lexical near-synonyms', in I N. Sudipa and G. Primahadi-Wijaya-R. (eds.) C*ahaya Bahasa: A Festschrift in Honour of Prof. I Gusti Made Sutjaja*, pp. 247–258. Denpasar: Swasta Nulus.

Quirk, Randolph et al. 1985. *A Comprehensive Grammar of the English Language.* London/New York: Longman.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*, version 3.4.3. Vienna: R Foundation for Statistical Computing.

Sinclair, J. 1966. 'Beginning the study of lexis', in C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins (eds.) *In Memory of J.R Firth*, pp. 410–30. Harlow: Longman.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: OUP.

Soares da Silva, A. 2015. 'Competition of synonyms through time: Conceptual and social salience factors and their interrelations', *Catalan Journal of Linguistics* 14, pp. 199–218.

Speelman, D., and D. Geeraerts. 2009. 'Causes for causatives: The case of Dutch *doen* and *laten*', in T. Sanders and E. Sweetser (eds.) *Casual Categories in Discourse and Cognition*, pp. 173–204. Berlin/New York: Mouton de Gruyter.

Tagliamonte, S. A., and R. H. 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), pp. 135–178.