

Review of Pettersson-Traba, Daniela. 2022. *The Development of the Concept of SMELL in American English. A Usage-Based View of Near-Synonymy*. Berlin: De Gruyter Mouton. ISBN: 978-3-11079-2201. DOI: <https://doi.org/10.1515/9783110792294>

Daniel Granados-Meroño
University of Murcia / Spain

The aim of the monograph is to provide a comprehensive insight into the development of the concept of SMELL in American English in the period ranging from the nineteenth century until 2009. By using a corpus-based approach, as well as a thoughtful and advanced deployment of statistical analysis, the goal is to observe the semantic evolution of five near-synonyms related to SMELL, namely *fragrant*, *perfumed*, *scented*, *sweet-scented*, and *sweet-smelling*.

Pettersson-Traba begins by acknowledging the difficulty in providing a complete definition of ‘synonymy’. Even if some dictionaries might define this notion as the linguistic phenomenon in which a word or expression means the same as another word or expression, she points out that “a partial degree of similarity is also considered for a word or expression to constitute a synonym of another term” (Cruse 2004: 157). These two views are used in the study to distinguish between ‘absolute synonymy’, which is found when a total similarity between two or more words takes place, and ‘partial synonymy’, which relates to contexts in which the similarity is not complete. Partial synonymy is much more frequent in language, while absolute synonymy is very rare (Cruse 2004: 157–158; Divjak and Gries 2006: 24; Liu 2010: 56–57; Taylor 2003: 264).

In Chapter 1, the author guides the reader through the most relevant schools and research approaches that have dealt with ‘lexical semantics’. Lexical semantics is defined as a field in linguistics that has been attempting to answer whether the semantic dimension of language is a purely linguistic feature, or it is rather influenced by encyclopaedic



knowledge —being thus relevant for the theoretical background of the study carried out in the monograph. In terms of research approaches, the most important basis for Pettersson-Traba's study are 1) distributional corpus-based approaches —which combine the interest in collocations and the use of more fine-grained statistical analyses in data retrieved from representative corpora— and 2) cognitive semantics —which is the most relevant theoretical framework in current research due to the importance of concepts such as 'prototypicality' (Rosch 1973) and 'entrenchment' (Langacker 1987).

In this first chapter, Pettersson-Traba introduces the aims, scope, relevant contributions, and structure of the study. *Fragrant*, *perfumed*, *scented*, *sweet-smelling*, and *sweet-scented* are selected as representations of the semantic field SMELL, which is interesting because of its richness in terms of near synonymy. The five near-synonyms are chosen due to their low degree of polysemy, which avoids discarding instances that denote a meaning related to other semantic fields.

In Chapter 2, the reader is provided with a classification of synonymy and an exhaustive review of the most relevant literature dealing with it. The types of synonymy which are most recurrently mentioned in classifications are 1) absolute synonymy, 2) cognitive synonymy, and 3) near-synonymy. Absolute synonymy accounts for those words or word senses identical on all four dimensions of meaning, namely, denotational, stylistic, expressive and collocational meaning (Leech 1990). Cognitive synonymy concerns pairs (or groups) of words that, despite being identical on the denotational dimension and mutually entailing one another, differ in non-denotational traits, such as connotation, register, style, or the language variety where they occur. Finally, near-synonymy, which is considered the most common type of synonymy, refers to those words that differ slightly in conceptual content and are not denotationally identical. Still, these synonyms are sufficiently similar to be interchanged in many contexts of use (Cruse 2004: 159). However, the boundaries between cognitive and near-synonymy are blurred and authors such as Edmonds and Hist (2002: 116–117) or Desagulier (2014: 153) argue for a two-fold division, namely absolute vs. non-absolute synonymy, which is the classification followed in the study.

Sections 2.2 and 2.3 provide a thorough review of the literature on the topic which is based on distribution usage methods to study synonymy. The review firstly points out that Divjak and Gries (2006) were some of the first who attempted to cluster potential near-synonyms in groups, rather than studying pairs of words, and included a wider range

of factors in the analysis, which required the use of more sophisticated multivariate techniques. Similarly, Gries and Otani's study (2010) on the near-synonyms set from the semantic domain of SIZE is also discussed in depth. Their study covers two sets—one comprising *little*, *small* and *tiny* and another including *big*, *great* and *large*—and analyses several factors (such as aspect, voice, or transitivity marking of the finite verb of the adjectives) at a morphological, syntactic, and semantic level. According to Pettersson-Traba, Gries and Otani (2010) is the most comprehensive work on synonymous adjectives, and thus, one of her main inspirations for the study on the methodological level. Finally, a previous study carried out by the author (Pettersson-Traba 2021) is mentioned as one of the few studies dealing with semantic change in near-synonymous adjectives diachronically. Pettersson-Traba (2021) examines the use of the above-mentioned synonyms related to SMELL by focusing on their modified nouns, which are classified into nine different categories. Results suggest that major changes took place in the nineteenth century, and it is hypothesised that these might be due to extralinguistic factors, such as those of industrialisation and mass production, which led to the introduction of artificially scented soaps and candles in the market.

In its first section, Chapter 3 deals more exhaustively with the synonym set (*fragrant*, *perfumed*, *scented*, *sweet-smelling* and *sweet-scented*) which is analysed in the study. It also provides the motivations behind the choice of SMELL as the object of study. The author also examines reference works to provide a preliminary idea of the meanings and contexts in which the synonym set is used. Dictionaries such as the *American Heritage Dictionary of English Language* (ACDOE),¹ the *Cambridge Dictionary* (CD),² or the *Merriam-Webster Dictionary* (MW)³ are used to provide insights on how these words differ between each other depending on the period of time. The study also shows the difficulties in determining the changes these words underwent and the (blurred) boundaries between their meanings.

Section 3.2 introduces the *Corpus of Historical American English* (COHA; Davies 2010) used in the analyses presented in Chapters 4–6. Pettersson-Traba grounds the selection of this database in the need of using a very large corpus due to the low frequency of the five near-synonymous adjectives under study. COHA fulfils this requirement, as it

¹ <https://www.ahdictionary.com/>

² <https://dictionary.cambridge.org/>

³ <https://www.merriam-webster.com/>

contains more than 475 million words from more than 100,000 individual texts, divided into four different genres: fiction, magazines, newspapers and non-fiction. Likewise, the corpus is suitable for a diachronic study because it covers the period 1810–2000 .

Section 3.3 explains the data annotation process, which consists in the manual revision of the POS tagging available for COHA, namely by excluding false positives of adjectives that are actually past participle verbs (*fragrant, scented*). The semantic domain is annotated using the *UCREL Semantic Analysis System* (USAS; Archer *et al.* 2003), together with a manual revision assisted by a more precise database, the *Historical Thesaurus of the Oxford English Dictionary* (HTOED).⁴ The remaining of Chapter 3 describes the wide range of variables included in the analysis of the first dataset (language-internal semantic, language-internal non-semantic and language-external variables). These variables and their levels are presented in Table 1 below.

Variable types	Variable	Variable levels	
Language-internal semantic variables	Sense	Natural	
		Artificial	
	Semantic category		Figurative
			Indeterminate
			ABSTRACT
			BODY AND PEOPLE
			CLEANING
			COSMETICS
			EARTH, ATMOSPHERE, AND WEATHER
			FOOD AND DRINK
OBJECT			
PLANTS AND FLOWERS			
SENSATION			
SPACE			
SUBSTANCE AND MATERIAL			
TEXTILE AND CLOTHING			
Animacy		Animate	
		Inanimate	
Concreteness rating		Average rating of concreteness from 1 to 5	
Concreteness binary		Concrete	
Countability		Abstract	
		Count	
		Non-count	
		Other	

Table 1: Variables and their levels in the first dataset (Adapted from Pettersson-Traba 2022: 95–96)

⁴ <https://historicalthesaurus.arts.gla.ac.uk/articles/>

Variable types	Variable	Variable levels
Language-internal non-semantic variables	Syntactic function	Attributive Predicative complement Postpositive Other
	Degree	Positive Comparative Superlative
	Collocate	Specific noun collocate (lemma)
Language-external variables	Period	Period 1 (1810–59) Period 2 (1860–1909) Period 3 (1910–59) Period 4 (1960–2009)
	Text-type	Fiction Non-fiction Periodicals

Table 1: Continuation

Chapter 4 deals with two closely related analyses: a semasiological analysis and an onomasiological one. Semasiology is the study of particular words and the sense or concepts that they designate, having a stronger interest in polysemy. The first study therefore focuses on the analysis of the near-synonymous adjectives over time to uncover potential changes in their prototypical structure. This has the purpose of determining whether any adjective within the set has a special effect on the semantic evolution of the concept of SMELL, or rather the whole set has a similar effect on it. Here, the frequency of use of the adjectives in COHA is analysed in regard to the changes caused by the variable *Period*. The second study, which analyses synonymy rather than polysemy, deals with the examination of various expressions which are used to designate a particular concept. As such, the starting point is based on the concepts or senses rather than on the words that designate them.

The results of the semasiological analysis provided in Section 4.3 show that, regarding the variable *Sense* across the four levels of *Period*, the adjective *fragrant* remains prototypical in the natural sense, but its use decreases significantly over time. We may witness a similar evolution in the figurative sense, while the indeterminate and artificial senses increase. In the case of *perfumed*, an increase in the use of three senses, namely artificial, indeterminate, and figurative is observed, while its use to denote natural aromas declines substantially. A similar trend can be appreciated for *scented*, while *sweet-scented* and *sweet-smelling* remain stable across the four periods. The latter is more prototypically used in the natural sense even in *Period 4*, despite the downward tendency.

The patterns arising from the analysis of the variable *Semantic category* are coherent with the analysis of the variable *SENSE*, as the levels corresponding to natural sense tend to decline, while the ones corresponding to the remaining levels increase or remain stable. For instance, in the case of *fragrant*, levels such as *PLANTS AND FLOWERS* or *EARTH, ATMOSPHERE AND WATER*, which clearly refer to the natural sense, are the most frequent.

In turn, the onomasiological analysis shows that *fragrant* is the most salient adjective across all five natural categories, and that all prototypically artificial categories, except *TEXTILE AND CLOTHING*, exhibit distributional changes over time. In particular, the frequency of *scented* increases at the expense of *perfumed* and *fragrant*, becoming the most salient adjective by *Period 4*. Similarly, the frequency of *perfumed* increases considerably and becomes almost as salient as *fragrant* in *Period 4* regarding the figurative category *ABSTRACT*. Finally, when used for semantic categories concerning indeterminate senses, *fragrant* slightly decreases over time, mainly in favour of *sweet-smelling*. These processes show that there exists some interrelation between the variables *Sense*, *Semantic Category*, and *Period*. In Chapters 5 and 6, Pettersson-Traba explores the nature, relevance and details of these interrelations.

Chapter 5 provides a comprehensive onomasiological analysis of the synonym set by means of multivariate approaches. It attempts to explain the motivations behind the patterns described in Chapter 4 and to find out whether any of the variables in Table 1 might entail proper predictors of the speaker's preference for one adjective. In this chapter, the author makes use of a statistical analysis by using multinomial regression models and a random forest analysis. Pettersson-Traba provides a detailed explanation of the statistics, which makes it easier for the reader to understand the interpretation of the results. The results from the random forest analysis show that *Semantic category*, *Sense*, and *Period* are the most important variables of predictors in a first model obtained through the multinomial regression analysis. These variables are precisely the ones included in the analyses in Chapter 4, which provides an additional ground to the idea that the pattern behind the diachronic changes might not be random. The variable *Collocate* is included later in the models and is shown to be significantly relevant, as it increases their prediction accuracy by around ten per cent.

Finally, an interesting insight in Chapter 5 is the plausible existence of a (probably still ongoing) process of substitution within the synonymy set, whereby *scented* gains

ground at the expense of *fragrant* and *perfumed*, as the semantic categories related to artificial, indeterminate and figurative senses increase, while the natural ones (which are closely related to *fragrant*) decrease dramatically.

By using a dataset of their noun collocates in an L5-R5 context window, Chapter 6 provides a more detailed discussion of the effects of the variable *Collocate* as regards the preference of speakers with the choice of adjectives in the synonym set. These are extracted automatically from COHA by using its collocates and POS-tag options. The study uses Semantic Vector Space (SVS) modelling of nouns collocates and measures the semantic (dis)similarity between the near-synonyms. The analysis draws on the collocational profiles and Pointwise Mutual Information (PMI) to identify prominent collocations of adjectives which need to comply with the criteria postulated by Baker (2017: 98–100).

The results of the SVS analysis are fed into cluster analysis to explore and interpret the (dis)similarities between the adjectives in different periods, which provide very interesting patterns. The collocational preferences of the adjectives included in the study (*fragrant*, *scented*, and *perfumed*) result in five clusters. On the one hand, we have *perfumed* and *scented* in P1 and P2 in Cluster 1, while *perfumed* in P3 and P4 and *scented* in P3 are in Cluster 3, nearly positioned to Cluster 2 including *scented* in P4. On the other hand, *fragrant* presents different behaviour in terms of collocational preferences, with P3 and P4 in Cluster 4, and P1 and P2 in Cluster 5 respectively.

Pettersson-Traba's results do not only confirm the two-sided pattern which is observed in previous chapters —1) decrease of natural senses-related adjectives (*fragrant*) and 2) increase of the other senses-related adjectives (*scented*, *perfumed*)— but also suggest that there is a specific period in which the shift is especially dramatic: between P2 and P3, as P1 and P2 tend to group together and be separated from P3 and P4.

The results from the SVS and cluster analyses allow the author to corroborate the historical, cultural, and social changes that might explain the patterns. Important social and technological changes took place during the period examined in the monograph, in particular in the USA, as a result of the First and Second Industrial Revolutions. Pettersson-Traba argues that this could possibly constitute the underlying motivations accounting for the rise in the use of SMELL. This hypothesis is further discussed in Chapter 6.

The study in Chapter 6 aims at testing that the First and Second Industrial Revolutions account for the rise of SMELL. To do this, the author uses the most relevant semantic categories taken from previous chapters to select some noun collocates which belong to the semantic categories in question, including the collocates of the 15 nouns most frequently modified by all five near-synonyms, among others. With this new dataset, the author aims to determine whether the patterns attested in the analyses developed in previous chapters are exclusive to these synonyms or are also attested in nouns not related to SMELL. The results are enlightening: the second-order collocates of near-synonyms are examined to pinpoint whether the conceptualisation of the semantic categories changes over time and whether these changes mirror those undergone by SMELL and the near-synonyms that designate it. Based on the data, the author considers that the development undergone by noun collocates in this category is probably related to developments in chemistry that took place during the Second Industrial Revolution. In turn, the remaining semantic categories show no major changes over time.

Finally, Chapter 7 provides a summary of the most relevant contributions of the monograph to the field of semantics, as well as some limitations of the study. Pettersson-Traba also suggests some future lines of research. For example, she considers that undertaking a cross-linguistic study of the equivalent terms of the adjectives in the synonym set in other languages from societies with similar sociocultural and technological developments would be interesting to further examine the hypotheses tested in the monograph.

I recommend Pettersson-Traba's monograph not only to those interested in historical semantics, synonymy or polysemy, but also to scholars interested in sociolinguistics. Chapters 4–6, which are the core of the monograph, constitute a very valuable source of information for those interested in making use of statistical analyses in their studies, as the chapters involve well-structured and comprehensive explanations in terms of methodology. Chapters 1–3 might be considered too long for some readers, as the author provides a very detailed review of the literature. However, given the conscientious and well-structured selection of works on the topic, these three chapters are unquestionably a useful reference for readers that might not be familiar with semantics and its historical evolution as a research field.

In short, this monograph is valuable not only for its academic relevance and interesting results, but also for its methodological explanations of the advanced statistical

analyses. Certainly, the two prestigious linguistic awards —namely, the *Book Award Aquilino Sánchez*⁵ and the *Leocadio Martín Mingorance Book Award for Theoretical and Applied English Linguistics*—⁶ that the monograph has received in 2023 constitute evidence of its high standard.

REFERENCES

- Archer, Dawn, Tony McEnery, Paul Rayson and Andrew Hardie. 2003. Developing an automated semantic analysis system for Early modern English. In Dawn Archer, Tony McEnery, Paul Rayson and Andrew Hardie eds. *Proceedings of the Corpus Linguistics 2003 Conference*. Lancaster: Lancaster University, 22–31.
- Baker, Paul. 2017. *American and British English: Divided by a Common Language?* Cambridge: Cambridge University Press.
- Cruse, Alan. 2004. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.
- Davies, Mark. 2010. *Corpus of Historical American English (COHA)*. <https://www.english-corpora.org/coha/>
- Desagulier, Guillaume. 2014. ‘Rather, quite, fairly, and pretty: Visualizing distances in a set of near-synonyms’. In Dylan Glynn and Justyna A. Robinson eds. *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*. Amsterdam: John Benjamins, 145–178.
- Divjak, Dagmar and Stefan Th. Gries. 2006. Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2/1: 23–60.
- Edmonds, Philip and Graeme Hirst. 2002. Near-synonymy and lexical choice. *Computational Linguistics* 28/2: 105–144.
- Gries, Stefan Th. and Naoki Otani. 2010. Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal* 34: 121–150.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Theoretical Prerequisites*. Stanford: Stanford University Press.
- Leech, Geoffrey. 1990. *Semantics: The Study of Meaning*. London: Penguin Books.
- Liu, Dilin. 2010. Is it a chief, main, major, primary, or principal concern? A corpus-based behavioral profile study of the near-synonyms. *International Journal of Corpus Linguistics* 15/1: 56–87.
- Pettersson-Traba, Daniela. 2021. *A Corpus-Based Study on near-Synonymy: The Concept Pleasant Smelling in 19th- and 20th-Century American English*. Santiago de Compostela: University of Santiago de Compostela dissertation.
- Rosch, Eleanor H. 1973. Natural categories. *Cognitive Psychology* 4/3: 328–350.
- Taylor, John R. 2003. Near synonyms as co-extensive categories: “high” and “tall” revisited. *Language Sciences* 3/25: 263–284.

⁵ <http://www.aelinco.es/en/i-premio-investigacion-aquilino-sanchez>

⁶ <https://aedeon.org/wp-content/uploads/listado-de-premios-actualizado-2024-marzo.pdf>

Reviewed by

Daniel Granados-Meroño

University of Murcia

Facultad de Letras

Campus de la Merced

Calle Santo Cristo

30001 Murcia

Spain

E-mail: daniel.granadosm@um.es