UMEÅ UNIVERSITY

Ambiguity at Work
Lexical Blends in an American English Web News Context

Daniel Kjellander

Department of Language Studies
Umeå 2022

Dissertation for PhD
ISBN: 978-91-7855-747-9 (print)
ISBN: 978-91-7855-748-6 (pdf)
Umeå Studies in Language and Literature 48
Cover design and title page photo: Daniel Kjellander
Author photograph: Simon Wessbo
Electronic version available at: http://umu.diva-portal.org/
Printed by: CityPrint i Norr AB
Umeå, Sweden 2022

To Madelene

Acknowledgements

I had a very vague conception of English linguistics when I began my English studies nearly fourteen years ago. Yet, I immediately experienced a strong feeling of homecoming. Little had I anticipated that I would end up writing a text of acknowledgements in a doctoral dissertation. Well, here I am. My profound gratitude to all the people that have guided and supported me through the years can barely be condensed in this limited space, but I'll do my best. There will be many exclamation marks. I do not apologize for that.

First, I would like to thank my supervisors Professor Marlene Johansson Falck and Associate Professor Maria Rosenberg for their hard work, extraordinary patience, and commitment to high academic standards. Thanks!

Thanks also to all the people at the Department of Language Studies at Umeå university! Thanks Maria Lindgren Leavenworth for your unwavering support throughout the doctoral education! Thanks Van Leavenworth for many inspiring discussions and help in the fulfillment of my teaching duties! Thanks Magnus for inflating (yep, that's the word) my poor little computer to become the beast it had to be to deal with my data. Thanks to Per (both B and A!), Karyn, Anders, Kristina, Johan, Nuno, Berit, Daniel, Elena (both G and L!), Hampus, and many others for inspiring conversations and support! You kept me floating!

A special thanks goes to Roger Mähler at Humlab, whose help was invaluable in the design of the data collection methodology! When all this is over, I look forward to improving my poor programming skills. I owe thanks to Professor Maarten Lemmens, Université de Lille, for crucial input on empirical issues at the early stages of the project. Sorry I didn't stick to "trunqites". Thanks to Dr. Mikko Höglund and Dr. Gregory Garretson for suggestions for improvement in connection to the mid-point and end-point seminars.

Finally, none of all this would have been possible without the support of my family. Thanks Mom and Dad for always being there! Thanks Simon for your support, especially towards the final stages of the writing process! Madelene! What can I say? Where would I be without you? Thanks for your constant support through the years! My boys Isak and Viktor! You are the gems in my life! Now my book is finished!

Table of Contents

List of Figures .... **iv**

List of Tables .... **vi**

Abbreviations .... **vii**

Abstract .... **viii**

1 Introduction **1**

1.1 Aims and research questions 3
1.2 On the empirical material 5
1.3 Organization of the thesis 6
1.4 Stylistic conventions 7

2 A Cognitive Linguistics framework **11**

2.1 Situating lexical blends in CL 12
2.2 Lexical meaning 13
2.3 Figurative meaning 14
2.3.1 Metaphor *14*
2.3.2 Metonymy *15*
2.3.3 Figurative meaning in research on lexical blends *16*
2.4 Language in use 17
2.5 The symbolic claim 18
2.6 Categorization 18

3 Encircling the domain of lexical blends **21**

3.1 A preliminary definition of lexical blending 21
3.2 Problems with the preliminary definition 23
3.2.1 Number of source words *23*
3.2.2 Truncation *24*
3.2.3 Issues of overlap *26*
3.2.4 Full overlap blends *28*
3.2.5 Intercalation: a special case of truncation and overlap *29*
3.2.6 Graphemic dependence *30*
3.2.7 Wordplay *31*
3.3 Further intricacies of defining blends 33
3.3.1 The semantics of composition *33*
3.3.2 Blend semantics *35*
3.3.3 Diachronic sensitivity *36*
3.3.4 The nature of blend segments *37*
3.3.5 Neighbor trouble: bordering word formation processes *39*
3.3.6 Neighbor trouble: different types of blending *40*
3.3.7 The data collection problem *41*
3.4 Overview of previous blend studies 43
3.4.1 Classificatory studies *45*
3.4.2 Structural analyses *49*
3.4.3 Psycholinguistic and experimental investigations *56*

3.4.4 Sociopragmatic studies ... *59*
3.4.5 Computational accounts ... *61*
3.5 Operational definition ... 63
3.6 Summary ... 65
4 Method ... 69
4.1 The NOW corpus ... 69
4.2 The subcorpus ... 70
4.2.1 Compiling the subcorpus ... *71*
4.2.2 Word class and bad data ... *75*
4.3 Finding the blends ... 84
4.3.1 The Candidate Words ... *86*
4.3.2 Truncation method ... *89*
4.3.3 Queries based on truncated forms ... *91*
4.4 Approaches to analyses of the blend data ... 92
4.5 Limitations ... 93
4.6 Summary ... 95
5 Identification of blend characteristics ... 97
5.1 The blend pond ... 97
5.1.1 The blend pond in numbers ... *98*
5.1.2 Duplicates ... *103*
5.1.3 Borderline cases ... *105*
5.2 Categories, domains, and series: a quantitative approach ... 108
5.2.1 Grammatical categories ... *109*
5.2.2 Semantic domain analysis ... *113*
5.2.3 Seriality from a quantitative perspective ... *124*
5.2.4 Summary of quantitative characteristics ... *130*
5.3 Blend characteristics from a qualitative perspective ... 131
5.3.1 Structural profiling ... *132*
5.3.2 Domain proximity ... *139*
5.3.3 Figurativity and iconicity of structural profiling and domain proximity *142*
5.3.4 Pseudomorphemic transfer ... *146*
5.3.5 Seriality from a qualitative perspective ... *149*
5.3.6 Summary of qualitative characteristics ... *153*
5.4 Blend characteristics in a systematically collected dataset ... 155
6 Ambiguities ... 159
6.1 Truncation ambiguity ... 159
6.2 Mode ambiguity ... 170
6.3 Source word ambiguity ... 178
6.4 Covert source ambiguity ... 184
6.5 Ambiguity as a functional characteristic of lexical blends ... 194
6.5.1 Ambiguity at work ... *195*
6.5.2 The epistemology of ambiguity ... *197*
6.5.3 Multistable meaning construal ... *201*
6.6 Summary of ambiguities in blending ... 203

7 Conclusion **205**
7.1 What blend characteristics can be identified through a systematic corpus method of data collection? 206
7.2 How is ambiguity realized in the blend data? 207
7.3 The nature of the data 208
7.4 Revisiting some central issues in blending 210
7.5 The dual model of blend classification 212
7.6 Contributions 214
7.7 Future prospects 216
Sammanfattning på svenska **219**
References **223**
Appendix 1. SQL scripts **237**
Appendix 2. UCREL Claws 7 Tagset. **241**
Appendix 3. The blend list. **245**
Appendix 4. USAS semantic categories. **254**
Appendix 5. Candidate Words. **259**
Appendix 6. Operationalized combining forms. **261**
Index **263**

List of Figures

Figure 1. AD blends.

Figure 2. Illustration of data transfer from NOW original text files to the SQL table lexicon.

Figure 4. Token distribution of the nn% tags.

Figure 5. Steps in the data collection procedure.

Figure 6. The three datasets.

Figure 7. Token frequency distribution of the 2,000 most frequent common nouns in the subcorpus.

Figure 8. Example screenshot of the dataset summary.

Figure 9. Quantitative distribution of the CaW queries output.

Figure 10. Type frequency proportions of CaWs and blends in the three datasets.

Figure 11. CaW duplicates.

Figure 12. Multiple functionalities of blends and source words.

Figure 13. Word class distribution in blends and source words.

Figure 14. Suggested semantic domains of DOUBLE BASSOON.

Figure 15. USAS main categories.

Figure 16. Semantic annotation of the concept of DEMOCRACY.

Figure 17. Proportion of USAS categories in the blend list.

Figure 18. Distribution of subdomains of the NAMES AND GRAMMATICAL WORDS category.

Figure 19. Distribution of truncation patterns in serial and solitary blends.

Figure 20. Distribution of graphemic overlap in serial and solitary blends.

Figure 21. Distribution of semantic domains in serial and solitary blends.

Figure 22. Distribution of figurativity and iconicity.

Figure 23. *Schnoodle* vs. *Addaction*.

Figure 24. *Frienemy* vs. *brainiac*.

Figure 25. Distribution of truncation patterns.

Figure 26. Alternative interpretations of standard truncation patterns.

Figure 27. Proportions of overlap in A|BC|D blends.

Figure 28. Proportions of overlap in ABD, ACD, AD, and Ax blends.

Figure 30. Fully graphemic and near-graphemic blends.

Figure 31. The qualitative phonetic difference between blends and source words in *Sportacus* and *Addaction* respectively.

Figure 32. The qualitative phonetic difference between blends and source words in *Canstruction* and *Spooktacular* respectively.

Figure 33. Representations of *tricknology* and *technology*.

Figure 34. Intercalative blends in the data.

Figure 35. Mean and median graphemic and phonemic IC ratios.

Figure 36. Schematic illustration of the influence of a covert source word.

Figure 37. Schematic illustration of the dual model of blend classification.

Figure 38. Multistable visual perception; the face-vase illustration of binocular rivalry.

List of Tables

Table 1. Stylistic conventions.

Table 2. Selected output from a lemma query of *train* in the subcorpus (10 out of 69 rows).

Table 3. Noun tags in the Claws 7 tagset.

Table 4. Example script output.

Table 5. Grammatical categories.

Table 6. Query output of the lemma *crook*.

Table 7. Examples of corpus forms without a lemma.

Table 8. Noun and verb lemmas in the subcorpus, and their corresponding occurrences.

Table 9. Selection point calculation.

Table 10. The three datasets constituting the blend pond.

Table 11. The 15 most frequent items retrieved in a wildcard query based on -*atorium*.

Table 12. Domain distribution among serial blends.

Table 13. Lexical neighborhood of *add*- and -*action*.

Table 14. Lexical neighborhood of *spin*- and -*inderella*.

Table 15. Output of *brain* and *maniac* in the USAS automatic semantic tagger.

Table 16. Serial blend formations.

Table 17. Graphemic blends.

Table 18. Frequency measures of *Compuware*, *Computrace*, and *Computrainers*.

Table 19. Wildcard query of the segment -*ania*.

Table 20. Frequency measures of *sneakernet*.

Abbreviations

CaW	Candidate Word
CL	Cognitive Linguistics
IDseq	Sequential Identification number
NOW	News On the Web corpus
PoS	Part-of-Speech
SP	Selection Point
SQL	Structured Query Language
SW	Source Word
Truncform	Systematically truncated candidate word
TextID	Text Identification number
UCREL	University Centre for Computer Corpus Research on Language (Lancaster University)
USAS	UCREL Semantic Analysis System
wID	Word Identification number

Abstract

The present study investigates the word formation process of lexical blending in the context of written US web news between January 2010–March 2018. The study has two interrelated aims. First, it aims to develop a transparent, rigid, and replicable method of data collection. This is motivated by a lack of systematicity of data collection procedures in previous research. Second, it aims to identify the characteristics of the retrieved blends; both generally and with a special focus on how ambiguity is realized. The data were collected from an offline version of the NOW corpus (News On the Web). A strict algorithm was devised to organize the data and to identify lexical blends among a large body of systematically collected word forms. Both automatic and manual procedures were employed in these tasks. The study is conducted within the framework of Cognitive Linguistics (CL). Semantic analysis is foregrounded in CL and language is considered perspectival, dynamic, non-autonomous, and experience-based. Furthermore, a Langackerian view on meaning is adopted in that symbolic potential is acknowledged in all resources and manifestations of language. Categorization is approached in accordance with the tenets of *prototype theory*, which acknowledges fuzzy category boundaries and gradual distribution of attributes.

The results of the study show that the data collection methodology is quantitatively robust, which offers the possibility to generalize the observations within the context of the chosen limitations. Consequently, the developed methodology may also be applied in future investigations. Second, quantitative analyses validate some previous assumptions about grammatical functions, semantics, and seriality in blending. Third, a set of qualitative characteristics are identified in the collected set of blends, which offers a comprehensive approach to describing blend formation in the given context. The characteristics *structural profiling* and *domain proximity* are suggested as prominent aspects of blend formation. Structural profiling is marked by prominent structural attributes such as similarity of source words and intricate patterns of amalgamation, but figurative strategies are also foregrounded. Domain proximity is described in terms of semantic similarity between the source words and an iconic relation between the fusion of structure and the fusion of concepts. The notion of *pseudomorphemic transfer* is used to capture blends that fall within the operational definition of the study but also seem to be connected to other morphological processes through the instantiation of morphological schemata. Blends clustered in series based on recycled truncated segments are revisited from a qualitative perspective, and it is claimed that the process of morphemization is likely influenced by the degree of morpheme-like character of a serially distributed segment. Furthermore, four types of ambiguity are identified in the blend data; *truncation ambiguity*, *mode ambiguity*, *source* *word ambiguity*, and *covert source ambiguity*. On the basis of the observed impact of ambiguity, it is suggested that the construal of meaning in lexical blending makes use of multistability, which is a perceptual phenomenon observed in, for instance, binocular rivalry. Taken together, the results constitute a background for suggesting a model of categorization divided into two levels of organization. This model of categorization is called *the dual model of blend classification*.

Keywords: lexical blending, data collection methodology, corpus-based investigation, ambiguity, figurativity, cognitive linguistics

1 Introduction

> Language is certainly vague and variable, but it is vague and variable in principled ways, which are at present imperfectly understood (Hanks, 2000: 208).

Words may have simple or complex meanings and forms, be borrowings or wholly new inventions. Many words rely on productive word formation patterns whereby entire words or parts of words are combined. They also occur in various stages of conventionalization; some are well-established and known to practically all fluent speakers of English, while others are newly formed expressions used only by small groups of people.

This dissertation investigates the combinatory word formation process of *lexical blending*, or *blending* for short. Words formed in this process are referred to as *blends*, or *lexical blends*, and they are often understood as the combination of two (or more) *segments* of words fused into one lexical item. A well-known example of an English blend is *brunch*, which is formed by combining the initial segment *br-* in *breakfast* and the final segment *-unch* in *lunch* (here represented as *breakfast* + *lunch*). Two other examples of blends are *smog* (*smoke* + *fog*) and *foolosopher* (*fool* + *philosopher*). The words constituting the origin of a blend are henceforth referred to as *source words*. It is generally agreed that an initial segment of the first source word and a final segment of the second source word are typically retained, but formations in which one of the source words is retained in full are often considered blends as well.

Discussions on blends sometimes use *portmanteau word* as an alternative term attributed to the author Lewis Carroll (1871) in his novel *Through the Looking Glass and What Alice Found There* (see Barnhart & Steinmetz, 1988). Carroll (1894) describes this term by saying that “[y]ou see it’s like a portmanteau – there are two meanings packed up into one word” (Carroll, 1894: 127), which should be metaphorically understood in relation to a 19th century luggage item that consisted of two parts that could be combined into one unit (see for instance Kemmer, 2003).

Blends sometimes attract attention in the media to the extent that they become buzzwords (e.g., Algeo, 1977; Lehrer, 2003, 2007; Lalić-Krstin & Silaški, 2018; Jurado, 2019). This might give them a discursive prominence possibly contributing to the assumption that blends are increasing, often in correlation with trends in popular culture and technological development (cf. Pound, 1914; Withington, 1932; Berrey, 1939; Bryant, 1974; Kemmer, 2003; Lehrer, 2007; Schmid, 2011; Danilović Jeremić & Josijević, 2019). Despite this presumed increase and popularity, blends have received little attention in linguistic research until recent decades (Kubozono, 1990; Lehrer, 1996; Kemmer, 2003; Lehrer, 2007; Renner, 2019). Hence, relatively little is known about the mechanisms behind this word formation process (Gries, 2006; Bauer, 2012; Balteiro & Bauer, 2019). Moreover, the complex nature of blends in combination with their low frequency and unclear role in linguistic organization have generally made them marginal in linguistic investigation (Kubozono, 1990; Lehrer, 1996; Kemmer, 2003; Ronneberger-Sibold, 2006; Hamans, 2010).

Blends are difficult to define and classify in relation to other word formation processes (Bauer, 1983; Cannon, 1986; López Rúa, 2004; Bauer, 2012; Beliaeva, 2014a; Renner, 2019). Some studies have addressed classificatory issues in both experimental and observational settings, the latter notably in the form of corpus investigations. For instance, Beliaeva (2014a) shows that collocational patterns (i.e., to what extent different words co-occur in context) differ between the source words of blends and so-called *complex clippings*, which are constructions formed by the initial segments of the sources (e.g., *digicam* from *digi~~tal~~ cam~~era~~*). Gries (2012) analyzes *breakpoints* (i.e., the point at which a source word is truncated) in lexical blends, speech error blends (i.e., expressions derived from conflicting concepts in speech production), and complex clippings. He shows that these categories differ significantly in terms of how their respective source words are truncated. Importantly, Gries (2012) stresses that patterns of distribution in blending are probabilistic, suggesting that reported tendencies are typically a matter of degree. This observation has consequences for analyzing blends. From the perspective of classification, it highlights a conflict between the diverse nature of blend data and efforts to describe blends in terms of clear-cut categories. Although linguistic investigation often struggles with issues of categorization (Taylor, 2003), blends thus seem to be a particularly challenging object of study (cf. Gries, 2012).

In the study of blends, the need for systematically collected empirical material can hardly be overestimated. Wulff and Gries (2019) argue that unsatisfactory data collection procedures in blend research may have "amplif[ied] certain effects, maybe because the people who identified the blends unwittingly were more likely to notice formations as blends if they exhibited the hypothesized structure" (Wulff & Gries, 2019: 12). Thus, there is a potential problem of validity in studies based on unsystematic data collection. Plausible generalizations are difficult to make, since it cannot be asserted that the data are representative. There is no reason to question the general awareness of the importance of systematicity in linguistic investigation. In blend research, the empirical gap may instead be a combination of attitudes towards blending as an insignificant linguistic phenomenon and the challenging technical problems of localizing them. Both these issues appear less encumbering as the significance of blending is increasingly recognized (e.g., Balteiro & Bauer, 2019) and as the technical development of recent decades offers new powerful digital tools. In accordance with this development, one of the central purposes of the present study is to investigate blends that have been collected using a strictly controlled procedure. Doing so may allow conclusions that enable generalization within the limitations of the data.

One way to tackle the descriptive and explanatory challenges of blending has been to broaden the investigatory scope. A growing number of studies indicate that blend formation patterns draw on multiple resources; both traditionally linguistic and more general cognitive ones (e.g., Kelly, 1998; Kemmer, 2003; Fandrych, 2008; Bauer, 2012; Gries, 2012; Lalić-Krstin & Silaški, 2018). The current investigation adopts a similarly broad approach, taking into account various cognitive perspectives, of which linguistic concerns are at the center.

1.1 Aims and research questions

The aim of this study is twofold. First, it aims to design a transparent, rigid, and replicable method of collecting lexical blends. This is important given the lack of systematicity in the data collection of much of previous research on blends (Wulff & Gries, 2019). The development of such a methodological procedure is thus a prerequisite for obtaining representative data, and the context is set to a corpus of written American English web news in the period 2010–2018. The choice of context is motivated by the observations that blends often occur in news media, and that the development of digital media on the Internet tends to promote lexical blending (e.g., Lehrer, 2007: 128; Schmid, 2011: 220). Second, the study aims to identify the characteristics of the collected blends. *Characteristics* is used here as an overarching term for any systematically distributed pattern observed in the data. The term *attribute* is sometimes applied to capture narrower and more specific patterns that are typically gradually distributed (see Taylor, 2003: 44). Central attributes of blends, such as patterns of truncation, grammatical function, and distribution of semantic domains are analyzed with the objective to compare the findings in the systematically collected data with previous assumptions (cf. Wulff & Gries, 2019). The analyses of the data also involve an attempt to discern characteristics on a more general categorical level, where several attributes may be identified.

This corpus-based study is conducted within the framework of *Cognitive Linguistics* (CL). By adopting a CL perspective on meaning, symbolic potential is acknowledged in all resources of language. Following this idea, it is assumed that if blending is specifically chosen as a communicative strategy, then this form of expression is considered meaningful *per se* (cf. Fillmore, 1982; Langacker, 2008). Thus, a lexically blended construct has conceptual implications that are specific to this word formation process, and cannot easily (if at all) be transferred to other strategies of communication (cf. Veale & Butnariu, 2010).

This investigation also aims at studying the unresolved meaning potential often observed in blends. Broadly speaking, there is often more than one simultaneously available interpretation, which is seen as an instantiation of *ambiguity*. For instance, in a blend such as *herstorical* (*her* + *historical*), the source word *her* evokes an alternative interpretation of *historical*, in which the pronoun *his* is made available in the initial segment of the seemingly unrelated lexical item *historical*. The formally irreconcilable interpretations of *his* (i.e., the male pronoun vs. the unanalyzable segment *his*-) constitute two layers of meaning existing simultaneously. Such constructions are thought of as instantiations of functionally applied ambiguity. In addition, the multiple meaning potential in a blend such as *herstorical* serves to highlight ideological concerns in a striking way (cf. Fandrych, 2008; López Rúa, 2010, 2012; Lalić-Krstin & Silaški, 2018). Therefore, the phenomenon of ambiguity is investigated as a potentially informative characteristic motivating the use of blending as a word formation process.

The means to meet the aims is to study the relations between structural, functional, and semantic patterns in the systematically collected set of blends. Characteristics potentially unique to blends are of special interest, as are questions concerning the distribution of the blends in the data. The study aligns with the view that general cognitive mechanisms are essential clues to disentangling some of the problematic issues of blending (cf. Kelly, 1998; Kemmer, 2003; Gries, 2012; Silaški & Đurović, 2013; Beliaeva, 2014a; Jurado, 2019). In this way, the investigation of blend characteristics and attributes takes a broad approach to the analyses of the data and acknowledges that there may be several interacting cognitive mechanisms at work in lexical blending.

The study is divided into two main parts. The first part aims to characterize the data in general terms, with the objective to arrive at statistically representative descriptions and conclusions. The second part is specifically oriented towards the ways in which ambiguity is applied in the data. These two lines of investigation attempt to answer the following research questions:

- What characteristics of blends can be identified through a systematic corpus method of data collection?
- How is ambiguity realized in the blend data?

1.2 On the empirical material

The study is based on written American English web news from January 2010 to March 2018. This temporal scope spans 99 months. The chosen delimitations are based on three main arguments. First, the decision to focus on news is motivated by the assumed strong communicative incentives of this type of media, and by reports of a relatively high degree of occurrence of blends in news media context (Cannon, 1986; Lehrer, 2007; Danilović Jeremić & Josijević, 2019). Second, it has been observed that lexical creativity thrives in the domain of electronic, or digital, communication (López Rúa, 2007), and blending would thus be potentially productive in such a domain (see also Schmid, 2011: 220). Third, due to the linguistic and cultural diversity of English, investigating an English language phenomenon without any limitation as to regional variety would question the representativity of the findings. Thus, the choice to study an American English context is a motivated delimitation based on variety, but also because the impact of American English on cultures worldwide makes it an important object of study (Fandrych, 2008). With these considerations in mind, the *News On the Web* corpus (NOW) (English-Corpora, n.d.) was selected as the empirical source. The chosen temporal scope of NOW reflects the intention to collect relatively recent data, while it also afforded a sufficient size.

The data collection procedure started with the retrieval of three independent lists of 100 randomly selected words each. Two of these lists were based on the 2,000 most frequent common nouns, while one was based on the 2,000 most frequent verbs. This division allows comparisons evaluating the statistical robustness of the data collection procedures. Limiting the data to nouns and verbs is motivated by an intention to collect data from attestedly large categories in order to direct the attention towards relatively conventionalized source concepts. In addition, it has been argued that blends tend to prefer nouns as source words, while verbs are uncommon (e.g., Cannon, 1986; Elisa, 2019; Renner, 2019). Hence, including verbs as potential sources may have a validational function with regard to their assumed rarity in blending, since the data of this investigation is systematically collected.

The randomly selected words were intended to constitute the basis for the retrieval of blends, and their preliminary status was to constitute potential source words. Three parallel datasets were established following an identical collection and organization procedure. The randomly selected potential source words, referred to as *candidate words* (CaWs), were then truncated systematically, which resulted in lists of segments, named *truncforms*. The truncforms were subject to queries generating word lists, from which blends were extracted manually, based on a set of qualitative criteria. The rigor of this data collection procedure is motivated by the intention to develop a technique to collect blends in a systematic manner, thus addressing this issue in previous blend research (cf. Wulff & Gries, 2019).

1.3 Organization of the thesis

Following this brief introduction, chapter 2 describes CL, which is the theoretical framework of the investigation. Central tenets of the broad approach of CL are presented in chapter 2 together with concerns that are particularly important for the study of blends. The concepts of meaning and categorization are given special prominence in the interface between CL and the word formation process of lexical blending.

Chapter 3 addresses previous research and the challenges of delimiting blends. A preliminary definition based on comparatively uncontroversial assumptions is provided in section 3.1. This definition serves as a background to further discussions on previously observed intricacies of lexical blends (3.2). It is shown in these discussions that the seemingly straightforward criteria of the preliminary definition are often problematic. Issues that are less closely connected to the preliminary definition – while yet important for the understanding of blending as a word formation process – are dealt with in 3.3. Semantic, diachronic, and categorical concerns are discussed here, but the problem of collecting blends for research purposes is also addressed (3.3.7). Once the phenomenon of lexical blending has thus been outlined, a thematic overview of previous studies is given in 3.4. One of the important objectives of the overview in 3.4 is to describe the development of methodologies to investigate blends, which is a key concern of this dissertation. Chapter 3 concludes by an operational definition of lexical blends in the current investigation, which is needed as a selectional tool in the systematic collection of blend data (3.5).

The methodology of the study is described in chapter 4. First, it presents the empirical source of the data, the NOW corpus. Then the following sections (4.2 and 4.3) deal with the funneling process leading to the list of blends that constitutes the primary material of the study. The rationale behind the analyses of the data are accounted for in 4.4, and the limitations of the methodology are discussed in 4.5.

The analyses and results of the investigation are presented in chapters 5 and 6. Chapter 5 is divided into three types of analyses and a brief concluding comment. The first one assesses the nature of the data and accounts for issues of delimitation (5.1). The second type of analysis is carried out from a quantitative perspective (5.2). Here, word class, semantic associations, and the occurrences of blends in series of analogically formed items are given special attention. The third type of analysis takes a qualitative approach to the blend data (5.3). This section describes a set of overarching characteristics that influence the data. The chapter is summarized in 5.4.

Chapter 6 addresses specifically the issue of ambiguity in lexical blends. Four types of ambiguity are described in one section each (6.1–6.4). A more general approach to ambiguity is discussed in section 6.5, in which the observations in 6.1–6.4 are combined with a broader perspective on ambiguity. An alternative model of classification is suggested in 6.5.2, and the findings of the current investigation are connected to observations of multistable perception in 6.5.3. Section 6.6 briefly summarizes the investigation of ambiguity.

The dissertation is concluded in chapter 7, in which previous research is connected to the observations of this study. The research questions are revisited and answered in the two first sections (7.1–7.2). An extended discussion of the data collection procedure is given in 7.3. In the final sections the findings of the present study are connected to previous research, and potential ways ahead are suggested.

1.4 Stylistic conventions

Terms are generally represented in italics when introduced, while conceptual domains thought of as networks of associations are given in small caps. Examples of lexical blends and their respective source words are presented according to the formula *blend* (*source* + *source*). When the blend examples occur within brackets, they are separated from the source words with the preposition *from* (e.g., *blend* from *source* + *source*). On a few occasions, a specific part of an expression has been underlined to signal that it requires special attention. Blends are usually defined only the first time they appear in the text, but at some occasions the definitions have been repeated for clarity. Blends functioning as proper nouns are capitalized.

In cases where truncation needs to be explicitly described, it is signaled by crossing out structure, as already seen in the *digicam* example (*digi~~tal~~* + *cam~~era~~*). Truncated segments are indicated using a hyphen (e.g., *digi*-). The position of the hyphen shows whether the segment is initially or finally derived. Thus, the final position in *digi*- means that the segment has an initial position in the source word *digital* and in the complex clipping *digicam*. Truncation is also connected to the notion of breakpoint, which divides a construction without indicating that a certain part should be omitted. Such breakpoints are marked with vertical bars (e.g., *br|unch* or *st|op|tion*).

The truncforms used in the data collection queries are stylistically distinguished from truncated blend segments. The truncforms are fragments derived from a frequency-based algorithm used in the systematic collection of blends.

Truncforms are indicated with an asterisk (e.g., *routi** derived from *routine*). In contrast, the broader concept of *segment* refers to the part of a lexical item that remains once structure has been omitted from a source word. Segments are marked with a hyphen (e.g., -*unch* derived from *lunch*).

The programming language SQL (Structured Query Language, described in 4.2.1) used in the collection and management of the data makes use of technical terms called *commands*. They are represented in bold small caps, for instance **SELECT**.

The three separate datasets in the data collection procedure are stylistically indicated. Their abbreviations are represented in small caps, and the two noun datasets based on identical sources of sampling are distinguished with numeric indices (e.g., NN_2). The stylistic conventions are summarized in Table 1.

Table 1. Stylistic conventions.

Function	***Stylistic device***	***Example***
Description of source words (I)	Italicized items within brackets	(*breakfast* + *lunch*)
Description of source words (II)	Italicized items within brackets	(*brunch* from *breakfast* + *lunch*)
Description of truncation	Crossed out structure in italics	*cam~~era~~*
Breakpoint	Vertical bar	*br|unch*
Truncform	Asterisk replacing removed structure	*routi**
Indication of direction of a truncated segment	Hyphen positioned in the place of removed structure	*cam-*
SQL command	Bold small caps	**SELECT**
Dataset	Small caps with numeric indices	NN_2

The mathematical convention to reduce fractions such as 3/6 to 1/2 is not applied in this text. The reason for this is that fractions often represent a sequence of graphemes or phonemes in the context of this investigation. As an illustration, the segment -*uce* constitutes 3/6 (three out of six graphemes) of *reduce*. Expressing this fraction as 1/2 may obscure the direct relation between the graphemic and numerical representations. For instance, if the segment -*uce* is compared to -*duce*, they are described numerically as 3/6 and 4/6 of *reduce* respectively, which is assumed to be easier to process that 1/2 vs. 2/3. Retaining the original number of total graphemes (i.e., the denominator 6 in the case of *reduce*) is, thus, assumed to make such comparisons more transparent. For reasons of legibility, fractions are represented numerically, while whole numbers 1-10 are spelled out except in references to chapters and sections.

Examples from the NOW corpus are rendered exactly as they appear in the so-called *expanded context* of the corpus interface. The expanded contexts consist of 200-300 words among which the queried items occur. The only exceptions to this stylistic convention are conventional uses of square brackets[1] and omission of spaces in quotes (i.e., "quote" instead of " quote ").

Finally, a brief remark should be made about the phonemic analyses of the investigation. The phonological representations are based on an assessment of suggestions provided by the three online dictionaries Merriam-Webster (n.d.-a), Cambridge Dictionary (n.d.), and Macmillan Dictionary (n.d.). Prosodic markers are not included in the representations unless this is needed for explanatory purposes. As an example, the blend *brunch* is represented as /brʌntʃ/.

[1] Square brackets are used in, for instance, ellipsis ([...]), changes between upper and lower case (e.g., [*A*]*nother* instead of an original *another*), or alternative punctuation (e.g., [,]).

2 A Cognitive Linguistics framework

The theoretical framework of the dissertation is Cognitive Linguistics (CL). In essence, CL sees all aspects of language as symbolic, and no distinct boundary is acknowledged between language processing and general cognitive abilities and processes. This position is captured in the claim that linguistic meaning is "encyclopedic and non-autonomous" (Geeraerts, 2006a: 5) with respect to other types of knowledge. Linguistic meaning is described in CL as deeply experience-based (e.g., Johnson, 1987; Langacker, 2002; Gibbs, 2003; Langacker, 2008), which has been demonstrated in metaphor studies displaying the close connection between linguistic expression and bodily experience (e.g., Lakoff & Johnson, 1980; Lakoff, 1987; Gibbs, 1992; Taylor, 2002; Gibbs & Colston, 2006; Gibbs & Van Orden, 2010). Moreover, a lexical item is usually connected to a set of conventionally determined associations, but its meaning is also open-ended in the sense that any type of experiential knowledge is activated as well (Langacker, 2008: 39). In general, linguistic meaning in CL could be described in terms of "dynamic, context-sensitive cognitive construal" (Fisher, 2010: 45).

Furthermore, CL stresses that "language is all about meaning" (Geeraerts, 2006a: 3), which means that a semantic perspective is central in CL studies. This does not mean that other levels of linguistic analysis are disregarded, but that resources of various kinds contribute to establish meaning. The interrelatedness between different aspects of language is profiled in this view, and the categorical discrimination between different language functions is thought of as matters of degree (Langacker, 2008). Therefore, the term *semantic* is used in a broad sense, including a wide range of functions and resources that combine to construe meaning. This is particularly important in a study on blends, since it has been shown that their categorical distributions involve fuzzy conceptual boundaries (López Rúa, 2004; Bauer, 2012; Renner, 2015) and that they draw on several cognitive and socio-pragmatic resources (Kelly, 1998; Fandrych, 2008; Lalić-Krstin & Silaški, 2018).

Taking a CL perspective implies a focus on *language in use* (Rosch, 1973, 1978; Tuggy, 1993; Langacker, 2002; Taylor, 2002, 2003; Tomasello, 2003; Geeraerts, 2006c; Tomasello, 2006; Langacker, 2008; Tomasello, 2008; Glynn & Fisher, 2010; Schmid, 2015, 2020). At the heart of this approach lies the claim that "language is a complex-adaptive, dynamic system: language is not stable, but rather it is continually adapted and changed in unpredictable ways by the activities of language users" (Schmid, 2020: 10), The focus on language in use also aligns with a non-autonomous view on language in that experiential and socio-pragmatic types of knowledge and activity are taken to influence linguistic structure. The basic tenets described above, and their consequences for a study on blends, are discussed in more detail in the following subsections.

2.1 Situating lexical blends in CL

Blends are generally dealt with within morphology and word formation research. However, they have usually been seen as marginal usage phenomena, and have been left aside by several of the main trends of linguistic research in the 20th century (Fandrych, 2008). Along the same line of reasoning, blends have often been considered creative exceptions with little import for general theories of language or for accounts concerned with linguistic structure (Lehrer, 1996; Kemmer, 2003; Ronneberger-Sibold, 2006). There are recurring reports attesting to the need for novel, or at least expanded, theoretical and methodological approaches to further develop the field of blend research (e.g., Cannon, 1986; Kelly, 1998; Fandrych, 2008; Bauer, 2012; Arndt-Lappe & Plag, 2013). For instance, investigations in recent decades highlight the impact of general cognitive mechanisms (Lehrer, 1996; Kelly, 1998; Gries, 2006; Jurado, 2019), the influence of lexical structure (Gries, 2004a, 2004b, 2006), pragmatic implications (Fandrych, 2008; López Rúa, 2010; Lalić-Krstin & Silaški, 2018), semantically explained effects (Lehrer, 1996, 2003; Beliaeva, 2014a), phonological constraints (Kubozono, 1990), and prosody as important sources of influence on blend formation (Arndt-Lappe & Plag, 2013).

A CL conception of meaning encompassing a wide array of cognitive mechanisms therefore seems well suited for the study of blends (cf. Kemmer, 2003; Frath, 2005; Gries, 2006; Silaški & Đurović, 2013). Indeed, a CL approach may even prove to be a "powerful kind of device that can accommodate these often unpredictable items" (Cannon, 1986: 748). Previous studies of blends within CL have reinforced this assumption (e.g., Kemmer, 2003; Gries, 2012; Beliaeva, 2014a; Lalić-Krstin & Silaški, 2018; Wulff & Gries, 2019), and this investigation intends to contribute further to this development.

The CL theory *conceptual blending* (Fauconnier & Turner, 2002) is sometimes mentioned in accounts of lexical blending (e.g., Kemmer, 2003; Schmid, 2011; Silaški & Đurović, 2013). In brief, conceptual blending addresses abstract processes of conceptual integration, whereby the fusion of two mental spaces results in a third blended conceptual space. Conceptual blending as presented in Fauconnier and Turner (2002, 2006) has been employed in several studies to capture the semantics of lexical blends (Kemmer, 2003; Brdar-Szabó & Brdar, 2008; Beliaeva, 2014a). Besides being applied in research on lexical blends, its broad theoretical scope also makes it a suitable framework for a wide variety of topics addressing conceptual fusion. It is, for instance, applied to research on compounding and metaphor (Taylor, 2002; Ungerer & Schmid, 2006). Despite a relatively clear distinction between conceptual and lexical blending, they share focus on a process whereby two conceptual entities combine to form a third conceptual entity (cf. Ronneberger-Sibold, 2006; Beliaeva, 2014a; Renner, 2015).

It is worth stressing that the term blending is used in several contexts applying different definitions. In this study, the terms blend and blending refer consistently to lexical blending. Examples of other uses are the aforementioned name of the theory *conceptual blending* (Fauconnier & Turner, 2002) and in the phenomenon *syntactic blending* (Algeo, 1977). If other uses of blending are intended, this is explicitly stated (see also 3.3.6 for a detailed discussion).

2.2 Lexical meaning

The adopted perspective on lexical meaning is based on Langacker's (2008: 49) description of domain activation. In short, networks of knowledge and associations called *cognitive domains* are activated by a lexical item (e.g., *chair*). Cognitive domains normally occur in complex matrices consisting of multiple interacting domains, and there are no inherent limitations as to what types of conceptions may be included (Langacker, 2008: 44). For the purposes of the current investigation, the term domain is thought to encapsulate the complexity expressed in the notions of domain matrix and complex domains. Hence, domain is used throughout the account as a term ranging from relatively restricted concepts such as spinnaker to (presumably) more broad concepts such as love. Such domains are seen as encyclopedic in that they relate directly to experiential knowledge, although they may be characterized by a high degree of conventionality in language use (cf. Schmid, 2020).

The activated domains have varying degrees of conceptual prominence, or *centrality* in the construal evoked by a lexical item. For instance, in *chair* a central activated domain may be the conception of A PIECE OF FURNITURE used as a seat, while a domain such as MEMBERSHIP IN AN ORCHESTRA is equally available but typically less central. Hence, lexical items such as *chair*, *love*, and *blue* function as "points of access to a network" (Langacker, 1987: 163), the latter expressed here as *cognitive domain* or *conceptual domain*. The activation of a certain domain via the use of a lexical item often follows a conventional pattern, but possible associations are open-ended, and depend largely on the needs of the language users. Therefore, the nature of cognitive domains in relation to the principles of activation makes lexical meaning a matter of balance between stability and flexibility (Langacker, 2008: 49).

In the domain activation model, the construal of meaning is sensitive to cognitive processing, context, and diachronic development. For example, the meaning of *chair* depends on who construes it, in what circumstances it is used, and how it can shift over time. Thus, it is not possible to uphold a sharp distinction between a decontextualized lexical item and how it is used, which can be described as the relation between semantics and pragmatics. Langacker (2008) argues that these analytic levels "form a gradation [...] with no precise boundary between the two.

But toward either extreme of the scale lie phenomena that are indisputably either semantic or pragmatic" (Langacker, 2008: 40).

Observations of context dependence in blending (e.g., Lehrer, 2007; Lalić-Krstin & Silaški, 2018) make the domain activation model suitable for investigating the semantics of blends. Indeed, the findings in previous research concerned with the socio-pragmatic qualities of blends highlight the importance to incorporate pragmatic meaning on a theoretical level (cf. Ronneberger-Sibold, 2006; Fandrych, 2008; Lalić-Krstin & Silaški, 2018).

2.3 Figurative meaning

The roles of metaphor and metonymy have been extensively studied in CL (e.g., Lakoff & Johnson, 1980; Lakoff, 1987; Gibbs, 1992, 1994; Taylor, 2002; Cameron, 2003; Gibbs, 2003; Taylor, 2003; Geeraerts, 2006c; Cameron & Maslen, 2010). An important tenet in CL is that metaphor and metonymy are intrinsically conceptual in nature (e.g., Lakoff, 2006), and that they are integral strategies to construe meaning in language (e.g., Taylor, 2003; Geeraerts, 2009). Metaphorical and metonymical meanings are subsumed under the idea of *figurative meaning* in the context of this investigation. Although a clear demarcation line cannot be drawn between metaphor and metonymy (e.g., Goossens, 1990; Barcelona, 2003a; Geeraerts, 2009), they are commented on separately for reasons of clarity.

2.3.1 Metaphor

The understanding of metaphor in this study follows the common CL description of metaphor as "mappings across conceptual domains" (Lakoff, 2006: 232), often realized as aspects of a concrete source domain (e.g., ICE) mapped onto an abstract domain (e.g., SOCIAL INTERACTION). For instance, an expression such as *break the ice* is a recurring way to express a social process whereby an interactional obstacle is overcome (example 1).

(1) Some people find it easier to break the ice on line [Sic.] rather than meeting in person for reasons that only they know (NOW).

There is nothing in this sentence that reasonably licenses a literal interpretation of actually breaking some piece of frozen water. Instead, *break the ice* is contextualized in (1) as an abstract social process that may be realized in an online context or meeting in person. Example (2) is an even clearer instantiation of a metaphorical construal of *breaking*. Because words are abstract symbols, they cannot be used to break ice in a literal understanding of the example.

(2) And there are suggestions of the actual words you can use to break the ice (NOW).

Both examples (1) and (2) are understood in the present context as instances of conceptualization of "one mental domain in terms of another" (Lakoff, 2006: 185). In this study the mapping is also understood as an instantiation of experiential meaning based on the bodily experience of ice together with sensory memories of actually breaking it (cf. Lakoff, 1987; Gibbs, 2003). Lakoff's (2006) designation of a mental domain corresponds roughly to the notions of domain matrix or complex domain as outlined in the introductory paragraph of 2.2. (cf. Taylor, 2003: 89). Concepts such as ICE and SOCIAL INTERACTION are here seen as complex domains although it is fully possible to conceptualize them as networks consisting of several subdomains in a matrix. In accordance with the outline in 2.2, domain is the applied term for ICE and SOCIAL INTERACTION in this investigation (see Goossens, 1990: 325 for a similar approach).

2.3.2 Metonymy

Metonymy has been described in CL both in terms of a mapping within a conceptual domain (e.g., Lakoff, 1987) or as semantic extension based on conceptual contiguity (e.g., Peirsman & Geeraerts, 2006). The former perspective on metonymy is chosen here, as it aligns well with Langacker's (2008) term *domain* as a basis for meaning construal (see also Barcelona, 2003b; Croft, 2006).

The understanding of metonymy as a type of mapping can be applied to example (3) in that the conceptual domain THE WHITE HOUSE is seen as a part of the conceptual domain US PRESIDENTIAL ADMINISTRATION. Since only a part of the intended conceptual domain is used to designate the entire domain, this type of metonymy is usually referred to as a PART FOR WHOLE metonymy (e.g., Lakoff & Johnson, 1980; Kövecses & Radden, 1998; Barcelona, 2003b; Ruiz de Mendoza, 2003).

(3) The White House says the designations will nearly double the amount of public land (NOW).

Analyzing the metonymic meaning in (3) as an extension based on contiguity instead implies a focus on the adjacency of the concepts of the White House and the US presidential administration. In this view, one entity (*the White House*) refers to another entity (*US presidential administration*), both of which are interconnected by association (cf. Taylor, 2003: 125). The PART FOR WHOLE configuration seems somewhat less prominent in (3) than in a metonymy such as ALL HANDS ON DECK, in which *hands* is taken to be an intrinsic part of a person and the human body. The conceptual contiguity of the two concepts in a metonymic expression may thus vary in degree (Peirsman & Geeraerts, 2006). Both the domain and contiguity approaches accord with a CL perspective on meaning (Geeraerts, 2009), although the analysis of mappings within a domain is often considered the standard view of metonymy in CL (Barcelona, 2003b).

Metonymy typically lies behind several types of mapping. Besides the aforementioned PART FOR WHOLE metonymy, other recurring types of mapping are PART FOR PART and WHOLE FOR PART (Kövecses & Radden, 1998; Barcelona, 2003b). More categorically specific types are, for instance, CONTAINER FOR CONTAINED (i.e., *glass* for *beverage* in 4), PRODUCER FOR PRODUCT (i.e., *Ferrari* for *car* in 5), and BODY PART FOR THE WHOLE PERSON (i.e., *hands* for *people* in 6) (see for instance Lakoff & Johnson, 1980; Barcelona, 2003b; Geeraerts, 2009).

(4) Ted ordered another glass with his steak (NOW).

(5) I wouldn't hesitate to bid on this beautiful Ferrari (NOW).

(6) Obama said all hands are required to solve the world's newest threats (NOW).

Even a traditional blend such as *brunch* may be analyzed in terms of a metonymic mapping. The amalgamation of the conceptual domains BREAKFAST and LUNCH presumably requires a selection of attributes that combine in the blend. While the temporal frame may be adjusted to a compromise between morning and noon, other associations are likely a matter of cultural negotiation. Associations that might be relevant could be dishes on the menu, expected level of formality, and spatial constraints. At the very least, it seems likely that only some of the associations connected to each source word are retained in the blend (hence a WHOLE FOR PART metonymy; cf. Kemmer, 2003: 84), but a contextual analysis typically adds further valuable information about what associations are foregrounded.

2.3.3 Figurative meaning in research on lexical blends

Figurative meaning in lexical blending has been noted by Kemmer (2003) in an analysis of the blend *glitterati* (*glitter* + *literati*). This blend "refers to a particular kind of elite group, socially sought after, who metaphorically and metonymically 'glitter': in clothes, jewelry, and more abstractly, beauty, wealth, fame and prospects" (Kemmer, 2003: 85). Hence, GLITTER may be a literal sub-domain of JEWELRY (cf. Barcelona, 2003b), which makes the *glitter* of *glitterati* an instance of a PART FOR WHOLE metonymy in which the sub-domain stands for the entire domain. At the same time, the literal concept of GLITTER is also mapped metaphorically onto abstract conceptual domains such as BEAUTY and WEALTH.

Renner (2015) mentions figurative meaning in connection to wordplay in both compounding and lexical blending. He refers to items such as *brain drain* as "metaphtonymic echo compounds" (Renner, 2015: 120), which highlights the difficulty to determine whether the expression reflects metonymy, metaphor, or both (cf. Goossens, 1990). Renner (2015) suggests that the French blend *midinette* (Fr. *midi* 'noon' + dînette 'light meal', meaning *dressmaker's apprentice*) instantiates such metaphtonymic meaning.

Jurado (2019) assumes that the first source word of blends such as *heartgasm* (*heart* + *orgasm*) and *bookgasm* (*book* + *orgasm*) are used metonymically. *Heart* is described as a metonym for LOVE and *book* refers to the activity of reading (Jurado, 2019: 17). More generally, objects are claimed to refer to activities in blends such as *eargasm* (*ear* + *orgasm*), where *ear* stands for the activity of hearing.

The studies by Kemmer (2003), Renner (2015), and Jurado (2019) mentioned above exemplify an awareness of metaphor and metonymy as resources for construing meaning in lexical blends. The centrality of figurativity in CL implies that blend research within this theoretical framework could benefit from further investigation of metaphor and metonymy. In order to address figurative meaning in blends, contextual analysis is central to identify how blends are construed in actual language use.

2.4 Language in use

The analyses of the investigation take as their starting point the assumption that the patterns of use characterize the blends in the data. This means that observations of frequency and collocational patterns are considered in the analyses, and that contextual analyses play an important role.

The emphasis on language in use should be understood from the perspective of contrasting views on language in 20th century linguistics. In frameworks inspired by Saussurean structuralism (Saussure, 1966) and Generative grammar (Chomsky, 1968) it has been common (and still is) to separate an abstract language system and the competence it relies on from the way it is used, or performed (Tomasello, 2003; Geeraerts, 2006c). Central to the usage-based approach applied in this investigation is the view on linguistic organization as a *bottom-up* endeavor rather than *top-down* (Langacker, 2002; Schmid, 2015, 2020). The bottom-up approach holds that linguistic patterns, including various types of constructions, are the result of generalizations, or schematizations, based on analogical processing of exposure to natural language (Goldberg, 2005). The term *entrenchment* (see for instance Schmid, 2015, 2020) is often used to capture the effect of frequency of occurrence in the establishment of linguistic structures through a bottom-up process.

2.5 The symbolic claim

It may seem uncontroversial that lexical items such as *chair*, *point*, and *horse* are considered linguistic symbols for the concepts they activate, but in CL other levels of linguistic structure are seen as symbolic as well. Notably, grammar is thought to carry meaning in and of itself (Langacker, 2002; Taylor, 2002; Goldberg, 2005; Langacker, 2008; Booij, 2010). The symbolic complexity is typically not evenly distributed, which Langacker (2002) exemplifies with morphemes and grammatical classes as "symbolically minimal [while] [g]rammatical rules and constructions reside in units that are both schematic and symbolically complex" (Langacker, 2002: 292).

In lexical blending, other resources are available as possible carriers of meaning. For instance, there is symbolic potential in the ways source words are truncated as well as in the strategies used to combine such truncated segments. These types of structure are thought to be relevant for the conceptual construal of blends, although their meaning potential may be less profiled and more schematic than the conceptual content activated by the source words. For instance, the overlapping structure of the source words in *relationshit* contributes meaning that goes beyond the combination of the concepts of relationship and shit. Thus, blending as a linguistic construction is not necessarily semantically equated with possible alternatives such as compounding (possibly *relationship shit* or *shit relationship*) or genitive constructions (e.g., *the shit of relationship*). Instead, form and semantics are closely connected in blends (e.g., Kemmer, 2003: 83; Veale & Butnariu, 2010: 403).

2.6 Categorization

The problem of defining and classifying blends is addressed from the perspective of *prototype theory* (cf. Rosch, 1973, 1978; Geeraerts, 1989; Taylor, 2003), which is also a central tenet in CL. A recurring term with several implications for prototype categorization is *attribute*. Taylor (2003) contrasts this term with the notion of *feature*, which is used in an Aristotelian, or classical, approach to categorization. From the classical viewpoint, a set of criteria need to be met in order for an entity to be included in a specific category. These criteria are sometimes referred to as *necessary and sufficient conditions*. These conditions, or features, are typically categorical rather than gradual in nature, which means that they are subject to logical analysis, or syllogisms of the following type:

a) All men are mortal.

b) Socrates is a man.

c) Therefore, Socrates is mortal.

(taken from Taylor, 2003: 35)

The logic of this famous syllogism relies partly on the assumption that the features of propositions a) and b) are either true or false, e.g., Socrates is either a man or not. Attributes, on the other hand, “are not the binary constructs of the classical approach” (Taylor, 2003: 44). Instead, they are gradual by definition, and when categorizing an object “it is not a question of ascertaining whether the entity possesses this attribute or not, but how closely the dimensions of the entity approximate to the optimum value” (Taylor, 2003: 44). This perspective of categorization allows analyses of items in connection to central prototypes, and assessment of attributes based on their gradual nature.

Applying a prototype theory categorization to blends implies that classification is not based on discrete features. A definition of what constitutes a blend depends instead on a holistic assessment of its attributes. Hence, attributes may be distributed in a set of blends in such a manner that few, if any, attributes are shared by all entities (cf. Bauer, 2012; Renner, 2015). Because of this complex category designation, operationalizations with explicit delimitations are important. Furthermore, generalizations based on an operationalized model are thought of as distributional tendencies.

The operational delimitations, or the assessment of category membership, contrast with the gradual distribution of the attributes in the data. Geeraerts (2006a: 150) illustrates this distinction in a concrete manner by saying that "[m]embership in the category bird is discrete; something is or is not a bird. But some birds may be birdier than others". Thus, a prototype categorization perspective on what counts as a blend typically leads to assessments of candidate lexical items along the lines of Wittgenstein’s (1968) notion of *family resemblance* (see also Brdar-Szabó & Brdar, 2008). In this perspective, category attributes are distributed unevenly, gradually, and in associative chains. As a result, “family resemblance categories can sometimes exhibit immense structural complexity” (Taylor, 2003: 111).

The tension between prototype categorization and the requirement for systematic austerity requires carefully designed methodological models. A strict operationalization of category membership is used in this study to establish whether an item is considered a blend. This is necessary to obtain a representative sample from the corpus. At the same time, the analyses of the blends are conducted from the perspective of a potentially gradual distribution. Using other types of data (for instance investigating British English web news instead) or changing the operational definition may result in other findings. Therefore, the transparency of the method is an important factor when discussing the findings of this investigation in a broader context.

3 Encircling the domain of lexical blends

Defining blends is a difficult task (Algeo, 1977; Bauer, 1983; Cannon, 1986; Kemmer, 2003; López Rúa, 2004; Bauer, 2012; Renner et al., 2012a; Renner, 2015; Balteiro & Bauer, 2019; Beliaeva, 2019). This chapter aims at summarizing previous research on blends, and what the descriptive challenges are. The chapter is structured so as to encircle, in consecutive steps, the research field of blends. This is necessary because of the uncertainties of even basic issues. For instance, there is disagreement on the issue of whether combinations of initial segments (i.e., complex clippings as outlined in the introduction) should be seen as a type of blend. Other areas of contention concern the status of items with no structural overlap and items in which only one of the source words is truncated.

Chapter 3 begins with a preliminary definition that outlines what is traditionally considered a lexical blend (3.1). However, this definition does not sufficiently capture the complexity of blends. Sections 3.2 and 3.3 therefore deal with issues that are either derived from problems of the preliminary definition or are more generally concerned with the nature of blends.

Section 3.4 provides an overview of previous research into the categories, the nature, and the functions of blends. The investigations are organized thematically into *classificatory*, *structural*, *psycholinguistic*, *sociopragmatic*, and *computational* approaches. This does not mean that any specific study discussed in this section is only concerned with one specific theme, but rather that its main approach fits best with the theme it has been connected to. The studies represent a selection of blend research intended to capture central ideas as well as controversies. Their methodological and theoretical implications are particularly important in the light of the methodological focus of the present study.

The encirclement is closed by an operational definition in 3.5, which functions as the basis for determining membership in the category of blends retrieved from the corpus (cf. Beliaeva, 2014b: 45).

3.1 A preliminary definition of lexical blending

Despite decades of blend research, there is still disagreement on how to delimit blends (Beliaeva, 2014a; Balteiro & Bauer, 2019; Beliaeva, 2019). Definitions often vary depending on theoretical standpoint, as exemplified by the lexical item *Amerindian*. The segment *Amer-* is sometimes considered an abbreviation of *American* in the phrase *American Indian* (e.g., Algeo, 1977). This interpretation has sometimes led to the exclusion of such forms from the category of lexical blends (e.g., Kelly, 1998; Dressler, 2000; Bat-El, 2006). Alternatively, *Amerindian* may be seen as a fusion of two paradigmatically related concepts

American and Indian, involving truncation of the first source word (Lehrer, 2007). Taking this position allows more readily the interpretation that *Amerindian* is in fact a blend. There are, thus, two different explanations for the same lexical item.

The morphemic status of the constituents is problematic when defining blends. Blend segments are typically distinguished from morphemes (e.g., Bauer, 1983; Kemmer, 2003; Fandrych, 2008; Correia Saavedra, 2016), and a lexical item is usually not considered a blend if both its constituents are morphemes. The nature of a blend constituent can be exemplified with a segment such as *-cation* (derived from *vacation* in the blend *staycation*). The semantic status of the segment barely allows lexical combination without reference to its source word (cf. Lehrer, 1996). There are numerous examples of blend segments with morpheme-like status (e.g., *-gate* from *Watergate*, *thon-* from *Marathon*, and *-erati* from *literati*). This property can be contrasted with a regular derivational suffix with comparatively stable semantics, such as *-ness*. The suffix needs no other operation than the instantiation of a morphological schema to form combinations such as *lightness* from *light* + *ness* and *happiness* from *happy* + *ness* (e.g., Langacker, 2002; Booij, 2010).

The *Amerindian* and *staycation* examples show that a definition of blending must rely on careful considerations. At the same time, a basic outline is needed to discuss these considerations effectively. Therefore, a preliminary definition is given below. Key characteristics of blends are covered by this definition, although there are demarcation problems with, for instance, truncation patterns and overlap. The criteria of the definition follow the ones provided in Gries (2006: 536), but they also align with the definition given in Algeo (1977: 55).

In order for a word to be considered a blend, it

(i) *should* be formed from at least two source words
(ii) *should* require at least one of the source words to be truncated
(iii) *should* imply removing a final segment of the first source word, and/or an initial segment of the second source word
(iv) *could* involve graphemic and/or phonemic overlap in the source words
(v) *could* involve wordplay

Criteria (i)–(v) capture a comparatively conventional understanding of blending. However, there are several issues connected to the criteria that make delimitation difficult, and a number of issues associated with defining blends are not addressed by criteria (i)–(v). Both these types of definitional issues are discussed in the following sections.

3.2 Problems with the preliminary definition

Criteria (i)–(v) in the previous section leave room for considerable flexibility, which may be said to reflect the heterogeneous nature of blends. However, this flexibility poses a problem for systematic investigation. It results in a definitional vagueness that challenges efforts towards coherent and consistent descriptions of blends (cf. Bauer, 2012). The following sections address recurring problems associated with the preliminary definition in 3.1.

3.2.1 Number of source words

According to criterion (i), blends usually have two source words. However, it is often pointed out that three or more source words are occasionally involved in blending (e.g., Cannon, 2000; Gries, 2012; Beliaeva, 2019). For instance, the blend *Intelevisionary* (from Lehrer, 2007: 120) combines the brand name *Intel* with *television* and *visionary*. Another example is *turducken* (from Kemmer, 2003: 72), which refers to a dish made of three types of meat (i.e., *turkey*, *duck*, and *chicken*). Both Lehrer (2007) and Kemmer (2003) describe such instances as relatively rare in their data (see also Renner, 2019: 30).

A blend can also include another blend as one of its source words (López Rúa, 2004; Lalić-Krstin & Silaški, 2018). Thus, such items may also have more than two source words. Some examples taken from Lalić-Krstin and Silaški (2018) are *Branalysis* (*Brexit* + *analysis*), *Bremorse* (*Brexit* + *remorse*), and *Brepression* (*Brexit* + *depression*). The source word *Brexit* (*British* + *exit*) in these examples is itself a blend. As long as it is analyzable as an amalgamation of *British* and *exit*, it brings its double nature to the new blends, thus accumulating the number of sources in them. In *Bremorse*, for instance, *British* and *remorse* are structurally overt. *Exit*, however, is only tacitly present in *Bremorse* given that the source word *Brexit* is recognized. In this view *Bremorse* is a construction based on the three source words *British*, *exit*, and *remorse*. Phenomena like these typically require contextual and diachronic considerations. *Bremorse* could simply be understood as a blend of *British* and *remorse*, but the implicit contextual reference to *Brexit* motivates the inclusion of *exit* (Lalić-Krstin & Silaški, 2018). Furthermore, the more conventionalized a blend becomes, the less psychologically realistic it is to assume that its source words are conceptually foregrounded in a new blended construction.

The issue of identifying source words goes even further. A polysemic, or even homonymic, potential appears to offer an available strategy to refer to several sources. A blend such as *Misstake* (example 7) illustrates the rich potential of several simultaneously activated concepts. The online dictionary Merriam-Webster provides four main definitions of *miss* (Merriam-Webster, n.d.-b). They are divided into a verb entry with ten sub-entries, two noun entries with three and four sub-entries respectively, and an abbreviation of the proper name *Mississippi*. The verb entry is oriented towards *failure* and *omission*, while the noun entries are expressions of either *failure* or *young woman*.

(7) *misstake* ($miss_1/miss_2/miss_n$ + *mistake*/*take*)

The blend *misstake* is used in many different ways. It occurs, for instance, as an artist name (MissTake, n.d.) and the name of a movie (Internet Movie Database, n.d.). Each of these occurrences can be associated with multiple meaning potentials of *miss* and *take*. This uncertainty may be used as a means to achieve an effect of unresolvability. Little, if any, effort has been made to study this type of source vagueness systematically.

Analyses of blends such as *Bremorse* and *Misstake* show that the criterion concerned with the number of source words (i) is potentially more complex than the preliminary definition may indicate. The rarity of blends such as *Intelevisionary* could possibly question the need to include items with more than two source words in a general definition of blends. However, it is shown here that a broader view on how to identify source words may imply that constructing blends with more than two source words might not be all that rare. The ingenuity of such alternative ways to form blends also stresses their potential impact on blending.

3.2.2 Truncation

The truncation of source words is one of the characteristics that sets blends apart from most other word formation processes. The criterion that requires truncation in at least one of the source words (ii) stresses the centrality of truncation in blending. Criterion (iii), which stipulates the reduction of a final and/or initial segment in the first and second source word respectively, further underscores the importance of truncation by highlighting how it is realized in blending. Truncation is sometimes referred to as *clipping* in blend research. In this study, however, the term truncation is used to avoid confusion with clipping as a word formation process (e.g., *ad* from *advertisement* or *copter* from *helicopter*). Despite considerable efforts to outline the nature of truncation, it remains a problematic issue.

The centrality of truncation is commonly acknowledged in blend research, although truncation is realized in highly complex ways. The final truncation of the first source word combined with the initial truncation of the second source word is often claimed to be the prototypical pattern (Kelly, 1998; Plag, 2003; Gries, 2004b; Hamans, 2010; Bauer, 2012; Gries, 2012; Arndt-Lappe & Plag, 2013; Beliaeva, 2014a; Renner, 2015 to name a few). This pattern is sometimes expressed schematically as AD blends (e.g., Plag, 2003; Beliaeva, 2014a). The A represents the preserved segment of the first source word AB, while D is the remaining segment of the second source word CD (see Figure 1). This representation does not take into account how much of a source word is removed, but only signals the direction of the truncation pattern of the blend. Typical examples of AD blends are *brunch* and *smog*.

A~~B~~ + ~~C~~D → AD
sm~~oke~~ + ~~f~~og → smog

Figure 1. AD blends.

Figure 1 reveals that a seemingly straightforward AD blend such as *smog* displays an ambiguity as regards truncation. A phonemic analysis of /smɒg/ may suggest that *fog* is responsible for both the nucleus and the coda in *smog* (i.e., /-ɒg/). However, the graphemic realization involves an overlap in the sharing of the grapheme *o*, which occurs in both source words. In blends with more complex patterns of truncation such as *chunnel* (*channel* + *tunnel*) and *foolosopher*, the level of structural ambiguity is even more significant (cf. Gries, 2004b). These examples illustrate the potential diversity of the AD pattern.

Although an AD structure is often considered prototypical (e.g., Kubozono, 1990; Plag, 2003; Renner, 2015), a quick glance at blends in the media reveals that it represents but one possibility (cf. Lehrer, 2007; Lalić-Krstin & Silaški, 2018). ACD and ABD patterns are reportedly recurring alternatives. In the ACD pattern the second source word is preserved (e.g., *Texodus* from *Texas* + *exodus*), while the first source word is retained in ABD blends (e.g., *Czechout* from *Czech* + *checkout*; both examples from Lalić-Krstin & Silaški, 2018:6).

A fourth pattern is evident from the much-debated structure of blends such as *stoption*, represented here as an A|BC|D pattern (see also 3.2.4). The annotation with vertical bars signals that the overlapping structure (BC) is potentially meaningful, while a regular compound such as *blackbird* would be coded without these bars (i.e., ABCD). The special status of the A|BC|D construction highlights not only whether truncation occurs or not, but also raises the issue of the nature of truncation in such blends (cf. Correia Saavedra, 2016: 58). If truncation does occur despite the preserved source word structure, then it likely requires an explanation more intricate than the permanent removal of structure (e.g., removing *l* from *lunch* in the blend *brunch*).

The intricacy of truncation patterns is also semantically related. The ABD example *Czechout*, for instance, is identified as a blend with truncated source words only from its graphemic realization (see also 3.2.6). Moreover, the fact that the breakpoint of *Czechout* coincides with the morphemic structure of *checkout* further adds to the complexity of the truncation. The morphemes *check* and *out* are combined in *checkout*, which resembles the structure of the blend. This may suggest that *Czech* and *out* are the constituents of *Czechout*, which would in principle make it a compound. However, *Czechout* is a construction originating from the context of critical movements towards the European Union (EU), in which the blend *Brexit* is the perhaps most well-known concept. The conceptual schema underlying *Brexit* and *Czechout* is the idea of a nation or a region formally leaving an overarching administrative organization. This schema is instantiated in constructions where the concept of leaving is structurally similar to the agentive concept, e.g., *Texas-exodus*, the *Czech republic-checkout*, or *Ireland-retire* (realized in the blend *Retireland* in Lalić-Krstin & Silaški, 2018). Accordingly, the analysis of *Czechout* as a blend is motivated by its sharing of a conceptual schema with other blends and not only by its formal constituent structure.

Generally speaking, there are several points of contention in blend research as regards truncation. The AD structure is perhaps the least controversial pattern, while there is disagreement on the status of ABD, ACD and A|BC|D patterns (cf. Kubozono, 1990; Plag, 2003). Furthermore, the aforementioned debate on the distinction between blends and complex clippings may be described, at least in terms of structure, as a matter of AC truncation, since only the initial segments of the source words are retained (cf. Beliaeva, 2014a).

The motivation behind truncation has not received an equal amount of attention. It has been pointed out that the tight formal integration resulting from the amalgamation of truncated source words may be iconically motivated (Kemmer, 2003; Ronneberger-Sibold, 2006; Brdar-Szabó & Brdar, 2008; Fandrych, 2008; Veale & Butnariu, 2010). Truncation has also been described as a means to elaborate transparency for various purposes (Ronneberger-Sibold, 2006; Gries, 2012). These two perspectives of truncation signal a need for further investigation on the distribution and function of truncation patterns. The overview in this section also suggests that a broader scope including semantic perspectives may contribute to a developed understanding of truncation in lexical blending.

3.2.3 Issues of overlap

Criterion (iv) in the preliminary definition states that segments may share structure in a blend, which is referred to as overlap. Most other attributes are not specific to blends in isolation but instead constitute variables that, taken together, are associated with this word formation process (cf. Bauer, 2012; Renner, 2015).

In contrast, overlap stands out as an attribute specific to blending, although its distribution varies. In combination with truncation, it connects to the iconic relation between the structural and conceptual fusion in blends. It has also been analyzed as a means to achieve wordplay (e.g., Renner, 2015). Although overlapping structures in blends may be complex, overlap is essentially a rather straightforward attribute, which can be realized graphemically, phonemically or both. However, two specific themes of overlap deserve special attention.

The first theme concerns the diverse and gradual nature of overlap (e.g., Algeo, 1977; Cannon, 1986; Renner, 2015). It is a trivial observation that there is more overlapping formal structure in *californication* (*California* + *fornication*) than in *Spinderella* (*spin* + *Cinderella*), but the realization of overlap involves additional perspectives. First, overlap is usually described as shared structure in the juncture of the (typically) two source words. Gries (2004b: 649) points out that this is a simplification. The dotted second line in the example *californication* above is intended to highlight the observation that the grapheme *a* in the final segment coincides in both source words. Similarly, the phoneme /s/ in *Spinderella* matches the phonemic realization of both source words. Thus, both *californication* and *Spinderella* display overlap that is not directly attached to the junctures of the source words. Second, overlap may be the end-point of a gradual realization of similarity employing several resources such as orthography, phonology, and prosody (Kemmer, 2003; Gries, 2004b). In addition, the phonetic realization is likely used to exploit similarity to various degrees (Kelly, 1998). The endpoint of such a continuum is full similarity, which coincides with the concept of overlap.

The other theme concerns how overlap influences and interacts with other properties of blends, notably truncation and graphemic dependence. Analyses of truncation patterns rely on the possibility to determine the point at which the source word is reduced. A blend such as *webzine* (*web* + *magazine*) is analyzed as an ABD blend, because the second source word (CD) is truncated between the second and third syllables *-ga-* and *-zine*. Hence, the resulting segment is *-zine*. In contrast, an analysis of *glocal* (*global* + *local*) allows multiple interpretations as a consequence of the overlapping structure. Each of its source words *global* and *local* potentially contribute 5/6 of the blend. Therefore, it is not possible to determine the precise origin of every grapheme and phoneme. It is here suggested that it should be considered an ACD blend, since the integrity of the second source word *local* makes this source word structurally profiled.[2] However, it is also technically plausible to analyze *glocal* as an AD blend if the contributing segments are thought to be (for instance) *glo-* from *global* and *-cal* from *local*. A

[2] A potential headedness of the right-hand source word does not contribute to the analysis in this case, as the source words of *glocal* are assumed to be in a paradigmatic rather than syntagmatic relation (cf. 3.3.1).

third option is to interpret it as an *intercalative* blend (i.e., structure from one of the source words has been intercalated into the other source word; see 3.2.5). Doing so implies a non-linear analysis in which *global* is seen as the matrix word and *local* contributes the inserted *c* or /k/. This example shows that the shared structure of *glocal* creates structural ambiguity as regards type of truncation pattern. In fact, choosing the AD interpretation involves ambiguity even within itself, since it cannot be determined whether the segment pairs are *glo-* + *-cal* or *gl-* + *-ocal*.

Graphemic blends (i.e., items that require a graphemic representation to be identified as blends, e.g., *shampagne* from *sham* + *champagne*) constitute a prime example of ambiguity caused by discrepancies between phonemic and graphemic overlap. The graphemic overlap of *Czechout* is restricted to the graphemes *C* and *ec* (disregarding the capitalization of *C* in the blend). This analysis depends on the assumption that *checkout* is the second source word (see 3.2.2), and that fragmented structure is acknowledged as overlap (i.e., *Czech-* vs. *check-*). In contrasts, the phonemic realization results in full overlap (/tʃɛkaʊt/ vs. /tʃɛkaʊt/). The blend thus reveals the full potential of overlap only once the phonemic realization is detected. This discrepancy of language modes enables the dual activation of the two conceptual domains CHECKOUT and CZECHOUT. Overlap is the discriminating factor in this process, while dual activation functions as its operational mechanism (cf. McMahan & Evans, 2018: 867).

The two issues outlined above point to a somewhat paradoxical quality of overlap. On the one hand, it seems to be relatively easy to locate and delimit overlap. On the other hand, its applications give rise to high degrees of complexity. The various suggestions in previous research of how to classify blends with overlap underscore the difficulty to account for this phenomenon.

3.2.4 Full overlap blends

The type of blend structure that fulfills the overlap criterion (iv) in the most extreme way is that in which both source words are preserved in full. Examples (8) and (9) illustrate this type of blend, in which the phonological and graphemic similarity of the source words is used to create a completely overlapping structure. The chosen term for this pattern is *full overlap blends*. If truncation is foregrounded, such blends are also referred to as A|BC|D blends (3.2.2).

(8) *stoption* (*stop* + *option*)

(9) *palimony* (*pal* + *alimony*)

Constructions of this type are generally considered typical examples of lexical blends. However, neither of the source words in *stoption* and *palimony* are truncated according to the argumentation of Algeo (1977), Gries (2006), and Beliaeva (2014a). If the requirement for truncation (criterion ii) should be strictly applied, these items would not be considered blends. The reason why they are usually seen as blends is likely due to their overlapping structure and wordplay potential.

The debate on the status of full overlap blends highlights two issues. First, full overlap blends illustrate the accuracy of a family resemblance view on blend categorization (cf. Brdar-Szabó & Brdar, 2008; Beliaeva, 2014b). A blend such as *palimony* seems to lack a key attribute, but it is commonly interpreted as a blend because it is similar to other types of blends in other respects. Consequently, the criteria in the preliminary definition need to be employed with care. The modal verb *should* in criteria (i)–(iii) suggests necessity, which is potentially problematic regarding the well-attested difficulties of classification (e.g., Cannon, 1986; Kelly, 1998; López Rúa, 2004; Bauer, 2012; Beliaeva, 2019).

Second, the truncation of full overlap blends seems more intricate than just a matter of preserving the structure of the source words (cf. 3.2.2). Claiming that no truncation occurs in full overlap blends suggests that the structures of the source words are intact (e.g., *stop* and *option* in *stoption*). However, the initial segment *st*- and the final segment -*tion* are not shared by the two source words in *stoption* but are only referable to the first and second source words respectively. The special status of these segments in the blend suggests that truncation cannot be entirely disregarded as a cognitive phenomenon in full overlap blends (Correia Saavedra, 2016).

3.2.5 Intercalation: a special case of truncation and overlap

In certain blends, one of the source words seems to be inserted into the other (e.g., *clownselor* from *clown* + *counselor*, and *Calcrete* from *calcium* + *concrete*). These items are called intercalative blends (Kemmer, 2003: 72). They are structurally connected to both truncation (criteria ii and iii) and overlap (criterion iv). Kemmer (2003) focuses on intercalation as a phonemic phenomenon and recognizes variation due to differing pronunciation conventions. For instance, *Celtadelic* (*Celtic* + *psychedelic*) is intercalative only in the case where the initial phoneme is /s/. In the current investigation, the definition is extended to both phonemic and graphemic intercalation. Subsequently, a blend such as *Spinderella* is interpreted as a phonemically intercalative blend. In contrast, *glocal* is both phonemically and graphemically intercalative.

Kemmer (2003) points out that the intercalative blends in her data are equally possible to analyze as non-intercalative. They thus follow the traditional truncation pattern of removing a final segment from the first source word and an initial segment of the second source word (e.g., *Calcrete* from *cal~~cium~~* + *~~con~~crete* instead of *~~calcium~~* + *~~con~~crete*). This double possibility poses a problem of categorization. If an item can be categorized simultaneously as X and non-X, the explanatory value of the category is questionable (Gries, 2004b: 644; Renner, 2019: 33). However, the concept of intercalation need not be either problematic or redundant. In fact, the ambiguity of intercalative blends such as *Dragula* (*drag* [*-show*/*-queen*/*-and-drop*] + *Dracula*) is a recurring phenomenon also found in other types of ambiguous blend structures, such as overlap, breakpoints and multiple source words. Still, the truncation criteria (ii and iii) in the preliminary definition are challenged by the structure of these items, especially criterion (iii) that stipulates the pattern of truncation and formal composition.

Intercalative structures hint at ambiguity as a central demarcation line between blending and other word formation processes. Intercalative blends are ambiguous and unresolvable because of their creative and complex lexical structure. In fact, they may even be functionally motivated by this mechanism, which parallels Renner's (2015) notion of functional ludicity (i.e., playfulness) and Ronneberger-Sibold's (2006) ideas regarding functionally tuned transparency.

3.2.6 Graphemic dependence

According to criterion (iv), source words of blends may overlap phonemically. In fact, the phonemic overlap sometimes results in homophony between the blend and one of its source words. Such formations depend on a written, or graphemic, representation to be identified as blends (cf. Lehrer, 1996; Renner, 2015). Hence, they are characterized by *graphemic dependence*. In speech, the blends in examples (10) and (11) (taken from Lehrer, 2007) are thought to coincide phonemically with the source words *fantasy* and *personality*. The written form instead reveals the discrepancy between the source words and the blends.

(10) *fantasea* (*fantasy* + *sea*)

(11) *pursonality* (*purse* + *personality*)

In Ronneberger-Sibold (2006) these items are called *inclusive blends*, while Lehrer (2007) refers to them as *orthographic blends*. Here, they are called graphemic blends (3.2.3). The broader term *graphemic* is chosen to emphasize that several graphemic resources may be used in these constructions. For instance, hyphenation, capitalization, and italization are attested strategies in the formation of blends. Example (12) (Algeo, 1977) illustrates how the use of hyphens can function as a distinguishing factor between the blend and one of its source words.

(12) *in-sin-uation* (*insinuation* + *sin*)

The basic assumption of graphemic dependence is that the blend and one of its source words are phonemically identical. However, such homophony does not necessarily have to arise, as seen in examples (10) and (11). Merriam-Webster makes a phonemic distinction between the consonant in *sea* /si/ and the one in the final syllable in *fantasy* /-zi/ (10). The corresponding discrimination in the Cambridge Dictionary is the length of their respective vowels (/iː/ vs. /i/). Comparing *purse* and the initial segment of *personality* (11) in a similar way shows that Merriam-Webster and the Cambridge Dictionary consider them as good as identical. Thus, there is reason to question the clear-cut designation of graphemic dependence. It may well be distributed gradually, with some blends displaying almost perfect homophony, while others are better described as near-graphemic blends (cf. Schmid, 2020: 220).

The phonemic overlap of graphemic blends is contrasted with the absence of phonemic truncation in such items. This causes a potential conflict with the emphasis on truncation in the preliminary definition. Example (12) is particularly interesting in this view. There seems to be no structural truncation in this blend, but the hyphens function instead to shift focus to an alternative reference to the *-sin-* segment of the source word *insinuation*.

3.2.7 Wordplay

It is widely recognized that wordplay is an intrinsic characteristic of blending, which is reflected in criterion (v) (e.g., Lehrer, 1996; Gries, 2006; Fandrych, 2008; Renner, 2015). However, the fact that wordplay in itself is difficult to define makes it a particularly problematic criterion.

The term wordplay is often connected to punning, in which "two meanings are overlaid on each other" (Tuggy, 2006: 178), but wordplay is achieved in several ways. Renner (2015) suggests four strategies to realize wordplay in blending: formal complexity (13), play with grammatical conventions (14), graphic play on words (15), and play with semantic relations (16). In Renner's (2015) view, formal complexity (13) is realized in many types of patterns. Overlap is a central characteristic, but several source words and intricate patterns of amalgamation (e.g., *humongous* from *huge* and *monstrous*) are also included in the notion of formal complexity.

(13) *glocal* (*global* + *local*)

Another type of wordplay is realized by playing with grammatical conventions. For instance, Renner (2015) claims that English compounds cannot be formed of two proper nouns. Thus, a blend such as (14) plays with such principles since both Brad and Angelina are proper nouns.

(14) *Brangelina* (*Brad* + *Angelina*)

Graphic play on words as a means to achieve wordplay (15) is described in Renner (2015) as items in which the blend and one of its source words are perfectly homophonous. Such blends also constitute prototypical examples of graphemic dependence as outlined in 3.2.6.

(15) *pharming* (*pharmaceutical* + *farming*)

Furthermore, he describes semantic play on words (16) as the replacement of a segment in a word with an antonymic concept. More broadly, Renner (2015) points to the gradual distribution of wordplay as means to either foreground or background the playful qualities of a blend.

(16) *underwhelm* (*under* + *overwhelm*)

Wordplay can be achieved by playing with recognizability of blends and their source words. Ronneberger-Sibold (2006) suggests that the truncation of a blend may be constructed so as to balance a playful concealment of the source word(s) with the possibility of successful decoding. If a blend is too easy to recognize, the wordplay effect is downgraded (Gries, 2012: 159). However, the punning character is also lost when the source words are truncated to the extent that the blend is impossible to decode (Renner, 2015: 129).

Yet another potential resource of wordplay is the structural similarity of the source words. In example (17) the blend is distinguished from the first source word in the fourth syllable, since the three first syllables are identical.

(17) *identikit* (*identity* + *kit*)

This delay implies a potential effect of surprise, as the nature of the blend eventually conflicts with the source word in question. If the deviating structure of the blend is similar to the corresponding source word structure (cf. /-ti/vs. /-kɪt/), there is a chance that the discovery of the blend is even further delayed, thus enhancing the lexical tease (Kelly, 1998: 587).

A common denominator of the suggestions of Kelly (1998), Ronneberger-Sibold (2006), and Renner (2015) is that wordplay is analyzed as a communicative function. It is applied in varying degrees, and wordplay strategies are often combined. Little systematic research on wordplay in blending has been conducted in the past (but see Renner, 2015), although most studies on blends attest to the centrality of this attribute. The complexity of the phenomenon in combination with difficulties to collect blend data could perhaps explain this gap in blend research. Although wordplay can be described as a key attribute of lexical blending, its usefulness as a definitional criterion as expressed in the preliminary definition (3.1) may be questioned.

3.3 Further intricacies of defining blends

Defining blends exceeds the five criteria of the preliminary definition. The following sections deal with issues that influence what is considered a blend, although these issues are not directly connected to the preliminary definition in 3.1. The first issue is broadly related to the construal of meaning in blending (3.3.1 and 3.3.2). Such issues have implications for the difference between blends and other types of constructions, but they are also relevant for the decoding and classification of blends. The second issue deals with diachronic concerns (3.3.3 and 3.3.4). This perspective addresses the low viability of blends, the gradual development of blend segments, and the processes of lexicalization that sometimes obscure the identity of an item as a blend. The third issue draws the attention to issues of delimitation as regards the fuzzy boundaries between blends and other word formation processes (3.3.5). Issues of delimitation also apply to the term blending itself (3.3.6): there are several types of blending, the borders of which are not always easily discriminated (see chapter 1). Finally, the vagueness of lexical blends has practical implications for systematic investigation. Perhaps most notably, the issue of how to ensure the systematicity of the data has emerged in recent years as an important matter to increase the descriptive precision in blend research (3.3.7).

3.3.1 The semantics of composition

Blends are sometimes analyzed in terms of *endocentric* and *exocentric* structure (Bat-El, 2006; Tarasova & Beliaeva, 2020). In contemporary word formation research, this terminological pair typically refers to headedness and semantic structure. *Endocentric* compositions are forms governed syntagmatically and semantically by the head of the composition, which, in English, is typically the rightmost component. Forms such as *kitchen towel* and *blackbird* are thus identified as endocentric because of their syntagmatic structure (modifier-head), but also because "as a whole [they] have the same semantic 'function' as their head constituents" (Olsen, 2015: 370). Put simply, a *kitchen towel* is a kind of *towel*, and a *blackbird* is a kind of *bird*.

In contrast, *exocentric* compositions do not follow the endocentric headedness pattern, neither syntagmatically nor semantically (Schmid, 2011; Olsen, 2015). Therefore, determining the syntagmatic head is not straightforward, and the semantics of the composition is not necessarily derivable from one (or any) of the source constituents. For instance, the compound *white-collar* refers to the domain of office work and is only metonymically related to its source words. Neither of the source constituents *white* and *collar* can be ascribed the label of head, and the semantics of the expression is compositionally opaque. As indicated by the Greek prefixes *exo-* ('outside') and *endo-* ('inside'), the semantic head of exocentric compositions is therefore typically found outside the literal meaning of the sources. These terms – endocentricity and exocentricity – often discussed in relation to compounding, are not directly transferable to blending, but they point nonetheless to important perspectives on the structure and semantics of blends. They also highlight the close relationship with compounding.

The definition of *endocentric blends* essentially follows the general morphological considerations described above. These blends are thought of as instantiations following English headedness rules, in which the rightmost component is the head of the blend (Kubozono, 1990; Plag, 2003; Bauer, 2006; Plag, 2015; Tarasova & Beliaeva, 2020). *Motel* (*motor* + *hotel*) is one such blend since a *motel* is a kind of *hotel*, thus displaying the headedness of the rightmost source word *hotel* and the modifying function of the leftmost source word *motor* (Bauer, 2006). Other examples of endocentric blends are *webinar* (*web* + *seminar*) and *netizen* ((*Inter*)*net* + *citizen*), which are a type of *seminar* and *citizen* respectively. Because of the *modifier-head* structure, endocentric blends are often equated with *syntagmatic blends* or *telescope blends*. These blends are sometimes seen "merely as contractions" (Algeo, 1977: 56) of already existing syntagmatic or lexical combinations. For this reason, some studies exclude them from the blend data (e.g., Kelly, 1998; Dressler, 2000).

Categorizing blends as *exocentric* equally draws on the notion of headedness, which is illustrated in Bat-El's (2006) categorization of the blends *alphameric* (*alphabetic* + *numeric*) and *tangemon* (*tangerine* + *lemon*). These blends are motivated as exocentric blends because the source words "have the same semantic status, and thus none of them serve as a head" (Bat-El, 2006: 67). This delineation of exocentricity does not focus on the idea of semantic headedness outside the scope of the source words, but rather signals that neither source word is semantically more important than the other. Another suggestion to account for blends not following the rightheadedness pattern of English composition is found in Algeo (1977), where the term *dvandva blend* is introduced as a special case of blends with non-synonymous source words in a paradigmatic relation.

Blends have also been categorized along the lines of *determinative* or *coordinative* structure (e.g., Bauer, 2012; Plag, 2015). *Determinative*, rather than *endocentric*, "emphasizes the type of relationship between the modifier and the head" (Schmid, 2011: 123), thus stressing the modifier-head relation. This designation may seem more flexible, as the less profiled semantic analysis allows for greater variation. All examples of endocentric blends above are therefore equally determinative blends. The term *coordinative* (instead of *exocentric*) explicitly focuses on the lack of headedness.

There is even greater terminological diversity in the coordinative type of blends, which have also been called *associative* blends (Algeo, 1977), *conjunctive* blends (Kelly, 1998), and *paradigmatic* blends (Bauer, 2006). *Paradigmatic* is the most recurrent term, and it is applied in this investigation to highlight the semantic relation classified as a paradigmatic one, in which the order of the two source words is as good as interchangeable (potentially producing what Algeo, 1977: 60 calls a *counterblend*). Kelly (1998) addresses source word order in conjunctive blends, and argues that "the first word represented in blends tends to be higher in frequency, contain fewer syllables, and denote more prototypical category members than the second word" (Kelly, 1998: 579).

3.3.2 Blend semantics

Previous blend research has mainly been oriented towards classification and analysis of form, while semantic concerns have not been satisfactorily explored (Gries, 2012). Still, efforts have been made to deal with issues concerning how meaning is construed in the process of blending. For instance, the source words of blends have sometimes been thought to be synonymous (e.g., Bergström, 1906). Algeo (1977) stresses that paradigmatic blends in particular are characterized by semantic relations of this sort. However, such ideas have been called into question in recent years (e.g., Brdar-Szabó & Brdar, 2008; Gries, 2012; Renner, 2015). Another topic with semantic implications concerns the observations of iconicity and sociopragmatic impact (e.g., Ronneberger-Sibold, 2006; Fandrych, 2008; Veale & Butnariu, 2010). In more general terms, Kemmer (2003) suggests that "[t]he semantics of a lexical blend is a coherent cognitive structure that selectively incorporates and integrates aspects of the semantics of the activated words" (Kemmer, 2003: 71). Her study explores meaning construal of lexical blends within the theory of conceptual blending (see 3.3.6). Thus, the semantics of blends is connected to an established and broad view on semantic issues of combinatory concepts.

The processing of blends has also been studied from the perspective of semantic priming. This line of investigation examines how the decoding of blend segments is influenced by the activation of related semantic domains (e.g., Lehrer, 1996; Lehrer, 2007; see also Schmid, 2020: 54). Analyzing semantic priming may be informative as to the management of reduced transparency of blend segments, but it is not automatically relevant for the semantic construal of the resulting blend. For instance, assuming that *swacket* is based on *sweater* and *jacket* is typically insufficient to determine whether it is a sweater with jacket attributes or vice versa. Neither does it reveal the degree of influence of the source concepts. For instance, is it a balanced mixture between the concepts of sweater and jacket, or is it a jacket with only a minor influence of sweater attributes? Another example is the semantic difference between the blends *Spanglish* (*Spanish* + *English*) and *Yinglish* (*Yiddish* + *English*). The formal structure of these blends is markedly similar in that both instantiate the schema LANGUAGE/REGION + (*E*)*nglish*. Furthermore, both these items are structurally ambiguous in that they allow an AD truncation interpretation, while they are also potentially intercalative (i.e., based on alternative insertions in *Yi*[*dd*]*ish* and *Spa*[*n*]*ish*). However, the online dictionary Merriam-Webster describes *Spanglish* as "Spanish marked by numerous borrowings from English [and] *broadly*: any of various combinations of Spanish and English" (Merriam-Webster, n.d.-c). In contrast, the same dictionary defines Yinglish as "English marked by numerous borrowings from Yiddish" (Merriam-Webster, n.d.-d). Thus, this dictionary suggests different ways to conceptualize two constructions that are strikingly similar in terms of their formal structure. The *swacket* and *Spanglish-Yinglish* examples therefore illustrate the dependence on encyclopedic knowledge to account for the semantics of blends.

3.3.3 Diachronic sensitivity

Blends are known to be particularly sensitive to diachronic development (Lehrer, 2007; Brdar-Szabó & Brdar, 2008; Hamans, 2010; Lepic, 2016). Commonly occurring instances such as (18) and (19) are usually defined as blends. However, blends often undergo a diachronic process of lexicalization as a consequence of frequent use (Arndt-Lappe & Plag, 2013; Lalić-Krstin & Silaški, 2018). This development over time may eventually obscure their nature as blends (Cannon, 1986; Lehrer, 1996, 2007). Instead, language users treat them just as any other word that is not further analyzable into its constituent parts.

(18) *motel* (*motor* + *hotel*)

(19) *smog* (*smoke* + *fog*)

Other blends occur in specialized contexts, and often fall out of use as their motivation for existence disappear (Brdar-Szabó & Brdar, 2008; Lalić-Krstin & Silaški, 2018). Lehrer (1996) even claims that "most blends are ephemeral and do not enter the language as conventional words" (Lehrer, 1996: 385; see also Cannon, 2000).

Both tendencies towards lexicalization and low viability make it challenging to study blends systematically. Items that become conventionalized potentially lose their status as blends, while those that remain highly specialized may be difficult to find.

3.3.4 The nature of blend segments

The status and terminology of blend constituents is a recurring theme in blend research. They are called segments in this study, since this term is considered relatively neutral and transparent (cf. Renner, 2015). However, there are alternative terms. The constituents may also be labeled *splinters* (e.g., López Rúa, 2004; Gries, 2006; Lehrer, 2007) or *fragments* (a term used to complement *segment* in Ronneberger-Sibold, 2006). Blend segments are typically assumed to have no or little intrinsic meaning but rely instead on successful recognition of their source words (e.g., Lehrer, 1996, 2003; Ronneberger-Sibold, 2006; Lehrer, 2007; Ronneberger-Sibold, 2012; Juhasz et al., 2016). However, it has been observed that they are highly dynamic, and that they sometimes transcend the borders to more regular morphemic processes. Such reports raise questions about how to deal with the diachronic aspect of blend segments.

Blends sometimes become the model for further constructions in serial distribution (Warren, 1990; Lehrer, 1998; Schmid, 2011; Lalić-Krstin & Silaški, 2018). As an illustration, we return to the blends based on the schema LANGUAGE/REGION + (*E*)*nglish*, e.g., *Spanglish*, *Yinglish*, *Singlish* (*Singapore* + *English*), and *Chinglish* (*Chinese* + *English*). The recurring segments in such series of constructions can be described as semiproductive items. The term *splinter* is sometimes reserved for such segments (Kemmer, 2003; Renner et al., 2012a). They are the base constituents in constructions showing some signs of productivity, but these segments do not have the independent semantics of highly conventionalized morphemes (for instance the derivational affix -*ness*). Their meaning still depends on successful decoding of the source word.

Occasionally however, such segments gradually acquire a higher level of productivity and a more marked morphemic status. They are then often referred to as *combining forms* (e.g., Bauer, 1983; Warren, 1990; Lehrer, 1998; Schmid, 2011). Such items have a lexical character, which makes them similar to free lexemes. However, they typically need to be combined with another lexeme. Therefore, they are thought to "occupy a borderline position between free and bound forms" (Schmid, 2011: 41). Some typical examples of such combining forms are -*gate* from *Watergate* (as in the *Watergate scandal*), -*holic* from *alcoholic*, and *burger* from *hamburger*. *Burger* is represented without hyphen, since the process has reached the end pole of a continuum in which it has become a regular lexeme in its own right (Soudek, 1971; Algeo, 1977; Lehrer, 2007; Booij, 2010). Once a blend segment has reached the status of morpheme, it eventually becomes unreasonable to claim that it can be the sole defining structure of a lexical blend. For instance, items such as *avocadoburger* or *teriyakiburger* function as morphologically analyzable constructions, either as instances of compounding or derivation.

The process whereby a blend segment gradually develops from a truncated source word to an item with morphemic status is here called *morphemization* (see also Frath, 2005; Hamans, 2010; Lalic-Krstin, 2014; Jurado, 2019). Other terms with relatively similar meanings are *secretion* (Jespersen, 1922; Warren, 1990) and *morphological lexicalization* (Bauer, 1983). The term morphemization is chosen because of its transparency and its analogical relation to the neighboring processes *lexicalization* and *grammaticalization*. Segments undergoing a process of morphemization typically display semantic shift (cf. McMahon, 1994; Hay, 2001). This implies a gradual diachronic change of meaning (Kemmer, 2003). The segment (or combining form) develops a more schematic semantic structure (Warren, 1990; Lehrer, 1998). For instance, items such as *workaholic* and *shopaholic* retain the concept of addiction preserved in the segment *-holic* (from *alcoholic*). The resulting meanings are *addiction to work* and *addiction to shopping* respectively. Thus, the specific concept related to the addiction to alcohol is lost in the semantic shift of *-holic*.

There is some disagreement as to the status of blend segments undergoing morphemization (Beliaeva, 2014a). Fradin (2015) suggests that they are to be seen as *fractomorphemes*, which is a term introduced by Tournier (1985). In this view, constructions based on fractomorphemes (e.g., *glitterati* and *digerati* based on *-erati*) are categorically distinct from blends, because blends are defined a priori as unique items, i.e., as forms (types) occurring only once. Subsequently, the seriality of an item such as *glitterati* is incompatible with the chosen definition of blends. However, there are several reports attesting to the fuzzy boundary between blending and more regular morphological processes. Blend characteristics such as wordplay, overlap, and source word similarity tend to continue to function in serially distributed items such as *-erati* (Lehrer, 1996; Kemmer, 2003; Lehrer, 2007; Lalić-Krstin & Silaški, 2018; Beliaeva, 2019). Therefore, addressing the interface between serial and seemingly unique constructions by adding further categories (e.g., fractomorpheme) does not resolve the issue of classification. Instead, it risks adding further complexity (Beliaeva, 2014a: 33).

Blend research has yet to account for the relation between lexical blending and morphemization (cf. Correia Saavedra, 2016). The balance between these processes is a dynamic and intrinsic part of the context of creative word formation (Lalić-Krstin & Silaški, 2018). Thus, a better understanding of this relation is necessary to increase the knowledge about how blends function with regard to other morphological processes.

3.3.5 Neighbor trouble: bordering word formation processes

It is sometimes assumed that blending is a type of compounding (Quirk et al., 1985; Lehrer, 1996; Ronneberger-Sibold, 2006; Lehrer, 2007; Ronneberger-Sibold, 2012; Plag, 2015). Olsen (2015) states that compounds are built on "the combination of two or more lexemes (roots, stems or freely occurring words) in the formation of a new, complex word" (Olsen, 2015: 364). Describing blends as a type of compound implies, therefore, that the combination of two different concepts is foregrounded. Other characteristics that separate blending and compounding are backgrounded in this general classification. The blend *motel* and its compound counterpart *motor hotel* can be given as an illustration. Their first occurrences in print are simultaneous, and date to 1925 (Barnhart & Steinmetz, 1988). They may thus be seen as two formal alternatives, of which *motel* happened to become the more frequently occurring construction.

López Rúa (2004) suggests a less hyponymic taxonomy. She focuses on parameters that afford an interpretation of blending "as a bridge between the devices of shortening and compounding" (López Rúa, 2004: 76). Furthermore, she suggests that the boundaries between these concepts are fuzzy, i.e., instantiations of lexical items can be placed along a continuum from shortenings to compounding via blending. López Rúa (2004) is, however, restricted to structural aspects of blend formation, and mainly investigates truncation.

Analyzing lexical blending as a subtype of compounding requires that the essential characteristics of the hyponym (lexical blending) are included in the hypernym (compounding). However, there are several reasons why this idea is problematic. First, blends and compounds have differing motives for formation. Beliaeva (2014a) shows that the source words of blends have specific patterns of occurrence. Kemmer (2003) points to the "global pattern-matching that allows similarities on many different dimensions to count" (Kemmer, 2003: 77). Kubozono (1990) point to the specific phonological constraints influencing blend formation. Bauer (2006) goes as far as to say that compounds are rule-governed, while blends are creativity-driven (see also Ronneberger-Sibold, 2012). These observations (among others) attest to the notion of different mechanisms underlying blending and compounding. Such reports seem difficult to reconcile with a hyponymic classification: if blends were to be considered compounds, the rule-governed characteristic would be complemented with the creativity-driven quality. However, this does not seem to be the case. The rule-creativity distinction is a matter of difference in nature, rather than one of further specialization. Second, compounds and blends differ functionally in several respects. Renner (2015: 123) argues that English compounds with coordinative interpretation are rare, while blends with this semantic structure amount to 25% of his data. This implies that the strong preference for right-headedness in English compounds (cf. Plag, 2003) is not matched in English lexical blends, which instead display considerable structural and semantic diversity. A further functional difference is identified in Kjellander (2018), which shows that compounds and blends formed from the same source words differ significantly in terms of word class distribution.

Finally, the debate on complex clippings and blends has a terminological implication that deserves a brief comment, since it may influence how complex clippings are understood. The alternative term *clipping compound* is sometimes used instead of complex clipping for items of the *digicam* type. These terms are more or less interchangeable in use, although some researchers (e.g., Bauer, 1983) have argued for a more inclusive interpretation of complex clipping. However, the choice of term may also indicate a slightly different focus. The phrases *complex clippings* and *clipping compounds* are both endocentric English compounds, which means that they are right-headed constructions. Thus, complex clipping is a kind of clipping, while clipping compound is a kind of compound. Depending on which term is chosen, the property of truncation (i.e., clipping) or the process of composition (i.e., compounding) is potentially profiled in *digicam*. The choice to use the term *complex clipping* demonstrates, accordingly, that the truncation pattern of this type of item is given prominence in this context.

3.3.6 Neighbor trouble: different types of blending

The delimitation of what is a lexical blend is further complicated by the existence of other uses of the term blending. Three neighboring applications – *syntactic*, *conceptual*, and *speech error* blending – are presented here. First, Algeo (1977) exemplifies *syntactic* blending with phrasal constructions such as *a new time low* (20). This type of fusion involves few of criteria (i)–(v) in the preliminary definition. It resembles lexical blending in terms of its amalgamation of structure and, possibly, its syntagmatic truncation.

(20) *a new time low* (*new low* + *all-time low*)

Example (21) is also interpreted as an instance of syntactic blending, but it equally exemplifies *idiom breaking* (cf. Alm-Arvius, 2012: 100), which mainly relies on the deliberate failure to meet expectations of a set phrase, thus evoking an effect of surprise. Despite an internal syntactic structure, it is often pointed out that idioms do not allow elaboration, and that they function as lexical items (Taylor, 2002: 175). Examples (21–22) highlight the fuzzy boundary between lexical and syntactic blending, by involving a lexicalized construction that combines with another conventional lexical item.

(21) *nom de net* (*nom de plume* + (*inter)net*) (from Kemmer, 2003)

(22) *friends of a feather shop together* (*birds of a feather flock together*) (from Alm-Arvius, 2012)

Second, conceptual blending is a theoretical approach spanning a broad range of phenomena of which lexical blending is but one. Confusion may arise as the narrower terminological frame of lexical blending incorporates the conceptual integration addressed in conceptual blending (see 2.1).

Third, so-called *speech error blends* are misproduced items typically resulting from conflicts of lexical activation (Lehrer, 1996; Harley, 2008; Gries, 2012). Hockett (1967) exemplifies this phenomenon with the utterance in (23).

(23) Don't shell so loud! (Hockett, 1967: 912)

The item *shell* is the unintentional blend of *shout* and *yell*. It is claimed to have been uttered by a father to his noisy children in a stressful situation, and it "is known to have been a lapse, not a witticism" (Hockett, 1967: 914).

Several 20th century accounts of blending make little distinction between the products of speech errors and the intentional amalgamation of two conceptual domains and their respective linguistic expressions (see for instance Bergström, 1906; Hockett, 1973). However, research in recent decades has shown that speech error blending differs from word formation (or intentional) blending (Gries, 2004b, 2012; Beliaeva, 2014a). This distinction does not preclude that speech error blends may later be used as intentional blends, but the distinction points to structural and semantic differences of formation between intentional blends and speech error blends.

3.3.7 The data collection problem

The nature and the quality of the data are central issues in all branches of research. Indeed, a claim of epistemic authority depends on notions such as whether the material is a representative sample, and whether the data collection can be replicated. While such principles seem straightforward, collecting data is a challenging procedure. Blend research is no exception in this respect. Indeed, in recent years the issue of data collection has emerged as a pressing concern with descriptive implications for blending (Kjellander, 2019; Renner, 2019; Renwick & Renner, 2019; Wulff & Gries, 2019).

An initial concern is related to the object of study: the complexity of blends makes them difficult to classify, which affects the selection and collection of data. The underlying and potentially circular mechanism can be expressed as a situation in which classification determines the collection of data, which in its turn forms the basis for classification. For instance, there is to this day no general agreement whether or not complex clippings should be considered a type of blend. Because there are observable differences in their patterns of formation and usage (e.g., Gries, 2012; Beliaeva, 2014a), the decision to include or exclude them has potentially significant effects on the nature of the findings. Considerable efforts have been made to remedy this issue. First, the attempts in contemporary research to identify structural patterns and tendencies have been crucial in the development towards a better understanding of blending. As more evidence concerning the characteristics of blends has been accumulated in various datasets, a more diverse, and yet clearer view of blends and bordering word formation processes has begun to emerge. Second, a broader perspective on blend research has enabled new and important insights into factors that govern blend formation. It has been shown that some factors transcend traditional (formal) types of linguistic knowledge, while they are still potentially important to disentangle some of the intricacies of blending. Third, and perhaps the key concern, the transparency of definitional criteria and methodological procedures has become a central issue. When deciding on how to define blends in an exploratory investigation, preliminary definitions need not be problematic as long as the definition and the procedures of data collection are transparent. Although classification has been downplayed to some extent in recent decades (e.g., Kelly, 1998; Gries, 2004b), it will probably remain a recurring issue as the knowledge of blends increases (cf. Beliaeva, 2019).

Sampling is another important issue in the data collection procedure. If a preliminary definition of blending is acknowledged and accounted for, the data collection still depends on procedures allowing an analysis of how the sample is related to a larger population of blends within a predefined set of limitations. For instance, although a well delimited set of blends is analyzed in López Rúa (2012), the question is how her findings apply to other similar formations within a larger scope. Moreover, the representativity of the selected sources (web pages assumed to be relevant for the topic) can also be questioned. Such sampling issues increase the risk that the retrieved blends become “opportunistic data samples” (Wulff & Gries, 2019: 1) confirming a hypothesis already at the point of selection. Given these points of criticism, López Rúa (2012) is still one of the rather few examples of blend studies that, at least to some degree, allow identification of the sources.

The situation in contemporary blend research appears slightly paradoxical. Existing datasets have been meticulously analyzed, and, therefore, a lot is known about these blends. At the same time, the origin and metacontexts of these datasets are often opaque, or at least not transparent enough to allow effective replication, which makes it hard to verify the quality of the information. Therefore, systematic data collection emerges as one of the pressing issues in contemporary blend research (see Wulff & Gries, 2019 for a further discussion).

To summarize the various data collection difficulties, the list below, while not exhaustive, intends to capture the key concerns of data collection in contemporary blend research. Points i) and ii) address fundamental epistemological notions, while points iii) and iv) are more oriented towards motivations for a reformed view on the collection of blends.

(i) *The origin of the data*: How does the data collection procedure control for parameters such as genre, speech community, and temporal aspects?

(ii) *Generalizability*: What generalization of the data is afforded in comparisons between an investigated dataset and blends in general?

(iii) *Neighboring phenomena*: How are the collected blends delimited from other related word formation processes that occur in the same data?

(iv) *Lexicalization*: To what extent are the blends in the data lexicalized, and how is the morphological status of a blend segment accounted for?

Issue (i) is addressed in this investigation by the use of a carefully designed data collection procedure. This procedure is described in detail in chapter 4 so as to be transparent and permit replication. Granted that the data are considered reliable, information on issue (ii) may emerge as a result of comparing the findings with previous research.

3.4 Overview of previous blend studies

The study of lexical blends ranges over approximately a century. Bergström (1906) and Pound (1914) are two early studies that have remained influential long after their year of publication (see for instance Kubozono, 1990; Kelly, 1998; Fradin, 2015). However, the use of blends is much older. Cannon (1986) reports examples from the 16th century such as *foolosopher* and *blatterature* (*blatter* + *literature*). Instances have also been found in “Vedic Sanskrit, Attic Greek, Old High German, Latin, French, Russian, and particularly in modern Germanic languages” (Cannon, 1986: 725). Hockett (1973) even relates blending to the emergence and development of human language, which is a suggestion that has attracted some criticism (e.g., McMahon, 1994).

Blends have often been described in terms of novelty, creativity, fashionable qualities, and increasing use (Withington, 1932; Wentworth, 1933, 1934; Berrey, 1939; Hockett, 1967; Bryant, 1974; Algeo, 1977; Kemmer, 2003). Their position in linguistics has long been uncertain, and there has been little agreement on the status of blends as a morphological phenomenon. Kubozono (1990: 1) argues that blending is a morphological process, since a certain amount of regularity can be observed. Fradin (2000) instead suggests that blends should be seen as *extragrammatical morphological operations* in contrast to regular grammatical morphology (see also Dressler, 2000; Ronneberger-Sibold, 2012). Kemmer (2003) summarizes the status of blends in the field of morphology claiming that blends "have been little more than a footnote in the study of morphology in modern linguistics" (Kemmer, 2003: 76).

From a structuralist perspective, it has been suggested that blending is a literary linguistic device with limited import for the language system (Quirk et al., 1985; Fradin, 2000; Fábregas & Scalise, 2012). Algeo (1977) says that "its very lack of subtlety probably accounts for its popularity" (Algeo, 1977: 47), and Pound (1914) calls blends "whimsical folk-etymological perversions" (Pound, 1914: 8). The increasing interest in blends in recent decades challenges such conceptions and instead claims that blending has important communicative functions (e.g., Lehrer, 1996, 2007; Fandrych, 2008; Lalić-Krstin & Silaški, 2018).

In the last three decades, the scope of blend research has broadened considerably. A few examples are mentioned here. Phonological constraints are investigated by Kubozono (1990), suggesting that syllable constituency and phonological length influence blending. Structural, contextual, and semantic implications for source word recognition are demonstrated by Lehrer (1996) in an experimental study drawing attention to blending as a *process*. Kelly (1998) explores source word order, breakpoints, and source word similarity using corpus data and broad cognitive explanations. Gries (2004a, 2004b, 2006, 2012) shows in a series of articles that blends are characterized by probabilistic tendencies of distribution. Examples of such tendencies are similarity of source words, differences between blends and complex clippings (see also Beliaeva, 2014a), and patterns behind source word order. Ronneberger-Sibold (2012) presents typologically motivated constraints in German, Farsi, and Standard Chinese, thus demonstrating the implications of crosslinguistic differences in blending. The impact of socio-pragmatic concerns is highlighted in several accounts (e.g., Lehrer, 2007; Fandrych, 2008; López Rúa, 2010, 2012; Lalić-Krstin & Silaški, 2018). The influence of prosody is investigated experimentally by Arndt-Lappe and Plag (2013), and Juhasz et al. (2016). These examples highlight the many potential perspectives on blending. However, Balteiro and Bauer (2019) stress that many fundamental issues remain largely unresolved. They point to motivational mechanisms, methodology, and systematic data collection as key issues calling for further attention.

The rapid development of digital technologies is another important factor in the expansion of blend research. Its impact is twofold. First, the emergence of the Internet and computer technologies offer unprecedented opportunities to collect and analyze data (e.g., Cook, 2012; Gries, 2012; Beliaeva, 2014a; Correia Saavedra, 2016; Roig-Marín, 2016; Fontaine, 2017). Many of the studies in the last decades would have been impossible without the use of digital tools. Second, the development of the Internet may also influence the object of investigation, i.e., language as it is used on the Internet (e.g., Schmid, 2011). It has been observed that Internet language is relatively free from linguistic gatekeeping, which increases variation and informality (e.g., Crystal, 2001a, 2011). This is exemplified by the creative use of non-standard spelling, as described in Shaw (2008) and Herring (2011). Therefore, the context sensitivity and informality of blending (e.g., Lehrer, 2003; Fandrych, 2008) imply a possibility that blends may fit particularly well in this language domain (Schmid, 2011). Furthermore, Internet language has been characterized as a written medium with qualities of spoken language (e.g. Crystal, 2001a, 2011; Herring, 2011; Cenoz & Bereziartua, 2016). This property aligns with reports of blending as a process exploiting both graphemic and phonological resources (Algeo, 1977; Plag, 2003; Gries, 2004b; Lehrer, 2007; Bauer, 2012; Gries, 2012; Beliaeva, 2014b).

The tendencies in blend research described above will be dealt with in detail in the following sections, as these tendencies constitute an important background of the current investigation. The account complements the descriptions of blends in 3.1–3.3 by offering a theoretical and methodological survey of blend research.

3.4.1 Classificatory studies

Classifying different types of blends may have been a natural starting point for blend research in the early 20th century, but this approach involves several issues that still remain unresolved (cf. Balteiro & Bauer, 2019; Beliaeva, 2019). The studies discussed in this section are selected examples of a mainly classificatory perspective. However, their contributions have implications for other viewpoints as well, such as analyses of blend structure and conditions for formation.

Bergström (1906) and Pound (1914) are two early studies of blends with considerable influence on later accounts. Bergström (1906) is mainly oriented towards speech error blends and argues that "[d]eliberate, careful speech, as a rule, does not commit any blending-blunders" (Bergström, 1906: 10). This view is reflected in the reference to language as an ideal system (*langue*) (Saussure, 1966) as opposed to imperfect natural language use (*parole*) (Bergström, 1906:

21). Bergström (1906) takes a diachronic perspective, and collects examples mostly from Swedish, Middle English and Early Modern English. The explanatory scope of Bergström (1906) is broad, including syntactic blending, which he describes as misproduced sentences. The extensive list of mainly historical examples takes the form of a prescriptive grammar of blends. There are few generalizations and explanatory attempts, while a syllable analysis constitutes an exception (Bergström, 1906: §93).

Pound (1914) is in many ways similar to Bergström (1906), but the structure of the investigation is somewhat more transparent. The text is an attempt to classify a set of collected blends, and the style is oriented towards a general commentary on the phenomenon of blending. A critical stance is taken in the introductory sections towards the negligence of blending in linguistic research. Furthermore, the widespread use of blends is pointed out by saying that ”[t]he professional writing fraternity has no monopoly, however, on the creation of amalgam words; they are popular with various other classes also” (Pound, 1914: 15). Even so, blends are described as childish language, and the text signals a prescriptive viewpoint. It is worth noting that Pound (1914) claims that the use of blends is increasing.

A more descriptive perspective towards classification is found in Algeo (1977), which is an account that can be said to have paved the way for much of the recent development of blend research. One of the central objectives in Algeo (1977) is to establish a taxonomy that clarifies the relation to other word formation processes. The data are collected from 25 publications listed in the introduction, while “others are of my own collecting” (Algeo, 1977: 47). The process whereby the blends have been selected is somewhat opaque, which is a problem shared with the employed sources (for instance Bergström, 1906; Pound, 1914; Withington, 1932; Berrey, 1939; Hockett, 1967; Bryant, 1974). In addition, several of the empirical sources either explicitly focus on speech error blends or make little distinction between speech error and intentional blends.

Algeo (1977) is divided into analyses of *structural* and *systemic* categories. There are six types of blends in the structural categories, and a seventh one including compounds and derivatives such as *meritocracy* and *squandermania*. The first blend type is that of *blends with overlapping*, which is exemplified with *slanguage* (*slang* + *language*). Algeo (1977) claims that there is no truncation in these blends, because the source words are preserved in their entirety. Graphemic blends and forms with truncation of only one source word are also included in this category if the other source word is retained in full. The second structural category is named *sandwich words* in Algeo’s terminology. These blends are of the intercalative type (e.g., *autobydography* from *autobiography* + ‘*by dog*’). The third structural category, *blends with clipping* are blends with truncation, but without overlap. Complex clippings (e.g., *agitprop* from *agitation* +

propaganda) are included in this category. The fourth category, *clipping at morpheme boundaries*, consists of items bordering on compounding on the grounds that their breakpoints occur at a morpheme boundary (e.g., *Oxbridge* from *Oxford* + *Cambridge*). *Blends with clipping and overlapping* encompass blends, whose "patterns of formation are quite varied" (Algeo, 1977: 62). It is exemplified with *motel*, *ambisextrous* (*ambidextrous* + *sex*) and *daffynition* (daffy + definition). The sixth category, *blends with imperfect overlapping*, problematizes the concept of overlap and stresses the fact that overlap is a matter of degree. The blend *chump* (*chunk* + *lump*) is taken as an example in which the vowel is described as the only instance of perfect overlap. The consonant clusters *-nk* and *-mp* are not identical but Algeo (1977) argues that they have several articulatory similarities that make them instances of imperfect overlap.

The systemic categories are mainly oriented towards semantic and functional concerns. Algeo (1977) makes distinction between so-called *telescope words* and *portmanteau words*. Telescopes are described as syntagmatic blends resulting from contraction of syntactic phrases, and their status as blends is questioned because of this quality. Structurally motivated haplologic constructions (e.g., *chicagorilla* from *Chicago* + *gorilla*) are described as a core subcategory, but the analysis also includes contractions such as *aren't* and *can't* as syntagmatic blends. Although the theoretical principle seems clear, borderline cases exist (Algeo, 1977: 61), and it is often difficult to determine the origin and syntagmaticity of the examples. Portmanteau blends or *associative blends* are instead claimed to be blends proper. The main argument is that they are parallel amalgamations of two semantically or structurally related concepts. Algeo (1977) points out semantic association between the source words as the typical case, but graphemic or phonemic similarity are also relevant characteristics. Examples from the presentation are *swellegant* (*swell* + *elegant*) and *needcessity* (*need* + *necessity*). Associative blend is used as an overarching term for blends that are typically in a paradigmatic relation. Blends with (near) synonymic source words constitute a category of their own in Algeo's taxonomy, while non-synonymous but potentially coordinative ones are called *dvandva blends* (e.g., *Hungarican* from *Hungarian* + *American*). Functionally similar source words resisting both syntagmatic and coordinative combination constitute exceptions to the paradigmatic type of associative blends (exemplified by *food* and *alcoholic* in *foodoholic*). Such items are called *jumbles* to emphasize their uncertain structural and systemic nature. *Indefinite composites* constitute an even more etymologically unclear type. Such items are seen as associative blends originating from a group of words linked together by phonemic and graphemic properties (see also Pound, 1914). Algeo exemplifies this type of item with *glop* (*glob* + *slop*), which is glossed as "viscous liquid, unappetizing food, offensive sentimentality" (Algeo, 1977: 60). He suggests that *glop* may be the result of combining segments retrieved from groups of words (e.g., *gland*, *glub*, *glass*, *glean*, *glide*, *chop*, *drop*, *flop*, *sop*, *blop*). In sum, Algeo (1977) largely continues the tradition of Bergström (1906) and Pound (1914), but the classification appears more systematically organized.

In the development of blend research, Cannon (1986) represents a methodological shift, especially as regards data collection. Three well-known dictionaries are chosen as empirical sources. This enables comparative analyses as well as replication. The database consists of 132 blends, which is smaller than in previous accounts. Despite the minor size, the possibility to refer the blends to the context of a larger corpus (13,805 words) allows quantitative analyses. Cannon (1986) outlines the complex nature of blending and discusses previous accounts critically, emphasizing the challenges of describing and explaining blends systematically. The investigated blends are addressed from the viewpoints of syllabic structure, overlap, and phonemic realization. The analyses of syllable structure and overlap follow essentially Algeo's (1977) categorization, while the analyses of phonological patterns display little systematicity.

The quantitative analyses in Cannon (1986) reveal a number of characteristics. It is observed that only 55% of the blends conform to the truncation pattern in which initial and final segments are applied in the first and second source word respectively (i.e., AD blends such as *varactor* from *var~~ying~~* + *~~re~~actor*). Full overlap blends constitute 23%, and 68% of the blends (including full overlap blends) display some kind of overlap. There is a clear dominance of nouns in the list of blends (118 nouns, 11 adjectives, and three verbs). This dominance of nouns is stronger among the blends than in the corpus as a whole (90% vs. 77.5%).

The semantics of the blends vary as to the relation between the source words. Most of the source words are in coordinate relation. They are claimed to be semantically similar in most cases, while not synonymous. The meanings of the resulting blends are argued to be non-compositional.

There are few instances of slang, obscene language, trademarks and witticisms, which may reflect the fact that dictionaries have been used as sources. Instead, 53% of the blends belong to the domain of commercial activities. Other significant domains are *science and technology* (32% of the entire blend list) and *sports, travel and entertainment* (17%).

Because of the complexity of lexical blends, Cannon (1986) expresses pessimism concerning the possibilities to account for blending "within the traditional framework of generative rules" (Cannon, 1986: 748). Instead, he calls for alternative approaches. Still, it is suggested that blends have become structurally simpler over time, and that blends are easy to create (Cannon, 1986: 745-746).

In a later study concerned with classification, López Rúa (2004) addresses blends from the perspective of prototype theory. In this study, blends are connected to abbreviations and compounds, with the intention to identify the characteristics that set blends apart from bordering types of word formation processes. This is motivated by a need to improve the precision in the definitions of blends. In brief, López Rúa (2004) categorizes lexical items in relation to a network of compounds, blends, and abbreviations. Six parameters are used in the category decisions: *(morphosyntactic) source type, pronunciation, orthography, degree of shortening, degree of phonic integration,* and *language mode*. These structural parameters form the baseline for a description in which blends are seen as "a bridge between [...] shortening and compounding" (López Rúa, 2004: 76).

In sum, this section has described some concerns of classifying blends. The perhaps most important issue is the heterogeneous nature of blends, which makes them difficult to classify into categories with distinct boundaries (cf. Bauer, 1983). This characteristic stands out in analyses of blend structure, but semantic classification seems no less challenging. Perhaps as a response to these difficulties, prototype categorization has been suggested as a viable approach (e.g., López Rúa, 2004; Brdar-Szabó & Brdar, 2008; Bauer, 2012; Renner, 2015; López-Rúa, 2019).

Another notable topic is the origin of the data. It is typically hard to determine how the blends have been collected in the studies presented here, and how well the data represent blends more generally. The suggested systems of classification inherit this property. Cannon (1986) stands out as an exception in this respect, but the use of dictionaries as sources is potentially problematic. There are numerous examples of informal and context-dependent blends: e.g., *cabarazy* (*cabaret* + *crazy*; Pound, 1914), *sarcastigator* (*sarcastic* + *castigator*; Withington, 1932), *Youngmobile* (*young* + *Oldsmobile*; Bryant, 1974), *Cocacolonization* (*Coca-Cola* + *colonization*; Algeo, 1977), *Apocalipstick* (*apocalypse* + *lipstick*; López Rúa, 2012), and *horrorgasm* (*horror* + *orgasm*; Jurado, 2019). Such blends normally do not enter dictionaries. Cannon's (1986) observation concerning the lack of witticism and slang may thus simply be a consequence of the nature of the data. All the same, the systematic approach to data collection in Cannon (1986) signals a shift in attitude towards blends as an object of linguistic investigation.

3.4.2 Structural analyses

Until the end of the 20th century, most accounts of blending are concerned with classification (Gries, 2004b, 2006). The basis for classificatory analyses is often the formal structure of blends. At this point in the development of blend research, there is a growing interest in the structural phenomena, and the analyses of structure often take a broader view than just the form of the blends. A number of publications explore distributional patterns of blend structure, and the notions of constraints and schemas gain prominence to explain how blends are formed.

It is worth noting that the increased interest in structural concerns of blending parallels two other trends. First, the 1980s and 1990s see the emergence of the digital revolution. Electronic corpora and powerful software offer new opportunities to collect and analyze empirical material in all domains of research. This is beneficial to studies on blend structure in many respects. The new digital tools allow more efficient procedures of data collection and analyses of the data. Second, the broader scope of theories and methodologies, ranging from the framework of Natural Morphology to Cognitive Linguistics, indicates the need for different analytic tools to capture the complex patterns of blending.

Kubozono (1990) investigates how phonological constraints influence blend formation. The investigated mechanisms are called *syllable structure constraint* and *length constraint*. The syllable structure constraint states that source word breakpoints occur between onset and rhyme in the structure [*onset*, *rhyme* [*peak*, *coda*]], and that the juncture of the blend is constructed to fill the available constituent slot, i.e., [*onset*-] + [-*rhyme*]. Kubozono (1990) bases the analysis on monosyllabic blends formed from monosyllabic source words (e.g., *Frax* from *Fritz* + *Max*). This limitation implies the exclusion of 20% of the original data prior to the analysis. Kubozono (1990) shows that 71% of the investigated blends follow this structural pattern if the juncture is unambiguous. The proportion is slightly higher (77%) if ambiguous junctures are counted (i.e., cases where structural overlap occurs). A comparison with a Japanese dataset suggests that the syllable constraint is operational in Japanese as well.

The length constraint is defined as a mechanism whereby the syllable length of the blend is determined by the syllable length of the second source word. Kubozono (1990) claims that nearly 80% of the blend data conform to this principle. This constraint is taken as supporting evidence in favor of the (semantic) right-headedness of English blends. Thus, Kubozono's (1990) study contradicts accounts that see endocentric blends as a peripheral type, or even reject their status as blends proper (cf. Algeo, 1977).

Kubozono (1990) argues that blending should be seen as a morphological phenomenon. Specific systematic linguistic patterns are taken to characterize blending (Kubozono, 1990: 1), which contrasts with previous ideas on the irregularity of blends. For instance, phonotactic and neighborhood constraints are considered *a priori* preliminaries. Phonotactic constraints define the phonological affordances of a specific language, thus limiting which blend structures are possible. Neighborhood constraints restrict potential blends to forms that are not identical to already existing lexemes (see also Lehrer, 2007; Kjellander, 2018). In accordance with this view, a more prominent role is suggested for blends, since blending is considered a systematic word formation process.

The focus on paradigmatic blends in Kubozono (1990) raises the question to what extent headedness is an influential factor on the length constraint, since paradigmatic blends are typically not endocentric (cf. Kelly, 1998; Lehrer, 2007; Bauer, 2012). Furthermore, Kubozono's (1990) data cover a broad temporal and contextual domain. The low viability and uneven socio-pragmatic distribution of blends (e.g., Cannon, 1986; Lehrer, 1996; Fandrych, 2008; Lalić-Krstin & Silaški, 2018) make this broad scope potentially problematic. Nonetheless, Kubozono's (1990) study shows the potential of statistical analysis in blend research. Although delimitations and data retrieval raise some concern, its central aim of systematically investigating blends makes his study an important contribution to blend research in general.

Kelly (1998) is another influential study pointing towards broader cognitive principles that can explain the structure of blends. Kelly's (1998) data are limited to items defined as contractions of conjunctive phrases. For instance, the origin of *fratority* is considered to be the phrase *fraternity and sorority*. This delimitation implies a focus on paradigmatic blends, while seemingly syntagmatic constructions are excluded. The study is presented as three investigations merged into one. The three themes are source word order, source word breakpoints, and phonological similarity of truncated fragments.

First, Kelly (1998) suggests that the first source word is more frequent, has a lower number of syllables, and is more conceptually prototypical than the second source word. The argument is exemplified with *salt* in the conjunctive phrase *salt and pepper*: *salt* is described as more frequent, shorter, and more conceptually salient than *pepper*. This assumption is tested quantitatively in his blend data. Statistical evidence shows that the hypotheses are generally born out, but he points out that the connections between the variables pose interpretive challenges. For instance, shorter lexical items tend to be more frequent, disregarding other factors. Second, it is argued that boundaries, i.e., the position of truncation in source words, "fall primarily at major phonological joints, such as syllable, onset, and rime" (Kelly, 1998: 579). This hypothesis suggests that blend breakpoints conform to the prevalent syllable structure in English (see also Kubozono, 1990). However, blends with overlap are excluded in the analysis. Therefore, it remains unclear to what extent his findings apply to this commonly occurring structure (cf. Cannon, 1986). Third, Kelly (1998) shows that there is phonological similarity between the junctures where the source word segments meet. He constructs a system of categories, in which sounds are given numbers in relation to their proximity in terms of phonetic realization. The numbers are then compared statistically, which results in a matrix allowing quantitative analysis. Little effort has been made to develop this idea further, but the methodological design of phonetic quantification remains a potentially informative approach. In fact, Algeo's (1977) notion of imperfect overlap as a bridge between similarity and overlap is partly operationalized in Kelly (1998). Gries (2004b), however, points out that phonological similarity reaches beyond the point at which the source words meet, which is not reflected in Kelly's (1998) study. Moreover, the exclusion of overlap blends in Kelly's (1998) study raises questions as to the relevance of this finding for blends generally.

In the analysis of breakpoints, Kelly (1998) makes a distinction between his data and research conducted on speech error blends. Thus, his study is another account signaling the shift in attitude towards a taxonomic division of intentional and speech error blends. Furthermore, Kelly (1998) demonstrates how blend research may have implications for linguistics in general. The breakpoint analysis contributes further empirical evidence supporting a hierarchical representation of the onset-rhyme structure in the much-debated issue on English syllable structure. Finally, although the methodological and theoretical contributions have made Kelly's (1998) study influential, the data remain problematic. The sources consist of Pound's (1914) collection of blends and entries from the Oxford English Dictionary (OED) tagged with either *blend* or *portmanteau*. These sources are an unfortunate combination of the questionable relevance of Pound's (1914) blends and the problems of dictionary data as discussed in 3.4.1.

In an investigation conducted within the framework of Natural morphology, Ronneberger-Sibold (2006) addresses the issue of whether the transparency of source word segments is elaborated to achieve communicative effects. More specifically, she investigates whether this aspect of blend structure is functionally motivated. The study also offers a taxonomy of a number of structurally motivated blend types. Speech error blends are excluded from the analysis of transparency of segments. So-called *regular formations* are also excluded, since they are seen as various types of compounding. Ronneberger-Sibold's (2006) corpus consists of 612 German blends. It is divided into two sub-parts differing as regards both origin and overall properties. The first one is a collection of 392 blends taken from Grésillon (1984). It consists of blends described as "literary in the widest sense, with a satirical tendency" (Ronneberger-Sibold, 2006: 165). An unspecified proportion of the blends in the Grésillon corpus is attributed to the 19th century poet and author Heinrich Heine, but journalistic media and children's literature also contribute to the list of blends. The second subcorpus is a compilation of brand names retrieved from 54 issues of a journal registering German brand names between 1894–1994. Both the origin and delimitation of the subcorpora are well defined and transparent enough for replication. However, the selection of the specific sources raises questions concerning the relevance of the data in a 21st century account of blend structure. Despite this caveat, comparing the different contexts of the subcorpora offer valuable insights into sociopragmatic motives for selecting a specific level of transparency in a blend.

The method of Ronneberger-Sibold (2006) involves the construction of four main categories, or blending techniques, and four additional subcategories. This taxonomy is intended to capture different perspectives of transparency in blend structure. Establishing the theoretical and methodological matrix is described as a process, whereby blending techniques are defined and associated with a degree of transparency. The main categories are *complete blends*, *contour blends*, *semi-complete blends*, and *fragment blends*, which are ordered according to decreasing transparency. *Complete blends* are described as items in which the source words are preserved in full (i.e., full overlap blends), e.g., Germ. *Kamelefant* (Germ. *Kamel* 'camel' + Germ. *Elefant* 'elephant'). This type is maximally transparent. *Contour blends* involve truncation to some extent, but the prosodic pattern of one of the source words, called the *primary word*, is retained, e.g., Germ. *Tomoffel* (Germ. *Kartoffel* 'potato' + Germ. *Tomate* 'tomato'). The transparency of such blends is related to their prosody, i.e., the rhythm and syllabic patterns are assumed to facilitate the process of decoding. These blends are relatively transparent. In *semi-complete blends* truncation occurs, but one of the source words are opaque, notably a fragment in initial position. This blending technique is exemplified with Germ. *denglisch* (Germ. *deutsch* 'German' and Germ. *englisch* 'English'). Ronneberger-Sibold (2006) points out that there are no prosodic cues scaffolding the process of decoding in this type of blend, and the level of transparency is generally lower than in complete blends and contour blends. The last category, *fragment blends*, contains forms constructed only with truncated segments. Such blends are the least transparent ones of the four main categories. She exemplifies this blend type with the tobacco brand name *Cujasuma* (Cuba + Java + Sumatra). This example indicates that a distinction is not made between blending and complex clipping, since all source words retain an initial segment (cf. *digicam* from *digi~~tal~~* + *cam~~era~~*).

Ronneberger-Sibold's (2006) analysis shows that the Grésillon subcorpus blends, called *satirical creations*, are categorized almost exclusively as complete blends (66.3%) and contour blends (30.6%). Based on this observation, Ronneberger-Sibold (2006) assumes that the literary and satirical creations can be decoded without extensive and specific previous contextual knowledge. Therefore, they are considered the most transparent ones. In contrast, the blends of the brand name subcorpus are rather evenly distributed between the four main categories. This diversity of the brand name blends is seen in the distribution of transparency as well. She takes this as an indication that brand names have a limited descriptive function, but are instead motivated by a "requirement [...] to be above all evocative" (Ronneberger-Sibold, 2006: 175).

Contrasting with most preceding studies on lexical blends, Kemmer (2003) offers a schema-based approach within a Cognitive Grammar framework. She argues for the use of the concept *constraints* rather than *rules*. It is claimed that schemas are particularly fit to analyze the intricate characteristics of blending, since they afford an "integrated and cognitively well-grounded account" (Kemmer, 2003: 80). The application of a phonological schema is exemplified with the blend *swooshtika* (*swoosh* + *swastika*). It is suggested that the two source words in combination evoke a phonological schema /swVS/. This schema licenses combinations of source words resulting in blends with the initial phonological cluster /sw/. Then, the following phoneme can be any vowel /V/, and any voiceless sibilant /S/.

Kemmer's (2003) schema analyses apply the theory of conceptual blending (Fauconnier & Turner, 2002). This theory is suggested to provide an adequate background to explanations of the intricate conceptual relations between source words and blends. She exemplifies a semantic schema analysis with the blend *glitterati* (*glitter* + *literati*), which is explained as "the glamorous elite" (Kemmer, 2003: 73). It is shown that the source word *literati* is the basis for a whole range of analogical formations such as *chatterati*, *liberati*, and *digerati*. The process behind the use of these forms is referred to as a *lexical schema*. This type of schema has the potential to result in morphological reanalysis of a segment, i.e., the lexical fragment in question may eventually acquire morpheme-like properties. She also highlights the iconic relation between the tight formal fusion and an equally tight conceptual integration in lexical blends.

Kemmer (2003) contrasts the schema analyses with other theoretical approaches. She points out that building-block theories of morphology have considerable difficulties to explain blending. There are three main arguments for this claim. First, blends are not constructions based on morphemes, but of "phonological strings that trigger meanings" (Kemmer, 2003: 77). Second, it is argued that building-block theories cannot satisfactorily account for non-linear concatenation, often including overlap. Third, blends are claimed to rely on pattern-matching rather than morphological rules. Kemmer (2003) also argue against an Optimality Theory (OT) approach, despite the fact that it is constraint-based. OT is described as an essentially generative framework focusing on innateness rather than learning. As a consequence, frequency effects cannot be properly accounted for in OT. Furthermore, the categorical representations of OT are described as less suitable to analyze blends than frameworks allowing analyses of degrees of fit (Kemmer, 2003: 80). Finally, it is claimed that OT is concerned with formal structure and does not convincingly address the semantics of blends.

Finally in this section, two studies by Gries (2004b, 2006) demonstrate the use of corpus data and statistical analysis in blend research. These studies provide important methodological models in the systematic exploration of blending. Gries (2004b) addresses previous assumptions on blending by the use of rigid statistical analyses. Before introducing the methodological procedure, he discusses previous research, and criticizes the empirical base of earlier accounts. The data collection in Gries (2004b) is based on four criteria: the number of source words, truncation type, conflation type, and overlap. The delimitation can be summarized as blends formed from two source words, which display some kind of truncation. The structure of the concatenation is preferably, but not necessarily, of the type in which a final segment of the first source word is combined with an initial segment of the second source word (cf. criterion (iii) in 3.1). Thus, the definition of the blends is based on their structure and is taxonomically rather open. This kind of definition enables the inclusion of a large quantity of items in the blend corpus. Despite these measures, the origin of the data remains problematic in this study as well. The blends are taken from 18 sources with little in common, of which 14 are publications that Gries (2004b) has previously criticized. Course material from a linguistics course, two dictionaries, and an Internet resource (Linguist List) are employed. The resulting corpus contains 585 blends, which he claims to be one of the largest blend collections at the time of publication. However, the different (and often opaque) principles of data collection make it difficult to control the variables influencing the blends in this corpus (cf. Wulff & Gries, 2019). Problematic issues of previous collection of blends (e.g., representativity and viability) also become inherited in Gries' (2004b) data.

The first case study in Gries (2004b) aims at investigating the implications of source word length. In general, the second source word "is longer and contributes more [of its structure] to the blend" Gries (2004b: 654). A blend such as *Hinglish* (*Hindi* + *English*) exemplifies such an observation. The findings in this case study are interpreted as evidence showing the importance of recognizability as a mechanism governing blend formation (Gries, 2004b: 661). The second case study in Gries (2004b) demonstrates the impact of structural similarity between the source words and the resulting blend. The results from both case studies are compared with parallel analyses of speech error blends. The comparison shows that speech error blends differ significantly from intentional blends.

In the later publication, Gries (2006) again points out that earlier accounts are largely classificatory. He claims that "there is only a handful of studies which attempt to tackle the issue of how blends are actually formed" (Gries, 2006: 537). The objective of the study is to analyze recognition points from a psycholinguistic viewpoint, and to compare blends and complex clippings as regards graphemic similarity. Both these topics are addressed from a corpus-linguistic perspective. The notion of *selection point* (SP) is used to address the issue of recognition point. SP is a measure of the amount of structural material a language user typically requires to decode a lexical item. Technically speaking, an SP of a word is the location of a truncation at which a truncated query generates the highest token frequency for the word in a corpus. An illustration of this calculation is found in Table 9 in section 4.3.2. Gries (2006) shows that there is a statistically significant correlation between the SP of first source word and its truncation in a blend. In contrast, the second source word typically retains slightly less material than specified in an SP analysis. Gries (2006) suggests that the difference in truncation pattern between the source words depends on a lower requirement for information quantity of the second source word. The first source word is assumed to prime the second source word, which compensates for a comparatively reduced amount of lexical material. Thus, truncation is not arbitrary, but follows traceable and identifiable principles. Gries (2006) shows statistically that complex clippings do not conform to these patterns of truncation.

Furthermore, Gries (2006) investigates whether source words in blends display greater phonemic and orthographic similarity than randomly chosen pairs of words. The same procedure is performed on complex clippings with the aim to examine whether possible patterns in blends and complex clippings coincide. Measures of similarity are calculated using bi-gram analyses, which show that similarity is more common among the blends than in random pairs of words, and that complex clippings do not follow the blend pattern in this respect. As a consequence, there are good reasons to distinguish between blends and complex clippings from the perspective of source word similarity as well.

The overview of previous research mainly concerned with the structure of blends has demonstrated that the various findings offer several informative perspectives. The studies mentioned in this section have, for instance, contributed to the disentanglement of issues relating to the phonemic and orthographic similarity of source words and blends. They have also provided empirical support for a constraint-based approach to blending, and they suggest several ways how to further the methodological development of blend research.

Although certain tendencies have been revealed, central questions regarding the motivation for discernible patterns remain largely unanswered. The datasets often have problems of traceability, and it is not clear to what extent the findings are generally applicable. One way to approach such questions is to focus on the users of language rather than on the observed instantiations of language alone. This is the theme of the following section.

3.4.3 Psycholinguistic and experimental investigations

The studies accounted for in this section are based on experimental data. In particular, semantic concerns are addressed with the methodological tools of psycholinguistic investigation. Key issues in the psycholinguistic study of blends are illustrated below in studies by Lehrer (1996), Silaški and Đurović (2013), and Wulff and Gries (2019). Experimental methodologies imply certain challenges to linguistic investigation. One such challenge is the *observer's paradox*. McMahon (1994: 234) illustrates this by saying that "sociolinguists would like to know how speakers speak when they are not being observed, but the only way to find out [how they speak] is by observing them". *Relevance* is another important issue. It concerns whether it can be asserted that the experiment constitutes a relevant representation of a phenomenon outside the test frame (see detailed discussions in Tomasello, 2008; Gries, 2012; Arndt-Lappe & Plag, 2013).

Lehrer (1996) represents a new direction in blend research. This study introduces a processing perspective relating to the use of blends in natural language occurrences. The test design can be described as a conventional psycholinguistic experiment in which test participants are presented with a set of blends they are asked to identify. The test participants are undergraduate students of linguistics, presented with questionnaires. These questionnaires are based on 71 blends retrieved from, for instance, journals, newspapers, and advertisements. The identification of the source words is assumed to be central for successful decoding. The data are analyzed statistically using qualitative interpretation following a psycholinguistic tradition (see for instance Harley, 2008). Lehrer (1996) shows that successful decoding of blends depends to a varying degree on context, semantic priming, frequency, lexical neighborhood, and how much of the source word is retained in the blend. Moreover, the identification of the source words of an unknown blend is remarkably facilitated if contextual cues are present. For instance, decoding *clamato* (*clam* + *tomato*) is facilitated if the semantic scope is narrowed with the accompanying string "a delicious appetizer" (Lehrer, 1996: 368). Furthermore, recognition of one of the source words is claimed to influence the decoding of the blend, e.g., if either *sweater* or *jacket* are recognized, the blend *swacket* is more easily understood. Frequent source words are shown to be more easily recognized, and the more unique they are (i.e., the fewer neighbors they have) the easier the identification. In general, the level of reduction influences the possibilities to decode a source word.

Lehrer (1996) applies important limitations to her data. She narrows down the scope to *lexical* blending, hence syntactic blends and speech error blends are excluded. The object of investigation is then further specified to *novel* blends, as opposed to already established and lexicalized items. Adding the criterion *novel* implies that a point in time is determined when the blend was coined. Normally, corpus investigations cannot identify such events, while the construction of novel forms is possible in an experimental setting.

In parallel with several other studies around the late 1990s and early 2000s, Lehrer (1996) discusses the somewhat problematic status of blending in morphology, and outlines a preliminary position of the mechanisms of blending in relation to morphological theory. She points out that blends are not formed by morphemes in the traditional sense. Instead, blends are subject to other processes that require methods not traditionally associated with word formation research.

The second study presented in this section, Silaški and Đurović (2013), investigates recognition and disambiguation of blends. The test participants are 95 students of English for Specific Purposes (ESP) at the Faculty of Economics at Belgrade university. They are between 21–23 years old and have considerable experience of formal training in English. The experiment begins with an introduction, in which the phenomenon of lexical blending is explained to the participants. After this presentation, they are asked to identify the source words of 33 blends in a questionnaire. These blends are "mostly collected from Investopedia.com site in 2011, filed under the heading Buzz Words" (Silaški & Đurović, 2013: 89). The participants are asked to explain the meaning of the blends in their native language Serbian with the intention to avoid semantic limitations due to a lack of proficiency in English. Then, the researchers analyze the results of the questionnaire from the perspective of conceptual blending (Fauconnier & Turner, 2002).

Silaški and Đurović (2013) conclude that proficiency level is important for successful decoding of novel English blends, which is claimed to have pedagogical implications in an ESP setting. Furthermore, it is claimed that the test participants seem to overlook the "semantic coherence" (Silaški & Đurović, 2013: 100) of the source words. This leads to unsuccessful decoding of a blend, although the source words are identified correctly. Silaški and Đurović (2013) exemplify with the blend *siliconaire* (*silicon*/*Silicon Valley* + *millionaire*). To be interpreted according to its conventional use, it requires knowledge of the financial situation of the Internet business in the 1990s. Thus, the semantic coherence of the source words depends on encyclopedic knowledge and knowledge of how *siliconaire* has been used. They also mention the phonological aspects of blending, which is to be understood as an additional comment on the impact of language proficiency on their results. Overall, background knowledge is emphasized as an important factor for successful decoding.

In a more recent study, Wulff and Gries (2019) draw attention to the problem of data collection by revisiting the findings of previous corpus-based studies. They examine experimentally the contribution of the shorter source word, the stress influence of the second source word, and the similarity between the source words and the blends. One of the main aims of Wulff and Gries (2019) is to test these findings experimentally to evaluate the impact of systematic data collection. They emphasize that the largely unsystematic data collection of previous studies is a problematic issue in blend research.

The experiment is designed as a questionnaire test, in which the test participants are asked to form blends of words presented in pairs. These pairs have fixed word order, which is stated in the instructions of the questionnaire. The selected words, all nouns, are taken from the semantic domains *fruit*, *vegetables*, *dog breeds*, and *car brands*. Choosing words from these domains is motivated by the assumption that they provide "plausible scenarios for intentional blending" (Wulff & Gries, 2019: 4). Several measures are taken to divert the participants' attention from the investigated task. For instance, math problems are also included in the questionnaire. The participants are divided into two groups, in which one is asked to answer in written form, while the other group is asked to spell out their answers before writing them down.

Wulff and Gries (2019) show that two out of three previous findings gain some support from the results. As regards the role of the shorter source word, it is shown that its proportional contribution may not be as significant as previously suggested. Wulff and Gries (2019) relate the stronger effect of contribution found in earlier studies to unsystematic procedures of selection, which have favored certain patterns already at the stage of data collection. The second issue is concerned with the influence of the stress pattern of the second source word. It is commonly assumed that the second source word determines the stress of the blend, which is reflected to some extent in the experimental data of Wulff and Gries (2019). Finally, the investigation of the similarity between the source words and the blends reveals that their experimentally derived results conform rather well to previous findings. However, their general conclusion is that the investigated effects seem stronger in earlier observational studies, despite the relative consonance between their results and previous assumptions. Accordingly, they stress that "much more such 'validational work'" (Wulff & Gries, 2019: 19) is required.

In sum, psycholinguistic approaches such as those presented in this section may contribute valuable insights into the mechanisms of blending, especially in terms of language processing. However, shifting the focus from the structure of blends to the users of language involves all possible influences of human existence on language. Sociolinguistic concerns are central issues from this viewpoint, and in the following section, studies involving a sociopragmatic perspective on blending are presented.

3.4.4 Sociopragmatic studies

The role of sociopragmatic concerns has been highlighted in several studies on blending (e.g., Cannon, 1986; Lehrer, 2007; Fandrych, 2008; López Rúa, 2010, 2012). Two studies with a specific focus on such issues are presented here.

Fandrych (2008) stresses the import of contextual and sociopragmatic factors on blending. It is argued that building-block theories of morphology are limited as explanatory tools for blending. She claims that "structural criteria are not sufficient for the description of non-morphemic word-formation processes" (Fandrych, 2008: 73). Instead, a broad interdisciplinary approach is advocated, as it adds cognitive and functional perspectives to the analysis of structure.

The data in Fandrych (2008) are analyzed contextually, and she illustrates how important information is potentially lost in purely structural analyses. A reference is made to the impact of American culture in general, which involves its assumed predilection for blending. Cognitive factors are highlighted in terms of the processing of transparency and the elaboration of satire in blends. The investigated blends are mainly connected to the domains of business, economics, and the media. Blends associated with the media domain are argued to be short-lived, since it is assumed that their motivation for existence is to achieve an eye-catching effect with a limited temporal and contextual scope. For this reason, such blends are claimed to resist processes of lexicalization (Fandrych, 2008: 75). As regards functional concerns, blends are described as technical and colloquial with a preference for attention-catching, especially in the form of cataphoric suspense, which is often observed in newspaper headlines. Fandrych (2008) argues for the importance of the written mode in blending as well as in other non-morphemic word formation processes.

Another example of a sociopragmatic approach to blending is found in López Rúa (2012). In this study, 190 names of musical bands and artists are examined in order to find out how lexical form is employed to break norms and construe identities. López Rúa (2012) analyzes the formal structure, the grammar, and the semantics of the blends in relation to sociopragmatic goals and intentions. Thus, the investigation has a partly structural focus, while its aims are oriented towards cultural and sociological issues.

The analysis of formal structure is focused on blend constituent types, for instance *full word* and *splinter*, and investigates degrees of overlap (called *integration* in her study). Seven categories are established based on the form of the items, for instance *splinter + full word* (e.g., *Apocalipstick* from *Apocalyptic/apocalypse + lipstick*) and *splinter + splinter* (e.g., *Enertia* from *energy + inertia*). The data are diverse, showing some preference for the patterns *word + splinter* (e.g., *Skillzophrenic* from *skill + schizophrenic*), which is claimed to constitute 28.4% of the data. The type *word + word* (i.e., full overlap blends) exemplified with *Humanimal* (*human + animal*) is also common in her data (21.1%).

The degrees of overlap are expressed in the three-level taxonomy *low, medium*, and *high integration*. Low integration is represented by constructions with no structural overlap of the source words. She also notes that wordplay is rather scarce in this category. For these reasons, low integration items are questioned as blends proper. López Rúa (2012) exemplifies this type of construction with the lexical item *Con-Dom* (*control* + *domination*, including a reference to *condom*), which is in fact a complex clipping. Medium integration involves a broad range of patterns of truncation and amalgamation. Overlap can be realized both graphemically and phonemically in this category, and it varies considerably in terms of degree. Examples are *Stormageddon* (*storm* + *Armageddon*) and *Nocternity* (*nocturnal* + *eternity*). High integration is characterized by striking phonic integration and wordplay. Typically, one of the source words is thought of as including the other in examples such as *Blastard* (blast + bastard), in which the difference between the blend and the second source word is minimal. The high and medium categories represent more than 90% of López Rúa's (2012) data.

The grammatical function is presented quantitatively. López Rúa (2012) uses the term *syntactic* blends for syntagmatic constructions such as *Sexorcist* (*sex* + *exorcist*) and claims that 67.4% of the blends belong to this category.[3] The remaining blends are called *coordinate* blends, e.g., *Flametal* (*flamenco* + *metal*). It is pointed out that the relation between the source words is typically either complementary (e.g., *Sentimetal* from *sentimental* and *metal*), hyperbolic (e.g., *Warmageddon* from *war* + *Armageddon*), or anti-thetical (e.g., *Sinocence* from *sin* + *innocence*). Hyperbolic and anti-thetical constructions are claimed to be employed in the formation of "a shocking combination of items which conveys a message of rebelliousness against the rules" (López Rúa, 2012: 31).

These sociopragmatically oriented studies exemplify how blending can be studied from the perspective of broad communicative and cultural motivations. Still, the data they analyze are no less problematic than in most other types of investigation (cf. Wulff & Gries, 2019). Studies with a quantitative ambition need to pay particular attention to sociopragmatic concerns in future research. Moreover, a perhaps even more pressing issue is the nature of the empirical material in sociopragmatic investigations. For instance, alternatives to corpus-based data such as structured group conversations or interviews could potentially contribute to more detailed and informative accounts of the norms, strategies, and goals that influence the use of blends.

3.4.5 Computational accounts

Blending has also attracted some attention within the field of computational linguistics. Blend studies in this tradition demonstrate both the kinship and terminological divergences with theoretical linguistics (cf. Gries, 2010). The computational perspective is important because of its implications for the methodological development of blend research. The growing number of studies

[3] Her use of the term *syntactic* may seem slightly confusing since *syntactic* blending usually refers to phrasal blends (see 3.3.6). However, the context shows clearly that *lexical* blending is intended.

using automated procedures to collect and analyze blends attest to the efficiency of these tools. It may even be argued that the boundaries between computational and theoretical traditions are best described as a continuum (Gries, 2010). For instance, Gries (2004b, 2006, 2012) and Renwick and Renner (2019) make extensive use of computational tools, while their analyses are largely conducted within theoretical linguistics traditions.

Cook and Stevenson (2010) address the issue of recognizability of source words in blends from a quantitative perspective. More specifically, they design a computational method to detect source words of novel blends automatically. Their data consist of 324 blends retrieved from the website *wordspy.com*. The blends are collected from a reference list of 1,186 lexical items claimed to be neologisms. The data are then tested as to automatic recognition, or identification, of source words. This is done by applying statistical algorithms employing pre-defined variables called *features*. These features consist of interpretations of previous findings in blend research re-modeled to fit their computational frame. They find that their automated method of detecting source words of blends has an accuracy of 37%–42%. A comparison is made with Lehrer's (2003) corresponding results of human performance on this task, in which source words are correctly identified in 34%–79% of the cases. Cook and Stevenson (2010) claim that this procedure has the capacity to identify source words significantly better than by chance. Moreover, it could potentially be extended to identify blends automatically in a corpus.

The theoretical framework of Cook and Stevenson (2010) differs in many respects from much of the research on blends described in the previous sections. This is perhaps most apparent in their presentation of the computational design, but the use of the terminology and subsequent definitions differ on a more profound level. Transferring previous findings into computational models may seem attractive, but there are inherent problems with this part of their method. For instance, the tentative nature of the results of Kubozono (1990) and Kelly (1998) risks multiplying errors if taken for granted in a digital representation such as in Cook and Stevenson (2010). Furthermore, the definition of a blend as truly novel is not resolved in the study, and the use of *prefix* and *suffix* to refer to blend segments introduces theoretical uncertainty and does not satisfactorily address the complexity of blend constituency. Nonetheless, the study contributes important perspectives on the methodological development of blend research.

Cook (2012) continues to explore the possibilities to collect blends using automated procedures. The data are collected from the social media platform *Twitter*. Blends are retrieved by identifying discursive markers of the type *jargon for*. Such expressions are assumed to be indicators of a possible blend in a Twitter message. Although the output includes a set of blends, it is questionable whether the data are informative, since there are reports indicating that blends are "typically presented without glosses or explanations" (Lehrer, 2007: 115; but see also Ronneberger-Sibold, 2006). Despite this problem, patterns of co-occurrence may yet provide insights into blending, although it remains to be explored in more detail whether such tools can function to detect blends automatically.

Computational approaches to investigating blends involve both promising prospects and challenges. The theoretical interface between computational and theoretical linguistics calls for careful considerations to avoid misunderstandings, but both these branches of linguistic investigation are likely important in the development of blend research.

3.5 Operational definition

The preceding sections concerned with the definition of blending and previous blend studies have shown the many challenges to describe and investigate blends. Although blends often seem to elude delimitation, an operationalization of the investigated object of study is nevertheless necessary in order to answer the research questions of the present study (cf. Renner, 2019: 29). The definition given in the list of criteria below functions as the background for the identification of blend characteristics and the investigation of ambiguity in the systematically collected data. In short, it provides the rationale for what is here considered a blend and what is not (cf. Geeraerts, 2006a: 150 on category membership). Criteria A and B encapsulate what is interpreted as a blend in this investigation, and they constitute the basis for the selection and interpretation of the data. The abbreviations SW1 and SW2 stand for *first source word* and *second source word* respectively.

Borderline cases are well attested in previous accounts (e.g., Bauer, 1983), which calls for a specific comment. One of the recurrent problems concerns morphemization: the gradual nature of segments developing into combining forms contrasts with the binary distinctions used when establishing category membership (cf. Schmid, 2020: 288-289). A rather conservative position is taken in the operational definition, in which a conventionally acknowledged combining form is not considered a blend segment, since it can be “analyzed as a morphologized unit” (Renner, 2019: 29). For instance, *shopaholic* is not interpreted as a blend, because *shop* is a conventional lexeme and *-holic* is widely interpreted as a suffix or as a combining form. Subsequently, *shopaholic* is analyzed as a derivation or a compound. The identification of combining forms (or other morphemic segments) is operationalized using the category assessment of three well-established English online dictionaries: Merriam-Webster Dictionary (Merriam-Webster, n.d.-a), Cambridge Dictionary (Cambridge Dictionary, n.d.), and Macmillan Dictionary (Macmillan Dictionary, n.d.). These dictionaries are thought to represent a contemporary and conventional description of English. If a specific segment is described as a combining form in any one of them, then this indication is considered enough to define it as a combining form. Examples of segments considered combining forms on these grounds are *-gate* and *robo-*. Therefore, blend candidates built on such segments as their only truncated structure are excluded from the list of blends. Items combined with, for instance, *-nomics*, and *-tainment* are instead seen as blends because these segments have no status of combining form in any of the three chosen dictionaries. This procedure may exclude constructions that still display blend characteristics, while at the same time items may be included that border on morpheme-based word formation processes. Nonetheless, rigid principles of membership categorization are required to ensure the systematicity of the data. Hence, the operational definition serves the purpose of striking a balance between fuzzy boundaries and the requirement for systematicity.

In sum, the operational definition consists of slightly elaborated criteria of the preliminary definition (3.1), extensions based on the overview of blend formation issues (3.2–3.3), and broad considerations of previous investigations (3.4). Truncation is foregrounded in criteria A.1–2 and B.1–2, while full overlap blends (3.2.4) are captured in A.3. Criteria A.4–5 specifically deals with the structural diversity of blend formation. The intricate issue of morphemization is addressed in criteria B.3–4. It is worth noting that wordplay has not been included in the operational definition because of a lack of sufficiently stringent tools to define it.

A. Collected items are considered blends:

1) if they include an initial truncated segment from SW1 and a final truncated segment from SW2, e.g., *brunch* (*breakfast* + *lunch*).

2) if only one of the SWs is truncated, either SW1 initially or SW2 finally (and potentially further SWs initially and/or finally), e.g., *webinar* (*web* + *seminar*) or *frienemy* (*friend* + *enemy*).

3) if none of the (two or more) SWs seems to be truncated, but the SWs instead share structure (overlap), e.g., *stoption* (*stop* + *option*) and *Intelevisionary* (*Intel* + *television* + *visionary*).

4) if the blending of the SWs is non-linear but yet displays truncation and/or overlap, e.g., *Caleaveformia* (*California* + *leave*).

5) if the graphemic and/or phonemic structure of at least one of the SWs is modified and one or all of criteria 1-4 still apply, e.g., *in-sin-uation* (*insinuation* + *sin*).

B. Collected items are *not* considered blends:

1) if there is no truncation and no overlap of the combined elements (e.g., *masterpiece*) or if the segments are otherwise established morphemes (e.g., *pre-owned*)

2) if the item consists of two segments of which both are truncated from the right, which is here referred to as complex clippings, e.g., *digicam* (*digi~~tal~~* + *cam~~era~~*)

3) if the only potentially truncated segment is a neo-classical combining form, e.g., *biotechnology*

4) if one of the constituents is represented as a combining form in any of the three online dictionaries Merriam-Webster, Cambridge Dictionary, or Macmillan Dictionary *and* the other constituent is at the same time morphologically analyzable, e.g., *mountainscape* formed analogically from *mountain* and *landscape* (cf. Cannon, 1986: 731; see also Appendix 6).

3.6 Summary

The preliminary definition in 3.1 and the list of problematic issues in 3.2 suggest that blend formation is both rather straightforward and strikingly complex. The criteria in 3.1 are usually accepted as a baseline definition making blends relatively easy to identify. Although there is variation, similar definitions are found in numerous studies (e.g., Kemmer, 2003; Plag, 2003; Bauer, 2006; Gries, 2006; Lehrer, 2007; Fandrych, 2008; Gries, 2012; López Rúa, 2012; Arndt-Lappe & Plag, 2013; Lepic, 2016; Beliaeva, 2019). The definitional vagueness of the preliminary definition as emphasized in 3.2 can be addressed from the perspective of prototype theory (Renner, 2015, 2019). For instance, criterion (iii) describing the “clipping of both source words at the inner edges” (Renner, 2015: 122) may be described as a central attribute, although there are numerous attestations of blends that do not strictly follow this principle (e.g., Cannon, 1986). With a prototype theory approach, the baseline definition may remain comparatively stable and simple, while variation is afforded. The issues in 3.3 stand out as harder to account for systematically than the criteria of the preliminary definition. The centrality they are awarded varies in previous research, and several of them are poorly investigated. Nonetheless, they constitute important facets to grasp the domain of lexical blending, which is why they have descriptive implications.

The overview of previous blend studies displays considerable theoretical and methodological diversity. This is no coincidence. Blends are enigmatic in many ways, and they have escaped viable descriptions for nearly a century (cf. Balteiro & Bauer, 2019; Renner, 2019). In fact, it might be argued that blend research is just as complex, and sometimes even confusing, as the object of investigation itself. This observation is mirrored by the difficulties to agree even on basic definitions.

A close examination of the area of blend research reveals that there are numerous contradictions and mismatches, but there are also clearly discernible tendencies. First, several accounts argue that it is not reasonable to classify blends according to an Aristotelian tradition devising clear-cut categories based on necessary and sufficient conditions (cf. Taylor, 2003). Blends are just too heterogeneous and too dynamic. Instead, other principles of categorization are suggested, notably along the lines of prototype theory and family resemblance (cf. Brdar-Szabó & Brdar, 2008; Beliaeva, 2014b; Renner, 2015, 2019). Second, analyses of formal structure have so far offered possible answers to some of the issues of blending. However, their limitations as regards (for instance) semantic analysis await to be addressed systematically (cf. Bauer, 2012; Gries, 2012). The advent of broader theoretical and methodological approaches in recent decades may be interpreted as a response to this issue. This broadened perspective signals a new direction in blend research, which is essential to the development towards a better understanding of lexical blending (Renner et al., 2012a; Balteiro & Bauer, 2019). Third, psycholinguistic, sociopragmatic, and computational perspectives are promising approaches that deserve further attention. As Wulff and Gries (2019) show, much could be gained simply by testing previous assumptions from other perspectives so as to arrive at more reliable conclusions. Finally, data collection has emerged as a problematic issue (Wulff & Gries, 2019). The methodologies of the last three decades have developed significantly, as can be observed in the increasing precision of analyses and discussions. Still, while quantitative methods have suggested new ways to collect, organize, and interpret data, issues of representativity and relevance need to be addressed, and, at least to some extent, resolved. If fundamental requirements on systematic data collection are not met in future blend research, it is questionable whether further quantitative investigations could go beyond confirmation of already established intuitions. Finally, the overview of previous research has motivated an operational definition that is slightly refined in comparison with the preliminary definition in 3.1. This definition is necessary in order to provide a transparent background to the method of the present study.

4 Method

This chapter describes the systematic method designed to collect and analyze the blend data. After having selected the NOW corpus as the empirical source, the corpus data were organized and prepared for efficient management, which resulted in a subcorpus delimited to written American English web news in the period 2010–2018. To address the aim of collecting the data in a rigorous way, an algorithm was then devised to retrieve a randomized sample from the subcorpus in which blends might occur. This sample was called *the blend pond*. Its size was the result of balancing between a large enough quantity to match the expected low frequency of blends and a small enough size to allow manual analysis. Once the blend pond was established the next task was to identify blends according to the operational definition (see 3.5). It is worth stressing that the systematicity of the data collection was an important concern all the way from the selection of the corpus to the identification of blends in the blend pond.

The investigation is based on the *News On the Web* corpus (NOW) (English-Corpora, n.d.), which is described in 4.1. The following section then addresses the organization of the corpus data (4.2), including applied terminology and technical concerns. Section 4.3 accounts for the retrieval of the blend data. It describes the methodological procedures from the point at which the candidate words (henceforth CaWs) were randomly selected to the construction of the final list of blends. The analytic implications of the theoretical framework are discussed in 4.4, and the limitations of the method are dealt with in 4.5. Finally, the chapter is summarized in 4.6.

4.1 The NOW corpus

NOW consists of language data from news texts published on the Internet. This context constitutes a suitable empirical source for investigating blends, as it is connected to both news media and the application of digital technologies (cf. Lehrer, 2007; Fandrych, 2008). NOW is a so-called *monitor corpus* that collects data continuously, and its size currently grows by up to 200 million words each month (English-Corpora, n.d.). The corpus data were retrieved from news publications starting in January 2010. At the time of data collection, it contained 5.8 billion words from 20 English speaking countries. At present, it consists of 14.6 billion words (March 2022). Its large size in combination with a relatively wide temporal scope match well the expected low frequency and viability of blends (cf. Lehrer, 1996; Cannon, 2000).

The NOW corpus is annotated for grammatical function (i.e., word class). All occurring items are assigned tags that mark this property. NOW uses the term *part of speech* for this purpose, and the abbreviated tag is *PoS*. Prior to the inclusion of an original text in the NOW corpus, the process of PoS annotation is carried out automatically by the software *Claws 7* (University Centre for Computer Corpus Research on Language (UCREL), n.d.-a) The accuracy of this software is approximately 96-97% according to UCREL at Lancaster University.

Queries in the NOW corpus are often run in an online interface that allows them to be filtered by region, temporal scope, and word class. Ideally, these filters would have provided sufficient tools to collect the US data limited to the stipulated period of time (i.e., 2010–2018). The available technology would, thus, align with the empirical requirements of the study. However, initial pilot tests revealed that the applied limitations of the data were too demanding in terms of technical power. For instance, restricting a query to a narrow temporal scope in combination with a broad category such as *noun* would typically result in system breakdown. Therefore, it was necessary to download the entire corpus to access it offline. An offline version would be technically manageable, and it would also imply consistency and offer full control of the delimitations of the data. This issue was confirmed by Professor Mark Davies (personal communication), who explicitly advised me to use an offline version after having analyzed the requirements of my study in relation to the constraints of the online interface of NOW.

The complete offline version of NOW containing 5.8 billion words was purchased and downloaded in April 2018. It consists of 334 compressed files in UTF-8 text format (.txt). The storage size of these files amounts to approximately 84 GB. Furthermore, they are tagged for region, which enabled US data to be identified already in the process of archiving. Effectively, this was the first procedure in the creation of the subcorpus.

The full offline access to the data involves a notable restriction. US policies require that such data should be modified so as not to be useful as a substitute for original publications. To meet this requirement, the corpus constructors have designed an algorithm that randomly destroys 5% of the corpus data in order to render the original texts commercially unattractive (Corpusdata.org, n.d.). This measure causes little damage to the quantitative information retrieved from the corpus, but there is still a small risk that some contextual information may be lost.

4.2 The subcorpus

A key concern was how to delimit the downloaded NOW corpus regionally and temporally to ensure efficient data management. It was also necessary to filter out fragmented data and to adapt the corpus material to a suitable software frame. These tasks required the implementation of specific procedures, which are described in 4.2.1. The resulting body of data is referred to as the subcorpus.

These data have obviously their own characteristics and limitations. Two technical limitations are dealt with in 4.2.2, while more general concerns of limitation are discussed in 4.5.

4.2.1 Compiling the subcorpus

Before describing the details of the construction of the subcorpus, some technical details about the chosen software environment need to be presented. The programming language used in the method of this investigation is SQL (Structured Query Language; /ˈsiːkwəl/). SQL is based on data stored in tables of rows and columns (henceforth called *SQL tables*) and so-called *scripts*, which are text instructions designed to retrieve information from the SQL tables. An important property of SQL is the efficiency with which databases in the form of SQL tables can be connected. This programming language is therefore suitable to manage NOW, since the architecture of this corpus is based on several large text databases. The SQL scripts are formalized sets of specialized *commands* (sometimes also called *keywords*) and syntactic rules. Within the constraints of the commands and the SQL syntax, virtually any kind of operation can be executed involving data retrieval from one or several databases. Commands are represented in bold capitals to mark their special technical status (e.g., **SELECT**). The names of specific SQL tables are consistently represented in italics.

Managing a software solution such as the one described here requires knowledge about a broad range of pre-defined concepts and how they can be applied. It is also necessary to know how to design scripts that are efficient and accurate for the type of information they are supposed to retrieve. This makes the procedure more demanding than using more conventional online interfaces, but it is also more flexible as to what information can be collected.

The first concern in the construction of the subcorpus was how to extract and organize the downloaded text files. The file names contained both regional and temporal information of the pattern *year-month-region*. For instance, the file "12-06-us.txt" contained US data from June 2012. Thus, these metadata were used to identify the relevant files and store them in appropriate hard drive folders. Doing so meant that regional and temporal limitations were realized physically. This solution eased the load on the computer processing, since a large part of NOW could be disregarded. Importantly, it also increased the technical reliability of the regional and temporal delimitations.

The ensuing procedure consisted of setting up the computer software environment. The chosen programming language SQL required a software platform in which data storage and queries could be managed. The program pgAdmin (PostgreSQL Tools, n.d.) was chosen to be this platform. A folder *Blend Project database* was created in pgAdmin for the purpose of storing the entire subcorpus. Then, a SQL table containing all items (types) was set up in the Blend Project database. This SQL table was called *lexicon*. As usage data would be a central part of the study, a SQL table was also created containing detailed database information about all tokens in the subcorpus. This SQL table was called *now_us_db*. These two SQL tables in the Blend Project database would contain the entire subcorpus.

Several operations were required to ensure that all relevant data from the NOW corpus text files would be imported correctly into the SQL tables of the Blend Project database. The US part of the downloaded NOW corpus contained 18 lexicon files. Each one of them was imported and tested separately to verify that the data in the SQL table *lexicon* would correspond to the original material. Potential duplicates resulting from import errors were identified and removed in this process. A considerable part of the original corpus files consisted of fragmented, irrelevant, and disrupted data. This material is referred to as *bad data* (Berman, 2013: 168). One source of bad data was the 5% information that was destroyed by NOW prior to the download of the offline version. Instead of original text, these data consisted of rows of “@” signs. Figure 2 visualizes schematically the process whereby the original data were imported into the SQL table *lexicon*.

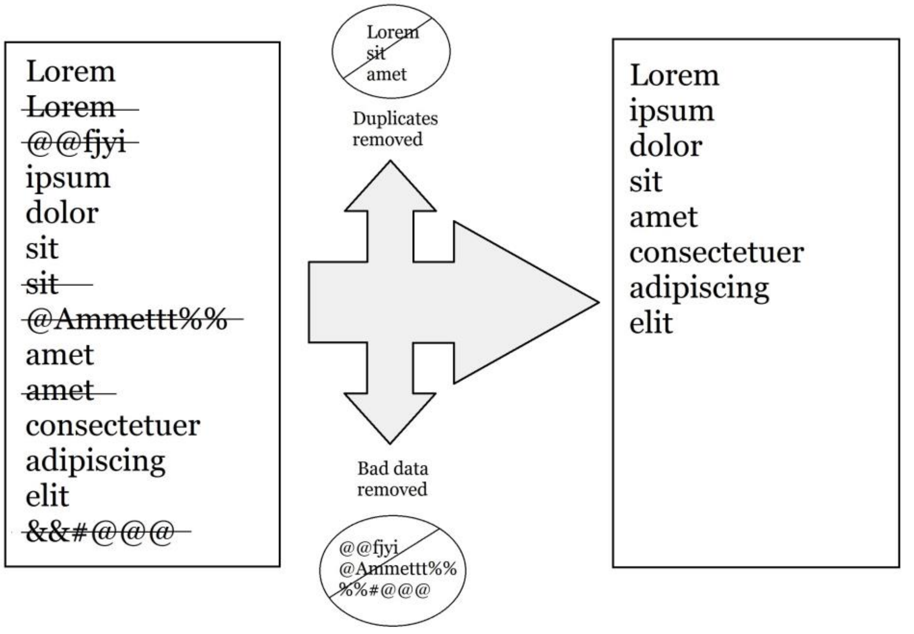


Figure 2. Illustration of data transfer from NOW original text files to the SQL table *lexicon*.

Constructing the SQL table *now_us_db* required somewhat less advanced operations, but the overall procedure was similar. Each of the 99 months in the offline version is represented by a text file. Thus, 99 files were loaded into the SQL table *now_us_db*. This process was continuously tested to ensure that the imported data were identical to the original text files.

A number of tests were performed before the SQL tables *lexicon* and *now_us_db* were finalized. The SQL table *lexicon* is a compilation of all 5.4 million forms occurring in the subcorpus. The database *now_us_db* is a sequentially ordered list of all tokens in the subcorpus, which amounts to 1.1 billion rows, or tokens. The combination of these SQL tables is illustrated in Figure 3. The arrows show how they are interrelated through the content of their respective columns (i.e., wID, Word, Lemma, PoS, Frequency, IDseq, and TextID; see below for an explanation of these terms). As the horizontal arrow in figure 3 shows, the column wID connects these two SQL tables, which enables association of all kinds of information in them. For instance, PoS information can be related to textID via wID to investigate variation in the distribution of, say, adverbials. Joining databases in queries in this way may imply a significant computational load, but the informative potential is considerable.

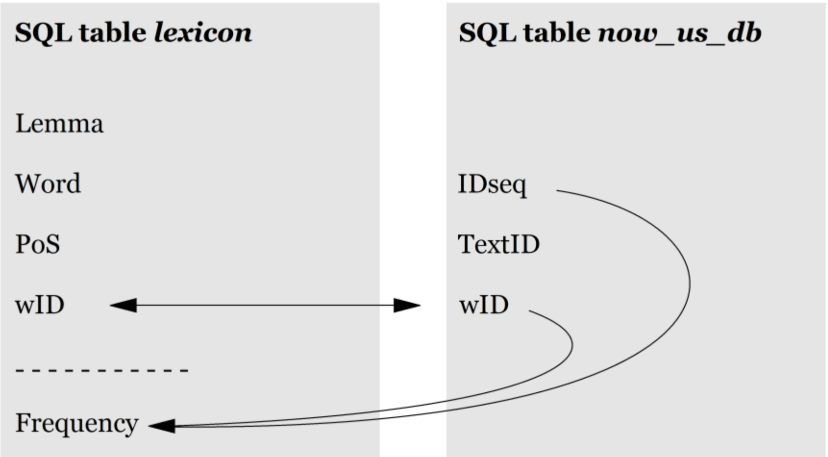


Figure 3. The combination of the SQL tables *lexicon* and *now_us_db*.

Lemma and Word

The term *lemma*, as applied in NOW, is a base form that underlies all instantiations of a word form. However, a lemma in NOW may encompass a wide variety of concepts, which are not automatically revealed in the form alone. As an illustration, *booked* occurs in the subcorpus both as a lemma in its own right and as an inflected word form of the lemma *book*. In contrast, *washed* only occurs as an inflected word form of the lemma *wash*. This displays the usage-driven annotation of NOW: the seemingly parallel forms *booked* and *washed* (i.e., inflected forms of *book* and *wash* respectively) do not have analogous patterns of occurrence, which results in their different annotation status.

The *word* level in NOW captures the textual instantiations of lemmas. It is case sensitive, but it does not distinguish between different PoS tags. For instance, forms such as *book*, *books*, *booked*, and *Book* are technically treated as different words in the corpus, while the noun *book* and the verb *book* are instead grouped together.

Word ID

The word ID (wID*)* in both SQL tables is a number (e.g., *2,940*), which is the numerical representation of textual forms (i.e., the NOW notion of *word*) combined with PoS annotation. NOW distinguishes between all PoS tag variants of a word, which implies that an item such as *train* is realized in several wIDs, since *train* occurs with several (sets of) PoS tags. Among its several recorded uses, *train* occurs 25,680 times as a singular common noun (PoS tag *nn1*) with the wID 2,940 (see row 1 in Table 2). Furthermore, rows 3 and 4 in Table 2 have identical word representations, but the PoS annotation is different, which results in two separate wIDs. In contrast, the case sensitivity of the word level is exemplified with rows 5 and 8, which also shows that the distinction between these wIDs (18,698 and 63,012 respectively) is based on the word level in this case.

Table 2. Selected output from a lemma query of *train* in the subcorpus (10 out of 69 rows).

	Lemma	***Word***	***PoS***	***wID***	***Frequency***
1	train	train	nn1	2,940	25,680
2	train	trained	vvn	4,799	15,262
3	train	train	vvi	6,461	9,819
4	train	train	nn1_vv0	9,604	7,360
5	train	train	vv0	18,698	2,722
6	train	trains	nn2_vvz	18,949	2,495
7	train	trained	vvn_vvd@	28,967	1,654
8	train	Train	vv0	63,012	395
9	train	TRAINED	vvn	679,122	10
10	train	TRAINS	vvz_nn2	754,811	1

IDseq
The most fine-grained level of analysis in NOW is represented by the sequentially ordered unique numbers (IDseq) assigned to each occurrence in the corpus. The downloaded offline version of NOW contains 5.8 billion IDseqs, while there are 1.1 billion ones in the subcorpus.

Frequency
In the *lexicon* SQL table *frequency* is short for *token frequency*, i.e., all occurrences of a specific wID. Depending on what is queried in the subcorpus, frequency data can be combined in various ways. For instance, if all occurrences of *book* as a verb are asked for, a script is designed to retrieve and summarize the frequencies of all *book* lemmas with the relevant PoS tags.

The frequency column in the *lexicon* SQL table is not included in the original NOW text files. It was imported into the *lexicon* SQL table by combining the information on every occurrence (see IDseq above) with each unique wID in the *now_us_db* SQL table (see the curved arrows in Figure 3). This adaptation was done to speed up queries. With these data permanently attached to the *lexicon* SQL table, most frequency queries could be run without joining the SQL tables, which saved a great deal of time and processing power.

TextID
The *text ID* column refers to the specific source from which a text body is retrieved, which is typically a published web article. Besides the possibility to locate the origin of any type of information, it could, for instance, be combined with frequency data to examine the distribution of a phenomenon in relation to its sources.

4.2.2 Word class and bad data
The nature of the subcorpus is in part a result of the empirical requirements of the study, but some specific characteristics are also inherited from the NOW corpus. One important concern for this investigation is the relation between the delineation of word class and the PoS annotation of NOW. Word class is framed here in two different ways with regard to how NOW is annotated. First, it is shown how the PoS tags in NOW were used to operationalize word class in this investigation. The presentation below accounts for how *common nouns* were dealt with. The delimitations of other word classes were based on the same type of analysis. Second, querying lexical items in the subcorpus provided PoS information that needed to be transformed into the operationalized word classes. The rationale and the procedures for this transformation are also dealt with in this section.

The procedures employed to deal with bad data constitute another important concern. Internet language reportedly includes disrupted and fragmented data (Crystal, 2001a). This is a challenge that needs to be tackled in this investigation too, although the sources of NOW are expected to be relatively conventional in terms of style and register compared to, for instance, chat forum language (cf. Crystal, 2001a, 2011).

Delineation of word class

According to the UCREL tagset information, common noun PoS tags in Claws 7 begin with the two letters *nn*. This information is sufficient to retrieve all items categorized as *common noun*, given that the basic definition of Claws 7 is acknowledged. The fourteen basic Claws 7 PoS tags beginning with *nn* are shown in Table 3. They occur both individually and in combinations with other PoS tags (see also Appendix 2).

Table 3. Noun tags in the Claws 7 tagset.

Common noun PoS tags	***Claws 7 definition***
nn	common noun, neutral for number (e.g., *sheep*, *cod*, *arm*)
nn1	singular common noun (e.g., *book*, *girl*)
nn2	plural common noun (e.g., *books*, *girls*)
nna	following noun of title (e.g., *M.A.*)
nnb	preceding noun of title (e.g., *Mr.*, *Prof.*)
nnl1	singular locative noun (e.g., *Island*, *Street*)
nnl2	plural locative noun (e.g., *Islands*, *Streets*)
nno	numeral noun, neutral for number (e.g., *dozen*, *hundred*)
nno2	numeral noun, plural (e.g., *hundreds*, *thousands*)
nnt1	temporal noun, singular (e.g., *day*, *week*, *year*)
nnt2	temporal noun, plural (e.g., *days*, *weeks*, *years*)
nnu	unit of measurement, neutral for number (e.g., *in*, *cc*)
nnu1	singular unit of measurement (e.g., *inch*, *centimeter*)
nnu2	plural unit of measurement (e.g., *ins.*, *feet*)

The example code below is included to show the structure of a basic SQL-script capturing all occurring PoS tags beginning with *nn* in the SQL table *lexicon*. It is worth noting that the percentage sign (%) after *nn* on the third row functions as a wildcard operator instead of the commonly occurring Boolean asterisk (*). Using the wildcard operator entails that all PoS information following *nn* is retrieved in the query, including all occurring combinations of PoS tags beginning with *nn*. Subsequently, this script captures many more PoS tags than just the basic ones in Table 3. In terms of stylistic conventions, italic type is not employed in SQL scripts, which is why *lexicon* and *nn* are in standard roman type.

```
SELECT pos, SUM(frequency) AS freq
  FROM lexicon
  WHERE pos LIKE 'nn%'
  GROUP BY pos
  ORDER BY freq DESC
  LIMIT 10
```

Running this script in the PGadmin environment generates an output in the form of a SQL table with rows and columns. The columns represent what types of information are asked for, and the rows represent its instantiations. More specifically, the above script:

- collects frequency data about each occurring PoS tag in the subcorpus (**SELECT** pos, [...] freq),
- defines from which SQL table the data were collected (**FROM** lexicon),
- defines which PoS tags should be included (**WHERE** pos **LIKE** 'nn%'),
- asks that all identical PoS tags should be grouped in one row (**GROUP BY** pos),
- asks that the output should be sorted in descending order from the most frequent PoS group (**ORDER BY** freq **DESC**), and finally
- asks that the output is limited to 10 rows, i.e., ten PoS tags (**LIMIT** 10).

The output produced by this script, as shown in Table 4, revealed that the tags *nn1* and *nn2* (*singular common noun* and *plural common noun* respectively) were by far the most frequent ones (127.5 and 47.7 million tokens respectively). Hence, a substantial number of common nouns in the subcorpus were retrieved already using this preliminary delimitation. The output presented in Table 4 also exemplifies combinations of PoS tag. For instance, a commonly occurring tag such as *nn1_jj* (row 6 in Table 4) combines *nn1* (singular common noun) and *jj* (general adjective). This tag expresses that a certain item can be categorized either as a noun or an adjective.

Table 4. Example script output.

	PoS	*Frequency*
1	nn1	127,455,922
2	nn2	47,746,560
3	nnt1	5,664,371
4	nn	3,533,071
5	nnt2	2,959,164
6	nn1_jj	2,923,055
7	nn1_vv0	2,891,305
8	nnu	1,892,704
9	nn1_np1	1,875,770
10	nnb	1,130,079
	total	***198,072,001***

A more extended examination of the distribution of PoS tags beginning with *nn* captured no less than 1,806 different tags (script P, Appendix 1). They were distributed over more than 210 million instances (tokens) spread over more than 1.2 million unique wIDs. As is shown in the frequency total of Table 4, the ten most frequent *nn%* tags covered >198 million of all tokens captured by *nn%* tags in the subcorpus. Hence, 94.2% of the total number of tokens with an *nn%* tag was covered by the ten most frequent PoS tags. The quantitative dominance of the ten most frequent *nn%* tags was less marked in the distribution of wIDs, of which 74.5% were identified with any of the PoS tags in Table 4.

Figure 4 is a visualization of the massive impact of the ten most frequent tags beginning with *nn* as regards token frequency. Even though only the 180 most frequent tags are retained in the graph, the chart line is barely distinguished from the x- and y-axes. If all 1,806 tags would have been included in the graph, the curve would practically have coincided with the axes. This illustration shows that most of the 1,806 *nn%* tags represent uncommon uses or even exceptions realized as a *long tail*, while the few most commonly occurring ones totally dominate the output. The numbers on the y-axis in Figure 4 represent million tokens per PoS tag, while those on the x-axis serve only as indications and have no intrinsic value as individual numbers in the illustration.

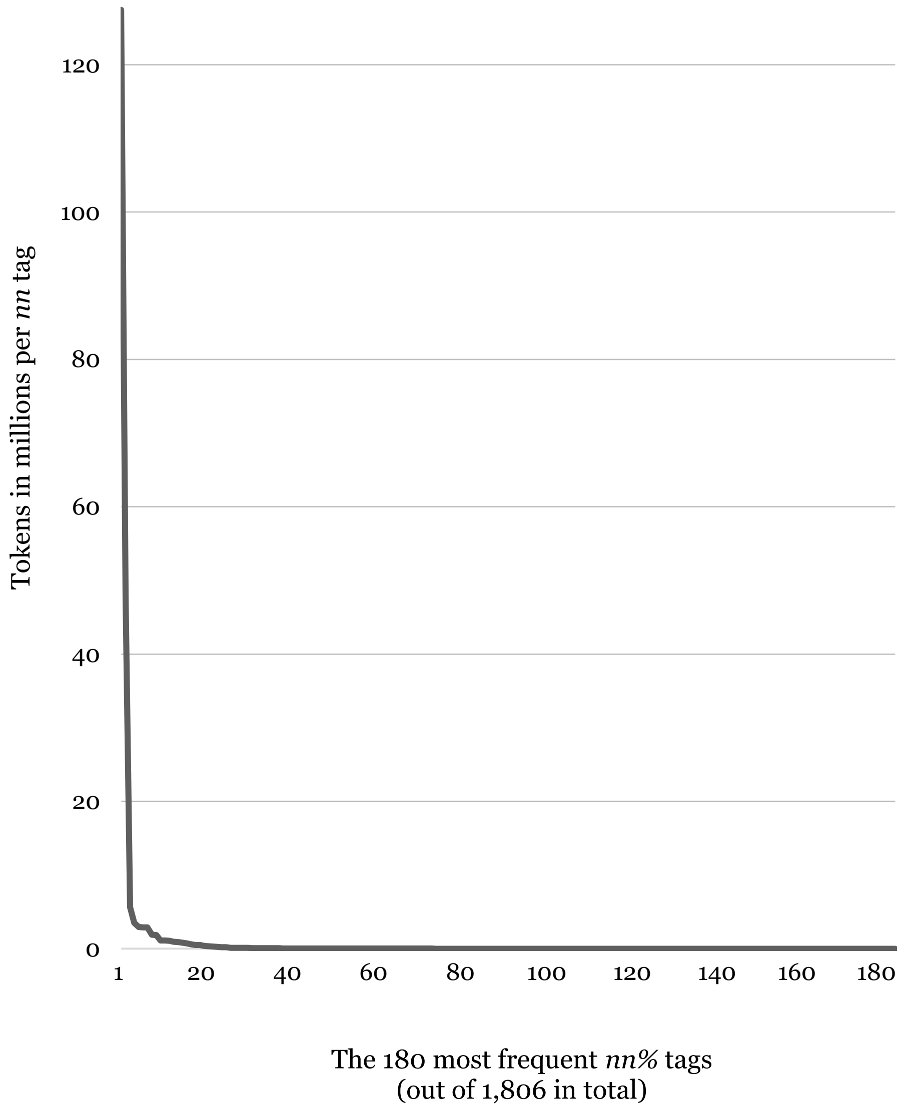


Figure 4. Token distribution of the nn% tags.

The interpretation of the PoS distribution (Figure 4) displays a few important characteristics of the dataset of methodological relevance. First, the ten most dominant tags retrieved by the *nn%* condition (**WHERE** pos **LIKE** 'nn%') capture a vast majority of the instances defined as common nouns according to the Claws 7 algorithm. The category *noun* thus has a very high level of prototype centrality within the given limitations of the empirical material. Second, the long tail of low frequency items is considerable, which mirrors the diversity of the data.

The examination of the *nn%* tags show that the applied condition satisfactorily covers *common noun* as an operationalized category in this investigation, while it does not exclude potential grammatical complexity. Instead, the grammatical complexity of the dataset naturally mirrors the well-attested grammatical complexity of English in general (cf. Quirk et al., 1985).

A final note on the topic of nouns is concerned with the category *proper nouns*. There are seven basic PoS tags for proper nouns in Claws 7: *np*, *np1*, *np2*, *npd1*, *npd2*, *npm1*, and *npm2* (see Appendix 2). These tags cover proper nouns such as *Mary* and *Paris*, but weekdays and name of months are included as well. Importantly, the structure of proper noun tags (beginning with *np*) implies that they are not captured by the *nn%* condition. Since previous studies point to the importance of proper nouns in blending (e.g. Ronneberger-Sibold, 2006; López Rúa, 2012), the technical exclusion of proper nouns may involve a potential loss of data.

The second major concern connected to word class is how the grammatical functions of individual lexical items are analyzed. This type of analysis relies on the operationalized delimitation of word class as described above. A considerable part of the data in the subcorpus is assigned several grammatical categories, which makes word class analyses demanding. This characteristic is reflected in both the corpus annotation and the analyses of the data. For instance, the blend *Spanglish* is marked as either noun or adjective, which is illustrated in examples (24) and (25). The construction in (24) signals a nominal function of *Spanglish*, whereas (25) shows that *Spanglish* is used as an adjectival modifier to the noun *slang*.

(24) I heard Spanglish constantly on the radio in salsa, bomba, plena and reggaeton lyrics (NOW).

(25) Not to mention all of the other Spanglish slang that is part of daily life in a place like Los Angeles (NOW).

The analysis of grammatical function was carried out using the PoS annotation system applied in the NOW corpus combined with contextual analyses. The operationalized grammatical categories are presented in the first two columns of Table 5 (*Category* and *Abbreviation*). The column *Claws 7 tags* lists the Claws 7 PoS tags that constitute the technical delimitation of each grammatical category. For instance, *Adjective* (row 4 in Table 5) is delimited by the Claws 7 tags *jj*, *jjr*, *jjt*, and *jk*. Blends and source words tagged with any (or several) of these tags are identified as instances of adjectival use.

Table 5. Grammatical categories.

Category	***Abbreviation***	***Claws 7 tags***
Common Noun	N	nd1, nn, nn1, nn2, nna, nnb, nnl1, nnl2, nno, nno2, nnt1, nnt2, nnu, nnu1, nnu2
Proper Noun	PN	np, np1, np2, npd1, npd2, npm1, npm2
Verb	V	vv0, vvd, vvg, vvgk, vvi, vvn, vvnk, vvz
Adjective	Adj	jj, jjr, jjt, jk
Adverb	Adv	ra, rex, rg, rgq, rgqv, rgr, rgt, rl, rp, rpk, rr, rrq, rrqv, rrr, rrt, rt
Pronoun	Pro	pn, pn1, pnqo, pnqs, pnqv, pnx1, ppge, pph1, ppho1, ppho2, pphs1, pphs2, ppio1, ppio2, ppis1, ppis2, ppx1, ppx2, ppy
Preposition	PP	if, ii, io, iw
Numeral	Num	mc, mc1, mc2, mcge, mcmc, md, mf

The analytic procedure is exemplified with the source word *exploitation* occurring in the blends *sexploitation, teensploitation* (*teen* + *exploitation*), and *jewxploitation*. The lemma *exploitation* is tagged in NOW with the PoS tags *nn1* (singular common noun), *nn2* (plural common noun), and *nnu* (unit of measurement). A token frequency analysis of *exploitation* shows that all 8,041 occurrences in the subcorpus except one (1) are tagged with either *nn1* or *nn2*. The remaining one is tagged with *nnu*. Based on these patterns, *exploitation* is assigned the grammatical category *common noun* (N).

However, not all blends and source words were analyzed as straightforwardly as *exploitation*. The lexical item *crook* will serve as an illustration. There are 3,469 occurrences with the lemma *crook* in the subcorpus, of which 3,442 (99.2%) are tagged as either singular or plural common nouns. The remaining instances are tagged as proper nouns (23 items or 0.7%) or verbs (4 items or 0.1%). Based on these observations, *crook* was assigned the grammatical categories “N, PN, V”, or *common noun*, *proper noun* and *verb*. Table 6 shows the content structure of the corpus query output based on the lemma *crook*. Each row represents a wID instance of *crook*, which is the realization in context presented in the column *Word*. The *PoS* column shows the respective Claws 7 tag, and *frequency* refers to the number of occurrences. The information in the column *PoS explanation* is not generated by the query but is included in the table for reference (see also Appendix 2).

Table 6. Query output of the lemma *crook*.

Lemma	***Word***	***PoS***	***Frequency***	***PoS explanation***
crook	crooks	nn2	1,220	plural common noun
crook	crook	nn1	838	singular common noun
crook	Crooks	nn2	736	plural common noun
crook	Crook	nn1	636	singular common noun
crook	Crook	np1	23	singular proper noun
crook	CROOK	nn1	10	singular common noun
crook	crooked	vvd%_jj	2	past tense of lexical verb (rarity marker) *or* general adjective
crook	CROOKS	nn2	2	plural common noun
crook	crooking	vvg	1	*-ing* participle of lexical verb
crook	crooked	vvd%_vvn%_jj	1	past tense of lexical verb (rarity marker) *or* past participle of lexical verb *or* general adjective

The low frequency of *crooked* in Table 6 may seem surprising, since it is a conventional inflected form of a relatively frequent word. The reason for the low frequency in this query is that NOW attributes *crooked* the status of lemma in its own right. Querying *crooked* as lemma produced 2,681 mainly adjectival occurrences distributed over seven wIDs. These adjectival items did not surface in the output presented in Table 6 because only the lemma *crook* was asked for. This observation highlights the need for caution in this type of analysis. It may be tempting to widen the analysis to include the lemma *crooked* because of its close relation to the lemma *crook*. However, Claws 7 records usage patterns, and *crooked* has apparently a significant enough level of usage integrity in the corpus to motivate it as an independent lemma. Therefore, expanding the analysis with the lemma *crooked* implies compromising with both the Claws 7 tagging principles and the rigorously employed data collection model. Such adaptations would put the validity of the quantitative analyses at risk.

Bad data

Despite the presumed conventional nature of the NOW data, a considerable portion of the downloaded lexicon material was bad data. Even though large quantities of irrelevant data were initially removed in the transfer process from the corpus text files to the corresponding SQL table (see Figure 2), the final *lexicon* SQL table still contained items that would contribute little to the investigation. Notably, they often had no lemma in the NOW corpus. Because of their low individual token frequencies, they were likely to be found in the long tail. In order to exclude such unconventional low frequency items in the random selection, an additional SQL script condition (**AND** lemma <> ' ') was added to the script designed to retrieve the CaWs.[4] Table 7 shows some examples of wordforms without a lemma in the NOW corpus.

Table 7. Examples of corpus forms without a lemma.

Word	***Lemma***	***Pos***	***Frequency***
Guild/AFTRA	-	nn1	13
Liiiiittle	-	nn1	2
Blues--	-	nn2	2
Dramatics	-	nn1_nn2	2
WavSTAT	-	nn1_vv0_np1	1
p-commerce	-	nn1_jj	1
Chih-ping	-	nn1	1
EYE/iStockPhoto	-	nn1	1
$6,495	-	nnu	1
LARIBA	-	nn1_np1	1

Due to the characteristics of the data, no simple random selection in this corpus would be unproblematic. Instead, the nature of the material and the tools to manage it highlight a recurring observation: although digital technologies offer new methodological possibilities, traditional methodological problems (such as word class category definition) still need to be addressed.

[4] The “<>” sign means “not equal to” and the two single quotes with nothing between them (i.e., ' ') signifies empty spaces in a SQL table row.

4.3 Finding the blends

Having outlined the environment in which the blends were collected, we now turn to the procedures by which the blends were retrieved from the subcorpus. There were three main tasks that needed to be resolved: how to select the CaWs (4.3.1), how to devise an algorithm that would truncate the CaWs in a systematic manner (4.3.2), and how to retrieve the textual material for the blend pond with the help of the truncated CaWs (4.3.3). Once these tasks were completed the identification of the blends according to the operational definition (3.5) could begin. The consecutive steps in the process of data collection are illustrated in Figure 5.

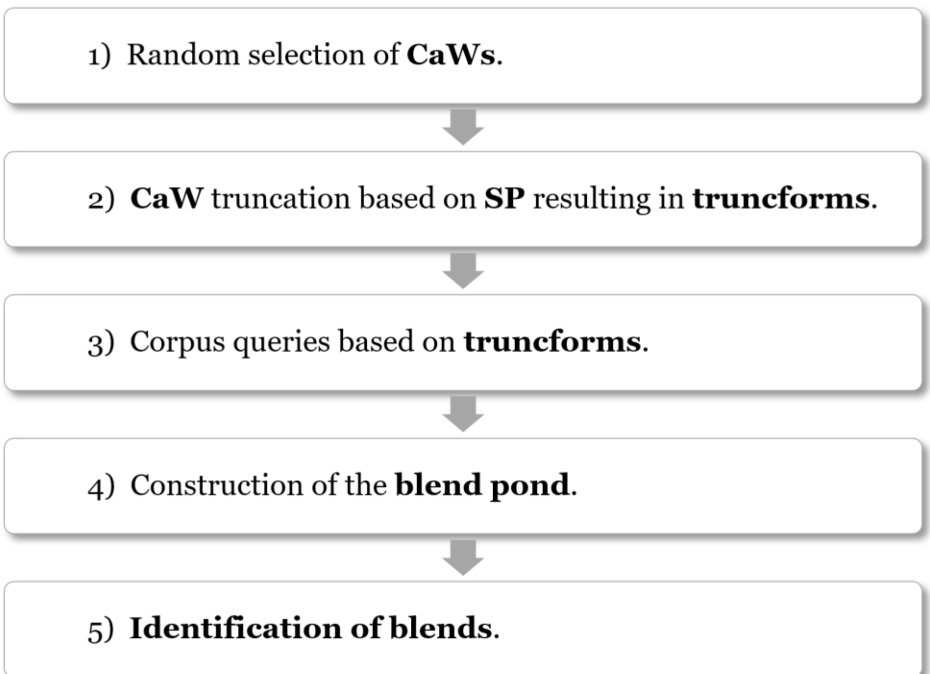


Figure 5. Steps in the data collection procedure.

One of the central methodological challenges was how to resolve the incompatibility of the huge corpus size and the requirement for manual identification of blends. This issue was addressed by retrieving a randomized sample (i.e., the blend pond) from the subcorpus that would make manual analysis feasible. Although considerably smaller, this sample would still have to contain large amounts of data, which is motivated by previous observations of the generally low level of occurrence of blends (Cannon, 1986, 2000). Therefore, the design of the sample was intended to strike a balance between a reasonable workload and a sufficient number of blends allowing quantitatively reliable analyses. The resulting sample database, or the blend pond, is both a small part of the subcorpus and the database from which the blend list is retrieved.

The first step of the data collection was a random selection of 100 words among the 2000 most frequent nouns and verbs respectively in the subcorpus (see SQL script H in appendix 1). These words were called candidate words (here CaW) since they were the potential candidates to form blends in the current data collection design.

The random selection of CaWs was repeated three times resulting in three separate datasets, of which two were based on noun CaWs, and one was based on verb CaWs (see Figure 6). There were several reasons for this type of iteration of the data collection procedure. First, comparing two separate collections of noun CaWs from an identical list of 2,000 items would be an indicator of the statistical reliability of the data collection model. Second, comparing the noun datasets with a verb dataset could be potentially informative as to the basis of blend formation. Previous research has suggested that verbs generally have a low tendency to form blends (Cannon, 1986; Elisa, 2019; Renner, 2019). Hence, it was considered interesting to see whether an attempt to foreground verbs as potential bases of blends would return differing results in comparison with the noun datasets. Repeating the data collection queries and organizing the output were relatively small efforts in comparison with the design and development of the data collection methodology. Therefore, the benefits of having significantly more blend data outweighed the extra work. Figure 6 is intended to visualize the relation between the noun and verb CaW datasets. The horizontal arrow between the noun datasets highlights the comparison between datasets produced under (as far as possible) identical conditions. The vertical arrow indicates the comparison between datasets influenced by different grammatical functions.

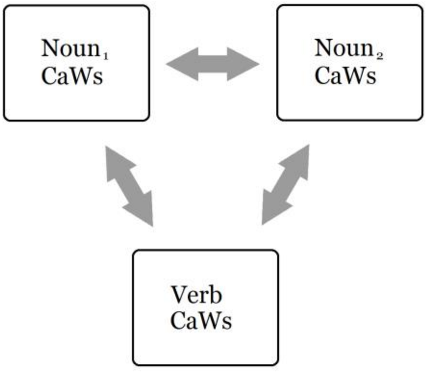


Figure 6. The three datasets.

Once the CaWs had been retrieved, they were truncated from right and left (RT and LT truncation respectively), which produced two *truncforms* per CaW. The truncforms would be the base for the queries for word lists potentially containing blends. For instance, a potential CaW *temperature* could result in the truncforms *tem** (RT) and **rature* (LT).

The chosen truncation principle, *selection point* (Gries, 2006; see section 2.2.2), was applied to all CaWs, resulting in 200 truncated items in each dataset prepared for queries in the subcorpus. The CaWs were truncated according to this strict algorithm to ensure that comparisons between CaWs could be monitored and that the output would follow systematic principles. Besides the importance for the validity of statistical analyses, a systematic truncation algorithm was necessary to make the study replicable.

When the subcorpus and the *truncforms* had been prepared, 600 queries were run in the subcorpus (see details in 4.3.3). The output of these queries was organized as word lists in Excel files. This collection, divided into three independent lists of words derived from each dataset, constituted the blend pond, and all its items were analyzed manually. Blends were then identified in the blend pond according to the operational definition in 3.5.

4.3.1 The Candidate Words

Investigating all occurrences of verbs and common nouns in the subcorpus was not a feasible option, given the elements of manual analysis in the method. Therefore, a random selection of data was necessary. The CaWs were the resulting units in this sampling process. The entire list of the 300 CaWs is found in Appendix 5.

The diversity of the subcorpus is illustrated with an examination of how the occurrences of noun and verb lemmas are distributed. Table 8 presents the lemmas of these grammatical categories in the subcorpus. The second column (*total number of lemmas*) shows how many items were captured in the subcorpus by queries for noun and verb lemmas respectively (script B, Appendix 1). The proportions of lemmas instantiated three times or less are accounted for in the third column, while hapax legomena lemmas (i.e., items occurring only once) are presented in the rightmost column. It should be pointed out that the hapax legomena are included in the measures of the column presenting items occurring three times or less. The proportions in Table 8 should be read as a funneling process from left to right, with *total number of lemmas* as 100% and the following query outputs as successively decreasing proportions.

Table 8. Noun and verb lemmas in the subcorpus, and their corresponding occurrences.

Grammatical category	*Total number of lemmas*	*Lemmas with three occurrences or less*	*Hapax legomena (one occurrence)*
Nouns	191,206	81,987 *(43%)*	44,395 *(23%)*
Verbs	35,213	14,199 *(40%)*	7,642 *(22%)*

The impact of the long tail of low frequency items is illustrated by the fact that 43% of the noun lemmas occur three times or less in the subcorpus, while restricting the query further reveals that as many as 23% of them occur only once. Thus, Table 8 shows that a considerable proportion of the nouns and verbs in the subcorpus consists of lemmas with low frequencies.

An investigation of the quantitative impact of the CaWs in the subcorpus showed that the 100 CaWs in each noun dataset covered 5.6% and 4.8% respectively of the total number of noun occurrences. The corresponding proportion in the verb dataset was 9.5%. In other words, the small selection of CaWs (lemmas) covered a large proportion of the occurrences in the subcorpus. This distributional pattern was expected given the limitation to collect the CaWs from the 2,000 most frequent lemmas. All the same, the uneven distribution of occurrences proved considerable even when applying this limitation.

The 2,000 most frequent common nouns occur 165,142,603 times (script G), which represents 80.3% of all the common noun instances in the subcorpus. The corresponding figures for the verbs are 94,452,988 tokens, which is 96.4% of the verb tokens in the subcorpus. These measures show that the Zipfian distribution characteristics observed in the subcorpus as a whole, is a rather prominent characteristic of the 2,000 most frequent nouns and verbs when analyzed separately. Figure 7 visualizes the distribution of these common noun lemmas in the subcorpus. The most frequent common noun represented in the graph is *year*, which occurs 2,249,523 times in the subcorpus. The corresponding item with the lowest frequency rank of the 2,000 common nouns is *cemetery*, which occurs 17,143 times. Thus, a random selection of these items is presumed to retrieve lemmas with at least some degree of conventionality.

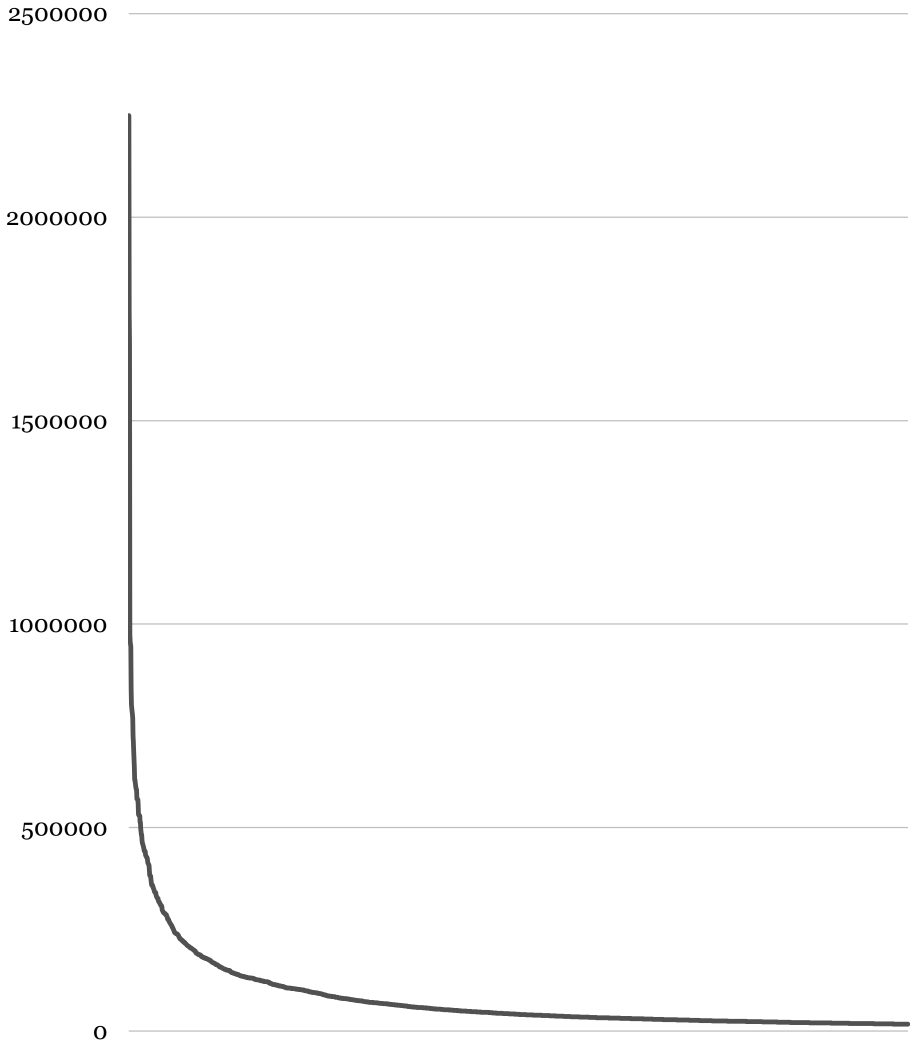


Figure 7. Token frequency distribution of the 2,000 most frequent common nouns in the subcorpus.

The random selection of the first dataset was performed in October 2018, as a SQL query (script H, Appendix 1), and the result was archived in an MS Excel file. The CaWs were then distributed in ten separate files containing textual and quantitative information generated by truncated queries based on the CaWs. These data (dataset NN_1) were subject to multiple analyses and pilot tests before the data collection was repeated twice, in August 2019 and September 2019, when the second and third datasets (VV and NN_2 respectively) were retrieved and organized.

4.3.2 Truncation method

Truncation of source words is a recurring issue in studies on blends. It has conventionally been assumed that this process is largely arbitrary (see for instance Bauer, 1983: 235), but several accounts in the growing domain of blend research point out that certain patterns in the structure of blends can in fact be identified (e.g. Kelly, 1998; Gries, 2004b; Bauer, 2012; Beliaeva, 2014a). Still, it is no trivial task to determine how much of a CaW should be truncated so as to make sure that the same type of truncation is applied to all the items in the data. This section outlines the rationale and the methodological choices of truncation as applied in the present investigation.

Statistical data in previous studies have shown that structural patterns in blends are typically a matter of probabilistic tendencies rather than categorical rules (e.g., Bauer, 2012; Gries, 2012). It follows from these observations that, while typical patterns of breakpoints can be identified (see for instance Kelly, 1998; Gries, 2004b), it is impossible to establish sharp categorical boundaries, or rules, predicting at what point a source word should split. Truncations of source words in blends are simply too complex. Thus, one of the methodological challenges was to devise an algorithm that could truncate the CaWs systematically with the intention to create the truncforms that would constitute the basis for the collection of blends. The design of this algorithm was based on two requirements. First, it had to be systematic in the sense that the truncforms should not be arbitrarily formed. Second, the truncforms should ideally reflect a maximally productive form retrieving as many blends as possible.

The chosen method to truncate the CaWs makes use of Gries' (2006) notion of selection point (SP; see also 3.4.2). Gries (2006) shows that the breakpoint at which a certain word is recorded as the most frequent item corresponds rather well with the truncation patterns of blending, which is why it was considered well suited for the purposes of the current investigation.

Table 9 illustrates a practical example of an SP calculation. The truncated form *routi** produced 41 wIDs in the subcorpus, and in a frequency analysis the CaW *routine* is the top frequent word (*Freq rank* in the table). Thus, the operationalized truncform is *routi**.

Table 9. Selection point calculation.

Truncation	*wID Freq*	*Token Freq*	*Freq rank*	*Truncform*
r	60,424	24,645,360	183	
ro	9,759	2,432,174	20	
rou	1,156	312,495	4	
rout	134	102,831	2	
routi	41	38,744	1	**routi***
routin	39	38,739	1	
routine	20	37,714	1	

Despite the technical advantages, using the concept of SP calls for a few comments. First, Gries (2006) suggests that there is a correlation between SP and the breakpoints of blends, which may imply that the procedure could appear circular as the query output would rely on a seemingly predefined narrow definition. However, this is not the case. The dataset of Gries (2006) may contain problematic issues as to the origin of the data, but the large size of the blend database ensures that his findings are statistically reliable (see also Wulff & Gries, 2019). Moreover, the complexity of breakpoints in combination with well-known patterns of structural overlap and sound similarity extend the structural potential of a truncform to include far more than the truncation alone might suggest.

Although the theoretical arguments were in favor of employing SP as a basis for a truncation algorithm, a number of pilot tests were carried out, in which a CaW such as *routine* was tested in queries based on, for instance, *routin**, *rout**, *rou**, and *ro**. The tests showed that more generous truncations (i.e., retaining more of the source words) had very little effect on the number of retrieved blends. Narrow truncations (such as *r** and *ro**), on the other hand, produced large outputs with considerable quantities of irrelevant data, thus increasing the manual workload without retrieving any further blends or other types of meaningful data.

Second, most CaWs in the data were not as straightforward in terms of SP calculation patterns as the one in Table 9 (and certainly not as in Gries' 2006 original demonstration of the source word *agitation*). Taking these comments into consideration, SP was considered sufficiently accurate in this context, especially as its frequency-based algorithm (scripts J, K, and L in Appendix 1) enabled the truncation to be controlled with quantitative measures. The use of a systematic procedure to generate the truncated forms was in fact necessary to allow quantitative analysis.

4.3.3 Queries based on truncated forms

This section outlines some technical and methodological aspects of the initial queries based on the truncforms. Running the SP scripts (J, K, and L in Appendix 1) implied time-consuming manual work since all 600 truncforms derived from the 300 CaWs were processed individually. The data needed to be organized and stored with care to make the analysis efficient, and to ensure that the information could be retrieved easily on any occasion. When all truncforms had been processed, the entire collection of the query outputs was organized in Excel files. The resulting material was the blend pond.

Queries were run for each truncform (script K, Appendix 1), which produced 600 lists ranging from only a few instances to several hundred items each. The CaWs were managed in groups of ten words stored in thirty Excel files containing the query output of 20 truncforms each. The output of the SQL queries in pgAdmin was copied into MSExcel spreadsheets, in which all retrieved items were analyzed. Identified blends and borderline cases were annotated and copied to separate tabs in a master document.

One tab in each of the thirty Excel spreadsheets was dedicated to a summary of quantitative measures with information about the number of retrieved lemmas and the number of blends. These summaries were then copied into the master document. Figure 8 shows a screenshot from one of the master document tabs containing the quantitative summary from the second dataset (VV) based on verb CaWs.

	A	B	C	D	E	F	G
1	**STATS for all vv% CaWs**						
2							
3	Total no of blend pond lemmas			**31474**			
4	Total no of lexical blends			**70**			
5	Proportion lexical blends			**0,22%**			
6							
7							
8	**1-10**						
9	Right	*vv% 1-10*		stock	withhold	disguise	compose
10	Trunc	Total number of lemmas	**1577**	149	7	32	18
11		Number of lexical blends	**4**	0	0	1	0
12							
13							
14	Left	*vv% 1-10*		stock	withhold	disguise	compose
15	Trunc	Total number of lemmas	**1488**	823	4	2	5
16		Number of lexical blends	**1**	1	0	0	0
17							
18		*Sum number of lemmas*	**3065**				
19		*Sum lexical blends*	**5**				
20							
21		Proportion lexical blends	**0,1631%**				

Figure 8. Example screenshot of the dataset summary.

Throughout the process of blend identification in the >100k word blend pond, a large number of SQL-queries were run to control the accuracy of the output. Contextual analyses, on the other hand, were carried out in the NOW online interface for reasons of efficiency. The initial destruction of 5% of the data (see 4.1) also involved a risk, however minimal, that contextual analyses would be affected in the offline version. For the contextual analyses, only US texts retrieved up to March 2018 were taken into account, so as to maintain the temporal and regional limitations of the subcorpus.

The final blend lists in the master document were complemented with information on grammatical category annotation, truncation patterns, graphemic dependence, intercalation, semantic domain, and the relation to their respective CaWs and source words. Each blend was also marked with information about its origin in the blend pond, so as to enable fast and easy reference to the source. These lists were subject to revision several times during the processes of analysis.

4.4 Approaches to analyses of the blend data

The various analyses of the collected blend data were carried out within the theoretical framework of CL outlined in chapter 2. This vantage point has several consequences for the interpretive processes and claims of the study. One particularly important issue concerns categorization, since much of the analytical work involved categorical assessment both as to what kinds of categories were considered, and how category delimitation was addressed. For these reasons, we return to some theoretical issues with a special focus on how the analyses of the blend data were dealt with.

First, the categorization of a lexical item such as *book* followed the principles of prototype theory, which also applies to metalinguistic concepts such as noun or lexeme (cf. Barsalou, 2012: 241). Hence, in the analyses of the data both instances of conventional metalanguage (e.g., *noun*) and suggested conceptual categories were thought of as more or less typical instantiations in relation to central prototypes (cf. Rosch, 1978: 39). As an illustration, a prototypical blend segment originates from a lexeme and derives its meaning from the decoding of its source word (Kemmer, 2003: 77). Such a segment (e.g., *br-* from *breakfast* in *brunch*) would not be assumed to have an independent conventionalized meaning. In contrast, *-erati* from *literati* (in for instance *glitterati*, *digerati*, and *twitterati*) might be seen as a less prototypical truncated segment, since the occurrence of series of items suggests that *-erati* could be used more independently than *br-*. Thus, there may be independent symbolic content in *-erati*, although its degree of conventionality is low. However, the segment *-erati* could still function as a component of a lexical blend from the point of view of prototype categorization.

Second, encyclopedic and pragmatic knowledge were taken into consideration when addressing the blend data (cf. Langacker, 2008: 38). The semantic configuration of a specific blend might therefore be influenced by the knowledge spectra of the users, and how it could be construed in a specific context. The latter perspective, the context, is directly observable in the current investigation. In contrast, encyclopedic knowledge of the affordances of a truncated segment may only be indirectly inferred in corpus examples such as (26) and (27).

(26) These days, the walkathon business boasts its own event producers, consultants, trade shows and technology vendors. [...] In the past five years, a period when many nonprofits saw contributions drop, "thon" revenue grew 13 percent, and participation grew even more (NOW).

(27) a proclamation to rename State College "City of THON" the weekend of Feb. 16, a super tradition, in honor of the 46-hour dance marathon (NOW).

The use of citation marks in the examples indicates a metalinguistic awareness of *-thon* as a segment that has unclear morphemic status and thus appears to require special attention in the text (cf. Lehrer, 1998).

Third, it is assumed that the conceptual integration of the source words in a blend may include metaphorical and metonymic configurations. These conceptual structures have been shown to be ubiquitous in language use (e.g., Lakoff & Johnson, 1980; Lakoff, 1987; Barcelona, 2003b; Croft, 2006; Lakoff, 2006; Peirsman & Geeraerts, 2006; Langacker, 2008), which is why they might be expected in this context as well. As pointed out in 2.3.3, some previous studies have highlighted the figurativity of blends, but this perspective also merits attention in the current investigation.

Finally, the symbolic claim of CL acknowledges meaning potential on several levels of conceptual organization. Analytic categories were, thus, not limited to issues of formal structure, but could also concern higher order schemas, such as constructional patterns (e.g., Goldberg, 2005; Booij, 2010). The functional use of ambiguity is another example of meaning potential residing in a broadly observed cognitive phenomenon (cf. Langacker, 2008: 55).

4.5 Limitations

There are obvious limitations of the empirical material despite the rigor applied to the various data collection procedures. Temporal, regional, and some distributional limitations have already been accounted for in this chapter. Another limitation is connected to the identification of blends in the blend pond:

the operational definition (3.5) serves the purpose of discriminating between blends and other items. Membership categorization operationalized in this way is necessary to obtain quantitatively analyzable data. Yet, it has been shown in chapter 3 that that blends often resist clear-cut classification, and that borderline cases are common. Hence, demarcation problems cannot be avoided, and call for analytic scrutiny and caution.

The decision to use *lemma* as the critical category in this investigation is motivated by a limitation of form and a limitation of procedure. First, the use of *lemma* as the level of analysis leaves out inflected and derived variants. Including forms such as *trains*, *trained*, and *trainer* would likely increase the dominance of already highly frequent items, which would potentially reduce the diversity of the data, and subsequently the range of identified blends. The *word* category in the NOW corpus (see 4.2.1) is case sensitive, which in turn would also constrain the output significantly. Second, the manual analysis required quantitative restrictions. Pilot studies showed that querying for the truncated form *eg** using *word* as the level of analysis returned 785 hits for manual analysis. Running the same query using *lemma* returned 323 hits. Furthermore, the graphemic variations *Eg**, *EG**, and *eG** were added to the analysis based on *word* so as to capture the greater variation in the *word* column.[5] These queries returned 822, 199, and 52 hits respectively. Thus, in order to encompass the same scope in the *word* queries as in the *lemma*-based query, the manual analysis would have to address 1,858 items, which is nearly six times the *lemma* output of 323 hits. Subsequently, the blend pond would likely grow to more than 600,000 items if the data collection would be based on *word* as the analytic level. Such an increase would make the amount of manual work unreasonable.

The issue of case sensitivity has several theoretical implications that need to be addressed separately, while the increased amount of data for manual analysis may possibly be resolved with more elaborated computational tools. The SP calculation procedure could, for instance, have been developed further to reduce much of the manual work. In addition, the manual analysis of the output of the queries could have been reduced by using a tagged reference lexicon filtering out a portion of the lexical items (cf. Renwick & Renner, 2019). Such an approach would likely increase the efficiency of the identification of lexical blends, but it could still not guarantee the elimination of a potential loss of data. Nevertheless, should the data retrieval be extended to a *word*-based data collection in future studies, extensive development of computational methods would be inevitable.

The registers and stylistic conventions of news texts constitute important discursive limitations to the material. Despite reports that blends are often found in the context of news media, studies in recent years have shown that other

[5] Other types of graphemes, such as hyphens, were excluded in this control test.

contexts may be equally productive in this sense (e.g., Cook, 2012; López Rúa, 2012; Renwick & Renner, 2019; Tarasova & Beliaeva, 2020). This does not challenge the methodological design described in this chapter, but instead suggests that alternative contexts may be subject to similar systematic collections of data. In fact, replicating the current methodology in other genres may contribute further information on sociopragmatic concerns of lexical blending (cf. Fandrych, 2008).

4.6 Summary

Chapter 4 has accounted for the rationale, procedures, and limitations of the chosen model to collect the data. The complexity of the model is a result of the requirement for systematicity in all stages of the data collection in combination with the low frequency and intricate characteristics of blends. The data were retrieved from the NOW corpus because of its considerable size and its relatively contemporary material. The retrieval of the specific data needed in this investigation proved to be highly demanding as regards processing power. Therefore, the corpus was downloaded as an offline version. The offline corpus did not come with a user interface of the type used in the online version. Instead, the management of the data was operated by the programming language SQL. The final corpus database consisted of a regionally and temporally limited body of data that had been organized in the pgAdmin software environment. This entity was called the subcorpus. A number of queries were designed to collect three randomized samples of nouns and verbs from the subcorpus. After having ensured that the selectional rationale was reasonable, three datasets were retrieved, which would be the material constituting the subject of manual identification of blends. This identification followed the principles of the operational definition as described in 3.5. The analyses of the blend data were then to be carried out from the background of a CL approach to meaning, structure, and categorization.

5 Identification of blend characteristics

This chapter addresses the general characteristics of the blends identified in the blend pond (4.3). In section 5.1 the basic properties of the blend pond are presented, such as the overall structure and size of the data, the proportion of lexical blends, and some important methodological considerations.

Section 5.2 takes a quantitative approach to the analysis of blend characteristics. The foci of attention are issues that have attracted much attention in previous blend research: *grammatical function*s, *semantic domains of usage*, and *seriality*. Section 5.3 deals with four characteristics emerging from the analyses of the data: *structural profiling*, *domain proximity*, *pseudomorphemic transfer*, and *seriality*. They originate as heuristics in the qualitative analyses and have proven to highlight key attributes of the blend data. One subsection deals specifically with figurativity and iconicity in the two characteristics structural profiling and domain proximity (5.3.3). Section 5.4 concludes the chapter and suggests some areas for further research.

Although the blend pond has been divided into three datasets for statistical reasons (see 4.3), the analyses following 5.1 address the data as a whole. When a distinction between the three datasets is motivated, they are here referred to as NN_1, VV, and NN_2, which is derived from the SQL script code for nouns and verbs (4.3). Their number indexing and order of presentation reflect the chronology of the data collection procedure as described in 4.3.

The analyzed blends are sometimes compared to those from other sources, henceforth marked with the degree sign (°) according to the pattern °*blend*.

5.1 The blend pond

As described in 4.3.3, the blend pond contains the output of the truncform queries of the subcorpus. This output consists of a large number of lexical items of various types. It includes conventional expressions such as *trendline*, *planetary*, and *symbolism*, as well as infrequent and nontraditional lexical items, for instance *silicone-gel-filled*, *40-milligram* and *urbanachampaign*. There are blends that match the operational definition in 3.5, for instance *egosystem* (*ego* + *ecosystem*) and *snowpocalypse* (*snow* + *apocalypse*), but also borderline cases that have not been included in the blend list, such as *helipad* (*heli*(*copter*) + *pad*) and *Koreagate* (*Korea* + (*Water*)*gate* (*scandal*)).

The presentation of this body of data marks the shift from automated data collection procedures to manual investigation. Both quantitative and qualitative analyses were made to examine its various properties.

5.1.1 The blend pond in numbers

The blend pond contains 100,658 lexical items in total, including duplicates between and within the datasets. The blend identification process (4.3) resulted in a list containing 206 blends. There are 23 duplicates among these, and two blends occurring three times. Removing the duplicates from this list would have reduced the number of blends to 179 unique items (206-23-(2*2)). The list of blends in Appendix 3 contains only these unique forms, but since the quantitative analyses are based on the total (type) occurrence of 206 blends, the list of unique items serves only the function of presenting the identified forms.

Table 10 presents the frequencies and proportions of the three datasets NN_1, VV, and NN_2. The column *mean output per CaW* shows the mean number of items that each query resulted in. Because each dataset is based on 100 CaWs, the column showing *number of lemmas* is by necessity the mean output multiplied with 100 (disregarding the approximation of the mean). The column displaying the *number of lemmas* highlights the actual number of lexical items analyzed manually in each dataset. The technical database term *lemma* is retained in Table 10, as well as in the rest of the text, in order to highlight the connection to the procedural descriptions in chapter 4. The presented numbers of blends include duplicates, and the *percentage of blends* column shows the result of calculating the proportion of blends per total output of lemmas. The percentage should be treated with some caution, since the sampling of the subcorpus is based on high frequency CaWs. Here it functions mainly as a coefficient enabling quantitative comparisons of the three datasets. A comparison of the proportion of blends in each dataset reveals whether the data collection procedure is statistically reliable. Nonetheless, it also gives an indication of the low frequency of blends in use, which is an issue that has been the subject of debate in previous research (e.g., Cannon, 1986, 2000).

Table 10. The three datasets constituting the blend pond.

Dataset	***Mean output per CaW***	***Number of lemmas***	***Number of Blends***	***Percentage of Blends***
NN_1	333	33,256	66	0.20%
VV	315	31,474	70	0.22%
NN_2	359	35,928	70	0.19%
Total	-	*100,658*	*206 (179)*	-

The lemma output shows little variation. Each of the datasets NN_1, VV, and NN_2 contributes 33.0%, 31.3%, and 35.7% respectively to the blend pond. These similarities are interpreted as evidence of the statistical soundness of the NOW corpus data, which means that the CaW selection (script H, Appendix 1) was applied to data that were well suited for random collection. Furthermore, the percentages of the blends in each dataset signal the reliability of both the automated data retrieval and the manual identification procedure based on the criteria described in 3.5.

The consistent output in the three datasets attests to the statistical robustness of the applied methodology. It is highly unlikely that the observed consistency could have been the result of inadequate criteria, and it is equally improbable that the precision of the proportions could have been obtained if the distribution of blends in the data had been unsystematic. In combination with the distributional tests of the subcorpus in chapter 4, this general overview of quantitative properties convincingly displays the empirical robustness of the investigation.

There is considerable variation as to how many lemmas each truncated form of every CaW produced. Figure 9 illustrates the distribution of the output in all three datasets. Each marker represents a CaW, and the 100 CaWs per dataset are ordered in decreasing type frequency from left to right. The datasets are distinguished by the three symbols lozenge, square, and triangle. The y-axis represents the number of lemmas produced by each CaW. For instance, the leftmost three data points show that each dataset contains one CaW that produced approximately 5,000 lemmas. These CaWs were *percent* (NN_1), *tell* (VV), and *school* (NN_2), and they contributed 5,309, 5,215, and 4,770 lemmas each, which constitutes 15.2% of the entire blend pond. The output of most other CaWs was substantially smaller. The graphic representation of the output distribution in Figure 9 is, therefore, characterized in terms of a long tail distribution (cf. 4.2.2 and 4.3.1).

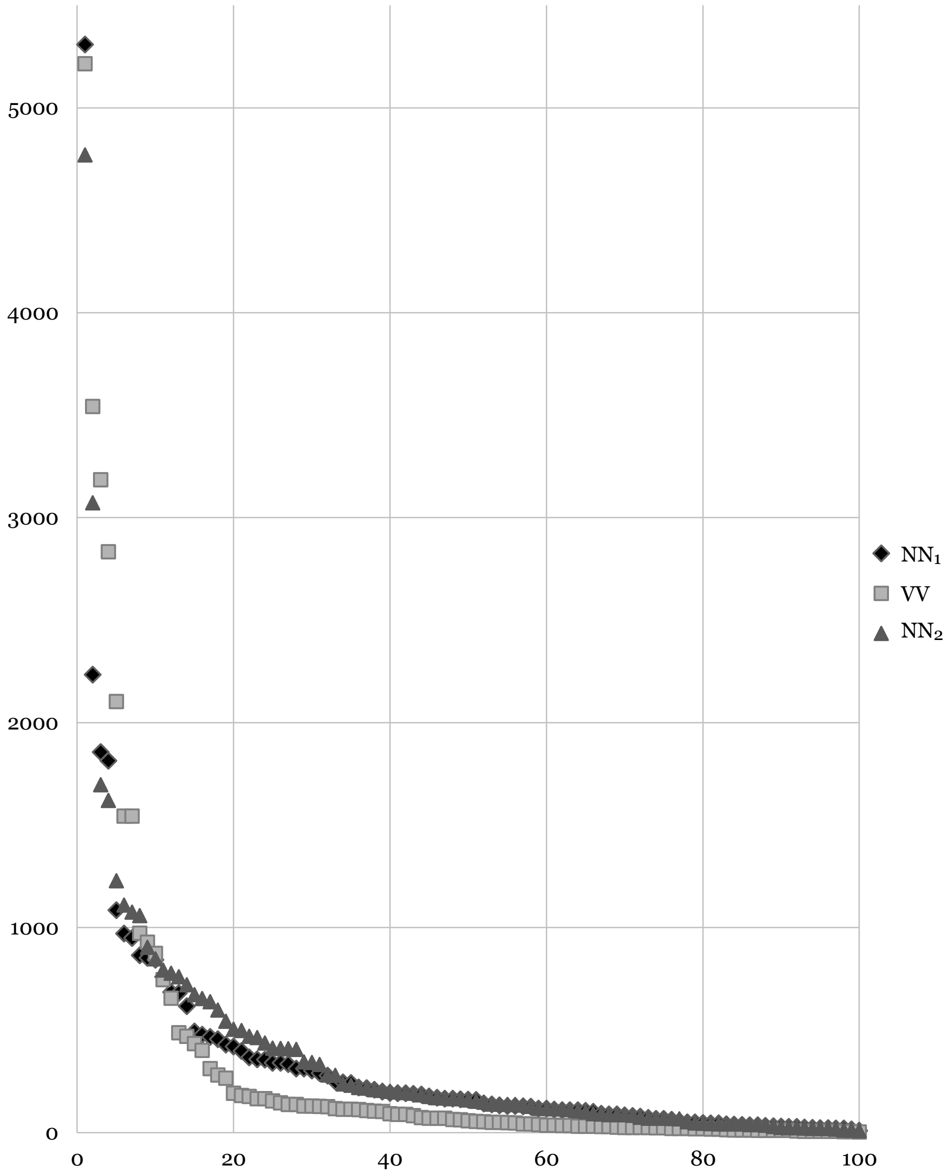


Figure 9. Quantitative distribution of the CaW queries output.

A minority of the CaWs have been responsible for a large part of the blend pond. The query outputs of the ten CaWs that produce the highest output numbers constitute 56.5% of the entire blend pond. A comparison between the datasets reveals that the tendency for quantitative dominance of a few high frequency items is even more marked in the VV dataset. The ten verb CaWs that generated the highest output made up 72.2% of the output of all verb CaWs. The corresponding proportions of the ten most productive CaWs in the noun datasets NN_1 and NN_2 are 50.4% and 48.4% respectively. This distribution pattern is slightly less prominent, while yet distinguishable, in the distribution of the identified blends.

In order to examine possible relations between distributional patterns of lemma output and identified blends, the CaWs were divided into groups of ten, based on their output frequency rank. The type frequency totals of CaWs and blends were calculated for each group, and were then divided by the type frequency totals of their respective dataset. Figure 10 shows the result of this analysis as proportions per dataset. The intervals on the x-axes (e.g., "11-20") represent the frequency rank of the CaWs included in this category. For instance, the CaWs with frequency rank 1-10 produced 50.4% of the total lemma output in the NN_1 dataset, while the proportion of identified blends in the same dataset and group is 34.8%. The proportions are presented in numbers for the CaWs with the highest output frequency in each dataset (i.e., the interval 1-10) so as to underscore the difference of this category in all three datasets. For reasons of legibility, explicit representations of the proportions are excluded in all other interval categories, since the general tendency is satisfactorily displayed without them.

The distribution of identified blends correlates quite well with the lemma output per CaW, which is illustrated in Figure 10. However, the correspondence between the lemma output and the blends seems less apparent in the high output CaWs (rank interval 1-10). This tendency could be an indicator of a frequency threshold, where blending may be less attractive as a word formation strategy. One possible explanation for the lower proportional representation of blends in the high output CaWs may be the constraining impact of lexical neighborhood (cf. Schmid, 2020: 54). The interval 41-50 equally stands out in all three datasets, as the proportion of blends is markedly higher in all three datasets. The mean query outputs for the NN_1, VV, NN_2 datasets are 175, 72.3, and 176.5 respectively, which might be an indication of a type frequency balance point where blends optimally occur. There is a similar tendency in the interval 61-70, although the low frequencies afford less generalization.

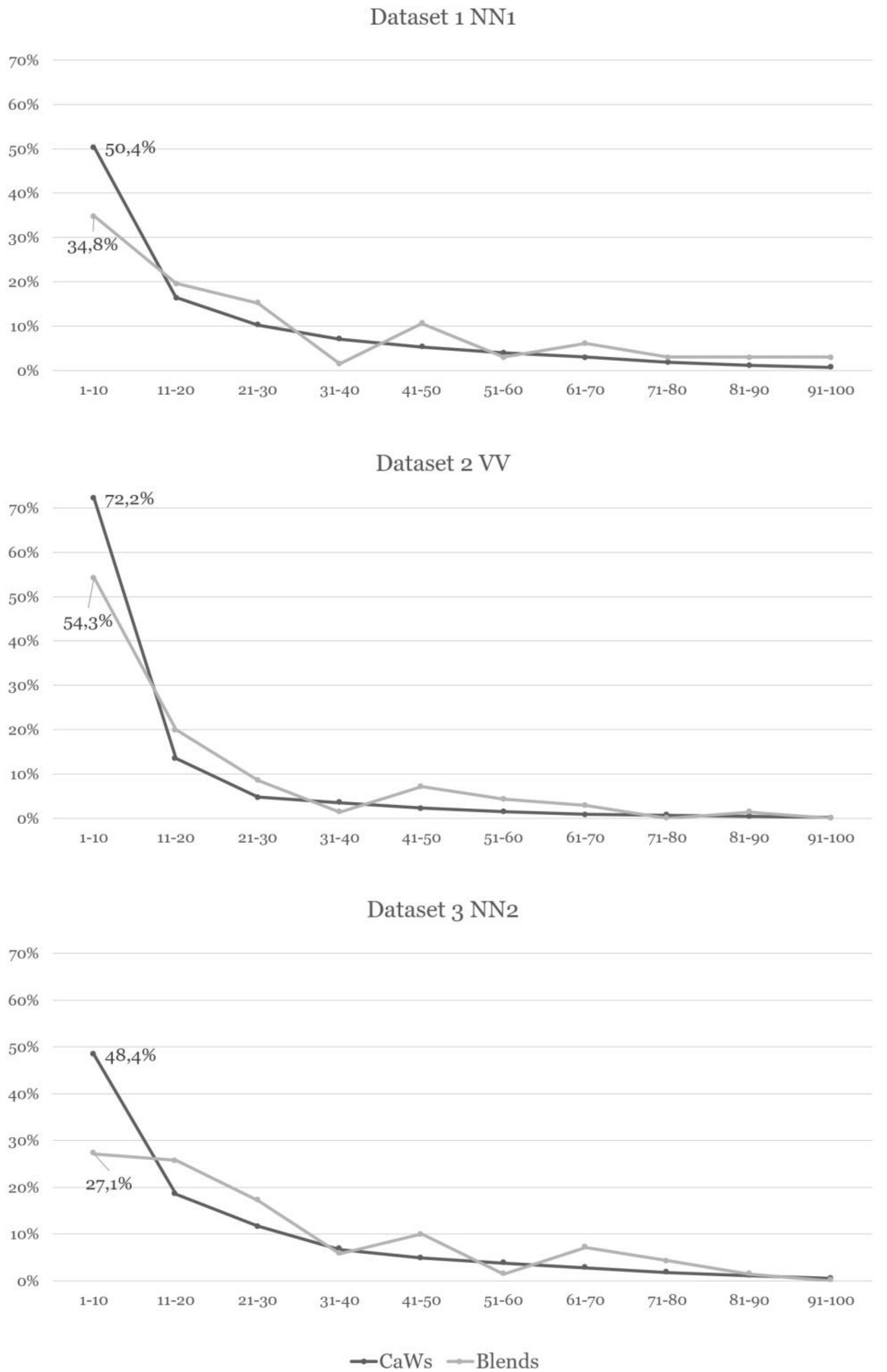


Figure 10. Type frequency proportions of CaWs and blends in the three datasets.

The quantitative observations presented here are particularly important, since they indicate the accuracy of the data collection model, which is central to the analyses of the investigation. The investigation of the quantitative properties of the blend pond shows that the data can be claimed to be systematic and, hence, that they are representative of language use within the scope of the subcorpus. Although the blend pond covers a large domain of language use, it should be emphasized that its characteristics are not assumed to be representative of American English in general, let alone English globally. It has been pointed out that blending seems particularly favored in English (Kemmer, 2003: 70), but the variation of English worldwide implies that it may not necessarily be distributed in the same way across different varieties. Moreover, the many reports of the informal character of blends suggest that usage patterns may also be sensitive to genre (e.g. Fandrych, 2008; Balteiro & Bauer, 2019).

5.1.2 Duplicates

Ideally, each CaW and query output item would represent a unique form in the blend pond. However, this was not the case. There were duplicates among both CaWs and blends, which raises important questions about the nature of the method and the data. The duplicate CaWs and blends are presented below, followed by a discussion on issues relevant for the investigation.

Five CaWs occur twice among the 300 CaWs in the blend pond: *other*, *intervention*, *English, roll*, and *date*. The design of the script collecting the CaWs (script H, Appendix 1) makes it technically impossible to retrieve duplicates within a dataset. Instead, the recorded duplicates occur between the three datasets. The items *other*, *intervention*, and *English* occur once in both the NN_1 and NN_2 datasets, while *roll* and *date* are found in the datasets NN_1/VV and NN_2/VV respectively. Figure 11 shows the duplicate CaWs, with *roll* and *date* marked to highlight their distributions.

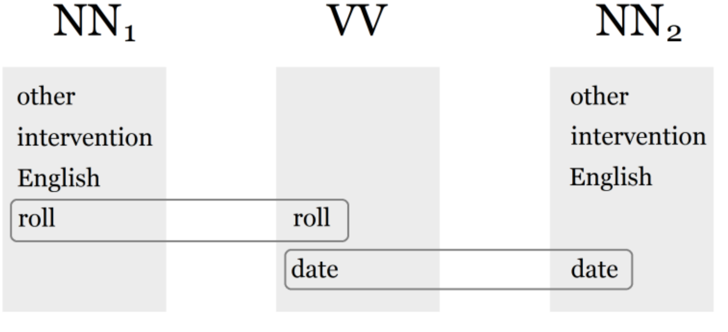


Figure 11. CaW duplicates.

Duplicates between the noun datasets were possible, since each randomly selected sample (100 CaWs) was taken from an identical set of the 2,000 most frequent items. The relatively large proportion of CaWs (5%) implies a comparatively high probability that an item would be retrieved more than once in iterated sampling procedures. It was perhaps less likely to find duplicates between a verb and a noun dataset, as these samplings were made from two different lists of words (i.e., high frequency verbs and nouns respectively). The fact that two duplicates were registered between the noun and verb CaW datasets is explained as instances of conversion, i.e., there were identical items tagged as both noun and verb. The duplicates *roll* and *date* were studied in detail to compare their word class distribution (script M, Appendix 1). It was found that *roll* is twice as common as a verb, and, conversely, that *date* is twice as common as a noun. This gives a preliminary indication that there is a potential for grammatical flexibility in the blend data.

The proportion of duplicates in the blend list was higher than among the CaWs. Removing duplicate items from the blend list would have reduced it by 13.1%, whereas a corresponding reduction of CaWs would have been 1.7%. This difference is expected, since blends derived from duplicate CaWs multiply in the blend list. For instance, the truncform **glish* resulting from the CaW *English* occurring twice among the CaWs. In the process of identifying blends, **glish* was responsible for five blends, which (necessarily) produced ten items in total registered as duplicates. Moreover, there is no technical obstacle to the occurrence of duplicate blends within datasets, which is reflected by the fact that four blends occur twice within the same dataset: *Manulife* (*manufacturer* + *life* (*insurance*)) and *permalloy* (*permeable* + *alloy*) in dataset NN_1, *teensploitation* in dataset VV, and *Dragonet* in dataset NN_2. In this part of the data collection procedure, it is worth noting that queries based on identical truncforms generate identical output.

The question then arises what analytical consequences the occurrence of duplicates has. Quantitative observations of attributes such as truncation patterns and intercalation rely on the rigorous quality of the data collection algorithms. Therefore, removing duplicates could have invalidated the quantitative results, since all occurrences are indisputably the result of the corpus queries. One way to reduce the number of duplicates in the blend list would be to collect all 300 CaWs in one query. However, such a solution would eliminate the possibility to evaluate the robustness of the data collection method, since comparisons between several samplings would no longer be possible (see 4.3 and 5.1.1). In addition, a single dataset design would still be no guarantee against duplicates, which means that the epistemological issue remains, albeit possibly to a lower degree.

For the above-mentioned reasons, duplicate CaWs and blends are retained in the data, because all occurrences carry information about the characteristics of blend formation. Qualitative observations may not be affected in any critical way by this choice, while keeping duplicates in the data is indispensable for the quantitative analyses.

5.1.3 Borderline cases

The assessment of the items in the blend pond involved dealing with word forms that were difficult to categorize in a straightforward manner. Such items are referred to as borderline cases. Some of them are included in the blend list, while others are not. All such decisions were based on the operational definition in 3.5.

Examples of borderline cases that are excluded are *heliport*, *Lewinskygate*, *dreamscape*, and *plantocracy*. They are formed from combining forms *heli-*, *-gate*, *-scape*, and *-cracy* in combination with a conventional morpheme. Thus, they are typical instantiations of the B.4 criterion in the operational definition, which excludes items based on a combining form and another morphologically analyzable segment (see 3.5). Despite their seemingly clear status in this investigation, they are not lightly dismissed as blends. For instance, *heliport* may be interpreted etymologically as an AD truncation structure (*heli~~copter~~* + *~~air~~port*). If the conventionalized uses of *heli-* and *port* (as operationalized here) are disregarded, it could be argued that *heliport* should be considered a blend. However, *heliport* was excluded from the list of blends, since both *heli-* and *port* were identified as items with morphemic status in the selected dictionaries.

A perhaps even trickier case is the lexical item *Dollywood*, which is the name of a theme park in Tennessee, USA (Dollywood Parks and Resorts, n.d.). The initial segment is based on the proper noun *Dolly* (*Parton*). Furthermore, *Dollywood* has a striking similarity with *Hollywood* and *Bollywood*, which are conventionally associated with entertainment and the film industry. Accordingly, there seem to be good reasons to categorize *Dollywood* as a blend. However, since Dolly is an independent proper noun and *-wood* is a conventional part of geographical names (e.g., *Greenwood*, *Fleetwood*, *Glenwood*), the final decision was to exclude it from the list of blends.

Other borderline cases are considered blends, while they are not seen as prototypical instances according to the operational definition. Among these items are *workahol* (*work* + *alcohol*), *snoratorium* (*snore* + ?*moratorium*/lat. *dormitorium*/*oratorium*), *probeware* (*probe* + *software*/*hardware*), and *scareware* (*scare* + *software*). The commonly occurring lexical item *workaholic* may suggest that *workahol* could be seen as an instance of backformation. Still, *workahol* is included in the blend list because of the metonymic references to *work* in terms of addictive behavior and *alcohol* as a prototypical addictive drug (cf. example 28).

(28) Work can only become "workahol" when people continue doing it in ways that are destructive to their family life and other relationships (NOW).

As regards s*noratorium*, the exact origin of the final segment is difficult to determine. Analyzing the context of the occurrences in the subcorpus suggests a bedroom organization designed, somewhat humorously, to mitigate the negative effects of snoring on cohabiting partners. The segment *-atorium* resembles *-orium* and *-ory*, which are locative noun suffixes commonly referring to a place for a specific activity. However, this similarity was not considered sufficient to account for the entire word form. Therefore, a wildcard query based on the segment *-atorium* was run in the subcorpus to further investigate *snoratorium*. This query revealed that *moratorium* is by far the most frequent item. There are also a few expressions of moderate quantity, and a long tail of low frequency items. The 15 most frequent items retrieved in this query are presented in Table 11.

Table 11. The 15 most frequent items retrieved in a wildcard query based on *-atorium*.

Rank	*Form*	*Frequency*
1	moratorium	4936
2	crematorium	391
3	natatorium	318
4	sanatorium	246
5	exploratorium	221
6	sportatorium	12
7	dreamatorium	12
8	comatorium	10
9	conservatorium	9
10	collaboratorium	7
11	oratorium	6
12	funkatorium	6
13	aquatorium	5
14	laboratorium	5
15	screamatorium	3

The unclear origin of *-atorium* and its similarity to a suffix could allow an interpretation of *snoratorium* as a kind of derivation. Investigating its frequency patterns and context suggested, however, that *-atorium* could equally be seen as a truncated segment of (lat.) *dormitorium* possibly also influenced by the lexical item *moratorium* (cf. Waring, n.d.).

Returning to the issue of segments with seemingly unclear morphemic status, *probeware* and *scareware* are included in the blend list although their final segment *-ware* resembles the conventional lexical item *ware*. However, examining the subcorpus contexts showed that all occurrences of *probeware* and *scareware* were related to computer software, or possibly hardware in the case of *probeware*. This specialized reference motivated the interpretation of *-ware* as a truncated segment rather than the broader use of *ware* as presented in the chosen dictionaries. For this reason, *probeware* and *scareware* are considered blends.

Another example of such an item is *Popmart*, which could be seen as a combination of the lexemes *pop* and *mart*.[6] However, there are explicit references to the trademarks *Kmart* and *Walmart* in the specific contexts in which *Popmart* are found (example 29). Thus, *Popmart* is retained in the list of blends as it is seen as a combination of *pop* and *Kmart*.

(29) Choosing to launch PopMart in Kmart played right into the tour's overall themes of the excesses of consumerism and pop culture. More simply though, when The Edge was asked by a correspondent from The Daily Show why they chose Kmart over Wal-Mart, he said only, “We don't like Wal-Mart as much; they're not as cool” (NOW).

In some cases, a segment of a source word was analyzed as a conventionalized clipping, which typically disqualifies it as a truncated segment according to the operational definition (3.5). If such a segment had been combined with another lexeme, the resulting composition was not included in the blend list (cf. Beliaeva, 2014a: 31 on *enviropig*). Examples of such items in the subcorpus are *envirosell* and *propwash*. Both *enviro* and *prop* occur in the corpus as independent lexemes, and so are the final segments *sell* and *wash*. *Envirosell* and *propwash* are therefore seen as instances of compounding (cf. Fábregas & Scalise, 2012: 130).

Youthquake (*youth* + *earthquake*) is an item that has been included in the blend list, although its structure resembles *envirosell* and *propwash*. *Quake* is registered as both a noun and a verb in Merriam-Webster (n.d.-a). The verb entry is associated to shaking, instability, and trembling, while the noun entry of *quake* is closely connected to *earthquake*. The examples of *youthquake* in the corpus

[6] As shown in (29), *Popmart* also occurs in the subcorpus as *PopMart*.

are mostly nominals alluding to powerful and rapid socio-cultural changes driven by cultures connected to youth and young people (see examples 30–31). The socio-cultural changes described in (30–31) are expressed by mapping certain concrete qualities (of large magnitude instability and trembling) onto abstract concepts (of socio-cultural change), which is a typical instantiation of metaphorical mapping in the CL framework (e.g., Taylor, 2003; Lakoff, 2006).

(30) Presley was the chief popularizer, the charismatic King whose gyrating hips set off a worldwide youthquake while he led a charge with fellow trailblazers such as Chuck Berry, Little Richard, and Bill Haley (NOW).

(31) The confidence of these highly educated young Arabs [...] has spread across the region like a rapid 60s youthquake. One young protester quotes President Obama's enabling "Yes we can" when I ask him why his generation rose up (NOW).

Examples such as (30–31) give reason to believe that the second source word of *youthquake* is indeed *earthquake* used metaphorically. This interpretation is supported by Macmillan Dictionary (n.d.), which claims that "[a]lthough productive manipulation of earthquake is well-attested [...] it's an interesting example of a blend based on a metaphorical, rather than literal, sense of the word" (Macmillan Dictionary, n.d. *youthquake*). Furthermore, the prosodic similarity of *youthquake* and *earthquake*, combined with the phonetic similarity of *youth* and *earth-*, also attest to the interpretation of *youthquake* as a blend of *youth* and *earthquake* (cf. Ronneberger-Sibold, 2006; Arndt-Lappe & Plag, 2013).

5.2 Categories, domains, and series: a quantitative approach

This section addresses the distribution of grammatical categories, semantic domains and seriality in the blend data. Thus, the focus is on the word class and contexts of blends, and the observation that some blends inspire the formation of other blends. These issues have been much debated in blend research (see for instance Renner et al., 2012b; Balteiro & Bauer, 2019). The concepts of word class, semantic domains and serial distribution capture significant usage patterns that contribute to the understanding of blending as a word formation process (cf. Schmid, 2015: 4). These analytic perspectives are studied as they emerge in the systematically collected data, which is an approach that may contribute to a more refined knowledge of how blends are used, and subsequently, how blending can be characterized in more detail.

The selected topics – grammatical categories, semantic domains, and seriality – are investigated quantitatively employing distribution patterns as the analytic model. A central concern is the systematic collection of the data, which reflects an empirical turn in blend research (e.g., Arndt-Lappe & Plag, 2013; Kjellander, 2019; Renner, 2019; Renwick & Renner, 2019; Wulff & Gries, 2019).

5.2.1 Grammatical categories

The analyses of grammatical categories were carried out on two levels. The first level concerns the registered grammatical function(s) in the subcorpus of any individual blend or source word. The second level summarizes and compares individual items and identifies patterns of word class distribution.

As described in 4.2.2, most of the word class data were initially processed using automated SQL script queries. However, preliminary evaluations of the output showed that manual monitoring and evaluation of all retrieved 618 items[7] was necessary because of source word uncertainties and contextual complexity. The difficulty to determine the source words of a blend is a well-known issue and was not a surprising observation (e.g., Lalić-Krstin & Silaški, 2018: 5). Perhaps more importantly, the grammatical complexity of the data (see the discussion on *crook* in 4.2.2) in combination with the reported 3–4% of inaccuracy in the Claws 7 PoS engine suggest that the automated analyses could be improved by manual control.

The grammatical complexity of the data is visualized in Figure 12, which summarizes the blends and source words associated with more than one grammatical category. The longest bar to the left represents the total number of blends (206 items). The source word data are derived from this number since each blend typically consists of two source words. The variation resulting from the occurrence of multiple source words and blends with the same source words was disregarded in this preliminary analysis, as it was assumed to have a small quantitative impact. Figure 12 shows that 30 blends were assigned several grammatical categories (i.e., multifunctional items), whereas the corresponding numbers for the source words are slightly higher. This difference may be the result of generally higher token frequencies of source words, which potentially allows greater functional variation.

[7] 206 blends plus two SWs per blend; not considering additional items resulting from ambiguity and several SWs.

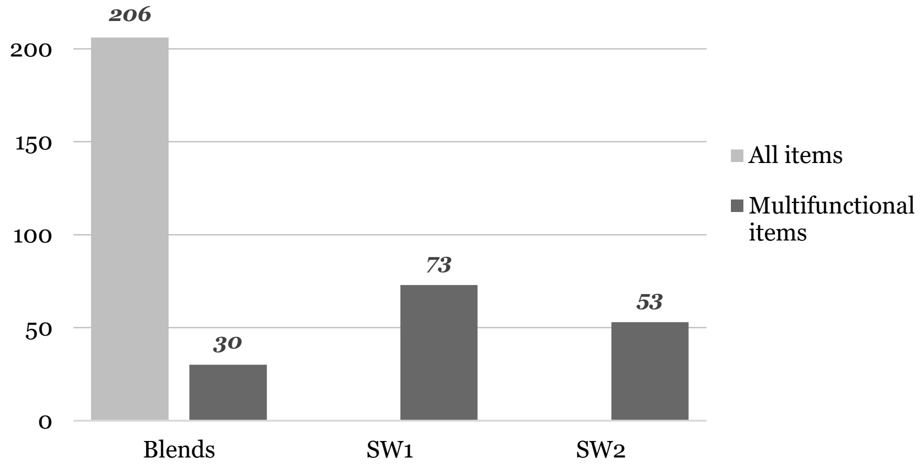


Figure 12. Multiple functionalities of blends and source words.

In comparison with the total number of identified blends (206 items), relatively few were grammatically ambiguous (15%). This degree of grammatical diversity was still considered large enough to require an analysis that goes beyond one-to-one relations between the analyzed items and their grammatical function(s).

To resolve the issue of grammatical complexity of individual items, the word class attributions were analyzed instead of the lexical items they were assigned to. For instance, the blend *Singlish* was assigned the categories of *noun* and *adjective* (N, Adj) in the blend list. Thus, *Singlish* contributes two word class markers to the distributional analysis. This way of analyzing word class distribution highlights the grammatical potential recorded in the corpus data. A quantitative effect of this operationalization is exemplified by the fact that the 66 blends of the first dataset (NN_1) were responsible for 78 word class markers in the word class distribution analysis.

Figure 13 shows all registered word class markers in the lists of blends and source words. The vertical bars represent actual numbers, which is represented in the figure with the common noun markers (N) amounting to 99, 152, and 171 for blends, source word 1, and source word 2 respectively. All grammatical categories presented in Table 5 (see 4.2.2) are represented in Figure 13 as well, including the tag PP (prepositions), which has no instantiations in the data. To underscore the observed relations between nouns and proper nouns, the exact numbers are given for the PN category as well. The perhaps most prominent characteristic of the data is the dominance of nominal functions. Figure 13 illustrates that the categories *common noun* (N) and *proper noun* (PN) are evenly distributed among the blends, while the source words have a clear preference for *common noun* (N).

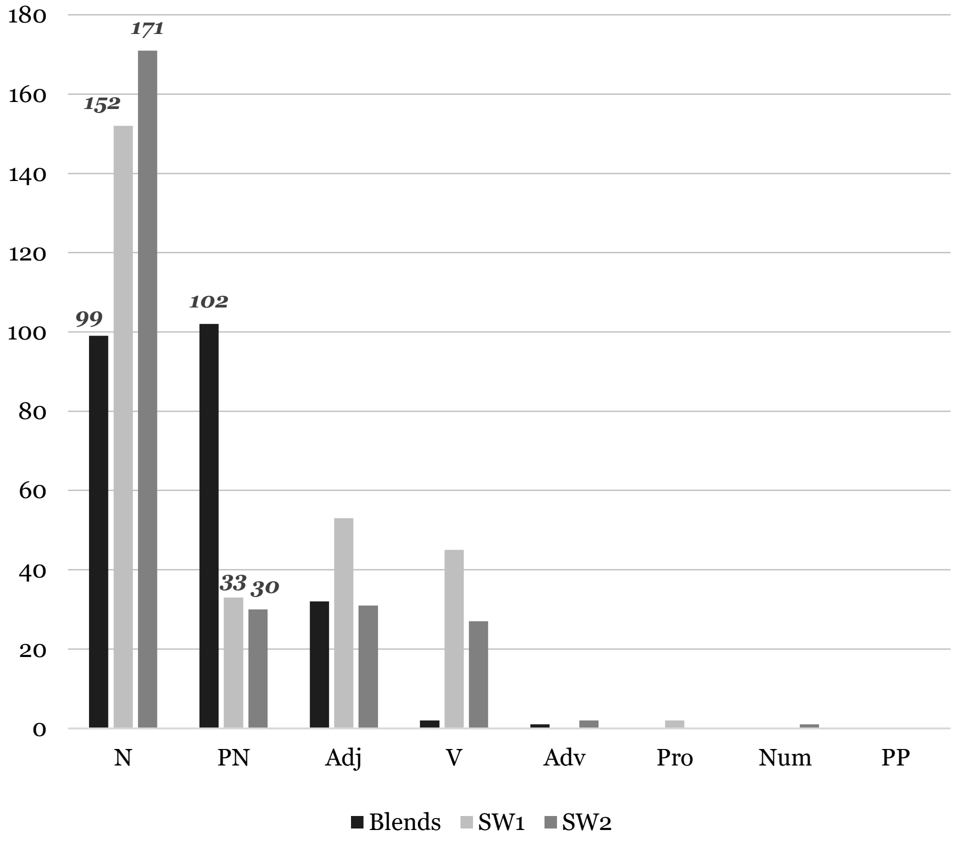


Figure 13. Word class distribution in blends and source words.

Blends such as *Websight* (*website* + *sight*) and *Elephunk* (*elephant* + *funk*) are examples of proper nouns (a company name and a record respectively) based on two common nouns (N+N → PN). The opposite (eponymic) strategy is observed in blends such as *godzillionaire* (*Godzilla* + *zillionaire*; example 32) and *masshole* (*Massachusetts* + *asshole*), in which one of the source words is a proper noun, while the resulting blend functions as a common noun (e.g., PN+N → N).

(32) I once referred to a highly paid athlete as a godzillionaire. I meant to say gazillionaire, but when I caught myself, I realized my “mistake” was better: the athlete's wealth was ginormous, as big and monstrous as the Japanese icon Godzilla (NOW).

A proper name in the sources may also result in a proper name blend, such as *Chevolution* (*Che* (*Guevara*)+ *revolution*) and *2pacalypse* (*Tupac* /tu:pæk/ + *apocalypse*), which are the names of a documentary movie and a record respectively (e.g., PN+N → PN). Blends have long been associated with nominal function (Cannon, 2000), and several studies point to the common use of blends as proper nouns (e.g., Ronneberger-Sibold, 2006; López Rúa, 2012; Danilović Jeremić & Josijević, 2019). The findings presented in the current investigation lend support to these observations.

Besides the strong nominal tendency, Figure 13 also displays an adjectival potential. Examples recorded as adjectives are *glocal*, *herstorical*, and *splatterific* (*splatter* + *terrific*). The adjectival function has received less attention in previous research, although some studies have noted this function (e.g., Cannon, 1986; Elisa, 2019). It is worth noting that several blends are marked as both nouns and adjectives, which is exemplified in *Spanglish* and *netizen*. These blends further highlight the impact of conversion in English, which is illustrated by the fact that almost 5 million tokens in the subcorpus were annotated as both nouns and adjectives.[8] *Spanglish* and the structurally similar forms based on the pattern LANGUAGE/DOMAIN + *English* are clear examples of conversion in the data, while *netizen* rather represent a grammatical flexibility observed in constructions shown in example (33).

(33) An online petition launched by the Thai Netizen Network (TNN) quickly gathered support (NOW).

The analysis of grammatical function showed that the blends in the data rarely function as verbs. Two blends were identified with verb function potential: *splurge* (*splash* + *surge*) and *dumbfounding* (*dumb* + *confounding*). These items were found in the VV and NN_2 datasets respectively. Thus, the CaWs appeared to have limited influence on the word class distribution in the data. This means that although there was verbal potential identified in the CaWs, this was not reflected in the blends.

In sum, the potential word class distribution of the blends was relatively evenly divided between the categories *common noun* and *proper noun*, with a slight additional potential for adjectival use. Adverbs, pronouns, and numerals were very rare even when blends and source words were combined, and no instances of prepositions were identified. The data suggest that verbs may occur as sources of blends (e.g., *Thinsulate* from *thin* + *insulate*, and *scattergories* from *scatter* + *categories*). However, the resulting blends do not designate events. These patterns of distribution correspond well with previous studies examining grammatical functions (e.g., Elisa, 2019; Renner, 2019).

The nature of the data collection procedure needs to be further explained in this context, since one of the selectional criteria was word class. The queries retrieving

[8] This information was retrieved in a separate subcorpus query using the wildcard condition 'WHERE pos LIKE 'nn%_jj%''.

the CaWs were restricted to retrieving nouns and verbs, which, in theory, may imply an output preference for these grammatical categories. However, the output lists resulting from the truncform queries (see step 3 in Figure 5, section 4.3) were not based on any such grammatical restriction. The script retrieving the lists of wordforms that would constitute the blend pond (script K, Appendix 1) only employed the form of the truncated segment as a selectional criterion (e.g., the truncform *routi** for the CaW *routine*).[9] This design of the script was based on the SP idea that the activation potential is solely based on effects of frequency (see Gries, 2006: 545). Accordingly, the output lists became grammatically diverse. As an illustration, example (34) lists the ten most frequent items in the output list generated by the truncform *fri** derived from the CaW *friend*. The fact that this CaW was retrieved as a noun (dataset NN_2) does not exclude adjectives (*fried* and *frightening*), adverbs (*friendly*), and proper nouns (*Friedman*) from being generated.

(34) friend, Friday, friendly, friendship, fried, Friedman, fringe, frightening, friction, Fridays

The remaining connection to word class of the truncforms is their activation potential of a specific lexical item according to the rationale of the SP analysis (see 4.3.2). Although the activation potential of *fri** favors the noun *friend*, there may be considerable grammatical variation in the query outputs.

5.2.2 Semantic domain analysis

This section begins with a delineation of how the idea of semantic domain is employed in the analysis. The theoretical delimitation and its operationalization are discussed, and the observed patterns of blends in the data and their implications for blend characteristics are also addressed.

Semantic domains are understood as "any knowledge configuration which provides the context for the conceptualization of a semantic unit" (Taylor, 2002: 196). For instance, understanding the conceptual domain (or semantic unit) DOUBLE BASSOON requires specific knowledge about WIND INSTRUMENTS and MUSICAL INSTRUMENTS in general, which represent two levels of semantic domains constituting a basic conceptual background to DOUBLE BASSOON (cf. Taylor, 2002: 195; 2003: 87-90). Figure 14 is a schematic illustration of the relation between the profiled concept of DOUBLE BASSOON and the semantic domains it relates to. It should be stressed that neither of the categories of MUSICAL INSTRUMENTS nor WIND INSTRUMENTS are as clear-cut and stable as Figure 14 might suggest. For instance, what counts as a MUSICAL INSTRUMENT may depend on perspectives such

[9] The SQL symbol used for wildcard queries is '%', which is seen in script K in Appendix 1. However, the truncform symbol '*' is used here to highlight the role of the truncform.

as cultural views, historical development, and even political ideologies. In other words, the “knowledge configuration” (Taylor, 2002: 196) exemplified in the domain MUSICAL INSTRUMENT involves a highly complex and dynamic network of ideas, which is subject to negotiation both synchronically and diachronically (cf. Croft, 2006: 274).

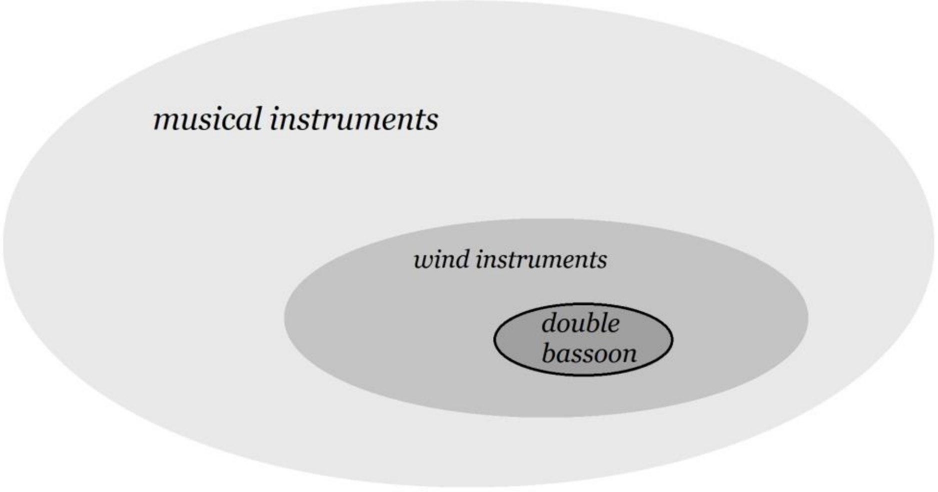


Figure 14. Suggested semantic domains of DOUBLE BASSOON.

Semantic domains were identified following the UCREL Semantic Analysis System (USAS) (University Centre for Computer Corpus Research on Language (UCREL), n.d.-d). USAS can be described as a matrix of 21 main categories referred to as “major discourse fields” (University Centre for Computer Corpus Research on Language (UCREL), n.d.-d) subdivided into 232 subcategories. The main categories are shown in Figure 15, and the entire set is provided in Appendix 4.[10] Each blend in the data was annotated with one of these main USAS categories.

As one of the available annotation resources, USAS also offers automatic semantic annotation (University Centre for Computer Corpus Research on Language (UCREL), n.d.-b). However, when the 179 unique blend forms were run in this automatic tagging software 172 instances were coded as Z99, which stands for ”[m]is-spellings or words that have not been included in the lexicon as yet” (University Centre for Computer Corpus Research on Language (UCREL), n.d.-c: 36). Therefore, the automatic tagging tool was not considered useful for the

[10] There is some variation as to the exact titles of the main categories on the USAS website. For instance, the (Q) category is labeled LANGUAGE AND COMMUNICATION in the presentation matrix on http://ucrel.lancs.ac.uk/usas/ (retrieved 2021-01-13), whereas a more detailed list downloadable (as a text file or a pdf file) from the same site uses the label LINGUISTIC ACTIONS, STATES, AND PROCESSES. The more detailed list is used throughout section 5.2.2.

current investigation, and the blends were instead analyzed and attributed a USAS tag manually.

A GENERAL AND ABSTRACT TERMS	**B** THE BODY AND THE INDIVIDUAL	**C** ARTS AND CRAFTS	**E** EMOTIONAL ACTIONS, STATES, AND PROCESSES
F FOOD AND FARMING	**G** GOVERNMENT AND THE PUBLIC DOMAIN	**H** ARCHITECTURE, BUILDINGS, HOUSES, AND THE HOME	**I** MONEY AND COMMERCE
K ENTERTAINMENT, SPORTS AND GAMES	**L** LIFE AND LIVING THINGS	**M** MOVEMENT, LOCATION, TRAVEL, AND TRANSPORT	**N** NUMBERS AND MEASUREMENT
O SUBSTANCES, MATERIALS, OBJECTS, AND EQUIPMENT	**P** EDUCATION	**Q** LINGUISTIC ACTIONS, STATES, AND PROCESSES	**S** SOCIAL ACTIONS, STATES, AND PROCESSES
T TIME	**W** THE WORLD AND OUR ENVIRONMENT	**X** PSYCHOLOGICAL ACTIONS, STATES AND PROCESSES	**Y** SCIENCE AND TECHNOLOGY
Z NAMES AND GRAMMATICAL WORDS			

Figure 15. USAS main categories.

Applying USAS categories in the operationalization of semantic domains is valuable from at least two viewpoints. First, USAS has been developed and tested for more than three decades, which has made it a reliable and balanced reference tool for semantic analyses. The intricacies of such analyses involve, for instance, decisions as to what concepts are considered related enough to form a category, what types of concepts and experiences are foregrounded in the taxonomy, and what level of detail is appropriate. The wealth of work employing USAS ensures that it has been thoroughly reviewed and that basic category delineation is expected to be consistent (see University Centre for Computer Corpus Research on Language (UCREL), n.d.-d for an overview). This does not entail that the content and structure of USAS is beyond dispute, but it provides a valuable starting point for research on semantic domains. Second, the widespread use of USAS potentially facilitates evaluation and replication, which is central not only to the current investigation. The adaptations of the USAS system to fit the data are largely matters of subcategorization, and this refinement typically does not alter the main category attributions applied in the distributional analyses. Therefore, there is reason to believe that replications and parallel investigations can be compared successfully with the observations in this study. However, despite these practical advantages of USAS, such a model should be used with some caution. From the theoretical perspective of this investigation, semantic domains are thought of as dynamic knowledge configurations that change over time and vary on both micro- and macro-levels. In addition, semantic domains typically form complex networks of meaning with a virtually limitless semantic potential (cf. Langacker, 2008: 44-47). Such complexity is unlikely to be captured by a category system built on classical categorization (cf. Taylor, 2003: 37-38). Therefore, the *discourse fields* of USAS are seen as reasonable conceptual approximations, allowing quantification of the data.

Figure 16 is an illustration of how the USAS categories were applied in practice. The example concept DEMOCRACY in the center of the figure is tagged as G1.2, which is the USAS code for the subcategory POLITICS. In its turn, the subcategory POLITICS is a part of the subcategory GOVERNMENT, POLITICS, AND ELECTIONS (G1), which pertains to the main category GOVERNMENT AND THE PUBLIC DOMAIN (G). As the main categories presented in Figure 15 constitutes the chosen level of analytic specificity, the concept DEMOCRACY would, thus, be annotated as a (G) item in this study.

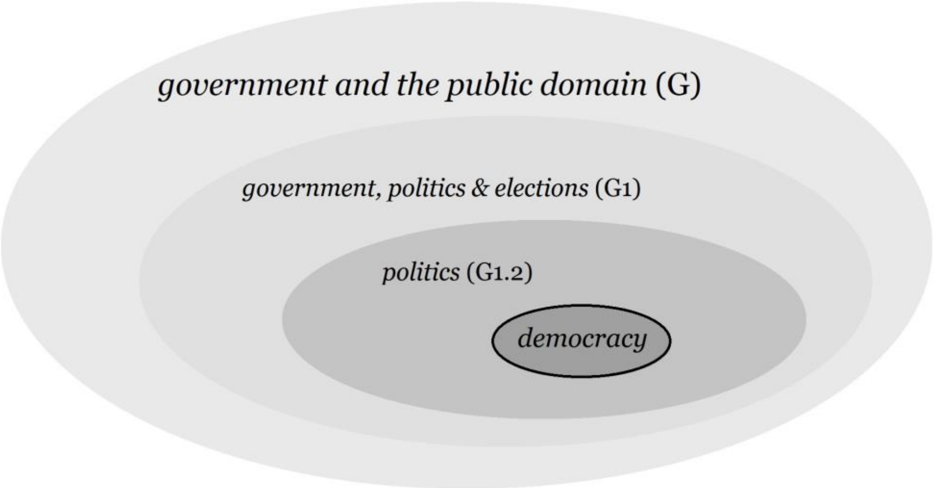


Figure 16. Semantic annotation of the concept of DEMOCRACY.

Example (35) is an extract from the NOW corpus, which highlights the importance of context in the annotation procedure. The reference to cyberspace reveals that *webiverse* ((*world wide*) *web* + *universe*) is associated to Internet technology, and thus qualifies as a SCIENCE AND TECHNOLOGY (Y) blend. The structurally similar (and serial) formations °*blogoverse* (*blog* + *universe*) and °*Twitterverse* (*Twitter* + *universe*) further underscore the association to the Internet domain.

(35) First there was the universe; then quantum theory pointed the way to multiverses. Now -verse attached to a noun means all the related members of that noun. Usually it's something in cyberspace: blogoverse, webiverse, Twitterverse (NOW).

Some minor modifications of the USAS categories were done to ground the semantic annotation of the blends as firmly as possible. For instance, the category of OTHER PROPER NAMES (Z3) was extended with a subdivision into CORPORATE BRAND NAMES (Z3.1), ARTIST ALIASES (Z3.2), EVENTS, MEDIA AND TRADEMARKED PRODUCTS (Z3.3), and NGOS; CHARITY, LOBBY, AND POLITICAL (Z3.4).[11] Such additions were made to increase the analytic precision, yet they do not surface in the distributional analysis. In some cases, existing USAS categories were extended, such as THE MEDIA: TV, RADIO & CINEMA *AND THE INTERNET (Q4.3), in which additional information follows the asterisk.

The overall distribution of the USAS categories in the blend list is shown in Figure 17. The proportions have been calculated as the number of items in each category divided by the total number of blends (206 items). For instance, 22 blends of all 206 blends in the blend list have been tagged with Y (SCIENCE AND TECHNOLOGY), which represents 10.7% of all blends. Categories representing 0.5% in Figure 17 consist of one blend. For reasons of legibility, the domains have been ordered in decreasing proportion, rather than the original alphabetical order of the codes in the USAS system (A-Z).

[11] NGO is a conventional abbreviation of *Non-governmental organization*.

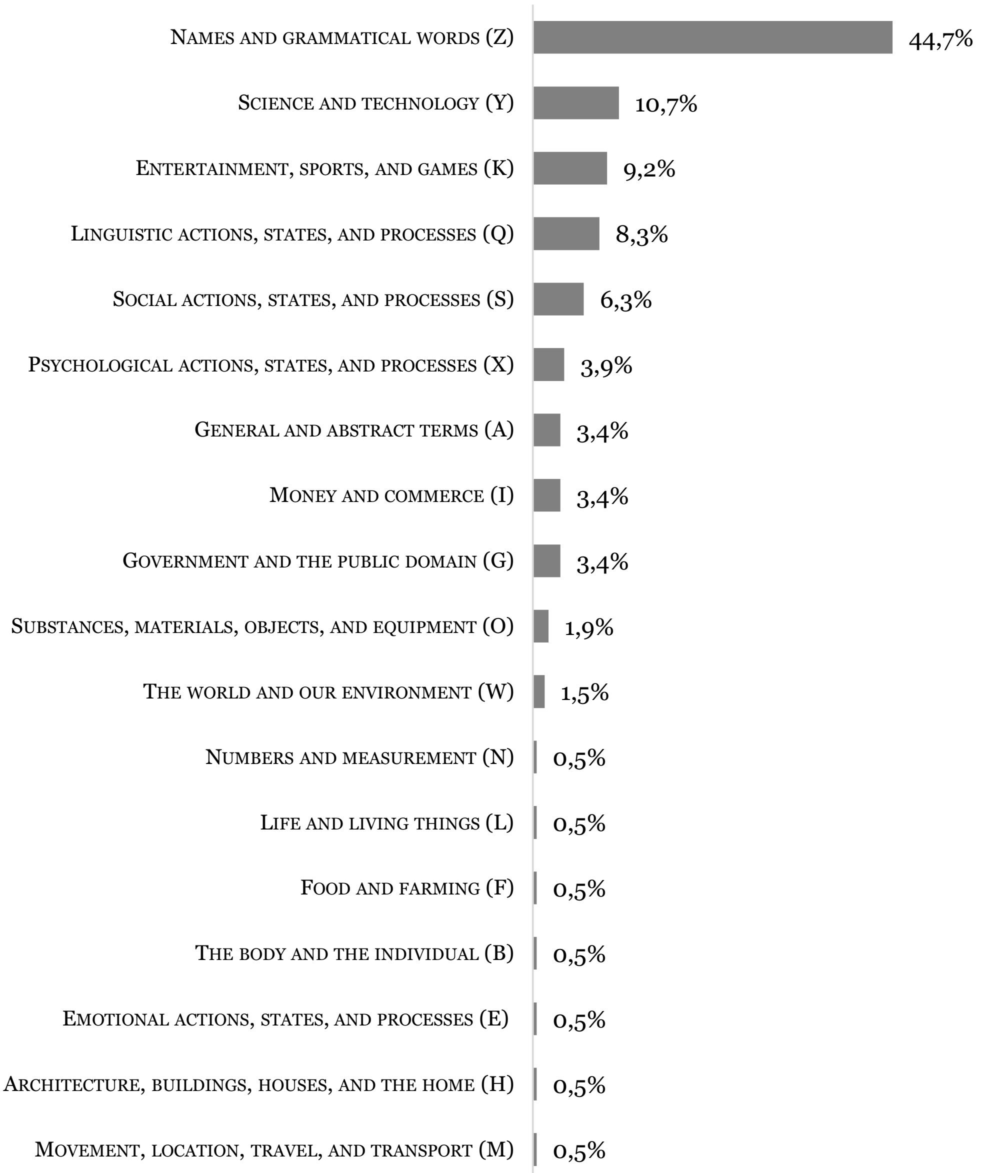


Figure 17. Proportion of USAS categories in the blend list.

The domain NAMES AND GRAMMATICAL WORDS (Z) is the most frequent category in the blend list (44.7% of all blends). Proper noun blends such as *Draganfly* (*Dragan* + *dragonfly*), *Canstruction* (*can* + *construction*), and *Netegrity* ((*Inter*)*net* + *integrity*) are examples from this group. The second most frequent domain SCIENCE AND TECHNOLOGY (Y) is still much less common (10.7%) than the (Z) domain. Examples from the (Y) domain are *webiverse* and *sneakernet* (*sneaker* + *Ethernet*/*Internet*). Blends associated with ENTERTAINMENT, SPORTS, AND GAMES (K) constitute 9.2% of the blend list, and two examples of such (K) domain blends are *disgraceland* (*disgrace* + *Graceland*) and *popera* (*pop* + *opera*).

Blends in the domains LINGUISTIC ACTIONS, STATES, AND PROCESSES (Q) and SOCIAL ACTIONS, STATES AND PROCESSES (S) are also relatively common (8.3% and 6.3% respectively). Example blends of the (Q) domain are *Singlish* and *sportcaster* (*sport* + *broadcaster*), while the (S) domain is exemplified with *frienemy* (*friend* + *enemy*) and *Chrislam* (*Christianity* + *Islam*). There is some variation between the datasets as to the (S) and (Q) domains, but 14.6% of all blends in the blend list are readily identified against one of these two backgrounds. There are thus indications in the data that a rather high proportion of the blends are referable to culture and communication.

Less prominent domains are PSYCHOLOGICAL ACTIONS, STATES, AND PROCESSES (X; e.g., *brainiac* from *brain* + *maniac*), GOVERNMENT AND THE PUBLIC DOMAIN (G; e.g., *diversicrats* from *diversity* + *bureaucrats*), GENERAL AND ABSTRACT TERMS (A; e.g., *egg-ceptional* from *egg* + *exceptional*), and MONEY AND COMMERCE (I; e.g., *godzillionaire*). The (X) domain consists of eight blends (3.9%), while the semantic tags (G), (A), and (I) are assigned to seven blends each (3.4%). Remaining domains registered in the analysis are typically represented with one or two blends. These overall patterns of semantic domain distribution of the attested blends generally correspond with similar findings in, for instance, Lehrer (2007), Fandrych (2008), Renner (2015), and Elisa (2019).

Further analyses show that all instances of the (Z) domain blends are tagged with one of the proper name subdomains PERSONAL NAMES (Z1), GEOGRAPHICAL NAMES (Z2), or OTHER PROPER NAMES (Z3). The only PERSONAL NAME tagged in the blend list is *Goracle* (*Gore* + *oracle*), while there are several GEOGRAPHICAL NAMES such as *Texico* (*Texas* + (*New*) *Mexico*) and *Calexico* (*California* + *Mexico*). The relatively generic category OTHER PROPER NAMES (Z3) is by far the most common tag among the blends (82 out of 92 instances of NAMES AND GRAMMATICAL WORDS). The influence of the (Z3) category in combination with its open character motivated, thus, a detailed analysis involving the addition of the aforementioned subdomains CORPORATE BRAND NAMES (Z3.1), ARTIST ALIASES (Z3.2), EVENTS, MEDIA AND TRADEMARKED PRODUCTS (Z3.3), and NGOS; CHARITY, LOBBY, AND POLITICAL (Z3.4). Applying a contextual analysis of the 82 blends in the (Z3) category shows that CORPORATE BRAND NAMES (Z3.1) and MEDIA AND TRADEMARKED PRODUCTS (Z3.3) are the dominant semantic domains. Figure 18 shows the distribution of all 92 NAMES AND GRAMMATICAL WORDS (Z) blends. The Z1 and Z2 domains are original USAS subcategories, whereas Z3.1 - Z3.4 are extensions of the subcategory of OTHER PROPER NAMES (Z3). The domains ARTIST ALIASES and

PERSONAL NAMES consist of one blend respectively: *Spinderella* (*artist aliases*) and *Goracle* (*personal names*).

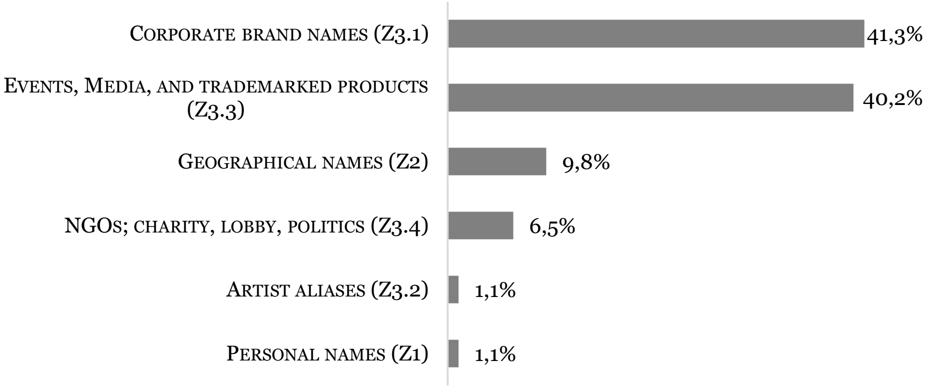


Figure 18. Distribution of subdomains of the NAMES AND GRAMMATICAL WORDS category.

Examples referred to as CORPORATE BRAND NAMES (Z3.1) blends are *Websight*, *Textainer* (*Texas*? + *container*), and *Netegrity* ((*Inter*)*net* + *integrity*). There is little in the structure of these items (besides the initial capitalization) that excludes the possibility of seeing them as common nouns. For instance, the structure of *Netegrity* resembles *netiquette* ((*Inter*)*net* + *etiquette*), but the attested occurrences in the corpus of *Netegrity* and *netiquette* reveal that they have different functions in use, which is why they have been attributed different semantic tags. Example (36) demonstrates how *Netegrity* is used in the subcorpus. As shown in this example, it seems clear that *Netegrity* refers to a specific company although the blended construction may have been intended to evoke associations to the *Internet* and *integrity*. *Netegrity* occurs only once within the temporal and regional limitations of the subcorpus, which makes it a hapax legomena in these data.

(36) an executive at Netegrity Inc. in Waltham, Mass. (NOW).

The blend *netiquette* as exemplified in (37) and (38) occurs both with and without quotation marks, which may suggest that there is variation as to its status as a conventionalized lexeme. It can be inferred from (37) that the author sees *netiquette* as a concept that has existed for some time ("is no longer"). In (38) *netiquette* occurs as an additional term to *conventions*. The quotation marks in this example indicates that it may be considered a specialized term that perhaps not all readers are familiar with. The contextual analysis suggests that it typically functions as a common noun. It has been tagged as SCIENCE AND TECHNOLOGY (Y) because of its close connection to the Internet domain.

(37) Good and proper netiquette is no longer an option for web enthusiasts (NOW).

(38) The conventions or “netiquette” that have developed for business email (NOW).

From the perspective of formal semantics, the categorical distinction between *Netegrity* and *netiquette* would likely be considered a pragmatic analysis, while the applied semantic tagging in the current analysis relies on a CL viewpoint on meaning. The CL perspective implies that no sharp distinction is made between a semantic and pragmatic analysis (cf. Langacker, 2008: 40), and since there are only indications in the data of *Netegrity* as a brand name, the meaning of the blend is interpreted accordingly (i.e., as a proper noun).

Figure 18 shows that blends tagged with EVENTS, MEDIA AND TRADEMARKED PRODUCTS (Z3.3) are practically as common as the CORPORATE BRAND NAMES blends (Z3.1). Some examples of (Z3.3) blends are *Vegemite* (*vegetarian* + *Marmite*), *Popmart* (*pop* + *Kmart*), and *spay-ghetti* (*spay* + *spaghetti*). Besides trademark product names such as *Invisalign* (*invisible* + *align*) and *Canadarm* (*Canada* + *arm*), there are several (typically also trademarked) names of artistic products (e.g., *2pacalypse*, *Elephunk*, *Youthanasia* from *youth* + *euthanasia*) and events (e.g., *Popmart*, *Canstruction*, *Spamalot* from *spam* + *Camelot*/*a lot*). The (Z3.3) and (Z3.1) categories not only constitute the main semantic domains of the NAMES AND GRAMMATICAL WORDS blends, but they are also semantically close to each other. Therefore, it could be argued that the dominance of the (Z) category expresses the influence of commercially related blends including artistic products and events. This reasoning aligns with previous studies such as Ronneberger-Sibold (2006) and Danilović Jeremić and Josijević (2019), which point to the frequent use of blends in commercial domains (see also Cannon, 1986, 2000; Lehrer, 2007; López Rúa, 2012).

An analysis of USAS subcategories was done in the SCIENCE AND TECHNOLOGY CATEGORY (Y) as well. The subcategory INFORMATION TECHNOLOGY AND COMPUTING (Y2 including subcategories) constitutes 68.2% (15 items) of the SCIENCE AND TECHNOLOGY category, with examples such as *netizen*, *webiverse*, and *faceblock* (*Facebook* + *block*). A further subcategorization into the added subcategories HARDWARE (Y2.1, represented only by *teledildonics* from *tele*(*communication*) + *dildo* + *electronics*), SOFTWARE (Y2.2, e.g., *probeware* from *probe* + *software*), and INTERNET PHENOMENA; SOCIOLOGICAL AND TECHNOLOGICAL (Y2.3, e.g., *sneakernet*) revealed some preference for the (Y2.3) subcategory (41% of the SCIENCE AND TECHNOLOGY category). However, at this level of analysis there are too few blends assigned with these tags to draw any major conclusions about their distribution. The largest category, INTERNET PHENOMENA; SOCIOLOGICAL AND

TECHNOLOGICAL (Y2.3), consists of nine blends, while (Y2.1), (Y2.2), and the more general (Y2) categories consist of six blends in total. The analysis of the SCIENCE AND TECHNOLOGY category shows that blends related to INFORMATION TECHNOLOGY AND COMPUTING are noticeable in the blend list, but they seem yet to have less influence than categories related to commercial concepts and activities. However, it is worth pointing out that CORPORATE BRAND BLENDS (Z3.1) such as *Computrace*, *Websight*, and *Netegrity* are closely associated with INFORMATION TECHNOLOGY AND COMPUTING, which stresses both the interconnections between domains and the difficulty to assign clear-cut semantic tags.

Overall, the distributional patterns of the semantic analysis suggest that the blends are often associated with commercial and technological concepts. This is represented in the analysis in that blends annotated as either NAMES AND GRAMMATICAL WORDS (Z) or SCIENCE AND TECHNOLOGY (Y) items make up 55.4% of the data. Commercial blends functioning as proper nouns (Z) are exemplified in (39) and (40). *Thinkfinity* (39) is described as a software product, which connects to both digital technologies and the commercial domain. The company name *Compuware* (*computer* + *software*; example 40) clearly connects to the domain of digital technology.

(39) They are seeking Common Core aligned resources, like those available through the Verizon Foundation's Thinkfinity platform, a free online resource that offers lesson plans and more (NOW).

(40) Another study by Compuware, a software and services provider, shows that 57 percent of users will not recommend a company with a bad mobile site (NOW).

The SCIENCE AND TECHNOLOGY (Y) blends *chrononaut* (*chrono-* + *astronaut*) and *teledildonics* are exemplified in context in (41) and (42). The use of brackets and quotation marks in (42) resembles the observation in (38), where additional contextual information and punctuation may serve to make sure the reader understands and pays attention to the specific meaning of the blend (see Cook, 2012 for an account in which such strategies are operationalized).

(41) if chrononaut Krikalev had spent his two years in space traveling at a smidge below the speed of light (almost 40,000 times faster than he was orbiting at) he would have returned to find two centuries or more had passed on Earth. # That's proper time travel (NOW).

(42) There's been a lot of buzz over the years about the growing field of “teledildonics,” an ugly-sounding portmanteau that refers to the intersection of sex toys, electronics, and computers (NOW).

The contexts of (39–42) are rather straightforward, but there are uses of blends that seem more ambiguous. *Technotopia* (*technology* + *utopia*) is a blend that appears challenging to annotate semantically because its usage patterns are ambiguous. In (43) *technotopia* functions as an abstract common noun, while it is the name of a series of published radio recordings (podcasts) in (44). Although the analyses in this study are synchronic, the temporal aspect of these examples may provide a preliminary clue to the different uses. Example (43) is dated April 21, 2014, while the time stamp of (44) is May 15, 2017. Hence, there is a possibility that the podcast referred to in (44) has made use of an already established common noun as exemplified in (43).

(43) imagine our robo-assisted future as a kind of technotopia, an immortal era of machine-assisted leisure (NOW).

(44) Technotopia is a podcast about a better future (NOW).

Having thus dealt with the blends pertaining to the most influential semantic domains, the remaining blends are dispersed across a broad spectrum of domains. Although these findings are largely in line with previous observations of the prominence of commercial and technological domains, a considerable part of the data also display variety as to contexts in which blends are found.

Finally, the influence of the NAMES AND GRAMMATICAL WORDS (Z) category in USAS is not only informative as regards the similarity with previous observations. The description of this category in the USAS system may offer an insight into why blends have been considered peripheral (cf. Kemmer, 2003; Fandrych, 2008), or even irrelevant in linguistic investigation (Ronneberger-Sibold, 2006: 159). The USAS system originally contains four (Z) subcategories with different types of PROPER NOUNS (Z0, Z1, Z2, and Z3), two categories named DISCOURSE BIN (Z4) and GRAMMATICAL BIN (Z5), three specific grammatical categories NEGATIVE (Z6), IF (Z7), PRONOUNS ETC. (Z8), and two categories called TRASH CAN (Z9) and UNMATCHED (Z99) (see Appendix 4). The depiction of the (Z) category in terms of bins and trash cans, including awarding it the last position in the matrix, signal its peripheral position in the USAS system. The fact that a significant proportion of the blends are captured by concepts referable to the (Z) category (an observation mirrored in previous studies) corresponds to the view on blends as marginal, peripheral, and with little import on the study of language. Ronneberger-Sibold (2006) captures this mechanism from a theoretical viewpoint suggesting that

> [s]ome linguists, particularly of generative inspiration, take blending merely as a matter of individual performance, comparable to slips of the tongue or poetry. In this view, blending has no status whatsoever in the linguistic competence of an average language user (Ronneberger-Sibold, 2006: 159).

In contrast to such reports, several studies argue for the considerable impact of blending in English (e.g., Kemmer, 2003; Lepic, 2016; Lalić-Krstin & Silaški, 2018). The rejection of the division between an abstract language system and language in use, which is a fundamental tenet in CL (Geeraerts, 2006c) seems, therefore, to be an important basis to broaden and deepen the knowledge of blends.

5.2.3 Seriality from a quantitative perspective

The three attributes truncation, graphemic overlap, and semantic domain distribution have received much attention in previous blend research and can be described as central issues in blending. In this section, these qualities are investigated in items that share source components in series such as *freakonomics* (*freak* + *economics*), *reaganomics* (*Reagan* + *economics*), and *clintonomics* (*Clinton* + *economics*). Almost half of the blends in the data (97 out of 206) belong to series of this type. The objective here is to find out whether these blends differ from apparently unique blend types in the subcorpus as regards truncation, graphemic overlap, and semantic domain distribution. Systematic investigation of blend seriality may potentially contribute to the development of well-motivated delineations of blends vis-à-vis affixations and compounds based on combining forms that have undergone the process of morphemization. More specifically, it may contribute to the discussion concerning whether items in such series should even be categorized as blends proper (see 3.3.4).

For reasons of efficiency, blends in the data that do not display seriality are, henceforth, called *solitary* blends. This is an operational term that should be understood from the perspective of the limitations of the study: while they are structurally unique types in the corpus data, it cannot be precluded that these blends occur in series in other contexts. Despite this caveat, the considerable size of the corpus and the systematic sampling procedure allow preliminary analyses of gradual tendencies of seriality in the material.

5.2.3.1 Truncation and seriality

Truncation of source words is attestedly a central characteristic of blends. Disagreement on this topic often concerns what types of truncation should be associated with blending. An inclusive approach is advocated in this investigation, although limitations apply, such as the double initial truncation of complex clippings (see B.2 in section 3.5). As part of the annotation of the data, all blends have been tagged according to the ABCD model (see section 3.2.2). For instance, a blend such as *Spinderella* has been tagged as an ABD blend because the first source word *spin* is retained in full (thus AB), while the second source word *Cinderella* is truncated initially (i.e., only D in CD is retained). The length of the source word is not taken into account in this annotation. The accumulated information about truncation patterns is then quantified in a distributional analysis. In this section the presentation focus on implications of seriality for truncation, but a detailed distributional analysis of truncation of the entire data is also given in 6.1.

Dividing blends tagged as serial vs. solitary enables the comparison of truncation distribution in those separate groups. Figure 19 illustrates the findings of this analysis. The bars represent the proportion in percentage of each truncation type within the serial and solitary blend groups respectively. For instance, 30% of the serial blends are coded as AD blends, whereas 21% of the solitary blends are of the AD type. These proportions have been marked explicitly in the figure to facilitate identification.

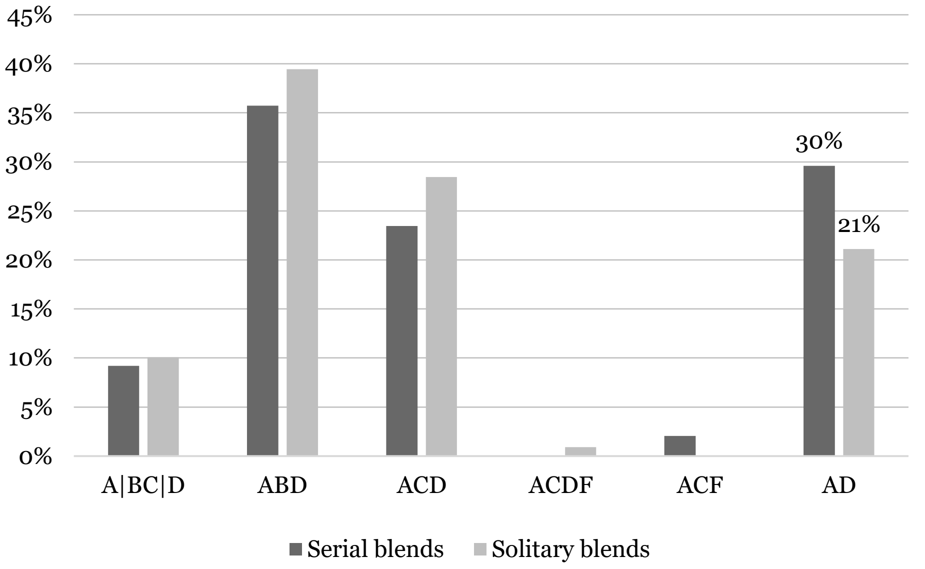


Figure 19. Distribution of truncation patterns in serial and solitary blends.

As Figure 19 shows, truncation patterns do not differ significantly between serial and solitary blends. If anything, the AD pattern, which is often emphasized as a blend-typical one, is even more prominent among the serial blends. Examples of serial AD blends are *teensploitation* and *technotopia*, while solitary examples are *stagflation* (*stagnation* + *inflation*) and *spork* (*spoon* + *fork*). The proportions of A|BC|D blends, which graphemically seem to retain the source words in full (e.g., *popera*) differ only marginally between serial and solitary blends. Hence, there seems to be little difference in terms of truncation patterns between serial and solitary blends in the corpus data.

5.2.3.2 Graphemic overlap and seriality

The second characteristic investigated in relation to seriality – graphemic overlap – is also commonly considered central to blending. In order to examine the impact of graphemic overlap in quantitative terms, each blend has been tagged with a number representing how many graphemes in the blend are shared in the source words. The measure of graphemic overlap is not restricted to linear structures, but the overlap is required to take place in the same syllable to be counted as such. For instance, the graphemic overlap of *Canstruction* is considered to be two graphemes: *c* and *n*. The grapheme *c* is the onset of the syllable /kæn/, while *n* constitutes its coda. A more strict interpretation of graphemic overlap would make only *n* the overlapping item, but this restriction would ignore the broader exploitation of similarity, which has been identified as an important characteristic of blending (e.g., Gries, 2006). In the *Canstruction* example, the similarity of *can* and *con-* displays a recurring structural pattern that licenses the interpretation of overlap in terms of entire syllables rather than in terms of single graphemes.

The distribution of graphemic overlap is summarized in Figure 20. The representation of overlap is ordered as increasing from left to right, beginning with no overlap (0 graphemes), and ending with the maximum number of overlapping graphemes registered in the data (7 graphemes). In the bars representing 4, 6, and 7 overlapping graphemes, overlap occurs in only one of the categories serial blends and solitary blends.

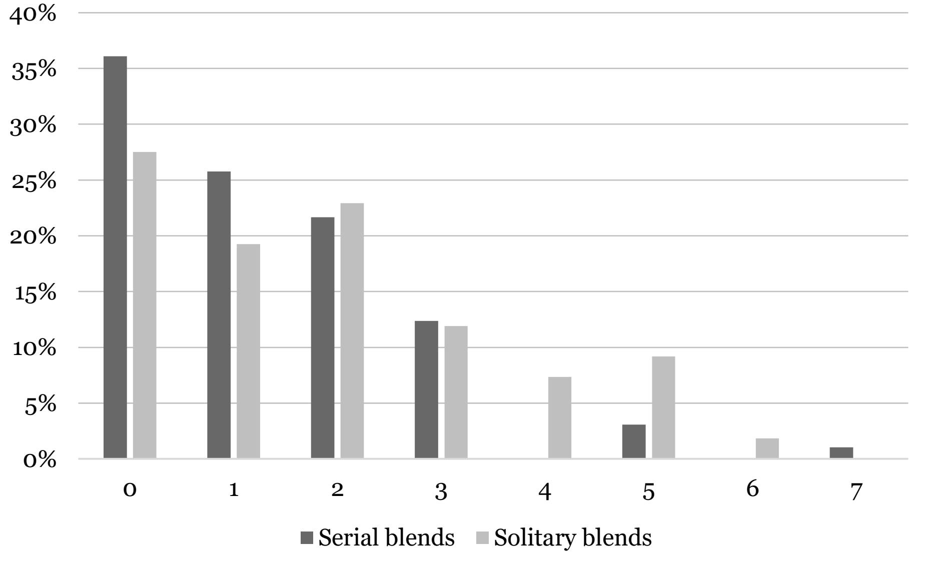


Figure 20. Distribution of graphemic overlap in serial and solitary blends.

The number of blends decrease in both of the categories serial and solitary blends as the amount of graphemic overlap increases. This observation illustrates that both serial and solitary blends generally employ modest levels of graphemic overlap. A preliminary interpretation of the comparison between serial and solitary blends in this respect suggests that serial blends differ only marginally from solitary blends as to graphemic overlap. However, calculating the means and median values shows that the mean values of graphemic overlap in serial and solitary blends are 1.27 and 1.87 respectively. The corresponding median values are 1 and 2. Thus, there are indications in the data that graphemic overlap is slightly more prominent in solitary blends than in serial ones.

5.2.3.3 Semantic domains and seriality

The third comparison contrasts the annotated semantic domains of the serial and solitary blends respectively. The analytic procedure is the same as in 5.2.2, and the distribution is represented in Figure 21. The semantic domains in Figure 21 are ordered in decreasing proportions of solitary blends. The proportions are calculated as parts of each category (i.e., serial vs. solitary blends). The serial blends follow roughly the generally observed patterns of semantic domain distribution (cf. 5.2.2). The category of NAMES AND GRAMMATICAL WORDS (Z) dominates both serial and solitary blends, and terms related to SCIENCE AND TECHNOLOGY (Y), ENTERTAINMENT, SPORTS AND GAMES (K), and SOCIAL ACTIONS, STATES AND PROCESSES (S) are rather common.

There is slightly less variation in distribution among the serial blends. This tendency is to be expected, since serial blends are often associated with the same semantic domain. The 11 blends of the -/ŋ/lish type (e.g., *Singlish* and *Spanglish*) stand out in this respect. They make up a cluster in the LINGUISTIC ACTIONS, STATES, AND PROCESSES (Q) category with significant impact on the distributional analysis. Although the other series of blends are less semantically homogenous than the -/ŋ/lish ones, solitary blends are still more dispersed. For instance, the solitary blends *splurge* (N), *schnoodle* (*schnauzer* + *poodle*) (L), *snoratorium* (H), *appletini* (*apple* + *Martini*) (F), *tragicomic* (*tragic*/*tragedy* + *comic*) (E), and *cosmeceuticals* (*cosmetics* + *pharmaceuticals*) (B) are single items that each contribute with the only data in their respective semantic domain.

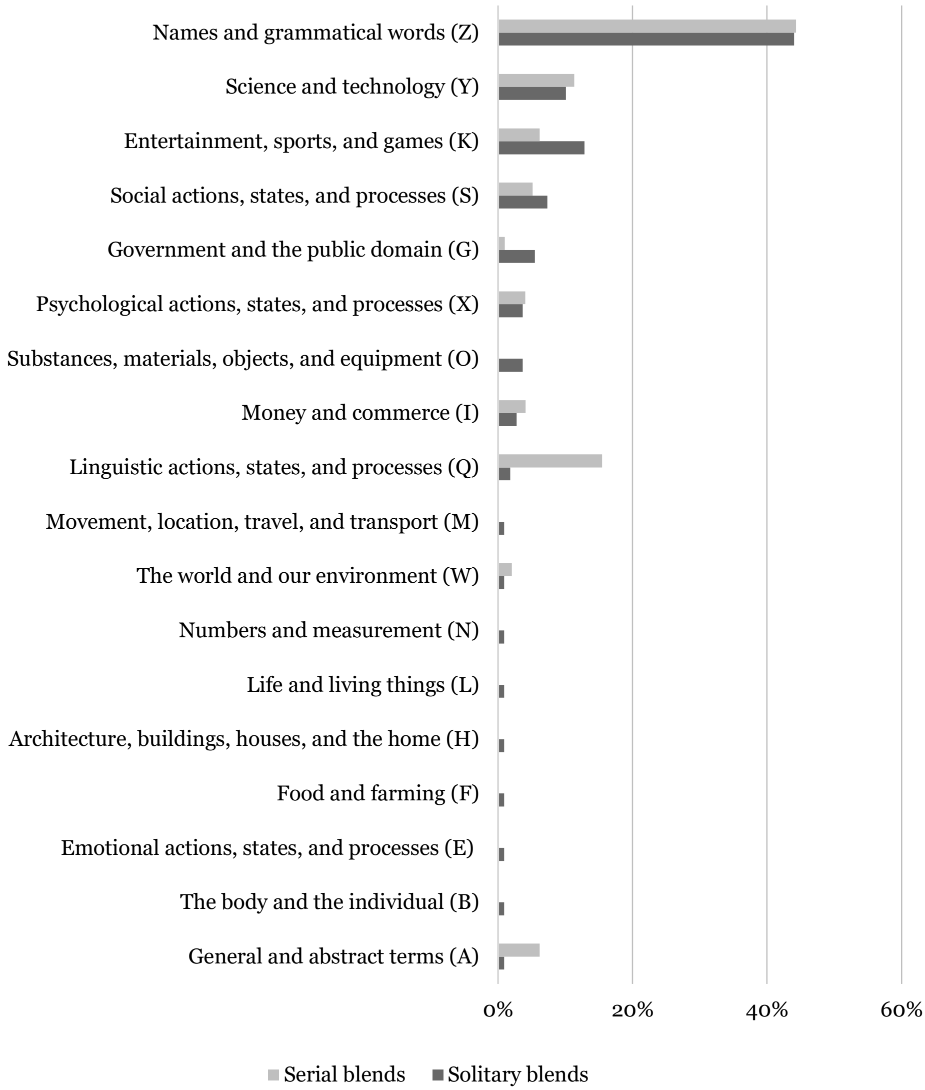


Figure 21. Distribution of semantic domains in serial and solitary blends.

The difference in distributional density should, however, not be exaggerated. In fact, more than half of the clusters of serial constructions (16 out of 27) contain blends with different semantic tags. Table 12 provides some examples from the data displaying such semantic diversity. It is also worth pointing out that the number of serial clusters (27) in itself is sufficiently high to make statistical analyses of distribution, even if all these clusters would be completely homogenous. Segments in the table with morphemic status have also been marked with a hyphen to signal their connection to the blends. The morphemic status of serial blend segments will be discussed in more detail in 5.3.5.

Table 12. Domain distribution among serial blends.

Segment	***Blends tagged with their symbol for semantic domain***
techno-	*technorati* (S), *Technovation* (Z), *technopreneurs* (I), *technotopia* (Y)
web-	*webinar* (S), *webisode* (Q), *webzine* (Q), *webtop* (Y), *Websight* (Z), *webiverse* (Y)
mass-	*Massport* (G), *masshole* (S)
-calypse	*2pacalypse* (Z), *snowpocalypse* (W; two duplicates), *Metalocalypse* (Z)
-caster	*sportcaster* (Q), *Netcaster* (Z)

The blends *Massport* (*Massachusetts* + (*air*)*port*/*port* (*authorities*)) and *masshole* exemplify how semantic domains are distinguished in this context. The WHOLE FOR PART metonymies in this pair of blends are related to different subdomains of the source word concept MASSACHUSETTS. Administrative aspects of MASSACHUSETTS are employed in *Massport*, which is a blend used in a context concerned with political and legal borders. The blend *masshole*, on the other hand, is more sociologically oriented in that the concept MASSACHUSETTS stands for regional and cultural identities related to the informal concept ASSHOLE. Consequently, *Massport* is categorized as a GOVERNMENT AND THE PUBLIC DOMAIN (G) blend, while *masshole* is connected to the domain SOCIAL ACTIONS, STATES, AND PROCESSES (S).

In sum, the quantitative analyses of seriality show that serial blends display similar patterns as solitary blends. This suggests that seriality does not in any apparent way influence the patterns of truncation, graphemic overlap, and distribution of semantic domains of the blends in the corpus data (cf. Beliaeva, 2014a: 33). These observations align with the assumption that characteristics associated with blending continue to be identifiable when blends start to develop series of constructions (e.g., Lehrer, 2007: 126; Beliaeva, 2019: 18). Hence, in the debate on seriality (see 3.3.4) this investigation has provided further support for the view that seriality is a characteristic that should be incorporated in a theory of lexical blending (cf. Lalić-Krstin & Silaški, 2018).

5.2.4 Summary of quantitative characteristics

The grammatical functions of blends and their source words have received attention in several studies (e.g., Cannon, 1986; Fontaine, 2017; Elisa, 2019; Renner, 2019). For instance, it has been assumed that blends are combinations of nominal source words (Cannon, 2000), while more recent studies suggest some variation in word class distribution (e.g., Elisa, 2019; Renner, 2019). However, the representativity of the data is a recurring issue in several accounts, and it is sometimes unclear whether observed patterns are biased by unsystematic data collection procedures (Wulff & Gries, 2019; see also 3.3.7).

The grammatical tendencies identified in this investigation are largely in accordance with tendencies found in previous studies (e.g., Elisa, 2019; Renner, 2019). The blends typically function as nouns, with a roughly equal distribution between common nouns and proper nouns. The source words are also typically nominal, and common nouns tend to be in majority. There is a modest occurrence of adjectival blends, while blends functioning as verbs are rare.

The grammatical multifunctionality of the blends is relatively low, which may be a result of their generally low token frequencies. The low numbers decrease the probabilities to register several ways to use a particular blend, and the entire body of blend data may subsequently appear less diverse. It may also be related to the syllabic structure of the blends. It has been shown that conversion in English tends to be favored for monosyllabic words (see for instance Sherman, 1975), which are uncommon in the blend data. The three occurrences of monosyllabic blends are *spork*, *splurge* and possibly *mearth*, which suggests that conversion, or grammatical multifunctionality, is restricted for this reason as well.

The distribution of semantic domains in lexical blends has also attracted interest in blend research (Lehrer, 1996; Kelly, 1998; Kemmer, 2003; López Rúa, 2012; Silaški & Đurović, 2013; Lepic, 2016; Lalić-Krstin & Silaški, 2018; Danilović Jeremić & Josijević, 2019). For instance, Kelly (1998) has suggested conceptual salience hierarchies to explain source word order, Lehrer (1996) has studied priming effects of source words, and Danilović Jeremić and Josijević (2019) have studied in which semantic domains blends tend to occur. Although previous research suggests some typical tendencies, distributional patterns of semantic domains have not yet been satisfactorily explored.

The findings here concerned with semantic domains also reflect previous observations in blend research (Cannon, 1986; Lehrer, 2007; Danilović Jeremić & Josijević, 2019 in particular). Commercial and technological domains are dominant, in which Internet related concepts are particularly profiled in the technological domains. A number of blends are associated with entertainment, sports, and artistic concepts. The prominence of these domains is contrasted with the wide variety of domains in which blends are actually found. Importantly, the categorization of the widely applied semantic analysis system USAS indicates why blends have been considered marginal in linguistic investigation, since categories that are described as peripheral in USAS are precisely those domains in which a large part of the blends are found.

The issue of seriality in blending largely concerns its relation to morphemization. The data underscore the importance of this issue, since almost half of the identified blends are in serial distribution. At some point morphological regularity and productivity may reduce typical characteristics of blends, although blending and other morphological processes seem to overlap in the process of morphemization (cf. Lehrer, 2007; Beliaeva, 2019). The analyses of blending in the subcorpus context contribute further support for this assumption.

The observations of quantitatively investigated attributes have in several ways confirmed previous findings. Therefore, this part of the study could be described as an example of "validational work" (Wulff & Gries, 2019: 19), as it is based on statistically representative data. However, it has also stressed some of the difficulties of analyzing blends quantitatively. The analyses of semantic domains have shown that semantic annotation involves challenges of delineation. The importance of contextual analyses has also been underscored. Hence, with the underlying assumption that the data have been quantitatively verified, the following section takes a qualitative approach to describe the identified blends.

5.3 Blend characteristics from a qualitative perspective

Blends are often characterized by a set of loosely connected characteristics that continue to challenge taxonomic attempts in word formation research (e.g., Bauer, 2012; Balteiro & Bauer, 2019; see also 3.2 and 3.3). The qualitative approach to the blends in the data involves an attempt to discern specific patterns that could be captured by some central common denominators as outlined below. These denominators are the four characteristics *structural profiling*, *domain proximity*, *pseudomorphemic transfer*, and *seriality* mentioned in the introduction. They are presented here alongside examples from the corpus. The discussion on figurativity and iconicity (5.3.3) follows immediately after the accounts of *structural profiling* and *domain proximity* since it connects closely to these characteristics.

5.3.1 Structural profiling

Structural profiling highlights the graphemic and phonemic properties of a blend, often manifested in wordplay (e.g., Ronneberger-Sibold, 2006; Gries, 2012; Renner, 2015). It is thought of as the close and often intricate interplay between structural characteristics (e.g., overlap) combined with networks of conceptual domains evoked by the blend and its source words (e.g., negative associations of the notion of *relationship* in the blend *relationshit*). The graphemic and phonemic properties are typically realized as similarities of source words and overlap, which influence the recognition of the source words and the decoding of the resulting blend (cf. Kelly, 1998; Gries, 2004b, 2006; Wulff & Gries, 2019). Blends identified as structurally profiled are typically formed from source words that are not conceptually connected in other respects than their coexistence in the blend.

In the blend *Crooklyn* (45) the entire first source word and 7/8 of the second source word are identifiable, and 4/8 of the graphemic structure consists of shared material (-*rook*-).

(45) *Crooklyn* (*crook* + *Brooklyn*)

The initial consonant of the blend remains phonetically a stop, although the articulation shifts from voiced bilabial to unvoiced glottal (see Kelly, 1998: 587 on degrees of phonetic similarity). Therefore, the character of the overlap in *Crooklyn* enables recognition of the source words, while there is still an element of play with transparency in the blend (cf. Ronneberger-Sibold, 2006). This suspended decoding combines with the graphemic similarities and the range of potential allusions emerging from the fusion of the concepts CROOK and BROOKLYN (cf. Ungerer, 2003: 557). For these reasons, *Crooklyn* is a good first example of a blend profiled by the structure of the source words in relation to each other as well as the resulting blend.

In *netiquette* (46) the source words are visible in their entirety (cf. section 3.2.4). Yet, the graphemic overlap seems less prominent (2/10 graphemes). A phonemic analysis (/nɛtɪkɛt/) somewhat increases the proportion of overlap (2/7 phonemes), since the total number of phonemes is lower than the total number of graphemes.

(46) *netiquette* ((*Inter*)*net* + *etiquette*)

In the first source word *net* 2/3 of the graphemes (as well as phonemes) function as overlap, while the corresponding proportion for the second source word is 2/9. In other words, there is considerable overlap in *netiquette* as well, although its distribution is more asymmetric than in *Crooklyn*.

The similarity and overlap identified in *Crooklyn* and *netiquette* are found in a considerable number of blends in the data. Examples (47–49) show how a single phoneme and grapheme may function as the only discriminating factor. In fact, segment pairs such as *-ship*/*-shit* (47), *-iction*/*-action* (48) and *-cop*/*-cup* (49) can even be described as minimal pairs.

(47) *relationshit* (*relation<u>ship</u>* + *<u>shit</u>*)

(48) *Addaction* (*add<u>iction</u>* + *<u>action</u>*)

(49) *Robocup* (*robo<u>cop</u>* + *<u>cup</u>*)

Prosodic similarity is also identified in the corpus data. In examples (50–52) the prosodic structure of the second source word is retained. For instance, the initial voiced glottal stop (/g/) in the blend *Goracle* (50) has likely limited influence (if any) of the prosody of the blend in comparison with its second source word *oracle*.

(50) *Goracle* (*Gore* + *oracle*)

(51) *Spinderella* (*spin* + *Cinderella*)

(52) *tricknology* (*trick* + *technology*)

The prosodic pattern of examples (50–52) aligns with the assumption that the second source word determines the stress of a blend (e.g., Arndt-Lappe & Plag, 2013; Wulff & Gries, 2019). However, *relationshit* (47) has the same prosody as the first source word, which parallels other blends in the data such as *faceblock*, *Elephunk*, and *Canadarm*. *Addaction* (48) and *Robocup* (49) are ambiguous as to stress inheritance: the stress of these blends coincides with both of their source words. This diversity of stress patterns may have other implications for blending, but it does not alter the status of these blends as profiled by their structure.

The special case of intercalation – which typically involves structural ambiguity – is a recurring strategy of amalgamation based on similarity (e.g., *tricknology* vs. *technology*). Hence, the occurrence of intercalative structures adds to the realization of structural ambiguity as a way to profile the structure of blends (see section 3.2.5 and Kemmer, 2003: 72). Intercalative patterns with varying degrees of structural ambiguity are identified in, for instance, *Addaction* (48), *Robocup* (49), *Spinderella* (51), and *tricknology* (52). As an illustration, *tricknology* includes a whole graphemic cluster of three coinciding consonants (*r*, *i*, and *k* in non-linear distribution), while a corresponding phonemic analysis (/tɹɪknɑlədʒi/ vs. /tɛknɑlədʒi/) reduces the potentially inserted items to /ɹ/ and /ɪ/. The close relation between *ck* and *ch* should be taken into account as well, since both these consonant combinations tend to result in a phonemic realization /k/ (Yavaş, 2006: 215). In contrast to this intricate pattern, *Spinderella* (51) is intercalative only phonemically with /p/ as the distinguishing item.

The relative visibility of the source words in blends such as *relationshit* (47), *Addaction* (48), and *Spinderella* (51) does not necessarily imply that the origin of the blend is easily recoverable (cf. Lehrer, 1996). Such blends typically constitute unconventional items that require some sort of process of decoding, which is often related to patterns of frequency of occurrence and the nature of the context in which they occur. For instance, a lemma-based comparison of *add* and *addiction* showed that *add* is more than 15 times as frequent as *addiction* in the data (379,290 vs. 24,789 hits). Therefore, it is reasonable to assume that there is a possible asymmetry of activation between those two items, which may obscure the identification of *addiction* as the first source word in *Addaction* (48). It is also worth noting that selecting *add* would allow the morphological reanalysis of *Addaction* as a compound of *add* and *action* (see further such examples in Bauer, 1983: 236). This mechanism constitutes a further potential obstacle in the successful decoding of the blend. Because of the conceptual distance of the source words, semantic priming (cf. Lehrer, 1996: 373) provides few clues as to source word recognition of such structurally profiled blends.

Another way to address the issue of source word recognition is to analyze the uniqueness of a specific segment structure (Professor Maarten Lemmens, personal communication). Such a uniqueness can be identified as the number of lexical neighbors or its phonemic conventionality (or both). It has already been shown that *add* may compete with *addiction* in terms of activation potential in *Addaction* (48). Querying the data specifically for the number of lemmas of the segments *add-* and *-action* returned 135 and 188 hits respectively, which illustrates the many possibilities of lexical activation of both these segments. Thus, there is considerable competition of type frequency. However, token frequency also influences the effects of lexical neighborhood. In Table 13, the ten most frequent items in the neighborhood of *add-* and *-action* are listed in order of decreasing token frequency. The source word *addiction* is outranked by the four items *add*, *address*, *additional*, and *addition*, while the six remaining neighbors still constitute fairly frequent competitors. The second source word *action* has greater impact in terms of token frequency, although its closest neighbors are rather frequent. The frequency analyses presented in Table 13 lend support to the assumption that the segments of *Addaction* have in fact a rather low level of uniqueness, and its constituent segments can therefore be described as unmarked. Only lowercase style is used in the table as it reflects the actual output of the corpus query.

Table 13. Lexical neighborhood of *add-* and *-action*.

***add-* neighbors**	***Frequency***	***-action* neighbors**	***Frequency***
add	376,290	action	225,001
address	169,394	reaction	54,298
additional	117,397	transaction	43,329
addition	117,221	interaction	32,722
addiction	24,789	attraction	15,761
additionally	20,455	fraction	12,915
added	17,636	satisfaction	11,117
addict	14,772	distraction	9,803
addictive	4,453	faction	8,153
additive	3,139	traction	6,399
add-on	2,717	extraction	4,356

In contrast, parallel analyses of *Spinderella* (51) show that this blend has higher levels of segment uniqueness. A query for lemmas based on *-inderella* returned only two hits, while *spin-* generated 177 types. Table 14 presents the output of the frequency queries of *spin-* and *-inderella*. *Spin-* has a number of potentially competitive lexical neighbors (e.g., *spine* and *spin-off*), but its neighbors are clearly less influential in terms of token frequency when compared with the influence of the segment neighbors of *Addaction* (48). The only two hits of the *-inderella* query display the dominance of *Cinderella*, but it is worth noting that the blend itself also surfaced in this query with five occurrences in the data. Lowercase is used to present the verbal data in Table 14 as well.

Table 14. Lexical neighborhood of *spin-* and *-inderella*.

spin- neighbors	***Frequency***	***-inderella neighbors***	***Frequency***
spin	28,929	Cinderella	1,863
spine	6,673	Spinderella	5
spin-off	5,912		
spinal	5,438		
spinach	2,277		
spinning	1,612		
spinner	1,376		
spina	547		
spinoff	525		
spindle	416		
spiny	392		

A comparison between tables 13 and 14 shows that the source words of *Addaction* and *Spinderella* display different patterns of frequency. The source words of *Addaction* have many lexical neighbors, which makes the blend potentially difficult to decode. Instead, the more marked source words *spin* and *Cinderella* make the blend *Spinderella* more transparent, especially thanks to the relative absence of neighbors to the segment *-inderella*. Similar patterns of influence on transparency from frequency effects and semantic neighborhood have been observed in, for instance, Lehrer (1996, 2007) and Kjellander (2018).

The source words of the blends in (45–51) exemplify the aforementioned lack of conceptual association. In the source word pairs *crook*/*Brooklyn*, *net*/*etiquette*, *relationship*/*shit*, *addiction*/*action*, *robocop*/*cup*, and *spin*/*Cinderella* there is virtually no inherent relation between the respective source concepts. Hence, the structural similarity of the source words contrasts with the conceptual disparity in these blends. For instance, there seems to be no intrinsic conceptual link between *shit* and the suffix *-ship* (47), but they are instead phonemically and graphemically similar. The structural intricacies thus establish the conceptual link between these otherwise unrelated source concepts. This observation is important given the traditional conceptualization of source words as generally synonymous (see Cannon, 1986 for a detailed account). Kubozono (1990) and

Brdar-Szabó and Brdar (2008) mention this idea as well, but argue that the semantic relationship between source words might vary.

The semantics of blends such as *relationshit* (47) and *Spinderella* (51) exhibit a markedly figurative potential of the blends vis-à-vis their source words (cf. Kemmer, 2003: 85). In *relationshit* the second source word *shit* functions as a metaphor for negative experiences associated with the second source word *relationship*. *Spinderella* combines selected schematic qualities from two separate domains, and the conceptual link between the source domains is established in the blend. *Spin-* is connected to the conceptual domain of DISC JOCKEY, which is described in Cambridge Dictionary (n.d.) as "someone who plays records and talks on the radio or at an event where people dance to recorded popular music, such as a nightclub" (Cambridge Dictionary, n.d. *disc jockey*). In this broad domain, *spin-* refers specifically to the spinning of a (traditionally *vinyl*) record on a record player. This narrow subdomain is used in a PART FOR WHOLE metonymy, in which the part *spin* stands for the whole conceptual domain DISC JOCKEY in the artist *Spinderella* (cf. Kemmer, 2003: 86 on the blend *glitterati*). In contrast, the segment *-inderella* refers to the Cinderella tale, which is conceptually distinct from the DISC JOCKEY domain. Abstract qualities related to *Cinderella*, such as for instance *fighting against bad odds* or *concealed excellence*, are mapped onto the disc jockey activity. Thus, *Cinderella* is used as a metaphor for certain qualities linked to the disc jockey *Spinderella*. These metonymic and metaphorical characteristics make *Spinderella* an example of a blend, in which multiple figurative strategies are employed.

Another blend in the data highlighting figurative strategies is *herstorical* (example 53).

(53) *herstorical* (*her* + *historical*)

In this construction, metonymy is employed to reinterpret one of the source words. A morphological reanalysis of *historical* is implicitly suggested in the form *herstorical* (53) through the exchange of the bigram *is* with *er* in the first syllable. The pronoun *his* emerging from this operation constitutes a PART FOR WHOLE metonymy in that *historical*, which is conventionally taken to represent *all* history, is in reality only a subdomain; i.e., the history of *men*, or *his* story in contrast to *her* story (cf. Curry, 2009; Lindsey, 2015).

A final note on structurally profiled blends concerns the distinction between syntagmatic and paradigmatic constructions. This issue is important given the recurring description of paradigmatic blends as blends proper displaying considerable wordplay, while syntagmatic blends are depicted as mere contractions of syntagmatic strings with modest structural import (e.g., Dressler, 2000; Plag, 2003; Bat-El, 2006). An analysis of collocational patterns related to

Crooklyn (45) showed that *crook* does not collocate with *Brooklyn* in any of the queries for the word classes *noun*, *verb*, or *adjective* among the 500 most frequent collocates in each word class (both left- and right-hand collocates included). These observations indicate that *Crooklyn* is not motivated as a contraction of a syntagmatic string *crook Brooklyn*. Instead, it resembles a blend construction, which Kelly (1998) refers to as “contractions of conjunctive phrases” (Kelly, 1998: 580), which is termed paradigmatic in these analyses.

A parallel analysis of *etiquette* showed that *net* occurs twice as an adjectival modifier in the compound *net etiquette*. One of these occurrences is presented in example (54).

(54) This, according to him, has also happened because we have leapfrogged from the PC to the mobile generation instantly -- where people lack inherent net etiquette and the lack of self restraint further complicates the scenario (NOW).

The related compound *internet etiquette* occurs 27 times in the analysis of the collocations in the corpus. The observed frequency further indicates the syntagmatic relation between *net* and *etiquette*. The meaning potential of *netiquette* is illustrated in example (55). Concepts such as rules, policing, and the wagging of fingers are explicitly mentioned. The second source word, *etiquette*, is associated with all these conceptions regardless of the combination with the first source word *net*.

(55) I didn't know there were rules about this. Are there? # -- Insta-confused # A: Yes, there are [...] And your petite netiquette gaffe pales in comparison to your friend's greater faux pas of calling you out for such a minor infraction (NOW).

Etiquette functions, thus, as the semantic head of the blend in the same manner as *hotel* in °*motel* (cf. Bauer, 2006). In (55), the whole blend functions as the modifier of a compound *netiquette gaffe*. This expression adds to the overall character of a normative concept with a potential to be violated. These collocational patterns suggest that *netiquette* could be described as a syntagmatic blend, while it still seems profiled by its structure.

In sum, the analyses of the subcorpus examples in this section demonstrate the potential of structural properties as a profiled characteristic in some blends. They also show how figurativity connects to the observed conceptual distance of these blends. Although structurally profiled blends typically share these attributes, there are considerable differences in distribution.

5.3.2 Domain proximity

Previous studies on blends occasionally report semantic similarity of source words as a distinct characteristic (Cannon, 1986; Kubozono, 1990; Cannon, 2000; Brdar-Szabó & Brdar, 2008). This property is seen in some of the current data as well. For instance, in *spork* (56) both of the source words *spoon* and *fork* refer to the specific domain of CUTLERY. Importantly, the conceptual similarities of the source words are not always mirrored in a similarity of form. Rather, the amalgamation in itself constitutes an iconic relation between /spɔɹk/ and the conceptual domain SPORK it refers to (cf. Ronneberger-Sibold, 2006).

(56) *spork* (*spoon* + *fork*; *cutlery*)

In the context of the current investigation, semantic similarity is conceived of as a correspondence between the conceptual domains evoked by (at least) two concepts. Such conceptual domains are mostly complex, they are thought to be construed on varying taxonomic levels, and the discrimination between domains is typically gradual (Barcelona, 2003b: 230; Langacker, 2008: 44-54). The stronger and more numerous associations of the compared domains, the more marked will the semantic similarity be. Therefore, semantic similarity is highly dynamic and variable, and the context of the blends are typically necessary to evaluate such relations. Blends formed from source words that are interpreted as similar in this respect are thought to be characterized by *domain proximity*.

A comparison between the blends *Texhoma* (57) and *stagflation* (58) will serve to exemplify how domain proximity is identified and elaborated. *Texhoma* is interpreted as a construction resulting from the combination of the domains *Texas* and *Oklahoma*, which in their turn are taxonomic sisters sharing certain characteristics in a superordinate domain of GEOGRAPHIC NAMES. In contrast, the source words of *stagflation* (*stagnation* + *inflation*) have less in common. *Inflation* is tagged semantically in USAS as I1.3 (MONEY: PRICE), while *stagnation* is potentially broader with the suggested USAS semantic tags A2.1- (AFFECT: MODIFY, CHANGE), O4.1 (GENERAL APPEARANCE AND PHYSICAL PROPERTIES), and X5.2- (INTEREST/BOREDOM/EXCITED/ENERGETIC). A superordinate domain encompassing both *stagnation* and *inflation* requires therefore a rather schematic concept such as possibly ABSTRACT CONCEPTS OF CHANGE, maybe with a preference for ECONOMIC CHANGE.

(57) *Texhoma* (*Texas* + *Oklahoma*; GEOGRAPHIC NAMES)

(58) *stagflation (stagnation* + *inflation;* ?ABSTRACT CONCEPTS OF CHANGE*)*

The differences between the source word pairs of *Texhoma* (57) and *stagflation* (58) are further illustrated by the suggestion of the USAS automatic semantic tagger, which refers *Texas* and *Oklahoma* to the domain of PERSONAL NAMES (coded as Z1, see Appendix 4). (University Centre for Computer Corpus Research on Language (UCREL), n.d.-b).[12] In contrast, *stagnation* is analyzed as AFFECT: MODIFY, CHANGE (A2.1), GENERAL APPEARANCE AND PHYSICAL PROPERTIES (O4.1) and INTEREST/BOREDOM/EXCITED/ENERGETIC (X5.2), while *inflation* is tagged as MONEY: PRICE (I1.3). It is worth noting that the blends *Texhoma* and *stagflation* are coded as OTHER PROPER NAMES (Z3) and UNMATCHED (Z99) respectively, which indicates both the difference as to conventionalization and the semantic similarity even between the source words and the blend of *Texhoma*.

Domain proximity is further illustrated by the differences in wordplay between the examples *Texhoma* (57) and *stagflation* (58). There is no overlap in *Texhoma*, and the truncation pattern is straightforwardly linear of the AD type (see 3.2.2). In *stagflation*, on the other hand, five out of 11 graphemes overlap (*-ation*) and the syllable structures of both source words are retained in the blend. The intercalative pattern of *stagflation* makes this blend structurally ambiguous, since it cannot be determined whether the cluster *-gfl-* is inserted into *stagnation*, or whether *-flation* originates from *inflation*. *Texhoma* seems mainly characterized by domain proximity when taking both semantic and structural characteristics into account, whereas *stagflation* is more oriented towards the notion of structural profiling (see 5.3.1).

The blend *spork* (56) follows the pattern of *Texhoma* (57) in many respects. The superordinate domain of the source words (CUTLERY) is readily identifiable, and the conceptual domains SPOON and FORK constitute conventional subdomains of CUTLERY. The truncation is of the AD type, and the overlap is limited to a graphemic overlap of *o*, which is not unambiguously an instance of overlap since the *o* of *spork* constitutes a part of the digraph *oo* in *spoon*.[13] The figurative quality often observed in blends characterized by structural profiling is limited in *spork*, the fused conceptual entities being essentially literal. The prosodic pattern of the blend is inherited from the second source word (*fork*).

Further examples of blends combining conceptual similarity and limited play with structure are *schnoodle*, *Calexico*, and *Chrislam* (examples 59–61). A shared superordinate domain is readily identifiable in these blends. Concerning structure, there is no overlap at all in *schnoodle* (59) and *Calexico* (60), while the graphemes *i* and *s* in *Christianity* and *Islam* are shared in *Chrislam* (61).

---

[12] Annotating both *Texas* and *Oklahoma* as *personal names* (coded as Z1, see Appendix 4) attests to the consistency of the automatic semantic tagger, but it also exemplifies its limitations, as pointed out in 5.2.2. There is a category *geographical names* (Z2) in the USAS category system, which would have been more adequate in these cases.

[13] Cf. the parallel issue in *Spanglish*. The *n* in Spanish could be seen as a graphemic overlap with the *n* in *English*, but the *ng* digraph in the latter functions in combination with *g* to represent /ŋ/. Therefore, it could be argued that there is not even a graphemic overlap in such cases. However, as the terminological choice (*graphemic* overlap) indicates, this orthographic phenomen does not lead to a non-overlap analysis here.

(59) *schnoodle* (*schnauzer* + *poodle*; DOG BREED)

(60) *Calexico* (*California* + *Mexico*; GEOGRAPHIC NAMES)

(61) *Chrislam* (*Christianity* + *Islam*; RELIGION)

The characteristics of *infomercial* (62) are closer to the patterns of *stagflation* (58). It seems somewhat less straightforward to identify a conceptual domain incorporating both source words of *infomercial* in a comparison with blends such as *spork* (56). A suggested concept of TYPE OF BROADCASTED MEDIA roughly captures the specific meanings of *information* and *commercial* in the resulting blend (cf. Merriam-Webster, n.d.-a).

(62) *infomercial* (*information* + *commercial*; TYPE OF BROADCASTED MEDIA)

The semantic similarity of the source words of *infomercial* partly depends on the specifying function of the blend. An understanding of the blend as a type of *commercial* implies that the source word *information* is narrowed down to a WHOLE FOR PART metonymy, in which *information* signals the use of *infomercial* in the context of various forms of advertising. As to its form, *infomercial* slightly differs from the *spork* type of blends (57, 59–61). Both source words of *infomercial* are truncated and combined according to the AD pattern, and the overlap is graphemically modest. However, the graphemes *o* and *m* are possible (while non-linear) overlapping elements. The phoneme /ʃ/ in the final part of *infomercial* could also have some import in this respect as it functions as the onset of the final syllable in both source words, i.e., /-ʃən/ in *information* and /-ʃəl/ in *commercial* (cf. Gries, 2004b: 649). Thus, the modest semantic similarity of the source words, in combination with at least some structural similarity, make *infomercial* less characterized by domain proximity than more prototypical items such as *spork*, *Texhoma*, *schnoodle*, *Calexico*, and *Chrislam*.

Blends characterized by domain proximity are often limited as regards structural characteristics such as overlap, similarity, and structural ambiguity. However, there are blends in the data, whose source words are conceptually close, while they also display structural wordplay. In *popera* (63) and *glocal* (64) the source words are recognizable in full, while only the grapheme *d* is missing in *frienemy* (65).

(63) *popera* (*pop* + *opera*; MUSIC GENRE)

(64) *glocal* (*global* + *local*; LOCATION)

(65) *frienemy* (*friend* + *enemy*; TYPE OF RELATION)

(66) *Spanglish* (*Spanish* + *English*; LANGUAGE)

Spanglish (66) displays a similarity between the sounds of the source words in the medial position of the blend, i.e., between /ŋ/ and /n/ (cf. Kelly, 1998: 587). In addition, *Spanglish* has a potential intercalative structure with *n* replaced by the consonant cluster *ngl* in the source word *Spanish*. Examples (63–66) underscore the diversity within the group of blends characterized by domain proximity: the semantic similarity and the modest play with structure vary as well as the combination of these two properties.

The conceptual fusion of source domains in blends such as *spork* (56), *Texhoma* (57), and *frienemy* (65) is prominent to the extent that it can be claimed to be mirrored iconically in the blended structure (cf. Kemmer, 2003; Ronneberger-Sibold, 2006; Veale & Butnariu, 2010). The categorical level of the source words and the blend is often the same in these blends, which produces relatively concrete blended concepts remaining in the category level of the sources (e.g., *spoon-fork-spork* or *Texas-Oklahoma-Texhoma*). Consequently, their figurative potential may be less profiled than in blends characterized by structural profiling. Domain proximity blends are instead either a level mixture of two concepts such as in *popera* (63) or some form of borderline entity as is the case in *spork* (cf. Bauer, 2012). In *schnoodle* (59) there is even a genetic mixture of two dog types, which typically becomes visible in the physical appearance (see Renner, 2020). Parallel instances occurring in other contexts are, for instance, ˚*tiglon* (*tiger* + *lion*) and ˚*zonkey* (*zebra* + *donkey*) (Merriam-Webster, n.d.-a). Cannon (2000: 955) highlights the conceptual competition of such forms in the blend pair ˚*geep* and ˚*shoat*, which refer to the crossbreeding of *sheep* and *goat*.[14]

5.3.3 Figurativity and iconicity of structural profiling and domain proximity

The analyses in 5.3.1 and 5.3.2 highlight how the two characteristics structural profiling and domain proximity are prominent in some blends, while other attributes may also be connected to these two main ones. Figurativity in structurally profiled blends such as *Spinderella* (51) and *herstorical* (53) is one such notable attribute. In blends such as *spork* (56) and *schnoodle* (59), which are characterized by domain proximity, an iconic relation is instead observed between the amalgamation of the blended form and its intended objects. The relation between figurativity and iconicity in connection to the main characteristics (structural profiling and domain proximity) is further analyzed here. This tendency is illustrated schematically in Figure 22.

[14] In Algeo's (1977) terms, these blends are instances of *counterblends*. ˚*Liger* (*lion* + *tiger*) (Merriam-Webster, n.d.-a) further exemplifies this pattern, with the addition that this counterblend is employed to signal the sex of the crossbred animal. The first source words in *tiglon* and *liger* refer to the male, and the second source word, subsequently, to the female.

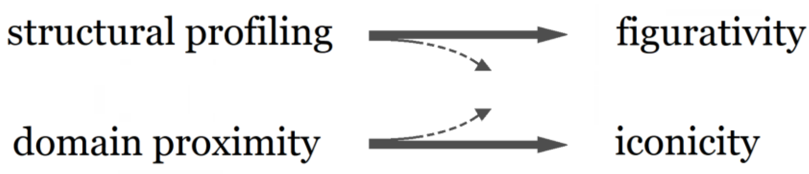


Figure 22. Distribution of figurativity and iconicity.

The pair of blends *schnoodle* (59) and *Addaction* (48) provide a typical example of how the distribution of figurativity and iconicity is linked to structural profiling and domain proximity. The domain proximity of a blend such as *schnoodle* can be contrasted with a structurally profiled blend such as *Addaction*, whose source words *addiction* and *action* seem less related. Figure 23 shows the differing types of fusion realized in *schnoodle* and *Addaction*. The structural as well as conceptual amalgamation of *schnoodle* is expressed as the sharing of domains (gray space in Figure 23). Both the source words and the blend can be referred to a superordinate conceptual domain of DOG BREED, which underscores the close relation between the source words and the blend. The result of mixing these dog breeds is mirrored in the blend, which points to iconicity as a marked semantic component. In contrast, *Addaction* has a prominent structural characteristic with properties such as similarity of source words, overlap, and intercalation (see 5.3.1).

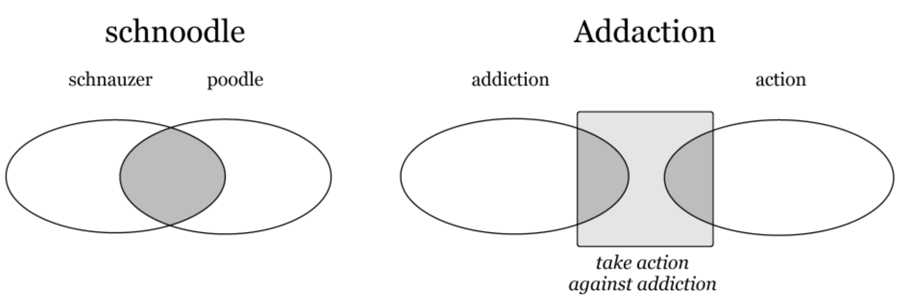


Figure 23. *Schnoodle* vs. *Addaction*.

The fusion of the conceptual domains ADDICTION and ACTION draws on a metonymic extension of the domain of ADDICTION relating to problems of addictive behavior typically requiring specific types of ACTION.

The structural fusion of *Addaction* is not equally mirrored conceptually, since it seems awkward (to say the least) to claim that the conceptual domain of ADDACTION is the necessary result of combining the concepts of ADDICTION and ACTION. Instead, *Addaction* functions as the conceptual link between the two otherwise comparatively unrelated source concepts (cf. Ungerer, 2003: 564). The *schnoodle-Addaction* pair is to some extent a prototypical example of the distribution of figurativity and iconicity as attributes connected to structural profiling and domain proximity respectively. However, these distributions are not always as straightforward.

Comparisons of the blends *frienemy* (65) and *brainiac* (67) will exemplify the gradual and complex nature of the distributions of figurativity and iconicity. In *frienemy* there is both overlap of the source words and similarity between the prosody of the blend and its source word *enemy* (cf. *popera* (63), *glocal* (64), and *Spanglish* (66) in 5.3.2). The conceptual domains of the source words *friend* and *enemy* are analyzed as taxonomic sisters with an immediate superordinate concept of TYPES OF RELATION. At the same time, the relatively broad concept of friend (expressed as RELATIONSHIP: GENERAL (S3.1) in USAS) is potentially restrained in *frienemy*, since *enemy* appears slightly more specialized (APPROACHABILITY AND FRIENDLINESS(-) (S1.2.1-) in USAS). Subsequently, it could be argued that certain aspects of *friend* are profiled in *frienemy*, and that *friend* is thus potentially metonymic in the blend. In contrast, the dominant characteristic of *brainiac* is thought to be structural profiling because of the similarity of the source words and the structural overlap in the blend.

(67) *brainiac* (*brain* + *maniac*)

Concerning the conceptual domains it refers to, there is a figurative relation between the source words *brain* and *maniac*. The physiological object *brain* stands for an abstract ability (intelligence) as realized in *brainiac*. Thus, *brain* as a source word of *brainiac* is thought to function as a metonymy of the BODILY OVER MENTAL type (cf. Kövecses & Radden, 1998: 65). Furthermore, *maniac* is also connected to *brain* in that insanity – whether intended as a medical concept or as an instantiation of hyperbole – is closely related to brain activity (cf. Merriam-Webster, n.d.-a). Therefore, *maniac* is metonymically related to both the literal *brain* (physiological object) and the metonymic extension of *brain* in *brainiac* (BODILY OVER MENTAL metonymy). The intricate meaning potential of *brainiac* as outlined here is reflected in the output of its source words *brain* and *maniac* in the USAS automatic semantic tagger. The USAS output is presented in Table 15. The minus sign (-) indicates that the opposite of SENSIBLE is intended, which suggests, for instance, qualities such as IRRATIONAL or UNREASONABLE. The plus symbol (+) underscores a generally positive characterization of the concept. The abbreviation *mf* means that no distinction is made between MALE and FEMALE. The superordinate categories of S1.2.6 connected to *maniac* are SOCIAL ACTIONS, STATES & PROCESSES (S1) and PERSONALITY TRAITS (S1.2).

Table 15. Output of *brain* and *maniac* in the USAS automatic semantic tagger.

	Coded output	***Explanation***
brain	B1 X2 X9.1+ X9.1+/S2mf	ANATOMY AND PHYSIOLOGY MENTAL ACTIONS AND PROCESSES ABILITY: ABILITY, INTELLIGENCE ABILITY: ABILITY, INTELLIGENCE/PEOPLE
maniac	S1.2.6-/S2mf B2-/S2mf	SENSIBLE/PEOPLE HEALTH AND DISEASE/PEOPLE

The comparison between *frienemy* (65) and *brainiac* (67) is illustrated in Figure 24. The analyses of these blends show that *frienemy* is characterized by domain proximity of its source words, while there may also be an element of structural intricacy in the blend. Yet, the conceptual relation between its source words is mainly about fusing two concepts that are taxonomically on the same level, which creates an iconic relation between *frienemy* and the conceptual domain it activates.

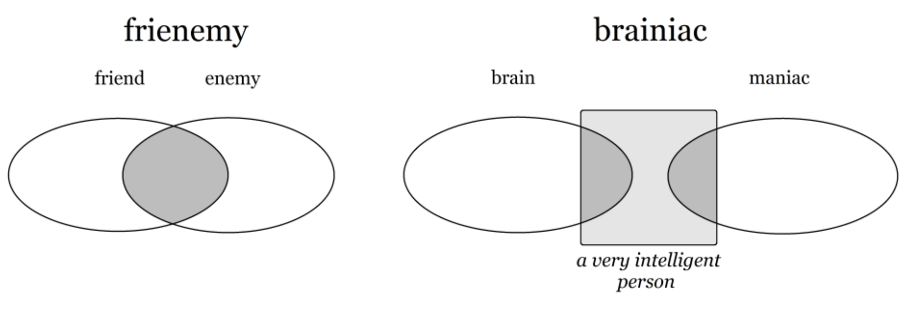


Figure 24. Frienemy vs. *brainiac.*

Brainiac resembles *frienemy* structurally in the near preservation of both source words, but its figurative potential appears more prominent. The conceptual domain BRAINIAC depends on metonymy in both source words. Although the visual representation is the same as in Figure 23 (*schnoodle* vs. *Addaction*), the comparison of *frienemy* and *brainiac* reveals that delineation may be less straightforward in some cases.

The *frienemy-brainiac* example should not be seen only in terms of the gradual distribution of domain proximity and structural profiling. These two characteristics are not thought of as a binary pair in the sense that they are mutually exclusive. They focus on two different sets of mechanisms that are typically seen in blending. The analyses in 5.3.1 and 5.3.2 show that *spork* (56) contrasts with *Spinderella* (51) in ways that are not random, and, importantly, are dependent on the interrelation between a number of seemingly disparate factors such as overlap, truncation, linearity, prosody, and figurativity.

5.3.4 Pseudomorphemic transfer

Wordplay and conceptual integration have often been referred to as central mechanisms in blending (e.g., Gries, 2006, 2012; Beliaeva, 2014a; Renner, 2015). The current section addresses blending strategies characterized less by wittiness and cleverness (cf. Algeo, 1977: 61), but strategies that resemble other morphological processes (Lehrer, 1996: 380). These strategies are most apparent in the morpheme-like qualities of the segments and the often markedly syntagmatic relation between the source words (e.g., *Vegemite*, *appletini* and *diversicrats*). The choice of term, *pseudomorphemic transfer*, is motivated by the intention to capture the resulting closeness to other concatenative word formation processes (such as for instance compounding and derivation), while there are still properties that separate this blend characteristic from more regular morphological processes (e.g., Lehrer, 2007; Beliaeva, 2019). The terminological choice *transfer* is intended to capture the idea of schematic influence on blending from entrenched morphological processes. Ronneberger-Sibold's (2006: 164) notion of *pseudosuffixation* is somewhat connected to pseudomorphemic transfer as it captures associations between existing suffixes and truncated segments. However, pseudomorphemic transfer is intended as a broader term not restricted to requiring the existence of a suffix that is identical to a specific truncated segment, or even suffixation as the only schematically transferred morphological process. The analytical category of *low integration* in López Rúa (2012) also connects to pseudomorphemic transfer in that overlap and wordplay are typically limited in low integration blends (see 3.4.4).

Two examples of blends in the data characterized by pseudomorphemic transfer are *Vegemite* (68) and *Dreamsicle* (69). The structural amalgamation occurs at syllable boundaries of the source words with virtually no overlap, which creates a first impression that the initial segment of the second source word is simply exchanged with another premodifier.

(68) *Vegemite* (*vegetarian* + *Marmite*)

(69) *Dreamsicle* (*dream* + *popsicle*)

Kemmer's (2003) term *substitution blend* is applicable to *Dreamsicle* in this sense, as it substitutes a "part of one lexeme with another whole lexeme" (Kemmer, 2003: 74). *Vegemite* does not fit equally well into this description, but the integrity of the segment *vege-* (e.g., Kjellander, 2018: 167) in combination with the abundance of forms based on both segments *vege-* and *-mite* (the latter segment referring to *Marmite*) indicate that there is an element of replacement at work in *Vegemite*. From the perspective of morphological analyzability, both *Dreamsicle* and *Vegemite* are instantiations of seemingly productive patterns, or schemas (cf. Booij, 2010). Thus, they signal the functional closeness to compounding and derivation.

The general pattern of morphemic substitution relies to some extent on the possibility to identify a morphemic segment (e.g., *dream-*) in the blended construct (e.g., *Dreamsicle*). This is possible in blends such as *Dreamsicle* (69), *appletini* (70), and *teensploitation* (71). Such blends even allow limited possibilities to identify alternative breakpoints (i.e., *dream-, apple-, teen-*).

(70) *appletini* (*apple* + *martini*)

(71) *teensploitation* (*teen* + *exploitation*)

Therefore, the source word truncation of these blends is morphologically motivated, and it could be argued that Algeo's (1977: 51) category *clipping at morpheme boundaries* should apply. However, similar instances such as *Vegemite* and *diversicrats* do not fit equally well into this delineation, and the fact that morpheme boundaries are identifiable in blends with extensive overlap further challenges Algeo's (1977) delineation. The latter observation is exemplified with blends such as *Alternet* (*alternative* + (*inter*)*net*), *televangelist* (*tele*(*phone*) + *evangelist*), and *californication*, which are characterized by structural profiling. These forms allow the identification of a morpheme depending on where a breakpoint is assigned (for instance, *-net*, *-evangelist*, *-fornication*). In some cases, alternative breakpoints allow the identification of more than one morpheme. Full overlap blends such as *sexploitation* are perhaps the most striking examples of this in that both source words are potentially recognizable in full (see 3.2.2 and 6.1).

The contrast between blends characterized by pseudomorphemic transfer (e.g., *Dreamsicle* and *appletini*), and blends with a stronger inclination towards structural profiling (e.g., *Alternet* and *Californication*) seems partly to be the degree with which a morphological schema is profiled in the blend. The morphological schema of replacement is an essential characteristic of *Dreamsicle*, while it is an optional way of interpretation in *Californication*. Importantly, the mechanism of substitution and the use of morpheme boundaries occur gradually, with prototypical as well as borderline cases.

The semantics of blends characterized by pseudomorphemic transfer is further highlighted by the observation that syntagmaticity is a recurring pattern in the data (cf. Bauer, 2006: 502; see also section 3.3.1). In examples (72–76), there is a potential to interpret the first source word as a modifier in a compound, or in a syntactic string. For instance, *permalloy* (72) could be paraphrased in the syntactic string *permeable alloy*, in which *permeable* functions as an adjective modifying the noun *alloy*. *Snowpocalypse* (73) follows a similar pattern in that it may be interpreted as a contraction of the metaphorical expression *snow apocalypse*. The order of the source words is thus restricted by headedness patterns usually giving conceptual prominence to the right-hand component (e.g., Bauer, 2006). Further instances of this pattern are *vactor* (74), *rolladium* (75), and *Teavana* (76), the latter being presumably a *Nirvana* (a place of euphoria) for tea enthusiasts.

(72) *permalloy* (*permeable* + *alloy*)

(73) *snowpocalypse* (*snow* + *apocalypse*)

(74) *vactor* (*vacuum* + *excavator*)

(75) *rolladium* (*rollerblade* + *stadium*)

(76) *Teavana* (*tea* + *Nirvana*)

The syntagmaticity of examples 72–76 is contrasted with blends characterized by structural profiling and domain proximity, the source words of which are often in a paradigmatic relation (e.g., *addiction*/*action* and *global*/*local*) (cf. Kelly, 1998; Bat-El, 2006).

The notion of pseudomorphemic transfer should be seen as an attempt to encompass several strategies that, taken together, form a blend characteristic identified as schematic similarity with conventional and productive morphological processes (cf. Kjellander, 2018 on the notion of *schema transfer effect*). The various types of functions influencing the classification of blends are exemplified with the discrimination between blends and complex clippings not only on mere structural grounds, but also on the basis of source word similarity (Gries, 2012), and collocational patterns of source words (Beliaeva, 2014a).

In the case of pseudomorphemic transfer, structural properties such as truncation patterns and the nature of the segments combine with semantic attributes in a characteristic that seems relatively reduced as to eye-catching blend properties. Investigating such a characteristic is important given the observation in Wulff and Gries (2019) that previous blend studies may have exaggerated certain properties of blend formation (see also Cannon, 1986). This tendency in blend research may be a consequence of analyzing datasets “accrued under less-than-ideal sampling conditions” (Wulff & Gries, 2019: 18), which stresses the need for systematic data collection in efforts to develop a better understanding of the motivating mechanisms and characteristics of blending.

5.3.5 Seriality from a qualitative perspective

This section revisits the recurring observation that blends often give rise to series of items based on a shared segment (e.g., Algeo, 1977; Lehrer, 1996, 1998; Kemmer, 2003; Lalić-Krstin & Silaški, 2018). The blends *Spanglish*, *Chinglish*, *Singlish*, *Hinglish*, and *Yinglish* (see 5.2.3) exemplify one such group of blends following a pattern LANGUAGE/NATIONALITY + *English*. This mechanism has several implications for the analysis of blending as a word formation process. First, seriality constitutes a challenge in the delineation of blending in relation to other word formation processes as it is often thought to blur the border between productive morphology and linguistic creativity (Bauer, 2006: 483). Second, it stresses the dynamic interplay between diachronic and synchronic perspectives on blending, which has been discussed in terms of the ephemeral nature of blends (e.g., Cannon, 1986; Bauer, 2006) and the morphemic development of blend segments (e.g., Warren, 1990; Lehrer, 1998, 2007). Third, it raises the question what mechanisms make certain blends produce such groups of blends while others do not. Lalić-Krstin and Silaški (2018: 7) suggest "immediate social context" as largely responsible for the patterns in their data. Because this study is mainly oriented towards synchronic analyses, the term *seriality* is chosen to emphasize the type of information that is retrieved from the empirical material.

Seriality is thought of as groups of blends built on the same source word (see 5.2.3). The analyses have resulted in 27 such groups containing at least two different blends with one common source (see Table 16). The number of identified blends vary in these groups, as well as the apparent relation between the serial segments. There are prototypical groups such as the aforementioned *English* blends (e.g., *Spanglish*, *Singlish*, *Yinglish*; see also 3.3.4). Other blends, for instance *Manulife* (*manufacturer* + *life* (*insurance*)) and *Metabolife* (*metabolism* + *life*) based on the segment *life*, raise questions about how semantically close the respective blends are. The serial interconnection between *Manulife* and *Metabolife* may even be uncertain, as *life* in *Manulife* refers to *life insurance*, whereas *life* in *Metabolife* seems rather a way to inspire general associations with health (cf. Ronneberger-Sibold, 2006: 165 on the use of *med* for Germ. *Medizine*).

The identified segments serving as source words for several blends are represented in Table 16 in the column *segment*. The hyphen of a segment indicates whether it occurs initially or finally in the blends. The segment *net-/-net* stands out in this list as it potentially functions both in initial and final position. The segments in Table 16 are ordered alphabetically, and the corresponding groups are named after the source word assumed to be responsible for the segment. The blends in the rightmost column constitute the respective identified serial items, with the exception that duplicates are represented only once. Proper nouns are capitalized.

Table 16. Serial blend formations.

Segment	***Source word group***	***Blends***
acti-	active/activity	Activision, Activia
-action	action	Constraction, Collaboraction, Addaction
-calypse	apocalypse	2pacalypse, snowpocalypse, Metalocalypse
-caster	broadcaster	netcaster, sportcaster
compu-	computer	Computrace, Computrainers, Compuware
eco-	ecological	eco-logical, ecologo
egg-	egg	eggstravaganza, eggs-actly, egg-cellent, egg-ceptional
execu-	executive	Executrain, Execunet
-/ŋ/lish	English	Spanglish, Chinglish, Singlish, Hinglish, Yinglish
identi-	identity	identikit, identigene
-life	life	Manulife, Metabolife
mass-	Massachusetts	Massport, masshole
net-/-net	(Inter)net	Navinet, netiquette, netizen, netcaster, Netegrity, Alternet, Execunet, sneakernet, netware,
-nomics	economics	freakonomics, reaganomics, clintonomics
-opera	opera	popera, hip-hopera
pop-	pop	Popmart, popera, poppermost
-(s)ploitation	exploitation	sexploitation, teensploitation, jewxploitation
-rati	literati	digerati, technorati
-/k/rete	concrete	Quikrete, Calcrete
sci-	science/scientific	Scitation, scientifiction
techno-	technology	technorati, technovation, technopreneurs, technotopia
tex-	Texas	Textainer[15], Texarkana, Texican, Texhoma, Texico
th(in)-	thin	thinspiration, Thinsulate
-ware	software/ (*possibly* hardware)	adware, mindware, Compuware, probeware, spyware, brainware, netware, scareware
web-	(world wide) web	webinar, webisode, webzine, webtop, Websight, webiverse
veri-	verify/verification /*lat.* Veritas	Verizon, Verisign, Verifone, Verichip, Verilux, Verisign
youth-	youth	youthquake, Youthcare, Youthanasia

[15] Here, the first source word is assumed to be *Texas*, but this interpretation is preliminary as the corpus data give no decisive information about the exact origin.

The data in Table 16 give rise to (at least) two important observations. First, the blends in serial distribution also display blend-typical characteristics such as structural attributes, figurativity, and iconicity. Thus, several of them could also be characterized in terms of structural profiling, domain proximity and pseudomorphemic transfer. Second, morphemic differences are registered between the serial segments. Serial blends are identified with morphemic bases (e.g., *-action* and *-opera*), combining forms (e.g., *techno-* and *eco-*), and non-morphemic fragments (e.g., *execu-* and *-nomics*). Following the operational definition in 3.5, morphemic segments require a truncated (non-morphemic) fragment in order to qualify as a blend (the result being otherwise typically a compound or a derivation). Therefore, serial blends with a morphemic segment are included as long as the other segment is non-morphemic. The blend °*Brexit* and the series of similar constructions exemplify this possibility (Lalić-Krstin & Silaški, 2018, see also 3.2.1). Kemmer (2003) suggests that the combination of segments with full lexemes may inspire the formation of groups of lexical items of similar types, and this seems to be the case in some of the blends in Table 16. However, the diversity, also including combining forms and non-morphemic fragments, indicates that there are various strategies at work in serially distributed blends.

The blends *youthquake* (77) and *egg-ceptional* (78) exemplify how prominent structural attributes are sometimes identified in serially distributed blends.

(77) *youthquake* (*youth* + *earthquake*)

(78) *egg-ceptional* (*egg* + *exceptional*)

The initial segment of *youthquake* involves overlap (*-th-* and /θ/ in both *earth* and *youth*) and prosodic correspondence between *youth* and *earth* as well as *youthquake* and *earthquake*. In *egg-ceptional* there is a near-correspondence between the consonants in the first syllable (/k/ vs. /g/), which are distinguished largely by their voicing. Moreover, *egg-ceptional* is graphemically intercalative, and the proximity in terms of phonetic realization of /ɪ/ and /ɛ/ suggests a near-phonemic intercalation as well.

Y*outhquake* (77) and *egg-ceptional* (78) also attest to the occurrence of figurativity in serial blends. In *youthquake* certain characteristics of *youth* are profiled, typically those related to sociological and psychological qualities. Thus, a WHOLE FOR PART metonymy (cf. Barcelona, 2003b: 238-240) is produced, where *youth* stands for the more restricted conceptual domains of YOUTH profiled in *youthquake*. In addition, *earthquake* functions metaphorically as (parts of) a concrete domain of experience mapped onto *youth* in the blend *youthquake* (e.g., a capacity for fast and powerful change). Subsequently, the figurative potential of *youthquake* involves both metonymic and metaphorical meaning structures. In *egg-ceptional* the broad and general expression *exceptional* is similarly metonymic. It profiles and hence metonymically represents exceptionally good qualities related to *eggs* such as their tastiness or esthetic qualities (cf. Barcelona, 2003b: 242-243). This account shows that *youthquake* and *egg-ceptional* are characterized by structural profiling along the same line as blends such as *relationshit* (47) and *Spinderella* (51) (see 5.3.1). Thus, the blend-typical structural intricacies of *youthquake* (77) and *egg-ceptional* (78) overlap with their serial realizations.

Turning to the issue of the morphemic status of serial segments, *-calypse* and *-rati* in Table 16 are considered truncated non-morphemic fragments, while *youth-* and *egg-* are conventional English morphemes. The segments *-calypse* and *-rati* resemble serial blend segments of the *-thon* and *-holic* type, which have been identified in studies concerned with morphemization (cf. Soudek, 1971; Warren, 1990; Lehrer, 1998; Kemmer, 2003; Lehrer, 2007; Lalic-Krstin, 2014; Jurado, 2019). The process of morphemization typically involves a semantic shift from a specific conceptual domain of the original source word (e.g., *Marathon*) to a schematic meaning associated with the segment (e.g., *-thon*, see 3.3.4). The result is sometimes a fossilization of the segment, which then typically becomes a combining form or a derivational affix (e.g., Lehrer, 1996: 361). The truncated segments *-calypse* and *-rati* are potential candidates to undergo morphemization given their conformity to morphonotactic[16] conventions in English, but this affordance is not applicable to *youth-* and *egg-* since they are already conventional morphemes (or lexemes).

The identified morphemic difference between *-calypse* and *-rati* vs. *youth-* and *egg-* suggests that there may be a potential for a functional difference between morphemic and non-morphemic blend segments in serial distribution. Only serial segments that are not already morphemes – and are morphophonotactically plausible – are likely subject for morphemization. The issue of why certain segments become serial is not answered by this observation, but it contributes an additional parameter to the analysis of seriality once it is established. The example of *-calypse*/*-rati* vs. *youth-*/*egg-* suggests that the morphemic status of the serial segments plays a role in the process whereby some serial segments continue to function in blending, while others undergo the gradual process of morphemization, sometimes resulting in a new conventionalized morpheme.

[16] The haplologic construction *morphonotactics* (*morpho-* + *phonotactics*) captures what phonotactic constraints could be expected to influence the structure of a morpheme.

5.3.6 Summary of qualitative characteristics

The four broadly sketched blend characteristics discussed in 5.3 constitute a complex network of mechanisms profiled to various degrees in the blend data. Some mechanisms stand out as highly profiled in a specific characteristic. One example is the iconic relation between conceptual and structural fusion as a part of the characteristic domain proximity. Other mechanisms are more diffusely distributed, such as figurativity, which seems to play a more central role in structural profiling but are observed in blends characterized by domain proximity and seriality as well. The complex distribution of mechanisms highlights the commonly observed difficulties to classify individual blends in hard-and-fast categories of the classical container type (cf. Kemmer, 2003; Taylor, 2003; Bauer, 2012). Instead, the blend characteristics discerned in the data intend to show how individual blends can be associated with one or several of these characteristics. These associations are typically a matter of degree. For instance, a blend such as *schnoodle* (59) appears more characterized by proximity of source word domains than *infomercial* (62) (see 5.3.2). Moreover, a blend can often be simultaneously associated with several characteristics. For instance, *popera* (63) is the blended result of two related source word domains *pop* and *opera* (i.e., music genres) which makes it characterized by domain proximity. At the same time, there is overlap in the source words, which is also fully visible in the blend.

As the term suggests, structural profiling is oriented towards structurally related mechanisms such as similarity, overlap, and intercalation. The uniqueness of segments varies, but it is an influential factor in the play with transparency, which has been shown to have implications for blending (e.g., Ronneberger-Sibold, 2006; Gries, 2012). What is notable about blends characterized by structural profiling is that source words are comparatively unrelated semantically, which runs counter to recurrent assumptions that source words should be semantically related (e.g., Cannon, 2000). Moreover, the conceptual link is often established figuratively in parallel with the profiled structure, which adds to the general potential of wordplay and wittiness.

Domain proximity captures the blend pattern in which source words are closely connected semantically. This pattern has been a recurring topic both historically (Bergström, 1906: 5) and in contemporary blend research (Kubozono, 1990; Cannon, 2000). The data of this investigation show that such relatedness is identifiable in certain blends, and that other mechanisms appear to be closely connected to this characteristic. A specific observation of the semantic structure of domain proximity is that both the source words and the resulting blend tend to be taxonomic sisters (e.g., *spork* (56)). The fused concepts in the data are typically concrete entities with limited figurative potential. Instead, the conceptual and structural fusion is typically iconic in the blends identified with domain proximity. Such iconicity has been reported in several studies (e.g., Kemmer, 2003; Ronneberger-Sibold, 2006; Brdar-Szabó & Brdar, 2008; Fandrych, 2008; Veale & Butnariu, 2010). This study claims that iconicity is distributed systematically as a mechanism with connections to domain proximity.

The perhaps most difficult characteristic to capture in the data is the tendency for certain blends to, at least seemingly, mimic conventionalized morphological patterns such as derivation and compounding. The term pseudomorphemic transfer is partly inspired by the notion of pseudosuffixation (Ronneberger-Sibold, 2006), and it is an extended version of *schema transfer effects* (Kjellander, 2018). The characteristic pseudomorphemic transfer is intended as a broad concept encompassing the influence of morphological schemas on the formation of blends. Analyzing the data from this perspective suggests that syllabic and morphological breakpoint patterns influence blend formation. Furthermore, the tendency for syntagmaticity – and thus typically right-headedness – highlights the close connection to compounding and perhaps to complex clipping. The conceptualization of pseudomorphemic transfer also involves observations of limited overlap and relatively fixed breakpoints. As a result, blends markedly profiled with this characteristic typically have a limited degree of qualities such as source word similarity, overlap, iconicity, and figurativity.

The last characteristic, seriality, differs from the other in that individual blends cannot be identified as serial. Instead, it is a characteristic that entirely depends on a broader context where other blends occur. Hence, it is a kind of encyclopedic knowledge about existing series of items that license the formation of similar constructions. Moreover, seriality is closely connected to the process of morphemization, which is largely a diachronic process. Because diachronic concerns are given limited emphasis in this study, the notion of seriality is limited to observations of forms that occur in series in the data. There are obvious problems with this approach, such as the possibility that unique blends in the data may have serially related blends identifiable in other contexts. Nonetheless, the data contain potentially valuable information on serial tendencies since the sampling of the subcorpus has been shown to be statistically valid. Hence, the distributional tendencies of seriality are expected to mirror tendencies of the subcorpus.

There are at least two important findings in this investigation on the issue of seriality. The first one is the lack of correspondence to other characteristics such as graphemic overlap and the distribution of semantic domains in the serial blends (5.2.3). This observation sheds further light on the question whether blends in serial distribution should be classified as blends proper. The findings of the current analyses suggest that seriality does not restrict other blend-typical characteristics. Thus, their serial connections to other items is no reason why they should not be considered blends. The second observation concerns the morphemic variation between serial segments. Here, it is suggested that there may be a systematic pattern behind what series of blends gradually loose some of the qualities motivating their status as blends. If a non-morphemic segment in serial distribution has a syllable structure paired with a potentially salient meaning that makes it a plausible candidate for morpheme status, then morphemization might occur. Subsequently, morphemization may gradually transform the serially distributed blends into compounds or derivations.

5.4 Blend characteristics in a systematically collected dataset

The nature of the blend pond constitutes a firm platform for the systematic investigation of lexical blends in the subcorpus. The comparison of the output from each of the three datasets has shown that the data collection methodology is robust. Observed patterns in the blend pond are therefore assumed to be representative of the subcorpus.

The quantitatively oriented observations on grammatical categories and semantic domains (5.2.1 and 5.2.2) are largely in line with previous findings in other contexts, while the quantitative analyses of seriality display patterns supporting the view that blending continues to be functional as a word formation process in instances of blends occurring in series. However, there are two notable observations in the qualitative and quantitative analyses of seriality. First, the morphemic diversity observed in the employed segments (5.3.4) suggests a possibility to explain morphonotactically why some serial blends may become subject to morphemization, in which they gradually lose their blending characteristics. Second, the decreasing graphemic overlap in serial blends (5.2.3.2), may indicate the influence from an ongoing process of morphemization, since the segment in question is increasingly conventionalized, and thus likely allows a decreasing amount of structural modification.

The four identified characteristics described in 5.3 are seen as useful conceptual tools in the detailed qualitative and quantitative analyses of the structure and semantics of the blends. They have been motivated by the need for comprehensive analyses of blends taking several cognitive factors into account, as emphasized in 3.6. The analyses show that the four characteristics are gradually distributed, occurring to a high degree in prototypical instantiations, while diluted in borderline cases. There is also considerable overlap in the data, with some blends associated with several characteristics. The four qualitatively analyzed characteristics are not intended as categories in which blends can be organized typologically but are thought of as patterns influencing the shape the blends take. This perspective resembles Kemmer's (2003) idea of "properties that tend to cluster in instances of blending" (Kemmer, 2003: 75). Hence, a classical container metaphor categorization is not considered sufficient to account for the observations here. Organizing the data from the perspective of prototype theory offers a better solution (see also Renner, 2015).

Figurativity and iconicity stand out as important types of semantic structures in the blend data. They seem to be in nearly complementary distribution, in which figurativity occurs mainly in blends characterized by structural profiling, while iconicity is typically associated with domain proximity. This seems logical since domain proximity blends have a limited choice of available source words for blending, which constrains the affordances of structural strategies. For instance, an object combining the qualities of a *spoon* and a *fork* (usually *spork*) will likely have to be based on the conventional lexical inventory of the domain CUTLERY. In contrast, blends characterized by structural profiling are instead constrained mostly by structural similarity, which offers considerable choice in the extensive English vocabulary. However, when the lexical structure guides the formation of blends (as is the case in blends characterized by structural profiling), it appears less likely that the source words are conventionally associated with the same semantic domain. Figurative strategies therefore seem almost inevitable to construe the fused meaning of such a blend (see Taylor, 2002: 500). For instance, to make sense of *Elephunk* (*elephant* + *funk*), a literal conceptual integration verge on the bizarre: a conceptual amalgamation of the animal *elephant* and the music genre *funk*. It seems more reasonable that certain attributes of the concept ELEPHANT are mapped metaphorically onto the target domain FUNK. This reasoning may be connected to Cannon's (2000) claim that "the etyma that are to be blended cannot be randomly selected, but should have some semantic and/or phonological similarities, ranging to partial rhyme" (Cannon, 2000: 952). The observations of figurativity and iconicity in this investigation underscore the experiential basis of meaning in blending. Approaching blends from a CL perspective is, thus, potentially effective, since its basic ideas on the centrality of experiential meaning offer theoretical tools to explore figurativity and iconicity.

In general, the data show signs of systematicity in the mechanisms associated with the attested blends. It is suggested that structural profiling and domain proximity represent two characteristics that make use of largely different strategies of formation. The other main characteristics, pseudomorphemic transfer and seriality, may not appear equally related to each other. Still, the potential morphemic implications of seriality involve the elaboration of morphemic qualities, which is central in pseudomorphemic transfer as well, albeit on a schematic level. The common denominator remains on an abstract level, since seriality occurs as an inter-blend phenomenon, while pseudomorphemic transfer is an intra-blend characteristic, although there is an indirect influence from other (morphemic) instantiations. For instance, the serially related blends *sexploitation*, *teensploitation*, and *jewxploitation* suggests a potential for morphemization of the segment -(*s*)*ploitation* based on the relation between these instantiations. In contrast, the blend *Dreamsicle*, analyzed in terms of pseudomorphemic transfer, displays an internal structure that schematically resembles compounding. At the same time, it has been observed that such combinations of whole lexemes often occur in series (Kemmer, 2003: 74).

In sum, chapter 5 shows that the applied systematic model to collect blends confirms several findings in previous blend research, while new information also emerges. A key component of the analyses is the representativity of the data, which gives empirical support to findings that coincide with previous research as well as those that contrast with common assumptions about blends. In particular, the qualitatively oriented observations in 5.3 display both complexity and systematicity in a fashion that merits further exploration. Despite the broad scope of the four main characteristics in 5.3, there is overlap between them in many respects, which makes individual blends intrinsically ambiguous. This quality may be thought of as a problem in a systematic investigation, and the analyses in chapter 5 give many examples in that direction. However, given the prominence of ambiguity in blending, this study takes the stance that ambiguity is in fact a central mechanism in lexical blending. The following chapter will focus precisely on this concept in relation to blends.

6 Ambiguities

This chapter is concerned with the patterns of ambiguity identified in the blends in the corpus data. It also addresses the implications of ambiguity on blend formation. The presentation is divided into thematically organized sections. Each of the sections 6.1–6.4 introduces a specific type of ambiguity and describes the attribute it is associated with. Four types have been identified: *truncation ambiguity*, *mode ambiguity*, *source word ambiguity*, and *covert source ambiguity*. They have emerged as particularly striking manifestations of ambiguity of the blends in the data, as displayed by truncation (cf. 3.2.2), graphemic dependence (cf. 3.2.6), intercalation (cf. 3.2.5), and implicit source words (cf. 3.2.1). Section 6.5 summarizes the findings and suggests a model for interpreting ambiguity in the blend data. In this final section, the epistemic question concerning the value of ambiguity in research in general (cf. Bauman, 1991) and in linguistic investigation in particular is also discussed.

6.1 Truncation ambiguity

Truncation – or the deletion of structure – stands out as one of the central attributes of lexical blending (see 3.2.2). Constructions such as *schnoodle* display a pattern, in which the truncation of the source words *schn|~~auzer~~* and *~~p~~|oodle* can be readily determined. However, two thirds of the blends in the data of this investigation do not have such evident breakpoints. Instead, they (139 items) display varying degrees of ambiguity in connection to truncation. For instance, the possible truncations of the source word *Brooklyn* in *Crooklyn* ranges from five graphemes (*~~Brook~~|lyn*) to only one (*~~B~~|rooklyn*), depending on how the blend is perceived. The expression *truncation ambiguity* is used for this phenomenon, which essentially concerns how source words are decoded in a blend. The structure of blends that display this type of ambiguity is argued to be based on truncation strategies involving meaning potential realized on a structural level.

The sharing of structure in a blend (e.g., -*rook*- in *Crooklyn*), which parallels truncation ambiguity, is usually referred to as overlap (see 3.2.3–3.2.5). In blends such as *relationshit* (*relationship* + *shit*) and *technovation* (*techno*(*logy*) + *innovation*) the overlapping structures -*shi*- and -*no*- afford different manifestations of truncation ambiguity. The position and the nature of the overlap are also potentially meaningful. In *relationshit*, the graphemic overlap constitutes 3/12 of the blend, and it occurs in the final syllable. This type of construction matches Kelly's (1998) description of certain blends in the spoken mode, in which "the speaker postpones, however momentarily, the listeners' recognition that they have been sidetracked" (Kelly, 1998: 587). *Relationshit* also likely requires a reanalysis of the segment -*shi*- as to where the source words have been truncated and to which source word is at work. *Technovation* has a slightly lower degree of graphemic overlap (2/12). However, the morphemic quality of *techno-* (a combining form according to Merriam-Webster, n.d.-a) has probably a more central role as it may reduce the perceived truncation ambiguity according to the breakpoint pattern *techno|vation* (*techno*(~~*logy*~~) + ~~*inno*~~*vation)*. Despite the interplay between truncation and overlap in blending, truncation is dealt with separately in order to highlight its ambiguity as exploited in the data. The main reason for this separation is that truncation and overlap profile different aspects of amalgamation in blending, i.e., the complex deletion of structure in the source words vs. the sharing of structure in the blend.

The distribution and characteristics of the truncation types provide important information about the nature of truncation in the data. The A|BC|D pattern (also referred to as *full overlap*, see 3.2.4) is perhaps the most intriguing case, but it is also important as it constitutes the endpoint on a continuum from perfect analyzability to complete obscurity of breakpoints. It has already been pointed out that the breakpoints are clearly traceable in *schnoodle*. In some blends the identification of truncated fragments is reasonably straightforward, although there may be some uncertainty of breakpoints (e.g., *calcium* and *concrete* in *cal|c?|rete*). Blends in which differences between spoken and written representations influence the analysis of breakpoint may increase the complexity of truncation ambiguity. For instance, the graphemic truncation of *spoon* in *spork* could either be *sp-* or *spo-* (*sp|o?|*~~*on*~~ + ~~*f*~~*|o?|rk*). However, the phonemic representation /spɔɹk/ suggests that the psychologically realistic truncation of *spoon* is *sp-*, while *-ork* is referred to *fork*. In other blends a fixed breakpoint cannot be determined (e.g., *Crooklyn* from *crook* + *Brooklyn* and *scientifiction* from *scientific* + *fiction*). Interestingly, while all other patterns are acknowledged as types of truncation, even in cases where nearly all of the truncated structure overlaps (e.g., *californication*), the extreme endpoint (i.e., the A|BC|D pattern) is sometimes claimed not to be characterized by truncation at all (see 3.2.4).

The importance of the A|BC|D pattern for the analysis of truncation ambiguity is illustrated by problematizing further the notion of breakpoint. Establishing a breakpoint may seem formally necessary to identify the truncation of a source word. For instance, the breakpoint in the blend *webzine* is determined on the grounds that there are no apparent structural connections between the initial segment *web-* and the two initial syllables *maga-* in *magazine*. As a consequence, the four final graphemes of ~~*maga*~~*zine* are retained, which makes *-zine* the final truncated segment in the blend.

In contrast, the overlapping structure of *scientifiction* allows no such formal identification of breakpoint, which may lead to the formally based conclusion that there is none (i.e., no truncation occurs). Contrary to this idea, it is here suggested that there are *multiple* breakpoints, and that these breakpoints contribute to a wordplay effect realized as ambiguity established in the complex realization of truncation (cf. Kelly, 1998).

Applying ambiguity as an explanatory model of truncation has implications not only for A|BC|D blends but extends to truncation patterns in blending generally. Blends such as *glocal* (*global* + *local*) and *collaboraction* (*collaboration* + *action*) are only minimally distinguished from the A|BC|D pattern, which could be rendered as *glo*|(*c*)|*al* and *collabora*|(*c*)|*tion*. Thus, truncation is almost as complex in these items as in the A|BC|D blends, with the difference that truncation cannot be disregarded on formal grounds in *glocal* and *collaboraction*. The truncation is equally complex in an ABD blend such as *Spinderella*. The graphemic overlap of *-in-*, suggesting at least the three options of truncation *-inderella*, *-nderella*, and *-derella*, is further complicated by the fact that the realization of the initial phoneme /s/ coincides in the blend and the second source word *Cinderella*. This further illustrates how the resulting truncation ambiguity can be realized on several levels. The phonemic structure of *Spinderella* also affords an intercalative interpretation, since /p/ could be considered an inserted phoneme in the first syllable of /sɪndəɹɛlə/ (i.e., /spɪndəɹɛlə/).

The intricacies of truncation ambiguity as sketched in this section have important implications for the distributional analyses of the standard truncation patterns in this investigation (cf. 3.2.2 and 5.2.3.) The annotation of the blend data is based on a holistic assessment including contextual properties and morphemic qualities of the truncated segments. These perspectives are crucial to arrive at psychologically plausible analyses of truncation patterns. Despite the necessary efforts to annotate the blends adequately, the inherent ambiguities of truncation are inevitably mirrored in the distributional analyses. Figure 25 presents the distribution of the five categories ABD, ACD, AD, A|BC|D, and Ax based on the annotation of the blend list. There are other and more elaborate systems to account for truncation (see for instance Beliaeva, 2014a), but a more minimalistic representation is chosen in this investigation to keep the analytical model slim. The descriptively reduced Ax tag is assigned to the blends *Texarkana* (*Texas* + *Arkansas* + *Louisiana*) and *teledildonics* (*tele*(*communication*) + *dildo* + *electronics*). The detailed annotation of the blend list (see Appendix 3) makes use of the additional letters EF to account for their truncation structure, but since there are only two unique forms (*Texarkana* occurs twice in the blend list) they are subsumed under the Ax tag in the figure. The bars in Figure 25 marked as *primary assessment* represent the results of the annotation of the standard truncation patterns. The striped bars show how many of the respective blends per truncation category that afford an alternative interpretation if only formal properties are considered. The y-axis represents the actual number of items, and the exact number per category is shown in italics on top of each bar. It should be stressed that the low level of ambiguous items in the A|BC|D category only indicates that other possible truncation patterns are typically insufficient to account for their structure. All the same, there is an inherent ambiguity in all A|BC|D blends, as previously exemplified with the blends *popera* and *scientifiction*. That is, *pop|era* and *p|opera* are simultaneously possible interpretations subsumed in the representation *p|op|era*.

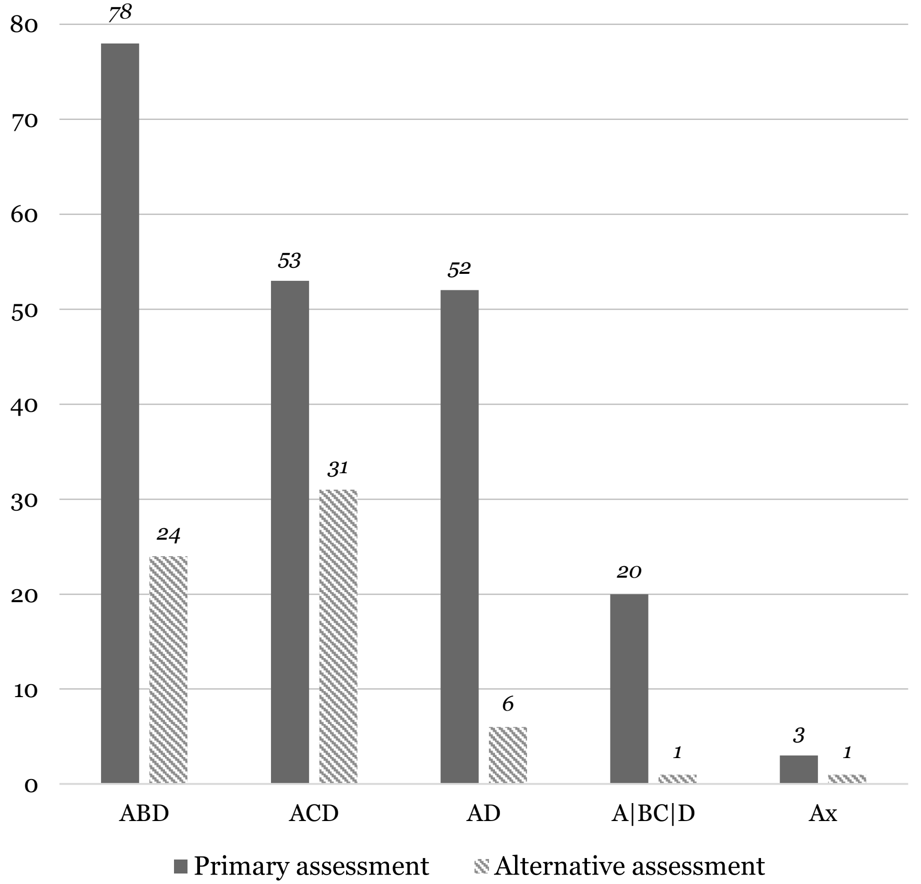


Figure 25. Distribution of truncation patterns.

Given the primary assessment, the most prominent truncation pattern is ABD (78 blends). Examples in this category are *webzine* (*web* + *magazine*) and *tricknology* (*trick* + *technology*). The first source word is kept intact in these blends. The second most prominent category is ACD (53 blends), which represents blends where the second source word is retained in full. Instances of ACD blends are *frienemy* (*friend* + *enemy*) and *faceblock* (*Facebook* + *block*). Taken together, the ABD and ACD blends constitute thus 131 out of the total of

206 blends in the data. The AD pattern, which implies truncation of both source words (e.g., *Mexicatessen* from *Mexican*/*Mexico* + *delicatessen*), occurs less than half as many times in the data (52 AD blends vs. 131 ABD and ACD blends). This distribution is notable given the traditional assumption that "the majority of blends combine the initial part of the first source word (W1) with the final part of the second source word (W2)" (Beliaeva, 2014a: 33-34). However, the quantitative measures of the alternative assessments in Figure 25 give an indication of the need for complementary analyses of the high levels of truncation ambiguity in the single truncation types (ABD and ACD).

The vast majority of alternative assessments are of the AD type pattern (54 out of 63 items in total). Should the distributional analysis be based on the alternative assessments, the AD blends could maximally constitute 100 of the 206 blends in the blend list. Figure 26 accounts for how the alternative assessments are distributed in relation to the primary annotation.

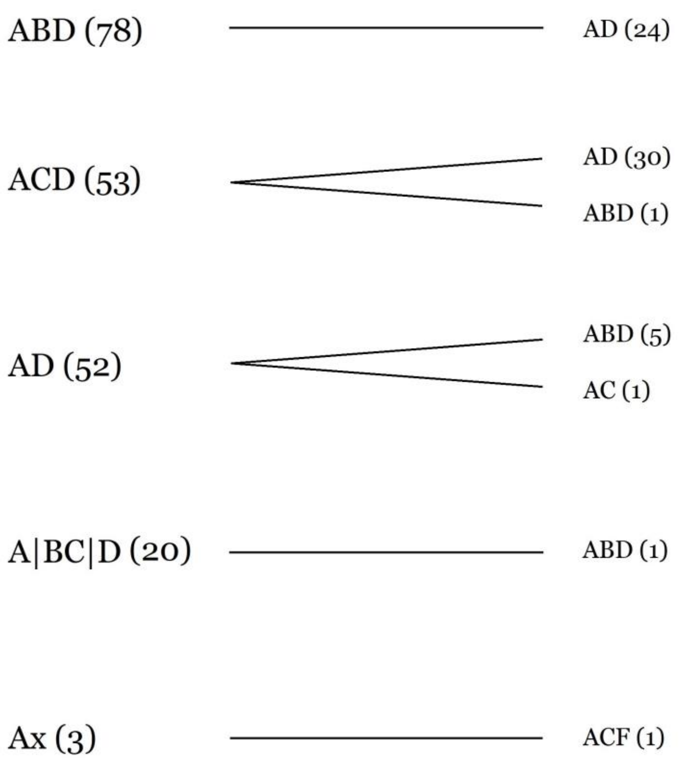


Figure 26. Alternative interpretations of standard truncation patterns.

The numbers in Figure 26 may give the impression that the AD truncation is more prominent than is suggested by the results of the primary assessment. However, formal affordances of AD truncations such as *execut|rain* (*executive* + *train*) and *Sens|urround* (*sense* + *surround*) remain largely on a structural level, while an analysis of their use suggests a more plausible ACD truncation pattern *execu|train* (79) and *Sen|surround* (80).

(79) ExecuTrain of Texas, a worldwide corporate training provider and world leader in business education training (NOW).

(80) Sensurround, I think they called it, with the whole cinema vibrating at a few hertz to give you a feeling of earthquake. Those were the subs that were in the Garage. The other thing that was kind of cool about it was they had stuck up little horns and speakers into corners that pealed off at certain frequencies. You actually got three-dimensional sound (NOW).

Some annotations of truncation patterns appear more ambiguous even when considering semantic qualities. ABD blends such as *brainiac* (*brain* + *maniac*) and *schlockumentary* (*schlock* + *documentary*) are reasonably represented as *brain|iac* (81) and *schlock|umentary* (82). However, the phonemic representations of these items may increase the impact of truncation ambiguity. In both cases the alternative breakpoints /bɹ|eɪniæk/ and /ʃl|ɑkjəmɛntəɹi/ seem plausible, thus making them AD blends according to the truncations *br|a*[*i*]*niac* and *schl|oc*[*k*]*umentary*.

(81) And the link between creativity [...] and mental disorders has been the subject of much study and forms part of our cultural lore. Consider the image of the “mad scientist,” or ranting brainiac, he said (NOW).

(82) the displeasure of seeing the new documentary titled “Merchants of Doubt.” The film's argument is the same as virtually every other left-wing *schlockumentary* that pops up on Netflix these days: Evil corporate interests are standing against positive social change. # On a purely theatrical basis, the film is a flop (NOW).

Other items in the blend list display yet more intricate patterns of truncation. For instance, *Scattergories* (*scatter* + *categories*) seems formally an unambiguous ABD blend (*Scatter|gories*). Still, the striking similarity between *Scatter* and the initial part of *categories* suggest that alternative solutions such as *S|cat*(*t*)*e*(*r*)*gories* (ACD) or *Sca|t*(*t*)*e*(*r*)*gories* (AD) may be possible (cf. Gries, 2004b: 653). The similarity affording this alternative truncation pattern is even more prominent if the phonemic representation /skætɚgɔɹiz/ is considered. In fact, it is only the difference between /ɚ/ and /ə/ that distinguishes /skætɚgɔɹiz/ from an A|BC|D pattern. The decision to let *Scattergories* be seen primarily as an ABD blend rests on the assessment that *scatter* is overtly foregrounded in the structure of the blend. Nonetheless, this item exemplifies the potential impact of truncation ambiguity.

The 20 A|BC|D blends shown in Figures 25 and 26 constitute 10% of the blend data, which suggests that they represent a rather common blending strategy in the corpus data. *Popera* and *palimony* could be described as prototypical, while there are borderline cases such as *splatterific* (*splatter* + *terrific*) and *Sense8* (*sensate* + *8* [*of August*]). *Splatterific* lacks the grapheme *r*, and it has thus been given an alternative assessment as an ABD structure (*splatter*|?*ific*). Yet, it is retained in the A|BC|D category on the grounds that the phonological realization of *terrific* is thought to be unobscured by the blend formation (cf. Lehrer, 2007: 118). *Sense8* is perhaps also possible to regard as a single truncation blend (ACD in this case), but the striking phonological impact is considered satisfactory for an A|BC|D interpretation. The use of the number *8* in *Sense8* implies a strong activation of phonemic decoding strategies to disentangle the unorthodox orthography (i.e., /eit/ in both *-ate* and *8*). It should be noted that there are instances with an A|BC|D resemblance in the data that have been interpreted as belonging to other truncation types. For instance, the ACD blend *collaboraction* (*collaboration* + *action*) differs both graphemically (an intercalated *-c-*) and phonemically (/eɪ/ vs. /æk/) from the A|BC|D pattern to the extent that it is motivated to consider it a single truncation blend (i.e., *action* is visible in full in the blend). Such items have not been given an alternative assessment because of the apparent structural difference from A|BC|D blends.

The inherent ambiguity of A|BC|D blends does not imply that the use of truncation ambiguity is consistent in these items. This is shown by displaying their variation of overlap. The proportion between the shared structure and the total number of graphemes has been calculated for each identified item, and the distribution is visualized in Figure 27. The percentage values represent how much of a specific blend is constituted by source word material that is shared in the blend. For instance, 20% of the blend *Thinsulate* is overlapping structure shared by the source words. Two duplicates (*popera* and *disgraceland*) are included as they constitute instantiations with quantitative import in the distributional analysis. Excluding them from quantitative analyses would risk invalidating the results, since they originate from the randomized sample of CaWs (see 5.1.2).

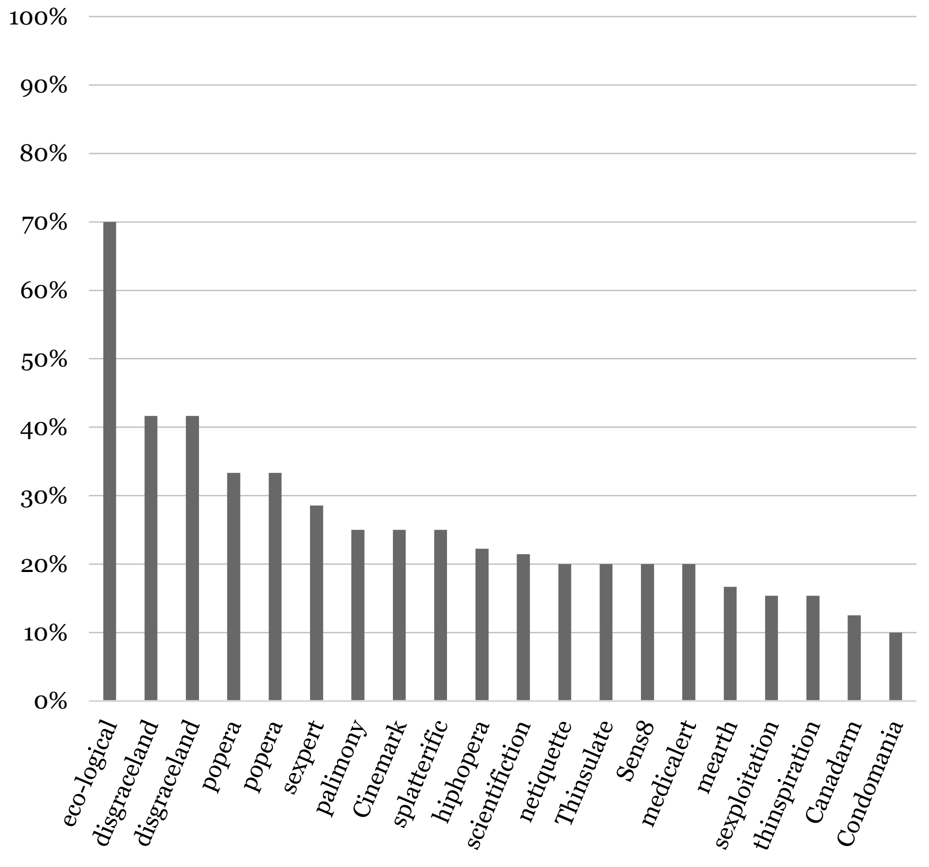


Figure 27. Proportions of overlap in A|BC|D blends.

In the minimal overlap of *Condomania* and *Canadarm*, the only overlapping graphemes *m* and *a* constitute 10% and 13% respectively of the entire graphemic representations of the blends. In contrast, the blend *eco-logical* (*ecological* + *logical*) consists of 70% shared overlapping structure. This high degree of overlap may motivate the use of a hyphen (-), since this grapheme facilitates the identification of the source word *logical*. The reference to *logical* would otherwise be difficult to trace in a (hypothetical) homographic blend *ecological*.

The distributional analysis of overlap in A|BC|D blends indicates that the ambiguity of truncation varies considerably in this category. However, a parallel analysis of overlap in blends with other truncation patterns (186 items) shows a similar distribution. The proportions (in decreasing order) of these blends are represented as a line in Figure 28 instead of bars. The major difference between A|BC|D blends and other truncation types is that the A|BC|D pattern requires shared structure, while this is optional in all other types of truncation. Accordingly, the flat line to the right in Figure 28 represents 66 blends with no registered graphemic overlap, of which 51 have been identified as blends characterized by pseudomorphemic transfer (e.g., *Popmart* from *pop* + *Kmart*, and *vactor* from *vacuum* + *excavator*). Besides the occurrence of blends with no overlap, a comparison between the distributional patterns in Figures 27 and 28 shows that the A|BC|D blends do not differ markedly in the amount of shared structure. These quantitative analyses of overlap thus constitute a preliminary measure of the potential distribution of truncation ambiguity. Still, it is important to keep in mind that other factors are important as well, such as the localization of breakpoints (cf. *relationshit*), the morphemic quality of segments (cf. *technovation*), and the relation between graphemic and phonemic realization (cf. *Spinderella*).

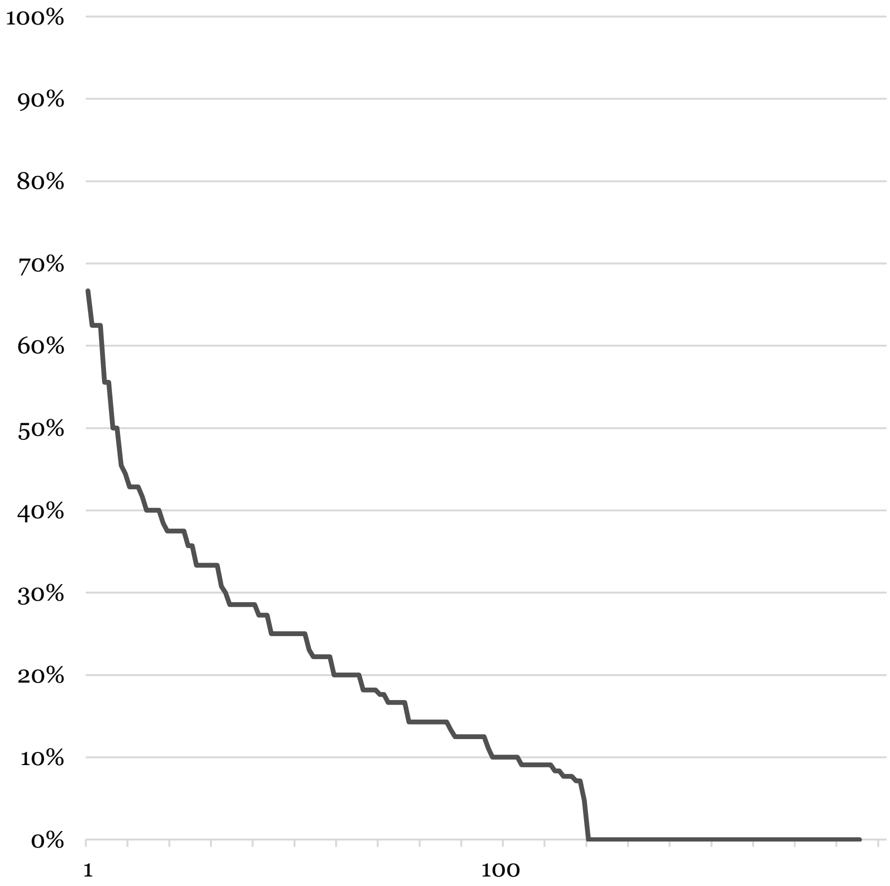


Figure 28. Proportions of overlap in ABD, ACD, AD, and Ax blends.

The analyses in this section are intended to outline general implications of truncation ambiguity emerging from the patterns of truncation in the data. Several tendencies of truncation are identified. One noteworthy observation is that the dominant strategy in the list of blends is to retain one of the source words in full. This accords with previous observations on blend structure, namely that truncation need not occur in both sources (e.g., Algeo, 1977; Gries, 2006; Lehrer, 2007). The observed patterns of distribution indicate that the AD pattern may not be quantitatively central in these data. Furthermore, the A|BC|D truncation type occurs rather frequently in the blend list (but see López Rúa, 2012 in which they are twice as common). The distribution of overlap in these blends indicates that truncation ambiguity varies from modest levels (e.g., *Canad|a|rm*) to a considerable part of the blend (e.g., *eco-|logical|*), but also that this diversity is discernible in the data as a whole. The A|BC|D structure can be contrasted with the AD, ABD, and ACD types because of the apparent preservation of source words, but the borderline cases represented by both included and excluded items in the process of categorization suggest that the difference between the truncation types is a matter of degree.

Truncation has been extensively discussed in blend research. It remains problematic in relation to classifying and defining blends, although there is some agreement on a few general assumptions (cf. Balteiro & Bauer, 2019). One of the main problems concerns the distribution of the standard truncation patterns (cf. Figure 25). As previously mentioned, the AD construction, in which the first source word is truncated finally and the second source word is truncated initially, has been considered the standard blend type. However, its prominent status remains an open question since few studies have addressed its assumed dominance in systematically collected datasets (cf. Wulff & Gries, 2019). It is worth noting that Cannon (1986) reported a similarly limited amount of AD blends in his systematically collected set of blends (see 3.4.1). The analyses of distribution presented here suggest that the AD pattern is not particularly prominent in the context of this study. However, the alternative assessment, implying a narrower formal analysis, would nearly double the number of AD blends (100 out of 206 blends). This highlights the impact of a broader analytic scope in the annotation of truncation patterns. It also shows that truncation ambiguity can be said to be a central characteristic of the blends in these corpus data. These emerging patterns are claimed to reflect characteristics of blends in the entire subcorpus, which gives an indication of how truncation ambiguity in blending functions in the context of American web news within the selected temporal frame.

It is a rather uncontroversial assumption that truncated segments function as segments with an activation potential for their source words (see for instance Lehrer, 2007: 116). This can be expressed in Cognitive grammar terms in that truncated blend segments "do not in themselves have symbolic status" (Taylor, 2015: 152). Therefore, claiming that both source words are preserved in full in A|BC|D blends implies that *popera* should be seen as a compound. Following this line of reasoning, the compound *pop opera* (example 83) should be analyzed in the same way as *popera*, which would go against their apparent structural differences (see also Correia Saavedra, 2016: 58).

(83) An epic pop opera surf rock concept album (NOW).

It is here suggested that this reasoning sets aside the fact that blending is the chosen word formation strategy in the recorded cases, regardless of what the reasons for it might be.[17] From a generative linguistics perspective, it could be argued that the blend is a deviating performance phenomenon while compounding represents the significant analytic level. A different approach is taken in the current investigation. The chosen construction (blending) is seen as meaningful in itself, which parallels research on the meaning of grammatical constructions (Langacker, 2002; Taylor, 2002; Goldberg, 2005; Langacker, 2008; Booij, 2010). This suggestion is exemplified by the structural ambiguity relying on truncation as a textual process with conceptual implications, which is a mechanism influencing the symbolic status of the blend. Furthermore, this stance relies on the assumption that the form of the blend cannot be separated from its semantic structure. Thus, blending is not randomly employed, but has symbolic potential in its own right, and truncation ambiguity is a meaningful and powerful submechanism of this strategy (see also Kemmer, 2003; Renner, 2015).

In sum, the analyses of the blends highlight the gradual nature of truncation ambiguity but also the impact of various strategies in its realization. Therefore, it seems reasonable to see truncation ambiguity as shifts of profile instead of only trying to account for a gradual decrease of structural reduction. For instance, the degree of source word transparency is rather high in *collaboraction*, while the simultaneous profiling of the source words implies the possibility of a wordplay effect. In this view, truncation becomes a dynamic process that does not necessarily entail the reduction of graphemic or phonemic structure, but can just as well be a matter of profiling certain segments of the lexical construct (cf. Langacker, 2008: 98 on the realization of grammatical categories as a matter of profiling).

This section has shown that truncation ambiguity depends on several different strategies of blend formation. One of these strategies is the recurring discrepancy between the written and spoken modes of language. The tension between these

[17] The blend *motel* is another authentic example of the initial co-existence of a blend and a compound form, which resulted in the blend becoming the conventional term (see 3.3.5).

two modes is exploited in another type of ambiguity, which is the topic of the following section.

6.2 Mode ambiguity

Several studies stress the that phonological aspects are important in the formation of lexical blends (e.g., Kubozono, 1990; Kemmer, 2003). At the same time, it has also been emphasized that the use of blends is characterized by a "heavy dominance of writing" (Cannon, 2000: 955; see also Gries, 2004b: 656). These claims point to the complex interrelation between sound and spelling in blends. This section addresses a phenomenon often observed in blends, in which conflicting domains are activated simultaneously by sound vs. spelling, hence creating an ambiguity arising from phonological inferences that conflict with graphemic representations (cf. Fandrych, 2008: 77). Blends exploiting this mechanism are called graphemic blends (see 3.2.6), and the concept of graphemic dependence is used to capture the key role of the written medium for the understanding of such blends (cf. Renner, 2015: 127). Examples are *Scitation* (*science* + *citation*) and *Sense8* (*sensate* + *8* [*of August*]). Their sound representation activates only one of the source words (*citation* and *sensate* respectively). In this investigation, observations of graphemic dependence suggest that ambiguity between modes of expression is an influential mechanism for blend formation. The chosen label for this phenomenon is *mode ambiguity*.

The pattern of graphemic dependence in blends such as *Scitation* is widely recognized (e.g., Algeo, 1977; Ronneberger-Sibold, 2006; Lehrer, 2007; Renner, 2015), but its motivations and potential impact on blend formation have not yet been sufficiently described. Renner (2015) highlights wordplay as a key explanatory factor in blending, suggesting that blends capture a communicative event "founded on the ludic exploration of the limits of verbal inventiveness and recognizability" (Renner, 2015: 131). Although Renner's idea aligns with the rationale behind mode ambiguity, it is suggested here that the more general cognitive principle of ambiguity is at work as an important characteristic of blending, and, more specifically, that mode ambiguity is what characterizes graphemic dependence.

Conflicts and connectivity of modes are attested in other contexts as well. Three observations are presented here to compare mode ambiguity to related phenomena. First, effects of the so-called *Stroop task* are frequently attested in psycholinguistic research since the publication of Stroop (1935). Harley (2008) explains these effects by saying that "[n]aming the color in which a word is written is made more difficult if the color name and the word conflict (e.g., 'red' written in green ink)" (Harley, 2008: 178). Thus, the interference of visual perception slows down the recognition of the graphemic representation. Second, cognitive connectivity is observed in studies on synesthesia, in which an inducer (e.g., graphemes) elicits a concurrent experience (e.g., color) in a mechanism that is sometimes described as an "alternative mode of perceiving (and conceptualizing) the world" (Ward, 2013: 69; see also Simner, 2012). Third, certain phonemic representations have been shown to be iconically related to specific shapes or concepts. Köhler's (1929) "bouba-kiki" study reveals a strong preference for sounds such as /k/ and /t/ to represent spiky and pointed objects such as the leftmost shape in Figure 29 . In contrast, comparatively soft labial sounds were associated with the shape to the right in Figure 29. The phenomenon is commonly referred to as *sound symbolism*. Ramachandran and Hubbard (2001), for instance, associate it with synesthesia, and Bross (2018) investigates sound symbolism from the perspective of correlations between length of vowels and length of corresponding objects.

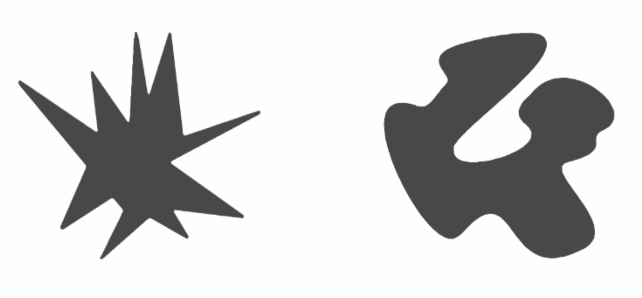


Figure 29. Figures exemplifying the Bouba-kiki effect.

These observations of visual interference, synesthesia, and sound symbolism highlight the potential of cognitive connectivity in language. The notion of mode ambiguity aligns with such observations in that it is motivated by conflicting modes of expression. Having thus outlined the concept of mode ambiguity and its plausibility as an analytic category, the following paragraphs describe the realization of graphemic dependence as the main source of mode ambiguity.

Graphemic blends are rare in the data, but their qualitative potential for cognitive intricacy is striking. Eight items have been identified in the list of blends. Table 17 presents these blends together with their source words (abbreviated as SW1 and SW2) and patterns of correspondence. The column *phonemic representation* offers the suggested identical representations of the blend and its phonologically similar source word. The blend *Robocup* occurs twice in the data but is represented only once in the table.

Table 17. Graphemic blends.

Blend	***SW1***	***SW2***	***Corresponding items***	***Phonemic representation***
Youthanasia	*youth*	*euthanasia*	Blend – SW2	/juːθəneɪʒə/
Draganfly	*Dragan*	*dragonfly*	Blend – SW2	/dɹægənflaɪ/
Websight	*website*	*sight*	Blend – SW1	/wɛbsaɪt/
Robocup	*Robocop*	*cup*	Blend – SW1	/ɹoʊbəkʌp/ but cf. /-kɒp/
Scitation	*science*	*citation*	Blend – SW2	/saɪteɪʃən/
eco-logical	*ecological*	*logical*	Blend – SW1	/iːkəlɒdʒɪkl/
Sense8	*sense*	*'sensate'* [/eɪt/]	Blend – SW2	/sɛnseɪt/

The items *Youthanasia* and *Websight* seem to display perfect phonemic similarity of the blend and one of the source words. However, not all blends in Table 17 are that straightforward. In *Robocup* there is a potential discrepancy between /kɒp/ and /kʌp/ in the final syllables of the source words. The proximity of the phonemes /ɒ/ and /ʌ/ has been considered minimal in this case, which is why *Robocup* is thought of as a graphemic blend. The blend *Sense8* requires a kind of phonemic feedback loop to retrieve the intended source word *sensate*, which is not orthographically explicit in the graphemic representation.

The conflict between two simultaneously activated conceptual domains can be further illustrated by an analysis of the blend *Websight*. A query for lemma tokens of the source word *website* returns 158,175 hits, which suggests that this item is a relatively conventionalized item in the subcorpus. The output of a parallel query investigating *Websight* amounts to only two hits. Hence, *website* is "more cognitively entrenched [...] than the resulting blend" (Kemmer, 2003: 71). In psycholinguistic terms, the precedence of *website* is overwhelming from a lexical selection perspective (see Harley, 2008: 173-174). Hearing or pronouncing /wɛbsaɪt/ will activate the conceptual domain WEBSITE instead of associations related to the blend *Websight* unless some other stimulus is provided. Therefore, decoding the (graphemically represented) item *Websight* presumably involves a point at which the phonemic realization activates the source word *website*, since the phonemic representations of the blend and its first source word are identical. The impact of this cognitive conflict is further emphasized by a potential need to figure out what is the intended meaning of the blend. The concept SIGHT related to *Websight* may give rise to a broad range of possible references, although it may have a specific technical meaning within a certain context. Nonetheless, even though a specific (intended) meaning may be known to a language user, the meaning potential activated by *sight* likely influences the construal of the whole concept.

The low frequency of graphemic blends in the data could be seen as a sign of peripheral status, which may question the potential impact of mode ambiguity. However, the existence of forms displaying only a minimal difference between the blend and one of its source words suggests that this type of ambiguity plays an important role. Blends such as *egg-cellent* (*egg* + *excellent*) and *ego-system* (*ego* + *ecosystem*) are distinguished from graphemic blends only by the voicing of the velar stop in the first consonant (/g/ vs. /k/). The fact that a hyphen is inserted in both these blends may even be an indication of a need to underscore the blended construct so as to avoid misinterpretation or suspicions that the word is misspelled (cf. *in-sin-uation* in Algeo, 1977: see also 3.2.6). Blends with one single phonetic property distinguishing them from full graphemic dependence are henceforth referred to as *near-graphemic blends*, and the phonetic property separating them from graphemic blends is called its *distinguishing sound*.

In order to investigate more closely the impact of mode ambiguity, a quantitative analysis was conducted including both graphemic and near-graphemic blends. The annotation required for this analysis was extended not only to distinguish between graphemic and near-graphemic blends, but also to discriminate between stressed and unstressed positions of the distinguishing sounds in near-graphemic blends. This subcategorization was motivated by the assumption that sound changes in unstressed syllables are generally taken to be harder to detect. Figure 30 presents the results of summarizing the blends falling into one of the categories *distinguishing sound in stressed syllable*, *distinguishing sound in unstressed syllable*, and *fully graphemic blends*. Example blends with a distinguishing sound in a stressed syllable are *Addaction* (in the rhyme), *Crooklyn* (in the onset), and *modfather* (in the onset). Instances where the change is located in an unstressed syllable are, for instance, *relationshit* (in the coda), *Canstruction* (in the rhyme), and *Kongfrontation* ([King] *Kong* + *confrontation*) (in the coda). The numbers attached to the right of each bar in the figure represent the actual number of items, including duplicates that were identified in each category.

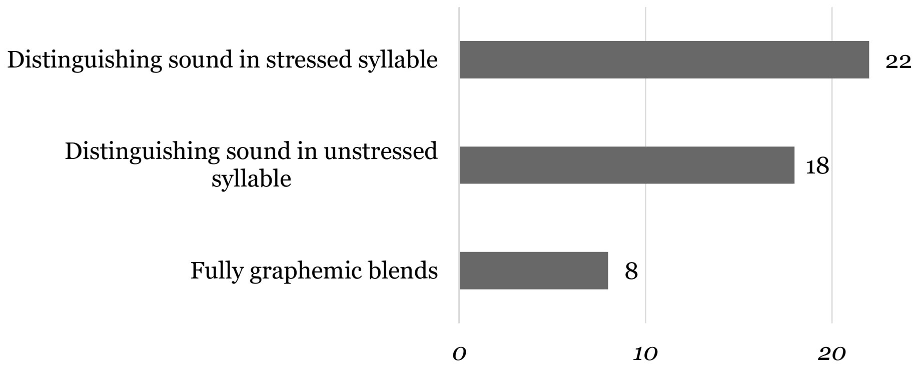


Figure 30. Fully graphemic and near-graphemic blends.

As Figure 30 shows, there are many more near-graphemic blends than fully graphemic blends. This distribution illustrates the impact of a prototype theory categorization of graphemic dependence. If only fully graphemic blends were included in the analysis, graphemic dependence might appear to be of limited import. However, if graphemic dependence is recognized in near-graphemic blends such as *egg-cellent* and *relationshit*, then a sizeable part of the blends in the data (48 out of 206 blends) are assumed to be influenced by this mechanism.

The gradual distribution of graphemic dependence is not restricted to the quantitative measure based on the detection of a distinguishing sound in a stressed or unstressed syllable. The overall structure may afford graphemic dependence even in blends with more than one distinguishing sound. For instance, the difference between *spaghetti* and *spay-ghetti* is represented phonemically with the exchange of the initial schwa /ə/ in the source word to /eɪ/ in the blend. These two sounds in the blend disqualifies it as a near-graphemic blend in the current analysis, but the difference motivating this discrimination seems minimal. The markedness of the overall structure of *spaghetti* is illustrated by the fact that only two alternative word forms were retrieved in a subcorpus wildcard query of *sp*ghetti*: the compound *spin-spaghetti* and the proper noun *Springhetti*.[18] Another notable example is *extravaganza*, which is typically represented phonemically as /ɪkstɹævəgænzə/ (Cambridge Dictionary, n.d.; Macmillan Dictionary, n.d.; Merriam-Webster, n.d.-a), while a phonemic representation of the three initial phonemes in the blend *eggstravaganza* would likely be /ɛgz-/. Based on this analysis, *eggstravaganza* is not considered a near-graphemic blend, since three sounds distinguish *extravaganza* from *eggstravaganza* (i.e., /ɪks-/ vs. /ɛgz-/). However, the actual realizations of the initial vowels are likely so close that it seems reasonable to acknowledge that

[18] A wider corpus search including hits later than March 2018 also included the blends °*spook-ghetti* (*spooky* + *spaghetti*) and °*SPA-ghetti* (*spa* + *spaghetti*). However qualitatively informative, these additional items have no quantitative importance in this case.

graphemic dependence has influenced this blend as well (Dr. Gregory Garretson, personal communication).

Furthermore, the analysis of near-graphemic blends did not consider the phonemes /tʃ/ and /dʒ/ as single entities, which means that some forms were excluded as near-graphemic blends. For instance, *Chinglish* (/tʃɪŋglɪʃ/ or /tʃɪŋlɪʃ/) was not annotated as near-graphemic, since /tʃ/ was seen as two sounds distinguishing it from the source word *English*. The reason for this category delineation is that lexical decoding based on sound differences was in focus. Although /tʃ/ and /dʒ/ are usually seen as single phonemes (i.e., *palato-alveolar affricates*, see for instance Yavaş, 2006) the possibility of a phonetic distinction between [t] and [ʃ] was considered more important here.

The quality of a sound change, for instance in terms of place and manner of articulation, may also influence the gradual distribution of graphemic dependence. Comparing the example pairs of blends *Addaction–Sportacus* and *Canstruction–Spooktacular* will serve to illustrate this observation. The blends *Addaction* and *Sportacus* are near-graphemic blends since they have only one sound that distinguishes them phonemically from their source words *addiction* and *Spartacus*. However, their sound changes differ in quality. The leftmost arrow in the IPA charts of Figure 31 indicates how the initial vowel of *Spartacus* changes from /ɑ/ to /ɔ/ in *Sportacus*. The chart to the right suggests the sound change that occurs in the blend *Addaction* in relation to *addiction*. The change of the vowel sound in *Sportacus* (i.e., from /ɑ/ to /ɔ/) constitutes a move from front position to back position, while the open-close distinction (vertical movement in the IPA chart) is changed only moderately. In *Addaction* the front position is retained in the sound change, but the vowel shifts instead from closed position to open position (i.e., from /ɪ/ to /æ/). The arrows in the figure suggest that the sound changes are qualitatively different. Still, the fact that the arrows are similar in size indicates that they may yet have a relatively equal degree of graphemic dependence.

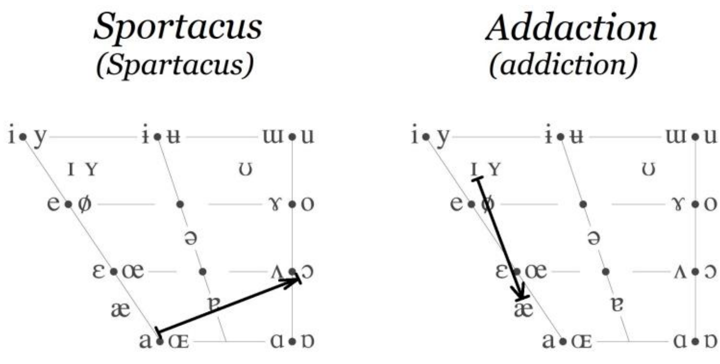


Figure 31. The qualitative phonetic difference between blends and source words in *Sportacus* and *Addaction* respectively.

The qualitative difference between the sound changes seems to have greater impact in the example pair *Canstruction–Spooktacular*. Figure 32 illustrates how the distinguishing sound in *Canstruction* moves towards open front (from /ə/ to /æ/), while the direction of the shift in *Spooktacular* is the opposite, i.e., towards closed back (from /ɛ/ to /uː/). The size of the arrows in Figure 32 suggests that the sound change in *Canstruction* may be perceived as smaller than in *Spooktacular*. Hence, *Canstruction* seems more graphemically dependent due to the smaller impact of the sound change of the distinguishing sound, and ambiguity of modes may therefore be a more dominant motivation factor in this blend.

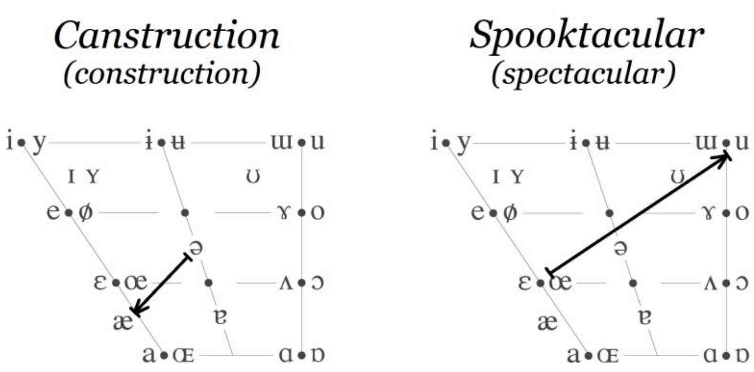


Figure 32. The qualitative phonetic difference between blends and source words in *Canstruction* and *Spooktacular* respectively.

All four blends in Figures 31 and 32 are annotated as near-graphemic blends since they have only one distinguishing sound between the blend and one of their respective source words. These differences of sound occur in stressed syllables in *Addaction–Sportacus* and in unstressed syllables in *Canstruction–Spooktacular*. Despite the systematic similarities of their number of distinguishing sounds and their position with regard to syllabic stress, they exemplify qualitative differences that may influence the impact of graphemic dependence.

Degree of markedness in the distinguishing sound is further illustrated by the blends *Crooklyn* (*crook* + *Brooklyn*) and *egg-cellent* (*egg* + *excellent*). The change of consonants between *excellent* and *egg-cellent* is voicing only, while the difference between *Brooklyn* and *Crooklyn* is more prominent. The manner of articulation changes from a bilabial voiced realization of /b/ in the front of the oral cavity to the unvoiced velar stop /k/ in the back of the mouth, which is unlikely to pass unnoticed in use.

The gradual distribution of blend characteristics, both generally (e.g., Gries, 2012) and concerning certain attributes (e.g., Algeo, 1977 on overlap), is reflected in the distribution of graphemic dependence as well. In the analyses presented here, the near-graphemic blends are interesting because of their striking similarity with one of the source words. The minimal differences between fully graphemic and near-graphemic blends in the data constitute an additional hint to the type of influence exerted by mode ambiguity. The activation tension emerging from the conflicting modes of realization can be elaborated with factors such as position, articulation quality, and stress of distinguishing sounds. These factors in combination determine the level of phonemic transparency of a particular blend.

In all, it has been shown in this section that graphemic dependence functions as a creative way to achieve ambiguity of modes. This mechanism can be framed as a cognitive challenge, in which one mode (speech) activates a relatively frequent concept and another mode (writing) activates a conflicting concept, often of marginal frequency. The observations of the gradual nature of graphemic dependence points to the explanatory potential to address mode ambiguity as a cognitive mechanism that influences lexical blending to various degrees.

The ambiguity treated in this section is realized between modes of expression, but ambiguity can also arise from the simultaneous activation of potential source words of blends. The following section investigates the impact of one such type of ambiguity, in which two explicitly recognizable source words seem to compete, affording two possible interpretations of the blend.

6.3 Source word ambiguity

Some blends appear to be optimized for simultaneous activation of the blend and one of its source words. For instance, the blend *Constraction* (*construction* + *action*) is highly similar to its first source word (*construction*), both in its orthography and its prosody. At the same time, the second source word (*action*) is fully visible in the blend. Hence, the striking similarity between one of the source words and the blend causes a conceptual conflict not only between these two items, but also in terms of which source word may be activated. This phenomenon is called *source word ambiguity* in this investigation.

Source word ambiguity is particularly notable in blends with an intercalative structure (cf. 3.2.5). As the term suggests, intercalation concerns constructions that have been modified in medial position by the insertion of structure from another source concept. A schematic representation of intercalation could be described as AB + CD > A[*cd*]B or C[*ab*]D, in which [*cd*] and [*ab*] symbolize any type, position, and amount of inserted or exchanged structure originating from the other source word. In the example *Constr*[*a*]*ction* the difference between the blend and its source word *construction* resides in the insertion of an *a* (from *action*) instead of the original *u*. However, intercalative blends typically also afford an alternative interpretation in which the blend is analyzed according to one of the standard truncation patterns as outlined in 3.2.2 (Kemmer, 2003; see also 3.2.5). Hence, *Constr*[*a*]*ction* (i.e., A[*cd*]B) could also be seen as an ACD blend with the breakpoint *Constr|action*. The schematic representation for these two simultaneously possible interpretations of *Constraction* is illustrated by the following formula:

$$\mathrm{AB} + \mathrm{CD} > \begin{cases} \mathrm{A}[cd]\mathrm{B} \\ \mathrm{ACD} \end{cases}$$

In this schematic example, the first source word (AB) has some degree of similarity with the intercalative interpretation of the blend (A[*cd*]B). The higher the degree of similarity, the more likely it is that the source word AB may be activated by the blend (A[*cd*]B). At the same time, the second source word (CD) is fully recognizable in the blend given the equally possible ACD interpretation. Thus, the more similar the A[*cd*]B and the AB source words, the more probable is a conflict of activation between the source words AB and CD. Because of this close connection between intercalation and source word ambiguity, intercalation is examined in detail as a means to investigate the affordances of source word ambiguity on the blend data.

Intercalation of structure has received rather little attention in previous accounts on lexical blending, but it has long been recognized as an available blend formation pattern. Algeo (1977) uses the metaphorical term *sandwich word* for blends such as °*daffynition* (*daffy* + *definition*) and °*insinuendo* (*insinuation* + *innuendo*).[19] Kemmer (2003) draws attention to the inherent structural ambiguity of intercalative blends. She describes intercalation in blending as "a minimally productive pattern, although the creative resources of the language certainly allow it" (Kemmer, 2003: 73; see also Cook & Stevenson, 2010: 130). Gries (2004b) questions the explanatory value of intercalation as an analytic category since intercalative blends can at the same time be seen as non-intercalative (cf. Kemmer, 2003: 72). Intercalative blends are also mentioned in Renner's study on wordplay (2015), subsumed under the category of formal complexity (see 3.2.7). These observations serve as the background for the assumption that the ambiguity of intercalative blends does not have to cause a taxonomic dilemma. Instead, this ambiguity may even be an attractive trait that motivates the use of intercalative structures in blending (cf. Giora et al., 2015; McMahan & Evans, 2018).

The annotation applied in this study distinguishes between graphemic and phonemic intercalation. These two types of intercalation need not necessarily co-occur. As pointed out in 6.1, *Spinderella* is graphemically an ABD blend in which the second source word is subject to truncation, while its phonemic representation could be seen as intercalative (/s[p]ɪndəɹɛlə/). The degree of intercalation may also vary graphemically and phonemically within a blend, which suggests that source word ambiguity may have a greater impact on either type of representation. For instance, *tricknology* is graphemically intercalative (i.e., *t*[*rick*]*nology*), but the similarity with its source word *technology* is smaller than in its phonemic counterpart. In Figure 33, this discrepancy is marked with gray color where the graphemic and phonemic structures are identical. The smaller number of differing items in the phonemic representation shows that *tricknology* is phonemically more similar to its source word *technology* than are the graphemic similarity of *tricknology* and *technology*.

[19] It is worth noting that the two examples °*autobydography* (*autobiography* + *by dog*) and °*in-sin-uation* (*insinuation* + *sin*) in Algeo (1977) are unambiguously intercalative. They resist any of the standard truncation patterns that are typically alternative interpretations of intercalative blends.

Graphemic representation	t	r	i	c	k	n	o	l	o	g	y
	t		e	c	h	n	o	l	o	g	y
Phonemic representation	t	ɹ	ɪ	k	n	ɑ	l	ə	d	ʒ	i
	t		ɛ	k	n	ɑ	l	ə	d	ʒ	i

Figure 33. Representations of *tricknology* and *technology*.

Technology (/tɛkˈnɑlədʒi/) and *tricknology* are assumed to have the same stress pattern (cf. Arndt-Lappe & Plag, 2013). Accordingly, the intercalated structure occurs in an unstressed syllable. This means that the similarity between /tɛknɑlədʒi/ and /tɹɪknɑlədʒi/ is presumably even more pronounced than shown in Figure 33. It has therefore a higher probability to pass unnoticed in speech. Accordingly, the potential for simultaneous source word activation – or source word ambiguity – seems greater phonemically, since /tɹɪk/ is fully recognizable and the blend /tɹɪknɑlədʒi/ is highly similar to the source word /tɛknɑlədʒi/.

Examining the blend list shows that 45 blends display some degree of intercalation. Some examples are *sp*[*ook*]*tacular, spa*[*y-*]*ghetti, intelli*[*s*]*en*[*s*]*e,* and *Spa*[*ngl*]*ish*. The degree of graphemic intercalation varies from a singular grapheme (e.g., *constr*[*a*]*ction*) to clusters of graphemes (e.g., *t*[*rick*]*nology*). Figure 34 shows that 22% of the blends in the data are intercalative to some extent. The presentation in the figure does not take into account how much structure is inserted in an intercalative blend. Neither does it distinguish between graphemic and phonemic intercalation.

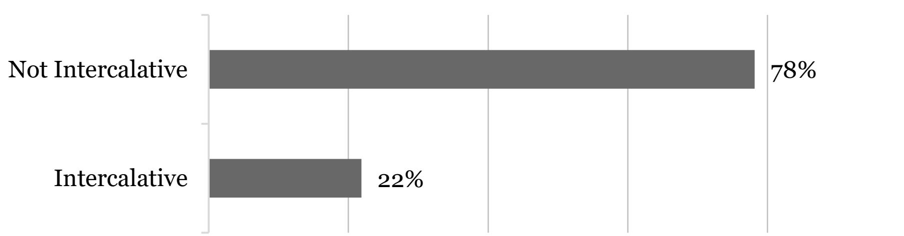


Figure 34. Intercalative blends in the data.

Constr[*a*]*ction*, *glo*[*c*]*al*, *Roboc*[*u*]*p*, and *Drag*[*a*]*nfly* are examples of blends clearly identified as intercalative. Besides their intercalative structure, they typically display a graphemic dependence to the extent that they are either near-graphemic or fully graphemic blends (see 6.2). More extensive insertions are found in *s*[*c*]*en*[*t*]*sational*, *s*[*pl*]*urge*, *faceb*[*l*]*o*[*c*]*k*, *stag*[*fl*]*ation*, and *Yi*[*ngl*]*ish*.

The blends *scentsational* and *faceblock* not only exemplify the possibility of several inserted constituents, but also display alternative intercalative structures such as *s*[*cent*]*sational* and *faceb*[*loc*]*k*.

Three blends are identified as intercalative based only on an analysis of phonemic structure: *Navinet* (/nævɪ[nɛ]t/), *Kongfrontation* (/k[ɒ?[ŋ]fɹənteɪʃən/), and *Spinderella* (/s[p]ɪndəɹɛlə/). Graphemic intercalation of structure does not occur in any of their respective source words *navigate*, *confrontation*, and *Cinderella*. There may be phonemic variation between /ɑ/ and /ɒ/ in the first vowel of *Kongfrontation* in the same way as in *confrontation* (cf. Merriam-Webster, n.d.-a), which potentially affects how many phonemes are perceived as intercalated (i.e., /k[ɒŋ]fɹənteɪʃən/ or /kɑ[ŋ]fɹənteɪʃən/).

The opposite pattern of intercalation is found in the three blends *e*[*ggs*]*travaganza*, *e*[*ggs-*]*actly*, and *e*[*gg-*]*ceptional*. They are clearly graphemically intercalative, but it is uncertain whether their phonemic representations qualify as intercalative. The source words *extravaganza*, *exactly*, and *exceptional* are represented with an initial phoneme /ɪ/ in Merriam-Webster (n.d.-a). This representation contrasts with the initial /ɛ/ in the source word *egg*. Thus, although *e*[*ggs-*]*actly* (cf. *exactly*) is an adequate example of intercalation, /[ɛ]gzæktli/ is not (cf. /ɪgzæktli/) as there is no preservation of an initial sound. As noted in 6.2, the phonemic representations may downplay the actual similarity between these three blends and their respective source words. Hence, there is still the possibility that speakers may perceive them as phonemically intercalative.

The discrepancies between graphemic and phonemic intercalation as exemplified here imply that the affordances of source word ambiguity may be unevenly distributed. An analysis of this distribution in the intercalative blends could provide valuable information about the general tendencies of source word ambiguity in the data. However, some kind of quantification is necessary to obtain the required information to conduct such a distributional analysis. To address this issue, the difference between the intercalative blends and the source words that functions as their matrix (e.g., *construction* in *Constraction*) has been quantified as a ratio. For each of the 45 intercalative blends the number of inserted or exchanged constituents (graphemes or phonemes) have been divided by the total number of constituents in the blend. Applying the example *Constraction*, the schematic representation of this division would be [*cd*]/A[*cd*]B. This measure is referred to as the *intercalation* (*IC*) *ratio*, and it is either graphemic or phonemic. The lower the ratio, the higher the degree of similarity between the blend and its matrix source word. This analysis gives a quantitative indication of how close the blend is to its matrix source word, and hence, how plausible it is that the double activation of source words plays a prominent role in the blend. For instance, there are ten graphemes in the blend *ego-system*, while only two of them (the graphemes *g* and -) distinguish *ego-system* from its matrix word *ecosystem*. Dividing the two differing graphemes with the total of 10 graphemes in the blend (2/10) results in a graphemic IC ratio of 0.2. In *Spooktacular* the graphemic IC ratio is 0.25, as expressed by *ook*/*Spooktacular* (3/12). Its phonemic IC ratio is based on the representations /spʊktækjəlɚ/and /spɛktækjəlɚ/. The exchanged /ʊ/ in the blend corresponds to the division of 1 by 11 constituents, which results in the ratio 0.09. Thus, the blend-matrix word relation is potentially more ambiguous in the phonemic representation of *Spooktacular* (IC ratio 0.09) than in its graphemic representation (IC ratio 0.25).

The distribution of the graphemic and phonemic IC ratios are shown in the box plot in Figure 35. The mean and median values indicated in the graph have a detailed level of approximation to enhance the illustration. The mean values of the IC ratios (indicated inside the boxes) differ only slightly between graphemic and phonemic representations (0.194 vs. 0.190). The median values (indicated outside the boxes) differ more (0.174 vs. 0.125). This discrepancy between mean and median values signals a greater diversity in the phonemic realizations, which is also visible in the overall broader range of the phonemic IC ratio.

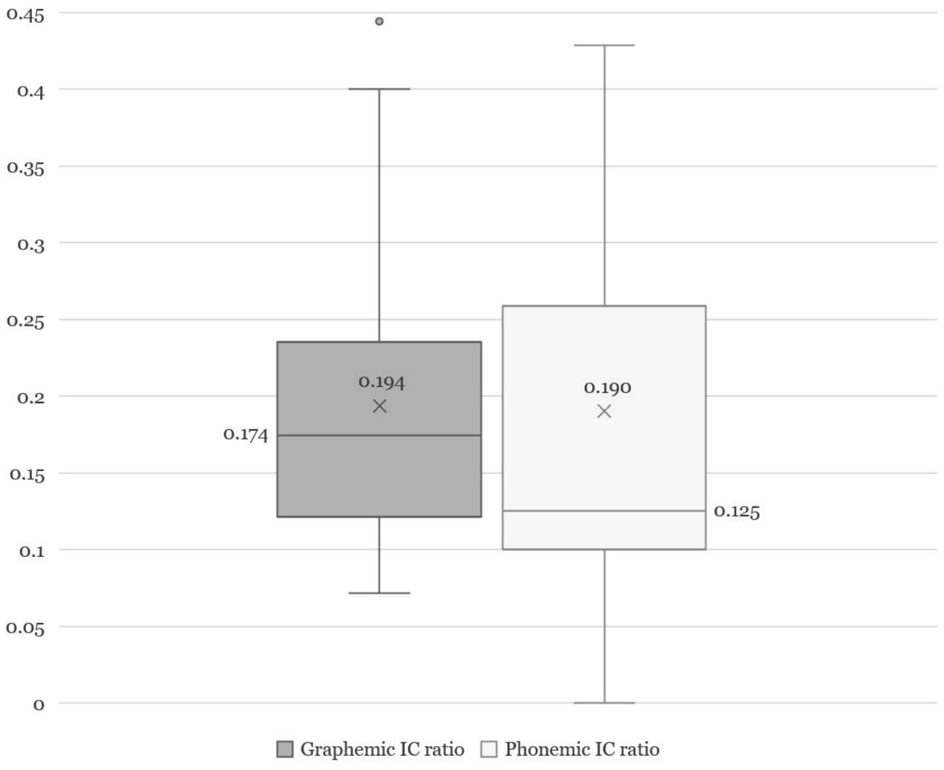


Figure 35. Mean and median graphemic and phonemic IC ratios.

The IC ratio analysis suggests that an optimal level of deviating graphemic structure in the intercalative blends may be centered around slightly less than 20% of the blend. Hence, in these data blends such as *herstorical* (graphemic IC ratio 0.18) and *intellisense* (graphemic IC ratio 0.17) stand out as more prototypical in quantitative terms than *Constraction* (graphemic IC ratio 0.08) and *eco-logical* (graphemic IC ratio 0.09). The lower median value of phonemic intercalation indicates that the phonology of the intercalative blends may afford more opportunities for source word ambiguity. However, the distribution of phonemic IC ratio also suggests more diversity as to the similarity between the blend and its matrix source word.

It should be noted that the IC ratio analysis only concerns quantification of proportions of structure. It is therefore possible that shorter blends may generally have higher IC ratios than might be expected from a qualitative analysis of the impact of source word ambiguity. For instance, the graphemic IC ratio of *glocal* is not particularly low (0.17), and yet only one constituent (*c*) distinguishes *glocal* from *global*. Both source words *global* and *local* are thus clearly recognizable, which makes source word ambiguity a plausible characteristic of *glocal*.

The activation of a source word via its structural similarity with the blend (e.g., *technology – tricknology*) may also be influenced by the absence of frequent lexical neighbors. A wildcard lemma query in the subcorpus of the segment **storical* returned 46 hits, of which all but one (*herstorical*) were based on *historical*. Extending the wildcard query to *h*storical* only resulted in *historical* and *herstorical*. In these queries, the token frequency of *historical* was 47,969 occurrences, while *herstorical* occurred only once. Taken together, 564 tokens of the remaining 44 items (e.g., *ahistorical*, *prehistorical*, and *pseudohistorical*) were found. These queries suggest that the source word *historical* has a very high activation potential in *herstorical* since there are few competing concepts. Hence, it is argued that *herstorical* is highly likely to afford source word ambiguity despite its average graphemic IC ratio (0.18).

The analyses of the blends in the data show that intercalation is relatively frequent, and that it is employed in flexible ways. The source word similarity exploited in these constructions varies both qualitatively and quantitatively. Hence, the relative prominence of intercalation suggests that the ambiguity it connects to may be a significant characteristic in lexical blending. This observation gives reason to believe that the simultaneous activation of source words in lexical blends should not be seen as accidental or a problem of categorization, but instead as a functionally applied resource.

The three types of ambiguity presented so far are predominantly related to the structure of the blends, which makes them primarily overt strategies of ambiguity. Truncation ambiguity concerns the formal possibilities of multiple breakpoints. Mode ambiguity deals with conflicts of activation between graphemic and phonemic representations. Source word ambiguity is realized in the simultaneous activation of two overtly represented concepts. However, ambiguity may also arise from a conflict of activation between an overtly discernible concept and one that is not explicitly expressed. This type of ambiguity is the topic of next section.

6.4 Covert source ambiguity

Source words of blends are commonly recognized through the process of decoding truncated segments (e.g., Lehrer, 1996; Ronneberger-Sibold, 2006; Juhasz et al., 2016). For instance, successful understanding of *digerati* requires the identification of *dig-* and *-erati* as *digital* and *literati* respectively. The fusion of these source words then constitutes the basis for the interpretation of what the blend *digerati* means. Some blends in the data seem to go beyond this basic process of construal, as they may conventionally be associated with an additional covert concept. One notable example is the blend *bodacious* (84). Merriam-Webster (n.d.-a) claims that *bodacious* occurred in print in 1832, and that it is likely formed from the source words *bold* and *audacious*. The dictionary entry also refers to a British dialectal form *boldacious*.[20] The vowel sound in the first syllable (/boʊdeɪʃəs/) also seems to be inherited from the source word *bold*. However, the use of *bodacious* in the corpus shows that it typically designates aspects of the *body*, often with sexual associations. Hence, BODY is seen as a covert source word of *bodacious*, and the ambiguity arising between *bold* and BODY is described here as *covert source ambiguity*. Blends such as *bodacious* characterized by this phenomenon are referred to as *covert source word blends*. In the following examples of such blends, the covert source word is represented with small caps within square brackets.

(84) *bodacious* (*bold* + [BODY] + *audacious*)

Covert source word blends typically exploit an activation potential afforded by structural similarity. In the case of *bodacious*, the American English clipping *bod* derived from *body* is identified graphemically in the initial syllables of the blend. The observed usage pattern of *bod* in the subcorpus is often oriented towards bodily appearance, which is illustrated in examples (85) and (86) from the subcorpus. Thus, the presence of the clipping *bod* in *bodacious* could partly explain the specific semantics of the blend (Professor Raymond W. Gibbs, personal communication).

[20] This etymology can be traced in Pound (1914: 44) as well, where an alternative spelling *bowldacious* occurs. Furthermore, Pound (1914) presents an analogical form *boldrumptious* (bold + rumpus + presumptuous), which also supports the idea that *bold* is the first source word of *bodacious*.

(85) Maybe, when she's trying to think of a reason she married him besides his rockin' bod, she likes to tell herself that he's endearingly goofy (NOW).

(86) Thought summertime fashion was all about a banging bikini bod? (NOW).

The impact of the covert source word BODY is further emphasized by how the blend was identified in the subcorpus. Following the principles of the SP analysis implies that the truncform *bod** was selected because a wildcard query returned the candidate word *body* as the most frequent item (see 4.3.2). The truncform *bod** then generated the list of wordforms in which *bodacious* was found. Given the assumption that the frequency rank of the SP analysis suggests an activation preference, BODY is clearly an available concept to be applied as a covert source word in *bodacious*.

A schematic representation of the function of covert source word blends is illustrated in Figure 36. Two source words (SW1 and SW2) are blended, and the structure of the blend appears to be the result of fusing these two concepts. However, a covert source word ([SW_{cov}]) influences the construal of the blend in ways that are not predictable given the representations of SW1 and SW2. Importantly, this influence depends on structural similarity to allow the simultaneous activation of [SW_{cov}] and the overt source word it resembles.

A hallmark of covert source word blends is the relative semantic dissimilarity of [SW_{cov}] vis-à-vis SW1 and SW2. In *bodacious*, for instance, the source words *bold* and *audacious* have rather generic meaning potentials commonly connected to fearless and challenging behavior or characteristics (e.g., Merriam-Webster, n.d.-a). The covert source word BODY is only indirectly related to *bold* and *audacious*. The influence of BODY (i.e., the [SW_{cov}]) thus narrows the semantic scope of the blend to the more specific patterns of use observed in the data. One of the results of this organization is that the covert source word may be realized as a WHOLE FOR PART metonymy in the blend, i.e., only certain selected components of the conceptual domain BODY are intended in the blend. It should be stressed that the structural similarity between the [SW_{cov}] and an overt source word constitutes an available potential that seems to be exploited only occasionally in the data.

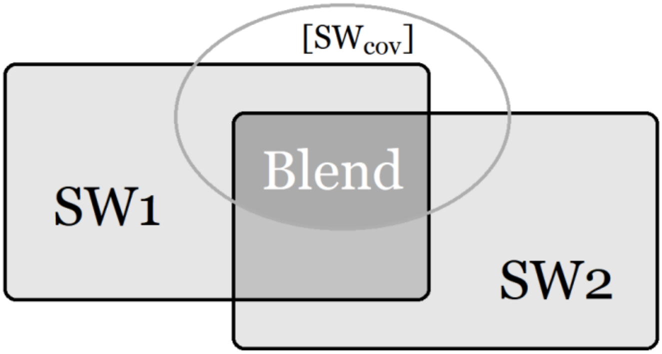


Figure 36. Schematic illustration of the influence of a covert source word.

The idea that secondary and covert sources may have semantic implications for blends has been stressed by, for instance, Veale and Butnariu (2010), who exemplify with the blend °*Chimerica*. They describe this blend as a coinage intended to capture the complex relationship between the US and China, but also point to the potential secondary meaning residing in the (graphemic) similarity to the name of the Greek mythological monster *Chimera* (Veale & Butnariu, 2010: 403). Merriam-Webster (n.d.-a) reports several uses of *chimera*, ranging from a direct reference to the mythological monster, a schematic concept of a compound monster, a symbol of unattainable goals, and a term in medical surgery. There is thus a rich figurative potential in the concept, which is exploited in Veale and Butnariu's (2010) example.

In quantitative terms, clear examples of covert source word blends are rare in the data, although covert source ambiguity seems to be at work in less prototypical instantiations as well. The serially related blends *Compuware*, *Computrace*, and *Computrainers* (87–89) are instances suggesting the activation of covert source words. They are all related to activities involving computers.

(87) *Compuware* (*computer* + [COMPANY] + *software*)

(88) *Computrace* (*computer* + [COMPANY] + *trace*)

(89) *Computrainers* (*computer* + [COMPANY] + *trainers*)

The segment *Compu-* is identified as a truncation of *computer* in these three blends. In *Compuware* (87) the source word *computer* is even suggested by the final segment *-ware*, which likely originates from either of the related concepts software or hardware (or both) (cf. Lehrer, 1996, 2007). However, detailed contextual analyses of examples (87–89) reveal strong associations with corporate business or company management. These analyses suggest that COMPANY functions as a covert source word activated by the initial cluster *comp-*.

There is also frequency-based support for the assumed covert influence of COMPANY. Examples (87–89) were found in the truncform queries based on the CaW *company*. The SP truncation from the right of *company* yielded *comp** as its truncform. Hence, the most frequent wordform in the query output in which examples (87–89) were found was *company*. Table 18 lists the frequency measures of the CaW and the three blends. These measures suggest that the activation potential of *company* is high in the blends. It is thus assumed that there is a covert source ambiguity established between *computer* and COMPANY, which is interpreted as a part of the semantic structure of these blends.

Table 18. Frequency measures of *Compuware*, *Computrace*, and *Computrainers*.

Wordform	***Frequency***	***Frequency rank***
company (CaW)	841,424	1
Compuware	67	162
Computrace	2	585
Computrainers	1	703

The lexicalized blend *splurge* (90) constitutes another example of a relatively prototypical covert source word blend. It parallels *bodacious* in several respects, although its etymology appears somewhat less clear (Barnhart & Steinmetz, 1988; Merriam-Webster, n.d.-a). *Splash* and *surge* are generic expressions with a rather broad semantic potential. Contextual analyses of the subcorpus as exemplified in (91–92) show that *splurge* commonly refers to the narrower concept of some kind of indulgence, typically in relation to money. Merriam-Webster (n.d.-a) makes an explicit reference to *spending*. The covert source word of *splurge* is thus assumed to be SPEND. The similarity between the initial consonant clusters *sp* of *splurge* and *spend* is thus exploited, and the narrower usage patterns cannot be predicted by the overt source words only.

(90) *splurge* (*splash* + [SPEND] + *surge*)

(91) I've sampled a little and have enjoyed it. some day soon i'll splurge a little on a more expensive bottle of it (NOW).

(92) Yes, I know sushi is one of the foods that everyone in LA feels is okay to splurge on, but I'm not one to spend more than $100 on sushi on any given occasion (NOW).

The frequency rank measures of *splurge* resemble other covert source word blends. The covert source word SPEND is also the CaW *spend* responsible for the retrieval of the blend *splurge*. The SP analysis of *spend* yielded *sp** as the truncform, thus any wordform in the subcorpus beginning with *sp*- has an activation potential of the conceptual domain SPEND.

The covert source word blends presented so far (examples 84, 87–90) are considered rather prototypical instantiations of covert source ambiguity. The blend list contains items that display less apparent – while yet noticeable – signs of employing this mechanism. Two such cases are *Condomania* (93) and *sneakernet* (94). *Condomania* (93) has initially been interpreted as a blend of *condom* and *mania*. However, a wildcard lemma query of the segment *-ania* revealed a large number of neighboring potential source word candidates in the subcorpus. Regional concepts stood out as the clearly dominating reference in this query. Accordingly, a regional association of the segment *-ania* could be a covert source word in *Condomania*.

(93) *Condomania* (condom + [-ANIA] + mania)

A preliminary semantic domain analysis of the *-ania* query shows that the 162 retrieved types are associated with one of the six semantic domains POLITICAL REGION, PROPER NOUN, OBSESSIVE BEHAVIOR, OTHER REGION, BIOLOGY, and GENERIC.[21] Table 19 presents the type frequency distribution of the output data, and the total token frequency of the blends pertaining to the particular semantic domain. The token frequency gives a rough indication of the prominence of the semantic domain associated with terms including the segment *-ania*. The frequency data in the table indicate that there could be an influence of concepts such as REGION or REALM in the decoding of wordforms including the segment *-ania*. POLITICAL REGION and OTHER REGION taken together constitute 54,585 tokens of the total of 64,568 returned hits. In other words, a concept that is not obviously associated with one of the semantic domains in Table 19 has a high probability to be influenced by the spatial notion of REGION. The blend *Condomania*, which is the name of a world-wide retailer of condoms and other sex related items, may be thought to express a manic interest in a certain category of objects. However, the dominance in terms of token frequency of the spatial meaning of *-ania* may yet have an influence on the blended concept.

[21] These semantic domains are not USAS categories but are preliminary conceptual domains used to capture the specific neighborhood patterns of *-mania*.

Table 19. Wildcard query of the segment *-ania*.

Semantic domain	***Type frequency***	***Examples***	***Token frequency of blends pertaining to the semantic domain***
POLITICAL REGION	62	Pennsylvania, Albania	54,046
PROPER NOUN	48	Shania, Vania, Xania	5,983
OBSESSIVE BEHAVIOR	39	megalomania, anglomania, fernandomania	3,827
OTHER REGION	6	Mundania, Ruritania, Arcania	539
BIOLOGY	5	leishmania, prytania, iguania	90
GENERIC	2	compania, crania (pl. form of cranium)	83
Total	*162*		*64,568*

Although the examination of the lexical neighborhood shows a potential for covert source ambiguity, it still seems somewhat unclear whether this mechanism is at work in *Condomania* since the context contributes little to the analysis. What makes *Condomania* a possible instance of covert source ambiguity is the potential impact of a whole group of items pertaining to the domain of regional concepts.

The structure of the blend *sneakernet* (94) differs slightly from the other examples of covert source word ambiguity in that both *sneaker* and its assumed covert source word SNEAK are fully discernible in the blend. *Sneakernet* is instead motivated as a covert source word blend based on contextual analyses, but it appears all the same less prototypical than previously described examples. Structurally related blends such as *Navinet* and *Execunet* (see 5.3.4) indicate by analogy that the segment *-net* may suggest a plausible breakpoint. The first source word would then be interpreted as *sneaker*, which in this blend typically refers to a type of shoe (cf. Ross, 2009: 4). The alternative breakpoint *sneak-* implies the segment *-ernet*, which is a possible truncation of either of the source words *Ethernet* or *Internet*. A structural analysis in isolation cannot resolve this ambiguity of truncation (cf. 6.1).

(94) *sneakernet* (*sneaker* + [?SNEAK] + *Ethernet*/*Internet*)

The meaning of *sneakernet* is described in the Cambridge Dictionary online as an expression referring to “electronic information being moved physically from one computer to another on a CD, disk, etc” (Cambridge Dictionary, n.d.; see also Ross, 2009). The usage patterns in the subcorpus largely accord with this definition, as exemplified in example (95).

(95) having full USB is really important for users in developing countries where sneakernet, the practice of sharing files by trading USB keys is more common than Internet access (NOW).

The meaning potential that could motivate the covert source word SNEAK is detected in instances such as (96) and (97). Both these examples from the subcorpus highlight the relation between *sneak* and *sneaker*, which evokes associations to stealth and furtiveness (see Merriam-Webster, n.d.-a on *sneaker*). Example (96) illustrates the topic of illegal file sharing in the use of expressions such as *fear of liability*. In this particular example, motivations for using *sneakernet* to exchange information are addressed. A risk of being detected when unlawfully sharing copyrighted material is downplayed but appears to be a possible motive for secrecy.

(96) “Teenagers and twenty-somethings I know routinely will go over to a friend's house with a terabyte drive to swap stuff,” he said. They choose the “sneakernet” approach less out of fear of liability than because it's so convenient (NOW).

In example (97) the prompt action indicated by *whisked the data out of the country* seems more associated to conflicting interests between the organizations. The choice of wording indicates that the FBI has acted despite an awareness of the dubious support of the New Zealand authorities. Using *sneakernet* as a part of the description may enhance the furtive character of the actions.

(97) the FBI sent material back to the States without the New Zealand police having any say in the matter and that the police force had effectively lost control over it. # The NZ government doesn't deny the FBI whisked the data out of the country via sneakernet (NOW)

Examples (95-97) show that the blend *sneakernet* may inherit both meanings of stealthily movement and the specific shoe type *sneaker*. This is all the more apparent as the metonymic relation is already established through the potentially silent walking with *sneakers* and the notion of stealth in *sneak*.

The frequencies of *sneak* and *sneaker* in the subcorpus suggests that the activation potential of *sneak* in *sneaker* is considerable. Table 20 lists the lexical neighbors of *sneakernet* returned in the truncform query of *sn**. The pattern in the table resembles the ones of *splurge* (*sp*–spend*) and *bodacious* (*bod*–body*), although the candidate word *snow* is the most frequent item in the truncform query of *sn**. *Sneak* is found to be the fifth most frequent wordform in the query output, and the 11,000 occurrences attest to its conventional status.

The structural similarity and the related semantics of *sneak* and *sneaker* make the frequency data less clear as regards the implications of SNEAK as a covert source word. As argued in relation to the schematic presentation of covert source ambiguity (Figure 36), the covert source word should be semantically different from either of the overt source words. The dissimilarity between *sneaker* and SNEAK seems less apparent than in *bold*–BODY and *splash*–SPEND. Hence, *sneakernet* is here considered as a less prototypical instantiation of covert source ambiguity.

Table 20. Frequency measures of *sneakernet*.

Wordform	***Frequency***	***Frequency rank***
snow	73,227	1
sneak	11,000	5
sneakers	5,776	9
sneaky	1,472	27
sneakily	189	103
sneaking	171	108
sneaker	150	116
sneakerheads	79	150
sneakerhead	53	174
sneakiest	36	203
sneakiness	26	224
sneak-peek	26	227
sneakernet	16	279

The last example in the current section of a blend possibly influenced by covert source ambiguity is the borderline case *snoratorium* (see 5.1.3). In (98) a suggestion of a covert source has been added.

(98) *snoratorium* (*snore* + [MORATORIUM] + lat. *dormitorium*/*oratorium*)

As the analyses in 5.1.3 show, the origin of the second source word in *snoratorium* is difficult to determine. The many available possibilities contribute to the operationalized ambiguity in the blend. It could be seen as a spatial concept resulting from the combination of the truncated segment *snor-* and the locative -(*a*)*torium* (example 99).

(99) Sending her husband off to a sound-proof “snoratorium” doesn't seem to [...] The couple allegedly joke with friends about sleeping in separate bedrooms (NOW).

Snoratorium could also be rephrased in a more abstract way as *snore moratorium* (cf. Waring, n.d.). The activation potential of *moratorium* is assumed to be notable, given its token frequency dominance in the wildcard queries of the segment *-atorium*. Despite this frequency pattern, the contextual sources in the subcorpus are scarce, hence the status of *snoratorium* as a covert source word blend remains somewhat unclear.

Before summing up on covert source ambiguity, the frequency measures based on SP deserve a comment. The SP analysis has the capacity to reveal quantitative tendencies in the data, but it should be used with caution. In the examination of covert source ambiguity the frequency measures used in the detection of SP serve mainly as a complement to contextual analyses of covert source words. As an illustration, wild card queries based on *add-* and *sing-* return wordforms that could function as covert source words to the blends *Addaction* and *Singlish*, such as *add*, *address*, *single*, and *sing*. The frequency ranks of these potential covert source words are high (the two most frequent lemmas of both queries). However, there are no contextual indications that *Addaction* should be influenced by *add* or *address*, or that *Singlish* would be covertly connected to *single or sing*. Hence, covert source ambiguity is not identified in *Addaction* and *Singlish*. The activation potential indicated by a high frequency ranking provides an affordance that may be applied by the users of a language, but there typically needs to be other types of indication as well, notably provided by the context.

The dependence on contextual analyses to identify possible covert source words highlights the affordances of a CL approach. Ronneberger-Sibold (2006) and Fandrych (2008) have already pointed out that generative frameworks have shown limited interest in blending as this theoretical orientation has framed blends rather as a performance phenomenon largely irrelevant for the study of "linguistic competence" (Ronneberger-Sibold, 2006: 159). From a generative point of view, it could be argued that the contextual observations and the frequency patterns of, for instance, *bodacious*, are merely coincidences or random pragmatic patterns having little (if any) import on the structure and meaning potential of a blend. This study takes a different position. It is suggested that the semantic potential of *bodacious* is closely related to how it is used (cf. Schmid, 2020). The observation of several items displaying similar characteristics is interpreted as an indication of a cognitive force realized by a lexical activation potential, which influences the meaning of the blend.

Covert source ambiguity relies on the basic (but not uncontroversial) assumption that word meanings are dynamic, encyclopedic, and closely related to conceptual construal and interaction in context (Taylor, 2003; Geeraerts, 2006a; Wierzbicka, 2006; Langacker, 2008; Svanlund, 2009; Bergen, 2012; Schmid, 2015; 2020 to name a few). According to this view, semantic structure is constantly elaborated on all potential levels of communication, and all available resources to construe meaning are potentially acknowledged. This leads us to consider the specific lexical conditions for covert source ambiguity.

The nature of the covert constituents is a central issue of covert source ambiguity. They need not necessarily be consciously profiled in the semantic structure of a blend. For instance, *sp* in *splurge* does not seem to have any formal connection to the (covert) concept of SPEND, but is instead suggested in Merriam-Webster (n.d.-a) and Barnhart and Steinmetz (1988) to originate from the (overt) source word *splash*. Similarly, the segment *bod* in *bodacious*, is not specifically related to the concept BODY, but is commonly seen as a truncation of *bold*.[22] However, there are several indications in the data suggesting that connections of the type SPEND–*splurge* and BODY–*bodacious* might be at work (see also Veale & Butnariu, 2010). The concepts SPEND and BODY thus remain covert potentials in the respective blends. However, their dependence on lexical neighborhood (e.g., *bod–body–bold*) makes them more specific than any general association that may be directly part of the blends *splurge* and *bodacious*. The covert reference seems to draw on lexical neighborhood effects, which have also been shown to have facilitatory effects on lexical retrieval in English (Andrews, 1997).

Blends have often been studied from the perspective of their complex formal characteristics, which is perhaps connected to the abundance of suggested types and classes in previous research. The insights into the limitations of formal classification (e.g., Kemmer, 2003; Gries, 2006) may be partly responsible for the increasing interest in broad approaches to blend research in the last decades. Covert source ambiguity aligns with this development in that it can be described

[22] The metaphorical use of expressions such as *bold move* is not relevant in this case, although a *bold move* may be referred to bodily experience ("choosing this as the debut single is a bold move" (NOW)).

as a mechanism operating on a level typically unavailable for analyses of form only. Hence, it contributes to the growing body of knowledge attesting to the need for multiple levels of analysis to describe lexical blending in adequate ways.

6.5 Ambiguity as a functional characteristic of lexical blends

This section sums up and comments on the findings concerned with ambiguity in the blend data. It starts with a discussion on the specific observations and implications of ambiguity (6.5.1) and moves on to discuss ambiguity from a more general epistemological perspective with relevance for understanding its role in blend research (6.5.2). The final subsection presents a cognitive model to accommodate ambiguity in lexical blending (6.5.3).

In brief, four main types of ambiguity have been detected in the data:

- *Truncation ambiguity*, which is oriented towards the formal fragmentation of the source words and the potential ambiguity emerging when the truncated segments are amalgamated in a blend.

- *Mode ambiguity*, which is realized by the elaboration of modes of expression, typically evoking conceptual conflicts.

- *Source word ambiguity*, which employs a potential processing conflict arising from the similarity of the source words and the blend.

- *Covert source ambiguity*, which is based on usage patterns connected to an influence of lexical neighborhood.

These mechanisms constitute parts of a suggested model to account for how ambiguity is realized in the blend data. The impact of the different types of ambiguity does not seem accidental but has been shown to be systematic. Ambiguity may therefore be seen as a characteristic with considerable importance in the formation of lexical blends.

The four observed ambiguities differ as to the type of conceptual organization they profile. Truncation ambiguity elaborates the formal, typically graphemic, structure. Mode ambiguity and source word ambiguity is realized in an intricate play with structure and meaning construal. Covert source ambiguity draws more on semantic-pragmatic factors influencing the construal of a blend, but the impact of lexical structure is also a component in the realization of the covert source strategy. These strategies set lexical blending apart from other word formation processes such as complex clipping, compounding and derivation.

The observations of ambiguity reported here may be a relevant factor for criterion v) in the preliminary definition, which states that wordplay could be involved in lexical blending (3.1). Although this criterion is commonly emphasized in blend research, few studies have systematically operationalized wordplay with the intention to examine how it is realized. Hence, wordplay remains to date largely a problem of definition. This is reflected in the current investigation in that the operational definition makes no a priori assumption about wordplay: i.e., it is not included as a criterion of what counts as a blend (see 3.5). However, the four detected ambiguities in this investigation could offer a tool for the operationalization of wordplay, both qualitatively and quantitatively. There may be other factors that influence the realization of wordplay (cf. Renner, 2015), but the four suggested ambiguities in this chapter might contribute to more precise definitions of wordplay and perhaps even provide a basis for distributional analyses.

6.5.1 Ambiguity at work

Previous studies have shown that unresolved and imprecise meaning is an intrinsic component of blending (e.g., Kelly, 1998; Ronneberger-Sibold, 2006; Renner, 2015). This component has been shown to be exploited systematically in various ways. For instance, Ronneberger-Sibold (2006, 2012) demonstrates how socio-pragmatic concerns motivate the elaboration of transparency. Renner (2015) shows how wordplay can be analyzed as a systematically and functionally employed strategy. Both these approaches exemplify strategies that connect truncation with ambiguity. The transparency of blends is intentionally challenged by the use of ambiguous truncation patterns (cf. Gries, 2012), while wordplay makes use of ambiguity to create humorous associations (cf. Brône & Feyaerts, 2003; Renner, 2015). These observations are expanded in here investigation in that ambiguity is assumed to be used in other ways as well, as outlined in the introduction to 6.5.

As suggested in 6.5, the four types of ambiguity registered here are informative in and of themselves, but they also attest to the influence of several types of cognitive factors involved in blending. In this sense, they provide additional information attesting to the claim of CL that "[l]inguistic meaning [...] involves knowledge of the world that is integrated with our other cognitive capacities" (Geeraerts, 2006a: 5). To highlight the gradual distribution from traditional linguistic functions to general cognitive processes, the four types of ambiguity are presented below as consecutive steps. This depiction is a schematic simplification, since the described processes are both intertwined and intrinsically complex, but it is a first step towards disentangling the main characteristics of ambiguity as it is realized in the data.

First, the fusion of form involving overlapping structure in blends such as *technovation* and *netiquette* may be described as a first level of ambiguity. In these blends, truncation is seen as the operator of ambiguity, while overlap is the consequence of the ambiguous formal structure, and the mechanism is therefore called truncation ambiguity. This type of ambiguity exploits the formal structure, especially in A|BC|D blends such as *netiquette*, as a means to achieve an attention-catching effect. Structural similarity, be it phonemic or graphemic, is a key factor in this process of playing with truncation patterns.

Second, mode ambiguity employs a psycholinguistically oriented lexical activation conflict established on a representation level, where the employed modes evoke conflicting concepts. However, not only prototypical graphemic blends such as *Scitation* are examples of operationalized mode ambiguity, but constructions such as *Kongfrontation* also exemplify this phenomenon, albeit to a lower degree.

Third, intercalative blends such as *Dragula* create an ambiguity residing in both structural and conceptual properties of the source words, hence the chosen term source word ambiguity. The initial segment *dra-* potentially activates two source domains simultaneously, and, importantly, there is nothing in the blend that provides unambiguous information on whether any one of them is to be considered the preferred interpretation (cf. Algeo, 1977 on *indefinite composites*). The ambiguity of such constructions seems to be highly marked as it employs several resources simultaneously. For instance, interpreting *Dragula* as the result of replacing the original *c* with *g* implies a focus on the literary figure *Dracula*. In contrast, if the initial segment *drag-* in *Dragula* is emphasized, the transgender concept *drag* (*queen*) or the computer related term *drag-and-drop* may instead be foregrounded.[23]

Fourth, there are indications in the data that source words can take part in the construal of a blend without being overtly represented in the formal structure of the blend. The ambiguity expressed in such blends, referred to as covert source ambiguity, stands out in this account as a phenomenon relying on partly different resources, perhaps most closely associated with lexical activation in parallel with mode ambiguity, but in more intricate ways. In a sense, this final level of analysis marks an endpoint in this account as it seems to rely less on the overt structural properties of a blend.

Finally, similarity in lexical blends has been explored in several studies in the last decades (e.g., Kelly, 1998; Gries, 2004a, 2006; Wulff & Gries, 2019). Such investigations have typically foregrounded its gradual nature and probabilistic distributions. Since ambiguity typically draws on similarity to give rise to

[23] I this case, the convention to capitalize proper nouns may perhaps influence the decoding in favor of *Dracula* as the intended source word.

conflicting interpretations, it comes as no surprise that ambiguity at work is also a matter of degree. Towards one end of the scale, the parallel layering of identical structures – whether driven by orthography, phonology, prosody, or any of these in combination – can even be thought of as realization of full similarity. Thus, the explanatory potential of ambiguity as outlined in this chapter closely connects to research on similarity in lexical blending: it is suggested that the observed ambiguities profile various aspects of meaning construal afforded by structural similarity. Where there is similarity, ambiguity may be exploited for communicative purposes.

These findings align with the symbolic claim of CL, which holds that all strategies in language are understood in the light of their symbolic functions (Langacker, 2008: 15). It has been shown in chapter 6 that ambiguity is not a randomly occurring phenomenon but instead a systematic characteristic of blend formation. Therefore, it is suggested that ambiguity in blending has symbolic potential on several levels and is significant enough to be understood as one of the cognitive forces motivating the existence of blends. In brief, ambiguity is inherently meaningful in blending.

6.5.2 The epistemology of ambiguity

In this subsection a more general epistemological perspective is taken, which is important in the outline of a model based on ambiguity as a functionally applied mechanism. This perspective connects to the CL approach to meaning in general and to blend research in specific.

From a sociologic and psychologic viewpoint, Bauman (1991, 2001) discusses the problems of unresolved symbols and their meaning in a Western philosophical tradition. He argues that the drive to avoid all kinds of uncertainty has been a recurrent and strong tendency for several centuries (see also McMahan & Evans, 2018). Bauman (2001) highlights the despair of

> the sceptics of the early modern era [who were] appalled and frightened out of their wits by their discovery and swore never to rest until the rock-steady foundations of clear and unambiguous knowledge of the world were found or constructed (Bauman, 2001: 59).

Further in the same essay, Bauman (2001) touches on the domains of language and semantics by claiming that the modern project, both from a historical and contemporary perspective, is preoccupied with

> a perfect, one-to-one fitting of names and things, words and meanings [...] a world in which there is an unambiguous [...] recipe for every situation and no situation without a recipe attached (Bauman, 2001: 65).

In the light of this epistemological tradition it is easy to understand a reluctance towards ambiguity and ambivalence in linguistic investigation as well. Oaks (2010: 3) captures this view by stating that "ambiguity represents a problem, a phenomenon that must be avoided if possible and resolved when it does occur" (see also Crystal, 2001b).

The importance of notions such as clarity, coherence, and definiteness can hardly be overestimated in the development of scientific knowledge, yet this attitude comes with inherent risks. As the CL enterprise stresses, language is typically not the ordered, static, and ideal construct that would naturally fit into such descriptions. Geeraerts (2006a) criticizes theoretical standpoints that view language ”as a more or less rigid and stable structure – a tendency that is quite outspoken in twentieth century linguistics.” (Geeraerts, 2006a: 4). Although elegance and minimalism are important and productive epistemological ideals, the complexity of the study object may conflict with too strong inclinations towards idealized models (cf. Hanks, 2000: 208).

In accounts of lexical blending, the issues of model design and possible consequences of too strict analytic reductionism have been recurring topics (Lehrer, 1996; Kemmer, 2003; Fandrych, 2008; Gries, 2012). At least two main lines of reasoning seem to be at hand. One is to consider blends as instantiations of creativity, slang, and deviation, which also often means that they are seen as marginal and not significant to the understanding of language. Another view is to acknowledge their complex nature, and to address their probabilistically distributed characteristics with the assumption that they are potentially informative in a wider linguistic sense as well (cf. Dressler, 2000; Gries, 2012). Both these types of (seemingly conflicting) attitudes pose epistemological and methodological challenges. Excluding blends from an account of regular word formation, also means that attestedly striking and effective ways of verbal communication are largely left unexplored. Including them, on the other hand, comes with considerable difficulties to explain their diversity (e.g., Beliaeva, 2019).

Contemporary blend research has reported ambiguity realized in structural overlap, source word origin, and semantic potential, yet little effort has been made to investigate the functional import of ambiguity in blending in systematic ways (but see for instance Renner, 2015). Although ambiguity is typically avoided in most domains of the western philosophical tradition, it is here shown that language users within the investigated context apply ambiguity in lexical blending in various creative ways.

A consequence of approaching blends from the perspective of ambiguity is that the well attested classificatory problems (e.g., López Rúa, 2004; Bauer, 2012) are inevitably included (cf. Hanks, 2000; Beliaeva, 2019). From the perspective of traditional classification, this may strike as a strategy undermining the objective to explain the systematicity of the object (if indeed this is feasible). As shown in this investigation, practically all observations of characteristics signal gradual distribution, but it does not stop there. The distribution of characteristics seems to overlap in complex ways, sometimes showing signs of systematic mechanisms, and sometimes not. The question is then whether it is a viable option to include ambiguity as a cognitive force in a theory of lexical blending. The answer to this question is that it depends on what is expected from the classification in this model. The fact that vagueness and complexity are incorporated is not a legitimate reason to reject the theory, as long as it effectively describes and explains the investigated object (cf. Wierzbicka, 1985). Blends have long been described as irregular and random. However, blend research in the last decades has shown that there is some degree of regularity in blending, and that the characteristics of blends tend to be gradually distributed. Therefore, it is suggested that attempts to classify lexical blends should direct the attention towards the cognitive mechanisms at work in individual blends as well as in groups of items, instead of trying to group occurrences of blends in strictly delimited container categories. In fact, not even the recognition of fuzzy conceptual categories may automatically resolve the issue of blend classification: overlapping characteristics and sometimes seemingly arbitrary patterns of distribution constitute significant obstacles to clear-cut categorization.

The suggestion to incorporate ambiguity in a theory of blending calls for an alternative approach to classification. Such an approach will have to be stringent, while at the same time need to accommodate and account for the complexities of blends. The observations in this investigation have highlighted two general tendencies, which may constitute the basis for a suggested model for classification. The first tendency is the relative uniformity of the blends. As a basis for blend identification, the rather clear-cut operational definition (3.5) resulted in a remarkable quantitative precision when comparing the three datasets (see 5.1.1). This accuracy shows that the basic criteria of the blend identification constitute a rather reliable means to describe the blends. Furthermore, the analyses of distribution of truncation patterns show that the four standard types (ABD, ACD, AD, and A|BC|D) cover most of the blend data. The alternative assessments do not influence this observation, as they are typically of one of the standard types. Recurring issues such as blends with more than two source words (as regards formal structure) and nonlinear patterns of amalgamation may be qualitatively accurate, but in the data of the current investigation they remain rather rare. The second tendency concerns the large set of mechanisms and constraints that interplay in complex ways. For instance, it has been shown that figurativity and iconicity connect to structural attributes, although their distributions are intricate and typically a matter of degree (5.3). The workings of the ambiguities described in chapter 6 constitute other examples of affordances that are occasionally used but are not always quantitatively prominent.

These two tendencies can be thought of as two levels of organization in lexical blending. The first level is described in terms of a basic schema. This schema functions as a conventionalized endoskeletal structure, which allows considerable modification by adding a range of various mechanisms. The operational definition in 3.5 or similar definitions of blends are typically manifestations of this first basic level. The second level is constituted by additional mechanisms. They may be used independently of each other, but they may also interact in patterns resembling complementary distribution. The second level is highly complex and consists of many different available strategies of formation. Examples of second level mechanisms are domain proximity, iconicity, intercalation, and graphemic dependence. Chapters 5 and 6 deal with several second level mechanisms, but it is assumed here that a list of possible mechanisms is open-ended in the suggested model. This double organization is called a *dual model of blend classification*.

Figure 37 illustrates schematically how the two levels of organization in the dual model could be understood. The first level is represented as the dark structure called *basic blend schema*. The small gray rounded squares are examples of second level mechanisms. The gray line between graphemic dependence and mode ambiguity indicates the inherent relation between these two characteristics. Empty slots in the basic blend schema mark that further additional characteristics are afforded, but their number is open-ended in the figure. Hence, the formation of lexical blends involves many possibilities available through the affordances offered by the two levels of this model.

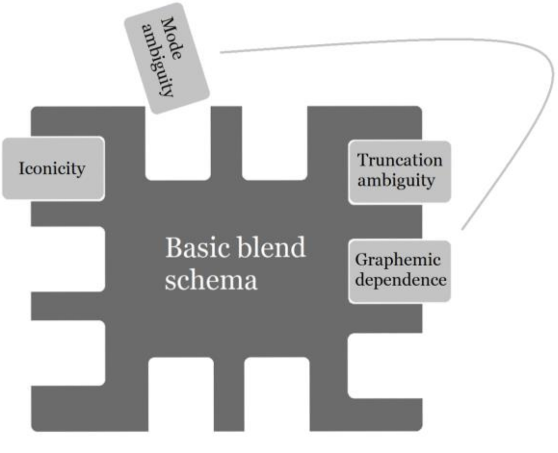


Figure 37. Schematic illustration of the dual model of blend classification.

The dual model has the capacity to account for the basic attributes of lexical blending, while it also deals with its inherent complexities. The fact that it is dual implies that the different levels can be addressed separately and need not automatically cause taxonomic conflicts. Finally, it should be noted that this perspective on blend classification presupposes a prototype theory perspective (cf. Taylor, 2003) acknowledging family resemblance as a concept capturing the taxonomic complexity of lexical blends.

6.5.3 Multistable meaning construal

It may seem challenging to account systematically for ambiguity in blending, but ambiguity is a recurring issue that has been investigated in other fields of research. Kornmeier et al. (2009) mentions the long tradition of studying ambiguity arising from perceptual instability. Phenomena concerned with the ways the brain manages and perceives several simultaneous sensory stimuli are given particular attention in the present study. They are commonly referred to as *multistable perception* or *multistability*.

Rubin (2003) explains multistability with the more specific observations of *binocular rivalry* (see also Levelt, 1968) saying that “a particular stimulus gives rise to two different interpretations that alternate over time” (Rubin, 2003: 289). Important characteristics “common to all forms of multistable alternation [are] *exclusivity*, *inevitability* and *randomness*” (Leopold & Logothetis, 1999: 260). *Exclusivity* captures the observation that perceptive interpretations are only available once at a time. *Inevitability* captures the somewhat controversial assumption that the alteration of stimuli dominance is largely automatic and obligatory (see Kornmeier et al., 2009 for a detailed discussion). *Randomness* highlights the idea that the alterations follow no principled pattern, which has also been the topic for debate (e.g., Leopold & Logothetis, 1999).

Multistability is often visualized with the famous *face-vase* representation of binocular rivalry (see Figure 38), but there are numerous examples in the literature illustrating the phenomenon (see for instance Levelt, 1968; Leopold & Logothetis, 1999). These visualizations share the characteristic that they are intrinsically ambiguous and require a focus on one visual interpretation at a time, hence there is a conflict of interpretation referred to as *rivalry*.

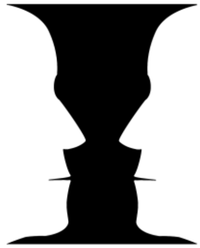


Figure 38. Multistable visual perception; the face-vase illustration of binocular rivalry.

The alternating foci of binocular rivalry have been employed in CL accounts to explain the notion of perceptual prominence and figure-ground reversal (e.g., Ungerer & Schmid, 2006: 164). Here, multistability – as exemplified with binocular rivalry – is applied to visualize how ambiguity may function in lexical blends. However, the possibility to connect ambiguity in blending to binocular rivalry also demonstrates the interconnectedness between a general cognitive mechanism and the processing of language.

The idea of multistable perception is developed in the context of blending to capture a specialized semantic phenomenon called *multistable meaning construal*. The term *meaning construal* refers to the dynamic character of meaning involving the influence of perspectival, encyclopedic, and multimodal resources (e.g., Langacker, 2002; Geeraerts, 2006a; Wierzbicka, 2006). The modifier *multistable* is intended to capture the idea of simultaneously activated conceptual domains, including structural categories understood as instantiations of schematic blend structures.

Multistable meaning construal can be applied to several of the central findings presented in chapters 5 and 6. In relation to truncation ambiguity, it resolves the question of whether A|BC|D blends display truncation or not. From the perspective of multistable meaning construal, this structure involves truncation as well, but the truncation pattern functions much like binocular rivalry. For instance, if *hip-hop* is profiled in *hip-hopera*, then *-era* functions as a truncated segment, while foregrounding *opera* makes *hip-h-* the truncated segment. The same rationale is applicable to blends that resemble A|BC|D blends, but are indisputably truncated, albeit minimally, such as *californication*. Establishing an unambiguous breakpoint is not possible in this blend, which illustrates that it may possibly be schematically closer to the A|BC|D structure, although the full preservation of the second source word *fornication* would result in the annotation ACD. Furthermore, an intercalative blend such as *glocal*, suggested to display source word ambiguity, is another instance of multistable meaning construal. In such constructions a part of the blend potentially activates two source words at the same time. In *glocal* this is exemplified by the combined segments *-lo-* and *-al*, which may activate both *global* and *local*. The multistability of *Websight* is manifested in the conflicting activations of the source word *website* (via the phonemic realization of the blend) and the graphemic reference to *sight*.

In sum, considering multistable meaning construal is potentially valuable in the analysis of patterns of ambiguity in blending. It has been shown that ambiguity is realized systematically in the data, since it comes with a considerable meaning potential. Multistable meaning construal subsumes both the broad CL tenets on meaning and the specific blend characteristics employing ambiguity on various levels. The connections to the extensive research on multistable perception in other fields may also offer possibilities to strengthen the empirical basis further.

6.6 Summary of ambiguities in blending

Chapter 6 has outlined the findings on ambiguity as a systematic and functionally applied characteristic of lexical blends. Four specific types of ambiguity have been identified in the blend data. These types of ambiguity operate on different levels of conceptual organization. They are typically established on the basis of various types of similarity between source words (notably graphemic, phonemic, and prosodic similarity), which previous research has shown to be a key attribute of blends.

The findings have then been discussed on a theoretical as well as an empirical basis, including epistemological concerns. The problem of categorization has been addressed by suggesting an alternative approach called the dual model of blend classification. This model offers a possibility to combine the seemingly straightforward attributes of blends and the complex characteristics that are also at work.

Ambiguity in blending has also been connected to other aspects of cognition, particularly as a special case of multistability labelled as multistable meaning construal.

7 Conclusion

The development of blend research during the 20th and the beginning of the 21st century presented in chapter 3 can be described in terms of a journey. The departure takes place in studies registering this peculiar linguistic phenomenon, often laying bare a prescriptive attitude that marginalizes its role in language (Bergström, 1906; Pound, 1914). The next stop – a structuralist influenced classificatory approach – is reached via a winding road of descriptions of attested forms as well as accounts treating it as a speech error phenomenon (Berrey, 1939; Hockett, 1973; Bryant, 1974). The classificatory studies establish a terminological network forming a preliminary fundament on which a theory of blending can be modeled (Algeo, 1977; Cannon, 1986). However, these studies reveal that blends are diverse to the extent that they seem to evade systematic description. The intriguing nature of blending inspires a continuation of the journey despite these discouraging observations, and it takes a somewhat new turn towards the end of the 20th and the beginning of the 21st century when blends are investigated as a phenomenon with wider cognitive and socio-pragmatic implications (Lehrer, 1996; Kelly, 1998; Kemmer, 2003; Gries, 2004a, 2004b, 2006; Ronneberger-Sibold, 2006; Brdar-Szabó & Brdar, 2008; Fandrych, 2008). Here, a variety of methodological and theoretical tools are tested. These more recent studies show that corpus data, psycholinguistic methods, and frameworks such as CL afford new opportunities to increase the understanding of blending. This is, very roughly, the outset of the current investigation.

In this study, lexical blending is considered to be a word formation process characterized by a large set of cognitive and socio-pragmatic mechanisms. Psycholinguistically oriented blend studies (e.g., Lehrer, 1996), cognitively inclined approaches (e.g., Kelly, 1998), and corpus-based investigations with a CL approach (e.g., Gries, 2004b, 2006) constitute examples of the epistemic background to the methodology and the following analyses. On a more theoretical level, general assumptions concerning conceptual structure, or *meaning*, are based on CL accounts such as Taylor (2003), Langacker (2008), and Schmid (2015, 2020). The schema-based approach to analyzing blends is mainly inspired by Kemmer (2003). Taken together, these accounts offer a broad explanatory frame of the study.

The research question aiming to identify blend characteristics in systematically collected corpus data has been met by addressing a number of perspectives in the investigation. First, the quantitative aspects of the data have been shown to be statistically reliable. The method of data collection is robust, and the operational definition has proven reasonable as a background to the following analyses (5.1). The validated data have then been tested as to a few recurring issues in blend research: the grammar, semantics, and potential seriality of blends (5.2).

Qualitative concerns have then been examined, which has resulted in the identification of a set of key general characteristics that have emerged in the analyses of the data (5.3). The second research question, concerned with how ambiguity is realized in the blend data, has been accomplished by showing that four types of ambiguity can be detected. These ambiguities exploit different cognitive mechanisms to arrive at multiple layers of simultaneously occurring meaning potentials (6.1–6.4). Furthermore, an alternative model for classifying blends has been suggested based on schematic representations of the fundamental functions in blending, combined with a broad set of additional mechanisms that come into play in different degrees (6.5.2). It has also been shown that systematically exploited ambiguity can be connected to the cognitive phenomenon of multistability, which has been studied in other fields of cognitive research (6.5.3). This observation aligns with the CL tenet of the non-autonomous nature of language. In an attempt to highlight the import of this general cognitive phenomenon in lexical blending, the concept of multistable meaning construal is suggested as a possible theoretical approach to explain ambiguity in blending.

The identified characteristics and the use of ambiguity are gradually distributed in the data. Moreover, their distribution commonly overlaps in the occurring blends. Therefore, it is reasonable to describe the observed patterns in chapters 5 and 6 in terms of characteristics that influence blend formation. The analyses of main characteristics and patterns of ambiguity also show that some mechanisms, such as figurativity and iconicity, tend to be systematically distributed.

The general picture emerging from the observed patterns is a language phenomenon that truly resists "the imposition of boundaries between language and other psychological phenomena" (Langacker, 2008: 8), which is why a CL approach is particularly appropriate. Following this overview, the two sections 7.1 and 7.2 address the two research questions in more detail (see 1.1).

7.1 What blend characteristics can be identified through a systematic corpus method of data collection?

On the basis of data that have been systematically collected and successfully validated, it has been possible to identify and describe several characteristics. *Structural profiling*, *domain proximity pseudomorphemic transfer*, and *seriality* are four schematic concepts that capture differing strategies of formation. Other patterns have also been identified, such as:

- the functional connection between figurative meaning and the profiling of structural strategies

- the functional connection between iconicity and domain proximity of source words

- a preference for single-source word truncation, which contrasts to the traditional conception of blending characterized by truncation of both source words

The analyses of serial blends suggest a potential systematicity in the process through which some serial blend segments undergo morphemization. The suggestion is that morphonotactic qualities and semantic potential influence the probability of morphemization to occur. Series of constructions based on segments undergoing morphemization may then eventually be interpreted as compounds or derivatives rather than blends. However, other types of data are likely required to investigate this suggestion further. The quantitative analyses of seriality contribute empirical support to consider blends in serial relation to be analyzed as instantiations of lexical blending, albeit to varying degrees.

The quantitative robustness of the data has enabled revisiting important assumptions on blending. It has been shown that previously reported distributions of grammatical potential are generally mirrored in the data (e.g., Cannon, 1986; Elisa, 2019; Renner, 2019), which contributes further empirical evidence to basic assumptions on grammatical tendencies in blending. Equally, the analyses of semantic domains largely support previous assumptions on the semantic functions of blends (e.g., Lehrer, 2007; Danilović Jeremić & Josijević, 2019). These observations in combination stress the connection in blending between grammatical functions (mostly nominal in the blend data) and semantic functions (mostly names related to business, technology, and the media in the blend data), which also attests to the CL claims concerning the close interrelatedness between syntagmatic function and conceptual structure (Langacker, 2008: 95).

7.2 How is ambiguity realized in the blend data?

At least four types of ambiguity are realized in the data: *truncation ambiguity*, *mode ambiguity*, *source word ambiguity*, and *covert source ambiguity*. The functions of these ambiguities differ in several ways, but one of the key observations is that they exploit distinct cognitive strategies. They operate on *formal*, *modal*, *lexical* and *conceptual* levels, which further stresses the broad explanatory scope of contemporary blend research conducted within a CL framework. Occurrences of blends may simultaneously instantiate several of these types of ambiguity, which makes blending a markedly multimodal language phenomenon.

The ambiguity on the formal level is realized in underdetermined truncation structures. The more extreme the truncation ambiguity, the more it connects to observations of multistable perception often exemplified with the phenomenon of binocular rivalry. Blends with completely overlapping structure constitute a prime example of this phenomenon as both source words can be perceived in full, which has sometimes led to the conclusion that there is no truncation in these items. It is argued here that truncation does in fact occur, but it is instead realized as a multistable structure.

Ambiguity manifested as conflicting modes of expression is realized in the simultaneous activation of the phonemic representation of a source word and the corresponding graphemic structure in the blend. The strategy to make use of graphemic dependence to achieve wordplay effects underscores the importance of mode ambiguity in blending.

The third type, source word ambiguity, resides in the double potential of lexical interpretation. An analysis of intercalative structures, which constitute the most prominent examples of this type of ambiguity, indicates that graphemic structures could possibly be the dominant systemic factor in realizing this ambiguity. This mechanism is also observed in instantiations of wordplay.

Covert source ambiguity stands out from the other variants in that it operates with (at least partly) concealed sources. The focus on covert sources implies that certain inferences can be drawn on the basis of patterns in the corpus data seen in analyses of context and possibly also implications of frequency on lexical activation.

It has been stressed that it is neither easy nor uncontroversial to apply the concept of ambiguity in systematic models (see 6.5.2). This study claims that incorporating ambiguity in a theory of lexical blending is motivated since ambiguity encompasses several characteristics of blends that are otherwise difficult to reconcile.

7.3 The nature of the data

The construction of the data collection process has been guided by an intention to make the study quantitatively defensible and possible to replicate. As shown in the presentation of the methodology, the combination of low frequencies and the lack of annotation of blends involves a number of methodological challenges. The perhaps most apparent challenge is how to address the need for manual analysis of large amounts of data, especially since efforts to retrieve blends automatically have not yet met the requirements for precision sufficiently well (see 3.4.5). To address the issue of manual workload, the concepts of CaWs and SP were introduced as operators of data retrieval. The chosen approach to collect three separate sets of data enabled the collected blends to be related to the randomly selected CaWs, and to compare the patterns of distribution between the three subsets of data (see 4.3). The evaluation of the output asserted the accuracy of the data collection model (see 5.1.1). Following this procedure, the manual work was possible to keep on a reasonable level, while the statistical value of the data would remain intact.

The second challenge is concerned with the nature of the corpus. The size of the NOW corpus and the technical requirements of the organization and analysis of the data have required the use of powerful digital tools with regard to both software and hardware. As a consequence, a considerable proportion of the research project has been dedicated to technical concerns. Much of the effort dedicated to these challenges is to do with the need to strike a balance between digital data and qualitative linguistic analyses. Time management quickly becomes an issue when dealing with the gigantic amounts of information in contemporary online corpora, especially if their annotation is insufficient for a certain purpose.

Although the application of the methodology has largely proven successful, the model could be developed and improved in many ways. For instance, it could involve further automated blend identification procedures including matching and exclusion of attested non-blends (Professor Maarten Lemmens, personal communication; see also Renwick & Renner, 2019). This would enable processing of considerably larger datasets with potentially higher statistical precision, without necessarily increasing the manual work. A reversed type of corpus matching, in which already attested blends are identified and stored separately could also contribute, albeit more moderately, to a shift towards more automated procedures. Annotation of existing blends remain, however, a long-term goal as there is currently little agreement as regards classification. A more overarching aim might be the development of a procedure that could be employed using less extensive technical requirements but still be robust enough to be statistically reliable. Such a tool might offer an opportunity to be used collectively so as to construct a systematic corpus of blends annotated for a broad range of variables including diachronic information. The methodology applied in the current investigation may perhaps inspire the further development of such a tool, and ultimately a corpus powerful enough to disentangle some of the riddles of blending.

7.4 Revisiting some central issues in blending

This section returns to the issues of defining blends as outlined in 3.2 and 3.3. The themes related to the preliminary definition in 3.1 are the number of source words, truncation, overlap, intercalation, graphemic dependence, and wordplay (3.2). More general issues in blending concern the semantics of blends, diachronic issues, the nature of blend segments, delineational problems, and data collection (3.3).

Definitions of blends in previous research commonly assert that the number of source words is potentially more than two (e.g., Cannon, 2000; Lehrer, 2007; Gries, 2012; López Rúa, 2012; Renner, 2019). In the data of this investigation only three such instantiations were found, of which two were duplicates. Hence, they are very rare in these data (see also Renner, 2015: 126). Other equally rare phenomena, such as fully graphemic blends or blends based on verbs as source words, are usually not included in basic definitions of blends, and yet they seem no rarer than blends with more than two source words. Thus, the question is whether it is motivated to include the possibility of more than two source words in basic definitions of lexical blends. This is not to downplay the affordance to include several source words in a blend. It is even suggested here that source words may be involved in a blend without necessarily being overtly identifiable (see 3.2.1 and 6.4). The argument is instead that the possibility of having more than two source words is one of many equally possible attributes of blends. Hence, if the dual model of blend classification is applied, the basic blend schema would only require the fusion of two source words (see Figure 37 in 6.5.2). The possibility of constructing a blend based on more than two source words would probably be better referred to as one of the available mechanisms pertaining to the additional mechanisms as outlined in 6.5.2.

Full overlap blends (or A|BC|D blends) such as *hip-hopera* constitute a challenge to analyses of formal structure (see 3.2.4). The perhaps most apparent issue concerns how truncation should be interpreted in such blends. This investigation suggests that truncation occurs in full overlap blends as an instantiation of the general cognitive principle of multistability (see 6.5.3). This approach to truncation highlights the dynamic nature of meaning and form of blends. Moreover, the ambiguity related to truncation is not restricted to full overlap blends but is acknowledged in less prototypical instantiations of ambiguous truncation patterns as well.

Blends depending entirely on the graphemic representation – called graphemic blends – are uncommon in the data, while their structure is strikingly attention-catching (see 3.2.6). It has been highlighted in the analyses in section 6.2 that only a minimal extension of the graphemic dependence increases the number of instantiations dramatically. For this reason, the notion of near-graphemic blends has been employed to stress the potential to make use of graphemic dependence even if the blend does not entirely rely on the graphemic representation (6.2). According to this view, the full impact of graphemic dependence would not be captured by a strict categorization including only fully graphemic blends. Instead, including near-graphemic blends is in line with the chosen prototype theory perspective. The distribution of graphemic dependence is gradual, and it has been shown that both quantitative and qualitative aspects of this mechanism influence its use in blend formation.

Classification is a recurring topic in blend research, and recent studies have shown a growing interest in categorizing blends from the perspective of prototype theory (e.g., Brdar-Szabó & Brdar, 2008; Beliaeva, 2014b; Renner, 2015, 2019). The findings presented in chapters 5 and 6 are accommodated within the basic tenets of prototype theory and family resemblance (e.g., Rosch, 1978; Taylor, 2003; Geeraerts, 2006b, 2009). The distributions of the identified characteristics are typically matters of degree. Moreover, any specific attribute may also occur in often seemingly arbitrary ways, while there are some signs of complementary distribution, for instance as regards structural profiling and figurativity. Hence, some of the blends stand out as prototypical instantiations, while others appear peripheral or even examples of borderline cases.

Another common issue in accounts of lexical blending concerns diachronic matters (3.3.3). The methodology applied in this study is largely designed to study blends synchronically. However, at least two issues that emerge from the analyses of the data may have import for diachronic concerns. First, the potential connection between morphemization and morphemic qualities of serial blend segments is a suggestion that concerns development over time (cf. Lehrer, 1998, 2007; Beliaeva, 2014a; Jurado, 2019). The synchronic data presented in 5.3.5 give a first indication that such a connection could exist. Testing and extending this suggestion in more detail could be done by collecting diachronic data of blends in serial distribution, and by further investigating whether serial blend segments with litte or no morpheme-like qualities are more prone to morphemization. Second, the detection of covert source words may also have diachronic implications. The observations described in 6.4 show how the concealed meaning potentials are manifested within the selected temporal scope, but it has not been considered meaningful to study diachronic development during the relatively short time span of the data of this investigation. However, almost two centuries have passed since the first records of the examples of covert source ambiguity *bodacious* and *splurge* (Merriam-Webster, n.d.-a). Their respective usage patterns have likely undergone change during this time, and a diachronic investigation of their use could contribute to the exploration of covert source ambiguity by providing information about how the specialized usage patterns (e.g., bodily reference in *bodacious*) have developed over time.

As regards distributional issues concerned with the division of syntagmatic and paradigmatic blends (3.3.1), the data show that the source words of blends characterized by pseudomorphemic transfer (5.3.4) tend to be in syntagmatic distribution. The other characteristics do not seem to impose this constraint, while paradigmatic distribution of source words seems generally more prominent in blends characterized by structural profiling and domain proximity. The recurring emphasis in previous research on paradigmatic blends as more typical instantiations of lexical blending (e.g., Kelly, 1998; Dressler, 2000; Bat-El, 2006) thus seems to coincide with the characteristics structural profiling and domain proximity.

The concept of wordplay is commonly ascribed to blending. All the same, it is a topic in need of further systematic investigation (3.2.7). As wordplay has not been operationalized in the current investigation, comments on it are of a more indirect nature. The perhaps most notable observations with relevance for wordplay are the intricate structures of structural profiling (5.3.1), the play with conflicting activation in mode ambiguity (6.2), and the ambiguity of source words (6.3). In all these manifestations, different cognitive functions serve to challenge the processing of linguistic structure. Further studies into the workings of wordplay could perhaps interpret such indications in more detail and expand the knowledge on the conceptual underpinnings as well as distributional characteristics of wordplay.

Finally, intercalation (3.2.5) has been included as one of the central issues in blending because of its potential in relation to ambiguity, but it is also closely connected to the notion of wordplay. Analyses of intercalation have shown that such structures can be realized both graphemically and phonemically (see 6.3). Moreover, there need not be a correspondence between these two modes of expression. The relatively high proportion of intercalative structures in the data suggests that intercalation may in fact be more pervasive in blending than suggested in Kemmer (2003).

7.5 The dual model of blend classification

The well-attested issues of blend classification have been extensively described in previous research (see chapter 3). It may be reasonable to question whether lexical blends are items that can be classified along the same lines as concepts such as inflection or word class. The characteristics that have been identified in chapters 5 and 6 accord with previous suggestions of prototype theory categorization (Renner, 2015, 2019), but this observation only partly helps to resolve the issue of classification. Therefore, an attempt has been made in the present study to outline a model that splits up the analyses of blends into two levels of organization (6.5.2). This suggestion is discussed here.

In general, models of classification typically need a capacity to deal with borderline cases and exceptions. As Gries (2012: 146) points out, not only blends are hard to categorize, but most linguistic phenomena need to tackle this issue. One of the problems of blends is that classical categorization usually results in too many exceptions for any suggested taxonomy to be plausible (cf. Kemmer, 2003: 71). Seeing blends as items that relate to a central prototype may evade some of the problems of exceptions, but then the question is what constitutes a prototypical blend. The findings here suggest that there are a broad range of constructs that could compete for prototype status, which indicates that it is in fact hard to justify a specific type of construction as a prototypical blend. An additional perspective to classify blends seems necessary, which has been addressed by suggesting the dual model of blend classification.

The chosen theoretical frame is another important concern of classifying blends. Should a set of delimitations be applied in a taxonomic attempt, they still depend on the chosen theory. This is not unique to the classification of blends. However, the level of disagreement on fundamental properties in blend research makes it difficult to reach a reasonable degree of acceptance even as to the identity of the object (Bauer, 2012). For instance, disregarding syntagmatic blends (e.g., Kelly, 1998; Dressler, 2000) implies excluding large amounts of data, and considering complex clippings as a type of blend (e.g., Algeo, 1978; Ronneberger-Sibold, 2006) involves the inclusion of several mechanisms that have been shown to separate blends from complex clippings in important ways (e.g., Gries, 2012; Beliaeva, 2014b).

The suggestion to divide blends into two levels of organization is motivated by a preliminary strategy to meet some of the challenges of the diversity of blends. The main outline of the suggestion – the dual model of blend classification – has already been sketched in 6.5.2, but a few additional comments are called for. The issue of truncation has been discussed extensively in previous chapters, and the data have provided insights into some distributional characteristics. An important observation is the consistency of the basic patterns AD, ACD, ABD, and A|BC|D. The latter pattern has sometimes been considered a special case, but the argument for multistable meaning construal (involving assigning the truncation pattern a symbolic status) makes it potentially both an ABD and an ACD pattern. Therefore, the stable constituent parts of this pattern are the symbols for the schematic segments A and D, which is why the A|BC|D pattern is thought of as an instantiation of the AD schema. Thus, the remaining (basic) schemas for truncation in blending are AD, ABD, and ACD. Should future studies show that there are strong reasons to consider complex clippings as a type of blend, the truncation pattern AC would, accordingly, be included in the depiction of the truncation schema.

Another fundamental characteristic is conceptual integration (e.g., Ronneberger-Sibold, 2006; Beliaeva, 2014a). For instance, figurative meaning and iconicity have been shown to provide conceptual affordances in blends. More research is needed to further outline the distribution of such strategies, but the fusion of concepts seems sufficiently central to be included as a component of the basic blend schema (see 6.5.2). There may be other characteristics that could also be included in the basic level of the dual model, such as overlap and grammatical function, but at this stage, truncation and conceptual integration are considered a reasonable point of departure. These characteristics are the two key qualities in the basic schema of a blend in this preliminary outline of the model.

The second level of the dual model is concerned with characteristics that follow more complex patterns of distribution. In this study, several such additional mechanisms have been identified. The aforementioned gradual distribution and overlap of characteristics imply that the blends are centered around some prominent prototypical characteristics, and that the interrelation between blends is best understood from the perspective of family resemblance (cf. Taylor, 2003; Brugman & Lakoff, 2006; Langacker, 2008). As a consequence, the categorization of blends is seen as dynamic and flexible, allowing regrouping depending on what perspectives of blending are investigated.

In sum, the classificatory approaches described in 3.4.1 have served the important purpose of suggesting terms and delineations available for taxonomic analyses. Studies concerned with the structure (3.4.2), the psychological underpinnings (3.4.3), and sociopragmatic concerns (3.4.4) have constituted the necessary background to analyze blend patterns and then to offer an alternative way to classify blends. Balteiro and Bauer (2019) stress the importance of all these perspectives, including revisiting the complex issue of classification. The model proposed here – the dual model of blend classification – should be seen as a theoretical suggestion intended for further investigation, in particular, the mechanisms it intends to describe merits further exploration.

7.6 Contributions

This research project has attempted to meet the aims and research questions described in the introduction. In addition, several noteworthy findings have emerged at various stages of the process. The results presented in the dissertation connects both to the specific subject area of lexical blending and to a broader context, by which it has been inspired in many ways. Some of the central contributions to the fields in which this dissertation is situated are listed below.

The descriptions of characteristics in chapters 5 and 6 constitute a suggested analytic frame for the blends in the study. However, given the large size of the corpus and its broad scope of registers, the identified characteristics have explanatory value for blends more generally. The complementary distributions of figurativity and iconicity in relation to structural profiling and domain proximity provide systematic tools to further explore the semantics of blends. Neither figurativity nor iconicity appears to be arbitrarily applied in blending, which adds a type of semantic systematicity to the previously observed patterns of structural regularity (e.g., Gries, 2006, 2012; Wulff & Gries, 2019). The observed quantitative preference for single truncation is another contribution that may influence the understanding of how blends are usually formed.

The realization of ambiguity contributes a perspective that has been advanced occasionally in previous research (e.g., Veale & Butnariu, 2010). The current investigation contributes a detailed account showing how ambiguity is exploited systematically and constitutes an intrinsic part of the formation of lexical blends. In CL terms, ambiguity has symbolic status in blending. Furthermore, the observations of ambiguity have been interpreted as instantiations of multistable meaning construal, which is connected to more general cognitive principles of multistable perception. Thus, multistable meaning construal constitutes an explanatory concept that could potentially be applied to other aspects of language that make use of ambiguity as a means of communication.

The systematic data collection of this study constitutes another important contribution to blend research. Until rather recently, the nature of the data has not been an issue of any particular concern in blend research. There may be several reasons for this, but in parallel with the growing interest in statistical models to investigate blends, it has been pointed out that the nature of the data requires more care (e.g., Arndt-Lappe & Plag, 2013; Wulff & Gries, 2019). As pointed out in Wulff and Gries (2019), the origin of the data employed in previous blend studies is often opaque, and there is potentially a risk that certain particularly attention-catching qualities are given a too prominent role. The systematic data collection procedure presented in chapter 4 is therefore an addition to the growing number of methodologies that serve to validate and increase the knowledge of blending. Importantly, this contribution relies on the fact that the data collection methodology has been shown to be robust (5.1.1).

The value of this investigation is perhaps primarily identified as a contribution to the general knowledge of lexical blends. As it relates to the study of word formation and morphology, it may provide important insights in these fields as well, perhaps especially on the issue of the relation between morphemization and blend classification.

There are also key implications for CL. The perhaps most notable contribution is the established connection between multistable perception and lexical blending. The widely acknowledged phenomenon of multistability in cognitive sciences offers a potential theoretical background in the exploration of further implications of multistable meaning construal. In addition, multistability as an interpretive framework of blends aligns well with the assumption that "linguistic structure is seen as drawing on other, more basic systems and abilities (e.g., perception, memory, categorization) from which it cannot be segregated" (Langacker, 2008: 8).

7.7 Future prospects

The following reflections mark the end of this research project, but they offer at the same time an outlook on issues that have emerged in the investigation. Only a few topics are suggested below, and there are surely other informative and valuable approaches that could develop the findings of this study further.

The systematic data collection procedure, perhaps along with the suggested improvements in 4.5, may be applied to other contexts than the one explored in this investigation. Replicating the data collection procedure in other contexts could contribute valuable information about possible discrepancies and similarities in the use of blends in different contexts. It could also add to the general knowledge of blending in that an increasing number of replications likely improves the validity of the accumulated results.

Metonymy and metaphor have been shown to underlie the semantics of the blends in the data. This appears to be a particularly promising subject area in blend research. Future investigations could expand the exploration of figurativity in blends to see, for instance, how certain tropes may be distributed in specific contexts, whether certain conceptual metaphors and metonymies influence figurativity in blends, and whether there are systematic differences between metaphors in blends and compounds. Such studies would likely require a focus on qualitative methods of analysis, but the representativity of the data should be transparent in all future investigation of blends.

The issue of seriality remains relevant for blend research. Not only does it seem impossible to predict which blends become models for similar constructions, but seriality also blurs the boundaries of word formation categories and highlights the need for both synchronic and diachronic approaches. Investigating further the potential impact of morphemic qualities in serial blend segments could contribute to a better understanding of what constraints operate in morphemization. Lalić-Krstin and Silaški (2018) report two examples of semantically bleached concepts inspired by the blend *Brexit*: *Mexit* (*Messi* + *exit*) and *Zumxit* (*Zuma* + *exit*). In both these examples, *exit* is no longer used to refer to a country leaving a higher order administration, which is the recurring pattern of other *Brexit* model blends such as *Auxit* (*Austria* + *exit*), *Dexit* (*Deutschland* + *exit*), *Fixit* (*Finland* + *exit*), *Frexit* (*France* + *exit*), and *Germexit*/*Gerxit* (*Germany* + *exit*) (examples taken from Lalić-Krstin & Silaški, 2018) However, *exit* was already an established conventional lexeme before the formation of *Brexit*. The observed variation in the formation of these blends will, therefore, unlikely have any dramatic effect on *exit*. Although this type of pattern appears reasonable, future studies could collect larger samples of serial blend data to investigate the potential impact of this issue from a quantitative perspective. The more overarching issue of why seriality comes about remains challenging, but an increased knowledge of morphemization could possibly provide useful information to disentangle the issue of seriality as well.

Sammanfattning på svenska

Föreliggande studie undersöker ett ordbildningsfenomen som på engelska går under benämningen *lexical blends*. Termens närmaste svenska motsvarighet är *teleskopord*, vilken syftar på ordsammansättningar där vanligen två ord kombineras genom att båda eller endast ett av dem trunkeras. Ett vanligt förekommande engelskt exempel är *brunch*, vilket är en teleskopsammansättning av *breakfast* (sv. *frukost*) och *lunch*. I *brunch* har således *breakfast* reducerats till *br-* och *lunch* till *-unch*. Betydelsen av denna ordbildningsprocess har diskuterats i tidigare forskning, och engelskan har lyfts fram som ett språk där teleskopord är särskilt ofta förekommande (Kemmer, 2003).

Studiens syfte är tvådelat. Dels avser den att konstruera en metod för att samla in teleskopord på ett systematiskt sätt, för att därigenom kunna dra generella slutsatser inom den valda kontexten utifrån ett representativt urval. Studiens andra delsyfte är att undersöka den systematiskt insamlade listan av teleskopord med avseende på övergripande karaktäristika samt med avseende på hur mångtydighet – eller *ambiguitet* (eng. *ambiguity*) – framträder i materialet. Studien genomförs bland annat mot bakgrund av att tidigare forskning inte i tillräckligt hög utsträckning har kunnat säkerställa representativiteten i insamlade data (Wulff & Gries, 2019), men också för att semantiska perspektiv på teleskopord framstår som ett område som behöver ytterligare uppmärksamhet (jfr. Gries, 2006; Renner, 2015; Balteiro & Bauer, 2019).

Det valda teoretiska ramverket är *Kognitiv Lingvistik* (eng. *Cognitive Linguistics* förkortat *CL*). Detta ramverk utgår från att språk inte särskiljs från övrig kognition (Lakoff & Johnson, 1980; Langacker, 2002; Taylor, 2002; Geeraerts, 2006c). Sammankopplingen av språk och övrig kognition i CL medför att kroppslig erfarenhet och encyklopedisk kunskap inkluderas i synen på språklig betydelsepotential. Det innebär också att språklig mening och struktur betraktas som dynamiska och kontinuerligt föränderliga i konventionaliseringsprocesser (Schmid, 2015, 2020). Det är från detta perspektiv därför naturligt att undersöka språk utifrån användande istället för att betrakta språk som ett idealiserat system av mening och struktur. Vidare betonas *semantik* – språklig mening – som det centrala perspektivet på språklig analys (Langacker, 2008). Denna breda ansats på språk och kognition studerar därför hur språklig mening konstrueras konceptuellt i ett flertal processer som tidigare strukturalistisk och generativ språkforskning har gett begränsad betydelse, t.ex. figurativitet, schematisk meningspotential och syntaktiska konstruktioner som meningsbärande. CL ger därmed möjlighet att undersöka ett brett spektrum av meningspotential i teleskopord, vilket har visat sig framgångsrikt i ett flertal studier under de senaste decennierna (Kelly, 1998; Kemmer, 2003; Gries, 2006; Beliaeva, 2014a). Slutligen är en språksyn i enlighet med CL inte förenlig med klassisk aristoteliansk kategorisering (Taylor, 2003). Kategorier bör istället företrädesvis beskrivas utifrån perspektivet prototypteori (Rosch, 1973, 1978; Geeraerts, 2006b), vilket utvärderar kategoritillhörighet holistiskt utifrån grader av likhet med en prototyp. Även i kategoriseringshänseende lämpar sig CL väl som teoretisk utgångspunkt för att studera teleskopord (jfr. Renner, 2015).

Ett återkommande problem med forskning på teleskopord är deras diversitet (Cannon, 1986; Beliaeva, 2014a, 2019). Svårigheten att enkelt sammanfatta vad som utgör ett teleskopord innebär således att det också är svårt att jämföra olika studier då utgångspunkten är oklar (Bauer, 2012). Denna avhandling ger därför först en preliminär definition av teleskopord, vilken sedan problematiseras utifrån olika perspektiv. Därefter görs en genomgång av huvudlinjerna i forskningen om teleskopord från 1900-talets början fram till rådande dagsläge. I denna översikt lyfts särskilt metodologiska frågeställningar och ansatser fram, då dessa utgör en grund för studiens epistemologiska överväganden kring teleskopord. Slutligen presenteras en operationell definition som utgör grundvalen för det systematiska insamlandet av data. Denna operationella definition ska förstås som avhandlingens explicita avgränsning av teleskopord. Transparensen i denna definition syftar bland annat till att göra datainsamlingen replikerbar.

Insamlandet av data har getts stort utrymme i avhandlingen, då metodens transparens är viktig för anspråket på materialets representativitet. Data har samlats in från onlinekorpuset *News On the Web corpus* (NOW; sv. ung. *nyheter på nätet*). NOW samlar sedan 2010 kontinuerligt in data från 20 engelskspråkiga regioner. I denna studie avgränsades det empiriska materialet till data från USA från och med januari 2010 till och med mars 2018. För att kunna hantera den stora datamängden (drygt en miljard ord) i kombination med snäva utsökningskriterier, laddades korpuset ned i en offlineversion. Därefter tillämpades en strikt algoritm för hur teleskopord skulle identifieras i ett slumpmässigt texturval från offlinekorpuset. Den lista av teleskopord som detta resulterade i analyserades därefter utifrån avhandlingens teoretiska premisser.

En primär kvantitativ analys av det empiriska materialet visade att den valda metoden för insamling och identifikation av teleskopord är tillförlitlig med avseende på kvantitativ precision. Utifrån denna bakgrund testades tre aspekter på teleskopord: ordklass, semantiska domäner och serialitet. Det sistnämnda berör observationen att delar av ett teleskopord kan uppvisa en viss begränsad produktivitet manifesterad i ordformer som följer ett mönster från ett ursprungligt teleskopord. Exempel på detta är begrepp som *Frexit* (*France* + *exit*) och *Bexit* (*Belgium* + *exit*) som följt på de ursprungliga *Brexit* (*British* + *exit*) eller *Grexit* (*Greece* + *exit*). Ordklass, semantiska domäner och serialitet har undersökts i tidigare forskning, och observationerna i denna studie bekräftar – med några tillägg och modifikationer – i stort vad som tidigare rapporterats.

I en kvalitativ analys av materialet framstod fyra övergripande tendenser som särskilt belysande. I en svensk översättning kan dessa kallas *profilering av struktur*, *konceptuell närhet*, *pseudomorfemisk transfer*, och *serialitet*. De två första tendenserna är direkt observerbara i enskilda teleskopord, och kan beskrivas som egenskaper som har att göra med hur ett teleskopord realiseras. I materialet har dessa tendenser visat sig koppla till figurativitet och ikonicitet. Pseudomorfemisk transfer är ett begrepp relaterat till teleskopord som schematiskt liknar närliggande morfologiska processer. Serialitet förstås i detta sammanhang som en encyklopedisk kunskap om ett specifikt ordbildnings-mönster, som realiseras som en tillgänglig affordans kopplad till en grupp av teleskopord. I studiens material finns till exempel orden *Spanglish* (*Spanish* + *English*), *Yinglish* (*Yiddish* + *English*) och *Singlish* (*Singapore* + *English*), vilka illustrerar denna potential.

Undersökningen av ambiguitet bland studiens teleskopord kan sammanfattas i fyra olika typer. Översatt till svenska kan dessa kallas *trunkeringsambiguitet*, *modal ambiguitet*, *källordsambiguitet* och *ambiguitet med dold källa*. Dessa typer av mångtydighet innefattar teleskopordens form såväl som begrepps-mässiga aspekter. Den sammantagna bilden som framträder i materialet är att ambiguitet spelar en central roll som motiv för användningen av teleskopord. Mot bakgrund av dessa iakttagelser diskuteras ambiguitet som en potentiell förklaringsmodell. Detta resonemang utmynnar i ett förslag på en tvådelad teoretisk modell för hur teleskopord kan definieras, samt hur observationerna av ambiguitet kan kopplas till multistabilitet av liknande typ som iakttagits i relation till andra kognitiva fenomen.

Avhandlingen sammanfattas slutligen i bokens sista kapitel. Här besvaras forskningsfrågorna i en mer formell tappning. Några centrala frågeställningar för framtida forskning preciseras, och studiens iakttagelser sätts i samband med tidigare forskning.

References

Algeo, John. (1977). Blends, a Structural and Systemic View. *American Speech* 52. 47-64.

Algeo, John. (1978). The Taxonomy of Word Making. *Word* 29. 122-131.

Alm-Arvius, Christina. (2012). *Comprehensive Semantics. Lexicon, Grammar, Text Cognition & the World*. Stockholm University: Unpublished manuscript.

Andrews, Sally. (1997). The Effect of Orthographic Similarity on Lexical Retrieval: Resolving Neighborhood Conflicts. *Psychonomic Bulletin & Review* 4. 439-461.

Arndt-Lappe, Sabine & Plag, Ingo. (2013). The Role of Prosodic Structure in the Formation of English Blends. *English Language and Linguistics* 17. 537-563.

Balteiro, Isabel & Bauer, Laurie (2019). Introduction. *Lexis. Journal In English Lexicology*. 14. 1-3. https://doi.org/10.4000/lexis.4085

Barcelona, Antonio. (2003a). *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*. Berlin: Mouton de Gruyter.

Barcelona, Antonio. (2003b). Metonymy in Cognitive Linguistics: An Analysis and a few Modest Proposals. *Amsterdam Studies in the Theory and History of Linguistic Science. Motivation in Language: Studies in honor of Günter Radden*., Vol. ed. by H. Cuyckens, T. Berg & R. Dirven, 223-255. Amsterdam: John Benjamins Publishing Company.

Barnhart, Robert K. & Steinmetz, Sol. (1988). *Chambers Dictionary of Etymology*. London: Chambers Harrap Publishers Ltd.

Barsalou, Lawrence W. (2012). The Human Conceptual System. *The Cambridge Handbook of Psycholinguistics*, Vol. ed. by M. J. Spivey, M. Joanisse & K. McRae, 239-258. Cambridge & New York: Cambridge University Press.

Bat-El, Outi. (2006). Blend. *Encyclopedia of Language & Linguistics (2nd ed.)*, Vol. 2 ed. by K. Brown, and A.H.Anderson, 66-70. Amsterdam: Elsevier.

Bauer, Laurie. (1983). *English Word-Formation*. Cambridge & New York: Cambridge University Press.

Bauer, Laurie. (2006). Compounds and Minor Word-Formation Types. *The Handbook of English Linguistics*, Vol. ed. by B. Aarts & A. M. S. McMahon, 483-506. Oxford: Blackwell Publishing Ltd.

Bauer, Laurie. (2012). Blends: Core and Periphery. *Cross-Disciplinary Perspectives on Lexical Blending*, Vol. ed. by V. Renner, P. Arnaud, V. Gast & F. Maniez, 11-22. Berlin: De Gruyter.

Bauman, Zygmunt. (1991). *Modernity and Ambivalence*. Oxford: Polity press.

Bauman, Zygmunt. (2001). *The Individualized Society*. Cambridge: Polity Press.

Beliaeva, Natalia. (2014a). A Study of English Blends: From Structure to Meaning and Back Again. *Word Structure* 7. 29-54.

Beliaeva, Natalia. (2014b). *Unpacking Contemporary English Blends: Morphological Structure, Meaning, Processing*: Victoria University of Wellington, Doctoral Dissertation.

Beliaeva, Natalia. (2019). Blending Creativity and Productivity: On the Issue of Delimiting the Boundaries of Blends as a Type of Word Formation. *Lexis. Journal In English Lexicology*. 14. 1-22. https://doi.org/10.4000/lexis.4004

Bergen, Benjamin K. (2012). *Louder than Words: The New Science of how the Mind Makes Meaning*. New York: Basic Books.

Bergström, Gustaf Adolf. (1906). *On Blendings of Synonymous or Cognate Expressions in English: A Contribution to the Study of Contamination*: Lund University, Doctoral Dissertation.

Berman, Jules J. (2013). *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*. Amsterdam: Elsevier Science & Technology.

Berrey, Lester V. (1939). Newly-Wedded Words. *American Speech* 14. 3-10.

Booij, Geert. (2010). *Construction Morphology*. New York: Oxford University Press.

Brdar-Szabó, Rita & Brdar, Mario. (2008). On the Marginality of Lexical Blending. *Jezikoslovlje* 9. 171-194.

Brône, Geert & Feyaerts, Kurt. (2003). *The Cognitive Linguistics of Incongruity Resolution: Marked Reference-point Structures in Humor*. KU Leuven, Departement Linguistiek.

Bross, Fabian. (2018*). Cognitive Associations between Vowel Length and Object Size: A new Feature Contributing to a Bouba/kiki Effect*. Paper presented at the Conference on Phonetics & Phonology in German-Speaking Countries, Humboldt-Universität zu Berlin.

Brugman, Claudia & Lakoff, George. (2006). Radial Network: Cognitive Topology and Lexical Networks. *Cognitive Linguistics: Basic Readings*, Vol. ed. by D. Geeraerts, 109-139. New York: Mouton de Gruyter.

Bryant, Margaret M. (1974). Blends are Increasing. *American Speech* 49. 163.

Cambridge Dictionary. (n.d.). Retrieved from https://dictionary.cambridge.org/ 2022-03-11.

Cameron, Lynne. (2003). *Metaphor in Educational Discourse*. London: Continuum.

Cameron, Lynne & Maslen, Robert. (2010). *Metaphor Analysis: Research Practice in Applied Linguistics, Social Sciences and the Humanities*. London: Equinox.

Cannon, Garland. (1986). Blends in English Word Formation. *Linguistics* 24. 723-753.

Cannon, Garland. (2000). Blending. *Morphology: An International Handbook on Inflection and Word-Formation*, Vol. ed. by G. Booij, C. Lehmann & J. Mugdan, 952-956. Berlin & New York: Walter de Gruyter.

Carroll, Lewis. (1871). *Through the Looking Glass and What Alice Found There*. Oxford: The Millenium Fulcrum Edition.

Carroll, Lewis. (1894). *Alice's Adventures in Wonderland and Through the Looking Glass*. New York: MacMillan & Co.

Cenoz, Jasone & Bereziartua, Garbiñe. (2016). Is Instant Messaging the Same in Every Language? A Basque Perspective. *Reading and writing* 29. 1225-1243.

Chomsky, Noam. (1968). *Language and Mind*. New York: Harcourt, Brace & World.

Cook, Paul. (2012). *Using Social Media to Find English Lexical Blends*. Paper presented to the 15th EURALEX International Congress (EURALEX 2012), Oslo, Norway, 2012.

Cook, Paul & Stevenson, Suzanne. (2010). Automatically Identifying the Source Words of Lexical Blends in English. *Computational Linguistics* 36. 129-149.

Corpusdata.org. (n.d.). *Full-text Corpus Data*. Retrieved from https://www.corpusdata.org/ 2021-02-18.

Correia Saavedra, David. (2016). Automatically Identifying Blend Splinters that are Morpheme Candidates. *Digital Scholarship in the Humanities* 31. 55-71.

Croft, William. (2006). Metonymy: The Role of Domains in the Interpretation of Metaphors and Metonymies. *Cognitive Linguistics: Basic Readings*, Vol. ed. by D. Geeraerts, 269-302. New York: Mouton de Gruyter.

Crystal, David. (2001a). *Language and the Internet*. Cambridge & New York: Cambridge University Press.

Crystal, David. (2001b). *Language Play*. Chicago: University of Chicago Press.

Crystal, David. (2011). *Internet Linguistics: A Student Guide*. New York: Routledge.

Curry, Ramona. (2009). Reviving the History, Revising the Historiography of Female Media Pioneers. *Journal of women's history* 21.188-203.

Danilović Jeremić, Jelena & Josijević, Jelena. (2019). To Blend so as to Brand: A Study of Trademarks and Brand Names. *Lexis. Journal In English Lexicology*. 14. 1-19. https://doi.org/10.4000/lexis.3732

Dollywood Parks and Resorts. (n.d.). Retrieved from https://www.dollywood.com/ 2021-06-23.

Dressler, Wolfgang U. (2000). Extragrammatical vs. Marginal Morphology. *Extragrammatical and Marginal Morphology*, Vol. ed. by U. Doleschal & A. M. Thornton, 1-10. Munich: Lincom Europa.

Elisa, Mattiello. (2019). A Corpus-based Analysis of New English Blends. *Lexis. Journal In English Lexicology*. 14. 1-28. https://doi.org/10.4000/lexis.3660

English-Corpora. (n.d.). *News On the Web corpus*. Retrieved from https://www.english-corpora.org/now/ 2022-03-11.

Fábregas, Antonio & Scalise, Sergio. (2012). *Morphology: From Data to Theories*. Oxford: Oxford University Press.

Fandrych, Ingrid. (2008). Pagad, Chillax and Jozi: A Multi-Level Approach to Acronyms, Blends, and Clippings. *Nawa: Journal of Language & Communication* 2. 71-88.

Fauconnier, Gilles & Turner, Mark. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Fauconnier, Gilles & Turner, Mark. (2006). Mental Spaces: Conceptual Integration Networks. *Cognitive Linguistics: Basic Readings*, Vol. ed. by D. Geeraerts, 303-372. New York: Mouton de Gruyter.

Fillmore, Charles J. (1982). *Frame Semantics*. Paper presented at the SICOL 1981.

Fisher, Kerstin. (2010). Quantitative Methods in Cognitive Semantics: Introduction to the Volume. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, Vol. ed. by D. Glynn & K. Fisher, 43-59. Berlin: de Gruyter Mouton.

Fontaine, Lise. (2017). The Early Semantics of the Neologism BREXIT: A Lexicogrammatical Approach. *Functional Linguistics* 4.

Fradin, Bernard. (2000). Combining Forms, Blends and Related Phenomena. *Extragrammatical and Marginal Morphology*, Vol. ed. by U. Doleschal & A. M. Thornton, 11-59. Munich: Lincom Europa.

Fradin, Bernard. (2015). Blending. *Word-Formation: An International Handbook of the Languages of Europe*, Vol. 1 ed. by P. O. Müller, I. Ohnheiser, S. Olsen & F. Rainer, 386-413. Berlin & Boston: De Gruyter Mouton.

Frath, Pierre. (2005). Why is there no Ham in a Hamburger?' A Study of Lexical Blends and Reanalysed Morphemisation. *RANAM: Recherches Anglaises et Nord Amércaines* 38. 99-112.

Geeraerts, Dirk. (1989). Prototype Theory. *Linguistics* 27. 587-612.

Geeraerts, Dirk. (2006a). Introduction: A Rough Guide to Cognitive Linguistics. *Cognitive Linguistics: Basic Readings*, Vol. ed. by D. Geeraerts, 1-28. New York: Mouton de Gruyter.

Geeraerts, Dirk. (2006b). Prototype Theory: Prospects and Problems of Prototype Theory. *Cognitive Linguistics: Basic Readings*, Vol. ed. by D. Geeraerts, 141-165. New York: Mouton de Gruyter.

Geeraerts, Dirk. (2009). *Theories of Lexical Semantics*. Oxford: Oxford University Press.

Geeraerts, Dirk (ed.) (2006c). *Cognitive Linguistics: Basic readings*. New York: Mouton de Gruyter.

Gibbs, Raymond W. (1992). When Is Metaphor? The Idea of Understanding in Theories of Metaphor. *Poetics Today* 13. 575-606.

Gibbs, Raymond W. (1994). *The Poetics of Mind: Figurative Thought, Language and Understanding*. Cambridge: Cambridge University Press.

Gibbs, Raymond W. (2003). Embodied Experience and Linguistic Meaning. *Brain and Language* 84. 1-15.

Gibbs, Raymond W. & Colston, Herbert L. (2006). Image Schema: the Cognitive Psychological Reality of Image Schemas and their Transformations. *Cognitive Linguistics: Basic Readings*, Vol. ed. by D. Geeraerts, 239-268. New York: Mouton de Gruyter.

Gibbs, Raymond W. & Van Orden, Guy. (2010). Adaptive Cognition without Massive Modularity. *Language and Cognition* 2. 149-176.

Giora, Rachel, Fein, Ofer, Kotler, Nurit & Shuval, Noa. (2015). Know Hope: Metaphor, Optimal Innovation and Pleasure. *Cognitive Linguistics and Humor Research*, Vol. ed. by G. Brône, K. Feyaerts & T. Veale, 129-146. Berlin & Boston: De Gruyter Mouton.

Glynn, Dylan & Fisher, Kerstin (eds) (2010). *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. Berlin: de Gruyter Mouton.

Goldberg, Adele. (2005). *Constructions at Work: the Nature of Generalization in Language*. New York: Oxford University Press.

Goossens, Louis. (1990). Metaphtonymy: the Interaction of Metaphor and Metonymy in Expressions for Linguistic Action. *Cognitive Linguistics* 1. 323-342.

Grésillon, Almuth. (1984). *La Règle et le Monstre: Le Mot-Valise. Interrogations sur La Langue à Partir d'un Corpus de Heinrich Heine*. Tübingen: Niemeyer.

Gries, Stefan Th. (2004a). Isn't that Fantabulous? How Similarity Motivates Intentional Morphological Blends in English. *Language. Culture, and Mind*, Vol. ed. by M. Achard & S. Kemmer, 415-428. Stanford, California: CSLI Publications.

Gries, Stefan Th. (2004b). Shouldnt it be Breakfunch? A Quantitative Analysis of Blend Structure in English. *Linguistics* 42. 639-667.

Gries, Stefan Th. (2006). Cognitive Determinants of Subtractive Word Formation: A Corpus-based Perspective. *Cognitive Linguistics* 17. 535-558.

Gries, Stefan Th. (2010). Corpus Linguistics and Theoretical Linguistics: A Love–Hate Relationship? Not Necessarily.... *International Journal of Corpus Linguistics* 15. 327-343.

Gries, Stefan Th. (2012). Quantitative Corpus Data on Blend Formation: Psycho- and Cognitive-Linguistic Perspectives. *Cross-Disciplinary Perspectives on Lexical Blending*, Vol. ed. by V. Renner, P. Arnaud, V. Gast & F. Maniez, 145-167. Berlin: De Gruyter.

Hamans, Camiel. (2010). The Productivity of Blending: Linguistic or Cognitive? Or How to Deal with Administrivia and Ostalgia. *Lingua Terra Cognita II. A Festschrift for professor Roman Kalisz*, Vol. ed. by D. Stanulewicz, T.Z. Wolański & J. Redzimska, 467-490. Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego.

Hanks, Patrick. (2000). Do Word Meanings Exist*? Computers and the Humanities* 34. 205-215.

Harley, Trevor A. (2008). *The Psychology of Language: From Data to Theory*. New York: Psychology Press.

Hay, Jennifer. (2001). Lexical Frequency in Morphology: Is Everything Relative? *Linguistics* 39. 1041-1070.

Herring, Susan C. (2011). Computer-mediated Conversation Part II: Introduction and Overview. *Language@Internet* 8. 1-12.

Hockett, Charles F. (1967). Where the Tongue Slips, There Slip I. *To honor Roman Jakobson*, Vol. ed. by M. Halle, 910-936. The Hague: Mouton.

Hockett, Charles F. (1973). *Man's Place in Nature*. New York: McGraw-Hill.

Internet Movie Database. (n.d.). *Misstake*. Retrieved from https://www.imdb.com/title/tt1290407/ 2021-05-25.

Jespersen, Otto. (1922). *Language, its Nature, Development and Origin*. London: George Allen & Unwin Ltd.

Johnson, Mark. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago Press.

Juhasz, Barbara J., Johnson, Rebecca L. & Brewer, Jennifer. (2016). An Investigation into the Processing of Lexicalized English Blend Words: Evidence from Lexical Decisions and Eye Movements During Reading. *Journal of Psycholinguistic Research*.1-14.

Jurado, Alejandro Barrena. (2019). A Study on the 'Wordgasm': The Nature of Blends' Splinters. *Lexis. Journal In English Lexicology* 14. 1-49. https://doi.org/10.4000/lexis.3916

Kelly, Michael. (1998). To "Brunch" or to "Brench": Some Aspects of Blend Structure. *Linguistics* 36. 579-590.

Kemmer, Suzanne. (2003). Schemas and Lexical Blends. *Amsterdam Studies in the Theory and History of Linguistic Science. Motivation in Language: Studies in Honor of Günter Radden*, Vol. ed. by H. Cuyckens, T. Berg & R. Dirven, 69-97. Amsterdam: John Benjamins Publishing Company.

Kjellander, Daniel. (2018). Cognitive Constraints in English Lexical Blending: A Data Collection Methodology and an Explanatory Model. *Pragmatics & Cognition* 25. 142-173.

Kjellander, Daniel. (2019). Gold Punning: Studying Multistable Meaning Structures Using a Systematically Collected Set of Lexical Blends. *Lexis. Journal In English Lexicology*. 14. 1-28. https://doi.org/10.4000/lexis.3962

Kornmeier, Jürgen, Hein, Christine Maira & Bach, Michael. (2009). Multistable Perception: When Bottom-Up and Top-Down Coincide. *Brain and Cognition* 69. 138-147.

Kubozono, Haruo. (1990). Phonological Constraints on Blending in English as a Case for Phonology-Morphology Interface. *Yearbook of morphology* 3. 1-20.

Köhler, Wolfgang. (1929). *Gestalt Psychology*. New York: Liveright.

Kövecses, Zoltán & Radden, Günter. (1998). Metonymy: Developing a Cognitive Linguistic View. *Cognitive Linguistics* 9. 37-78.

Lakoff, George. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Lakoff, George. (2006). Conceptual Metaphor: The Contemporary Theory of Metaphor. *Cognitive Linguistics: Basic Readings*, Vol. ed. by D. Geeraerts, 185-238. New York: Mouton de Gruyter.

Lakoff, George & Johnson, Mark. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Lalić-Krstin, Gordana. (2014). Forms and meanings of the source word armageddon in English lexical blends. *Engleski jezik i anglofone književnosti u teoriji i praksi: Zbornik u čast Draginji Pervaz [English Language and Anglophone Literatures in Theory and Practice: Festschrift in Honour of Draginja Pervaz]*, Vol. ed. by T. Prćić & M. Marković, 257-273. Novi Sad: Filozofski fakultet.

Lalić-Krstin, Gordana & Silaški, Nadežda. (2018). From Brexit to Bregret. *English Today* 34. 3-8.

Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar*. Stanford, California: Stanford University Press.

Langacker, Ronald W. (2002). *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Berlin: Mouton de Gruyter.

Langacker, Ronald W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Lehrer, Adrienne. (1996). Identifying and Interpreting Blends: An Experimental Approach. *Cognitive Linguistics* 7. 359-390.

Lehrer, Adrienne. (1998). Scapes, Holics, and Thons: The Semantics of English Combining Forms. *American Speech* 73. 3-28.

Lehrer, Adrienne. (2003). Understanding Trendy Neologisms. *Italian Journal of Linguistics* 15. 371-384.

Lehrer, Adrienne. (2007). Blendalicious. *Lexical Creativity, Texts and Contexts*, Vol. ed. by J. Munat, 115-133. Amsterdam: John Benjamins Publishing Company.

Leopold, David A. & Logothetis, Nikos K. (1999). Multistable Phenomena: Changing Views in Perception. *Trends in Cognitive Sciences* 3. 254-264.

Lepic, Ryan. (2016). *Lexical Blends and Lexical Patterns in English and in American Sign Language*. Paper presented to the Online Proceedings of the Tenth Mediterranean Morphology Meeting MMM10 7-10 September 2015, Haifa, 2016.

Levelt, Willem J. M. (1968). *On Binocular Rivalry*. The Hague: Mouton.

Lindsey, Treva B. . (2015). Post-Ferguson: A “Herstorical” Approach to Black Violability. *Feminist studies* 41. 232-237.

López-Rúa, Paula. (2019). From Carmageddon and Invizimals to SimCity and Digimon: Blending Patterns in Videogame Titles. *Complutense Journal of English Studies* 27. 183-204.

López Rúa, Paula. (2004). The Categorial Continuum of English Blends. *English Studies* 85. 63-76.

López Rúa, Paula. (2007). Keeping up with the Times: Lexical Creativity in Electronic Communication. *Lexical Creativity, Texts and Contexts*, Vol. ed. by J. Munat, 137-159. Amsterdam: John Benjamins Publishing Company.

López Rúa, Paula. (2010). Making and Breaking the Rules: Lexical Creativity in the Alternative Music Scene. *Language Awareness* 19. 51-67.

López Rúa, Paula. (2012). Beyond All Reasonable Transgression: Lexical Blending in Alternative Music. *Cross-Disciplinary Perspectives on Lexical Blending*, Vol. ed. by V. Renner, P. Arnaud, V. Gast & F. Maniez. Berlin: De Gruyter.

Macmillan Dictionary. (n.d.). Retrieved from https://www.macmillandictionary.com/ 2022-03-11.

McMahan, Peter & Evans, James. (2018). Ambiguity and Engagement. *American Journal of Sociology* 124. 860-912.

McMahon, April M. S. (1994). *Understanding Language Change*. Cambridge: Cambridge University Press.

Merriam-Webster. (n.d.-a). Retrieved from https://www.merriam-webster.com/ 2022-03-11.

Merriam-Webster. (n.d.-b). *miss*. Retrieved from https://www.merriam-webster.com/dictionary/miss 2020-04-10.

Merriam-Webster. (n.d.-c). *Spanglish*. Retrieved from https://www.merriam-webster.com/dictionary/Spanglish 2021-04-15.

Merriam-Webster. (n.d.-d). *Yinglish*. Retrieved from https://www.merriam-webster.com/dictionary/Yinglish 2021-04-15.

MissTake. (n.d.). *Home [Facebook Page]*. Facebook. Retrieved from https://www.facebook.com/misstaketattoo/ 2021-05-25.

Oaks, Dallin D. (2010). *Structural Ambiguity in English: An Applied Grammatical Inventory*. London & New York: Continuum.

Olsen, Susan. (2015). Composition. *Word-Formation: An International Handbook of the Languages of Europe*, Vol. 1 ed. by P. O. Müller, I. Ohnheiser, S. Olsen & F. Rainer, 364-386. Berlin & Boston: De Gruyter Mouton.

Peirsman, Yves & Geeraerts, Dirk. (2006). Metonymy as a Prototypical Category. *Cognitive Linguistics* 17. 269-316.

Plag, Ingo. (2003). *Word-Formation in English*. Cambridge: Cambridge University Press.

Plag, Ingo. (2015). English. *Word-Formation: An International Handbook of the Languages of Europe*, Vol. 4 ed. by P. O. Müller, I. Ohnheiser, S. Olsen & F. Rainer. Berlin & Boston: De Gruyter Mouton.

PostgreSQL Tools. (n.d.). *pgAdmin: Open Source Community Project*. Retrieved from https://www.pgadmin.org/ 2018-09-04.

Pound, Louise. (1914). Blends: Their Relation to English Word Formation. *Vol. 42. Anglistische Forschungen*. Heidelberg: Carl Winter's Universitätsbuchhandlung.

Quirk, Randolph, Svartvik, Jan, Leech, Geoffrey & Greenbaum, Sidney. (1985). *A Comprehensive Grammar of the English Language*. London & New York: Longman.

Ramachandran, Vilayanur S. & Hubbard, Edward M. (2001). Synaesthesia: A Window into Perception, Thought, and Language. *Journal Of Consciousness Studies* 8. 3-34.

Renner, Vincent. (2015). Lexical Blending as Wordplay. *Wordplay and Metalinguistic /Metadiscursive Reflection: Authors, Contexts, Techniques, and Meta-Reflection*, Vol. ed. by A. Zirker & E. Winter-Froemel, 119-133. Berlin & Boston: Walter de Gruyter GmbH.

Renner, Vincent. (2019). French and English Lexical Blends in Contrast. *Languages in contrast* 19. 27-47.

Renner, Vincent. (2020). An Ecosystem View of English Word-Formation. *Mental Lexicon* 15. 4-20.

Renner, Vincent, Arnaud, Pierre & Maniez, François. (2012a). Introduction: A Bird's-eye View of Lexical Blending. *Cross-Disciplinary Perspectives on Lexical Blending*, Vol. ed. by V. Renner, P. Arnaud, V. Gast & F. Maniez. Berlin: De Gruyter.

Renner, Vincent, Maniez, François & Arnaud, Pierre (eds) (2012b). *Trends in Linguistics. Studies and Monographs: Cross-Disciplinary Perspectives on Lexical Blending*. Berlin: De Gruyter Mouton.

Renwick, Adam & Renner, Vincent. (2019). New Lexical Blends in the Simpsons: A Formal Analysis of English Nonce Formations and their French Translations. *Lexis. Journal In English Lexicology*. 14. 1-16. https://doi.org/10.4000/lexis.3829

Roig-Marín, Amanda. (2016). 'Blended' Cyber-neologisms. *English Today* 32. 2-5.

Ronneberger-Sibold, Elke. (2006). Lexical Blends: Functionally Tuning the Transparency of Complex Words. *Folia Linguistica* 40. 155-181.

Ronneberger-Sibold, Elke. (2012). Blending Between Grammar and Universal Cognitive Principles: Evidence from German, Farsi, and Chinese. *Cross-Disciplinary Perspectives on Lexical Blending*, Vol. ed. by V. Renner, P. Arnaud, V. Gast & F. Maniez. Berlin: De Gruyter.

Rosch, Eleanor. (1973). On the Internal Structure of Perceptual and Semantic Categories. *Cognitive Development and the Acquisition of Language*, Vol. ed. by T. E. Moore, 111-144. New York, San Francisco & London: Academic Press Inc.

Rosch, Eleanor. (1978). Principles of Categorization. *Cognition and Categorization*, Vol. ed. by B. B. Lloyd & E. Rosch, 27-48. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Ross, John. (2009). *Network Know-How: An Essential Guide for the Accidental Admin*. San Francisco: No Starch Press.

Rubin, Nava. (2003). Binocular Rivalry and Perceptual Multi-stability. *Trends in Neurosciences* 26. 289-291.

Ruiz de Mendoza, Francisco. (2003). The Role of Mappings and Domains in Understanding Metonymy. *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*, Vol. ed. by A. Barcelona. Berlin & Boston: De Gruyter.

Saussure, Ferdinand de. (1966). *Course in General Linguistics*. New York: McGraw-Hill.

Schmid, Hans-Jörg. (2015). A Blueprint of the Entrenchment-and-Conventionalization Model. *Yearbook of the German Cognitive Linguistics Association*. Vol. 3:1 ed. by P. Uhrig & T. Herbst, 3-26.

Schmid, Hans-Jörg. (2020). *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment*. Oxford: Oxford University Press.

Schmid, Hans-Jörg. (2011). *English Morphology and Word-Formation: An Introduction*. Berlin: Schmidt.

Shaw, Philip. (2008). Spelling, Accent and Identity in Computer-Mediated Communication. *English Today* 24. 42-49.

Sherman, Donald. (1975). Noun-Verb Stress Alternation: An Example of the Lexical Diffusion of Sound Change in English. *Linguistics* 13. 43-72.

Silaški, Nadežda & Đurović, Tatjana. (2013). Of “siliconaires” and “millionerds”: How ESP Learners Understand Novel Blends in English. *Iberica*. 85-105.

Simner, Julia. (2012). Defining Synaesthesia. *The British Journal of Psychology* 103. 1-15.

Soudek, Lev. (1971). The Development and Use of the Morpheme Burger in American English. *Linguistics* 9. 61-89.

Stroop, John R. (1935). Studies of Interference in Serial Verbal Reactions. *Journal of Experiential Psychology* 18. 643-662.

Svanlund, Jan. (2009). *Lexikal etablering: En korpusundersökning av hur nya sammansättningar konventionaliseras och får sin betydelse*. Stockholm: Acta Universitatis Stockholmiensis.

Tarasova, Elizaveta & Beliaeva, Natalia. (2020). On Twittizens and City Residents: Experimental Study of Semantic Relations in English Compounds and Blends. *Mental Lexicon* 15. 21-41.

Taylor, John R. (2002). *Cognitive Grammar*. Oxford: Oxford University Press.

Taylor, John R. (2003). *Linguistic Categorization*. Oxford: Oxford University Press.

Taylor, John R. (2015). Word-Formation in Cognitive Grammar. *Word-Formation: An International Handbook of the Languages of Europe*, Vol. 1 ed. by P. O. Müller, I. Ohnheiser, S. Olsen & F. Rainer. Berlin & Boston: De Gruyter Mouton.

Tomasello, Michael. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, Massachusetts: Harvard University Press.

Tomasello, Michael. (2006). Usage-based Linguistics: First Steps Toward a Usage-based Theory of Language Acquisition. *Cognitive Linguistics: Basic Readings*, Vol. ed. by D. Geeraerts, 439-458. New York: Mouton de Gruyter.

Tomasello, Michael. (2008). *Origins of Human Communication*. Cambridge, Massachusetts: MIT Press.

Tournier, Jean. (1985). *Introduction Descriptive à la Lexicogénétique de l'Anglais Contemporain*. Genève & Paris: Champion, Slatkine.

Tuggy, David. (1993). Ambiguity, Polysemy, and Vagueness. *Cognitive Linguistics* 4. 273-290.

Tuggy, David. (2006). Schematic Network: Ambiguity, Polysemy, and Vagueness. *Cognitive Linguistics: Basic Readings*, Vol. ed. by D. Geeraerts, 167-184. New York: Mouton de Gruyter.

Ungerer, Friedrich. (2003). The Conceptual Function of Derivational Word-Formation in English. *Anglia* 120. 534-567.

Ungerer, Friedrich & Schmid, Hans-Jörg. (2006). *An Introduction to Cognitive Linguistics*. Harlow, UK: Longman, Pearson Education Ltd.

University Centre for Computer Corpus Research on Language (UCREL). (n.d.-a). *CLAWS part-of-speech tagger for English*. Retrieved from http://ucrel.lancs.ac.uk/claws/ 2021-02-18.

University Centre for Computer Corpus Research on Language (UCREL). (n.d.-b). *Free USAS English web tagger*. Retrieved from http://ucrel-api.lancaster.ac.uk/usas/tagger.html 2021-06-30.

University Centre for Computer Corpus Research on Language (UCREL). (n.d.-c). *Introduction to the USAS Category System*. Retrieved from http://ucrel.lancs.ac.uk/usas/usas_guide.pdf 2021-06-30.

University Centre for Computer Corpus Research on Language (UCREL). (n.d.-d). *USAS*. Retrieved from http://ucrel.lancs.ac.uk/usas/ 2021-06-30.

Veale, Tony & Butnariu, Cristina. (2010). Harvesting and Understanding On-line Neologisms. *Cognitive Perspectives on Word Formation*, Vol. ed. by A. Onysko & S. Michel, 399-421. Berlin & New York: Mouton De Gruyter.

Ward, Jamie. (2013). Synesthesia. *Annual Review of Psychology* 64. 49-75.

Waring, Vernon. (n.d.). *A Snoratorium*. Retrieved from http://www.mywordwizard.com/a-snoratorium.html 2021-06-24.

Warren, Beatrice. (1990). The Importance of Combining Forms. *Trends in Linguistics. Studies and Monographs [TiLSM]: Contemporary Morphology*, Vol. ed. by W. Dressler, H. Luschutzky, O. Pfeiffer & J. Rennison, 111-132. Berlin & Boston: De Gruyter Mouton.

Wentworth, Harold. (1933). Twenty-nine Synonyms for 'Portmanteau Word'. *American Speech* 8. 78-79.

Wentworth, Harold. (1934). Blend-words in English Glossaries Illustrative of *Linguistic Growth*. New York: Cornell University.

Wierzbicka, Anna. (1985). *Lexicography and Conceptual Analysis*. Ann Arbor: Karoma.

Wierzbicka, Anna. (2006). *English: Meaning and Culture*. Oxford: Oxford University Press.

Withington, Robert. (1932). More “Portmanteau” Coinages. *American Speech* 7.200-203.

Wittgenstein, Ludwig. (1968). *Philosophical Investigations*. Oxford: Blackwell.

Wulff, Stefanie & Gries, Stefan Th. (2019). Improving on Observational Blends Research: Regression Modeling in the Study of Experimentally-elicited Blends. *Lexis. Journal In English Lexicology*. 14. 1-21. https://doi.org/10.4000/lexis.3625

Yavaş, Mehmet. (2006). *Applied English Phonology*. Oxford: Blackwell Publishing.

Appendix 1. SQL scripts.

Script A. Script calculating the number of lemmas (excluding empty slot lemmas).

```
1   SELECT COUNT(lemma) AS lemma
2   FROM lexicon
3   WHERE lemma <> ''
```

Script B. Number of common noun lemmas

```
4   WITH noun_lemma_count AS
5       (
6       SELECT lemma, SUM(frequency) AS freq
7       FROM lexicon
8       WHERE pos LIKE 'nn%'
9       AND lemma <> ''
10      GROUP BY lemma
11      ORDER BY freq DESC
12      )
13      SELECT COUNT(lemma) AS no_of_nouns
14      FROM noun_lemma_count
```

Script C. Low frequency lemmas in the subcorpus

```
15  WITH hapax AS
16      (
17      SELECT lemma, SUM(frequency) AS freq
18      FROM lexicon
19      WHERE pos LIKE 'nn%'
20      AND lemma <> ''
21      GROUP BY lemma
22      ORDER BY freq DESC
23      )
24      SELECT lemma, freq
25      FROM hapax
26      WHERE freq < 4
27      ORDER BY freq DESC
```

Note: The assigned number on row 26 [**WHERE** freq < 4] retrieves all instances with 3 occurrences or less. Subsequently, the value ‘4’ is changed to ‘2’ to retrieve instances of hapax legomena (i.e., one token).

Script D. Tokens of common noun and verb lemmas (excluding empty slot lemmas; section 4.2.2) 'nn%' in line 30 was exchanged with 'vv%' in the verb query.

```
28  SELECT SUM(frequency) AS token_freq
29  FROM lexicon
30  WHERE pos LIKE 'nn%'
31  AND lemma <> ''
```

Script E. Script calculating the total number of tokens in the subcorpus (excluding empty slot lemmas).

```
32  SELECT SUM(frequency) AS token_freq
33  FROM lexicon
34  WHERE lemma <> ''
```

Script F. Calculating the token frequency of the CaWs.

```
35  SELECT SUM(frequency) AS frequency
36  FROM lexicon
37  WHERE lemma IN ('testimony','track', [...], 'literature','major')
```

Note: For each of the three datasets, all 100 text strings (CaWs) were pasted within the brackets of the WHERE...IN-clause. For reasons of legibility, only two initial and two final items have been retained in this representation. The 96 omitted items are represented on line 37 by the ellipsis symbol [...].

Script G. Tokens of lemmas among the 2,000 most frequent common nouns.

```
38  WITH top2000 AS
39          (
40          SELECT lemma, SUM(frequency) AS freq
41          FROM lexicon
42          WHERE pos LIKE 'nn%'
43          AND lemma <> ''
44          GROUP BY lemma
45          ORDER BY freq DESC
46          LIMIT 2000
47          )
48          SELECT SUM(freq) AS top2000_freq
49          FROM top2000
```

Script H. Randomized selection of Candidate Words (CaWs)

```
50	WITH random_nouns AS
51			(
52			SELECT lemma, SUM(frequency) AS frequency
53			FROM lexicon
54			WHERE pos LIKE 'nn%'
55			AND lemma <> ''
56			GROUP BY lemma
57			ORDER BY frequency DESC
58			LIMIT 2000
59			)
60			SELECT lemma
61			FROM random_nouns
62			ORDER BY random ()
63			LIMIT 100;
```

Note: For the verb CaW dataset the PoS condition on line 54 was altered to WHERE pos LIKE 'vv%'. In parallel, the chosen column name (lines 50 and 61; 'random_nouns') was changed accordingly in the work process to avoid confusion, although this change did not have any technical import on the function of the script.

Script J. SP calculation of CaWs (I) Right hand truncations and their respective frequencies.

```
64	SELECT sp.sp, COUNT(*) AS wID_freq, SUM(frequency) AS token_freq
65	FROM lexicon
66	JOIN selection_points('routine') sp
67	  ON lemma LIKE sp.sp || '%'
68	GROUP BY 1
69	ORDER BY 1
```

Script K. Truncation calculation of CaWs (II) Rank calculation for every truncation. This script was also used in the final truncform queries.

```
70	SELECT lemma, SUM (frequency) AS freq
71	FROM lexicon
72	WHERE lemma LIKE 'routi%'
73	GROUP BY lemma
74	ORDER BY freq DESC
```

Note: The output of script K produces a frequency ranking list that groups all structurally identical lemmas (see line 73). Therefore, each lemma typically contains a number of wIDs representing its varying grammatical uses, which should not be surprising given the grammatical diversity of most lemmas. For instance, 'routi%' produces 12 lemmas in the ranking list, but the total number of wIDs shown in the list is 41 (see table 8, section 4.3.2).

Script L. SP calculation of CaWs (II) Left hand truncations and their respective frequencies.

```
SELECT reverse(sp.sp), COUNT(*) AS wID_freq, SUM(frequency) AS token_freq
FROM lexicon
JOIN selection_points(reverse('major')) sp
  ON reverse(lemma) LIKE sp.sp || '%'
GROUP BY 1
ORDER BY 3 DESC
```

Script M. Analysis of conversion patterns of *date* and *roll.*

```
81 SELECT lemma, SUM (frequency) AS freq
82 FROM lexicon
83 WHERE lemma = 'date'
84 AND pos LIKE 'vv%'
85 GROUP BY lemma
86 ORDER BY freq DESC
```

Note: The search terms on line 83 were both 'date' and 'roll' applied in their respective queries. To retrieve their uses as nouns and verbs, the strings 'nn%' and 'vv%' on line 84 were employed respectively (see 5.1.2).

Script N. Collecting word class information.

```
SELECT lemma, pos, SUM (frequency) AS freq
FROM lexicon
WHERE lemma IN ('example1', ' example2', ' example3')
GROUP BY lemma, pos
ORDER BY freq DESC
```

Script P. Investigating PoS tag distribution.

```
SELECT pos, COUNT(wid) AS nr_wIDs, SUM(frequency) AS token_freq
FROM lexicon
WHERE pos LIKE 'nn%'
GROUP BY pos
ORDER BY token_freq DESC
```

Appendix 2. UCREL Claws 7 Tagset.

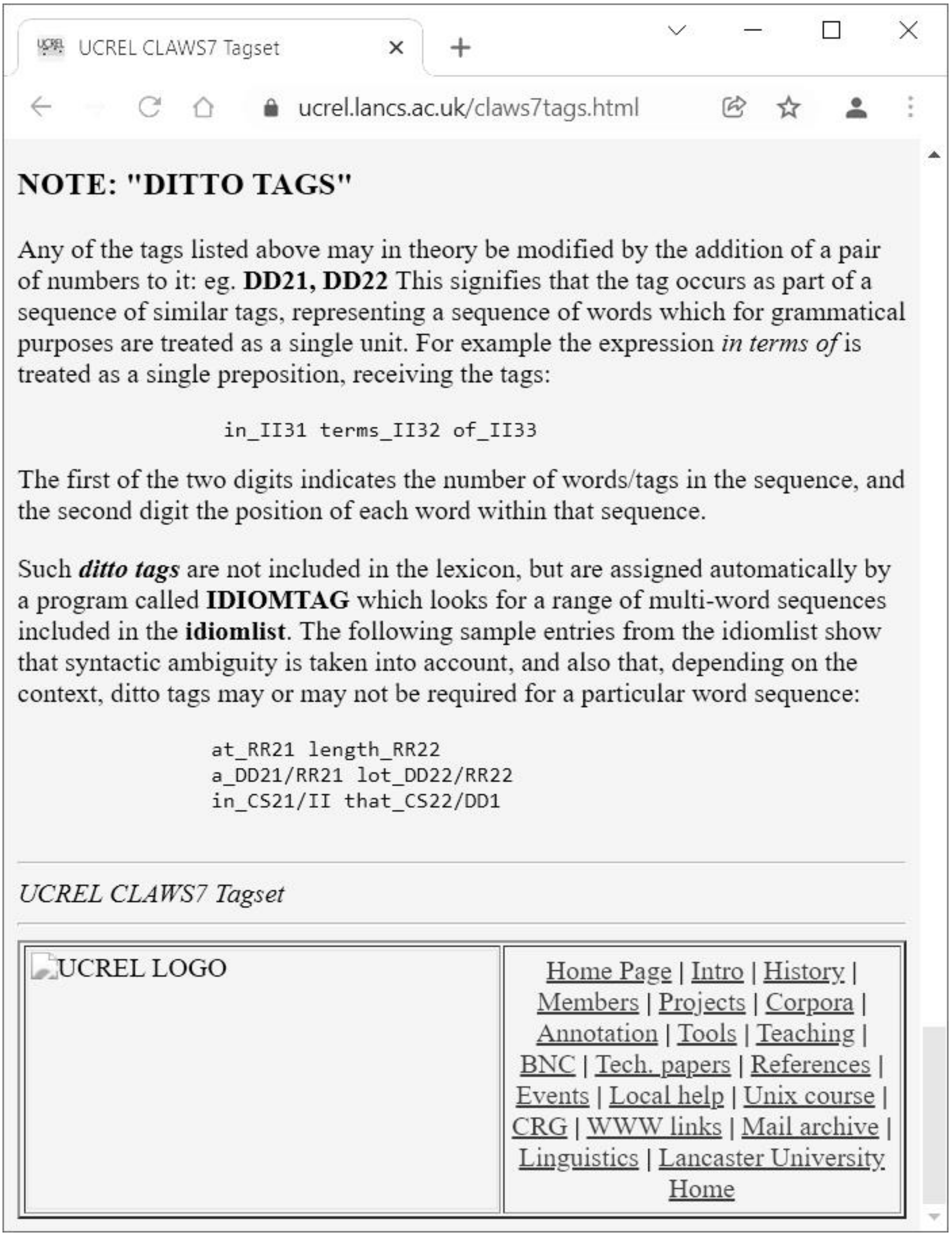
The entire tagset in Appendix 2, including the final text reproduced as a screenshot, is the work of UCREL. Retrieved from http://ucrel.lancs.ac.uk/claws7tags.html (2018-10-09).

APPGE	possessive pronoun, pre-nominal (e.g., my, your, our)
AT	article (e.g., the, no)
AT1	singular article (e.g., a, an, every)
BCL	before-clause marker (e.g., in order (that), in order (to))
CC	coordinating conjunction (e.g., and, or)
CCB	adversative coordinating conjunction (but)
CS	subordinating conjunction (e.g., if, because, unless, so, for)
CSA	as (as conjunction)
CSN	than (as conjunction)
CST	that (as conjunction)
CSW	whether (as conjunction)
DA	after-determiner or post-determiner capable of pronominal function (e.g., such, former, same)
DA1	singular after-determiner (e.g., little, much)
DA2	plural after-determiner (e.g., few, several, many)
DAR	comparative after-determiner (e.g., more, less, fewer)
DAT	superlative after-determiner (e.g., most, least, fewest)
DB	before determiner or pre-determiner capable of pronominal function (all, half)
DB2	plural before-determiner (both)
DD	determiner (capable of pronominal function) (e.g any, some)
DD1	singular determiner (e.g., this, that, another)
DD2	plural determiner (these,those)
DDQ	wh-determiner (which, what)
DDQGE	wh-determiner, genitive (whose)
DDQV	wh-ever determiner, (whichever, whatever)
EX	existential there
FO	formula
FU	unclassified word
FW	foreign word
GE	germanic genitive marker - (' or's)
IF	for (as preposition)
II	general preposition
IO	of (as preposition)
IW	with, without (as prepositions)
JJ	general adjective
JJR	general comparative adjective (e.g., older, better, stronger)
JJT	general superlative adjective (e.g., oldest, best, strongest)
JK	catenative adjective (able in be able to, willing in be willing to)
MC	cardinal number, neutral for number (two, three..)
MC1	singular cardinal number (one)

MC2	plural cardinal number (e.g., sixes, sevens)
MCGE	genitive cardinal number, neutral for number (two's, 100's)
MCMC	hyphenated number (40-50, 1770-1827)
MD	ordinal number (e.g., first, second, next, last)
MF	fraction, neutral for number (e.g., quarters, two-thirds)
ND1	singular noun of direction (e.g., north, southeast)
NN	common noun, neutral for number (e.g., sheep, cod, headquarters)
NN1	singular common noun (e.g., book, girl)
NN2	plural common noun (e.g., books, girls)
NNA	following noun of title (e.g., M.A.)
NNB	preceding noun of title (e.g., Mr., Prof.)
NNL1	singular locative noun (e.g., Island, Street)
NNL2	plural locative noun (e.g., Islands, Streets)
NNO	numeral noun, neutral for number (e.g., dozen, hundred)
NNO2	numeral noun, plural (e.g., hundreds, thousands)
NNT1	temporal noun, singular (e.g., day, week, year)
NNT2	temporal noun, plural (e.g., days, weeks, years)
NNU	unit of measurement, neutral for number (e.g., in, cc)
NNU1	singular unit of measurement (e.g., inch, centimetre)
NNU2	plural unit of measurement (e.g., ins., feet)
NP	proper noun, neutral for number (e.g., IBM, Andes)
NP1	singular proper noun (e.g., London, Jane, Frederick)
NP2	plural proper noun (e.g., Browns, Reagans, Koreas)
NPD1	singular weekday noun (e.g., Sunday)
NPD2	plural weekday noun (e.g., Sundays)
NPM1	singular month noun (e.g., October)
NPM2	plural month noun (e.g., Octobers)
PN	indefinite pronoun, neutral for number (none)
PN1	indefinite pronoun, singular (e.g., anyone, everything, nobody, one)
PNQO	objective wh-pronoun (whom)
PNQS	subjective wh-pronoun (who)
PNQV	wh-ever pronoun (whoever)
PNX1	reflexive indefinite pronoun (oneself)
PPGE	nominal possessive personal pronoun (e.g., mine, yours)
PPH1	3rd person sing. neuter personal pronoun (it)
PPHO1	3rd person sing. objective personal pronoun (him, her)
PPHO2	3rd person plural objective personal pronoun (them)
PPHS1	3rd person sing. subjective personal pronoun (he, she)
PPHS2	3rd person plural subjective personal pronoun (they)
PPIO1	1st person sing. objective personal pronoun (me)
PPIO2	1st person plural objective personal pronoun (us)
PPIS1	1st person sing. subjective personal pronoun (I)
PPIS2	1st person plural subjective personal pronoun (we)
PPX1	singular reflexive personal pronoun (e.g., yourself, itself)
PPX2	plural reflexive personal pronoun (e.g., yourselves, themselves)
PPY	2nd person personal pronoun (you)
RA	adverb, after nominal head (e.g., else, galore)

REX	adverb introducing appositional constructions (namely, e.g.,)
RG	degree adverb (very, so, too)
RGQ	wh- degree adverb (how)
RGQV	wh-ever degree adverb (however)
RGR	comparative degree adverb (more, less)
RGT	superlative degree adverb (most, least)
RL	locative adverb (e.g., alongside, forward)
RP	prep. adverb, particle (e.g about, in)
RPK	prep. adv., catenative (about in be about to)
RR	general adverb
RRQ	wh- general adverb (where, when, why, how)
RRQV	wh-ever general adverb (wherever, whenever)
RRR	comparative general adverb (e.g., better, longer)
RRT	superlative general adverb (e.g., best, longest)
RT	quasi-nominal adverb of time (e.g., now, tomorrow)
TO	infinitive marker (to)
UH	interjection (e.g., oh, yes, um)
VB0	be, base form (finite i.e., imperative, subjunctive)
VBDR	were
VBDZ	was
VBG	being
VBI	be, infinitive (To be or not... It will be...)
VBM	am
VBN	been
VBR	are
VBZ	is
VD0	do, base form (finite)
VDD	did
VDG	doing
VDI	do, infinitive (I may do... To do...)
VDN	done
VDZ	does
VH0	have, base form (finite)
VHD	had (past tense)
VHG	having
VHI	have, infinitive
VHN	had (past participle)
VHZ	has
VM	modal auxiliary (can, will, would, etc.)
VMK	modal catenative (ought, used)
VV0	base form of lexical verb (e.g., give, work)
VVD	past tense of lexical verb (e.g., gave, worked)
VVG	-ing participle of lexical verb (e.g., giving, working)
VVGK	-ing participle catenative (going in be going to)
VVI	infinitive (e.g., to give... It will work...)
VVN	past participle of lexical verb (e.g., given, worked)
VVNK	past participle catenative (e.g., bound in be bound to)

VVZ	-s form of lexical verb (e.g., gives, works)
XX	not, n't
ZZ1	singular letter of the alphabet (e.g., A, b)
ZZ2	plural letter of the alphabet (e.g., A's, b's)



UCREL CLAWS7 Tagset

ucrel.lancs.ac.uk/claws7tags.html

NOTE: "DITTO TAGS"

Any of the tags listed above may in theory be modified by the addition of a pair of numbers to it: eg. **DD21, DD22** This signifies that the tag occurs as part of a sequence of similar tags, representing a sequence of words which for grammatical purposes are treated as a single unit. For example the expression *in terms of* is treated as a single preposition, receiving the tags:

```
in_II31 terms_II32 of_II33
```

The first of the two digits indicates the number of words/tags in the sequence, and the second digit the position of each word within that sequence.

Such ***ditto tags*** are not included in the lexicon, but are assigned automatically by a program called **IDIOMTAG** which looks for a range of multi-word sequences included in the **idiomlist**. The following sample entries from the idiomlist show that syntactic ambiguity is taken into account, and also that, depending on the context, ditto tags may or may not be required for a particular word sequence:

```
at_RR21 length_RR22
a_DD21/RR21 lot_DD22/RR22
in_CS21/II that_CS22/DD1
```

UCREL CLAWS7 Tagset

UCREL LOGO

Home Page | Intro | History | Members | Projects | Corpora | Annotation | Tools | Teaching | BNC | Tech. papers | References | Events | Local help | Unix course | CRG | WWW links | Mail archive | Linguistics | Lancaster University Home

Appendix 3. The blend list.

Blend	*Dataset*	*CaW*	*SW1*	*SW2*
2pacalypse	VV	collapse	Tupac ('2pac')	apocalypse
Accuweather	NN2	weather	accurate	weather
Activia	NN2	activity	active	bacteria
Activision	NN1,NN2	action	active/activity	vision
Addaction	NN1	action	addiction	action
adware	NN1	hardware	Ad	software
agritainment	VV	agree	agriculture	entertainment
Alternet	NN2	net	alternative	net
Amasia	VV	amass	America	Asia
Amerindian	NN1	Indian	American	Indian
Amerisource	NN2	source	America/American	source
appletini	NN1	app	apple	martini
bodacious	NN1	body	bold	audacious
brainiac	NN2	brain	brain	maniac
brainware	NN2	brain	brain	software?
Calcrete	VV	calculate	calcium	concrete
Calexico	VV	call	California	Mexico
californication	VV	call	California	fornication
Canadarm	NN2	form	Canada	(robot) arm

Canstruction	NN1	action	can (tin)	construction
Chevolution	VV	check	Che	revolution
Chinglish	NN1,NN2	English	Chinese/China	English
Chrislam	NN2	Islam	Christianity	Islam
chrononaut	VV	chronicle	chrono-	astronaut
Cinemark	VV	park	cinema	Mark (proper noun)
clintonomics	NN2	economics	Clinton	economics
Collaboraction	NN1,VV	action	collaboration	action
Computrace	NN1	company	computer	trace
Computrainers	NN1	company	computer	trainers
Compuware	NN1	company	computer/company	software
Condomania	NN2	condition	condom	mania
Constraction	NN1	action	construction	action
cosmeceuticals	NN2	cost	cosmetics	pharmaceuticals
Crooklyn	NN1	crowd	crook	Brooklyn
digerati	VV	digest	digital	literati
disgraceland	VV,NN2	disguise	disgrace	Graceland
diversicrats	NN2	diversity	diversity	bureaucrats
Draganfly	NN1	drive	Dragan	dragonfly
Dragonet	NN1,NN2, NN2	drive	dragon	Dragnet (tv series)
Dragula	NN1,NN1	drive	drag	Dracula
dramady	NN1	drive	drama	comedy

Dreamsicle	NN1	drive	dream	popsicle
dumbfounding	NN2	funding	dumb	confounding
eco-logical	NN1	local	ecological	logical
Ecologo	NN2	logo	ecological	logo
egg-cellent	NN1	egg	Egg	excellent
egg-ceptional	NN1	egg	Egg	exceptional
eggs-actly	NN1	egg	Egg	exactly
Eggstravaganza	NN1	egg	Egg	extravaganza
egosystem	NN1	egg	Ego	ecosystem
Elephunk	VV	think	elephant	funk
Execunet	NN2	net	executive	net
Executrain	NN2	brain	executive	train
faceblock	VV,NN2	stock	Facebook	block
feedom	NN2	fee	Fee	freedom
freakonomics	NN2	economics	freak	economics
frienemy	NN2	friend	friend	enemy
frightmare	NN2	friend	fright	nightmare
Fundred	NN1	hundred	Fun	hundred
glocal	NN1	local	global	local
godzillionaire	VV	go	Godzilla	zillionaire
Goracle	VV	go	Gore	oracle
herstorical	NN1	local	Her	historical

Hinglish	NN1,NN2	English	Hindi	English
hip-hopera	NN2	camera	hip-hop	opera
Hungaroton	NN1	hundred	Hungary	Qualiton
Identigene	NN2	idea	identity/ identification	gene
Identikit	NN2	idea	identity	kit
Idioteque	NN2	idea	idio-/idiot	discoteque/ biblioteque
infomercial (misspelled "infomerical")	NN1	local	info(rmation)	commercial
Intellisense	NN2	sense	intelligence/-nt	sense
Invisalign	NN2	sign	invisible	align
jewxploitation	VV	station	jew/-ish	exploitation
Kongfrontation	VV	station	Kong	confrontation
Manulife	NN1,NN1	manufacturer	manufacturer	life
masshole	NN2	mass	Massachusetts	asshole
Massport	NN2	mass	Massachusetts	airport
mearth	VV	mean	Me	earth
Medicalert	NN2	expert	medical	alert
Metabolife	NN1	life	metabolism	life
Metalocalypse	VV	collapse	metal	apocalypse
Mexicatessen	VV	lessen	mexican/ mexico	delicatessen
mindware	NN1	hardware	mind	software
modfather	NN1,NN2	other	mod	godfather

Moneygram	NN2	ram	money	telegram
Navinet	VV	navigate	navigate	net
Netcaster	NN2	net	Net	broadcaster
Netegrity	NN2	net	Net	integrity
netiquette	NN2	net	Net	etiquette
Netizen	NN1,NN2	citizen	Net	citizen
Netware	NN2	net	Net	software?
Optimark	VV	park	optimal/ optimize	mark
palimony	NN1	testimony	Pal	alimony
permalloy	NN1,NN1	percent	permable	alloy
Plantronics	VV	plan	plane	electronics
popera	NN1,NN2	population	Pop	opera
Popmart	NN1	population	Pop	Kmart/ Walmart
poppermost	NN1	population	pop(ular)	uppermost/ toppermost
probeware	VV	probe	probe	software
Provenge	VV	prove	pro-	avenge/ revenge
Quikrete	NN1	question	quick	concrete
reaganomics	NN2	economics	Reagan	economics
relationshit	NN1	relation	relationship	shit
Robirds	VV	rob	robot	birds
Robocup	NN1,VV	robot	robocop	cup
rolladium	NN1,VV	roll	rollerblades/ rollerskate	stadium

scareware	NN2	school	scare	software?
Scattergories	NN2	school	scatter	categories
scentsational	NN2	school	scent	sensational
schlockumentary	NN2	school	schlock	documentary
schnoodle	NN2	school	schnauzer	poodle
scientifiction	NN2	school	scientific	fiction
Scitation	VV	station	science	citation
Screamadelica	NN2	school	scream	psychadelica
screenagers	NN2	school	screen	teenagers
Sense8	NN2	sense	sensate	8 (date: Aug 8)
Sensurround	NN2	sense	sense	surround
sexpert	NN2	expert	Sex	expert
sexploitation	VV	station	Sex	exploitation
Simulink	VV	think	simulation/ simulate	link
Singlish	NN1,NN2	English	Singapore	English
Sinterstation	VV	station	sinter	printerstation
sneakernet	NN2	snow	sneak	Ethernet/ Internet
snoratorium	NN2	snow	snore	?moratorium/ lat.dormitorium/ oratorium
snowpocalypse	VV,NN2	collapse	snow	apocalypse
Spamalot	VV	spend	spam	Camelot (cf. *a lot*)
spamvertised	VV	spend	spam	advertised

Spanglish	NN1,VV, NN2	English	Spanish	English
spay-ghetti	VV	spend	spay	spaghetti
Spinderella	VV	spend	Spin	Cinderella
splatterific	VV	spend	splatter	terrific
splurge	VV	spend	splash	surge
Spooktacular	VV	spend	spooky	spectacular
spork	VV	spend	spoon	fork
Sportacus	VV	spend	sport	Spartacus
sportcaster	VV	spend	sport	broadcaster
spyware	VV	spend	Spy	software
stagflation	NN1	population	stagnation	inflation
teavana	VV,NN2	tell	Tea	Nirvana
technopreneurs	VV	tell	technology	entrepreneurs
technorati	VV	tell	technology	literati
technotopia	VV	tell	technology	utopia
technovation	VV	tell	technology	innovation
teensploitation	VV,VV	station	teenage/-r	exploitation
Teledildonics (3 SWs)	VV	tell	telecommunication	dildo + electronics
televangelist	VV	tell	telephone/tele	evangelist
Tenncare	VV	tell	Tennessee	health care
testimoney	NN1,VV	testimony	testimony	money
Texarkana (3 SWs)	VV,NN2	tell	Texas	Arkansas + Louisiana

Texhoma	VV	tell	Texas	Oklahoma
Texican	VV	tell	Texas	Mexican
Texico	VV	tell	Texas	New Mexico
Textainer	NN1	text	Texas?	container
Thinkfinity	VV	think	think	infinity
thinspiration	VV	think	Thin	inspiration
Thinsulate	VV	think	Thin	insulate
tragicomedy	NN2	tragedy	tragic/tragedy	comedy
tragicomic	NN2	tragedy	tragic/tragedy	comic
tricknology	NN2	trial	trick	technology
trimnastics	NN2	trial	Trim	gymnastics
vactor	NN1	factor	vacuum	excavator
varactor	NN1	factor	variable	reactor
Vegemite	NN1,VV	vegetable	vegetarian/-able	Marmite
Verichip	VV	verify	verification	microchip
Verifone	VV	verify	verification	telephone
Verilux	VV	verify	veritas (lat.)	lux (light intensity)
Verisign	VV,NN2	verify	verify/veritas (lat.)/verified	sign
Verizon	VV	verify	veritas (lat.)	horizon
webinar	NN1	web	Web	seminar
webisode	NN1	web	Web	episode
webiverse	NN1	web	Web	universe

Websight	NN1	web	website	sight
webtop	NN1	web	Web	desktop
webzine	NN1	web	Web	magazine
workahol	NN2	school	work	alcohol
Yinglish	NN1,NN2	English	Yiddish	English
Youthanasia	NN1	youth	youth	euthanasia
Youthcare	NN1	youth	youth	healthcare
youthquake	NN1	youth	youth	earthquake

Appendix 4. USAS semantic categories.

The following list of semantic categories is based on UCREL Semantic Analysis System (USAS), which is available on http://ucrel.lancs.ac.uk/usas/ (retrieved 2021-01-12). Extensions have been made in accordance with the nature of the empirical material of the present study. These instances are marked with *. For the original USAS list of semantic categories, refer to http://ucrel.lancs.ac.uk/usas/.

(A) GENERAL & ABSTRACT TERMS

A1 General		
		A1.1.1 General actions, making etc.
		A1.1.2 Damaging and destroying
	A1.2 Suitability	
	A1.3 Caution	
	A1.4 Chance, luck	
	A1.5 Use	
		A1.5.1 Using
		A1.5.2 Usefulness
	A1.6 Physical/mental	
	A1.7 Constraint	
	A1.8 Inclusion/Exclusion	
	A1.9 Avoiding	
A2 Affect		
	A2.1 Affect: Modify, change	
	A2.2 Affect: Cause/Connected	
A3 Being		
A4 Classification		
	A4.1 Generally kinds, groups, examples	
	A4.2 Particular/general; detail	
A5 Evaluation		
	A5.1 Evaluation: Good/bad	
	A5.2 Evaluation: True/false	
	A5.3 Evaluation: Accuracy	
	A5.4 Evaluation: Authenticity	
A6 Comparing		
	A6.1 Comparing: Similar/different	
	A6.2 Comparing: Usual/unusual	
	A6.3 Comparing: Variety	
A7 Definite (+ modals)		
A8 Seem		
A9 Getting and giving; possession		
A10 Open/closed; Hiding/Hidden;Finding; Showing		
A11 Importance		
	A11.1 Importance: Important	
	A11.2 Importance: Noticeability	
A12 Easy/difficult		
A13 Degree		
	A13.1 Degree: Non-specific	
	A13.2 Degree: Maximizers	
	A13.3 Degree: Boosters	
	A13.4 Degree: Approximators	
	A13.5 Degree: Compromisers	
	A13.6 Degree: Diminishers	
	A13.7 Degree: Minimizers	
A14 Exclusivizers/particularizers		
A15 Safety/Danger		

(B) THE BODY & THE INDIVIDUAL
B1 Anatomy and physiology
B2 Health and disease
B3 Medicines and medical treatment
B4 Cleaning and personal care
B5 Clothes and personal belongings
(C) ARTS & CRAFTS
C1 Arts and crafts
(E) EMOTIONAL ACTIONS, STATES & PROCESSES
E1 General
E2 Liking
E3 Calm/Violent/Angry
E4 Happy/sad
E4.1 Happy/sad: Happy
E4.2 Happy/sad: Contentment
E5 Fear/bravery/shock
E6 Worry, concern, confident
(F) FOOD & FARMING
F1 Food
F2 Drinks
F3 Cigarettes and drugs
F4 Farming & Horticulture
(G) GOVERNMENT & THE PUBLIC DOMAIN
G1 Government, Politics & elections
G1.1 Government etc.
G1.2 Politics
G2 Crime, law and order
G2.1 Crime, law and order: Law & order
G2.2 General ethics
G3 Warfare, defence and the army; Weapons
(H) ARCHITECTURE, BUILDINGS, HOUSES & THE HOME
H1 Architecture, kinds of houses & buildings
H2 Parts of buildings
H3 Areas around or near houses
H4 Residence
H5 Furniture and household fittings
(I) MONEY & COMMERCE
I1 Money generally
I1.1 Money: Affluence
I1.2 Money: Debts
I1.3 Money: Price
I2 Business
I2.1 Business: Generally
I2.2 Business: Selling
I3 Work and employment
I3.1 Work and employment: Generally
I3.2 Work and employment: Professionalism
I4 Industry
(K) ENTERTAINMENT, SPORTS & GAMES
K1 Entertainment generally
K2 Music and related activities
K3 Recorded sound etc.
K4 Drama, the theatre & show business
K5 Sports and games generally
K5.1 Sports
K5.2 Games
K6 Children's games and toys

(L) LIFE & LIVING THINGS
L1 Life and living things
L2 Living creatures generally
L3 Plants
(M) MOVEMENT, LOCATION, TRAVEL & TRANSPORT
M1 Moving, coming and going
M2 Putting, taking, pulling, pushing, transporting &c.
M3 Movement/transportation: land
M4 Movement/transportation: water
M5 Movement/transportation: air
M6 Location and direction
M7 Places
M8 Remaining/stationary
(N) NUMBERS & MEASUREMENT
N1 Numbers
N2 Mathematics
N3 Measurement
N3.1 Measurement: General
N3.2 Measurement: Size
N3.3 Measurement: Distance
N3.4 Measurement: Volume
N3.5 Measurement: Weight
N3.6 Measurement: Area
N3.7 Measurement: Length & height
N3.8 Measurement: Speed
N4 Linear order
N5 Quantities
N5.1 Entirety; maximum
N5.2 Exceeding; waste
N6 Frequency etc.
(O) SUBSTANCES, MATERIALS, OBJECTS & EQUIPMENT
O1 Substances and materials generally
O1.1 Substances and materials generally: Solid
O1.2 Substances and materials generally: Liquid
O1.3 Substances and materials generally: Gas
O2 Objects generally
O3 Electricity and electrical equipment
O4 Physical attributes
O4.1 General appearance and physical properties
O4.2 Judgement of appearance (pretty etc.)
O4.3 Colour and colour patterns
O4.4 Shape
O4.5 Texture
O4.6 Temperature
(P) EDUCATION
P1 Education in general
(Q) LINGUISTIC ACTIONS, STATES & PROCESSES
Q1 Communication
Q1.1 Communication in general
Q1.2 Paper documents and writing
Q1.3 Telecommunications
Q2 Speech acts
Q2.1 Speech etc: Communicative
Q2.2 Speech acts
Q3 Language, speech and grammar
Q4 The Media
Q4.1 The Media: Books
Q4.2 The Media: Newspapers etc.
* Q4.3 The Media: TV, Radio & Cinema *and the Internet

(S) SOCIAL ACTIONS, STATES & PROCESSES

- S1 Social actions, states & processes
  - S1.1 Social actions, states & processes
    - S1.1.1 General
    - S1.1.2 Reciprocity
    - S1.1.3 Participation
    - S1.1.4 Deserve etc.
  - S1.2 Personality traits
    - S1.2.1 Approachability and Friendliness
    - S1.2.2 Avarice
    - S1.2.3 Egoism
    - S1.2.4 Politeness
    - S1.2.5 Toughness; strong/weak
    - S1.2.6 Sensible
- S2 People
  - S2.1 People: Female
  - S2.2 People: Male
- S3 Relationship
  - S3.1 Relationship: General
  - S3.2 Relationship: Intimate/sexual
- S4 Kin
- S5 Groups and affiliation
- S6 Obligation and necessity
- S7 Power relationship
  - S7.1 Power, organizing
  - S7.2 Respect
  - S7.3 Competition
  - S7.4 Permission
- S8 Helping/hindering
- S9 Religion and the supernatural

(T) TIME

- T1 Time
  - T1.1 Time: General
    - T1.1.1 Time: General: Past
    - T1.1.2 Time: General: Present; simultaneous
    - T1.1.3 Time: General: Future
  - T1.2 Time: Momentary
  - T1.3 Time: Period
- T2 Time: Beginning and ending
- T3 Time: Old, new and young; age
- T4 Time: Early/late

(W) THE WORLD & OUR ENVIRONMENT

- W1 The universe
- W2 Light
- W3 Geographical terms
- W4 Weather
- W5 Green issues

	(X) PSYCHOLOGICAL ACTIONS, STATES & PROCESSES		
		X1 General	
		X2 Mental actions and processes	
			X2.1 Thought, belief
			X2.2 Knowledge
			X2.3 Learn
			X2.4 Investigate, examine, test, search
			X2.5 Understand
			X2.6 Expect
		X3 Sensory	
			X3.1 Sensory: Taste
			X3.2 Sensory: Sound
			X3.3 Sensory: Touch
			X3.4 Sensory: Sight
			X3.5 Sensory: Smell
		X4 Mental object	
			X4.1 Mental object: Conceptual object
			X4.2 Mental object: Means, method
		X5 Attention	
			X5.1 Attention
			X5.2 Interest/boredom/excited/energetic
		X6 Deciding	
		X7 Wanting; planning; choosing	
		X8 Trying	
		X9 Ability	
			X9.1 Ability: Ability, intelligence
			X9.2 Ability: Success and failure
	(Y) SCIENCE & TECHNOLOGY		
		Y1 Science and technology in general	
*			*Y1.1 Science fiction
		Y2 Information technology and computing	
*			*Y2.1 Hardware
*			*Y2.2 Software
*			*Y2.3 Internet phenomena; sociological and technological
	(Z) NAMES & GRAMMATICAL WORDS		
		Z0 Unmatched proper noun	
		Z1 Personal names	
		Z2 Geographical names	
		Z3 Other proper names	
*			*Z3.1 Corporate brand names
*			*Z3.2 Artist aliases
*			*Z3.3 Names of events, media and trademarked products
*			*Z3.4 Names of NGOs; charity, lobby, and political
		Z4 Discourse Bin	
		Z5 Grammatical bin	
		Z6 Negative	
		Z7 If	
		Z8 Pronouns etc.	
		Z9 Trash can	
		Z99 Unmatched	

Appendix 5. Candidate Words.

NN_1

testimony
track
prosecutor
routine
citizen
pipeline
banking
miss
voting
body
label
hardware
action
gain
mouth
consumption
intervention
avenue
chemistry
manufacturer
prize
summit
shadow
vegetable
expectation
custom
mood
concentration
bear
population
property
mail
relation
youth
company
bull
auction
suicide
belt
strength
indian
drive
egg
employer
touchdown
helicopter
task
base
percent
english
union
crowd
dollar
visitor
patient
command
permit
suspect
offer
being
court
biology
depression
marketplace
cable
button
status
comfort
script
factor
tourist
life
temperature
harassment
scope
rule
corruption
understanding
set
sentiment
disease
text
web
element
song
hundred
app
tradition
match
mayor
roll
question
local
comparison
smile
robot
pitch
other
literature
major

VV

stock
withhold
disguise
compose
perpetuate
attribute
amass
agree
cripple
check
stimulate
digest
curate
meet
progress
bake
date
call
transgender
invoke
entitle
verify
stuff
conduct
tune
navigate
roll
defend
rob
rebuild
establish
reproduce
pose
outnumber
anticipate
mean
chuck
fabricate
prove
mislead
speculate
attach
conceal
nurture
go
enjoy
highlight
stall
unlock
mourn
calculate
veto
spearhead
doubt
rally
renovate
collapse
probe
unload
interpret
liberate
line
blur
park
station
organize
bypass
cruise
lessen
write
disable
utter
confine
reiterate
wield
rotate
present
dot
spark
undergo
chronicle
spend
soften
reliever
hide
satisfy
blend
wound
enact
bond
attest
tell
swap
sponsor
investigate
bleed
plan
stab
think
oppose

NN2

cost
measurement
string
other
shoe
english
editor
form
marijuana
poverty
analysis
funding
activity
legislator
soldier
pirate
affair
hearing
trial
independence
immigration
blue
degree
economics
floor
symbol
trend
planet
territory
asset
agency
friend
fight
coalition
intervention
tragedy
brain
involvement
chair
lady
source
jury
logo
condition
net
palm
dragon
proposal
layer
sense
bullet
professional
volume
focus
capability
album
destination
expert
diversity
uncertainty
facebook
organ
hole
date
press
mass
islam
threat
accuracy
camera
weather
sign
meal
evening
school
silver
protester
ram
consideration
cowboy
maintenance
cub
land
snow
accountability
pass
judgment
climate
incentive
relief
protection
anger
vaccine
self
fee
idea
custody
offender
phenomenon
consultant

Appendix 6. Operationalized combining forms.

The table illustrates how borderline cases have been assessed according to criterion B.4. The focus of attention below is combining forms, but there are also examples of affixes in the list (e.g., *tele*-). The online versions of the dictionaries have been employed. Segments with at least one designation as a combining form (marked in the table as x) in one of the dictionaries have been operationalized as a combining form in the study.

	Merriam-Webster	**Cambridge Dictionary**	**MacMillan Dictionary**
-gate	x	x	x
-ware	-	-	-
-(a/o)holic	x	x	x
-scape	x	x	x
-erati	-	-	-
-topia	-	-	-
robo-	-	-	x
-tron	(x)	-	-
-tronics	-	-	-
-nomics	-	-	-
petro-	x	-	-
perma-	(x)	-	-
-bot	x	x	x
-cracy	x	-	-
-tainment	-	-	-
-(a)thon	x	x	x
-pocalypse	-	-	-
agri-	x	-	-
-(o)naut	-	-	-
tele-	x	x (prefix)	x (prefix)
tragi-	-	-	-
heli-	x	-	-
-copter	x	x	x

Index

amalgamation, 16, 23, 26, 31, 40-41, 61, 133, 139, 142-143, 146, 156, 160, 199
associative blends, 35, 47
binocular rivalry, 201-202, 208
blend pond, 69, 85-86, 91, 97-100, 103, 155
candidate words, 5, 69, 85
Claws 7, 70, 76, 79-82, 109, 241
clipping, 7, 24, 40, 46, 53, 61, 65, 107, 147, 154, 184, 194
clipping compound, 40
cognitive domains, 13
combining forms, 37, 63, 105, 124, 151, 261
complex clippings, 2, 21, 26, 40, 42, 44, 55-56, 65, 124, 148, 213
compounding, 12, 17-18, 34, 37, 39-40, 47, 49, 52, 107, 146-147, 154, 157, 169, 194
conceptual blending, 12-13, 35, 41, 54, 58
conceptual domain, 13, 15, 113, 137, 139, 141, 143, 145, 152, 172, 185, 188
conceptual fusion, 12, 27, 142
conjunctive blends, 35
coordinative, 34-35, 39, 47
counterblend, 35, 142
covert source ambiguity, 159, 184, 186-189, 191-193, 196, 207, 211
determinative, 34
distinguishing sound, 173-174, 176-177
domain proximity, 97, 131, 139-145, 148, 151, 153, 156, 200, 206-207, 212, 215
dual model of blend classification, 200, 203, 210, 212-214
dvandva, 34, 47
endocentric, 33-34, 40, 50-51
exocentric, 33-35
expanded context, 9
figurative meaning, 14, 17, 207, 214
fractomorphemes, 38
full overlap, 28-29, 48, 53, 60, 64, 160, 210
graphemic dependence, 30
IC ratio, 182-183
iconicity, 35, 97, 131, 142-144, 151, 153-154, 156, 199, 200, 206-207, 214-215
indefinite composites, 47
intercalation, 29, 179
long tail, 78-79, 83, 87, 99, 106
metaphor, 11-12, 14, 17, 137, 155, 216
metonymy, 14-17, 137, 141, 144-145, 151, 185
mode ambiguity, 159, 170-171, 173, 177, 196, 200, 207-208, 212
monitor corpus, 69
morphemic status, 22, 37-38, 93, 105, 107, 128, 152
morphemization, 38, 63-64, 124, 131, 152, 154-156, 207, 211, 215-216
multistable meaning construal, 202-203, 206, 213, 215-216
multistable perception, 7, 201-203, 208, 215
near-graphemic blends, 31, 173-175, 177, 210
News On the Web corpus (description), 5, 69
observer's paradox, 57
paradigmatic blends, 35, 51, 137, 212
portmanteau, 1, 47, 52, 122
prototype theory, 18-19, 48, 65-66, 92, 156, 174, 201, 211-212
pseudomorphemic transfer, 97, 131, 146-148, 151, 154, 156, 167, 206, 212
sandwich words, 46
schema, 22, 26, 36-37, 53-54, 147-148, 154, 200, 205, 210, 213-214

selection point, 55, 86, 89
semantic domain, 92, 113, 119, 124, 127, 129, 156, 188-189
semantic neighborhood, 136
semantic shift, 38, 152
seriality, 38, 97, 108, 124-127, 129, 131, 149, 152-156, 205-207, 216
solitary blends, 124-129
sound symbolism, 171
source word ambiguity, 159, 178-181, 183, 189, 194, 196, 202, 207-208
speech error, 2, 40-41, 45-46, 52, 55, 57, 205
splinter, 37, 60
structural fusion, 143, 153
structural profiling, 97, 131, 140, 142-145, 147-148, 151-153, 156, 211-212, 215
Structured Query Language (description), 8, 71
synesthesia, 171
syntactic blending, 13, 40, 46, 61
telescope words, 47
truncation ambiguity, 159-161, 163-165, 167-169, 196, 202, 207-208
truncforms, 5, 7, 86, 89, 91, 104, 113
UCREL, 70, 76, 114-115, 140, 241, 254
uniqueness, 134-135, 153
USAS, 114-119, 121, 123, 131, 139-140, 144-145, 188, 254
wildcard, 76, 106, 112-113, 174, 183, 185, 188, 192
wordplay, 17, 22, 27, 29, 31-32, 38, 60, 64, 132, 137, 140-141, 146, 153, 161, 169-170, 179, 195, 208, 210, 212
Zipfian distribution, 87

Studier i språk och litteratur från Umeå universitet
Umeå Studies in Language and Literature

Publicerade av Institutionen för språkstudier, Umeå universitet
Published by the Department of Language Studies, Umeå University

Redaktion/Editor: Maria Lindgren Leavenworth

Distribuerade av/Distributed by: eddy.se ab
P.O. Box 1310, SE-621 24 Visby, Sweden
E-mail: order@bokorder.se
Phone: +46 498 253900
Fax: +46 498 249789

1. Elena Lindholm Narváez, *'Ese terrible espejo'. Autorrepresentación en la narrativa sobre el exilio del Cono Sur en Suecia*. Diss. 2008.
2. Julian Vasquez (ed.), *Actas del Simposio Internacional "Suecia y León de Greiff (1895–1976)"*. 2008.
3. Dorothea Liebel, *Tageslichtfreude und Buchstabenangst. Zu Harry Martinsons dichterischen Wortbildungen als Übersetzungsproblematik*. Diss. 2009.
4. Dan Olsson, *„Davon sagen die Herren kein Wort". Zum pädagogischen, grammatischen, und dialektologischen Schaffen Max Wilhelm Götzingers (1799–1856)*. Diss. 2009.
5. Ingela Valfridsson, *Nebensätze in Büchern und Köpfen. Zur Bedeutung der Begriffsvorstellungen beim Fremdsprachenerwerb*. Diss. 2009.
6. Per Ambrosiani (ed.), *Swedish Contributions to the Fourteenth International Congress of Slavists (Ohrid, 10–16 September 2008)*. 2009.
7. Therese Örnberg Berglund, *Making Sense Digitally: Conversational Coherence in Online and Mixed-Mode Contexts*. Diss. 2009.
8. Gregor von der Heiden, *Gespräche in einer Krise. Analyse von Telefonaten mit einem RAF-Mitglied während der Okkupation der westdeutschen Botschaft in Stockholm 1975*. Diss. 2009.
9. Maria Lindgren Leavenworth, *The Second Journey: Travelling in Literary Footsteps*. 2010.
10. Niklas Torstensson, *Judging the Immigrant: Accents and Attitudes*. Diss. 2010.
11. Van Leavenworth, *The Gothic in Contemporary Interactive Fictions*. Diss. 2010.
12. Heidi Hansson, Maria Lindgren Leavenworth & Lennart Pettersson (red.), *Regionernas bilder. Estetiska uttryck från och om periferin*. 2010.

13. Anette Svensson, *A Translation of Worlds: Aspects of Cultural Translation and Australian Migration Literature*. Diss. 2010.
14. Mareike Jendis, Anita Malmqvist & Ingela Valfridsson (Hrsg.), *Text im Kontext 9. Beiträge zur 9. Arbeitstagung schwedischer Germanisten, 7.–8. Mai 2010, Umeå*. 2011.
15. Nicklas Hållén, *Travelling Objects: Modernity and Materiality in British Colonial Travel Literature about Africa*. Diss. 2011.
16. Stephanie Fayth Hendrick, *Beyond the Blog*. Diss. 2012.
17. Misuzu Shimotori, *Conceptual Contrasts: A Comparative Semantic Study of Dimensional Adjectives in Japanese and Swedish*. Diss. 2013.
18. Tove Solander, *"Creating the Senses": Sensation in the Work of Shelley Jackson*. Diss. 2013.
19. Helena Eckeskog, *Varför knackar han inte bara på? En studie om arbete med läsförståelse i åk 1–2*. Lic. 2013.
20. Katarina Kärnebro, *Plugga stenhårt eller vara rolig? Normer om språk, kön och skolarbete i identitetsskapande språkpraktiker på fordonsprogrammet*. Diss. 2013.
21. Ingalill Gustafsson, *www.lektion.se – din kollega på nätet. En studie av lärares lektionsförslag i skolämnet svenska (skrivande)*. Lic. 2013.
22. Moa Matthis, *"Take a Taste": Selling Isak Dinesen's Seven Gothic Tales in 1934*. Diss. 2014.
23. Anna Maria Hipkiss, *Klassrummets semiotiska resurser. En språkdidaktisk studie av skolämnena hem- och konsumentkunskap, kemi och biologi*. Diss. 2014.
24. Maria Levlin, *Lässvårigheter, språklig förmåga och skolresultat i tidiga skolår – en undersökning av 44 elever i årskurs 2 till 3*. Diss. 2014.
25. Janet Enever, Eva Lindgren & Sergej Ivanov (eds.), *Conference Proceedings from Early Language Learning: Theory and Practice*. 2014.
26. Eva Lindgren & Janet Enever (eds.), *Språkdidaktik: Researching Language Teaching and Learning*. 2015.
27. Hanna Outakoski, *Multilingual Literacy Among Young Learners of North Sámi: Contexts, Complexity and Writing in Sápmi*. Diss. 2015.
28. James Barrett, *The Ergodic Revisited: Spatiality as a Governing Principle of Digital Literature*. Diss. 2015.
29. Hilda Härgestam Strandberg, *Articulable Humanity: Narrative Ethics in Nuruddin Farah's Trilogies*. Diss. 2016.
30. Berit Aronsson, *Efectos pragmáticos de transferencias prosódicas del sueco al español L2: Implicaciones para la clase de español lengua extranjera*. Diss. 2015.
31. Karyn Sandström, *Peer Review Practices of L2 Doctoral Students in the Natural Sciences*. Diss. 2016.
32. Godelinde Perk, *Julian, God, and the Art of Storytelling: A Narrative Analysis of the Works of Julian of Norwich*. Diss. 2016.

33. Sergej Ivanov, *A Transnational Study of Criticality in the History Learning Environment*. Diss. 2016.
34. Mai Trang Vu, *Logics and Politics of Professionalism: The Case of University English Language Teachers in Vietnam*. Diss. 2017.
35. Parvin Gheitasi, *Say It Fast, Fluent and Flawless: Formulaicity in the Oral Language Production of Young Foreign Language Learners*. Diss. 2017.
36. Yvonne Knospe, *Writing in a Third Language. A Study of Upper Secondary Students' Texts, Writing Processes and Metacognition*. Diss. 2017.
37. Elena Rassokhina, *Shakespeare's Sonnets in Russian: The Challenge of Translation*. Diss. 2017.
38. Anita Malmqvist & Asbjörg Westum (red.), *Nya familjekonstellationer – nya benämningar? Familjen och dess medlemmar i finskan, ryskan, svenskan och tyskan*. 2019.
39. Monica Egelström, *Samma lärare – olika praktiker? En studie av literacy och meningsskapande i grundskolans tidiga ämnesundervisning*. Diss. 2019.
40. Kristina Belančić, *Language Policy and Sámi Education in Sweden: Ideological and Implementational Spaces for Sámi Language Use*. Diss. 2020.
41. Moa Sandström, *Dekoloniseringskonst. Artivism i 2010-talets Sápmi*. Diss. 2020.
42. Ej utkommen/Not published.
43. Nuno Marques, *Atmospheric and Geological Entanglements: North American Ecopoetry and the Anthropocene*. Diss. 2020.
44. Spoke Wintersparv, *Teaching the Reading Experience: Upper Secondary Teachers' Perspectives on Aesthetic Aspects of Literature Teaching*. Diss. 2021.
45. Elena Glotova, *Soundscapes in Nineteenth-Century Gothic Short Stories*. Diss. 2021.
46. Baran Johansson, *Two Languages, Two Scripts: Bilingual and Biscriptal Children with and without Reading Difficulties Read and Write in Persian (L1) and Swedish (L2)*. Diss. 2022.
47. Justin Zelime, *Contrasting Language-in-Education Policy Intentions, Perceptions and Practice: The Use of English and Kreol Seselwa in the Seychelles*. Diss. 2022.
48. Daniel Kjellander, *Ambiguity at Work: Lexical Blends in an American English Web News Context*. Diss. 2022.

Skrifter från moderna språk (2001–2006)

Publicerade av Institutionen för moderna språk, Umeå universitet
Published by the Department of Modern Languages, Umeå University

1. Mareike Jendis, *Mumins wundersame Deutschlandabenteuer. Zur Rezeption von Tove Janssons Muminbüchern*. Diss. 2001.
2. Lena Karlsson, *Multiple Affiliations: Autobiographical Narratives of Displacement by US Women*. Diss. 2001.
3. Anders Steinvall, *English Colour Terms in Context*. Diss. 2002.
4. Raoul J. Granqvist (ed.), *Sensuality and Power in Visual Culture*. 2002. NY UPPLAGA 2006.
5. Berit Åström, *The Politics of Tradition: Examining the History of the Old English Poems* The Wife's Lament *and* Wulf and Eadwacer. Diss. 2002.
6. José J. Gamboa, *La lengua después del exilio. Influencias suecas en retornados chilenos*. Diss. 2003.
7. Katarina Gregersdotter, *Watching Women, Falling Women: Power and Dialogue in Three Novels by Margaret Atwood*. Diss. 2003.
8. Thomas Peter, *Hans Falladas Romane in den USA*. Diss. 2003.
9. Elias Schwieler, *Mutual Implications: Otherness in Theory and John Berryman's Poetry of Loss*. Diss. 2003.
10. Mats Deutschmann, *Apologising in British English*. Diss. 2003.
11. Raija Kangassalo & Ingmarie Mellenius (red.), *Låt mig ha kvar mitt språk*. Den tredje SUKKA-rapporten. / *Antakaa minun pitää kieleni*. Kolmas SUKKA-raportti. 2003.
12. Mareike Jendis, Anita Malmqvist & Ingela Valfridsson (Hg.), *Norden und Süden. Festschrift für Kjell-Åke Forsgren zum 65. Geburtstag*. 2004.
13. Philip Grey, *Defining Moments: A Cultural Biography of Jane Eyre*. Diss. 2004.
14. Kirsten Krull, *Lieber Gott, mach mich fromm… Zum Wort und Konzept „fromm" im Wandel der Zeit*. Diss. 2004.
15. Maria Helena Svensson, *Critères de figement. L'identification des expressions figées en français contemporain*. Diss. 2004.
16. Malin Isaksson, *Adolescentes abandonnées*. Je *narrateur adolescent dans le roman français contemporain*. Diss. 2004.
17. Carla Jonsson, *Code-Switching in Chicano Theater: Power, Identity and Style in Three Plays by Cherríe Moraga*. Diss. 2005.
18. Eva Lindgren, *Writing and Revising: Didactic and Methodological Implications of Keystroke Logging*. Diss. 2005.
19. Monika Stridfeldt, *La perception du français oral par des apprenants suédois*. Diss. 2005.
20. María Denis Esquivel Sánchez, *"Yo puedo bien español". Influencia sueca y variedades hispanas en la actitud lingüística e identificación de los hispano-americanos en Suecia*. Diss. 2005.
21. Raoul J. Granqvist (ed.), *Michael's Eyes: The War against the Ugandan Child*. 2005.
22. Martin Shaw, *Narrating Gypsies, Telling Travellers: A Study of the Relational Self in Four Life Stories*. Diss. 2006.

Umeå Studies in Linguistics (2001–2006)

Publicerade av Institutionen för filosofi och lingvistik, Umeå universitet
Published by the Department of Philosophy and Linguistics, Umeå University

1. Leila Kalliokoski, *Understanding Sentence Structure: Experimental Studies on Processing of Syntactic Elements in Sentences*. Diss. 2001.
2. Anna-Lena Wiklund, *The Syntax of Tenselessness: On Copying Constructions in Swedish*. Diss. 2005.
3. Fredrik Karlsson, *The Acquisition of Contrast: A Longitudinal Investigation of Initial s+Plosive Cluster Development in Swedish Children*. Diss. 2006.

PHONUM (1990–2005)

Publicerade av Institutionen för lingvistik, Umeå universitet (1990–1998) och av Institutionen för filosofi och lingvistik, Umeå universitet (1999–2005)
Published by the Department of Linguistics, Umeå University (1990–1998) and by the Department of Philosophy and Linguistics, Umeå University (1999–2005)

1. Eva Strangert & Peter Czigler (eds.), *Papers from Fonetik –90 / The Fourth Swedish Phonetics Conference, Held in Umeå/Lövånger, May 30–31 and June 1, 1990*. 1990.
2. Eva Strangert, Mattias Heldner & Peter Czigler (eds.), *Studies Presented to Claes-Christian Elert on the Occasion of his Seventieth Birthday*. 1993.
3. Robert Bannert & Kirk Sullivan (eds.), *PHONUM 3*. 1995.
4. Robert Bannert, Mattias Heldner, Kirk Sullivan & Pär Wretling (eds.), *Proceedings from Fonetik 1997: Umeå 28–30 May 1997*. 1997.
5. Peter E. Czigler, *Timing in Swedish VC(C) Sequences*. Diss. 1998.
6. Kirk Sullivan, Fredrik Karlsson & Robert Bannert (eds.), *PHONUM 6*. 1998.
7. Robert Bannert & Peter E. Czigler (eds.), *Variations within Consonant Clusters in Standard Swedish*. 1999.
8. Mattias Heldner, *Focal Accent – f_0 Movements and Beyond*. Diss. 2001.
9. Mattias Heldner (ed.), *Proceedings from Fonetik 2003: Lövånger 2–4 June 2003*. 2003.
10. Felix Schaeffler, *Phonological Quantity in Swedish Dialects: Typological Aspects, Phonetic Variation and Diachronic Change*. Diss. 2005.