Derived nouns and events: Eventuality-related meaning in non-deverbal nominalizations

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Philosophie (Dr. phil.)
durch die Philosophische Fakultät der
Heinrich-Heine-Universität Düsseldorf

vorgelegt von
VIKTORIA CHRISTINE SCHNEIDER
aus
LANGENFELD

Betreuer: Univ.-Prof. Dr. Ingo Plag

Düsseldorf, April 2025

D61

Acknowledgments

I want to thank my supervisor and mentor Univ.-Prof. Dr. Ingo Plag for making this project possible in the first place. He is an outstanding supervisor and researcher. I am thankful for his support, not only professional but also on a personal level. I went through a very dark period of my life, and he did not only support me but also became a guiding hand in overcoming adversities and helping me through the final phase of my dissertation.

I want to thank my mentor Univ.-Prof. Dr. Heidrun Dorgeloh for emotional support. She always had advice for any struggles that were inevitable in the realization of such a project.

I would like to thank my second examiner, PD Dr. Martin Schäfer, for being not only available for the assessment process, but also a team member with valuable input, especially on the computational part of my work.

I would like to thank my colleagues Dr. Sven Kotowski and Dr. Lea Kawaletz and for valuable input, not only for my work, but for the whole project.

Special thanks go to my colleague Dr. Dominic Schmitz. He not only read several parts of this work a hundred times, but I also learned a lot about statistics and creating nicer plots.

Many thanks to my colleagues in the department of English Linguistics at Heinrich Heine University Düsseldorf for their valuable feedback at any stage of my work. Their constant input helped a lot in developing the ideas as they are now realized.

I would like to thank our (former) student assistants for their valuable help in data cleaning and proofreading.

I want to thank my husband, Florian Hofmann, for supporting me over the past years. He always believed in me and the never-ending supply of chocolate was more than helpful. It was not the first or last night we spent working on projects together.

Special thanks go to my parents for always believing in me.

This research is part of the CRC Project C08/DFG Project `The semantics of derivational morphology' - which I am delighted to be a part of.

Abstract

The present work contributes to the general knowledge on word-formation processes by investigating nominalizations with the suffixes *-ee*, *-ment* and *-ation* in English. The word-formation processes with these suffixes are well researched for deverbal bases, as verbs are eventive in nature and the most productive word class for such nominalizations. However, *-ee*, *-ment* and *-ation* also take nominal bases. This dissertation offers the first systematic study on denominal nominalization processes with these three suffixes. Frame semantic analyses showed that the semantic structure of nouns have eventive elements. The word-formation process changes the reference from the the noun to the eventive elements available in the base. Furthermore, computational analyses indicate that the word class of the base is not the most influential factor for a successful nominalization process. The overall results show that noouns are possible bases for eventuality-related nominalizations indicating that a semantic account of such nominalizations is important.

Die vorliegende Arbeit leistet einen Beitrag zum allgemeinen Wissen über Wortbildungsprozesse durch die Untersuchung von Nominalisierungen mit den Suffixen *-ee*, *-ment* and *-ation* im Englischen. Die Wortbildungsprozesse mit diesen Suffixen sind für deverbale Basen bereits erforscht, da Verben ereignishaft sind und somit die produktivste Wortklasse für solche Nominalisierungen darstellen. Allerdings können *-ee*, *-ment* and *-ation* auch Nomen als Basis haben. Diese Dissertation bietet die erste systematische Untersuchung zu solchen denominalen Nominalisierungs-prozessen mit diesen drei Suffixen. Die semantische Dekomposition nominaler Basen zeigt, dass Nomen auch ereignisbezogene Elemente in ihrer semantischen Struktur haben, welche dann benutzt werden können, um die neue Nominalisierung zu bilden. Außerdem zeigen computationelle Methoden, dass die Wortklasse der Basis nur eine geringe Rolle spielt. Die Ergebnisse dieser Arbeit ermöglichen einen tieferen Einblick in die Wortbildungssemantik und bestärken die Rolle der Semantik bei Wortbildungen allgemein.

Contents

List of Figures

List of Tables

List of abbreviations

AIC	Akaike Information Criterion
DL	Dicriminative Learning
DS	Distributional Semantics
DSM	Distributional Semantic Model
GL	Generative Lexicon
LCS	Lexical Conceptual Structure
LDA	Linear Discriminant Analysis
LDL	Linear Discriminative Learning
LNRE	Large-Number-of-Rare-Events Theory
NDL	Naive Discriminative Learning

t-SNE t-distributed Stochastic Neighbor Embeddings

VIF Variance Inflation Factors

Symbols and notational conventions

$\wedge$	conjunction (‘and’)
$\vee$	disjunction (‘or’)
[...]	ellipsis in a quote or corpus attestation
♦	morphologically related
SMALL CAPITAL LETTERS	a. frame attributes
	b. variables in statistical models
italics	a. frame types
	b. examples within the text
	c. technical terms
boldface	example words in context

Chapter 1

Introduction

Dipping into the world of word-formation processes, this work focuses on the process of derivation: More precisely, derivational processes with the suffixes *-ee*, *-ment*, and *-ation*. As most of the research on derivations deals with verbs as bases (e.g., Grimshaw 1990; Alexiadou 2001; Lieber 2016; Plag et al. 2018), the central question of this dissertation is how the derivational process works with nominal bases. Research on derivational processes claims that nouns are also possible bases (e.g., Plag 1999, 2004; Bauer et al. 2013; Fábregas 2024), but no systematic study on the semantics of English nominalizations based on nominal bases has been conducted yet. All derivatives with the three suffixes under investigation in this dissertation create so-called 'eventuality-related nominalizations'.

Eventuality-related nominalizations are nominalizations that denote either a (sub-)eventuality or a participant of an eventuality. The term 'eventuality' here includes events, processes and states (for a fine-grained distinction of the different types of eventualities see, e.g., Bach 1986; Van Valin & LaPolla 2002). The category of eventuality-relatedness is a category which includes several more fine-grained classes. For example, in the literature on nominalizations, eventuality-relatedness is further specified as participant, result, or eventive readings.[1] The different types of readings are merged together in the umbrella term 'eventuality-related'. The examples in (1) and (2) illustrate eventuality-related nominalizations. Example (1) refers to a participant of an eventuality. Example (2) denotes a whole eventuality.

[1] See Lieber 2017 for an overview of the several diverging classification systems for the semantic categorization of nominalizations.

(1) *Salary and bonus information for every* ***employee***. (COCA; see Davies 2008; BLOG, 2012)

(2) *Markham sets down the rules about park* ***befoulment***. (Plag et al. 2018: 474)

The important observation for the two examples is that both nominalizations refer to eventive structures contained in the base word. More precisely, the *employing-action* is denoted by the base *employ*. The nominalization, in turn, denotes a participant in the *employing-action* of the base verb, an *employee*. The example with the suffix *-ment*, in turn, creates a transpositional reading as it denotes the same eventuality as its base verb *befoul*. Hence, an eventuality -related nominalization is a nominalization that denotes either parts of an eventive structure, for example participants in an action, or sub-eventualities like processes, or even a complex eventuality.

The identification of the word class of the base is not a trivial task. First, the word class of a base is often not unambiguously identifiable due to many conversions in English. Determining the direction of conversion, for example, is a verb converted into a noun or a noun converted into a verb, poses a difficult problem for the identification of the word class of a base (see, e.g., Balteiro 2007; Bram 2011, Plag 2018: 87, see also Barbu Mititelu et al. 2023). Second, less ambiguous cases indicate that the majority of English word-formation processes operate on more than one word class (see, e.g., Plag 1999, 2004, 2018: 87, Bauer et al. 2013: ch. 10). Thus, the focus on deverbal nominalizations in the literature on word-formation processes does not provide a complete picture of eventuality-related nominalizations.

The examples given up to this point are all deverbal and related to the eventualities denoted by the verbal base (e.g., Haspelmath 2001; Van Valin & LaPolla 2002; Szabó 2015; Moltmann 2019). Some cases of clearly denominal nominalizations are given in Example (3).

(3) a. *ozonation, sedimentation*
    b. *biographee, debtee*

For instance, the nominal base *sediment* is clearly non-eventive. The resulting derivative from the nominalization process *sedimentation* has, nonetheless, an eventive reading. It is unclear where the origin of the eventive-nature of the denominal derivatives lies. To date, there is no systematic study on denominal nominalizations with eventuality-related readings available. My thesis focuses on non-deverbal, eventuality-related nominalizations with the suffixes *-ee*, *-ment*, and *-ation*.

Different approaches to the semantics of nominalizations might suggest different solutions to the problems posed by the word class of the base in the examples in (3). For example, syntactic as well as as lexicalist morpheme-based approaches can be applied. Syntactic approaches, for example Alexiadou (2001) or Borer (2013), work with functional projections which are responsible for eventive or participant readings of a nominalization. These approaches do not capture the semantic side of nominalizations, i.e., how it is possible to create the reading of the nominalization. In lexicalist morpheme-based approaches, for example the framework of Lieber (2004; 2016), affixes are characterized to come with a semantic representation of their own. Suffixes are interpreted to add feature specifications like +DYNAMIC to induce an eventive semantic reading. The semantic structure of the base and the derivative remains underspecified. For a closer discussion of different approaches to word-formation processes, see Chapter 2.1.

Other approaches, for example word-based approaches or computational approaches are, among others, concerned with the semantic structure of the base and its nominalization. Some word-based approaches see affixes as no linguistic signs on their own (e.g., Koenig 1999; Booij 2010). Complex words, like nominalizations, are handled as words that have meaning. Following the idea of a word-based approach, semantic frames decompose the semantic structure of a word (e.g., Barsalou 1992a,b; Petersen 2007; Löbner 2013, 2014, 2017; Petersen & Gamerschlag 2014). This decomposition allows to represent the semantic structure of a word which then, in turn, can be used to model word-formation processes. In other words, the approach admits the modeling of the semantic structures of bases and their nominalizations. Research on nominalizations in a frame semantic approach shows that a reference shift from base to derivative is dependent on the semantic information given by a base (cf. Kawaletz 2023). The reference-shifting approach used by Kawaletz (2023) for the analysis of deverbal derivatives with the suffix *-ment* successfully adds to the understanding of the semantics of word-formation processes and is extendable to nominalizations with other suffixes and different word classes of bases.

Computational approaches, for example distributional semantics, analyze the semantics of words in vector representations (e.g., Mikolov et al. 2013b,a; Baroni et al. 2014; Boleda & Herbelot 2016; Lapesa et al. 2018; Boleda 2020; Wauquier 2022; Bonami & Guzmán Naranjo 2023). The usage of these models to word-formation processes contribute to the understanding of the semantics of word-formation processes in general (e.g., Lapesa et al. 2018; Wauquier 2022; Bonami & Guzmán Naranjo 2023). For instance, the semantic relation of nominalizations and their bases.

Another computational approach to analyze complex words is that of discriminative learning (cf. Baayen et al. 2011, 2019b). This approach is based on cognitive learning theory (cf. Rescorla & Wagner 1972; Wagner & Rescorla 1972). Discriminative learning simulates the learning process in the mental lexicon. Implementations illustrate, among other findings, that pseudowords have meaning (e.g., Chuang et al. 2021; Schmitz et al. 2021; Schmitz 2022), or can disambiguate similar word forms in their semantic neighborhood (e.g., Schmitz et al. 2023).

Based on previous research and in contrast to syntactic and lexicalist approaches, a word-based approach and two computational approaches constitute the main approaches of the present thesis. The applied approaches, i.e., the reference-shifting approach, and distributional semantics as well as linear discriminative learning, are expanded upon in the following subsections.

1.1 Semantic decomposition

I decided to use a reference shifting approach represented in semantic frames to decompose the semantic structure of nominal bases and model their derivatives. This approach has already proven to be successful for deverbal nominalizations with the suffix *-ment* (Plag et al. 2018; Kawaletz 2023). The main idea of the approach is that the word-formation process induces a reference shift from the reading of the base to the reading of the derivative. In order to reveal the semantic structures of the base which are then, in turn, usable for a reference shift, the semantic structure of the base needs to be decomposed. In other words, the semantic decomposition of a base reveals eventualities and eventive elements, e.g., participants in an eventuality, which are then possibly referred to by the derivative. Another argument in favor of the use of this reference shifting approach is that it is extendable to nominalization processes with different suffixes as well as to different word classes of the bases. The main question for the application of the reference shifting approach is:

> With nominal bases requiring eventive structures to successfully shift the reference from the noun to the nominalization, how can this shift be modeled?

The aim of the first part of my dissertation is to analyze the semantic structures of nouns as bases by decomposing their semantics and to describe the induced reference shifts that create the reading of eventuality-related nominalizations. Non-eventive nominal bases need deeper semantic decomposition than verbs in order to reveal the eventive structures that are used for word-formation processes. The semantic decomposition is formalized in semantic frames illustrating the reference shift from base to derivative in Chapter 4.

1.2 Computational methods

The second part of my dissertation presents two computational approaches, distributional semantics using a pre-trained *fastText* implementation (Bojanowski et al. 2016; Mikolov et al. 2018) and linear discriminative learning (cf. Baayen et al. 2016, 2019b). The semantics of words are represented in so-called word vectors. Broadly speaking, the semantics of a word are computed via their co-occurrences with words or forms. The computational output represents the semantics of a word in vectors in several dimensions. The resulting word vectors are then used for further analyses. Different measures, for example, cosine similarities or the Euclidean distance of two vectors, are used to compute the semantic similarities of words. The distributional approach has already proven to be a useful tool for word-formation processes (see, e.g., Lapesa et al. 2018; Wauquier et al. 2018; Huyghe & Wauquier 2020; Kotowski & Schäfer 2023). The main question for the computational analyses develops from the idea that deverbal derivatives should be more similar to their verbal bases than denominal derivatives to their nominal bases. That is, verbs are eventive in nature and thus easier identified as the base compared to nouns as bases. In other words, the idea is that deverbal derivatives show higher levels of similarity with their verbal bases, as verbs are directly denoting eventualities and nouns need further semantic decomposition to reveal the eventive elements referred to by a derivative. Hence, the question is:

> Are denominal derivatives as similar to their nominal bases as deverbal derivatives to their deverbal bases?

Cosine similarities of the word vectors for bases and derivatives function as a measure of the similarities of derivatives and bases. The cosine similarities then entered a beta regression model to investigate whether the word class of the base is an influential factor.

A second analysis with the word vectors was performed: *t*- distributed Stochastic Neighbor Embeddings and linear discriminant analyses. The idea behind these analyses is to see whether visible clusters are observable in the data. The assumption is that a visible difference between deverbal and denominal nominalizations exists.

Concluding, the applied approaches point into the direction that the word class of a base word is not the central factor for a successful word-formation process. It rather seems that previous claims about the importance of semantic restrictions of a word-formation process are far more important than the syntactic category of the base (e.g., Barker 1998; Plag 2004). The creation of an eventuality-related nominalization is possible if the base provides the eventive elements required for the word formation process.

The next chapter, Chapter 2, provides the necessary theoretical background for the studies performed. Chapter 3 illustrates the data retrieval as well as methodological considerations. Chapters 4 and 5 deal with the semantic decomposition and the derivational process in a reference shifting approach modeled in semantic frames. Chapter 6 contains the analysis of the *fastText* implementation, and Chapter 7 presents the linear discriminative learning implementation. Chapter 8 compares both computational implementations and discusses the findings. Chapter 9 combines the implications from both approaches, the reference shifting approach and the computational implementation. Chapter 10 concludes my thesis.

Chapter 2

Challenges for the interpretation of eventuality-related nominalizations

Eventive nominalizations are words which are built by morphological processes such as derivation or conversion.[2] The resulting word formed by these nominalization processes is a noun. The derivational process for nominalizations can, for example, be created by several suffixes. These nominalizing suffixes can be divided into three different categories: event nominalizations with the suffixes *-al, -ance, -ation, -ing, -ment,* nominalized person nouns with the suffixes *-ant, -arian, -ee, -eer, -er, -ese, -ess, -i, -ist*, and abstract notions, collectives, or locations with the suffixes *-dom, -ery, -hood, -ism, -ity, -ness, -ship, -age* (cf. Plag 2018: ch. 4). This dissertation focuses on event nominalizations, henceforth eventuality-related nominalizations, with three suffixes: the person noun suffix *-ee* and two event nominalizers, *-ment* and *-ation*.

The base words for eventuality-related nominalizations can either be verbs, nouns, adjectives, or other word classes (e.g., Plag 1999, 2004; Bauer et al. 2013; Plag 2018). However, only verbal bases for several nominalizations processes are well researched. The other word classes which

[2]An earlier version of parts of Section 2.1, Section 2.2.1 and Section 2.3.1 have already been published in Schneider 2023. An earlier version of parts of Section 2.3.2 have already been published in Kotowski et al. 2023.

can serve as bases for derivational processes are not as well researched and no systematic study is available for English nominalizations based on other base words than verbs. Furthermore, research on nominalizations found that not only the word class of a base plays a role for word-formation processes, but the semantics of the base and the resulting nominalization also play a central role for these processes. This chapter will first show how semantics interact with derivational processes followed by a description of methods applied in the research on derivational semantics. Previous research on the affixes investigated in this dissertation will be summarized.

2.1 Eventuality-related nominalizations

As this dissertation analyzes the semantics of eventuality-related nominalization, the term 'eventuality-related' has to be defined first. Eventualities are complex semantic entities of which participants, like AGENTS and PATIENTS, or even sub-eventualities like RESULT-STATES, are part of. To refer to such participants, sub-eventualities and eventualities also in the broader sense, including processes and states, the term 'eventive elements' will be used. For the creation of the meaning of an eventuality-related nominalization, either the eventuality, one of its sub-eventualities, or one of its participants from the semantic representation of the base word is required (e.g., Barker 1998; Plag 2004; Plag et al. 2018; Kawaletz 2023). Table 2.1 illustrates possible eventuality-related nominalizations based on the verb *employ*.

TABLE 2.1: Derivatives based on the verb *employ*.

derivative	eventive element
employer	AGENT
employee	PATIENT
employment	EVENTUALITY

It can be observed that the nominalizations with the different nominalizing suffixes, *-er*, *-ee*, and *-ment*, refer to different eventive elements from the eventuality denoted by the base. The eventive elements are inherited from the semantic representation of the base verb (Löbner 2013; Plag et al. 2018; Kawaletz 2023). The derivative with the suffix *-er* denotes the AGENT, and the derivative with the suffix *-ee* denotes the PATIENT of the *employing-action*.[3] The derivative with the suffix *-ment* denotes the whole eventuality.

As illustrated, the derivational process of eventuality-related nominalizations is straightforward with verbs, as verbs are eventive in nature (e.g., Van Valin & LaPolla 1997, 2002; Haspelmath 2001; Szabó 2015; Moltmann 2019). Verbs denote eventualities which can be decomposed into several eventive elements and eventive structures that are relevant for the construal of the reading of the derivatives: The semantic representation of the verb *employ* in Table 2.1 consists of an eventuality and its participants. The word-formation process can refer to one of these eventive elements to create the meaning of the derivatives.

Turning to nominal bases, nouns are not always as straightforwardly eventive as verbs. Especially nouns which denote concrete objects or artifacts seem not to be eventive at all. Nonetheless, it is possible to take a noun like *biography* as the base word for a nominalization to form the derivative *biographee*. The derivative shifts the reference from the book or its content to a person, the person whose life is described. In order to achieve this change in meaning, a participant of an eventuality which fulfills the requirements of the word-formation process with the suffix *-ee* is needed. This referential shift is similar to the verbal process in the example derivative *employee* which denotes the PATIENT of the *employing-action*.

Nevertheless, although the referential shift from the object *biography* to a participant of an eventuality in the derivative *biographee* seems similar to the process with verbal bases, it is unclear where the eventuality with the required participant comes from. Thus, the process does not seem to be as straightforward with nominal bases as it is with verbal bases. The problem has been mentioned rather vaguely in the literature:

[3] An alternative analysis would say that an *employee* is someone who is in a state of being employed. A detailed analysis of *employee* is beyond the scope of this study. I use this derivative only for exemplification of my approach.

> The verbal relation is implied by context or can be inferred from the nature of the non-verbal base. [...] Such interpretations follow from the sort of activities that the base nouns could conceivably be involved in. (Bauer et al. 2013: 233)

> [...] a crucial ingredient in my analysis is that denominal action nouns lack any head that describes the eventuality, and all differences between them and deverbal nominalisations follow from there. (Fábregas 2024: 238)

It remains unclear if and where an eventuality in a nominal base is embedded in the semantic representation of the nominal base. More precisely, the semantic side of the derivational process of an eventuality-related nominalization with a nominal base remains undefined.

One approach that does incorporate eventualities in lexical entries in nouns is Pustejovsky's 1996 Generative Lexicon (for different approaches on eventualities in nouns see, e.g., Larson 1998; Winter & Zwarts 2012). For example, for the artifact noun *biography*, the approach by Pustejovsky would posit two eventualities in the lexical entry. In one eventuality, the artifact came into being as a result of the writing of the book (AGENTIVE-QUALE). The second eventuality is what Pustejovsky (1996: ch. 6) has labeled 'qualia' (more specific: TELIC-QUALE) and other people have called 'affordance' (see, e.g., Löbner 2013: 315). Affordances describe what artifacts are used for and how they function as an integral part of the meaning of the noun. The second eventuality would thus describe the presumed usage of a book, that is, the reading of the book. Both eventualities relate to the idea by Bauer et al. (2013) that the base noun could be involved in an activity. The analysis by Pustejovsky does not go into detail about the eventive structures. For the purpose of this study, the eventualities in the nominal base need to be further decomposed as the word-formation with a nominalizing suffix can refer to a whole eventuality, an embedded sub-eventuality, or a participant taking part in the eventuality. Hence, further decomposition of the eventuality is needed to identify the eventive elements required for the word-formation process.

Another often used approach for the analysis of the word-formation process of nominalizations is distributed morphology (see, e.g., Bobaljik 2013 for an overview). In distributed morphology, word-formation is seen as part of syntax, and morphology is seen as the realization of the process (Bobaljik 2013). The formalization of word-formation processes is variable as the success or failure of the analyses depends on the embedding of hierarchical structures of several ongoing processes into the grammar system. These hierarchical structures might also include allomorphy and other morphological interactions (e.g., Bobaljik et al. 2000; Embick 2003; Harbour 2007; Embick 2010; Arregi 2012; Bobaljik 2012; Bonet & Harbour 2012; Moskal 2015a,b). For word-formation processes, two premises are said to be important (see, e.g., Chomsky 1970; Pesetsky 1996; Marantz 1997): Morphologically irregular nominalization is limited to root nominalization, and lexical decomposition is syntactic (cf. Bobaljik 2013).

The meaning of the base and the outcome of a word-formation process are not considered in distributed morphology. The approach analyzes the syntactic possibility of a nominalization, more precisely, the possibility of a newly derived word to be used in the syntactic structure. However, this handling of nominalizations as purely syntactic and morphological possibilities cannot explain why denominal nominalizations which are based on non-eventive nouns like *biographee* or *pixelation* are possible. Only the possibility of these nominalizations to be morphologically built and syntactically used is of interest in distributed morphology, not their meaning. The present thesis aims to find a way to analyze the nominalization process in general as well as working out an explanation for which features a given word must exhibit so that it can be used as a morphological base for the word-formation process. This analysis needs a semantic perspective and not a purely syntactic/morphological analysis. Furthermore, previous research on nominalizations from a semantic perspective (e.g., Barker 1998; Plag 2004; Lieber 2016; Plag et al. 2018; Kawaletz 2023) has shown that semantic constraints apparently play a role for word-formation processes. Semantic approaches on word-formation include, but are not limited to, semantic decomposition of the base and the derivative to look into the semantic elements of the process (see Section 2.2).

Lexicalist morpheme-based approaches are another option to analyze nominalizations but from a different perspective, a compositional one (see, e.g., Lieber 2004, 2016). Nominalizing affixes are analyzed to have a semantic representation of their own which they contribute to the word-formation process. For example, an affix can have a feature specification

+DYNAMIC to mark eventive semantics. Such analyses lead to underspecifications of the semantics of the resulting nominalizations leaving the relationship of base and derivative remains undefined.

All the approaches described so far do either focus on the general make-up of deverbal nominalizations or on nouns, i.e., whether nouns are eventive or not. None of these approaches go into detail about the semantics of denominal nominalizations. This lack of knowledge about denominal nominalizations on the one hand and their relation to deverbal nominalizations on the other hand gives rise to the following questions on the semantics of denominal eventuality-related nominalizations:

RQ1 Where in the semantic representation of the nominal base word can we find eventive elements that enable an interpretation of denominal eventuality-related nominalizations, even if the nominal base is non-eventive or not straightforwardly eventive?

RQ2 Do denominal and deverbal eventuality-related nominalizations behave differently with regard to

a. the decomposition of the semantic structure of the base?
b. the possibility of the denominal nominalizations to refer to the same eventive elements as deverbal nominalizations?
c. the importance of the word class of their semantic base?

In order to investigate these questions, approaches looking at the semantics are needed. More precisely, the semantics of the nominal base and the semantics of the nominalizations need detailed inspection. This being the case, methods from the field of 'derivational semantics' (e.g., Plag 2004; Lieber 2004, 2016; Löbner 2014; Plag et al. 2018; Wauquier et al. 2018; Wauquier 2020; Kawaletz 2023) have shown to be successfully applicable for the aims of this dissertation. The next section will describe the derivational semantic approaches which are used in this dissertation to analyze eventuality-related nominalizations.

2.2 Methods in derivational semantics

Kotowski & Plag (2023) define derivational semantics as follows:

> These different facets of the mapping from form to meaning are related to a set of arguably broader problems of derivational morphology. One concerns affix-base interaction and the degree to which the semantics of a derivative can be modeled compositionally via the semantics of the base and the semantics of the affix. (Kotowski & Plag 2023: 3)

Several problems of derivational morphology can be tackled by methodological approaches including semantics. Phenomena which can be investigated with derivational semantic methods are, for example, polysemy, affix rivalry, or form-meaning mismatches (Kotowski & Plag 2023).

The field of derivational semantics can be divided into different subfields of research as pointed out in the following (from Kotowski & Plag 2023):

- aspectual properties of conversion
- mismatches of zero-derived nouns and verbs
- semantic rivalry
- semantic coherence of morphological categories
- organization of semantic knowledge
- distinction of inflection/derivation

Different perspectives can be taken in derivational semantics to investigate linguistic phenomena: One can either look from a word-based perspective (e.g., frame semantics, Section 2.2.1), a morpheme-based perspective (e.g., Lieber 2004, 2016), or a distributional perspective (distributional semantics, Section 2.2.2). For the work at hand, the chosen perspective is concerned with the relation of words and their forms and meanings in the lexicon (cf. Kotowski & Plag 2023).

A central question of derivational semantic research is: How does a derivative get its meaning? In other words, which semantic information from the base word is potentially transferred to the derivative? In the case of denominal eventuality-related nominalizations, the interaction of base and derivative is of crucial importance. That is, like deverbal nominalizations, the meaning of denominal derivatives is based on the denotation of the base. Consequently, an approach which decomposes the semantic structure of the base is required to successfully depict the semantic change. Furthermore, the semantic similarity between nominal base and denominal derivative is of interest due to the assumption that their semantics are related, or rather that the reading of the derivative is dependent on the eventive elements in the semantic structure of the base.

A semantic decompositional approach is needed to decompose the semantic structures of base and derivative. More specifically, the approach will be used to describe which eventive elements are possibly used for denominal nominalizations and where they are to be found in the semantics of the base by decomposing the semantic structure of the base. Frame semantics (see, e.g., Petersen 2007; Löbner 2013, 2014; Petersen & Gamerschlag 2014; Löbner 2017, 2018; Plag et al. 2018; Kawaletz 2023) is an approach which can describe derivation as the possibility to induce referential shifts. The architecture of this approach will be used in this dissertation for the semantic decomposition.

The assumption that the denotation of the base is interrelated with the denotation of the base, as the applied decompositional approach suggests, will be tested by computational models. For this aim, distributional semantics (see, e.g., Boleda & Herbelot 2016; Lapesa et al. 2018; Boleda 2020; Huyghe & Wauquier 2020; Wauquier 2020, 2022) will be used to compare the semantics of denominal derivatives and their bases to reinforce their semantic similarity observed by the findings of the referential shifts (e.g., Löbner 2013). In other words, the reference shifting approach suggests that the eventive elements required for the derivational process are already in the semantic structure of the base and, thus, the base and the derivative should be semantically similar. Furthermore, deverbal and denominal derivatives are compared to each other to see whether they behave similarly with regard to their semantic similarity with the base.

Another computational approach, discriminative learning (cf. Baayen et al. 2016, 2019b, Section 2.2.3), will be used for the analysis of eventuality-related nominalizations and their bases. This approach is included in this dissertation for three reasons: First, the semantics of words are represented by word vectors estimated by a method relying on cognitive learning models, which makes a comparison of the results of such an implementation with results from a different computational implementation extremely interesting. Second, discriminative learning does not only concern the semantics of words in their contexts but also accounts for the resonance of a word with the whole mental lexicon. The interpretation of the denominal nominalization in a simulated mental lexicon is especially interesting as most of the derivatives in the data are neologisms. Moreover, not only the semantics of the derivatives in isolation are measured, but also their resonance with the simulated mental lexicon. Third, the mathematical operations behind the approach are more easily interpretable by the researcher in comparison to deep learning networks and hence more understandable.

2.2.1 Frame semantics

The semantic frames used in this dissertation are based on Barsalou's (1992a, 1992b) theory of cognitive frames. The general idea behind semantic frames is that meanings are concepts, and that such concepts in human cognition are stored as 'frames' (e.g., Petersen 2007; Löbner 2013, 2014; Gamerschlag et al. 2014). Research on deverbal nominalizations has shown that frame semantics is a useful tool to analyze the semantic representation of a base word and its derivative in a unified format (e.g., Löbner 2013; Schulzek 2014; Kawaletz & Plag 2015; Plag et al. 2018; Schulzek 2019; Kawaletz 2023). The usage of semantic frames for derivation is built on the assumption of referential shifts (cf. Löbner 2013). That is, the newly created reading of a derivative is built by a shift of meaning from the denotation of the base word to an element included in the semantic structure of the base. Hence, decomposition of the semantics of a base is needed to identify the elements in their semantic representation to be used by the derivational process to shift the meaning from base to derivative. The analysis of the semantics of denominal derivatives and their nominal bases applied in this thesis will decompose the semantics of the bases to indicate the reference shifts to the elements denoted by the derivative.

The central building blocks of a frame semantic analysis are attribute-value structures as known from other frameworks (e.g., HPSG: Pollard & Sag 1994; Riehemann 1998; Koenig 1999; Müller 2015; the inclusion of phonological, syntactic and semantic properties: Sag & Wasow 1999; Sag 2012; and other constraint-based approaches to grammar: see Bonami & Crysmann 2016 for an overview). Attributes describe a concept and each attribute then takes a specific value (Löbner 2013: ch. 12). For example, the sentence *John employed Paul* describes an employing-eventuality that is specified for the attributes AGENT with the value *John* and PATIENT with the value *Paul*. The semantic representation of the verb *employ* can be depicted in an attribute-value-matrix (AVM) as in Figure 2.1:

$$\boxed{0}\begin{bmatrix} \textit{employing-action} \\ \text{AGENT } \boxed{1}\ \textit{John} \\ \text{PATIENT } \boxed{2}\ \textit{Paul} \end{bmatrix}$$

$$\text{REF} = \boxed{0}$$

FIGURE 2.1: Example AVM for the verb *employ*.

The AVM describes an *employing-action* as the semantic representation of the verb *employ*. The action is specified for two attributes, an AGENT and a PATIENT. These two attributes are further specified with the values *John* and *Paul*. The numbers in the AVM are indices which serve as labels for and reference to the elements in the frame. The index $\boxed{0}$ labels the *employing-action*, the index $\boxed{1}$ labels the AGENT, and the index $\boxed{2}$ labels the PATIENT. Reference of the lexeme is specified by the REFERENCE-attribute (REF). For the lexeme *employ*, reference is on the whole event, i.e., on node $\boxed{0}$. Frames are recursive as a value can have further attributes which then have further values (Löbner 2013: ch. 12).

Moving on to the description of derivatives which can be participant-denoting in semantic frames, a shift of reference is necessary from the action itself to a specific participant taking part in the eventuality denoted by the base. This referential shift of the base word to the derivative is indicated in the REFERENCE-attribute (REF) at the bottom of the semantic frame representation (cf. Plag et al. 2018, Kawaletz 2023). Figure 2.2 depicts an AVM for the derivative *employer*.

$$\boxed{0}\begin{bmatrix} \textit{employing-action} \\ \textsc{agent}\ \boxed{1}\ \textit{John} \\ \textsc{patient}\ \boxed{2}\ \textit{Paul} \end{bmatrix}$$

$\textsc{ref}=\boxed{1}$

FIGURE 2.2: Example AVM for the derivative *employer*.

In order to form the derivative *employer* from the base verb *employ*, the reference shifts from the *employing-action*, indexed with $\boxed{0}$, to the AGENT-attribute, indexed with $\boxed{1}$. Hence, the meaning of the derivative describes the AGENT of the *employing-action*, *John*, and not the whole action of the verb *employ*. The derivative *employee*, on the other hand, would shift the reference of the base to the PATIENT-attribute *Paul*, indexed with $\boxed{2}$.

The extension of the approach of referential shifts in semantic frames to denominal nominalizations is straightforward. Frames allow for flexible zooming in and out concerning the semantic representation of a word (e.g., Gamerschlag et al. 2013; Naumann 2013; Kawaletz 2023). For example, it is possible to include world knowledge into the frame semantic representation of a word (e.g., Andreou 2017a). This flexibility to include more or less information in the representation if necessary is needed for the analysis of nominal bases. The eventive elements required for the referential shift induced by the derivational process might be deeper embedded in the semantics of the noun (cf. Pustejovsky 1996: ch. 6). Furthermore, the flexibility of frames allows to incorporate knowledge, for example, about the purpose of a human-made artifact or concept, e.g., *a book* or *a pixel*.

The derivational processes can induce referential shifts to different elements denoted by the semantic representation of the base (cf. Kawaletz 2023). These different shifts are formalized in lexical rules (see, e.g., Plag et al. 2018; Schulzek 2019; Kawaletz 2023). The formalization in such rules is, again, similar to lexical rules found in other theoretical approaches as HPSG (e.g., Pollard & Sag 1994; Riehemann 1998; Koenig 1999; Müller 2015) or Lieber's Lexical Semantic Framework (Lieber 2004, 2016). The outcome of the application of a lexical rule leads to a new lexical entry which is related in form and meaning to the input (cf. Sag & Wasow 1999). The derivational process with one suffix can lead to several such rules.

These rules can then be organized in an inheritance hierarchy for indicating and summarizing the possible meaning shifts by a derivational process (e.g., Plag et al. 2018; Kawaletz 2023) similar to other approaches (Riehemann 1998; Koenig 1999; Desmets & Villoing 2009; Booij 2010; Tribout 2010; Bonami & Crysmann 2016).

The frame semantic analyses in this dissertation will illustrate different referent shifts found for denominal eventuality-related nominalizations with the suffixes *-ee*, *-ment* and *-ation*. Inheritance hierarchies will be used to organize the different lexical rules posed by the possibility of inducing different shifts in the generalization sections for each individual suffix under investigation in Chapter 4.

2.2.2 Computational methods: Distributional semantics

Computational methods, like distributional semantics, can be used for semantic analyses. The underlying hypothesis of distributional semantics is the so-called 'distributional hypothesis' (see, e.g., Harris 1954; Boleda & Herbelot 2016; Boleda 2020). This hypothesis states that a difference in meaning is represented in a difference in distribution. Hence, if words do not occur together in the same contexts, the semantics of these words are expected to be different. On the other hand, if two words are often used together, their semantics are expected to be similar.

Following the idea of the distributional hypothesis, the similarity and dissimilarity of words in context can be captured in so-called word vectors which, in turn, reflect the semantics of a word. Word vectors are created via the occurrences of words in context with the word the vector is computed for, i.e., the target word. That is, if a word occurs in the context, the reading of the target word is associated with the word in context. With word vectors, each target word is represented by a string of numbers, representing the co-occurrences of the target word with other words in the context. Words can then be compared to each other by comparing their respective vectors.

The sentences in (1) serve as a toy example. The target words, for which word vectors are computed, are *dress* and *banana*.

(1) a. *dress*: *The dress was in the closet, with the t-shirt and the skirt.*
b. *banana*: *The banana was as tasty as the apple and the peach.*

The words *t-shirt* and *skirt* appear in the context of *dress*, and the words *apple* and *peach* occur as context for the word *banana*. There are different ways to obtain word vectors (e.g., Boleda & Herbelot 2016; Boleda 2020, for an overview). The simplest way is to count word occurrences (for a critical discussion of why predicting word vectors is better than counting words, see, e.g., Baroni et al. 2014). Table 2.2 shows a toy count for the target words *dress* and *banana* and the context words *t-shirt, skirt, apple, peach* and, *cat*. The words *dress* and *skirt*, for example, co-occur about fifty times together in contexts, whereas *dress* and *peach* do not occur together. The word *banana* is, in the toy example, used in the same contexts as the words *apple* and *peach* but not with the words *skirt* or *cat*. The row of the counts of occurrences for the target words with the context words constitutes the words' vector.

Table 2.2: Toy example of a computation of word vectors for the target words *dress* and *banana*.

	t-shirt	skirt	apple	peach	cat
dress	33	50	2	0	0
banana	4	0	46	38	0

Obviously, this toy calculation is not representative for the semantics of a word. Distributional semantic methods operate on large corpora as the basis to compute word vectors. The final matrix of vectors computed from the corpus is a so-called 'vector space'. The word vectors of the target words are computed on the basis of the whole corpus. The word vectors for *dress* and *banana* in the toy example consist of five dimensions as five words were checked for co-occurrences. However, more than five dimensions are necessary for a reliable word vector, i.e., a valid computation of the semantics of a word. How many dimensions are needed and how these dimensions are computed depends on several factors.

There are different approaches for the computation of semantic word vectors, i.e., how to deal with the context information for the computation. For example, continuous bag of words (CBOW, Mikolov et al. 2013b,a), skip-gram (e.g., Mikolov et al. 2013b,a), or instance vectors (Lapesa et al. 2018), to name only a few. The difference in these methods for the computation of vectors lies in the way they incorporate the context of a given word. CBOW predicts the target word by the sum of words in its context, whereas skipgram predicts a target word based on the individual words nearby. Instance vectors are computed for each token of a target word by calculating the average of the vectors of the words in the context of the target word token.

The choice of a computational model for the prediction of target word vectors is of importance. Different computational models can make use of different underlying algorithms. Some of these models offer pre-trained data like, for example, *fastText* (Bojanowski et al. 2016; Mikolov et al. 2018), *Word2Vec* (Mikolov et al. 2013b,a), or FRACSS (Marelli & Baroni 2015). It is also possible to make use of the algorithms provided by these models to train a new model. There are also further approaches (see, e.g., Landauer & Dumais 1997; Jones & Mewhort 2007; Shaoul & Westbury 2010, to name only a few) for obtaining word vectors.

Two further factors have to be decided on: the dimensionality and the context window of semantic word vectors. The dimensionality depends on the research question. For example, *fastText* models commonly use 300 dimension. The context window, i.e., how many content words on the left and on the right of the target word are included in the computation, is variable. Additionally, the decision of the inclusion or exclusion of words in the window, e.g., whether to use only content words, or all words, or also grammatical information, is dependent on the research question and the method of choice. There are many different models and approaches available, and the implementation of new ideas into models is constantly developing. The implementation used in this dissertation are described in Chapter 6.

Distributional Semantics has already been used to investigate different research questions in linguistics (for a detailed overview, see, e.g., Boleda & Herbelot 2016; Boleda 2020). For the present thesis, the work on derivational semantics is of special interest. For example, Wauquier et al. (2018), Huyghe & Wauquier (2020), and Wauquier (2020, 2022) investigated meaning differences in derivatives in French. The studies found that distributional semantics can help to distinguish the readings created by different suffixes and can show where suffixes cluster to create closely-related readings, for example, agentive readings. As the present study deals, among others, with different nominalizing affixes, distributional semantics can show whether the affixes behave differently with regard to their semantic similarity of the base words with different word classes as bases. Distributional methods can also be helpful to disambiguate different readings of a word, including derivatives and affixes (for more on affix polysemy, see, e.g., Lieber 2004; Bauer et al. 2013; Lieber 2016; Plag 2018), or to deal with different base types (e.g., Kotowski & Schäfer 2023). The distributional approaches used in this dissertation will compare nominalizations based on nouns and on verbs to see whether both types of nominalizations are comparably similar to their bases.

Several different measures are available to analyze the similarities and dissimilarities of word vectors and, thus, the semantic similarities and dissimilarities of the semantics of words. The studies in this dissertation use cosine similarities to determine the similarity of a derivative and its base word, following previous studies on derivational semantics (cf. Sitikhu et al. 2019; Huyghe & Wauquier 2020). A higher cosine similarity reflects a higher similarity of the two compared words.

The eventuality-related nominalizations in this dissertation are analyzed with different distributional methods. For this study, one major issue for the work with word vectors is that most of the denominal nominalizations are low in frequency, some are even hapax legomena. Resulting from this fact, a model which can estimate new word vectors on the basis of already existing ones is needed. I decided to use a pre-trained vector space due to the effectiveness of the already performed distributional analysis. In theory, the training of new vectors per Word2Vec (Mikolov et al. 2013a,b) would have been an alternative. However, the low frequency of many of the derivatives under investigation might result in incomplete and unreliable semantic representations in a word vector.

A small pilot study was performed using FRACSS (Marelli & Baroni 2015). With FRACCS, it is possible to get an individual vector for the base word and one for the investigated suffix. The word vector for the nominalization is then the base word vector plus the suffix vector. The idea is that language is decomposable into smaller components, i.e., bases and suffixes, to create new readings, i.e., derivatives. For example, the vector for *bananas* would be the word vector $\overrightarrow{banana}$ plus the plural suffix vector $\overrightarrow{plural}$ resulting in the vector $\overrightarrow{bananas}$. The composition of derivatives by their base and suffix is especially effective to estimate word vectors for low frequent nominalizations. More precisely, training the model with existing nominalizations and receiving a vector for the suffix enables new compositions of nominal bases and the suffix. Unfortunately, the pilot study with the suffix *-ee* was not successful. In the pre-trained data set of FRACCS only a few *-ee*-formations of my deverbal training data were contained. Hence, training for a reliable suffix vector for *-ee* was not possible. Without a reliable suffix vector, the composition for the denominal derivatives by summing up the word vector and the suffix vector would not end in reliable vectors useful for further steps of analyses.

Another method that allows for the creation of new word vectors is *fastText* (Bojanowski et al. 2016; Mikolov et al. 2018). *FastText* (Bojanowski et al. 2016) is an open source library with an implementation for python and pre-trained vector spaces. For the creation of the vectors for the low frequent nominalizations in this dissertation, a pre-trained model containing subword information was used. Due to the inclusion of this subword information via character n-grams (also called n-graphs) in the implementation, it is possible to compute new word vectors for words that are not in the pre-trained *fastText* vector space. This is crucial for the distributional analysis as many derivatives, especially the denominals, are often relatively low in frequency and therefore not found in pre-trained vector spaces. The vector for the low frequent nominalization *biographee*, for example, contains the information of the n-grams, here exemplary trigrams, *#bi/bio/iog/ogr/gra/rap/aph/phe/hee/ee#*. The resulting vector is thus based on the semantic information given by the smaller parts of the words. The computational details for the analysis of eventuality-related nominalization are given in Chapter 6.

Not only *fastText* including subword information was used for the estimation of word vectors, but also a second computational model: linear discriminative learning. The next section will explain the idea of discriminative learning in detail, preserving the technical details for Chapter 7. The comparison of two different computational approaches, i.e., *fastText* and discriminative learning, is made to see if both show similar results, allowing for insights on derivational semantics from two different computational perspectives: *fastText*, a deep learning perspective, and linear discriminative learning, a discriminative-cognitive perspective.

2.2.3 Computational methods: Discriminative learning

Discriminative learning is a cognitive-based approach, in contrast to models like FRACCS and *fastText* which are deep learning approaches. The estimation of word vectors from a discriminative perspective is cognitively more plausible or, at least, better interpretable (cf. Baayen et al. 2016, 2019b). Besides providing information on words in isolation, discriminative learning also includes holistic information on the entire mental lexicon as well as on interrelations of its entries. Discriminative learning allows us to compare implementations modeling whole words in Chapter 7 to implementations of other computational models like *fastText* (Chapter 6). Implementations in which the information about the suffix is added to the derivative, and implementations which are fully decompositional, i.e., they take only bases and affixes, not the derivatives into account, are also computationally possible and comparable to each other (cf. Stein 2023). Due to its relatively simple architecture, discriminative learning is linguistically transparent and can be interpreted more clearly than deep learning methods. In discriminative learning, two computational methods are used, Naive Discriminative Learning (NDL, Baayen et al. 2016) for vector computation and Linear Discriminative Learning (LDL, Baayen et al. 2019b) for the simulation of the mental lexicon.

Naive Discriminative Learning (NDL, Baayen et al. 2016) operates on the Rescola-Wagner rules (Rescorla & Wagner 1972; Wagner & Rescorla 1972) and creates semantic word vectors on the basis of these rules. It is grounded in theories on cognitive mechanisms. Words or grammatical/inflectional functions are seen as cues. These cues may or may not occur with the target words which are the outcome of the NDL estimation of word vectors. The word vectors consist of so-called association weights between a target word and cues. This process is similar to the calculated co-occurrences of target word and context word by other distributional computations. The association weight of a cue increases every time the pertinent outcome and cue co-occur, while it decreases every time the pertinent outcome occurs but the cue does not occur. Hence, each encounter of an outcome with cues leads to a recalibration of all association weights of the relevant outcome. At the end of this recalibration process, a stable end-state is reached. As every outcome is connected to all cues via their pertinent association weights, by taking these weights together one obtains a semantic vector representing the semantics of an outcome. NDL shows to be successful for the computation of word vectors for linguistic phenomena (see, e.g., Baayen et al. 2011, 2016; Tomaschek et al. 2021; Nieder et al. 2022; Arndt-Lappe et al. 2022; Schmitz 2023; Schmitz et al. 2023). The study in this dissertation uses NDL vectors which are then used to train a linear discriminative learning model.

Linear Discriminative Learning (LDL, Baayen et al. 2019b) uses already computed semantic vectors, e.g., from NDL,[4] as basis to linearly map form onto meaning and vice versa. Previous studies using LDL showed that the approach can be used to successfully simulate comprehension and production processes. LDL assumes that meaning itself is a dynamic concept, being emergent from the context in which words are being used. This computational model can, for example, give insight into the semantics of pseudowords (Chuang et al. 2021; Schmitz et al. 2021; Schmitz 2022) or role nouns (Schmitz et al. 2023), the nature of durational differences between different types of /s/ (Schmitz et al. 2021; Schmitz 2022), or the acoustic duration of derived words (Stein & Plag 2021; Stein 2023). Due to the low frequency of many of the nominalizations under investigation, a model that can deal with such infrequent words is needed. The estimation of the word vectors of the infrequent nominalizations is similar to the computation of word vectors for pseudowords in LDL. Thus, semantic word vectors for low frequent derivatives and their bases are retrievable from the implementation. Furthermore, similar to the subword information in *fastText*, the semantic vectors in LDL are based on trigrams. This makes both approaches comparable to each other.

One term used in NDL/LDL is *lexome*, i.e., the basic semantic unit which corresponds to whole words or morphological functions. Two types of lexomes are considered in discriminative learning: content lexomes, and inflectional and derivational lexomes (see, e.g., Chuang et al. 2021). Content lexomes describe units that contain semantic content, no matter if the form is morphologically simplex or complex, like *dress* and *dresses*.

[4]One can also use vectors created by other computational models, such as *fastText* (Bojanowski et al. 2016; Mikolov et al. 2018), *Word2Vec* (Mikolov et al. 2013b,a), FRACSS (Marelli & Baroni 2015), or others.

For the study at hand this means that bases and derivatives are both content lexomes. Inflectional lexomes reflect inflectional functions, for example, tense, number, or aspect. Derivational lexomes, as their name suggests, reflect derivational functions. The nominalizations investigated in this dissertation are built by content lexomes, i.e., the base, and derivational functions, i.e., the suffix, to create new content lexomes, i.e., the derivative. I will stick to the terms derivatives, bases, and suffixes in my dissertation, although the idea of lexomes instead of these categories is not to be discarded (see Chapter 7).

Chapter 7 presents the LDL implementation. The technical details of the implementation are given in the pertinent chapter. The LDL implementation allows a direct comparison of the computational perspectives of LDL and *fastText* in Chapter 6.

2.3 Derivational suffixes under investigation

2.3.1 The suffix *-ee*

Phonetically, the suffix *-ee* is autostressed (e.g., Plag 2018: 89). Verbal bases are frequent and nominal bases are not uncommon for formations with the suffix *-ee*. Looking at the semantics, the reading of a derivative with the suffix *-ee* is rather clearly discernable, i.e., it denotes a participant. The derivational process follows three known restrictions. First, the derived noun denotes a sentient participant of an eventuality. For example, the PATIENT in the *employing-action* (see, e.g., Table 2.1) is necessarily a sentient being. The sentience constraint is an important semantic requirement for derivatives ending in the suffix *-ee*, but it can be violated in domain-specific terminology (Barker 1998: 710).

The second constraint is that the base of the nominalization needs to have an eventuality in its semantic structure. Barker (1998) calls this constraint “episodic linking”:

> The intuition behind episodic linking is very simple: the referent of a noun phrase headed by an *-ee* noun must have participated in an event of the type corresponding to the stem verb. For example, in order to qualify as a *gazee* it is necessary to participate in a certain role in a gazing event. This requirement is a crucial part of explaining how the meaning of an *-ee* noun can depend on the meaning of its stem without depending on the syntactic argument structure associated with the stem. (Barker 1998: 711)

The necessity of episodic linking shows that eventualities and their participants are indispensable for the creation of the reading of derivatives with the suffix *-ee*. More specifically, these derivatives denote a participant of the eventuality in the base word. Due to the other requirements of the word-formation process with *-ee*, not all participants involved in the eventuality in the base word are possible candidates for the word-formation process. The same restrictions hold for denominal *-ee* derivatives (Barker 1998; Plag 2004).

Third, the participant denoted by the suffix *-ee* is restricted by a lack of volitionality. Hence, the participant which is described by the derivative must be non-volitional. This lack of volitionality is usually defined as the lack of control the *-ee* participant has over an eventuality. The lack of volitional control is not an absolute property but variable to a certain degree, in contrast to the sentience requirement or the episodic linking (Barker 1998: 719). More precisely, a participant does not have to be completely non-volitional but rather more non-volitional than other possible targets for the suffix *-ee*. For instance, a PATIENT is less volitional than an AGENT and thus more likely referred to by the nominalization with the suffix *-ee*. As a consequence to the gradability of volitionality, exceptions to the volitionality constraint can be found. For example, *escapee* denotes the AGENT of the *escaping-action*. This violation is possible for *-ee* derivatives with bases that are either intransitive verbs or verbs that can be interpreted as having a non-volitional subject participant (Barker 1998: 719f.). For new formations with the suffix *-ee*, the volitionality constraint is expected to be satisfied. However, if no other non-volitional participant than the AGENT is in the semantic structure of the base, the AGENT is referred to by the nominalization with *-ee*.

This dissertation wants to shed light on the nominalization process with nominal bases. For nominalizations with the suffix *-ee*, denominal examples are attested in the literature. The list in (2) illustrates some *-ee* nominalizations with a nominal base.

(2) Non-deverbal *-ee* derivatives (Barker 1998; Plag 2004; Mühleisen 2010)

aggressee	*adressee*	*tutee*	*wardee*
biographee	*asylee*	*bargee*	*benefactee*
debtee	*blind datee*	*covenantee*	*custodee*
laryngectomee	*executionee*	*festschriftee*	*inquisitee*
mentee	*letteree*	*malefactee*	
patentee	*moneylendee*	*optionee*	
return	*philanthropee*	*politicee*	

Summarizing, previous research has established three restrictions for nominalizations with the suffix *-ee*, which hold for all kinds of bases (Barker 1998; Plag 2004):

1. the reading of the derivative is a participant involved in an eventuality in the base word
2. the participant must be sentient (sentience requirement)
3. the participant lacks volitional control with some exceptions, for example, intransitive verbs (non-volitionality constraint)

2.3.2 The suffix *-ment*

Nominalizations with the suffix *-ment* are less clearly restricted than nominalizations with the suffix *-ee*. The suffix *-ment* is said to show little productivity as it is a non-native suffix (Bauer et al. 2013: ch. 10). However, corpora show quite a few attestations of *-ment*-derivatives (cf. Kawaletz 2023). Mostly verbal and bound bases are attested for the derivational process with *-ment*. Other categories of base words, for example, adjectives or nouns, are rare but attested in corpora. As most non-native suffixes, *-ment* appears to have preference for disyllabic bases and iambic feet (Bauer et al. 2013: ch. 10).

Basically, nominalizations with the suffix *-ment* denote processes or results given by the base (Plag 2018: ch. 4). A large-scale study about the semantics of nominalizations with the suffix *-ment* by Kawaletz (2023) focuses on new *-ment*-formations based on verbal bases. Due to the eventive nature of verbs (e.g., Van Valin & LaPolla 1997, 2002; Haspelmath 2001; Szabó 2015; Moltmann 2019), it is unsurprising that verbal bases are productive for nominalizations with *-ment*. Overall, verbs are the most common word class to serve as bases for *-ment*-formations (e.g., Bauer et al. 2013; Plag 2018). Kawaletz (2023) made use of the reference shifting approach in semantic frames adapted in this dissertation (see Section 2.2.1). The analysis of deverbal *-ment* nominalizations shows that the readings of the resulting derivatives are versatile. The process with *-ment* can target several elements in the semantic representation of a base word.

The successful creation of a nominalization with the suffix *-ment* is generally restricted by the semantics of the base. If a base has a specific element, e.g., a RESULT-STATE, the nominalization process can shift the reference from the denotation of the base to the result as in *annoyment*.[5] If a second element for the derivational process with the suffix *-ment* is available in the semantic representation of the base, *-ment* can shift the reference on this element as well. In some cases, a *-ment* nominalization is clearly used to distinguish it semantically from more common nominalizations (e.g., *revealment* instead of *revelation*, Bauer et al. 2013: ch. 10). Hence, nominalizations with *-ment* are highly polysemous as *-ment* is not rigorously restricted to specific eventive elements. Only one restriction seems to hold: Nominalizations with *-ment* cannot refer to animate entities, i.e., not to AGENTS, PATIENTS, or other animated participants of an eventuality (Kawaletz 2023).

For denominal *-ment*, no preset list of attestations is available. Denominal *-ment* formations are indeed rare and complex to retrieve from corpora. One denominal nominalization with the suffix *-ment* is explicitly given in Bauer et al. (2013), namely *illusionment* (see Chapter 4, Section 4.2.1.1). More attestations of denominal *-ment* were found in corpora. The data retrieval is described in Chapter 3 as well as the concrete number of attestations.

[5] For more possible readings of *annoyment* see Kawaletz (2023).

2.3.3 The suffix *-ation*

The suffix *-ation* is, just like the suffix *-ment*, non-native and therefore said to be less productive compared to native suffixes. The derivational process with *-ation* does mostly operate on verbs and bound bases (Bauer et al. 2013: ch. 10). The preference for verbal bases lies, again, in the fact that verbs are eventive in nature (e.g., Van Valin & LaPolla 1997, 2002; Haspelmath 2001; Szabó 2015; Moltmann 2019). A verbal base simplifies the analysis of eventuality-related nominalizations due to the eventive nature of the verbal base. Although the suffix is, in relation to native suffixes, less productive, new formations are still found in corpora.

The suffix *-ation* has some variants: *-ation, -cation, -ion, -ition, -iation, -sion, -ution, -tion* and it is the default nominalizer for verbs ending in *-ize* or *-ify* (Plag 1999: 68f.; Bauer et al. 2013: ch. 10). Plag (1999: 207) states that "every noun in *-ation* is a possible candidate for the back-formation of a verb in *-ate*." Furthermore, the line of argumentation for *-ation* as a suffix on its own are examples like *declare* ♦ *declaration* where no verb in *-ate* exists. These observations, as well as the fact that many nominalizations ending in *-ation* do not have an attested verbal base ending in *-ate*, lead to the decision to treat *-ation* as a suffix on its own, similar to Plag (1999, 2018), for instance.

Nominalizations with the suffix *-ation* usually denote events or results of processes, but locations or means can also be found (Plag 2018; Stein 2023). For instance, derivatives in *-ation* can denote "'provide with X' (ornative), [...] or 'make into X' (resultative)" (Plag 2018: 93) readings. Previous findings on *-ation*-nominalizations indicate that a nominalization "suggest that action nominals in *-ation* merely denote an Event having to do with the entity denoted by the base" (Plag 1999: 209). More precisely, an eventive structure must exist in the nominal base in order to make the nominalization possible. This is in line with the analyses and their findings in this thesis for all suffixes.

A list of typical denominal attestations with the suffix *-ation* is given in (3). Note that this list is not exhaustive. The data used in this dissertation is described in Chapter 3.

(3) Denominal *-ation* derivatives (Bauer et al. 2013; Plag 2018)

artefaction	*epoxidation*	*metalation*
ozonation	*placentation*	*sedimentation*

Nominalizations ending in *-ation* based on chemical substances (Chapter 4.3.1.1) and technical terms (Chapter 4.3.2) are exemplary analyzed. This analysis illustrate the semantic shift from a nominal base to a derivative in a semantic decomposition approach.

2.4 Summary and road map

Eventuality-related nominalizations are dependent on eventive elements in the base which can be referred to by the derivative. Due to the necessity of eventive elements in the semantic representation of the base word, verbal bases are more often used and can be analyzed in a more straightforward manner as they are eventive in nature (e.g., Van Valin & LaPolla 1997, 2002; Haspelmath 2001; Szabó 2015; Moltmann 2019). Consequently, research tends to focus on deverbal formations, only stating that the process has to work similarly with nouns. This leaves the field without an account for denominal eventuality-related nominalizations.

The first part of my thesis will deal with the question how eventuality-related nominalizations work with nominal bases. More precisely, where the eventive structures required for the word-formation process are to be found in the semantic structure of the base. Frame semantics are used for the decomposition of the semantic structures of bases and derivatives (Chapter 4). The frame semantic analysis helps to explore the eventive structures in the nominal base. These eventive structures, in turn, are needed for the word-formation process with the suffixes *-ee*, *-ment*, and *-ation*. Applying the reference shifting approach which proved to be useful for the analysis of deverbal *-ment*-derivatives (e.g., Plag et al. 2018; Kawaletz 2023), the reference shifts are induced by the word-formation processes. The word-formation process operates on the semantic structures of the base to create the reading of the derivative. In other words, the nominalization operates on semantic structures already given by the base. The frame semantic notation helps to visualize the semantic structure of the base as well as the reference shift from base to derivative.

The second part of this dissertation deals with the application of computational methods to test how semantically similar base and derivative are. As the reference shifting approach analyzes the derivational process as being dependent on the semantic structure of the base, base and derivative are supposedly semantically related. The aim of this study is to find out how strong this relation is, if denominal nominalizations are as semantically similar to their bases as deverbal nominalizations are, and if the word-formation processes with different suffixes show different degrees of similarities of denominal and deverbal nominalizations. Two different computational implementations are be modeled: First, a deep learning algorithm, i.e., *fastText* (Chapter 6), and second, a discriminative and cognitive-based approach, i.e., discriminative learning (Chapter 7).

Chapter 3

Methodology: General aspects and data collection

In order to find different types of denominal nominalizations, data from several corpora was collected: the British National Corpus (BNC; Davies 2004-), the Corpus of Contemporary American English (COCA; Davies 2008-), the News on the Web Corpus (NOW, Davies 2016-) and iWeb: the intelligent Web-based Corpus (iWeb, Davies 2018-).[6] For the data collection, the available web interface of these corpora was used. The search strings '*ee', '*ment' and '*ation' were used to compile all nominalizations ending with the suffixes under investigation: *-ee*, *-ment*, and *-ation*. The resulting data sample was pre-cleaned by manually excluding bases which only occur as verbs for the denominal data sets. For example, *employ* as the verbal base for *employee*, or *acquire* as the verbal base for *acquirement*. The verbal bases were retained for the computational analyses in Chapters 6 and 7.

For the determination whether a base word should be classified as a noun or a verb, first, the *Oxford English Dictionary online* (OED 2025) was consulted. For some words, the categorization was clear, e.g., for the noun *debt*. However, as conversion in English is productive, many bases are attested as verbs and nouns, for example, *charge* (see, e.g., Bauer et al. 2013; Plag 2018; Barbu Mititelu et al. 2023). In Table 3.1, exemplary words, attested derivatives, and the possible word classes of their base words are illustrated. For example, the word *debtee* remained in the data set

[6]The data as well as the scripts used for the individual analyses are available here: `https://osf.io/kaqsv/`.

of denominal nominalizations. Its base is clearly denominal because the base word *debt* is only attested as a noun. For *covenantee* and *chargee*, different criteria are needed in order to decide whether the base word for the resulting derivative is a verb or a noun.

TABLE 3.1: Derivatives and the attested word classes of their base words.

derivative	base	word class of base
debtee	*debt*	noun
covenantee	*covenant*	noun or verb
chargee	*charge*	noun or verb

Three criteria were applied to ensure that only nominalizations with nominal bases entered the analysis. Potential bases that met at least one of the following criteria were excluded to create a final denominal data set with derivatives that are based on nouns.

The first criterion is a frequency criterion. Theories of the mental lexicon state that a word with a higher frequency is more prominent in the mental lexicon of speakers (Plag 2018: 46; Warren 2013: ch. 3). Regarding the distinction of word classes, this implies that in cases of a form which can be either a verb or a noun, the more frequent representation in the speaker's mind is more likely to be chosen as a base word for the derivational process (Warren 2013: 136). Since the knowledge of the stored words in the mental lexicon is gradual and speaker-dependent, an indication for the possibility of a word stored as a verb and/or a noun had to be defined. The likelihood of the prominence of the representation in the mental lexicon of the verbal form increases with a higher relative frequency of the verb. The frequency for the analysis was measured in percentages, comparing the frequencies of the occurrences of a word form as a noun to the frequencies of occurrences of a word form as a verb. If the verbal form of the base represents over 30 percent of all tokens of the base forms, the base was excluded from the data set. The threshold value of 30 percent was set arbitrarily to ensure that the noun is generally more frequent in use. Broadly speaking, it is assumed that derivational processes operate on the more frequent type of word class.

The frequency criterion, based on a large corpus, is also supported by the Large-Number-of-Rare-Events theory (LNRE, e.g., Baayen 2001; Evert

& Baroni 2007). LNRE modeling is based on the distribution of frequencies assuming that the distribution of a word in relation to a different word remains the same, disregarding the actual number of observations, i.e., tokens in the data. For the decision of the word class of the base this means that the relation of the verbal and the nominal form of a word stays the same, independent of the corpus. This being the case, if a noun is distributionally clearly more represented than the corresponding verb, the base is considered to be denominal.

The application of the frequency criterion proceeded as follows: When a base word was listed not only as a noun but also as a verb, or an adjective, in the OED (2025), the frequencies of the base words were checked in COCA (Davies 2008-) and COHA (Corpus of Historical American English; Davies 2010-). The automatic word class tagging of the bases in the corpora was used to decide how many word forms are attested as nominal or verbal. The frequencies from COHA were also considered, because the dominant word class, depicted in the frequency of noun and verb in COCA, might have changed due to conversion and lexicalization. Notably, the frequencies of both corpora show roughly the same relative frequencies regarding the word class of the bases in the data.

An example of the application of the frequency criterion is given in Table 3.2. Since the verb *charge* as a potential base word for the derivative *chargee* has an overall frequency of 31 percent of all tokens, *chargee* was excluded from the denominal data set due to the possibility of being deverbal. The distributions of the possible word classes for the base word *covenant* are different. The word *covenant* is only a verb in 8 percent of all tokens. Consequently, the derivative *covenantee* was included in the data set as it is clearly denominal.

However, the percentage criterion is not always reliable on its own in deciding whether a derivative is created on a verbal or a nominal form of the base word, thus, the criterion needs additional considerations. The unreliability of the percentages solely might be due to a frequency threshold of the proportion versus the absolute frequency. That is, in some cases, the verb takes less than 30 percent of all tokens, but it is in absolute numbers highly frequent. For example, in COCA, the word *interview* has a low frequency for the verbal base form in terms of proportion, with 9 percent of all tokens, but overall a high frequency of occurrences as it occurs 6783

TABLE 3.2: Number of attestations by word class in COCA given as tokens and percentages.

	tokens	percentages
charge		
V	8265	31
N	26469	69
covenant		
V	187	8
N	2223	92

times as a verb and 73551 time as a noun in the corpus. The bases which show a high overall frequency of the verbal form of the base were excluded after the pre-sorting via the frequency criterion. The high overall frequency of a verb was defined by a threshold of 5000 occurrences of the verb. This threshold was introduced to distinguish whether the derivative is really based on the noun and not the verb. More precisely, with a high frequency of the verb overall, it remains unclear if the interpretation of the word class of the base is nonetheless biased to be a verb due to its many occurrences in language use.

The second criterion is based on definitions from the OED (2025) and applied after the frequency criterion. Definitions of the remaining derivatives were searched for in the OED (2025). This was done to check whether the reading of a derivative is clearly related to a verbal base although the nominal form of the base is more likely to be the base. For example, *internee*, which is defined as “a person who is imprisoned or otherwise confined” (OED 2025) is clearly based on the verbal reading of *intern* and not on its nominal reading, as the reading of the derivative is related to the denotation of the verb *to intern someone* and not to the semantics of the noun *intern*: *trainee*.

The third criterion excludes bases if a nominal base word was described as being obsolete in the OED (2025), like *legate*, the derivative was excluded as it was not possible to determine whether the derivative is based on the obsolete noun or on the remaining verb.

For the comparisons of denominal and deverbal eventuality-related nominalizations in computational models (Chapter 6 and Chapter 7), not only the denominal data was needed but also deverbal data. Deverbal derivatives were needed to see whether denominal and deverbal nominalizations show different similarities of derivatives and their bases. The deverbal data was retrieved by the sorting of the data for denominal nominalizations, excluding verbal bases. More precisely, the deverbal data which was not needed for the reference shifting approach was included for the computational studies. As verbs are more likely the base of eventuality-related nominalizations due to their eventive nature (Van Valin & LaPolla 1997; Haspelmath 2001; Van Valin & LaPolla 2002; Szabó 2015; Moltmann 2019), only a random sample of the deverbal data was used. For the randomization, the excluded data was used. For example, *charge* was excluded from the set of denominal derivatives due to a high relative frequency of the verb related to the noun. Thus, *charge* was a candidate to be included in the data for deverbal nominalizations. This procedure also ensures that no derivative is tagged as denominal and deverbal at the same time.

An overview of the final data sets for the derivatives with the suffixes *-ee*, *-ment*, and *-ation* is given in Table 3.3. The denominal data was used for the reference shifting approach after the above described cleaning processes. The deverbal data were only used in the computational approaches to make denominal and deverbal derivatives comparable. Examples of denominal and deverbal derivatives with the suffixes are given in (1).

TABLE 3.3: The data of all three suffixes ordered by category: denominal/deverbal.

	denominal	deverbal
-ation	67	72
-ee	46	312
-ment	29	273
total	142	657

(1) denominal

a. *-ee*: *biographee, covenantee, debtee*

b. *-ment*: *illusionment, devilment, rascalment*

c. *-ation*: *concertation, instrumentation, ozonation* deverbal

a. *-ee*: *appointee, devotee, employee*
b. *-ment*: *accomplishment, development, enjoyment*
c. *-ation*: *avocation, beneficiation, idolization*

Chapter 4

Frame semantic analysis

This chapter deals with a frame semantic analysis.[7] Examples for nominalizations with all three suffixes are given. Section 4.1 deals with the suffix *-ee*, Section 4.2 illustrates nominalizations with *-ment*, and Section 4.3 analyses the word-formation process with the suffix *-ation*.

4.1 Nominalizations with the suffix *-ee*

Nominalization with the suffix *-ee* usually refer to sentient, non-volitional participants in an eventuality denoted by the base. In Figure 4.1, a frame representation of the derivational process with the suffix *-ee* is shown. It includes the constraints given by Barker (1998) (see Section 2.3.1). The frame illustrates which elements are needed in the semantic representation of the base to successfully form a nominalization with the suffix *-ee*. An element of this descriptive generalization is an eventuality in the base. This eventuality then includes a participant which, in turn, is referred to by the derivative.

[7] An earlier version of parts of Section 4.1 have already been published in Schneider (2023). An earlier version of parts of Section 4.2 have already been published in Kotowski et al. (2023). This section presents an altered version of parts of the previously published version.

$$\begin{bmatrix} \textit{lexeme} & \\ \text{PHON } \boxed{x} & \\ \text{CAT N} & \\ \text{SEM } \boxed{0} & \begin{bmatrix} [...] & \\ [...] & \begin{bmatrix} \textit{eventuality} & \\ \text{PARTICIPANT } \boxed{1} & \begin{bmatrix} \textit{entity} \\ \text{SENTIENCE } \boxed{2}\ \textit{yes} \\ \text{VOLITIONALITY } \boxed{3}\ \textit{no} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{REF} \neq \left\{\boxed{1}\right\} & \end{bmatrix}$$

FIGURE 4.1: Generalized base for the process for denominal *-ee* nominalizations.

The type *lexeme* at the top of the frame stands for the derivative and base under investigation. This base has the phonological form[8] (PHON) and the syntactic category noun (CAT N).

Additionally, the base contains a semantic attribute (SEM) which represents the semantics of the base. In order to successfully derive an *-ee* nominalization, the semantic representation of the base needs to contain an *eventuality*. Importantly, the base word does not have to consist of this eventuality directly in the node indexed with $\boxed{0}$, but the eventuality can be more deeply embedded in the semantic representation of the base word. This potential embedding is represented by the three dots ([...]) in SEM $\boxed{0}$, which serve as placeholders for a potentially different semantic type of the base. Moreover, this eventuality in the base needs to have one PARTICIPANT such as the one indexed with $\boxed{1}$. The PARTICIPANT is an *entity* which has to be sentient and non-volitional. The SENTIENCE-attribute indexed with $\boxed{2}$ has the value *yes* for indicating that the entity is sentient. The VOLITIONALITY-attribute indexed with $\boxed{3}$ has the value *no*, thus indicating that the entity described is non-volitional. The PARTICIPANT indexed with $\boxed{1}$ meets the restrictions of the suffix *-ee* described in Section 2.3.1 and is

[8]The phonological form given in Figure 4.1 is only short-hand for a more complex representation that would have to include, for instance, the necessary adjustments in stress. As this study focuses on semantics, the phonology of *-ee*-derivatives is not described in detail (see, e.g., Bauer et al. 2013: 227).

hence required for the referential shift to create the meaning of the *-ee* nominalization.

Figure 4.2 shows the derivational process with the suffix *-ee* . The nominalization is a *lexeme* with a phonological representation (PHON) and the syntactic category noun (CAT N). The attribute M-BASE describes the morphological base of the derivative, as in Figure 4.1.

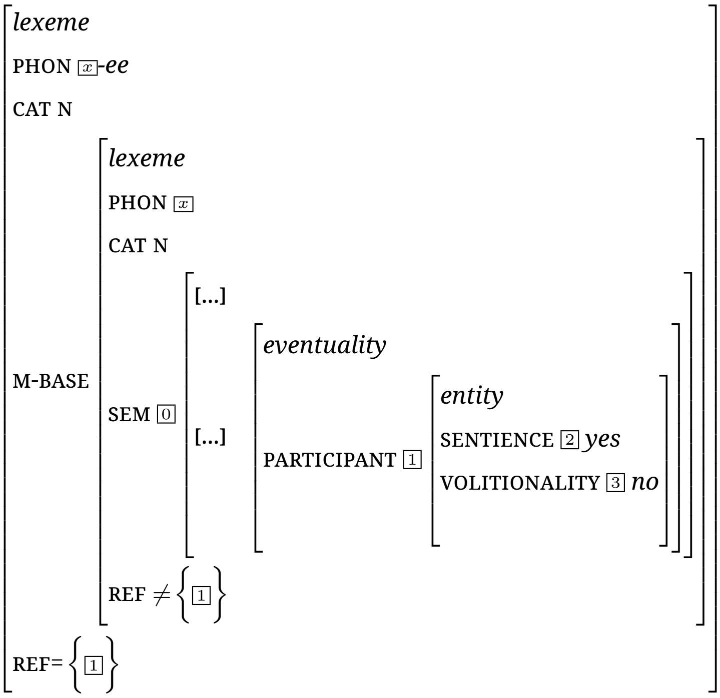

FIGURE 4.2: Generalized derivational process for denominal *-ee* nominalizations.

At the bottom of the representation of the base and the derivative, the potential referents of the lexemes are indicated by a REFERENCE-attribute (REF) for each lexeme. The reference of the base word is not fixed in the generalized frame with the exception that it cannot refer to the PARTICIPANT indexed with 1. The reference of the derivative, however, is the node indexed with 1, as the reference shifts towards the PARTICIPANT.

The frame in Figure 4.2 is the template for the frame representations of the individual words in the remainder of this section. The analysis will show that the derivational process is not always as straightforward as assumed in the generalized frame.

For the semantic decomposition of the individual bases, the definitions of the base words in the *Oxford English Dictionary* (OED, 2025) were used as a starting point. This source was chosen, as it is based on solid empirical evidence (i.e., a large range of actual attestations). The entries are accessible for verification and provide paraphrases detailed enough to derive semantic descriptions from. The usage of derivatives is illustrated with attestations from the *Corpus of Contemporary American English* (COCA, Davies 2008-) and the *British National Corpus* (BNC, Davies 2004-).

For eventuality-related nominalizations with the suffix *-ee*, the question of the base is not a trivial one. Some of the bases serve as stems in a syntagmatic process (Bauer et al. 2013: ch. 23). For example, the base *debt* is the stem for the derivative *debtee*. With other bases, the process is a paradigmatic one. That is, the base for the derivative already has a (pseudo-)suffix, for example, *tutor* for *tutee*. This is then a paradigmatic relation (Bauer et al. 2013: ch. 23).

This decision of the process, syntagmatic or paradigmatic, is also a problem for morpho-phonology (cf. Bauer et al. 2013: ch. 9). For example, with paradigmatic related forms, the stress is assigned via the related forms and not by the usual valid phonological rules. For the paradigmatically transferred stress pattern, different names are coined in the literature. Bauer et al. (2013) summarize these names: ‘stem selection’ (Raffelsiefen 1998), ‘paradigm uniformity’ (Steriade 2000), ‘multiple correspondence’ (Burzio 1994), and ‘split-base’ (Steriade 1999).

The different morphological patterns, i.e., syntagmatic and paradigmatic processes, are found for eventuality-related nominalizations with the suffix *-ee*. Following this line of thought, Section 4.1.1 deals with the syntagmatic process and Section 4.1.2 with the paradigmatic process. Section 4.1.3 will explain a different pattern of *-ee*-derivatives, violating the volitionality constraint, and Section 4.1.4 sums up the findings of the semantic analysis using the reference shifting approach.

4.1.1 Syntagmatic process

4.1.1.1 *debtee*

The first derivative under investigation is *debtee*. Its base word is the noun *debt*. A *debtee* denotes a person as illustrated in (1):

(1) *The “**debtee**” is an old word for the creditor or payee.* (BNC, see Davies 2004-, ACAD, 1992)

As indicated in the definition in (1), the *debtee* has an eventuality-related interpretation, as it denotes an entity to whom the money is owed. This eventuality suggests a potential *paying-eventuality*. More precisely, the *debtee* can be construed as the recipient in a future money transfer eventuality. Since *debtee* denotes a participant in an eventuality, the assumption is that the eventuality to induce the referential shift of the derivative is in the semantic representation of the base word *debt*. This assumption leads to the question where the eventuality and the required participant for the word-formation process with *-ee* is to be found in the nominal base. The paraphrases in the OED (2025 s.v. *debtee*) already provide a clue for the eventuality in the semantic representation of the base word *debt*:

1. *That which is owed or due: a sum of money or a material thing; a thing immaterial.*
2. *A liability or obligation to pay or render something; the condition of being under such obligation.*

In the second definition, *debt* is defined as an obligation. Semantically, obligations can be analyzed as states, i.e., non-dynamic eventualities with at least one participant. I will refer to this eventuality as an obligation-state. This link leads to the representation of the base in Figure 4.3 and of the derivative in the frame in Figure 4.4. Base and derivative have a phonological representation (PHON) and the syntactic category noun (CAT N). The base *debt* in Figure 4.3 is at the same time included in the derivative as the morphological base (M-BASE) of the derivative in Figure 4.4.

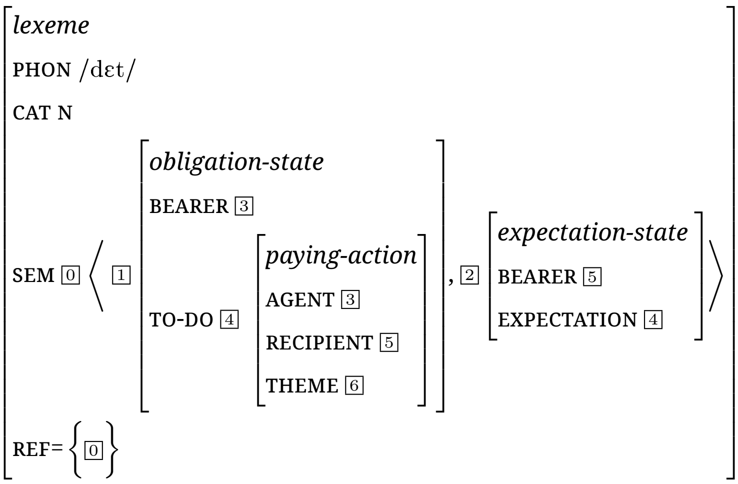

FIGURE 4.3: Representation of the lexeme *debt*.

The semantics of the base (SEM) is depicted as a so-called multi-AVM and consists of two frames with different source nodes, indexed with [1] and [2]. Both AVMs illustrate states, an *obligation-state* ([1]) and an *expectation-state* ([2]). First, the *obligation-state* has two attributes; a BEARER[9] of the state, indexed with [3], and an attribute TO-DO, indexed with [4]. The attribute TO-DO specifies what the BEARER has to do in order to fulfill their obligation. The TO-DO attribute has the value *paying-action*, which is a sub-eventuality in the semantic representation of *debt*. This *paying-action* has at least three attributes; an AGENT indexed with [3], a RECIPIENT indexed with [5] and a THEME indexed with [6]. The indices indicate that the BEARER of the *obligation-state* is the same person as the AGENT of the *paying-action* indicated by [3]. The BEARER in the *obligation-state* is the one who has to pay.

The second state in the base is an *expectation-state* which describes the circumstance that the *paying-action* ([4]) has not been finished. The *debtee* expects to get their money back. The *expectation-state* has at least two attributes, a BEARER indexed with [5] and an EXPECTATION indexed with [4]. Note that the BEARER of the *expectation-state* is co-indexed with the RECIPIENT of the *paying-action* ([5]) and the EXPECTATION is co-indexed with the whole *paying-action* ([4]).

[9]The label BEARER is only one option to name the participant in a state. Different notations for the participant in a state are known (see, e.g., Van Valin & LaPolla 2002). The label of the participant does not play a crucial role for the analysis provided in this thesis, as long as this participant role fulfills the sentience and volitionality restrictions of *-ee* in pertinent cases.

Moving on to the process of derivation represented in Figure 4.4, the question arises of how the word-formation process with *-ee* knows which node to pick as the reference node (REF). The restrictions for nominalizations with the suffix *-ee* state that the referent has to be sentient and non-volitional (Section 2.3.1). In the semantic representation of the base *debt*, two possible targets for the shift of reference are present, the BEARER of the *obligation-state* and the BEARER of the *expectation-state*, as BEARERS are sentient and non-volitional. The participant indexed with [3] is not only a BEARER of the *obligation-state* but also an AGENT in the embedded eventuality, the *paying-action*. Hence, the participant in [3] is not only a non-volitional participant as a BEARER in the *obligation-state* but also a volitional participant, an AGENT in the *paying-action*. The BEARER of the *expectation-state*, on the other hand, is co-indexed with the RECIPIENT of the *paying-action*. As the BEARER of the *obligation-state* is co-indexed with the AGENT of the *paying-action*, it is more likely for the derivative to refer to the BEARER of the *expectation-state*. The entity indexed with [5] is only represented as a non-volitional participant. If two possible targets exist, the more non-volitional participant is chosen as a referent for the word-formation process, according to the volitionality constraint. Hence, [5] is the reference node for the interpretation of *debtee*. Note that the derivative does refer to the element indexed with [5] in the *expectation-state* [2] as the payment of the *paying-action* has not been completed.

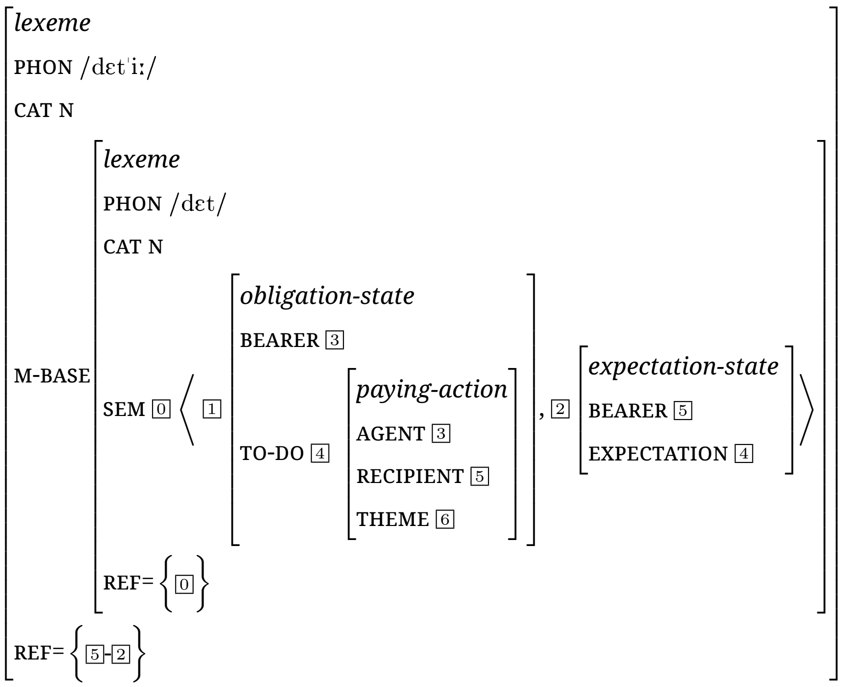

FIGURE 4.4: Representation of the lexeme *debtee*.

The frames for *debt* and *debtee* illustrate the derivational process on an eventive nominal base. Accordingly, an eventuality exists in the semantic representation of a nominal base, here two states. The eventuality and its participants provide possible targets for the word-formation process.

4.1.1.2 *biographee*

As shown in 4.1.1.1, the derivational process of *-ee*-derivatives with nominal bases can be rather straightforward in the case of eventive bases. However, what happens if a base is not clearly eventive? This will be illustrated by the derivative *biographee*. The examples in (2) illustrate the use of this derivative:

(2) a. [...] version of the Petraeus biography, whose biographer is admitted to have lain with the **biographee** in the Biblical sense. (COCA, see Davies (2008-), BLOG, 2012)

b. [...] who noted that 'politicians are happiest when talking, at their most miserable when making up their minds.' His **biographee** would dissent (BNC, see Davies (2004-), MAG, 1985-1994)

The base for the derivative is the noun *biography*. The base noun is not clearly eventive, as it denotes, among other things, an object. As Pustejovsky (1996: ch. 8) already pointed out, artifacts, like *books*, also serve as objects which contain information. Such objects which can be of two different types are so called ‘dot-objects’. *Books*, for example, can be described as physical objects and as containing information at the same time. This also applies to the specific type of book, *biography*, which is not only a concrete object but also a thing that contains information. This information concerns the life of a person and the derivative *biographee* denotes this person. The referential target of the suffix -*ee* is the person about whose life a book is written.

Moreover, a *biography* has to come into being and it is used for a specific purpose. These two processes are eventualities that are part of the semantic representation of the base word *biography*. These different aspects of the meaning of the base *biography* and the derivative *biographee* are represented in the reference attributes REF for the base in Figure 4.5 and for the derivative in Figure 4.6.

The semantic representation of the base *biography* is more intricate than that of *debtee*. The base noun *biography* is labeled as a *text* in the semantic representation, as the frame for the base illustrates *biography* as an information object. This *text* has the attributes PRODUCTION indexed with [1], AFFORDANCE indexed with [2], and TOPIC indexed with [4]. The structure in the base is based on Pustejovsky (1996: ch. 6). This illustration of the base shows which elements can be assumed in a nominal base in general. A look at the properties of the participants in these eventualities will show which of them are compatible with the suffix -*ee* . The PRODUCTION-attribute is analogous to the AGENTIVE-QUALE in Pustejovsky, and the AFFORDANCE-attribute is analogous to the TELIC-QUALE in Pustejovsky. Hence, a biography is a *text* about something (TOPIC) which is produced by someone (PRODUCTION/AGENTIVE) for a special purpose (AFFORDANCE/TELIC).

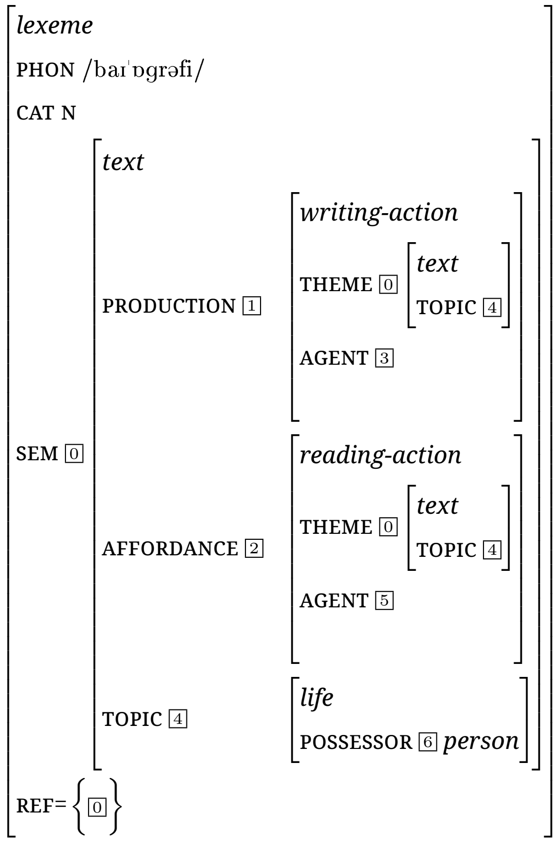

FIGURE 4.5: Representation of the lexeme *biography*.

The PRODUCTION-attribute and the AFFORDANCE-attribute have eventualities as their value. The PRODUCTION is a *writing-action* and the AFFORDANCE is a *reading-action*. Both eventualities contain an AGENT; in the *writing-action* the agent is the writer of the text and in the *reading-action* the agent is the reader of the text. The two AGENT-attributes refer to different persons and are therefore given different index-numbers ([3], [5]).[10] In contrast, the THEME-attributes (indexed with [0]) represented in both eventualities refer to the same entity, namely the *text* itself, which is the referent of *biography*. This THEME contains the TOPIC-attribute indexed with [4].

[10]The frame-format does not prohibit co-indexation of these two elements. The writer can read their own work as well. The non-co-indexed version is probably the more common interpretation.

Finally, in order to account for the specific genre of text instantiated by *biography*, the topic-attribute is typed *life*. In turn, the type *life* has a POSSESSOR,[11] indexed with $\boxed{6}$, which is a person.

Figure 4.6 shows the semantic frame for the derivative *biographee* including the morphological base (M-BASE) in Figure 4.5. The REFERENCE-attribute is specified for an element with the required semantics to form the meaning of the derivative *biographee*. The reference is on the element indexed with $\boxed{6}$, the POSSESSOR. Of the seven different nodes in this frame, only this one adheres to the restrictions posed by *-ee* . All other nodes are either volitional (AGENT: $\boxed{3}$, $\boxed{5}$) or non-sentient (THEME $\boxed{0}$, PRODUCTION $\boxed{1}$, AFFORDANCE $\boxed{2}$, TOPIC $\boxed{4}$). Thus, the aforementioned restrictions on derivatives with the suffix *-ee* successfully narrow down which part of the base's semantic structure can be used as reference for the derivative.

[11]The label POSSESOR is only one option to name the participant in $\boxed{4}$. Other analyses for this participant might be available. The label of the participant does not play a crucial role for the analysis provided in this thesis.

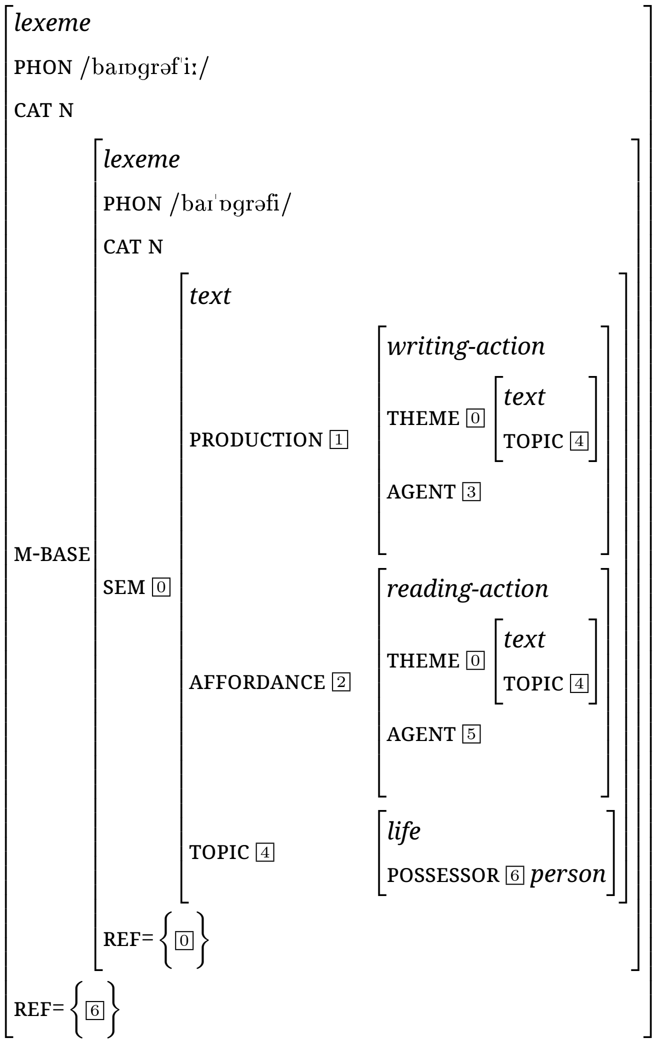

FIGURE 4.6: Representation of the lexeme *biographee*.

4.1.1.3 *covenantee*

The next derivative analyzed is *covenantee* as in the examples in (3). The base is the noun *covenant*.

(3) a. It was argued on behalf of the respondents that the doctrine applied to a covenant which was imposed for the benefit of the trade of the **covenantee** and which either forbids the covenantor to carry on his trade or restricts the way in which he may carry it on. (BNC, see Davies (2004-), ACAD, 1991)

b. However, even if the **covenantee** would not have entered the agreement without the covenant, the contract may not be invalidated as a whole [...]. (BNC, see Davies (2004-), ACAD, 1991)

According to the definition from the OED (2025 s.v. *covenant*), a covenant is “a mutual agreement between two or more persons to do or refrain from doing certain acts”. The derivative *covenantee* denotes a person who is under an obligation to fulfill such an agreement, regardless of whether the agreement authorizes or prohibits something. Interestingly, *covenantee* can refer to two entities, namely to both parties that are involved in the covenant. This is opposed to *debtee* and *biographee*, for which only one possible referent can be identified. Figure 4.7 illustrates the frame for the base *covenant* and Figure 4.8 the frame for the derivative *covenantee*.

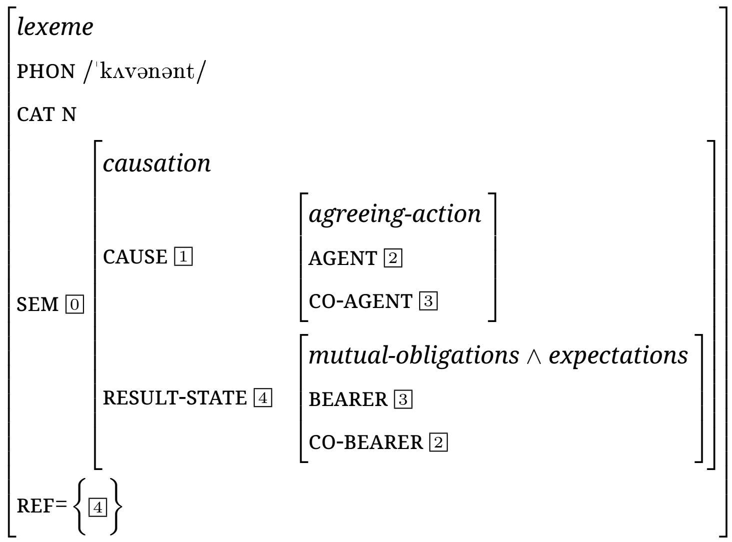

FIGURE 4.7: Representation of the lexeme *covenant*.

Focusing on the representation of the base, a *covenant* is a different type of obligation than the one in the base *debt* described above (see section 4.1.1.1). The obligations denoted by a *covenant* are *mutual-obligations* and *expectations*. These *mutual-obligations* and *expectations* are analyzed as a state in $\boxed{4}$, which is a sub-eventuality that results from a *causation*-eventuality. The *causation*-eventuality consists of a CAUSE, which is an *agreeing-action* and a RESULT-STATE. The *agreeing-action* has at least two attributes, an AGENT indexed with $\boxed{2}$, a CO-AGENT indexed with $\boxed{3}$. Two agents are realized, because the agreement on the *covenant* as the RESULT-STATE is made by two parties. These parties can be on equal level as assumed in the frame notation. The RESULT-STATE has at least two attributes, a BEARER indexed with [2], and a CO-BEARER indexed with [1]. The two BEARERS are co-indexed with the two AGENTS in the *agreeing-action*.

Two possible targets for the suffix *-ee* are available in the M-BASE in Figure 4.7, as two non-volitional sentient beings are represented in the RESULT-STATE of the base, the BEARER [3] and the CO-BEARER [2]. The REFERENCE-attribute (REF) in the frame in Figure 4.8 shows that the meaning of the derivative *covenantee* can either be on the element indexed with [2] or [3] in the RESULT-STATE [4] ([2]-[4], [3]-[4]). Interestingly, both participants are symmetric in the frame representation, as both occur as AGENTS in the CAUSE and as BEARERS in the RESULT-STATE. It is not possible to tell apart which participant an *-ee* form refers to because they behave identically in all sub-eventualities. Thus, two possible referents in their non-volitional reading in the sub-eventuality [4] are the expected result.

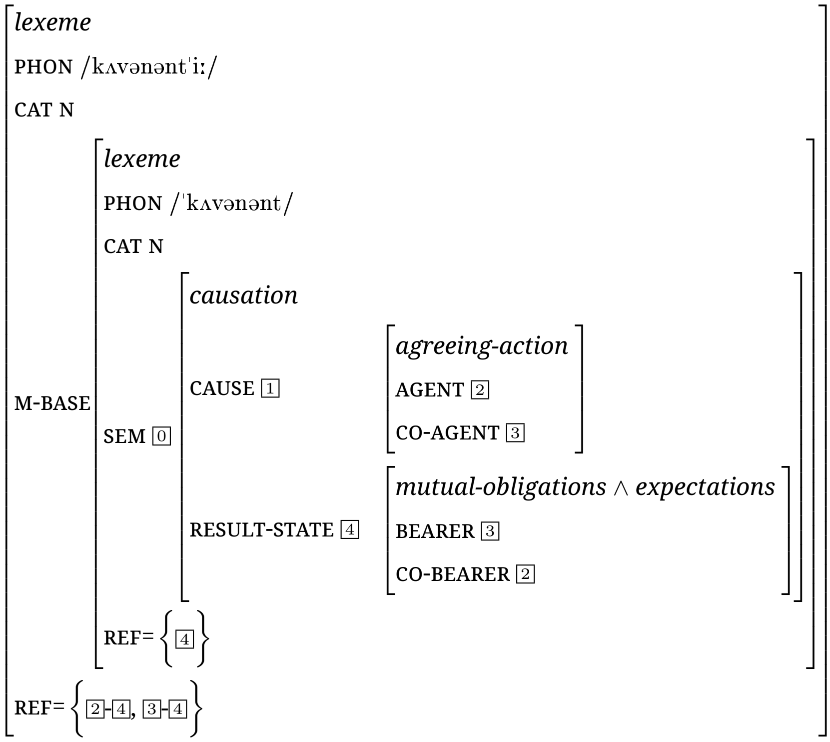

FIGURE 4.8: Representation of the lexeme *covenantee*.

4.1.2 Paradigmatic process

The data on denominal *-ee*-derivatives show an interesting subset with *-er/-or* forms as their bases. For example, the derivative *tutee* is most likely constructed on the noun *tutor*. According to Bauer et al. (2013: 524) these nouns are paradigmatically related to each other. This relation a-rises when the new word has the same base but a different suffix. This is especially interesting as the data show that many *-ee*-derivatives come into being not by suffixation to a base (i.e., a syntagmatic process) but by a paradigmatic process (see also Bonami & Guzmán Naranjo 2023). Many bases in this category are loan words from Latin or French. Some of these loan words seem to contain a suffix although they are loaned as a whole word. For example, the word *tutor* does not have an English base *tut*. Nonetheless, formations like *tutee* are possible. These formations seem to exchange the non-existent suffix rather than be added to the roman base, hence a paradigmatic process.

The analysis proposed here will show that the derivational process operates on the same eventualities. Hence, the paradigmatic process is shown by the switch of the reference from one participant to another in the same eventuality. Some of the bases in this subset appear to be derived but are non-derived. For example, according to the OED (2025), the word *tutor* was loaned from French or Spanish as a complex word and not derived within the morphological system of English. Nonetheless, it is possible to form *tutee* parallel to *tutor* by changing the apparent suffix although the *-or* form is non-derived. The semantic representation of the *-er/-or* bases and the reference shifts to a different participant are more straightforward due to the fact that the eventuality is already given by the respective *-er/-or* forms and is therefore more readily available.

4.1.2.1 *tutee*

Two examples are illustrated in detail, *tutee* and *mentee*. Examples of the derivative *tutee* are given in (4):

(4) ‘I used to have a **tutee** who lived there.’ she told Guido. ‘A tutee.’ Guido laughed at that. ‘Little miss schoolmistress.’ (BNC, Davies 2004-, FIC, 1993)

The derivative is based on the noun *tutor*. The noun *tutor* is sufficiently more frequent than the verb to assume that the noun serves as the base. The word *tutor* occurs as a loan word around 1500[12] and *tutee* in the 1920's (OED 2025 s.v. *tutor*). The verb *tutor* occurred in 1590. However, the meaning of the verb is definitely based on the eventuality given by $tutor_{N}$. Due to the fact that the verb is built on the eventuality denoted by the loan word and the immensely higher frequency of the noun, the assumption of a paradigmatic relation between the two nouns, *tutor* and *tutee*, is straightforward. Additionally, this relation is visible due to the change of the alleged suffix. Figure 4.9 illustrates the frame for *tutee*.

$$\begin{bmatrix} \textit{lexeme} \\ \text{PHON /tjuː ˈtiː/} \\ \text{CAT N} \\ \text{M-BASE} \begin{bmatrix} \textit{lexeme} \\ \text{PHON /ˈtjuːtə/} \\ \text{CAT N} \\ \text{SEM } \boxed{0} \begin{bmatrix} \textit{teaching-action} \\ \text{AGENT } \boxed{1} \\ \text{PATIENT } \boxed{2} \end{bmatrix} \\ \text{REF=} \left\{ \boxed{1} \right\} \end{bmatrix} \\ \text{REF=} \left\{ \boxed{2} \right\} \end{bmatrix}$$

FIGURE 4.9: Representation of the lexeme *tutee*.

The *teaching-action*[13] is directly accessible from *tutor*. The reference of *tutor* is on the element indexed with $\boxed{1}$, as *tutor* denotes the AGENT of the *teaching-action*. The derivative *tutee*, on the other hand, has a different REFERENCE-attribute which points to a different entity, namely the PATIENT of the *teaching-action* indexed with $\boxed{2}$. The two lexemes refer to different participants in the same eventuality to create the meanings of *tutor* and *tutee*.

[12]It occurred even earlier in nowadays obsolete contexts.

[13]The special interaction of student and tutor is probably not entirely described with the label *teaching-action*, as this relation may contain more than teaching. However, for the illustration of the derivational process, the label is sufficient.

4.1.2.2 *mentee*

A second example of an *-ee*-derivative with a pseudo-derived base is the derivative *mentee* in (5):

(5) One corporate lawyer scratched his tradition of grabbing Christmas drinks with a female **mentee**, and opted for the safer alternative of lunch. (COCA, see Davies 2008- NEWS, 2018)

The derivative *mentee* denotes the person who is supervised. The related word *mentor* is, as well as *tutor*, a loan word from French. The noun *mentor* is sufficiently more frequent than the verb, which leads to the assumption that the noun serves as the base. The first attestation of *mentor*$_{\mathrm{N}}$ in the OED (2025 s.v. *mentor*) is from 1750 and *mentee* occurs in 1965. The verb *mentor* occurred in 1918 and is based on the eventuality given by *mentor*$_{\mathrm{N}}$. The paradigmatic relation between the two nouns is straightforwardly visible due to the change of the alleged suffix and the operation on the same eventuality.[14]

The semantic representation of *mentee* in Figure 4.10, then, looks similar to the one of *tutee*. The specification of the eventuality is changed into a *supervising-action*. *Mentor* denotes the AGENT and *mentee* the PATIENT. Both forms operate on the same eventuality but target a different participant of this eventuality (*supervising-action*). Thus, the word-formation process with the suffix *-ee* works similarly for the bases in section 4.1.1 and *-er*/*-or* bases as a non-volitional and sentient participant is referred to.

[14] Some might argue that the verb *mentor* also exists and the addition of the suffix *-ee* for the form *mentoree* is evidence for a verbal base in the first place. However, the applied frequency criterion described in Chapter 3 results in mentor$_{\mathrm{N}}$ as being far more frequent and hence the base for *mentee* and *mentoree*.

$$\begin{bmatrix} \textit{lexeme} \\ \text{PHON } \text{/mənˈtiː/} \\ \text{CAT N} \\ \text{M-BASE} \begin{bmatrix} \textit{lexeme} \\ \text{PHON } \text{/ˈməntə/} \\ \text{CAT N} \\ \text{SEM } \boxed{0} \begin{bmatrix} \textit{supervising-action} \\ \text{AGENT } \boxed{1} \\ \text{PATIENT } \boxed{2} \end{bmatrix} \\ \text{REF=} \left\{ \boxed{1} \right\} \end{bmatrix} \\ \text{REF=} \left\{ \boxed{2} \right\} \end{bmatrix}$$

FIGURE 4.10: Representation of the lexeme *mentee*.

4.1.3 Volitional participant denotation: *bargee*

The restrictions for the suffix *-ee* described in Section 2.3.1 are that the derivational process with the suffix *-ee* requires a participant which has to be sentient and non-volitional. These restrictions also underlie all of the examples formalized in sections 4.1.1 and 4.1.2 and can be regarded as the basis of the regular semantics of *-ee*-nominalizations. However, exceptions to the non-volitionality constraint can be found in the literature on *-ee* for deverbal and denominal nominalizations. More precisely, Barker (1998) pointed out that some derivatives with *-ee* can also refer to volitional entities. One example is *escapee*, which, just as the also existing derivative *escaper*, denotes the AGENT of the *escaping-action* from the verb *escape*. According to Barker (1998), this is possible for *-ee*-derivatives with bases that are either intransitive verbs, or verbs that can be interpreted as having a non-volitional subject participant, like *stand*, for example. The data on denominal derivatives (Barker 1998; Plag 2004; Mühleisen 2010; Chapter 3) also shows *-ee*-derivatives which denote AGENTS like *bargee* in the examples in (6).

(6) a. It is often a matter of difficulty to find out exactly what will have to be paid for haulage, toll and wharfage, and this fact increases the office labour in sending out tenders. Then again, in dealing with the **bargee** himself there is often trouble. The owner of a couple of narrow boats and his crew are at a great disadvantage when pitted against such competitors [...] (BNC, see Davies 2004-, MAG, 1982)

b. 'I'm a **bargee**, owner-operator, that's my ship down there,' she said, pointing down the side of the quay. (BNC, see Davies 2004-, FIC; 1985-1994)

The base for *bargee* is the noun *barge*, whose semantic structure is represented in the frame in Figure 4.11. The frame specifies that *barge* is a lexeme with a phonetic representation (PHON) and the syntactic category noun (CAT N). The semantics of *barge* (SEM $\boxed{0}$) is an *entity* as the lexeme *barge* denotes a concrete object. Two attributes are contained in the semantic representation, a PRODUCTION-attribute ($\boxed{1}$) which is analogous to the AGENTIVE-QUALE in Pustejovsky (1996: ch. 6), and an AFFORDANCE-attribute ($\boxed{2}$) which is analogous to the TELIC-QUALE in Pustejovsky (1996: ch. 6). The entity in $\boxed{0}$ is produced (PRODUCTION $\boxed{1}$) by being built (*building-action*) by an AGENT ($\boxed{3}$), and the entity is created for a specific purpose (AFFORDANCE $\boxed{2}$) which is a *shipping-action*. The *shipping-action* is performed by an AGENT ($\boxed{4}$)[15] and transfers a THEME ($\boxed{5}$) which is usually of the type *freight*. The REFERENCE-attribute (REF) indicates that the lexeme *barge* denotes the *entity* in SEM $\boxed{0}$.

[15] It is, in theory, possible that the AGENTS in $\boxed{3}$ and $\boxed{4}$ are the same person, which would be indicated by co-indexation.

lexeme
PHON /bɑːˈʤ/
CAT N
SEM [0]
  boat
  PRODUCTION [1]
    building-action
    AGENT [3]
    THEME [0]
  AFFORDANCE [2]
    shipping-action
    AGENT [4]
    THEME [5] *freight*
REF= { [0] }

FIGURE 4.11: Representation of the lexeme *barge*.

The frame in Figure 4.12 specifies that the derivative *bargee* is a lexeme with a phonetic representation (PHON) and the syntactic category noun (CAT N). The lexeme is a derivative, i.e., it has a morphological base (M-BASE) which is the lexeme *barge* in Figure 4.11. The difference of the two lexemes lies in their denotation, described in the REFERENCE-attribute (REF): The base denotes the entity indexed with [0] and the derivative denotes the element indexed with [4], the AGENT of the *shipping-action*, hence the person steering the barge. As no sentient, non-volitional participant is part of the semantics of the base, the volitional, sentient participant indexed with [4] is used for the derivational process instead. Theoretically, the participant indexed with [3], the AGENT of the *building-action*, could be a potential referent. However, no textual evidence was found that *bargee* refers to the builder of a barge but to its bargeman.

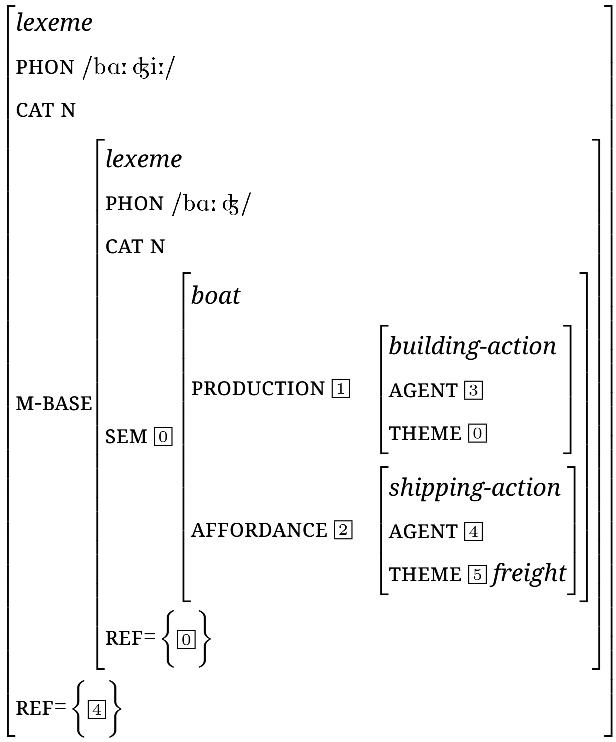

FIGURE 4.12: Representation of the lexeme *bargee*.

4.1.4 Summary of readings

It was shown that derivatives with the suffix *-ee* refer to sentient, non-volitional participants in an eventuality denoted either by non-derived bases as in Section 4.1.1 or by paradigmatically related bases in Section 4.1.2. Nevertheless, as for deverbal derivatives in *-ee*, exceptions can be found to the non-volitionality constraint for denominal derivatives as shown in Section 4.1.3. Thus, two different semantic representations can be achieved by a derivational process with the suffix *-ee*: A non-volitional or a volitional, sentient element in the base's eventuality. Such divergent semantic classes of a word-formation process can be captured in inheritance hierarchies (Riehemann 1998; Koenig 1999; Booij 2010; Bonami & Crysmann 2016; Plag et al. 2018). Figure 4.13 shows such an inheritance hierarchy for denominal nominalizations derived by the suffix *-ee*. The abbreviation *n-n-lfr* stands for noun-to-noun lexeme formation rule and indicates that the categories under this node describe denominal nominalizations. The hierarchy severs the phonological component (PHON) of the word-formation process from different semantic (SEM) categories.

On the left side of the inheritance hierarchy, the phonological realization (PHON) of the derivative is described. The phonology of the base is modified by adjusting the phonological form according to the information of the suffix, in this case the suffix *-ee* . On the right side, the semantic specification of a lexeme-formation rule is illustrated. The left node describes the *regular semantics* investigated in Sections 4.1.1 and 4.1.2. In order to derive the *regular semantics*, a *non-volitional-participant-noun*, the base (M-BASE) must provide but not necessarily denote an eventuality with a participant which is non-volitional and sentient (PARTICIPANT). This constraint is described by the reference (REF) which is on the element indexed with $\boxed{x}$. The individual derivatives discussed in this dissertation have their reference on a participant which meets the constraints given in the *regular semantics* node in the inheritance hierarchy. The derivatives discussed in sections 4.1.1 and 4.1.2, e.g., *biographee* and *tutee*, connect the phonological component of the word-formation process with its regular semantics. These forms are listed below each other.

The right semantic node illustrates the semantic make-up of the exceptions as in Section 4.1.3. The participant which is referred to by the *-ee*-derivative is sentient and volitional, which leads to a *volitional-participant-reading* as in *bargee*.

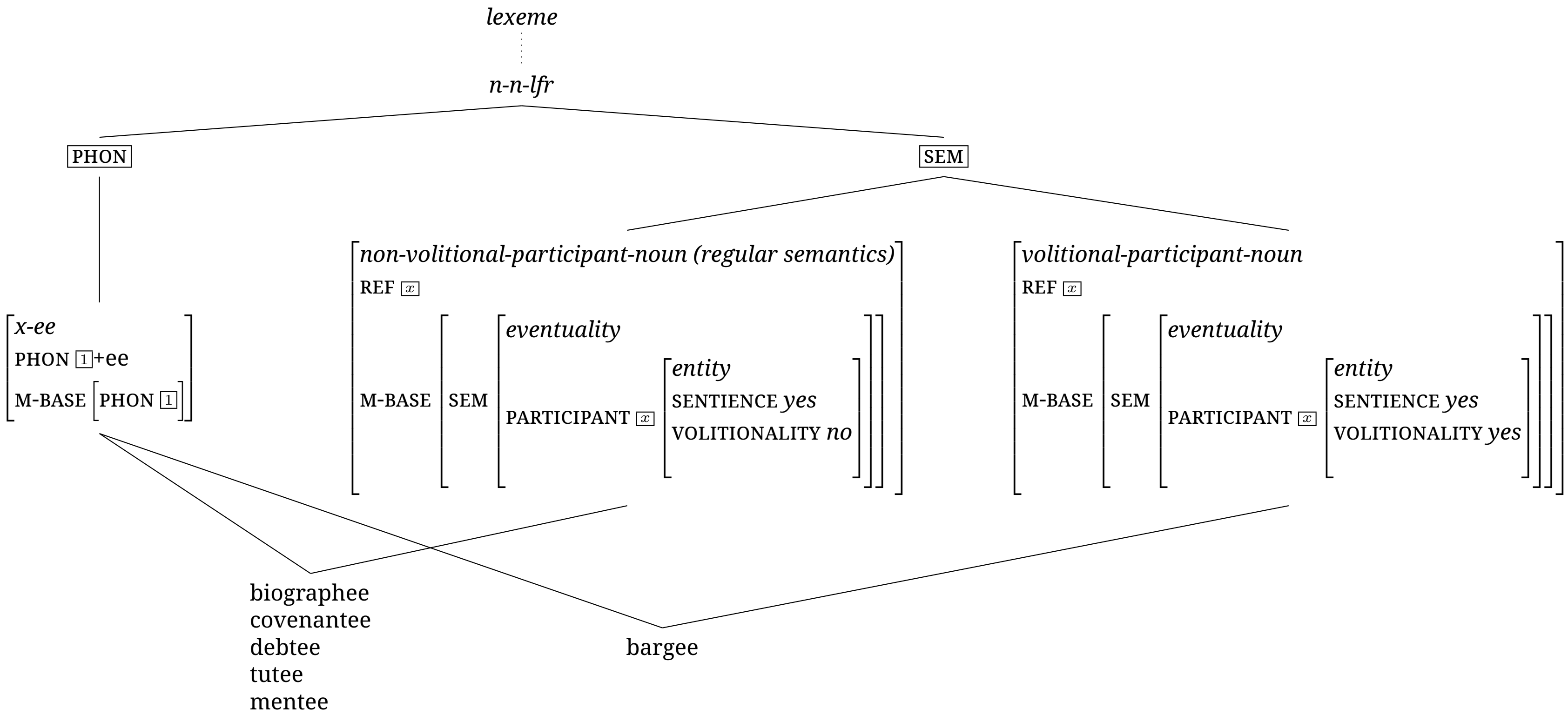

FIGURE 4.13: Inheritance hierarchy of lexical rules for *-ee*.

4.2 Nominalizations with the suffix *-ment*

The suffix *-ment* has the potential to induce a range of referential shifts, where eventuality-related elements in the base verb frame serve as the target. Denominal formations, on the other hand, pose a problem to the reference shifting approach, as it relies on decidedly eventive structures in the base. These eventive structures are not as straightforwardly identifiable in the nominal base as nouns do usually not denote eventualities. Further decomposition of the base is needed, similar to the analysis of nominalizations in *-ee* in Section 4.1. Section 4.2.1 will show the successful application of the reference shifting approach to psych nouns as bases. I chose to call the base category psych nouns analogous to Kawaletz (2023: ch. 5). Usually, psych nouns have a change in the psych state of an experiencer, e.g., *illusionment*, *allurement*, *concernment*. More precisely, the underlying assumption is that all psych expressions that serve as base for nominalizations with the suffix *-ment* operate on change-of-psych-state-causation structures. Section 4.2.2 will deal with the issue of non-eventive attitudinal nouns as nominal bases. All attitudinal person nouns attested as base for nominalizations with the suffix *-ment*, e.g., *devilment*, *rascalment*, *bastardment*, allow for analyses as participants of activities. Attitudinal nouns are purely entity-denoting nouns that typically refer to a person (see, for example, Schmid 1999; Paradis 2008; Morzycki 2009).

4.2.1 Psych nouns as bases

Turning to psych nouns as bases for eventuality-related nominalizations, the examples in (7) illustrate some of them. The reference shifting approach applied in this dissertation models the semantic structure of psych nouns as bases for a derivational process with the suffix *-ment* like psych verbs.

(7) allure, anger, concern, confusion, illusion

The term psych nouns is related to psych verbs as discussed in Kawaletz (2023: ch. 5). Psych verbs are verbs that denote an emotional or psychological event or state, for example, frighten or fear. Psych verbs have widely been discussed with their syntactic properties in Italian (Belletti

& Rizzi 1988). The research on psych verbs is, however, often focused on their atypical argument realization.

The examples in (8) illustrate psych nouns. The definitions from the OED for *illusion* and *allure* already show a change-of-psych state eventuality. For illude the state changes to an illuded state ('one is illuded') and for allure the change of state includes a change in 'attraction'. Thus, psych nouns behave similar to psych verbs.

(8) a. The action of illuding, the condition of being illuded; that whereby one is illuded. (OED 2025 s.v. *illusion*)
    b. Attraction, fascination, allurement; an instance of this (OED 2025 s.v. *allure*)

The semantic analysis of psych nouns in this thesis is based on the frame semantic analysis by Kawaletz (2023: ch. 5) for psych verbs. The frame in Figure 4.14 describes a psych verb which is of the type *lexeme*. It has a phonological representation (PHON), a syntactic category (CAT V) and semantics (SEM).

$$\begin{bmatrix} \textit{lexeme} \\ \text{PHON } \boxed{x} \\ \text{CAT V} \\ \text{SEM } \boxed{0} \begin{bmatrix} \textit{change-of-psych-state causation} \\ \text{AGENT } \boxed{1} \\ \text{CAUSE } \boxed{4} \begin{bmatrix} \textit{eventuality} \\ \text{STIMULUS/INSTUMENT } \boxed{2} \end{bmatrix} \\ \text{EFFECT } \boxed{5} \begin{bmatrix} \textit{change-of-psych-state} \\ \text{RESULT-STATE } \boxed{6} \begin{bmatrix} \textit{psych-state} \\ \text{EXPERIENCER } \boxed{3} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{REF=} \left\{ \boxed{0} \right\} \end{bmatrix}$$

FIGURE 4.14: The semantic representation of an experiencer psych verb base (adapted from Kawaletz (2023: 131)).

The psych verbs denote *change-of-psych-state causation* events. In Figure 4.14, a generalization over such events is introduced in the SEM-attribute. The complex causative event is made up of two sub-events, i.e., CAUSE and

EFFECT. Both sub-events contain typical (but not necessarily obligatory) participants, namely, AGENT, STIMULUS, INSTRUMENT, and EXPERIENCER. In a marginalized structure, a psych-causation frame is event-structurally fixed and includes a causative sub-event and a second sub-event, i.e., an EFFECT, during which the EXPERIENCER attains a *psych-state* as the result of the whole *change-of-psych-state causation*. The semantics of the psych verb *to enrage*, for example, would type the RESULT-STATE [6] value as, roughly, a *furious-state*.

Example (9) shows the OED definition for the nominalization *allurement*. The definition indicates that a change-of-state is happening.

(9) Alluring faculty or quality; attractiveness, appeal; fascination, charm. Also: an instance of this. (OED 2025 s.v. *allurement*)

The prototypical derivative in Figure 4.15 is a lexeme with a phonological representation (PHON) and a syntactic category (CAT N), analogous to the analysis in Kawaletz (2023: ch. 5). The semantics of the morphological base (M-BASE) are embedded. As indicated by the co-indexation of their respective semantics attributes (SEM [0]), the derived lexeme inherits the entire semantic structure of its base. However, the two lexemes differ with regard to their possible referent nodes indicated in the REFERENCE-attribute (REF), i.e., the base verb can only denote the complex event indexed with [0]. The resulting *-ment*-derivative is polysemous between different eventuality-related readings (REF = {[0], [2], [4], [5], [6]}). Besides making reference to the sub-eventualities *change-of-psych-state* [5] and *psych-state* [6], a derivative like *enragement* can also refer to the whole event [0], to the CAUSE [4], or to the event participant STIMULUS/INSTRUMENT [2] (see, Kawaletz 2023: ch. 5 for details).

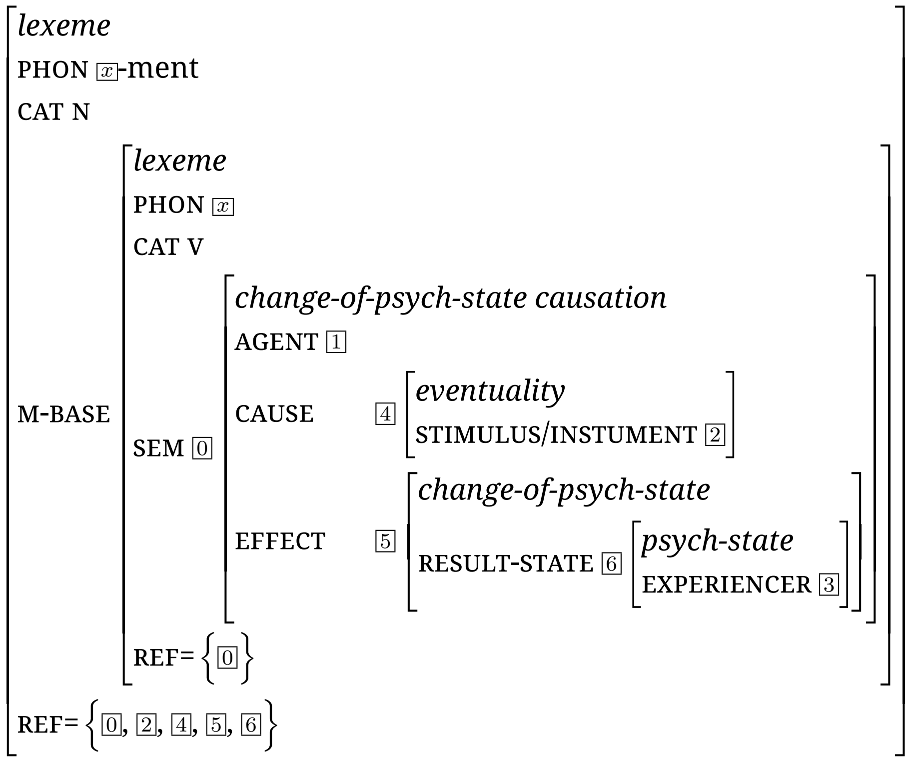

FIGURE 4.15: The word formation with *-ment* on object experiencer psych verb bases (adapted from Kawaletz (2023: 131)).

To summarize, the lexeme frame in Figure 4.15 illustrates the process of *-ment*-nominalizations on "object experiencer psych verb bases" (cf. Kawaletz 2023: ch. 5). This process can be modeled as possible referential shifts on the base semantics and can unproblematically be extended to other deverbal nominalizations.

Turning to the psych nouns as bases, the assumption is that psych nouns have the same make-up as psych verbs. More precisely, the eventuality in a psych noun is a *change-of-psych-state causation*. Thus, the derivational process with the suffix *-ment* with psych nouns as bases is similar to nominalizations with psych verbs as bases.

4.2.1.1 *illusionment*

The first *-ment*-derivative under investigation is *illusionment*, which is based on the noun *illusion*. In order to understand the meaning of *illusionment*, example (10) depicts the usage of base and derivative. In (10-a), *illusionment* is defined as an individual's complex mental state that consists of several illusions, i.e., false beliefs of reality. In contrast, *illusionment* in (10-b) refers to a more complex eventuality, in which an individual creates such a false belief of reality. While the possibility of further related senses of the derivative is not excluded, these two attestations exhaust the readings found.

(10) a. [...] a system of intertwined fundamental illusions that had always been lived within [...] This way of being that one recognizes only retrospectively may be called **illusionment**[...][16]

b. Winnicott's emphasis of the importance of the baby's capacity for **illusionment** draws directly on Freud's description of the baby's ability magically to conjure up a phantasy or hallucination of the mother's breast before it eats it.[17]

The two examples in (10) refer to different components of causative events that affect psychological states: Either the whole complex of a causative macro-event in example (10-b), or merely one element of such an event, its RESULT-STATE in (10-a). Applying the reference shifting approach to the derivative *illusionment*, the base provides the eventive elements the derivative potentially shifts the reference to.

Starting with the semantic structure of the noun, *illusion* is itself an eventive noun. The readings of the base noun *illusion* illustrated in (11) show that the semantics of the base is best understood as a change of a psych state causation eventuality, analogous to the psych verb semantics illustrated in Figure 4.14 based on the analysis in Kawaletz (2023). In example (11-a), *illusion* denotes the RESULT-STATE of a causation eventuality, i.e., a *false-belief-state*. In contrast, *illusion* in example (11-b) denotes the STIMULUS/INSTRUMENT element in a causation eventuality that brings something about, or is used for bringing about, the change of state. The nature of *illusion* as a pseudo-nominalization explains some parts of the conceptual structure assumed.[18] Thus, in example (11-a), the whole

[16]Margulies, A. 2018. Illusionment and Disillusionment: Foundational Illusions and the Loss of a World. *Journal of the American Psychoanalytic Association* 66(2): 289.

[17]Minsky, R. 2014. *Psychoanalysis and Gender: An Introductory Reader*. 2nd ed. NY: Routledge. [via www.googlebooks.com, n.p.]

[18]Although the lexeme appears to be a nominalization itself, with the object-experiencer verb *to illude* (roughly meaning 'to trick, to deceive someone') as its potential base, it is more likely that the noun was loaned directly from French (see also the

causative component as well as the eventive elements engaged in it are left implicit, while in example (11-b) the experiencer is not spelled out. The example in (8-a) is taken from the OED (2025 s.v. *illusion*).

(11) a. She wasn't [...] under the **illusion** that marriage was a relationship characterized by endless bliss and romance. (COCA, see Davies 2008-, WEB, 2012)

b. I am surprised Jean hasn't tried to use an **illusion** to appear and sound how he used to. (iWeb, see Davies 2018-)

Building on these considerations, the frame in Figure 4.16 models *illusion* as a simplex lexeme with a semantic structure that is analogous to the one for morphological bases of psych verb nominalizations as in Figure 4.14. Figure 4.16 includes specifications of the phonology (PHON), the syntactic category (CAT), the semantics (SEM), and possible referents (REF) in the form of attributes embedded in the semantics structure of the lexeme.

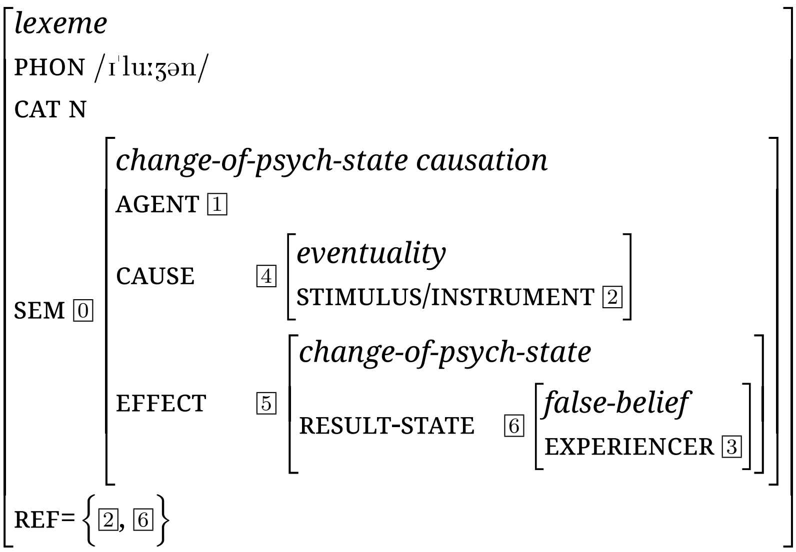

FIGURE 4.16: Representation of the lexeme *illusion*.

Figure 4.16 analyzes the meaning of *illusion* as the potential to make reference to different nodes in the structure of a *change-of-psych-state causation* eventuality, indexed with 0. More precisely, *illusion* is polysemous and its denotation is context-dependent. This polysemy of the base goes

OED online entry (OED 2025 s.v. *illusion*). Nominalizations based on psych verbs do often not refer to the whole causation eventuality but only to sub-parts of this eventuality (cf. Kawaletz 2023: ch. 5).

hand in hand with different possible denotations of the underlying *change-of-psych-state causation* eventuality. The meaning of *illusion* is different from other subtypes of *change-of-psych-state causation* events because of the specification of the RESULT-STATE (indexed [6]). Reflecting the semantics of an illusion, this RESULT-STATE is specified as a *false-belief* and it takes the EXPERIENCER as participant (indexed [3]). In other words, an eventuality of the type *illusion* will result in an EXPERIENCER holding a false belief of reality. The REFERENCE-attribute (REF) spells out the referential potential of the lexeme *illusion*. As illustrated by the examples in (11), *illusion* refers either to the RESULT-STATE (indexed [6]; in (11-a)) or to the STIMULUS/INSTRUMENT (indexed [2]; in (11-b)) in the denoted eventuality.

Turning to the *-ment*-nominalization *illusionment*, the examples in (10) illustrate that the derivative can either refer to a *change-of-psych-state causation* eventuality or to the RESULT-STATE of such an eventuality. The semantics of the base word *illusion* include such a *change-of-psych-state causation* eventuality. Hence, the required elements for the reference shift induced by the word-formation process with the suffix *-ment* to form *illusionment* are given in the semantic structure of the base *illusion*. Figure 4.17 depicts the semantics of the lexeme *illusionment* as a complex lexeme in an AVM. The semantic frame for the lexeme *illusion* in Figure 4.16 is included as the morphological base (M-BASE) of the nominalization.

The frame in Figure 4.17 reflects the assumption of the reference shifting approach that *-ment* has no lexical meaning of its own but operates on the base's semantics to change the reference from the denotation of the base to the denotation of the *-ment*-derivatives. Hence, eventuality-related readings of a derivative are based on eventive elements provided by the eventuality denoted by the semantic structure of the base. First, the *-ment*-derivative *illusionment* does not specify a separate semantic contribution but operates on the base semantics of *illusion*. This is indicated by co-indexation as [0] of the base's and the derivative's respective SEM-attributes. Second, as shown in Figure 4.16, the semantics introduced by the base provides the *change-of-psych-state causation* eventuality that the attested readings of *illusionment* call for. Importantly, modelling the base-derivative pair *illusion-illusionment* as sharing the same semantic structure does not entail that the two lexemes are semantically completely synonymous to each other. In the frame for the derivative in Figure 4.17, the differences in attested readings between the two forms are captured via the referential potentials indicated in the REFERENCE-attributes. First, base and derivative share the capacity to refer to the RESULT-STATE [6]. Second, however, the base *illusion* can denote the STIMULUS/INSTRUMENT [2] of the eventuality but not the complex causation eventuality indexed with [0], while the derivative *illusionment* shows the reverse potential. Importantly, the referential space in the frame in Figure 4.17 (i.e., REF= {[0], [6] }) is based on the specific attestations found for *illusionment*, i.e., more readings referring to different eventive elements in the frame are imaginable.

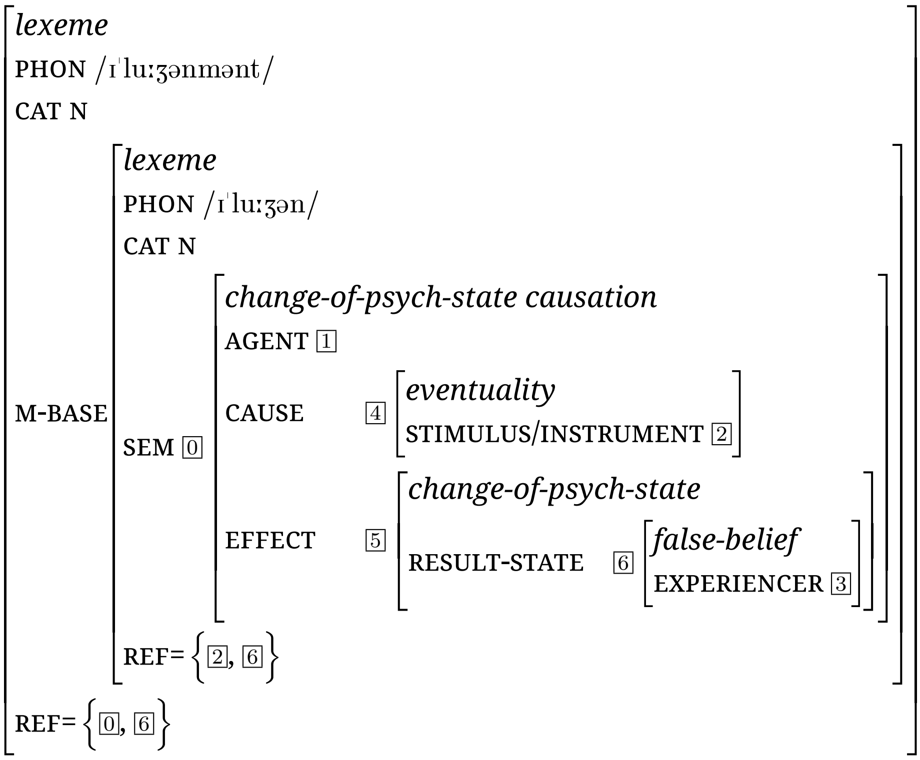

FIGURE 4.17: Representation of the lexeme *illusionment*.

4.2.1.2 *allurement*

The next exemplary nominalization for a semantic analysis is *allurement*, which is based on the psych noun *allure*.[19] The examples in (12) illustrate the meaning of the base.

(12) a. The new-ish: Cutscenes are often viewed as the tedious buzz-killers of gaming, so it's rare to see a title that embraces them so wholeheartedly as Asura's Wrath. It's a different sense of player agency than gamers might be used to, and "different" holds some **allure**. (COCA, see Davies 2008-, BLOG, 2012)

b. And then what did the first eye see? Perhaps it was the image of love, or terror, or just some blurry blip relayed to a ganglion somewhere. But imagine the pressure to create enough **allure** to convince the eye it was worth staying! (COCA, see Davies 2008-, FIC, 2004)

The derivative *allurement* refers to the state of being attracted or captivated by someone or something as in example (13-a). It can also refer to the act of enticing or attracting someone or something, usually by appealing to their desires or interests as in (13-b). Both physical and non-physical attraction can be described by *allurement*.

(13) a. It was the sweet **allurement** of the mimosa tree in full bloom that finally overcame my fears. (COCA, , see Davies 2008-, WEB, 2012)

b. The charm of this being, usually grum, more odd than pretty, lay in her eyes, a pair of large green eyes, cloudy-green, sides-of-a-fish as the Romanian says, heavy-lashed and overbrowed, with a gaze somewhat lost. Did any other **allurement** weave the net that caught the duke? Possibly. (COCA, see Davies 2008-, FIC, 2014)

Interestingly, the examples in (12) show the same behavior for the base noun *allure* as the derivative in (13). Even the definition of *allure* in example (8-b) shows a semantic connection of base and derivative. In general, *allure* and *allurement* imply a strong attraction or appeal that draws

[19] The verb *allure* also exists. However, it is infrequent in the corpora used to determine the word class of a base as described in Chapter 3.

someone or something towards a particular object or goal. Base and derivative have the potential to refer to the same elements in the frame, i.e., the derivative shows a clear tendency for a transpositional reading. To what extent both words are interchangeable in contexts is not of interest for the analysis at hand.

Figure 4.18 illustrates the semantic representation of the psych noun *allure*. It resembles the structure for psych nouns and the analysis of *illusion* (Section 4.2.1.1).

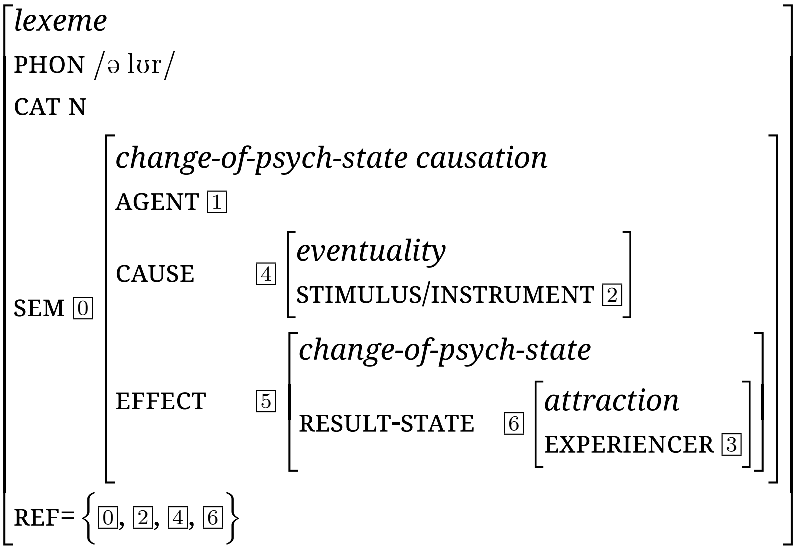

FIGURE 4.18: Representation of the lexeme *allure*.

The derivative *allurement* is based on the psych noun *allure* and Figure 4.19 depicts its semantic analysis. As *allurement* is based on *allure*, the representation in Figure 4.18 is included as morphological base (M-BASE) for the nominalization. The difference lies in the form, as the derivative includes the suffix *-ment* , and the referential potential. The examples in (13) show that its meaning is transpositional. It refers ot the same eventive elements the base *allure* can refer to (REF 0, 6). It can not refer to the INSTRUMENT/STIMULUS indexed with 2, probably due to the requirements for nominalizations with *-ment* (see Chapter 2, Section 2.3.2), i.e., *-ment* cannot shift the reference on an animated entity, and a STIMULUS can be animated. This restriction does not forbid the reference shift onto other elements than the ones indicated in the REFERENCE-attribute given by the semantic representation of the base.

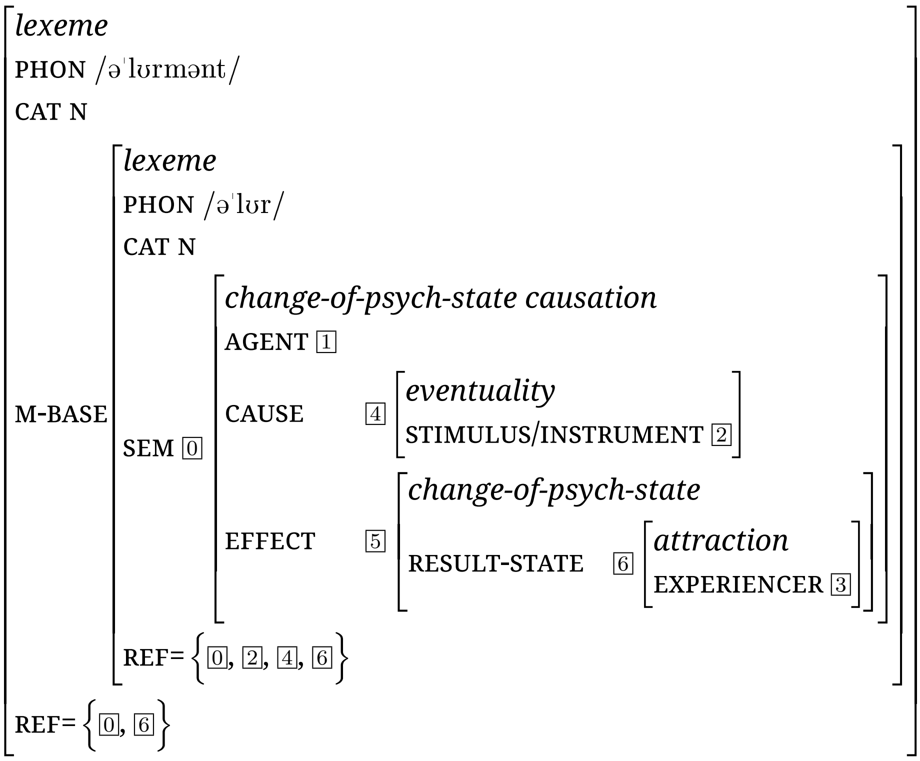

FIGURE 4.19: Representation of the lexeme *allurement*.

Summarizing, the nominalizations *illusionment* and *allurement* operate on eventive elements given by their base words *illusion* and *allure*. The reference shifting approach applied for the suffix *-ee* (Section 4.1) and for *-ment* by Kawaletz (2023) is applicable for nominalizations operating on psych nouns. The next chapter will show the operationalization of nominalizations with the suffix *-ment* on different nominal bases, i.e., attitudinal nouns.

4.2.2 Attitudinal nouns as bases

Attitudinal nouns mostly denote people, and the semantics is specialized in a way that they profile the speaker's stance toward single (or highly restricted sets of) behavioral or character traits of the entities denoted by such attitudinal nouns (see, for example, Schmid 1999; Paradis 2008; Morzycki 2009). The examples in (15) give some attitudinal nouns in the data with the suffix *-ment*.

(14) bastard, butcher, devil, imbecile, rascal

First, these traits tend to be gradable, and and this is the reason why attitudinal nouns tend to be gradable as well. This gradability is illustrated in example (15-a) showing that all five listed nouns are compatible with the degree modifiers *total*, *real*, and *complete*. The attestation in example (15-b) illustrates that *devil* displays the same behavior as attitudinal nouns. Second, compatibility with the progressivized copula, i.e., *being a X*, as in example (16-a), and with command imperatives as in example (17-a) show that attitudinal nouns allow for agentive contexts, a diagnostic that speaks in favor of events as part of their semantics (see Lakoff 1966; Maienborn 2003 on these test environments).

(15) a. You are a total/real/complete genius/bastard/rascal/bugger/imbecile etc.
b. Well, it wasn't always so nice either, for she was pretty tough at times too. A **real devil** she could be, the one I got for a wife. (COCA, Davies 2008-), WEB, 2012)

(16) a. You are being a genius/bastard/rascal/bugger/imbecile etc.
b. Either way, he does seem to take delight in **being a devil**, as his various antics – dressing up like a cop; chopping off someone's hair; strapping his dog to the car roof – all too handily reveal. (COCA, Davies 2008-), BLOG, 2012)

(17) a. (Don't) be a genius/bastard/rascal/bugger/imbecile etc.
b. **Be a devil** and stop being so staid... (BNC, see Davies 2004-, FIC)

The noun *devil* as well as the noun *rascal* behave like attitudinal nouns. The next sections (4.2.2.1, 4.2.2.2) analyze the semantic representation of these nouns. Furthermore, the corresponding nominalizations *devilment* and *rascalment* hold for exemplary nominalization with the suffix *-ment* based on attitudinal nouns.

4.2.2.1 *devilment*

The derivative *devilment* can denote events, while its base *devil* is an entity-denoting noun, as the examples in (18) illustrate again.

(18) a. Well, it wasn't always so nice either, for she was pretty tough at times too. A **real devil** she could be, the one I got for a wife. (COCA, Davies 2008-), WEB, 2012)

b. Either way, he does seem to take delight in **being a devil**, as his various antics – dressing up like a cop; chopping off someone's hair; strapping his dog to the car roof – all too handily reveal. (COCA, Davies 2008-), BLOG, 2012)

c. A very wicked or cruel person; (in Middle English sometimes) a man of gigantic stature or strength, a giant. In weakened sense: a person who is very difficult to deal with (OED 2025 s.v. *devil*)

d. A sin, vice, or evil quality personified; a personification of a particular undesirable quality by which a human being may be possessed or actuated.(OED 2025 s.v. *devil*)

The examples in (19), however, show that the derivative *devilment* can be found in at least two different readings. In (19-a), the derivative denotes an activity, as indicated by the predicate *goes on* and the predicative complement *what we call deceptive practices*. Example (19-b), in contrast, is less transparent, because *devilment* denotes a property or characteristic of the addressee rather than an activity. The definition provided by the OED (2025 s.v. *devil*) in (18-d) underlines the denotation of *devilment* in the other examples. In other words, the reading of *devilment* denoting characteristics of the entity described by the base nouns is more hidden in the semantic structure of the base, compared to the activity-reading.

(19) a. [...] the biggest **devilment** that goes on in these elections are what we call deceptive practices – people are going to get robocalls [...] (COCA, see Davies 2008-, WEB, 2012)

b. She's of a mind it'll wash any **devilment** right out of me. (COCA, see Davies 2008-, FIC, 2019)

c. Action performed by, or characteristic, of the Devil or a devil; evildoing, mischief; an instance of this. Also: a devilish, rascally, or rash disposition; mischievousness; wild spirits. (OED 2025 s.v. *devilment*)

Neither of the two examples in (19) make concrete reference to *the Devil* in the religious sense of God's adversary, and neither do any of the 44 *devilment*-attestation in COCA (Davies 2008-). The definition in (19-c) by the OED (2025 s.v. *devilment*) states the same, rather, it is the (mostly) negatively evaluated behavior of, or characteristics associated with, *the Devil* that are metaphorically shifted to more general actors or bearers. Unsurprisingly, this metaphorical shift already operates on the base *devil*. For example, the OED (2025 s.v. *devil*) definitions in example (18-c) and (18-d) acknowledge both the senses of a 'wicked or cruel person' and of an 'evil quality personified'. In these usages, *devil* functions as an attitudinal person noun.

Building on these considerations, *devilment* is clearly based on the attitudinal noun *devil*. Given the characteristics of attitudinal nouns, any decompositional approach will have to account for the base gradable structures. Furthermore, the approach has to account for the intensity of the characteristics of the noun as well as for its eventive meaning components.

Figure 4.20 illustrates the semantic representation of the nominal base *devil* as a type of *lexeme* with its phonological properties (PHON) and its syntactic category (CAT N). The semantics of the word (SEM) are further specified as *person*. This label was chosen as a devil is an animated entity which acts like a person. The *person* contains a node for ACTIVITY ∧ HABITUALITY. This node denotes the characteristics of the base noun as well as its activity component inherent to attitudinal person nouns. For a devil, the ACTIVITY ∧ HABITUALITY are further specified as ⟨*wickedness, evilness etc.*⟩ according to the definitions in the OED (2025 s.v. *devil*) in examples (18-c) and (18-d). These properties of a devil are PROFILED PROPERTIES (PROF-PROP).

$$\begin{bmatrix} \textit{lexeme} \\ \text{PHON } /\text{ˈdɛvəl}/ \\ \text{CAT N} \\ \text{SEM } \boxed{0} \begin{bmatrix} \textit{person} \\ \text{ACTIVITY} \wedge \text{HABITUALITY } \boxed{1} \begin{bmatrix} \langle \textit{wickedness, evilness etc.} \rangle \\ \text{PROF-PROP } \boxed{2} \begin{bmatrix} \text{M-DIM} \begin{bmatrix} \textit{prop-scale} \\ \text{DEGREE } \boxed{3} \begin{bmatrix} \alpha \\ \copyright_{s,Rel}\ (\boxed{3},\boxed{4}) > \end{bmatrix} \\ \text{THRESHOLD-DEGREE } \boxed{4}\ \beta \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{REF} = \boxed{0} \end{bmatrix}$$

FIGURE 4.20: Representation of the lexeme *devil*.

In order to account for their scalar nature, the properties take property scales (*prop-scale*) as measure dimensions (M-DIM) and map degrees on said scales (cf. Kotowski 2023). Following standard assumptions on scalarity (see, e.g., Solt 2015; Kennedy & McNally 2005), degrees on an open property scale have to exceed some comparison degree for an entity to count as, here, for example, *evil* or *wicked*. The semantic frame in Figure 4.20 therefore introduces a THRESHOLD-DEGREE-attribute and build on a two-place COMPARATOR-attribute (see Löbner 2017). This comparator ($\copyright_{s,Rel}$ ($\boxed{3}$,$\boxed{4}$) $>$) states that the value α of DEGREE exceeds the value β of THRESHOLD-DEGREE on the property scale they apply to. In other words, the properties of *rascal* are of a higher value than the threshold. In used notation, © stands for an comparator, *Rel* for relation, and *s* for sort: thus, a comparator establishes a relation between elements of the same sort (such as colors, materials, heights, temperatures etc.). The values the comparator takes as input are co-indexed here. Broadly speaking, $\copyright_{s,Rel}$ *(*$\boxed{3}$,$\boxed{4}$*)* $>$ could be repeated as an attribute of value β. This would be redundant, however, as co-indexation within the comparators themselves declares which values are to be compared (see Löbner 2017 for details). The REFERENCE-attribute (REF) specifies the denotation of the base *devil*, i.e., the node indexed $\boxed{0}$, a *person*.

Crucially, attitudinal nouns such as *devil* neither denote activities nor properties as such but are best analyzed as a *person* which bears certain kinds of activities. Therefore, the reference of the noun *devil* is not on the property $\boxed{2}$, but the referential node is indexed with $\boxed{0}$ as indicated in the REFERENCE-attribute, the *person* itself bearing the ACTIVITY ∧ HABITUALITY-attributes.

The decomposition of the semantics of the base illustrated in the semantic frame thus captures three key ingredients of attitudinal (person) nouns. First, it includes the systematic possibility to refer to properties of person-entities. Second, these profiled properties are analyzed as scalar attributes that include dedicated measure dimensions. Third, with respect to properties of events, the analysis shows that the meaning of such nouns can be captured in a straightforward manner by a PARTICIPANT-attribute of an event-semantic structure. Broadly speaking, the denotation of attitudinal nouns is dependent on the characteristics and behaviors of the referent, e.g., on how a *devil* usually behaves.

Figure 4.21 illustrates the semantic frame for the derivative *devilment*. The semantic representation (SEM) of the derivative is the same as that of the morphological base (M-BASE) *devil* in Figure 4.20. The possible referential shifts are indicated in the REFERENCE-attribute (REF).

$$\begin{bmatrix} \textit{lexeme} \\ \text{PHON } /\text{ˈdɛvəlmənt}/ \\ \text{CAT N} \\ \text{M-BASE} \begin{bmatrix} \textit{lexeme} \\ \text{PHON } /\text{ˈdɛvəl}/ \\ \text{CAT N} \\ \text{SEM } \boxed{0} \begin{bmatrix} \textit{person} \\ \text{ACTIVITY} \wedge \text{HABITUALITY } \boxed{1} \begin{bmatrix} \langle \textit{wickedness, evilness etc.} \rangle \\ \text{PROF-PROP } \boxed{2} \begin{bmatrix} \text{M-DIM} \begin{bmatrix} \textit{prop-scale} \\ \text{DEGREE } \boxed{3} \begin{bmatrix} \alpha \\ ©_{s,Rel}\,(\boxed{3},\boxed{4}) > \end{bmatrix} \\ \text{THRESHOLD-DEGREE } \boxed{4}\ \beta \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{REF} = \boxed{0} \end{bmatrix} \\ \text{REF} \left\{ \boxed{1}, \boxed{2} \right\} \end{bmatrix}$$

FIGURE 4.21: Representation of the lexeme *devilment*.

The semantic frame in Figure 4.21 represents the activity- and property-readings of *devilment* (see the examples in (19)) corresponding to different nodes in the semantics of the base *devil*: first, the *activity* node indexed with [1] and, second, the node depicting the set of devil-properties [2]. Both nodes are available for the word formation process with the suffix *-ment*, and the context defines whether *devilment* refers to the activity-reading ([1]), or to the property-reading ([2]). The reference shift is illustrated in REFERENCE-attribute (REF). The derivative *devilment* refers to the element indexed with [1] (the whole ACTIVITY ∧ HABITUALITY) ([1]) or the element indexed with [2] (PROFILED-PROPERTIES).

The analysis of denominal nominalizations with the suffix *-ment* is similar to the proposed analyses of *-ment* on verbal and eventive nominal bases (i.e., Section 4.2.1 and Kawaletz 2023), as it makes use of referential shifts and relies on the base to provide the necessary eventive structures for the derivational process. Crucially, the feasibility of the referential shifting approach for *-ment* on attitudinal nouns is entirely reliant on the semantics of the base structure. The semantic decomposition of *devil* warrants assumptions of eventive elements as inherent parts of the base structure. The next Section 4.2.2.2 shows that the analysis of attitudinal nouns as bases for derivational processes with *-ment* is not only an individual case for *devil* and *devilment* but extendable to other attitudinal nouns as well.

4.2.2.2 *rascalment*

A second nominalization based on an attitudinal noun in the data set with the suffix *-ment* is *rascalment*. The example in (20-a) is the only found attestation of the derivative. The definition taken from the OED (2025 s.v. *rascalment*) in example (20-b) is the same as for *rascality*.

(20) a. When oul' Molly was a girl, Peig said, she was full of spirits and up to all the **rascalment** of the day.[20]

b. The world of rascals; rascals collectively; the rabble. (OED 2025 s.v. *rascalment*)

[20] McGill, Bernie. 2010. The Butterfly Cabinet: A Novel. London: Headline Review. [via www.googlebooks.com, n.p.]

The meaning of *rascalment* can only be understood if the addressee is familiar with the base word and its denotation in example (21). The noun *rascal* refers to a mischievous or dishonest person as in (21-c) who behaves in a playful, sometimes harmful way. It can denote someone who is up to no good, often engaging in minor acts of deception or trickery for their own amusement or benefit. It can also denote someone who is playful and impish, but not necessarily malicious. According to the meaning of *rascal* and *rascalment*, the semantic analysis is highly similar to the analysis for *devil* and *devilment* in the previous section (4.2.2.1).

(21) a. Barney, the **rascal**, brought it back with him yesterday and carried it about in his pocket all evening, never thinking of it once, " Johanna explained, shaking her fist at that guilty person just coming in. (COCA, see Davies 2008-, WEB, 2012)
b. Later they try and fail to reach Bastien on his mobile. That **rascal**, one would think that a Sunday evening could be the perfect time to bring it down a notch, to slow down just a bit.(COCA, see Davies 2008-, FIC, 2019)
c. An unprincipled or dishonest person; a rogue, a scoundrel. (OED 2025 s.v. *rascal*)

Figure 4.22 illustrates the semantic representation of *rascal* in a frame. The structure is similar to the structure of *devil* in Section 4.2.2.1. The semantics are specified as a *person* which participates or bears an ACTIVITY and a HABITUALITY.

$$\begin{bmatrix} \textit{lexeme} \\ \text{PHON } /\text{ˈræskəl}/ \\ \text{CAT N} \\ \text{SEM } \boxed{0} \begin{bmatrix} \textit{person} \\ \text{ACTIVITY} \wedge \text{HABITUALITY } \boxed{1} \begin{bmatrix} \langle \textit{trickery, mischievousness, villainousness, etc.} \rangle \\ \text{PROF-PROP } \boxed{2} \begin{bmatrix} \text{M-DIM} \begin{bmatrix} \textit{prop-scale} \\ \text{DEGREE } \boxed{3} \begin{bmatrix} \alpha \\ \copyright_{s,Rel} (\boxed{3},\boxed{4}) > \end{bmatrix} \\ \text{THRESHOLD-DEGREE } \boxed{4}\ \beta \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{REF} = \boxed{0} \end{bmatrix}$$

FIGURE 4.22: Representation of the lexeme *rascal*.

Note that attitudinal nouns such as *rascal* neither denote activities nor properties as such, but are best analyzed as ACTORS of certain kinds of activities, or as entities that bear a property. Hence, the REFERENCE-attribute (REF) indicates that the element indexed with $\boxed{0}$, i.e. the *person*, is denoted. In other words, a *rascal* either takes part in an ACTIVITY or bears specific HABITUALITIES.

Similar to the semantic decomposition of *devil*, *rascal* as attitudinal noun profiles single or highly restricted sets of properties. A *rascal* is an entity with relatively high degrees of trickery, mischievousness, villainousness, etc. as displayed in the character or behavior of a *rascal*. In Figure 4.22, these degrees of behavior are represented in PROF-PROP, indexed with ($\boxed{2}$) as an attribute, where the set ⟨*trickery, mischievousness, villainousness, etc.*⟩ serves as representatives for a whole property bundle for *rascal*.

Turning to the derivative *rascalment*, the semantic frame in Figure 4.23 illustrates its semantic representation. The frame of the morphological base (M-BASE) *rascal* from Figure 4.22 is included as it provides the required eventive structures. The REFERENCE-attribute indicates the reference shift from the base $\boxed{0}$ to the derivative that denotes either the whole ACTIVITY ∧ HABITUALITY ($\boxed{1}$) or the PROFILED-PROPERTIES ($\boxed{2}$).

lexeme
PHON /ˈræskəlmənt/
CAT N
M-BASE [
  lexeme
  PHON /ˈræskəl/
  CAT N
  SEM [0] [
    person
    ACTIVITY ∧ HABITUALITY [1] [
      ⟨*trickery, mischievousness, villainousness, etc.*⟩
      PROF-PROP [2] [M-DIM [
        prop-scale
        DEGREE [3] [α ; ©$_{s,Rel}$ ([3],[4]) >]
        THRESHOLD-DEGREE [4] β]]]]
  REF = [0]]
REF = {[1],[2]}

FIGURE 4.23: Representation of the lexeme *rascalment*.

4.2.3 Summary of readings

It was shown that derivatives with the suffix *-ment* can refer to different nodes in the semantic structure of the nominal base. Furthermore, two distinct categories of base nouns are identified: psych nouns and attitudinal nouns. For psych nouns (Section 4.2.1), the nominalization with the suffix *-ment* can refer to either the whole semantic structure denoted by the base or to the RESULT-STATE inherent in the semantic representation of the base noun. As *-ment* is highly polysemous (cf. Kawaletz 2023), more readings than the ones discussed are highly likely to be found. Second, attitudinal nouns (Section 4.2.2) as bases are attested on two distinct readings as well: One reading refers to the whole semantic structure of attitudinal base nouns, i.e., the activity and the person node together, whereas it is also possible for a *-ment*-formation to refer to the PROFILED PROPERTIES.

Again, such divergent semantic classes of a word-formation process can be captured in inheritance hierarchies (Riehemann 1998; Koenig 1999; Booij 2010; Bonami & Crysmann 2016; Plag et al. 2018). Figures 4.24 and

4.25 show such inheritance hierarchies for denominal nominalizations derived by the suffix *-ment*. The inheritance hierarchies are split by the category of the base noun, i.e., psych nouns as bases are represented in Figure 4.24 and attitudinal nouns as bases in Figure 4.25. The abbreviation *n-n-lfr* stands for noun-to-noun lexeme formation rule and indicates that the categories under this node describe denominal nominalizations. The hierarchy severs the phonological component (PHON) of the word-formation process from different semantic (SEM) categories. First, the semantics are divided into the different base nouns, i.e., psych nouns and attitudinal nouns. Next, the two distinct readings for each category of base noun are given.

Starting with the description of the semantic node for psych nouns, the nominalization with *-ment* can, first, create a transpositional reading and refer to the whole semantic structure denoted by the base noun. Second, the RESULT-STATE of a *change-of-psych-state* can be referred to by the nominalization. In the semantic nodes in the inheritance hierarchy for attitudinal nouns, the derivative with *-ment* can refer to the ACTIVITY and the PROFILED-PROPERTIES node denoted by the semantics of the base.

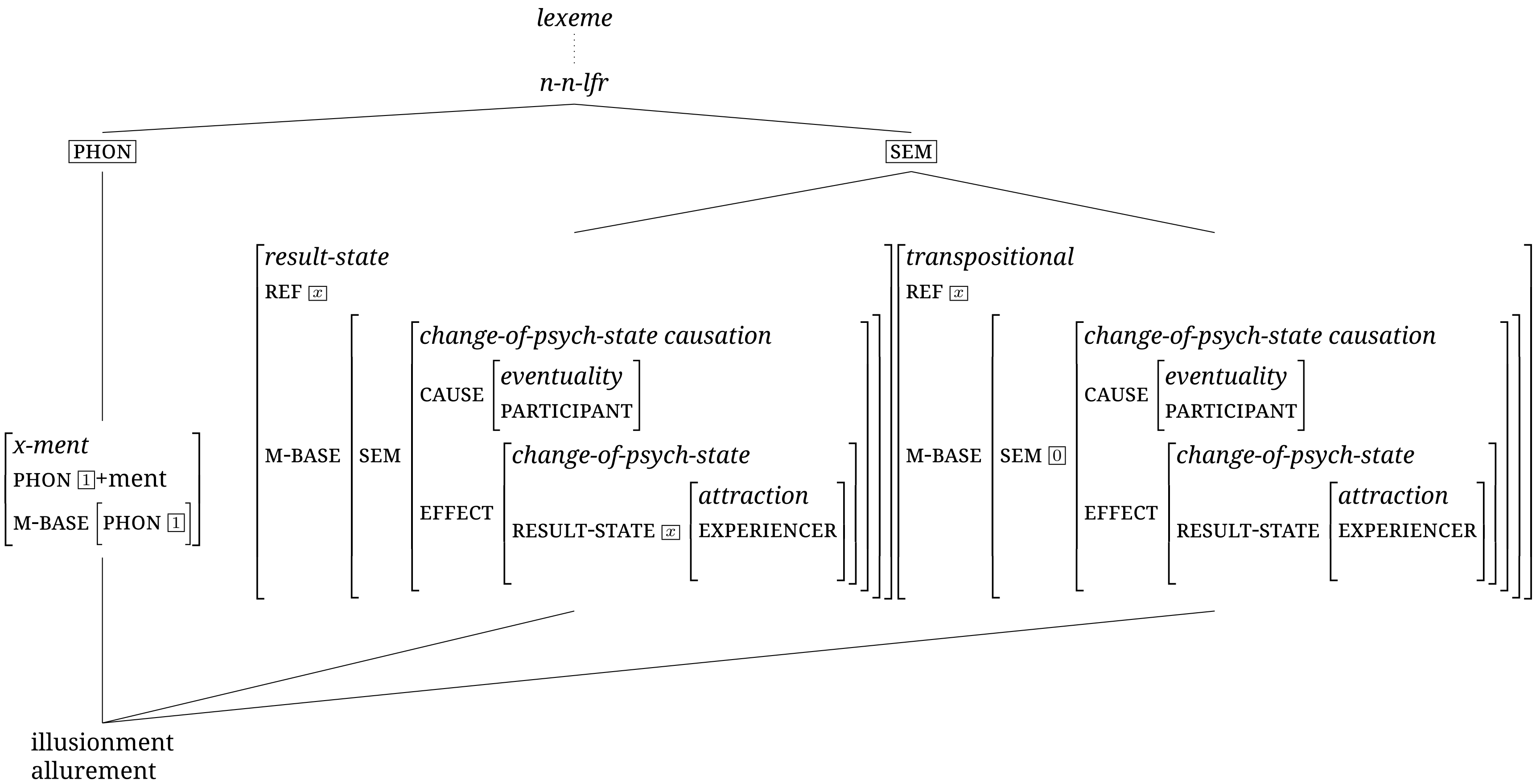

FIGURE 4.24: Inheritance hierarchy of lexical rules for psych nouns in *-ment*.

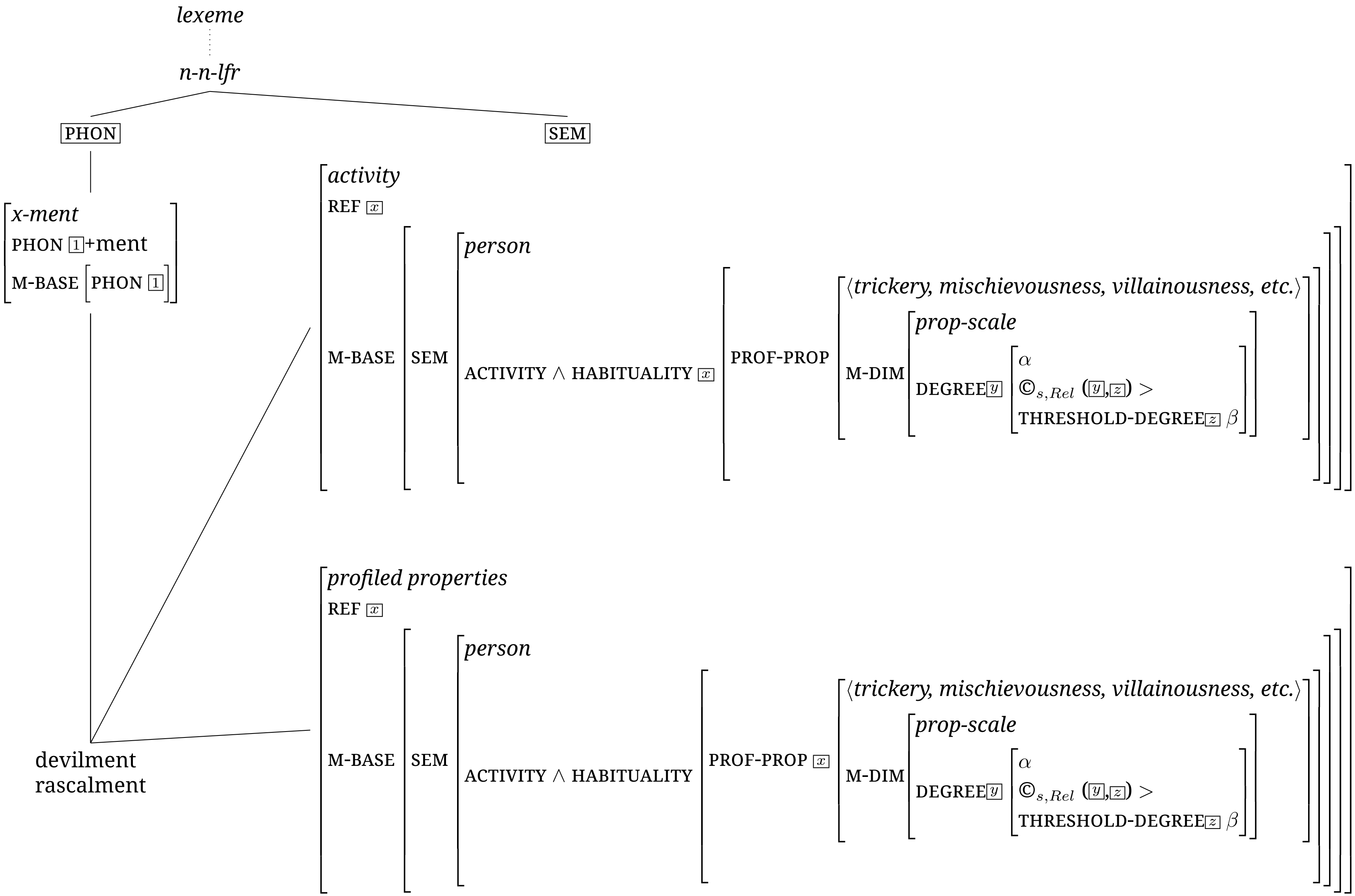

FIGURE 4.25: Inheritance hierarchy of lexical rules for attitudinal nouns in *-ment*.

4.3 Nominalizations with the suffix *-ation*

Nominalizations with the suffix *-ation* usually denote results of processes (e.g., *sedimentation*), locations (e.g., *exhibition*) or means (e.g., *allocation*) (Plag 1999: ch. 7; Bauer et al. 2013: ch. 11; Plag 2018: ch. 4). Additionally, an ornative reading in the sense 'provide with X' is to be found (e.g., *epoxidation*) (Plag 2018: 93; Plag 1999: 207). Note that the suffix is not split into *-ate*, a verbalizing suffix, and *-ion*, a nominalizing suffix, in this thesis. It is seen as one suffix that directly attaches to the nominal base. Arguments in favor for this treatment are that many nominalization in *-ation* do not have a verbal base ending in *-ate* (e.g., *declare* ♦ *declaration*) and backformation from an *-ation* derivative into a verb with *-ate* is productive (cf. Plag 1999, 2018).

I will exemplify the word-formation process with two chemical substances as bases in Section 4.3.1, *ozonation* and *hydroxylation*, and two technical terms as bases, *pixelation* and *brecciation* in Section 4.3.2.

4.3.1 Chemical substances as bases

Chemical substances are productive with the suffix *-ation* (cf. Plag 2018: 91). It is possible for the word-formation process to refer to eventive elements in the base to create the meaning of the eventuality related nominalization. The process is analogous to the previously analyzed data in *-ee* (Section 4.1) and in *-ment* (Section 4.2).

The nominalization process with chemical substances as bases is exemplified by the analysis of *ozone* ♦ *ozonation*, and *hydroxyl* ♦ *hydroxylation*. The examples in (22) give the definition for the two bases by the OED (2025).

(22) a. An allotropic form of oxygen, O3, which is a bluish toxic gas with a characteristic sharp odour, is a powerful oxidizing agent, and is produced from molecular oxygen by electrical discharge and in the upper atmosphere by ultraviolet light. Also colloquial: fresh, invigorating air. (OED 2025 s.v. *ozone*)

b. The radical HO or OH, consisting of an atom of hydrogen in combination with an atom of oxygen, which is a constituent of a vast number of chemical compounds. (OED 2025 s.v. *hydroxyl*)

These definitions illustrate which elements have to be included in the semantic representation of chemical substances. Chemical substance have molecules ('O3', 'HO or OH') and a purpose ('is a powerful oxidizing agent', 'constituent of a vast number of chemical compounds'). Both attributes are included in the description of their semantics in the following generalized representation of chemical substances as well as in the pertinent sub chapters (*ozonation* 4.3.1.1, *hydroxylation* 4.3.1.2).

Figure 4.26 illustrates the semantic representation of a chemical substance in a frame notation. Similarly to the already described representations, the *substance* is of the type *lexeme* with a phonological representation (PHON), the syntactic category noun (CAT N) and its semantics (SEM).

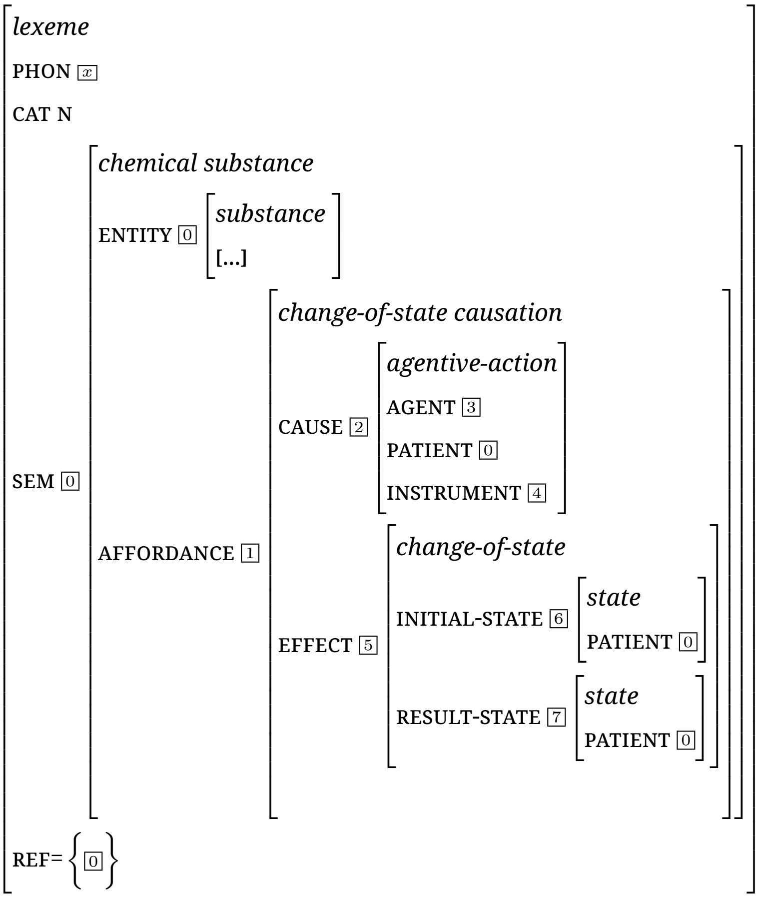

FIGURE 4.26: Representation of a prototypical chemical substance. The frame representation is based on *change-of-state verbs* in Plag et al. (2018) and Kawaletz (2023: 58).

The first attribute in the semantic representation is an ENTITY (0). The representation has several abbreviated attributes. This abbreviation is represented with three dots ([...]) to indicate that a more fine-grained representation of the semantics is possibly included. For instance, the substance has a certain aggregation state or can be described as having particular molecules. However, the further specification of these features of a chemical substance are of no importance to the semantic analysis proposed in this thesis. Thus, these features are abbreviated in the semantic representation of the base.

The semantic representation of the prototypical *chemical substance* has an AFFORDANCE-attribute (1). This attribute represents the purpose of the substance. That is, the semantic structure of chemical substances includes the information that the substance can be used for something.

This idea is based on the TELIC-QUALE in Pustejovsky (1996: ch. 6). Similar to bases like *biography* in Section 4.1.1.2, this usage of the substance is an AFFORDANCE ([1]). Notably, this node contains the eventive structures on which the denotation of the later nominalization-process is dependent on. More precisely, the potential usage of a chemical substance bears an eventive structure: a *change-of-state causation*, which, in turn, enables the reference shift from the denotation of the base to the newly created reading of the derivative. This line of thought underlines previous findings which"suggest that action nominals in *-ation* merely denote an Event having to do with the entity denoted by the base" (Plag 1999: 209).

The included *change-of-state causation* follows the structure of change-of-state verbs in Kawaletz (2023: ch. 4). That is, the substance is involved in a *change-of-state* when it is used for a chemical process, its AFFORDANCE. This *causation* has a CAUSE ([2]) and an EFFECT ([5]). The CAUSE is usually an *agentive-action* with an AGENT ([3]), a PATIENT ([0]), and an INSTRUMENT ([4]). The co-indexation of the substance and the PATIENT with the index number [0] indicates that the PATIENT of the *agentive-action* is the *chemical substance*.

The EFFECT ([5]) is a *change-of-state*. The INITIAL STATE ([6]) as well as the RESULT-STATE ([7]) indicate the change of the substance (PATIENT ([0])) from one state to another. Thus, the substance is undergoing a change from one *state* to another *state*. Several possible states are conceivable for the substance. For instance, a change in *physical-form* (Kawaletz 2023: p. 62f.), like *ozone* as an 'allotropic form of oxygen, O3', or a *come-into-being* eventuality (Kawaletz 2023: p. 87), like *hydroxyl* in a different 'chemical compound', are possible and plausible states.

Another option for the *change-of-state causation* is given in Figure 4.27. The possibility of the word-formation process with chemical substances as bases gives rise to an ornative reading of the derivative. Some theories of semantic interpretation on ornative readings model ornative readings similarly to locative readings (e.g., Jackendoff 1990, 1991; Plag 1999). That is, the ornative reading can be semantically described as a change in location of the chemical substance, for example, a change to a chemical compound. Furthermore, the distinction of locative and ornative readings is often ambiguous (cf. Plag 1999: 127). The proposed solution for this issue is to combine locative and ornative readings into one semantic representation. The lexical conceptual structure (LCS) in 4.1 by Plag (1999: 129) unifies locative, ornative and causative readings:

LCS of locative/ornative/causative -*ize* verbs

$$([\ [\quad]_{\text{BASE}} - ize]_V)$$
$$\text{NP}_i \ \_\ \text{NP}_{\text{Theme}}$$
$$\text{CAUSE}\Big([\quad]_i, [\text{GO}([_{\text{Thing, Property}}]_{\text{Theme/Base}}; \quad [\text{TO}[_{\text{Thing, Property}}]_{\text{Base/Theme}}])]\Big) \tag{4.1}$$

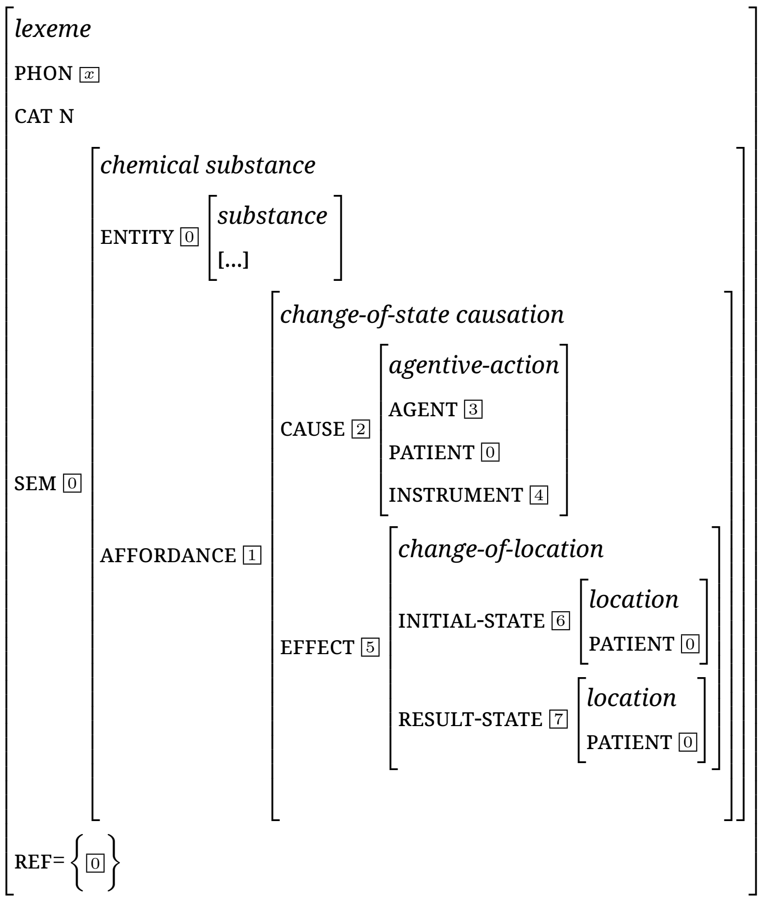

FIGURE 4.27: Representation of a prototypical chemical substance. The frame representation is based on *change-of-state verbs* in Plag et al. (2018) and Kawaletz (2023: 58).

Locative, ornative and causative readings are all characterized by a GO TO function. Thus, not only locative and ornative readings are represented by the same semantic structure, but so are causative readings. Before turning to this unification, the frame semantic representation in Figure 4.27 displays the formalization of possible ornative readings laid out by the structure of a chemical substance: a *change-of-location* in the EFFECT ($\boxed{5}$) of the *change-of-state causation*.

Turning back to the unification of locative, ornative and causative readings, the GO TO function describes all three readings. For instance, the verb *nuclearize* has a locative/ornative and causative reading, expressed by a different order of arguments in the LCS (for a more detailed discussion of examples, see Plag (1999: 130ff.)). This unification of readings results in the frame semantic representation in Figure 4.28. The EFFECT $\boxed{5}$ includes a *change-of-state* and a *change-of-location*. Both sub-eventualities are separated with the logical operator OR ($\vee$). This separation indicates that either a *change-of-state* or a *change-of-location* is happening, using either a *state* or a *location* in the INITIAL-STATE and the RESULT-STATE. Importantly, the *change-of-state* is represented with the *states* and the *change-of-location* is expressed by the *locations*.

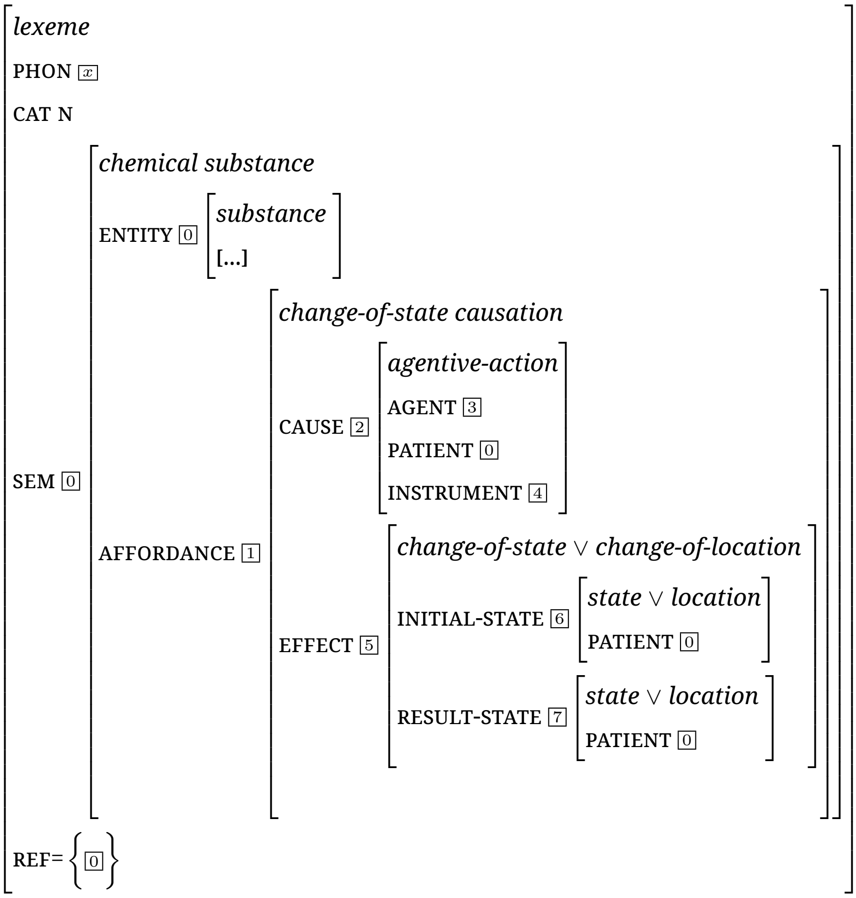

FIGURE 4.28: Representation of a prototypical chemical substance. The frame representation is based on *change-of-state verbs* in Plag et al. (2018) and Kawaletz (2023: 58).

The REFERENCE-attribute (REF) indicates that the denotation of the chemical substance is the element indexed with (0). This is the ENTITY in the representation.

Turning to the reference shifts induced by the word-formation process with the suffix *-ation* with chemical substances as base, the examples in (23) give the definition for the two nominalizations exemplified: *ozonation* and *hydroxylation*.

(23) a. Impregnation or treatment with ozone; reaction with ozone, esp. in an ozonolysis process. Also: conversion of oxygen into ozone. (OED 2025 s.v. *ozonation*)

b. The introduction of a hydroxyl group into a molecule or compound. (OED 2025 s.v. *hydroxylation*)

The definitions illustrate that three readings of a nominalization based on chemical substances are possible: process (‘an ozonolysis process’), resultative (‘introduction of a hydroxyl group into a molecule or compound’), and ornative readings (‘impregnation with ozone’). Thus, three reference shifts are possible.

Figure 4.29 illustrates the reference shifts. The process reading shifts the reference to the affordance ([1]), the *change-of-state causation*. The resultative reading shifts the reference from the base to the RESULT-STATE indexed with [7]. The ornative reading shifts the reference to the RESULT-STATE indexed with [7], similar to the resultative reading. The difference to the resultative reading is that the ornative reading refers to a *location*, replacing the *state*.

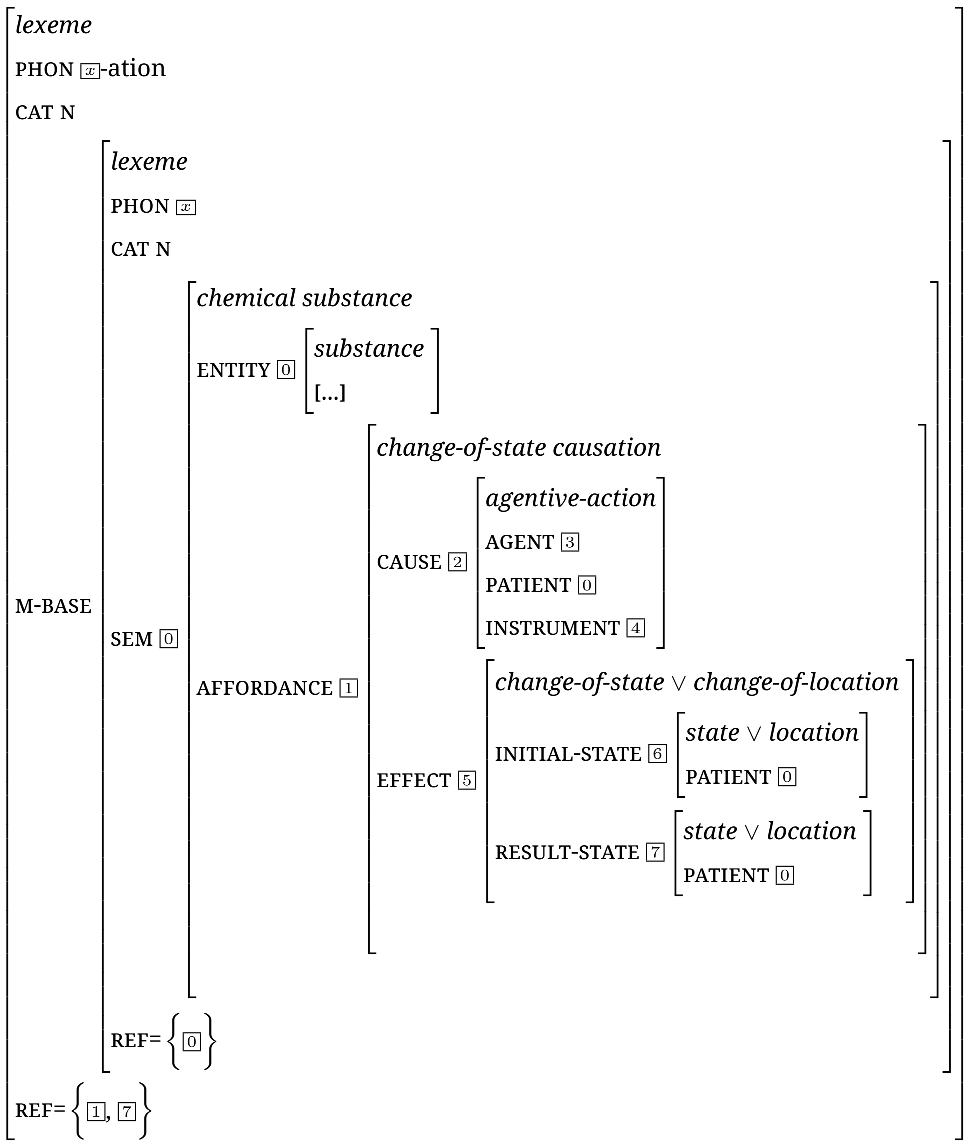

FIGURE 4.29: Representation of a denominal derivative with a chemical substance as base. The frame representation is based on *change-of-state verbs* in Plag et al. (2018) and Kawaletz (2023: 58).

The nominalization process based on chemical substances is described by examples in the following. First, the process from *ozone* to *ozonation*, and second, the process from *hydroxyl* to *hydroxylation* is analyzed.

4.3.1.1 *ozonation*

The first -*ation*-derivative based on a denominal chemical expression is *ozonation*. The base for the nominalization is the noun *ozone*. The examples in (24) illustrate that ozone is used as a chemical substance with molecules (‘*ozone* in the air’) and a purpose (‘has a deleterious effect on stratospheric *ozone*’).

(24) a. In the Netherlands, a government health institute warned of high levels of smog due to **ozone** in the air in parts of the country. (COCA, see Davies 2008-, NEWS, 2019)

b. Nitrous oxide produced in this way may escape up through the ground surface and into the atmosphere, where it has a deleterious effect on stratospheric **ozone** and moreover poses a serious threat as a potent agent for global warming. (COCA, see Davies 2008-, ACAD, 2018)

The frame in Figure 4.30 represents the semantic structure of the base noun *ozone* and has the same attributes as the generalized frame in Figure 4.28. It can be applied to chemical bases, here to *ozone*. The frame shows the *lexeme* with its phonological representation (PHON) and indicates the syntactic category of the base which is a noun (CAT N). SEM describes the semantics of the base. The lexeme is further defined as an ENTITY (0), which is a *substance*. The notation of the characteristics of this substance are abbreviated ([...]). The substance is further defined as *ozone*.

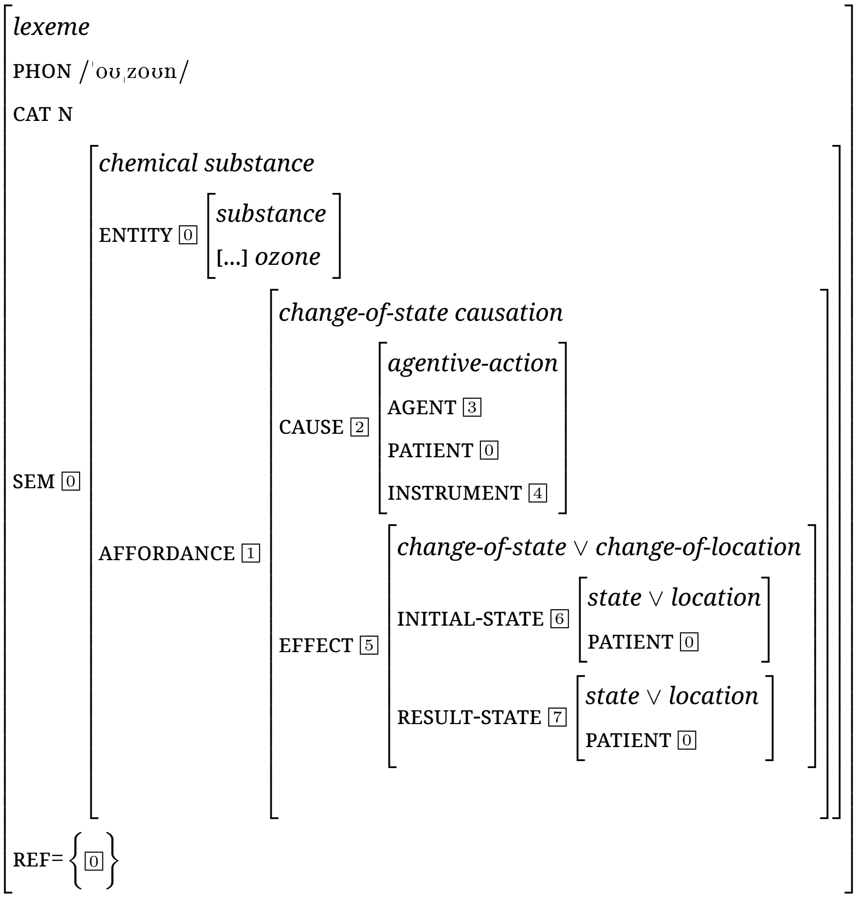

FIGURE 4.30: Representation of the lexeme *ozone*.

A second attribute of the semantic structure of the base is labeled AFFORDANCE. In the case of chemical substances like *ozone*, the *affordance* is the usage of the substance for chemical processes. This usage gives rise to a structure of an eventuality in the nominal base. The AFFORDANCE is further specified as a *change-of-state causation*. The *change-of-state causation* has two attributes: CAUSE ([2]) and EFFECT ([5]). The CAUSE is specified as an *agentive-action* in which an AGENT, ([3]) a PATIENT ([0]) and an INSTRUMENT ([4]) are involved. The co-indexation of the PATIENT with the index [0] indicates that the base word itself is involved in the *agentive-action*. Note that also an inchoative reading is plausible. In such a reading, the CAUSE would neither contain an AGENT nor an INSTRUMENT.

The EFFECT ([5]) is a *change-of-state* ∨ *change-of-location*. This string allows for a resultative as well as an ornative (locative) reading. Both options include an INITIAL-STATE ([6]) and a RESULT-STATE ([7]). The *states* as well as the *locations* belong together, distinguishing resultative readings from ornative ones.

The REFERENCE-attribute (REF) indicates that the denotation of the base *ozone* is on the element indexed with $\boxed{0}$, the *chemical substance*, which is at the same time the PATIENT of the *agentive-action* and the *change-of-state* ∨ *change-of-location* eventualities.

The examples in (25) show the different readings of the nominalization *ozonation*. The reference shifts from a chemical substance to a chemical process in example (25-a). Example (25-b) illustrates a result-state reading and the example in (25-c) an ornative reading.

(25) a. Evaluation of membrane filtration and **ozonation** processes for treatment of reactive dye wastewater. (COCA, see Davies 2008-, ACAD, 2015)

b. **Ozonation** has also been found to improve the taste, odor and color of water. (COCA, see Davies 2008-, ACAD, 1998)

c. Dissolved iron, manganese, and hydrogen sulfide must be oxidized into solid particles by chlorination or **ozonation** before they can be removed by a media filter. (COCA, see Davies 2008-, MAG, 1996)

Ozonation is in Figure 4.31 of the type *lexeme*. The phonological representation (PHON) is given as well as the information that the derivative is a noun (CAT N). The semantic representation of the base noun *ozone* (Figure 4.30) is provided as morphological base (M-BASE) of the derivative. More precisely, the possibility to refer to eventive elements provided by the base is crucial for a successful word-formation process. The derivational process shifts the reference from the base to the readings of the nominalization. In other words, the different readings of the derivative *ozonation* are based on and dependent on the semantic structure of the base *ozone*. The REFERENCE-attribute (REF) illustrates the reference-shift from the node indexed $\boxed{0}$ to the nodes indexed with $\boxed{1}$, $\boxed{7}$.

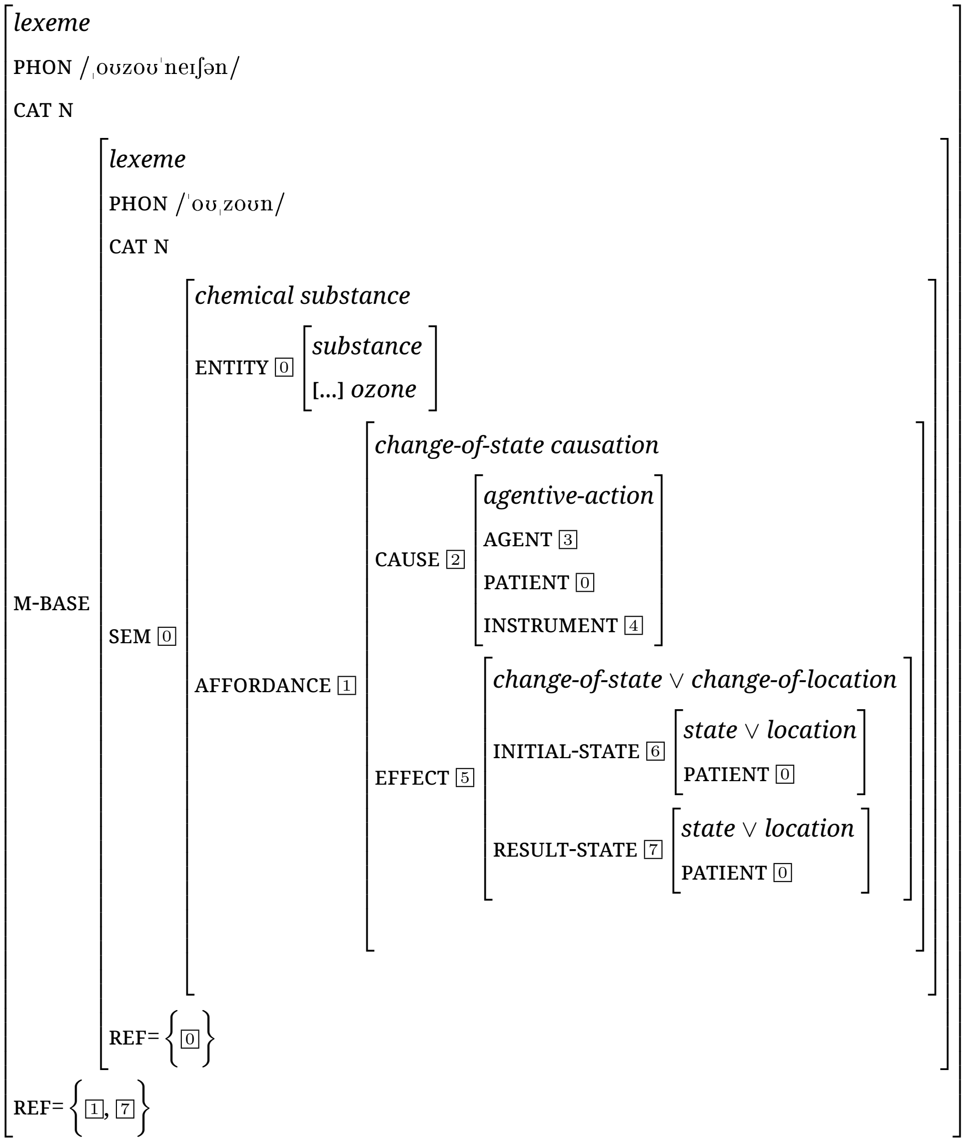

FIGURE 4.31: Representation of the lexeme *ozonation*.

One of the readings of the derivative is the process reading as in example (25-a): It induces a shift of the reference to the element indexed with $\boxed{1}$, AFFORDANCE. By referring to the AFFORDANCE attribute, the whole *change-of-state causation* is referred to by the derivative.

Another reading of *ozonation*, as in example (25-b), is a resultative reading. The reference shifts to the element indexed with $\boxed{7}$, the RESULT-STATE. For the resultative reading, the RESULT-STATE is a *state* in which the base, the chemical substance, is after the successful nominalization process.

A further possible reading is an ornative reading as illustrated in example (25-c). Again, the reference shifts to the element indexed with [7], the RESULT-STATE. The difference to the resultative reading is that in the ornative case the location changes and not the state (cf. Plag 1999). In other words, the base is moved from the *location* in the INITIAL-STATE ([6]) to the *location* in the RESULT-STATE ([7]). The *states* for the resultative reading as well as the *locations* for the ornative readings are dependent on each other. That is, the *states* belong together and the *locations* belong together, a combination of *state* and *location* or vice versa is not planned.

4.3.1.2 *hydroxylation*

The second *-ation*-derivative under closer investigation is *hydroxylation*. The base for the nominalization is the noun *hydroxyl*. Examples in (26) illustrate that the base is a chemical substance. As such, it has specific molecules ('(OH) group') and a purpose ('distribution of **hydroxyl** in garnets').

(26) a. The upshot of this is that, if you look back at figure 1, you'll have to replace every **hydroxyl** (OH) group that you see [...] (COCA, see Davies 2008-, FIC, 2019)

b. The distribution of **hydroxyl** in garnets from the subcontinental mantle of southern Africa. (COCA, see Davies 2008-, ACAD, 2018)

Figure 4.32 illustrates the frame semantic analysis of the lexeme *hydroxyl*. The semantic representation is taken from the generalized frame 4.26 and is alike to the *ozone* in 4.30. Only the type of substance in the ENTITY-attribute ([0]) changes from *ozone* to *hydroxyl* in Figure 4.32.

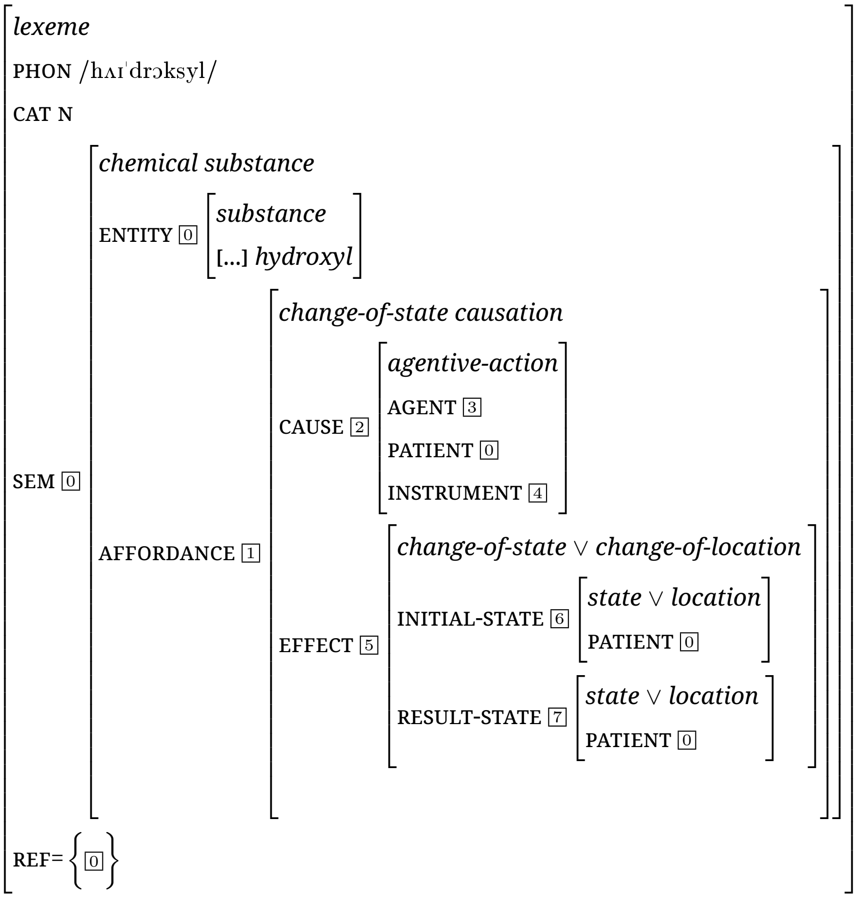

FIGURE 4.32: Representation of the lexeme *hydroxyl*.

The examples in (27) show the different readings of the nominalization *hydroxylation*. The reference shifts either from the chemical substance to the process as in example (27-a), to a resultative reading as in (27-b) or an ornative reading as in (27-c).

(27) a. P450s play an important role in various **hydroxylation** and oxidation processes including secondary metabolism as well as the breakdown of toxins and other xenobiotic compounds 35. (COCA, see Davies 2008-, WEB, 2012)

b. Their functions are potentially affected by their individual structures, their **hydroxylation** patterns as well as their functional groups glycosylated and/or alkylated. (COCA, see Davies 2008-, ADAC 2014)

c. [...] and this enzyme is responsible for the **hydroxylation** of 25(OH)D to form 1,25(OH)2D.
(iWeb, see Davies 2018-, jcancer.org, 2016)

Figure 4.33 indicates the reference shift from the base *hydroxyl* to the derivative *hydroxylation*. The semantics of M-BASE are the same and provided by the base *hydroxyl* in Figure 4.32.

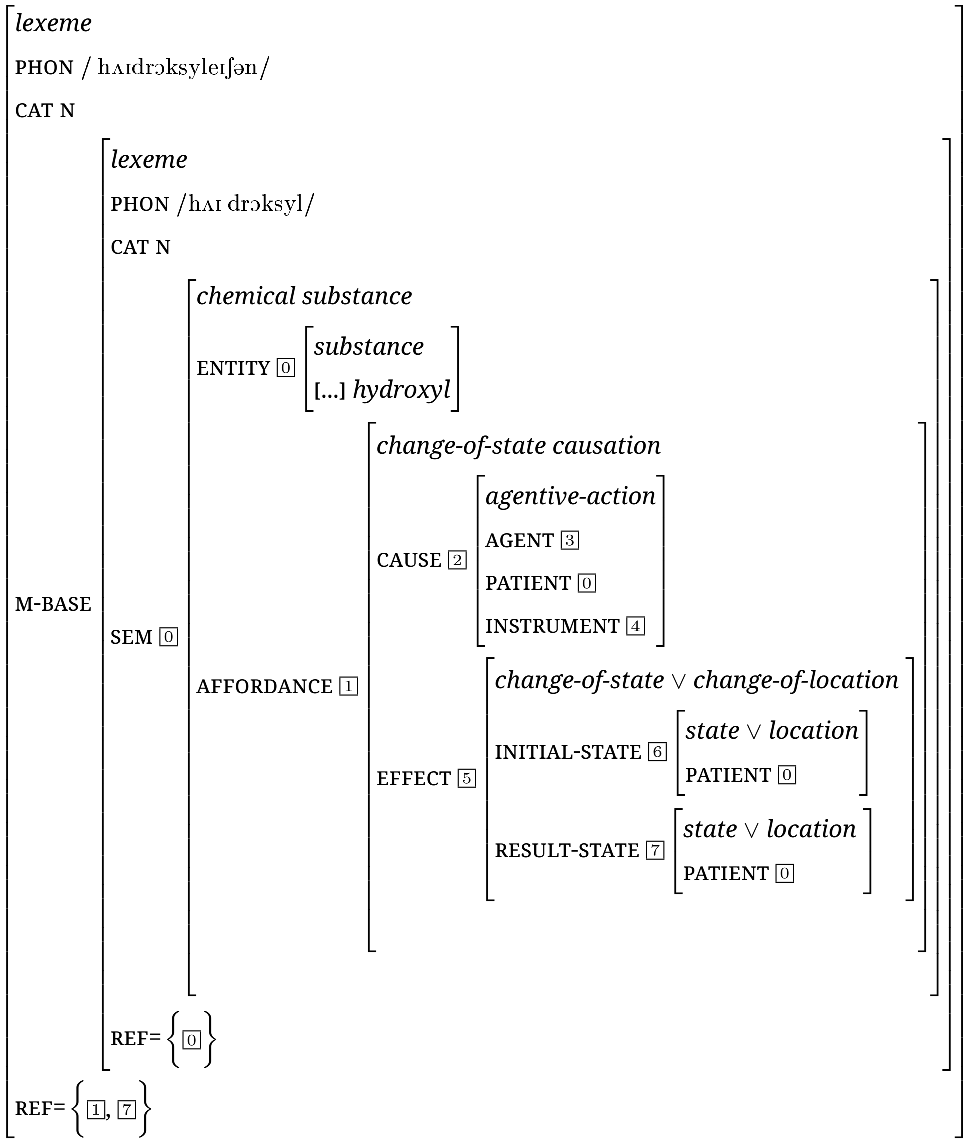

FIGURE 4.33: Representation of the lexeme *hydroxylation*.

The reference shift from the denotation of the base, [0], to the elements indexed with [1], [7] is indicated in the REFERENCE-attribute (REF). For a process reading, the reference shifts to the AFFORDANCE eventuality, i.e., a *change-of-state causation*. The resultative reading refers to the *state* in the RESULT-STATE, indexed with [7]. The ornative reading shifts to the *location* in the RESULT-STATE ([7]). Notice that the distinguishing element of a resultative and an ornative reading is the difference of the reference in the RESULT-STATE, one referring to a different *state*, the other referring to a different *location*.

4.3.2 Technical terms as bases

The nominalizations with the suffix *-ation* can have technical terms as bases. Technical terms are defined as follows:

> Nom d'action technique: Nom peu transparent pour un public non initié, dénotant une action précise complexe, dont la réalisation et la connaissance nécessitent un savoir acquis et qui est spécifique à un domaine particulier. Les noms d'action techniques appartiennent aux domaines de l'industrie, de l'agriculture et de l'artisanat. (Wauquier 2020: 220)

> Translation: Technical action nouns: A largely nontransparent noun for an uninitiated audience, denoting a complex, precise action, the knowledge and performance of which requires an acquired knowledge specific to a certain domain. Technical action nouns belong to the fields of industry, agriculture and craftsmanship. (Wauquier 2020: 220) (translated via deepL)[21]

The nominalization process with technical terms as bases is exemplified by the analysis of *pixel* ♦ *pixelation*, and *breccia* ♦ *brecciation*. The examples in (28) give the definition for the two bases by the OED (2025).

(28) a. Each of the minute areas of uniform illumination of which the image on a television, computer screen, etc., is composed; (also) each of the minute individual elements in a digital image. (OED 2025 s.v. *pixel*)

[21] https://www.deepl.com/de/translator.

b. A composite rock consisting of angular fragments of stone, etc., cemented together by some matrix, such as lime: sometimes opposed to conglomerate, in which the fragments are rounded and waterworn. (OED 2025 s.v. *breccia*)

These definitions illustrate which elements have to be included in the semantic representation of technical terms. One element is the denotation of the term ('uniform illumination [...] is composed', 'a composite rock') and another the purpose ('individual elements in a digital image', 'cemented together by some matrix'). Both attributes are included in the description of their semantics in the following generalized representation of technical terms as well as in the pertinent sub chapters (*pixelation* 4.3.2.1, *brecciation* 4.3.2.2).

Figure 4.34 illustrates the semantic structure of a potential technical base. This structure is similar to the representation for chemical substances in Section 4.3.1. The type of the technical term is a *lexeme* with a phonological description (PHON), the syntactic category of a noun (CAT N) and a semantic representation (SEM). The semantic structure is further specified containing an ENTITY and an AFFORDANCE-attribute. Differently to the modeling of chemical substances, the ENTITY remains an *entity* ($\boxed{0}$), as the types to occur (e.g., *pixel*, *breccia*) allow for more flexibility than the frame for chemical substances that are all substances (e.g., Kawaletz 2023: 151f.). The attributes of the *entity* are abbreviated ([...]). For instance, a pixel could be specified with an attribute that states whether it is lit or not.

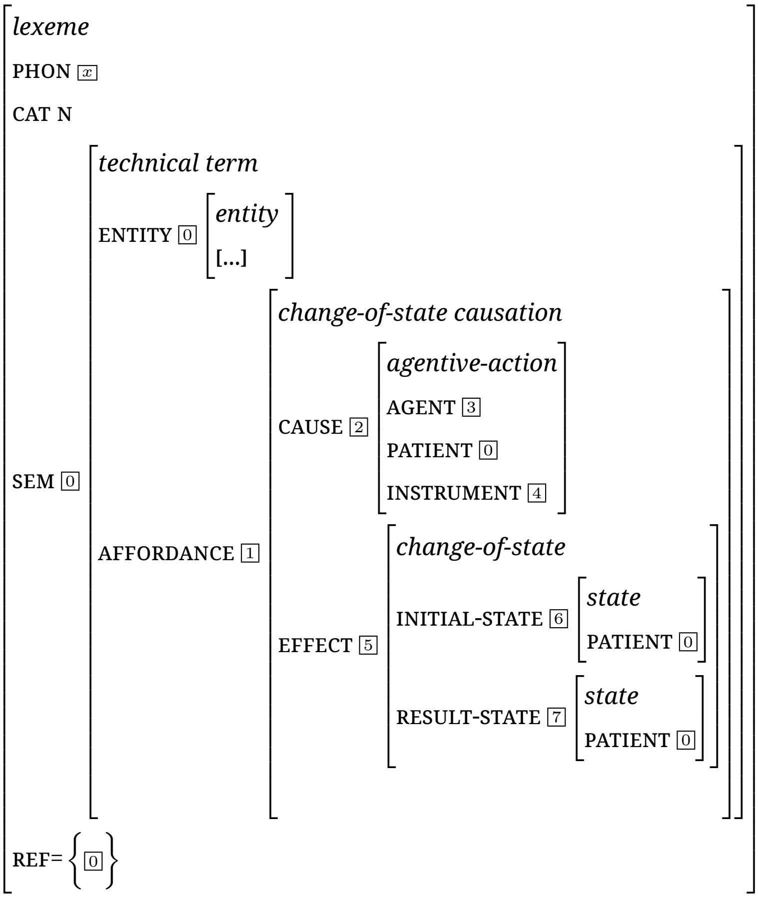

FIGURE 4.34: Representation of a generalized nominal base of a technical term. The frame representation is based on *change-of-state verbs* in Plag et al. (2018) and Kawaletz (2023: 58).

The eventuality in the AFFORDANCE-attribute is a *change-of-state causation*. This interpretation is ajar to change-of-state verbs (e.g., Kawaletz 2023: ch. 4), and conceptualized equally to the *chemical substance* bases in Section 4.3.1.

The examples in (29) illustrate the definitions of the nominalizations *pixelation* and *brecciation* which are based on technical terms.

(29) a. The capturing or reproduction of an image in the form of pixels, resulting in a grainy picture; spec. deliberate blurring of parts of a televisual or photographic image for purposes of censorship or to maintain the anonymity of the subject. (OED 2025 s.v. *pixelation*)

b. The fragmentation of a rock; also, a brecciated condition. (OED 2025 s.v. *brecciation*)

The reference shift can create two different readings: a process reading ('deliberate blurring of parts') and a resultative reading ('the fragmentation of a rock').

Figure 4.35 represents a generalized frame for a potential nominalization with the suffix *-ation*. The morphological base (m-base) is the same lexeme as described in Figure 4.34. Thus, the base provides all required elements for the word-formation process. The crucial change in the readings from the m-base to the nominalizations is the reference shift which is indicated by the reference-attribute. The reference shifts from the element indexed with $\boxed{0}$ to the elements indexed with $\boxed{1}$ or $\boxed{7}$ creating different readings. The shift to the affordance ($\boxed{1}$) denotes a process reading as it refers to the *change-of-state causation* eventuality. The shift to the result-state ($\boxed{7}$) changes the reference from the base to a resultative reading of the derivative.

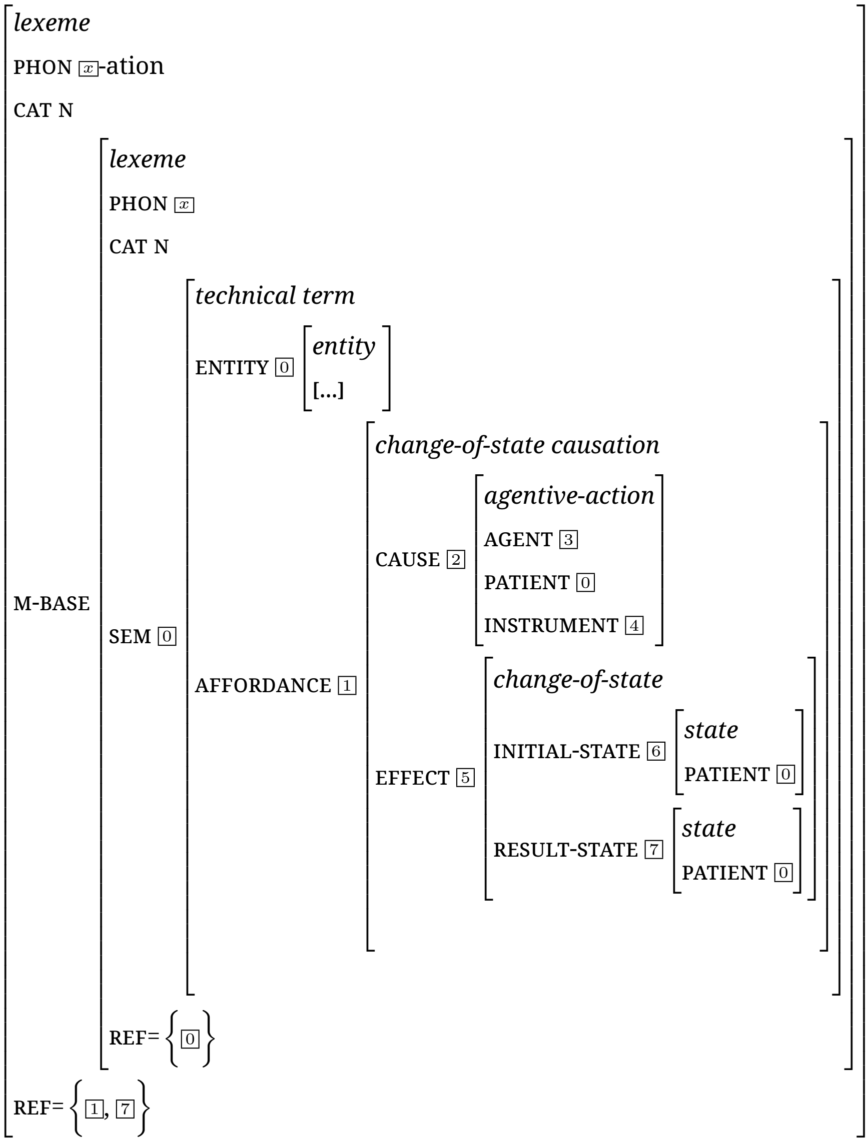

FIGURE 4.35: Representation of a generalized denominal derivative of a technical term. The frame representation is based on *change-of-state verbs* in Plag et al. (2018) and Kawaletz (2023: 58).

4.3.2.1 *pixelation*

The first technical nominalization with the suffix *-ation* is *pixelation*. The base for the nominalization is the noun *pixel*. As illustrated in example (30-a). The noun describes a technical term for small dots which can also be metaphorically used as in example (30-b).

(30) a. Combining standard data of human color perception with the energy it takes to create a red, blue, or green **pixel**, the team, led by Johnson Chuang at Simon Fraser, found colors that required the same amount of energy to produce while being easiest to distinguish from each other. (COCA, see Davies 2008-, ACAD, 2009)

b. Beside the vanquished star, a small dot emerged: a single lit **pixel**. (COCA, see Davies 2008-, FIC, 2019)

The frame in Figure 4.36 represents the semantic structure of the base noun *pixel*. It is based on the generalized frame for technical terms in Figure 4.34. The only difference is that the frame for *pixel* specifies the ENTITY (0) with its specifications ([...]) to be a *pixel*.

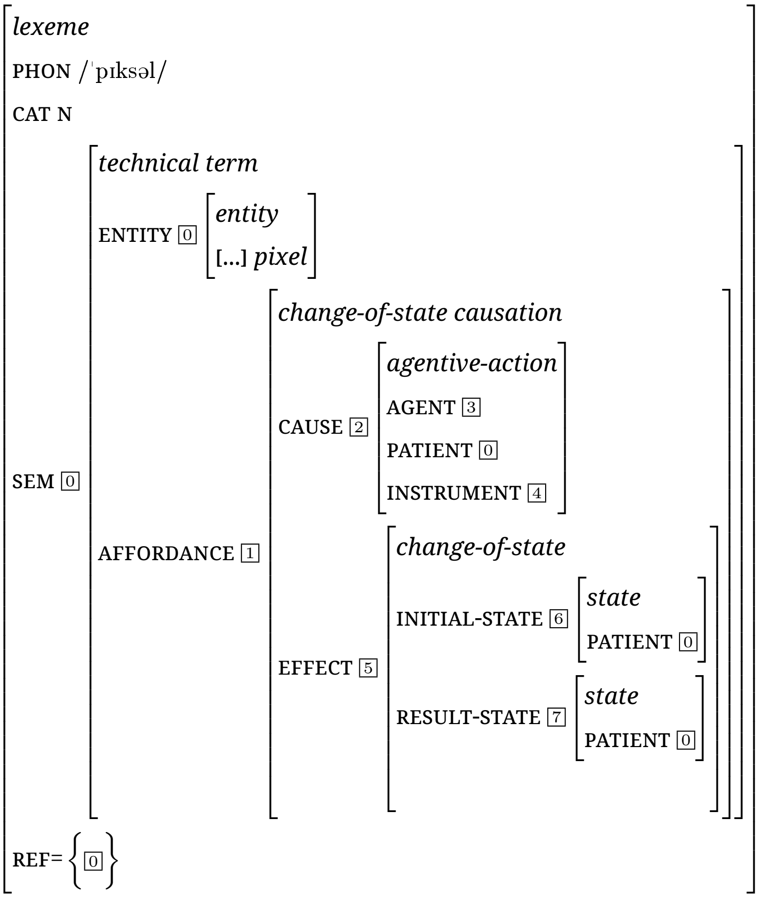

FIGURE 4.36: Representation of the lexeme *pixel*.

The REFERENCE-attribute (REF) illustrates that the denotation of the base is on the elements indexed with 0 in the semantic representation. The co-indexation of the referred ENTITY is a *pixel*, which is involved in the CAUSE (2) as well as in both states in the EFFECT (5).

The examples in (31) show the different readings of the nominalization *pixelation*. The reference shifts from the denotation of the noun itself to the process in example (31-a). In example (31-b) the reference of the nominalization is on the RESULT-STATE.

(31) a. This process of scanning and **pixelation** has the effect of smearing images and edges by averaging the gray scale and hue across each pixel. (COCA, see Davies 2008-, ACAD, 2000)

b. The result is a stereoscopic 3D display that has almost zero crosstalk and no noticeable **pixelation**, providing a smooth, strain-free experience. (COCA, see Davies 2008-, NEWS, 2014)

Figure 4.37 illustrates the derivative *pixelation*. The representation of *pixel* serves as morphological base (M-BASE) for the reference shifts. The reference shift is indicated in the REFERENCE-attribute (REF). The denotation of the derivative can be on the element indexed with [1], or [7]. The shifted reference to the AFFORDANCE corresponds to a process reading as in example (31-a). The other possible reference shift is a shift to the RESULT-STATE as in example (31-b).

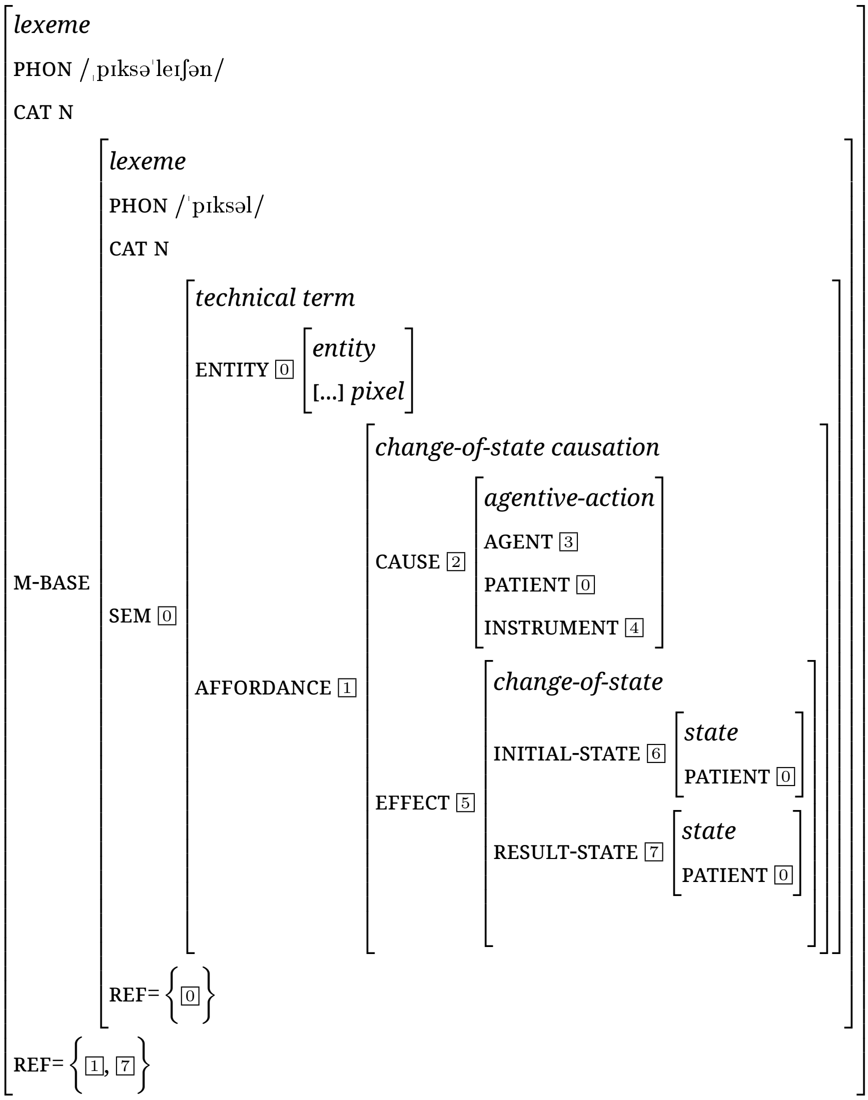

FIGURE 4.37: Representation of the lexeme *pixelation*.

4.3.2.2 *brecciation*

The second nominalization with the suffix *-ation* is *brecciation*. The base for the nominalization is the noun *breccia*, which describes a type of a composite rock. It is a technical term as it describes a special way of how stones build such a breccia. The examples in (32) illustrate the denotation of the base noun. Both examples refer to the entity, i.e., the composite rock.

(32) a. A chunk of **breccia** can be an entire archive of bits and pieces, compressed to a single stone. (COCA, see Davies 2008-, FIC, 2019)

b. After the waters receded, the slow process of **breccia** formation began again, sealing the sandy lens within Lang Trang I. (COCA, see Davies 2008-, ACAD, 1991)

The frame in Figure 4.38 represents the semantic structure of the base noun *breccia*. The representation for technical terms in Figure 4.34 serves as a blueprint. The specification of the ENTITY (0) as a *breccia* is added.

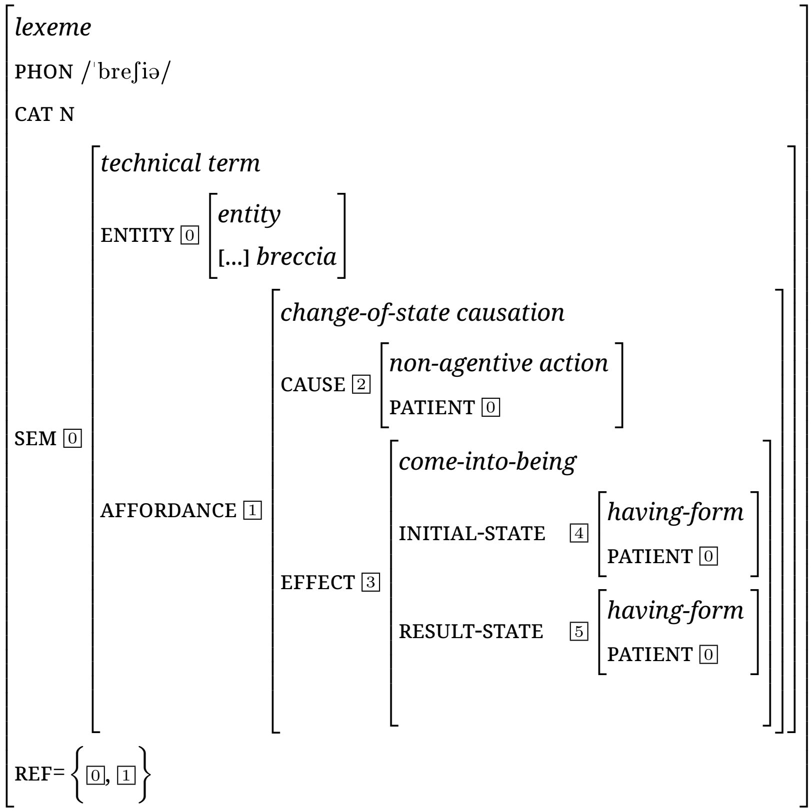

FIGURE 4.38: Representation of the lexeme *breccia* based on *change-of-state verbs* with a *come-into-being* structure (Kawaletz 2023: p. 87).

The AFFORDANCE is a *change-of-state causation*. The CAUSE is in this case a *non-agentive action* with a PATIENT, the rock, (0), as only participant.

The EFFECT also changes to a *come-into-being* eventuality which has two different states: an INITIAL-STATE (4) and a RESULT-STATE (5). Both states are further specified as *having-form*. The PATIENT is involved in both states, as it changes from one form to the other form. The REFERENCE-attribute

(REF) indicates that the denotation of the noun is on the elements indexed with 0.

The examples in (33) show the different reading of the nominalization *brecciation*. The reference shifts to the process inherent in the semantic structure of the base noun in example (33-a). The second reading of the derivative is a resultative reading as in (33-b).

(33) a. Cross-bedding and lamination within intraclasts show that they started as centimeter- to decimeter-thick cemented horizons, whose tops were primary sedimentary interfaces, and whose sides resulted from **brecciation**. (COCA, see Davies 2008-, WEB, 2012)

b. Other parts of the story are the specific chemical composition and impactite **brecciation**: rocks consisting of broken, angular fragments cemented together by a fine-grained matrix. (COCA, see Davies 2008-, MAG, 2006)

Figure 4.39 shows the frame semantic representation of the derivative *brecciation*. The representation of *breccia* in Figure 4.38 serves as morphological base (M-BASE). The base provides the eventive elements required for the word-formation process.

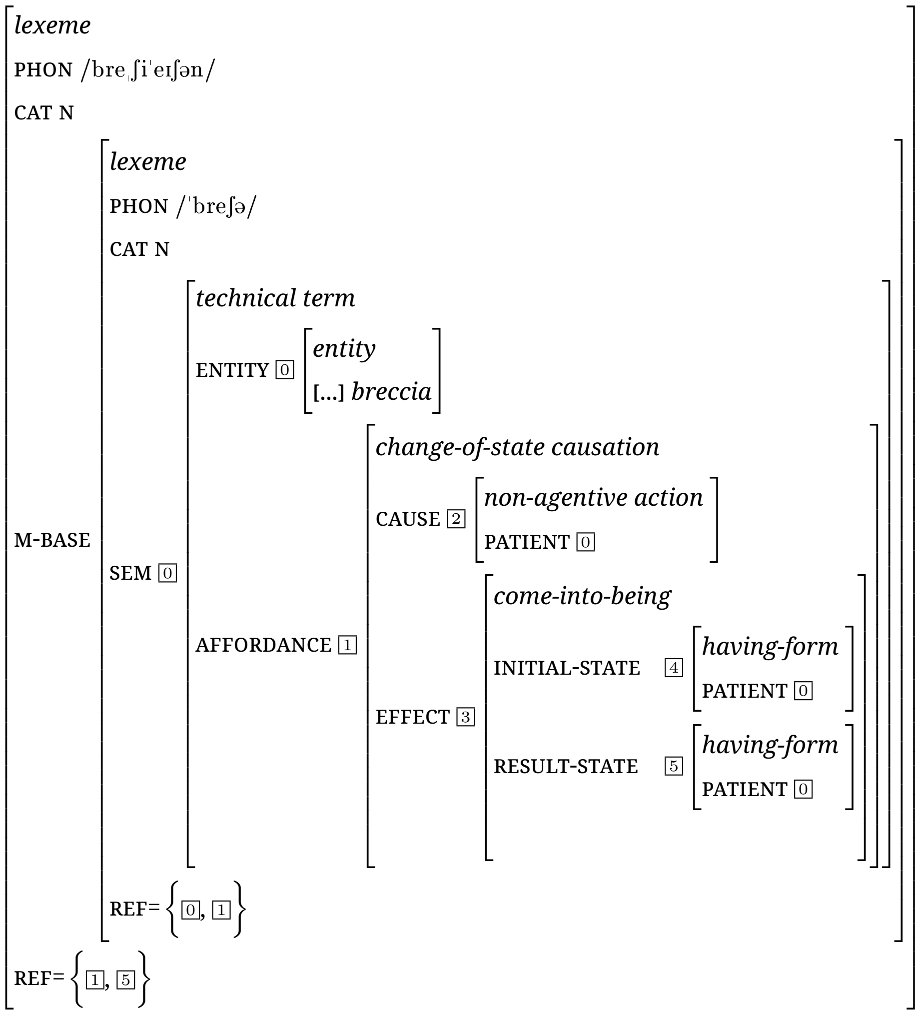

FIGURE 4.39: Representation of the lexeme *brecciation*.

The reference shift from base to derivative is illustrated in the REFERENCE-attribute (REF). The elements indexed with 1 and 5 are referred to by the nominalization. The reference shift to the *change-of-state causation* eventuality in the AFFORDANCE (1) creates a process reading of the derivative as in example (33-a). The other possible reading refers to the RESULT-STATE (5) as in example (33-b).

4.3.3 Summary of readings

The analyses of the examples of *-ation*-nominalizations are based on chemical substances and technical terms. Two derivatives of each category were modeled in a frame semantic framework. The findings for chemical expressions display three different readings: process, resultative and ornative. For the technical terms, process and resultative readings are attested.

The nominalizations with the suffix *-ation*, thus, show different possibilities for a reference shift. Such divergent semantic classes of a word-formation process can be captured in inheritance hierarchies (Riehemann 1998; Koenig 1999; Booij 2010; Bonami & Crysmann 2016; Plag et al. 2018). Figures 4.40 and 4.41 illustrate such inheritance hierarchies for denominal nominalizations derived by the suffix *-ation*. The abbreviation *n-n-lfr* stands for noun-to-noun lexeme formation rule and indicates that the categories under this node describe denominal nominalizations. The hierarchies sever the phonological component (PHON) of the word-formation process from different semantic (SEM) categories. The REFERENCE-attributes (REF) in the semantic (SEM) nodes indicate that the reference of each reading is on the element indexed with $\boxed{x}$.

Starting with the inheritance hierarchy in Figure 4.40, the process readings are depicted. The left semantic node represents the process reading arising with all four nominalizations: The reference shifts to the AFFORDANCE-attribute. Thus, the process reading refers to a *change-of-state causation* eventuality.

Figure 4.41 illustrates the reference shifts to resultative and ornative readings by *-ation*-nominalizations. The resultative reading in the left semantic node shows a reference shift to the RESULT-STATE consisting of a state. All examples can refer to this node in the semantic representation of the base. The ornative reading in the right semantic node also refers to the RESULT-STATE. In this reading, the result is not a state but a location. Only the exemplary chemical substances show ornative readings. Nonetheless, it is imaginable that other *-ation*-nominalizations like the technical terms can also show a reference shift to an ornative RESULT-STATE.

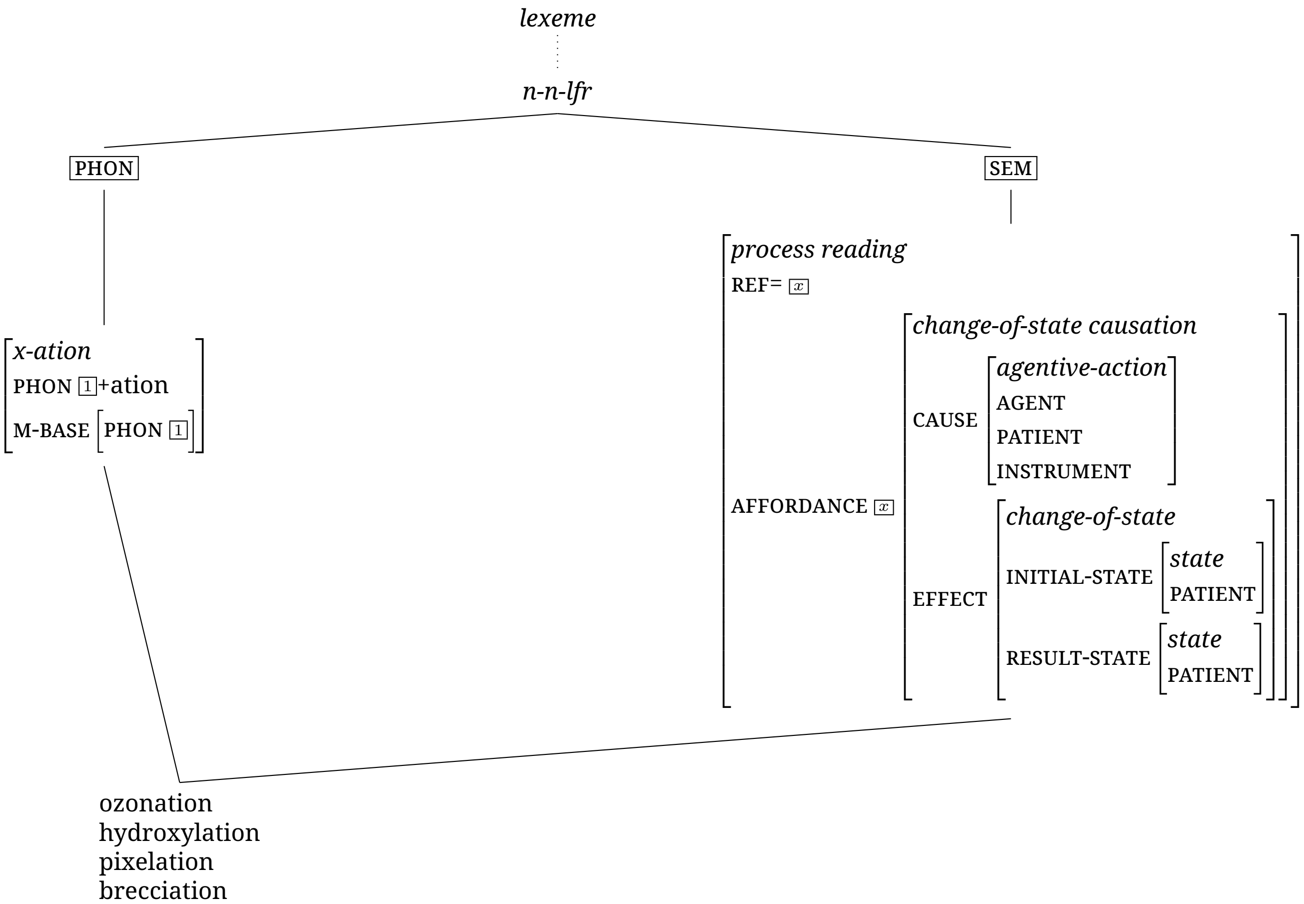

FIGURE 4.40: Inheritance hierarchy of lexical rules for *-ation* for process readings.

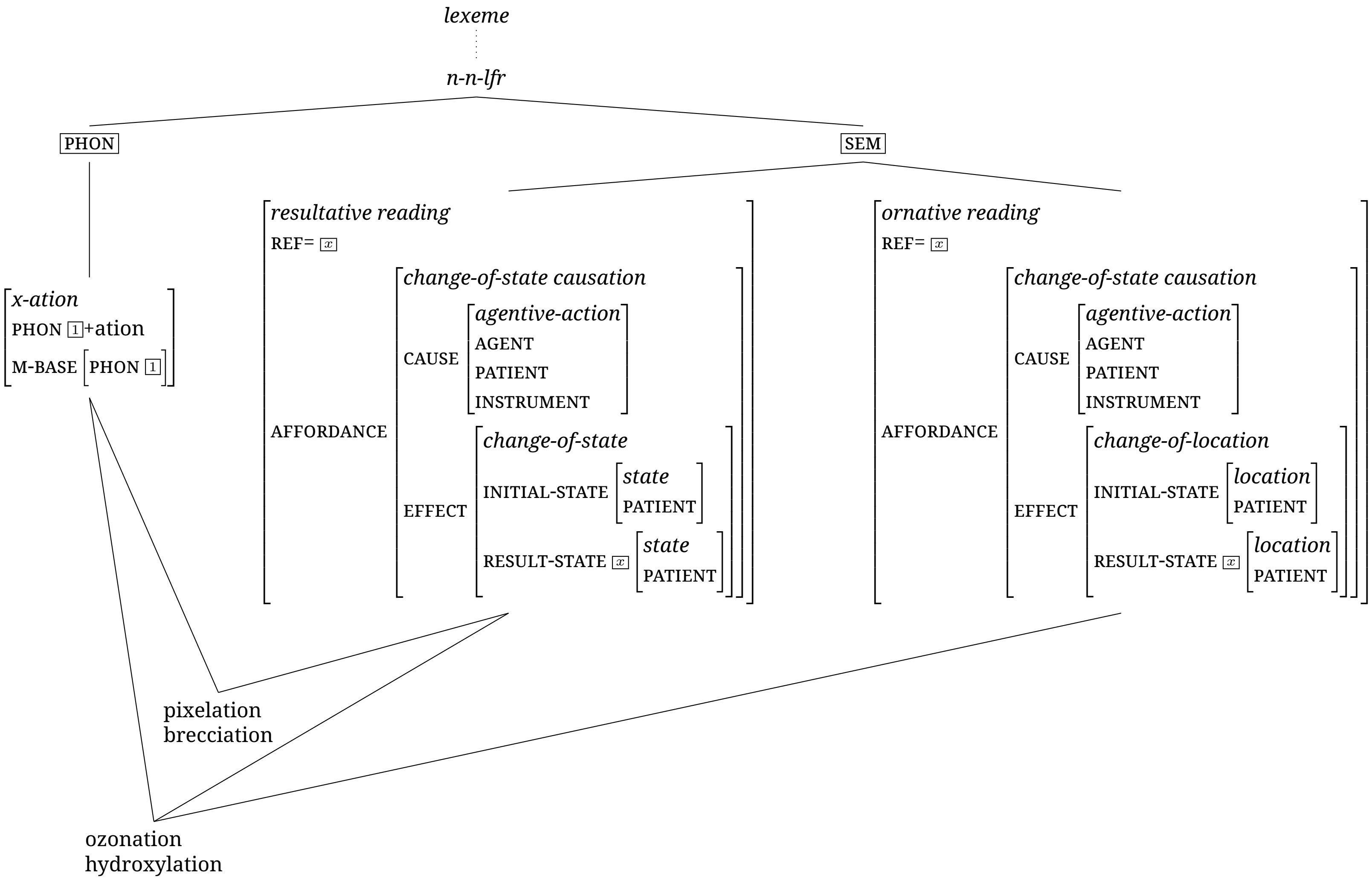

FIGURE 4.41: Inheritance hierarchy of lexical rules for *-ation* for resultative and ornative readings.

Summarizing, the results of the frame semantic analyses proposed for eventuality-related nominalizations with the suffix *-ation* indicate that the nominal base contains the semantic element required for the word-formation process with the suffix. Broadly speaking, the nominalizations are possible because the nominal base contains the required eventive structures and makes them available for the word-formation process. The reference shifts to a process, a resultative, or an ornative reading are dependent on the semantic structure of the base. These findings indicate that the word-formation processes work similar to the deverbal formations (cf. Kawaletz 2023).

Chapter 5

Discussion: Frame semantics

The aim of the semantic analysis using a reference shifting approach was to investigate denominal nominalizations with the suffixes *-ee*, *-ment*, and *-ation*.[22] The research questions for this part of the dissertation asked whether it is possible to find an eventuality in the nominal base to extend the reference shifting approach from verbal bases (cf. Kawaletz 2023) to nominal bases. Furthermore, the analysis showed how the assumed referential shift induced by the word-formation process can be explained and modeled.

A word-formation process can attach to several different word classes (cf. Plag 2004). However, eventuality-related word-formation processes clearly prefer verbal bases. This preference is grounded in the fact that verbs denote eventualities themselves. Nouns, on the other hand, usually denote things or participants in eventualities (see, e.g., Haspelmath 2001; Van Valin & LaPolla 2002; Szabó 2015; Moltmann 2019). Moreover, conversion in English makes it in many cases impossible to decide unambiguously to which word class the base belongs (see Chapter 3). However, for the semantic reference shifting approach in this dissertation, i.e., the reference shifting approach, the word class of the base is not of paramount importance as the analyses in Chapter 4 showed.

[22] An earlier version of parts of Section 5.1 and Section 5.4 have already been published in Schneider (2023). An earlier version of parts of Section 5.2 and Section 5.4 have already been published in Kotowski et al. (2023).

5.1 Discussion *-ee*

The analysis of denominal derivatives with *-ee* demonstrates that nouns, as well as verbs, can provide the eventive elements that are required for the word-formation process. In contrast to verbs, deeper decomposition of the semantics of nouns is necessary to reveal the required eventive elements (cf. Plag 2004). Consequently, the word class of a base does not have a central role for eventuality-related nominalizations. Rather, the semantics provided by the base word are crucial. Assuming that the semantics are central for word-formation processes in general, every word that has the eventive elements for the process in its semantic representation can serve as a potential base, regardless of its word class. The word-formation process can then shift reference to an eventive element in the eventuality provided by the semantic representation of such a base.

The semantic approach applied is one of semantic decomposition and reference shifting and thus similar to the approach by Plag et al. (2018) and Kawaletz (2023) for deverbal nominalizations with the suffix *-ment*. This approach allows to decompose the base into its semantics which consists of eventive elements like (sub-)eventualities and participants. In some cases, the eventive elements that are used for the nominalization process are deeply embedded in the semantic structure. Applying the analysis to denominal derivatives and their base words demonstrates that seemingly non-eventive nouns like *biography* have eventualities embedded in their semantic representation. The REFERENCE-attribute allows for the inclusion of the denotation of base and derivative in one frame. Importantly, the applied approach is a non-coercive approach. Coercion of eventive elements is not needed, as such elements have been shown to already be provided by the respective bases. Thus, the analyses presented in this paper are in line with frameworks that assume embedded eventualities in the semantics of non-eventive nouns (see, e.g., Pustejovsky 1996: ch. 6).

The analyses of denominal nominalizations with the suffix *-ee* illustrate that the constraints posed by Barker (1998) for deverbal derivatives also hold for denominal ones. That is, a denominal nominalization with the suffix *-ee* refers to a participant in the semantic structure of the base. The referred entity has to be sentient and non-volitional. As Barker (1998) already pointed out, the non-volitionality constraint is gradable and of the observed constraints the one that can be violated.

As a consequence of the successful violation of the volitionality constraint for suffixation with *-ee*, for example *bargee* (Section 4.1.3), the inheritance hierarchy in Section 4.1.4 does not only depict the *regular semantics* for derivatives with the suffix *-ee*, but also the possibility of such derivatives to refer to volitional participants. The restrictions that a volitional participant can only be the target for a nominalization with the suffix *-ee* if no non-volitional participant is present in the semantic representation of the base word is often found for deverbal nominalizations (e.g., *escapee*, cf. Barker 1998). The volitionality constraint as such is a gradable constraint, stating rather that the most non-volitional participant is taken for the nominalization with *-ee* . Thus, the violation of the volitionality constraint is not surprising, as this requirement can be violated if only a volitional participant is available in the semantic structure of the base. This violation is also observable for denominal nominalizations with the suffix *-ee*. Only a single occurrence of such a volitionality violation is found in the data analyzed here, *bargee*. For bases lacking a non-volitional participant in their semantic representation, affix competition between *-ee* and *-er* might be involved in the process of *-ee* referring to a volitional participant as well (cf. Barker 1998).

5.2 Discussion *-ment*

The second denominal nominalizations under investigation in this dissertation are built with the suffix *-ment*. The analysis of *-ment*-derivatives shows that the reference shifting approach can also be applied for denominal nominalizations. The word-formation process is dependent on the semantic structures provided by the morphological base. The data for denominal *-ment*-formations show two prominent types of bases, i.e., psych nouns (Section 4.2.1) and attitudinal nouns (Section 4.2.2). The different base types presented different vantage points for the modeling of nominalization semantics. The analyses of the psych nouns showed that the nominalizations are straightforwardly eventuality-related. Their meaning is best described as referencing nodes of a causative event given by the psych noun as base. The generalizations over the analyzed attestations show that derivative and base share the same underlying semantic structure (and partly even allow for referencing the same nodes). In contrast, the attitudinal nouns are person nouns that allow for systematically analyzing their denotation as ACTORS of (habitual) activities with lexeme-specific eventuality properties. In consequence, this allowed us to model the activity-reading of nominalizations with *-ment* based on attitudinal nouns as referential shifts to the activity-node or the PROFILED PROPERTY-node provided by the semantic structure of the base.

The analysis of *-ment*-formations are in line with the finding in this thesis that the semantic compatibility of nominalizations with the suffix *-ment* and their nominal bases relies on peculiar semantic structures already inherent in the semantic structure of the base noun.

Irrespective of the base, Plag et al. (2018) and Kawaletz (2023) show that assigning *-ment* a semantic representation is far from trivial. In particular, attempts at coming up with a unitary meaning that would capture the polysemy of derivatives with the suffix *-ment* in a satisfactory way are unsatisfying. The authors suggest a word-based reference shifting approach from base to nominalization as applied in this dissertation. The given analysis of nominal bases showed that the reference-shifting approach used for verbal bases and their derivatives is extendable to nominal bases. Furthermore, the inheritance hierarchy in Section 4.2.3 (based on, e.g., Riehemann 1998; Koenig 1999; Booij 2010; Bonami & Crysmann 2016; Plag et al. 2018) completes the picture of denominal *-ment*-nominalizations. For the analysis at hand, the inheritance hierarchy splits up in the different types of bases, i.e., psych nouns and attitudinal nouns, and then the different possibilities for the reference shift induced by the suffixation with *-ment*. Given the existence of property readings based on attitudinal nouns, however, the generalization that all *-ment*-derivatives either denote eventualities or their participant has to be extended to properties assigned to activities or persons, at least on the assumption that properties and eventualities are distinct ontological categories (see e.g., Metzger et al. 2019; Moltmann 2019).

5.3 Discussion *-ation*

The third category of eventuality-related nominalizations under investigation are those with the suffix *-ation*. The general make-up of the applied reference shifting approach is similar to the analyses made for *-ee* and *-ment*. An eventive structure is required in the base in order to be able to form derivatives with *-ation*. The generalized frames for chemical expressions in Section 4.3.1, the generalized frame for technical expressions in Section 4.3.2, and the individual analyses of the example words under investigation show that the required eventive structures for the word-formation process are inherent in the decomposed semantic structure of the base. In other words, the nominal bases for *-ation*-nominalizations contain eventive structures revealed by an AFFORDANCE-attribute which are available for the derivational process.

Nominalizations with the suffix *-ation* are said to create either events or results of processes, but locations or means can also be found (Plag 1999: ch. 7; Bauer et al. 2013: ch. 10; Plag 2018: ch. 4). Furthermore, derivatives in *-ation* can denote "'provide with X' (ornative), [...] or 'make into X' (resultative)" readings (Plag 2018: 93). Specifically, chemical or other substances as bases are used with *-ation* to denote results of processes, without first verbalizing the noun with *-ate* (Plag 2018). Chapter 4.3 analyzed chemical and technical terms. The observations are in line with previous findings on the suffix, as the chemical as well as the technical based *-ation*-nominalizations denote either processes, i.e., the AFFORDANCE-attribute giving rise to an eventuality, or result-states provided by the semantic structure of the pertinent base. Nominalizations based on expressions can also refer to ornative readings. These readings are formalized as a change in location, similar to previous analyses of ornative cases (e.g., Jackendoff 1990, 1991; Plag 1999).

The different readings as possible outcomes of the word-formation process with the suffix *-ation* are illustrated in several inheritance hierarchies in Chapter 4, Section 4.3.3. The inheritance hierarchy is based on previous analyses of polysemy in word-formation processes (e.g, Riehemann 1998; Koenig 1999; Booij 2010; Bonami & Crysmann 2016; Plag et al. 2018). Nominalizations with *-ation* create either *process*, *resultative* or *ornative* readings. These hierarchies define which elements are required by the nominalization in order to create the reading of the nominalization in a systematic way. Importantly, the possibility of including lexeme formation rules also for deverbal nominalization is given and it is up to further research to complete the whole picture of possible readings. The only requirement for word-formation processes in general is that nominalizations are dependent on the existence of required eventive elements in the base, independent of the word class of the base.

5.4 Nominal bases

The frame semantic analysis applied in this dissertation is the reference shifting approach by, for example, Plag et al. (2018) and Kawaletz (2023). This approach can be transferred successfully from the analysis of verbs to the analysis of nouns as bases for nominalizations. Many different types of nominal bases can be used for the nominalization process with the suffixes *-ee*, *-ment* and *-ation*: For example, information objects like a *biography*, eventive nouns like *debt*, psych nouns like *illusion*, attitudinal nouns like *devil*, as well as chemical substances like *ozone* and technical terms like *pixel*.

Given the eventive nature of the derivatives in question, a potential drawback of the reference shifting approach is the possible post-hoc assignment of eventive structures to the base in order to make reference shifting work. This holds in particular for non-eventuality-denoting bases such as attitudinal person nouns. Importantly, the eventualities in none of the base nouns are induced or coerced by the word-formation process itself. The category of the base noun, be it concrete/informative objects or paradigmatically related loan words, psych nouns or attitudinal nouns, or chemical and technical terms, does not play a role. The underlying assumption is that there are reasons independent from the word-formation processes to assume eventuality-structures as inherent to the base semantics, including in particular linguistic environments that select for eventualities or indicate agentivity.

The decomposition of bases via frames is a highly fruitful approach in laying bare such structures. The findings presented in this dissertation are in line with analyses that take eventualities to be inherent in certain non-eventive nouns that feature as input to eventive structures, be they perceived as dynamic meaning construals such as metonymical shifts (see, e.g., Baeskow 2021 for a recent proposal) or as core features of lexical entries (as, e.g., in the Qualia structure in Pustejovsky 1996; see also the general remarks in Bauer et al. 2013: 233).

More generally, the findings in the analyses of word-formation processes with different suffixes support views that caution against analyzing the word class of the base as primary feature for the word-formation process regarding its potential to serve as base of a word-formation process (see Barker 1998; Plag 2004). The results of the analyses are in line with previous research (e.g., Plag 1999, 2004) as they reveal that the semantic compatibility of eventuality-related nominalizations with their nominal bases relies on peculiar semantic structures of the base. The fact that all the suffixes under investigation clearly prefer verbal over nominal bases can be explained by, first, the semantic categories of such nominalizations as eventuality-related, second, their compatibility with verbs as the one syntactic category whose members prototypically denote eventualities, and third, their compatibility with fewer nominal bases due to the lack of inherent event-semantic components in the case of many nouns (see Van Valin & LaPolla 1997; Haspelmath 2001; Van Valin & LaPolla 2002; Szabó 2015; Moltmann 2019 on ontological preferences of word classes).

Taking a closer look at the ontology of word classes, verbs are usually said to denote eventualities and nouns to denote entities (e.g., Van Valin & LaPolla 1997; Haspelmath 2001; Van Valin & LaPolla 2002; Szabó 2015; Moltmann 2019). The nominalizations with the suffixes under investigation in this dissertation, *-ation*, *-ee* and *-ment*, do all denote eventualities. This fact explains why verbal bases are preferred by the suffixes: Verbs denote eventualities which are in turn taken for the nominalization process. In other words, the nominalization process is more straightforward with verbal bases due to the already existing eventuality in the base verb.

However, the analysis of eventuality-related nominalizations with the analyzed suffixes is also possible with non-eventive base nouns. The noun as a base has to be further decomposed in its semantics in order to find the eventuality the nominalization process operates on. Regardless of the need of further semantic decomposition, the nominalization process with nominal bases resembles the nominalization process with verbal bases.

The resemblance of the nominalization processes with verbal and nominal bases leads to the question whether the word class of the base is of any interest for the word formation process at all. The deeper decomposition of nominal bases might explain why verbs are preferred as bases for the nominalization process. Nevertheless, the process is not blocked with nominal bases. To sum up, one may conclude that the ontology of the word class of the base words leads to a preference of verbal bases but does not rule out other bases, like nouns.

Similar problems of mismatches between eventuality-denoting derivatives and their (non-eventive) nominal bases arise with other suffixes like *-age, -ance, -er, -ure* (Plag 2004, 2018: ch. 4). The reference shifting approach proposed and applied in this dissertation can potentially be used for the analysis of other nominalizations as well. Moreover, further research might be able to analyze adjectival bases for the word-formation processes with suffixes that give rise to eventuality-related nominalizations by transferring the successful application of the reference shifting approach for verbal and nominal bases to adjectival ones.

Chapter 6

FastText implementation

The suffixes *-ation*, *-ee*, and *-ment* are investigated in a distributional semantic approach as described in Chapter 2, Section 2.2.2.[23] Computational methods were chosen as a second approach applied in this dissertation to test the findings of the frame analysis in Chapters 4 and 5 which show that the derivative is semantically dependent on eventive elements in the semantic structure of the base. Whether derivatives and bases are indeed semantically similar as suggested by the reference shifting approach and if a difference between nominal and verbal bases is to be found (RQ2) will be investigated.[24] A difference between the semantic similarities of denominal and deverbal derivatives and their bases is expected to be found because of the observation that nominal bases need more decomposition to locate the eventive elements required for the word-formation process. The data for all three suffixes are analyzed individually in Sections 6.2.1.1, 6.2.1.2, and 6.2.1.3.

6.1 Methodology

6.1.1 Word vectors

In order to be able to perform a distributional semantic analysis, word vectors which represent the semantics of a word are needed for each target word, i.e., each derivative and its base. As a starting point, vectors

[23]An earlier version of parts of Sections 6.2.1.1 and 6.2.1.3 are to be published, preprint available Schneider (forthcoming). The version in this dissertation contains crucial changes.

[24]The data as well as the scripts used for the individual analyses are available here: `https://osf.io/kaqsv/`.

of denominal derivatives and their nominal bases and deverbal derivatives and their verbal bases in *-ee*, *-ment*, and *-ation* were required. The list of the denominal derivatives described in Chapter 3 was used. The deverbal derivatives and their bases were random samples of hits of the derivatives in the used corpora (BNC, Davies 2004-; COCA, Davies 2008-; NOW, Davies 2016-; iWeb, Davies 2018-) and manually checked for having a verbal base, as described in Chapter 3. The size of the data sets for each suffix are given in Table 6.1.

Table 6.1: The data of all three suffixes ordered by category, denominal/deverbal.

	denominal	deverbal
-ation	67	72
-ee	46	312
-ment	29	273
total	142	657

The word vectors for the derivatives and their bases in *-ee*, *-ment*, and *-ation* were created using *fastText* (Bojanowski et al. 2016), a python package which includes, but is not limited to, pre-trained vector spaces for several languages (Mikolov et al. 2018). These pre-trained word vectors are of 300 dimensions. For the computation of the required word vectors for the derivatives and bases in this study, which were not in the pre-trained data as most of the derivatives are highly infrequent (cf. Chapter 3), the pre-trained *common crawl subword model* was used. This pre-trained vector space contains 2 million pre-trained word vectors and subword information based on character n-grams consisting of *3*-, *4*-, *5*-, and *6*-grams (cf. Mikolov et al. 2018). For example, the word *biography* consists of nine trigrams *#bi/bio/iog/ogr/gra/rap/aph/phy/hy#*. The hash mark represents the beginning or the end of a word. The word vector for the word *biography* does now not only include the information about its distribution in context, but also the information that the trigrams are the form of the word. This means, for example, that there is a semantic connection via the form to the words *biography* and *biology* as both also include the trigrams *#bi* and *bio* (for more information on the mathematical operations for the word vector creation, see, Mikolov et al. 2013a, 2018).

The fact that the syntactic distribution of nouns and verbs is different (see, e.g., Baroni & Lenci 2010) might influence the distribution of the base words and the derivatives for the deverbal data because all nominalizations are nouns and thus probably not entirely comparable with verbal bases. This reasoning cannot be completely left out of sight. However, the inclusion of the subword information ensures that the base and its derivative shall be related at least in their form as both contain the same *n*-grams of the base.[25]

6.1.2 Statistical data analysis

The statistical analysis is twofold. First, beta regression models are fit to see how predictive the variable WORD CLASS OF BASE is for the cosine similarities of derivatives and bases (Section 6.2.1). The beta regression models contain control variables to control for possible influences on the variable of interest. Second, a *t*-distributed Stochastic Neighbor Embedding (*t*-SNE) and a linear discriminant analysis (LDA) are performed (Section 6.2.2). This analysis is performed with the vectors computed for the derivatives. The aim is to see whether patterns in the data can be found as well as to see whether denominal and deverbal derivatives are distinguishable from a discriminant view.

6.1.2.1 Beta regression

For the analysis of the semantic similarity of the derivatives and bases, cosine similarity was used as measure. A higher cosine similarity of the word vectors expresses a higher semantic similarity of the words' semantics, whereas a lower cosine similarity expresses a lower similarity (cf. Sitikhu et al. 2019; Huyghe & Wauquier 2020). A cosine similarity of 1 would express that both words are semantically the same. The cosine similarities are used to compare denominal derivatives and their nominal bases as well as deverbal derivatives and their verbal bases to each other. The cosine similarities of the word vectors for the derivatives and their bases were computed in Python (van Rossum & Drake 2009).

[25] Sometimes, it is not the case that the whole base is represented in the derivative, see, e.g., the pair *biography* and *biographee*, where not all graphemic information is retained in the nominalization.

Apart from the word class of the base, a number of variables may influence the cosine similarity of derivatives and bases. For example, the relative frequency of derivative and base (e.g., Hay & Baayen 2003) as well as the polysemy of the base word (e.g., Lieber 2004; Melloni 2007; Lieber 2016) may be influential factors on the similarities of derivatives and their base words. I decided to treat the variable WORD CLASS OF BASE as the variable of interest because the analyses investigate whether there is a differences of denominal and deverbal nominalizations regarding their similarity to the base. The variables BASE POLYSEMY and RELATIVE FREQUENCY were included in the statistical analysis as control variables. The inclusion of control variables aims to ensure that a found effect of the variable of interest is not an artifact of a confounding variable. Therefore, potentially confounding variables are introduced as control variables. In turn, the analysis will show whether these variables have an effect on the cosine similarity of derivatives and bases and if their inclusion show a problematic correlation with the variable of interest.

The variable WORD CLASS OF BASE is the central variable of interest in the statistical models, as RQ2 is directly concerned with a comparison of denominal and deverbal nominalizations. The coding of this variable was based on the procedure in Chapter 3. The expectation was that deverbal derivatives and their verbal bases are generally more similar to each other than denominal derivatives and their nominal bases. This expectation arises because verbs denote eventualities (see, e.g., Van Valin & LaPolla 1997; Haspelmath 2001; Szabó 2015; Moltmann 2019) just like eventuality-related nominalizations do. In other words, due to the semantic nature of verbs, a direct link of the eventuality in the verbal base to the nominalization is established. For nouns, on the other hand, the link of the eventive elements of the nominal base to the nominalization is not as straightforward as nouns do usually not denote eventualities. The analyses are thus expected to show an overall higher semantic similarity of deverbal nominalizations and their verbal bases compared to denominal nominalizations and their nominal bases.

A second expectation was built on ontological considerations of the common denotation of verbs and nouns as well as on findings of the readings created by the individual suffix by the reference shifting approach: The cosine similarities of derivatives and bases should look different for the data in *-ee* compared to the data of the other two suffixes, *-ation* and *-ment*. That is, data in *-ee* should show a higher cosine similarity of denominal derivatives and bases compared to the other two suffixes because *-ee* creates participant readings (cf. Barker 1998, and the analysis in Chapter 4) that are usually denoted by nouns. Nonetheless, as the participants needed for the interpretation of a nominalization in *-ee* are embedded in the eventuality denoted by the base, an overall higher similarity of deverbal derivatives and verbal bases is expected. To put it another way, nominalizations in *-ee* were expected to show a higher similarity of denominal derivatives and their nominal bases in opposition to the denominal data with the other two suffixes, but the deverbal derivatives should be more similar to their verbal bases compared to denominal derivatives. The suffixes *-ation* and *-ment*, on the other hand, create eventive readings which denote, for example, processes or states, and should hence show higher cosine similarities for deverbal derivatives and bases compared to *-ee* (cf. Plag 2018: ch. 4, and the analysis in Chapter 4). Furthermore, a lower similarity of denominal derivatives in *-ment* and *-ation* and their bases is expected.

I included the variable RELATIVE FREQUENCY as a control variable in the statistical analysis. A higher frequency of the base goes together with a higher segmentability and semantic transparency of the derivative, which results in a clearer connection between the derivative and its base (cf. Hay & Baayen 2003). Thus, a higher relative frequency was expected to increase the cosine similarity of derivative and base. For example, the derivative *government* is more frequent than its base *govern*. This leads to a lower segmentability and a lower semantic transparency of the derivative because the derivative *government* is lexicalized and hence only segmentable with further reasoning. In other words, a higher frequency of the derivative compared to its base results in a weaker association of the derivative with its base word. A low-frequent derivative like *devilment* shows the opposite picture. The base noun *devil* is more frequent than the derivative *devilment*. The derivative is highly segmentable and semantically transparent because the base *devil* is noticeable. Broadly speaking, a low-frequent derivative results in an interpretation connected to the denotation of the base word and is thus semantically more transparent (cf. Plag et al. 2018). For the study in this chapter, a higher RELATIVE FREQUENCY is expected to increase the cosine similarity of derivatives and base because it means that the derivative is lower in frequency than its base. Contrarily, the cosine similarity is expected to be decreased by a lower RELATIVE FREQUENCY, i.e., the base is less frequent than the derivative. The RELATIVE FREQUENCY was computed by dividing the frequency of one particular base by the frequency of the corresponding derivative. Note that relative frequency was log-transformed following standard procedures to avoid issues of unreliable model estimates (cf. Baayen 2008: 71).

The other control variable I included is BASE POLYSEMY. A higher polysemy of the base word was expected to decrease the cosine similarity of derivative and base based on the following reasoning: If a base has more than one reading, the similarity of the derivative and the base should decrease as the derivative is usually not based on all readings of the base word (see, e.g., Lieber 2004; Melloni 2007; Lieber 2016). For example, the base *biography* only has one synset attested in *WordNet* (cf. Fellbaum 1998; Princeton University 2010) and is thus not polysemous. The noun *devil*, on the other hand, has five synsets attested and thus shows polysemy. For the derivational process this means that a nominalization with the word *biography* only has to consider one specific reading of the base which can be used to create a new word. In the case of *devil*, one (more specific) reading has to be identified, which is then used for the derivational process. As several readings of a polysemous base are often not clearly distinguishable, it might be unclear on which reading the process is based on. Furthermore, due to the fact that more than one reading is available for the process, it is to be expected that the semantic similarity of derivative and base decreases if the base offers more than one reading in the first place.

Moreover, the polysemy of the base should already be included in the computed word vectors. That is, if a base occurs in more than one reading, the representation of the word in its vector reflects all readings in one vector. The polysemy of the base represented in the word vector is expected to lead to a lower cosine similarity of derivative and base because the derivative is not based on all the different readings included in the base's vector.

Table 6.2: Dependent variable, variable of interest, control variables and their expected effects.

dependent variable	expectation
COSINE SIMILARITY	The cosine similarity of derivatives and bases indicates how similar one derivative is to its base.
variable of interest	**expectation**
WORD CLASS OF BASE	Verbal bases are more similar to their derivatives due to their denotation of eventualities.
control variables	
RELATIVE FREQUENCY OF BASE/DERIVATIVE	A higher relative frequency goes together with a higher segmentability and more transparency and thus an increase of cosine similarity.
BASE POLYSEMY	A higher polysemy of the base goes together with a decrease of cosine similarity as the derivative is (often) based on one reading of the base.

The variables of interest in Table 6.2 were used as predictors in beta regression models, performed in R (R Core Team 2025), using the betareg package (Cribari-Neto & Zeileis 2010). The aim was to find out which factors influence the similarity between base and derivative. Beta regression was chosen as the statistical tool of choice as the cosine similarities in this study were in the interval of [0,1]. There was no output with negative cosine similarity values. The dependent variable for the models is the cosine similarity of derivative and base. A beta regression model for each suffix, *-ee*, *-ment*, and *-ation*, was fitted including precision phi components which are interpreted as follows: "a low precision coefficient means that the beta regression model estimates the values of this predictor to be more dispersed around the coefficient's mean than in the case of a predictor with a high precision coeffic[i]ent." (Plag et al. 2017: 202). The precision of a variable was included if the model with the precision reached a lower AIC value and a higher log-likelihood according to standard procedures (cf. Baayen 2008: ch. 6; Winter 2019: ch. 15).

Strong correlations of independent variables might bring collinearity into the model, which might in turn lead to unreliable results (Tomaschek et al. 2018). Hence, the independent variables were first tested for problematic correlation coefficients, i.e., $|rho| \geq 0.5$. For *-ee* (Section 6.2.1.1), no problematic correlations of the independent variables were observed. For the models using the data with the suffixes *-ment* (Section 6.2.1.2) and *-ation* (Section 6.2.1.3), some variables of interest correlated with each other. To check for collinearity issues in the fitted models, their VIF-values (variance inflation factors) were computed (Zuur et al. 2010; Fox & Weisberg 2019). Overall, multiple beta regression models were fitted, including and excluding variables and phi coefficients. More precisely, in order to find the best model for each suffix, I fitted the models with step-wise exclusion of first phi components and then variables of interest to determine the best model fit. I chose this procedure of testing which variables and phi components to retain to ensure that the reported models contain as much information as needed but not more than required. The final models reported in this dissertation have the best fit regarding the log-likelihood and the AIC values as well as no problematic correlation of the variables indicated by the VIF-values (cf. Plag et al. 2017; Zuur et al. 2010; Fox & Weisberg 2019). The detailed formulas for the best fitting beta regression models are reported in the pertinent subsections.

6.1.2.2 t-SNE and LDA analyses

In order to see whether the vectors for the derivatives themselves show patterns, for example, whether the derivatives show distinct clusters for denominal and deverbal nominalizations, t-distributed stochastic neighborhood embedding analyses (t-SNE, for more information on t-SNE analyses, see, e.g., van der Maaten & Hinton 2008; van der Maaten 2014; Krijthe 2015; Shafaei-Bajestan et al. 2022a,b; Schmitz et al. 2023; Schäfer 2025) for the estimated semantic vectors of the derivatives is performed. A t-SNE analysis reduces the dimensions of the vectors, 300 dimension in the *fastText* implementation in this dissertation. For the t-SNE analysis in this dissertation, two dimensions are retained for the analysis (Rtsne package, Krijthe 2015; gdsm package, Schmitz & Schneider 2022). The choice for two dimensions was made due to two reasons: First, two dimensions are the standard parameter in the pertinent literature. Second, the visualization of a two dimensional space is far more straightforward to interpret than the output of a three (or more) dimensional space.

The output of the analyses shows the clustering behavior of all derivatives in the data set. The data is sorted into denominal and deverbal derivatives. A t-SNE analysis for the data of each individual suffix, *-ee*, *-ment*, and *-ation*, is performed. The visible clusters of the analyses are then inspected further.

To ensure whether the visible clusters of the t-SNE analyses are meaningful, a linear discriminant analysis (LDA, for more information on LDA, see, Xanthopoulos et al. 2013; Trendafilov & Gallo 2021; Shafaei-Bajestan et al. 2022a; Schäfer 2025) is performed. The LDA in this dissertation is based on the two retained t-SNE dimensions. LDA predicts whether a derivative in the data set is denominal or deverbal. If clusters are visible in the t-SNE visualization, the prediction of the LDA should resemble these clusters. More precisely, LDA predicts whether there is a difference between denominal and deverbal derivatives. Taking the results of the t-SNE visualization as well as the LDA, one may or may not assume the existence of differences between deverbal and denominal derivatives.

First, the results of the beta regression analyses for the three suffixes will be reported in Section 6.2.1. Section 6.2.2 will show the results from the t-SNE and LDA analyses.

6.2 Results

6.2.1 Beta regression analyses

The raw cosine similarities for all suffixes ($n_{\text{denominal}} = 142$, $n_{\text{deverbal}} = 657$) sorted by the word class of the base is shown in Figure 6.1. The cosine similarities of denominal derivatives and their nominal bases are in the plots in the upper panel, and the cosine similarities of deverbal derivatives and their verbal bases are given in the plots in the lower panel. The suffixes are color-coded.

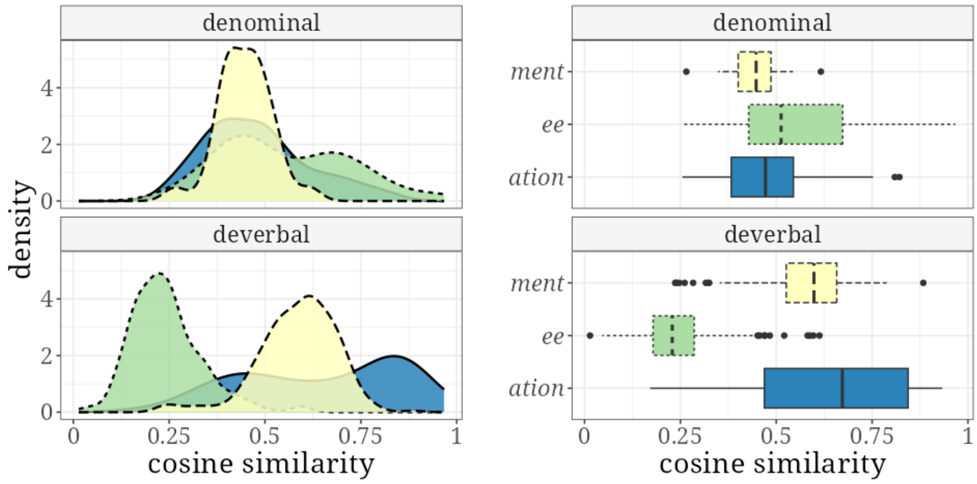

Figure 6.1: Cosine similarities of derivatives and bases for all suffixes. The cosine similarities of the denominal data are given in the upper panels and the cosine similarities for the deverbal data are given in the lower panels. The suffixes are color-coded: *-ation* in blue, *-ee* in green, *-ment* in yellow. Additionally, the suffixes are given distinctive line types to ensure the readability in a gray-scale version. The suffix *-ation* is represented by a solid line, *-ee* by a densely dashed line, and *-ment* by a dashed line.

Two densities look as if they could be bimodal, the denominal density for *-ee* and the deverbal density for *-ation*. Both densities were tested, in Section 6.2.1.1 for the data with the suffix *-ee* and in Section 6.2.1.3 for the data with *-ation*, and the test results of both tests indicate that the distribution of both densities is not bimodal. The data for deverbal *-ation* shows the largest interquartile range, whereas the data for *-ment* shows the lowest interquartile range. For the deverbal data, a clear difference is observed. The deverbal data with the suffix *-ee* show a lower cosine similarity than the data for *-ment* and *-ation* do. The deverbal data for *-ation* show the highest cosine similarities of all.

The denominal sets in the upper panel of the plots do not look strikingly different. A look at the p-values from a Wilcoxon-test is given in Table 6.3. As two comparisons are made per suffix for denominal and deverbal data, i.e., denominal and deverbal *-ee* is compared to *-ation* and *-ment*, *-ation* to *-ee* and *-ment*, and *-ment* to *-ee* and *-ation* , a Bonferroni correction of the comparisons is needed for the denominal and the deverbal data (cf. Baayen 2008: 114; Winter 2019: 176). The p-values of the Wilcoxon-tests are divided by 2 as two comparisons for each suffix are made. This results in a lower significance point. The test for the denominal data shows that while the data for *-ation* and *-ment* are indeed not significantly different ($p > 0.025$), the *-ee* data are significantly different from the data of the other two suffixes under investigation ($p < 0.025$ to *-ation* and to *-ment*).

TABLE 6.3: Comparison of denominal sets by p-values retrieved from Wilcoxon-tests. Significance codes: '***' $p < 0.0005$, '**' $p < 0.005$, '*' $p < 0.025$.

	-ee		*-ment*	
-ation	0.02	*	0.54	
-ee			0.01	*

The deverbal boxes in the plots in the lower panel illustrate that the deverbal data show more variation across the suffixes and they also show more variation compared to the denominal data. The p-values of the Bonferroni corrected Wilcoxon-tests for the deverbal data sets are given in Table 6.4. The deverbal data for the suffixes *-ation* and *-ment* are significantly different from each other ($p < 0.025$). The difference of the *-ee* deverbal data to the data of the other two suffixes is even more pronounced ($p < 0.0005$ for *-ation* and *-ment*).

TABLE 6.4: Comparison of deverbal sets by p-values retrieved from Wilcoxon-tests. Significance codes: '***' $p < 0.0005$, '**' $p < 0.005$, '*' $p < 0.025$.

	-ee		*-ment*	
-ation	< 2.2e-16	***	0.01	*
-ee			< 2.2e-16	***

6.2.1.1 Results *-ee*

The density of the raw cosine similarities for the data in *-ee* ($n_{\text{denominal}} = 46$, $n_{\text{deverbal}} = 312$) sorted by the word class of the base and corresponding box plots is shown in Figure 6.2. The orange area represents the density for the cosine similarities of denominal derivatives in *-ee* and their nominal bases. The purple area depicts the density of the cosine similarities of deverbal derivatives in *-ee* and their verbal bases. The data of the deverbal comparison is found to be in a lower range of the cosine similarity compared to the data of the denominal comparison. Furthermore, the orange area has two peaks, but a dip test for the cosine similarities of the denominal derivatives and their nominal bases indicates that the data are not bimodally distributed (dip test, $p = 0.1452$; diptest package, Maechler 2020). The denominal derivatives and bases are clearly more semantically similar to each other than the deverbal pairs. The difference of the cosine similarities between denominal and deverbal pairs is significant (Wilcoxon test, $p < 0.001$).

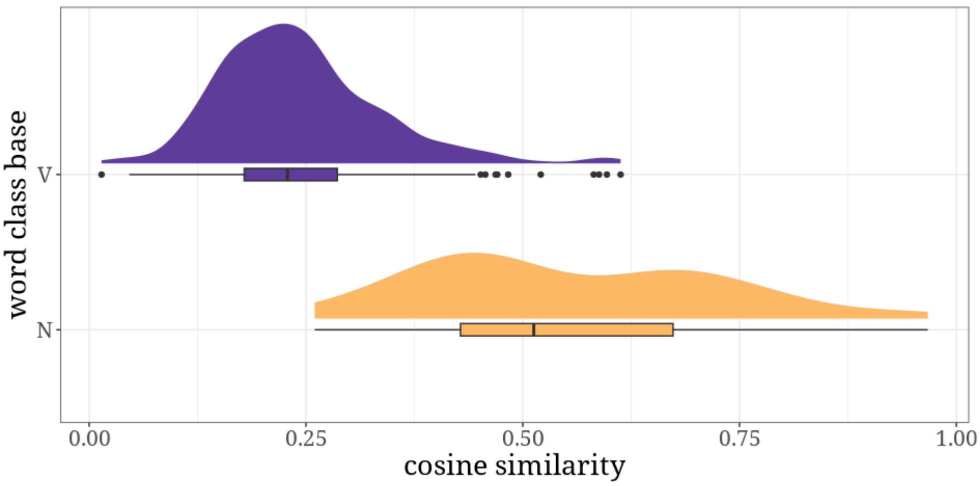

FIGURE 6.2: Cosine similarities of derivatives and bases for the suffix *-ee*. The cosine similarities of the denominal data are given in orange and the cosine similarities for the deverbal data are given in purple.

Before a beta regression model was fitted, correlation coefficients of independent variables were checked (SfL package, Schmitz & Esser 2021). No problematic correlations were observed. Several different models were fitted by the procedure explained above until a model with the best performance was found. The best fit was determined via AIC-value and log-likelihood of the models (cf. Baayen 2008: ch. 6; Winter 2019: ch. 15). The VIF-values of the final model confirm that the final model has no collinearity issues (cf. Zuur et al. 2010; Fox & Weisberg 2019). The formula of the final model is given in 6.1. The model includes precision specifications of all variables (cf. Ferrari & Cribari-Neto 2004; Plag et al. 2017).

$$\text{COSINE SIMILARITY} \sim \text{WORD CLASS OF BASE} + \text{RELATIVE FREQUENCY} + \text{BASE POLYSEMY} \mid \text{WORD CLASS OF BASE} + \text{RELATIVE FREQUENCY} + \text{BASE POLYSEMY} \quad (6.1)$$

The estimates of the final model and their p-values are given in Table 6.5. The variable of interest, WORD CLASS OF BASE reaches significance. The control variable RELATIVE FREQUENCY reaches significance. The control variable BASE POLYSEMY shows no significant effect, but it improves the model fit significantly.

TABLE 6.5: Fixed-effect coefficients and p-values as computed by the final beta regression model for the *-ee* data fitted to the cosine similarity values. Significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.

Coefficients	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.41	0.11	3.63	0.00	***
WORD CLASS OF BASE V	-1.21	0.10	-12.78	< 2e-16	***
RELATIVE FREQUENCY	-0.05	0.01	-3.72	0.00	***
BASE POLYSEMY	0.01	0.00	1.84	0.07	.
Phi coefficients	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**	
(Intercept)	2.02	0.26	8.02	1.06e-15	***
WORD CLASS OF BASE V	0.49	0.23	2.14	0.03	*
RELATIVE FREQUENCY	0.05	0.03	1.44	0.15	
BASE POLYSEMY	0.03	0.01	2.25	0.02	*

The variable WORD CLASS OF BASE reaches significance and its effect is depicted in Figure 6.3. The cosine similarity of base and derivative is higher when the base is a noun. This finding is contra the expectation. Deverbal derivatives were expected to show a higher cosine similarity with their bases compared to denominal derivatives. The higher cosine similarities for denominal derivatives and their nominal bases can be explained by the ontology of word classes (e.g., Van Valin & LaPolla 1997; Haspelmath

2001; Van Valin & LaPolla 2002; Szabó 2015; Moltmann 2019). That is, nominalizations with the suffix *-ee* create participant readings. Those readings are more noun-like than verb-like. The effect of the participant readings overrides the expected effect by verbal bases that the eventuality for the word-formation process is directly linked to the verbal base. In other words, nominalizations with the suffix *-ee* denote participants of an eventuality and are not as eventive as verbs.

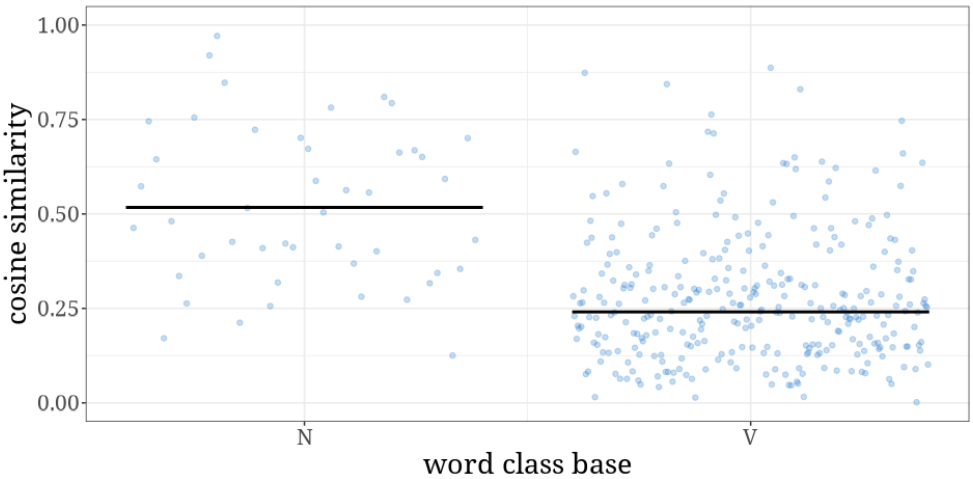

FIGURE 6.3: Significant effect of WORD CLASS OF BASE on the cosine similarity of derivatives and bases for the suffix *-ee*.

6.2.1.2 Results *-ment*

The density of the raw cosine similarities for the data in *-ment* ($n_{\text{denominal}} = 29$, $n_{\text{deverbal}} = 273$) sorted by the word class of the base is shown in Figure 6.4. The orange area shows the cosine similarity of denominal derivatives with *-ment* and their nominal bases and the purple area illustrates the cosine similarity of deverbal derivatives with *-ment* and their verbal bases. The densities suggest a difference in the data sets evoked by the word class of the base. The box plots show that the denominal pairs have a median cosine similarity of about 0.45 and the deverbal pairs of about 0.6. A Wilcoxon test shows that the difference between the two types word classes of the base is significant (Wilcoxon test, $p < 0.001$).

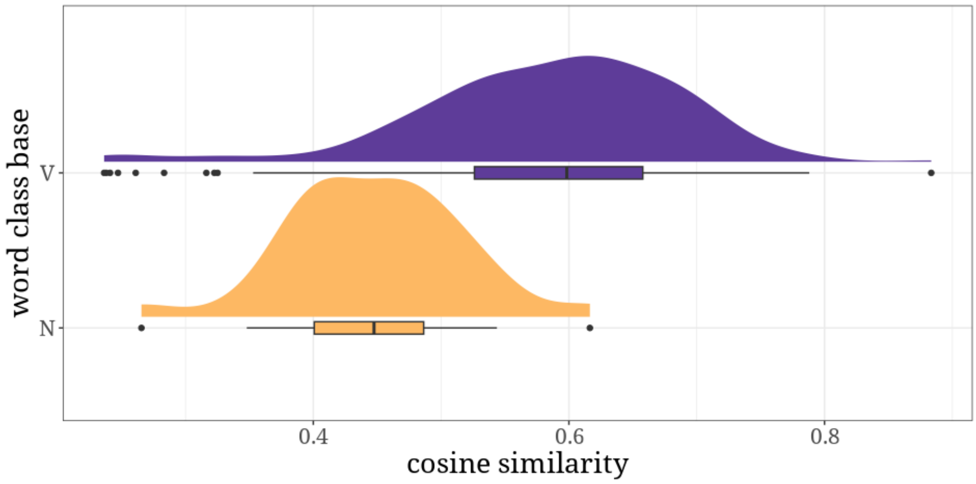

FIGURE 6.4: Cosine similarities of derivatives and bases for the suffix *-ment*. The cosine similarities of the denominal data are given in orange and the cosine similarities for the deverbal data are given in purple.

The data were checked for problematic correlations of the variables of interest with the control variables: WORD CLASS O OF BASE, RELATIVE FREQUENCY, and BASE POLYSEMY. The correlation test (SfL package, Schmitz & Esser 2021) revealed a potentially problematic correlation of the variables RELATIVE FREQUENCY and WORD CLASS OF BASE ($rho = -0.5$). Several different models were fitted by the procedure explained above until a model with the best performance was found. The best fit was determined via AIC-value and log-likelihood of the models (cf. Baayen 2008: ch. 6; Winter 2019: ch. 15). The VIF-values of the final model show, however, that the final model has no collinearity issues as the values for all included variables are below the value of 3 (cf. Zuur et al. 2010; Fox & Weisberg 2019). The formula of the final model is given in 6.2. The model includes a precision specification for WORD CLASS OF BASE (cf. Ferrari & Cribari-Neto 2004; Plag et al. 2017).

$$\text{COSINE SIMILARITY} \sim \text{WORD CLASS OF BASE} + \text{RELATIVE FREQUENCY} + \text{BASE POLYSEMY} \mid \text{WORD CLASS OF BASE} \quad (6.2)$$

The estimates of the final model and their p-values are given in Table 6.6. All variables in the model reach significance. The significant effects of the precision estimates indicate that only highly significant effect estimates represent an effect. Hence, the effect of the variable RELATIVE FREQUENCY is not interpreted as significant.

TABLE 6.6: Fixed-effect coefficients and p-values as computed by the final beta regression model for the *-ment* data fitted to the cosine similarity values. Significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.

Coefficients	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.09	0.10	0.96	0.34	
WORD CLASS OF BASE V	0.33	0.10	3.42	0.00	***
RELATIVE FREQUENCY	-0.02	0.01	-2.38	0.02	*
BASE POLYSEMY	-0.03	0.01	-3.64	0.00	***
Phi coefficients	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**	
(Intercept)	3.90	0.26	14.99	< 2e-16	***
WORD CLASS OF BASE V	-0.71	0.27	-2.58	0.01	**

Figure 6.5 shows the effect of WORD CLASS OF BASE on the cosine similarities of *-ment* derivatives. The cosine similarity increases when the word class of the base is a verb. The effect is as expected. The eventive nature of *-ment*-derivatives is related to the eventuality denoted by the base and verbs denote eventualities (e.g., Van Valin & LaPolla 1997; Haspelmath 2001; Van Valin & LaPolla 2002; Szabó 2015; Moltmann 2019). The similarity of derivatives and bases is higher for deverbal derivatives and their verbal bases due to the eventive nature of verbs.

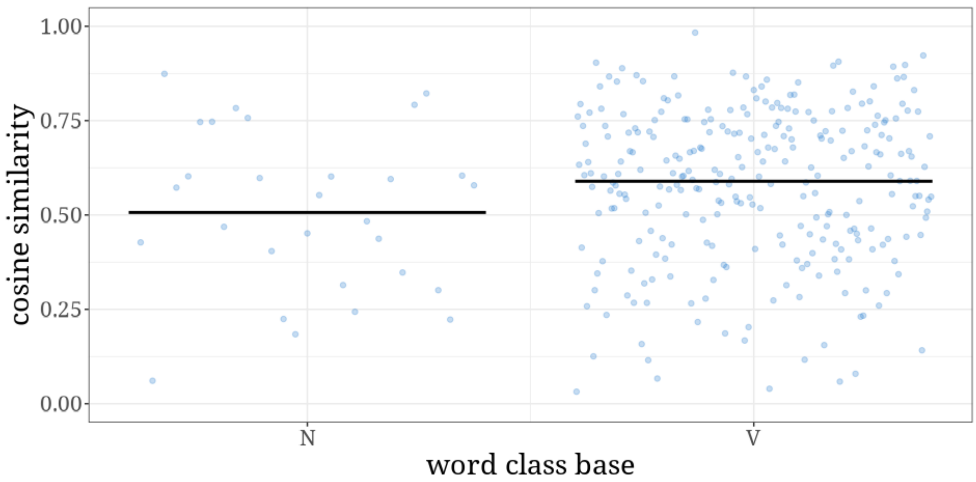

FIGURE 6.5: Significant effects of WORD CLASS OF BASE on the cosine similarity of derivatives and bases for the suffix *-ment*.

6.2.1.3 Results *-ation*

The density of the raw cosine similarities for the data in *-ation* ($n_{\text{denominal}} = 67$, $n_{\text{deverbal}} = 72$) sorted by the word class of the base is shown in Figure 6.6. The orange area shows the density of the cosine similarities for denominal derivatives and their bases. The purple area illustrates the density of the cosine similarities for deverbal derivatives and their verbal bases. A dip test for the deverbal data in the purple area indicates that the data are not bimodally but unimodally distributed (dip test, $p = 0.4654$; diptest package, Maechler 2020). The box plots show that the cosine similarity of the denominal derivatives and their bases in orange have a median of about 0.5. The cosine similarities of deverbal derivatives and their bases in purple have a median around 0.7. The difference is significant (Wilcoxon test, $p < 0.001$).

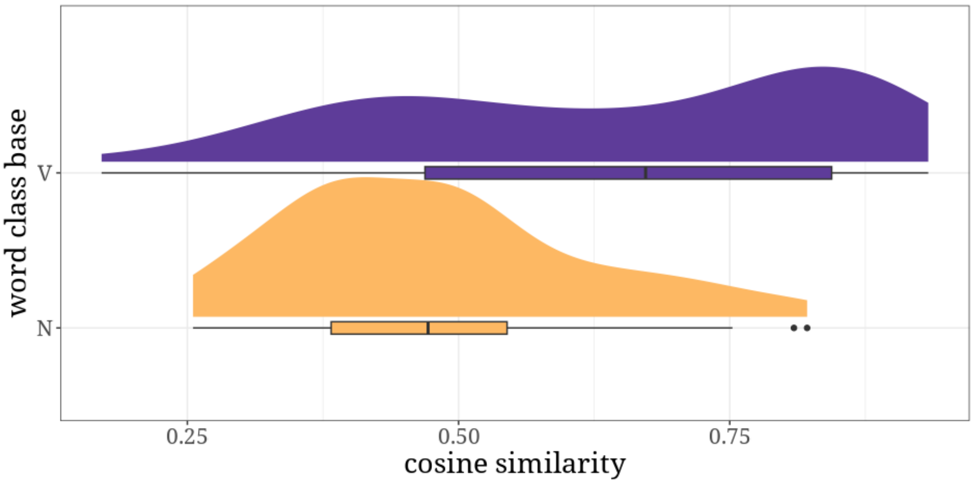

FIGURE 6.6: Cosine similarities of derivatives and bases for the suffix *-ation*. The cosine similarities of the denominal data are given in orange and the cosine similarities for the deverbal data are given in purple.

The correlation check of independent variables (SfL package, Schmitz & Esser 2021) showed a problematic correlation of RELATIVE FREQUENCY with the other two variables of interest BASE POLYSEMY ($rho = -0.62$) and WORD CLASS OF BASE ($rho = -0.71$). Several different models were fitted by the procedure explained above until a model with the best performance was found. The final model has the best fit according to AIC-value and log-likelihood of the models (cf. Baayen 2008: ch. 6; Winter 2019: ch. 15). The VIF-values of the final model show that the final model has no collinearity issues as the values for all included variables are below the value 3 (cf. Zuur et al. 2010; Fox & Weisberg 2019). The formula of the final model is given in 6.3. The model includes precision specifications of WORD CLASS OF BASE and BASE POLYSEMY (cf. Ferrari & Cribari-Neto 2004; Plag et al. 2017).

$$\text{COSINE SIMILARITY} \sim \text{WORD CLASS OF BASE} + \text{RELATIVE FREQUENCY} + \text{BASE POLYSEMY} \mid \text{WORD CLASS OF BASE} + \text{BASE POLYSEMY} \quad (6.3)$$

The estimates of the final model and their p-values are given in Table 6.7. The control variables RELATIVE FREQUENCY and BASE POLYSEMY reach significance. The significant effects of the precision estimates indicate that only highly significant effect estimates represent a true effect. Hence, the effect of the variable BASE POLYSEMY is not interpreted as significant.

TABLE 6.7: Fixed-effect coefficients and p-values as computed by the final beta regression model for the *-ation* data fitted to the cosine similarity values. Significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.

Coefficients	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.34	0.09	3.67	0.00	***
WORD CLASS OF BASE V	0.24	0.13	1.86	0.06	.
RELATIVE FREQUENCY	-0.07	0.02	-3.85	0.00	***
BASE POLYSEMY	-0.06	0.03	-2.42	0.02	*
Phi coefficients	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**	
(Intercept)	2.55	0.21	11.94	< 2e-16	***
WORD CLASS OF BASE V	-0.93	0.24	-3.94	8.27e-05	***
BASE POLYSEMY	0.19	0.08	2.45	0.01	*

The finding for the variable WORD CLASS OF BASE for nominalizations in *-ation* is unexpected. It was expected that deverbal derivatives show a higher similarity with their bases than denominal derivatives. The expectation is based on the ontology of words, more precisely, on the fact that verbs denote eventualities and nouns do not (e.g., Van Valin & LaPolla 1997; Haspelmath 2001; Van Valin & LaPolla 2002; Szabó 2015; Moltmann 2019). As the variable shows no significant effect, the word class of the base seems not to be a predictor for the cosine similarity of derivatives and bases. Broadly speaking, the word class of a base is not an influential factor for the derivational process with the suffix *-ation*.

6.2.2 t-SNE and LDA analyses

In order to see whether the vectors for the derivatives themselves show a pattern, a t-distributed stochastic neighbor embedding (t-SNE) analysis was performed. Such an analysis reduces the dimensions of the vectors, 300 dimensions in this study (for more information on t-SNE analyses, see, e.g., van der Maaten & Hinton 2008; van der Maaten 2014; Krijthe 2015; Shafaei-Bajestan et al. 2022b; Schmitz et al. 2023; Schäfer 2025). Two dimensions are retained for the analysis (Rtsne package, Krijthe 2015;

gdsm package, Schmitz & Schneider 2022). This analysis was included to investigate if there is a pattern to be found which distinguishes denominal and deverbal derivatives in general. In order to see whether the found clusters have significance, an LDA was performed. The findings for the data of the distribution for the derivatives suggests a difference of denominal/deverbal for *-ee* and *-ment* due to the significance of the variable WORD CLASS OF BASE, but not for *-ation*.

6.2.2.1 Results *-ee*

Figure 6.7 illustrates the *t*-SNE analysis for the derivatives in *-ee*. No visible clusters of the derivatives are observable. The vectors for the denominal derivatives are in the same area as the vectors for the deverbal derivatives.

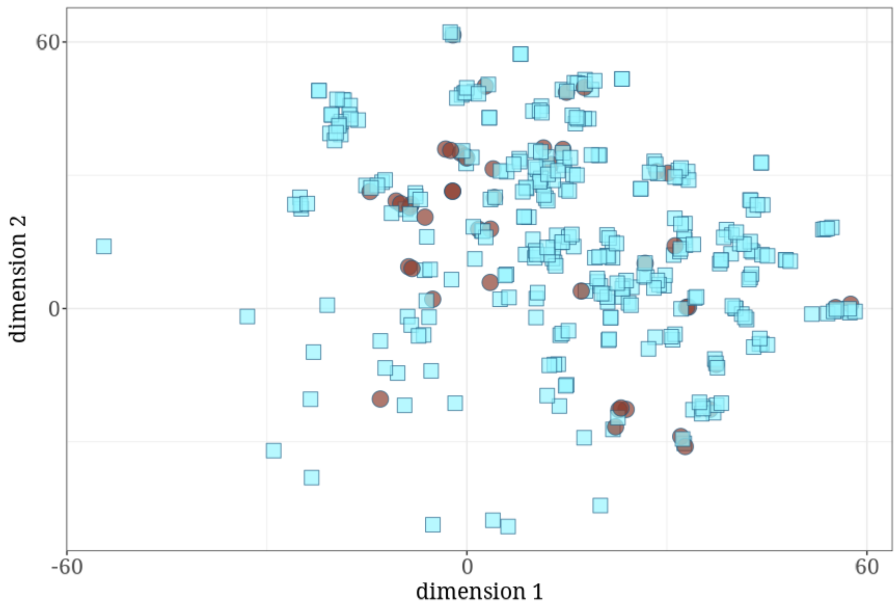

FIGURE 6.7: *t*-SNE for the suffix *-ee*. Red circles are denominal derivatives and blue squares are deverbal derivatives.

The LDA distribution of denominal and deverbal derivatives is depicted in Figure 6.8. The bars of the denominal data are completely covered by the bars of the deverbal data.

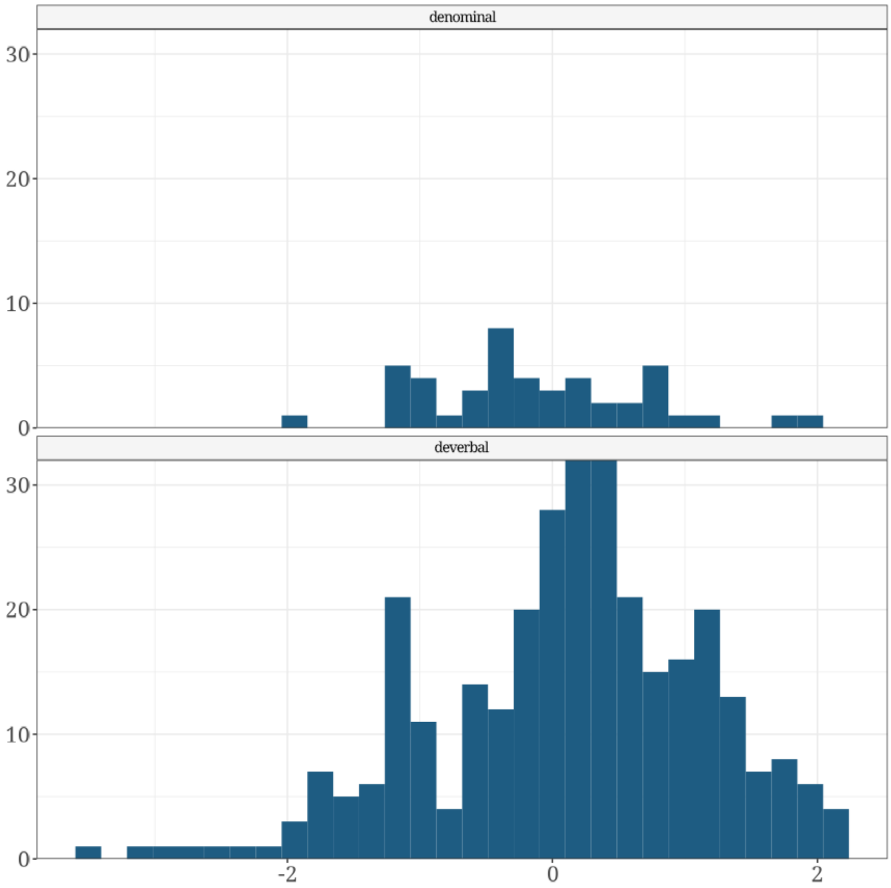

FIGURE 6.8: LDA histogram for the suffix *-ee*, displaying the distribution of the data.

Table 6.8 depicts the prediction of the LDA. All words are predicted to be deverbal.

TABLE 6.8: LDA prediction for *-ee*.

	Reference	
Prediction	N	V
N	0	46
V	0	312
Accuracy	0.8715	
95% CI	(0.8324, 0.9044)	
No Information Rate	1	
P-Value [Acc > NIR]	1	
Kappa	0	
Mcnemar's Test P-Value	3.247e-11	
Sensitivity	NA	
Specificity	0.8715	
Pos Pred Value	NA	
Neg Pred Value	NA	
Prevalence	0.0000	
Detection Rate	0.0000	
Detection Prevalence	0.1285	
Balanced Accuracy	NA	
'Positive' Class	N	

The results of the t-SNE analysis as well as the results from the LDA show that there is no clear distinction of denominal and deverbal nominalizations with the suffix *-ee*. The word class of the base seems not to be the most crucial factor for the analyses of derivatives. The clustering of the derivatives shows that the influence of the suffix leads to semantic similarity independent of the syntactic category of the base.

6.2.2.2 Results *-ment*

The t-SNE analysis for *-ment*-derivatives is illustrated in Figure 6.9. Two clusters are visible: one for denominal derivatives in the upper right corner and one for deverbal derivatives in the left side of the plot.

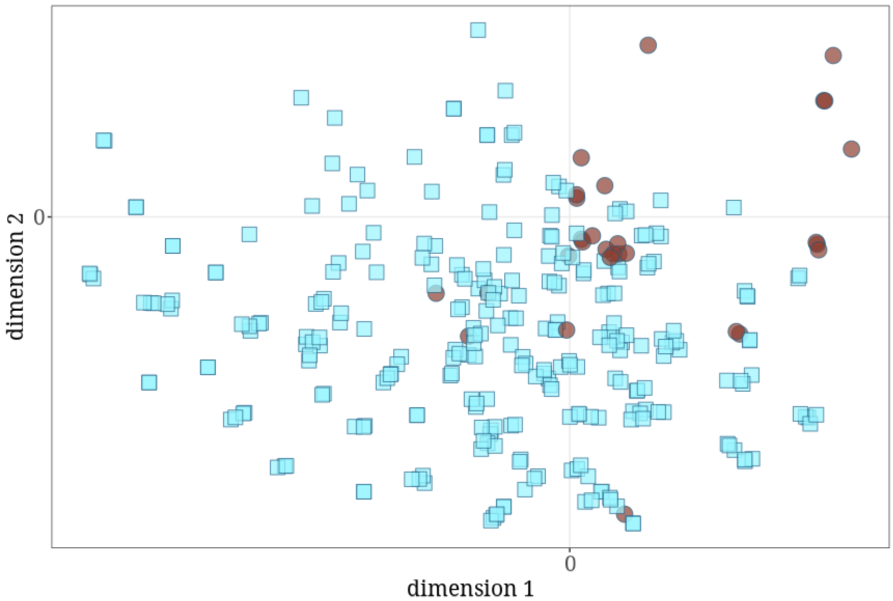

FIGURE 6.9: t-SNE for the suffix *-ment*. Red circles are denominal derivatives and blue squares are deverbal derivatives.

Figure 6.10 shows the words in the visible clusters. For convenience, some examples are given in (1).

(1) a. denominal *-ment*
rascalment, bridement, chairment

b. deverbal *-ment*
overtreatment, undertreatment, reenlistment

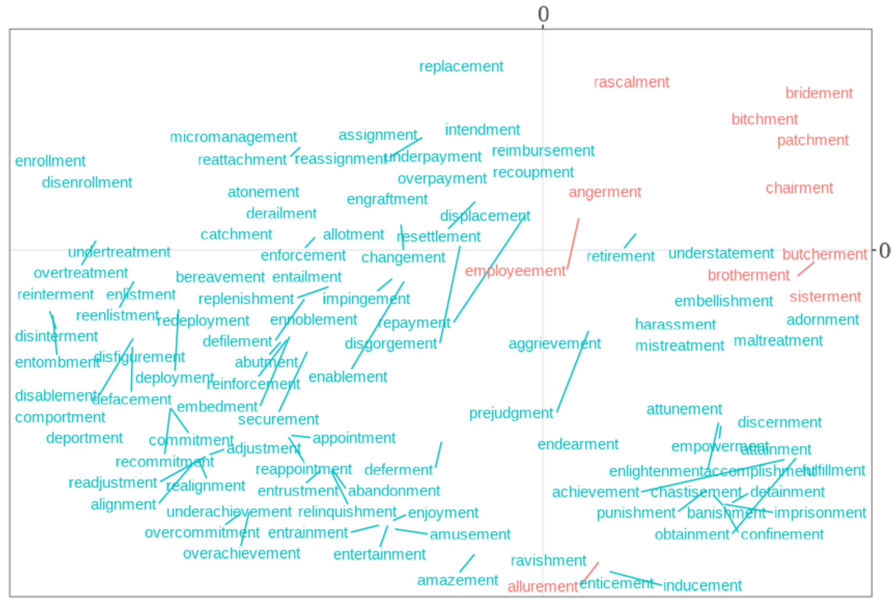

FIGURE 6.10: t-SNE for the suffix -*ment*. Red words are denominal derivatives and blue words are deverbal derivatives.

Figure 6.11 depicts the LDA for -*ment*-derivatives. Similar to the results from the t-SNE analysis, a difference between denominal and deverbal derivatives emerges from the data. The vectors of denominal nominalizations show data points further to the left whereas deverbal nominalizations show some data further to the right. There is an overlap of these two categories of nominalizations in the middle range.

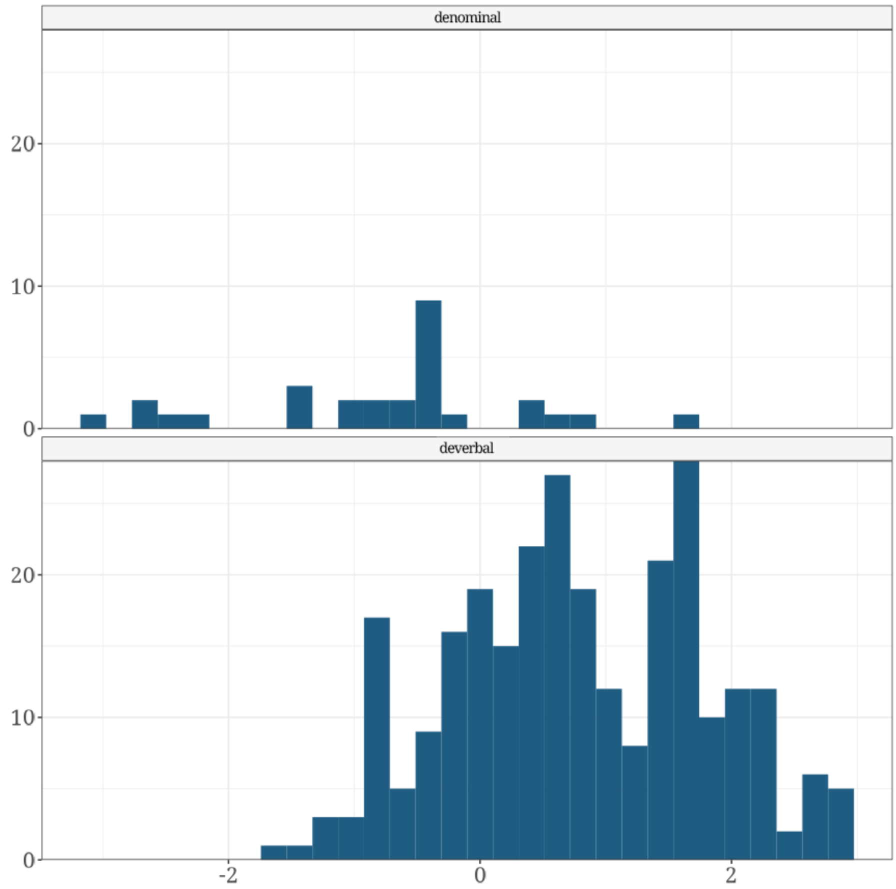

FIGURE 6.11: LDA histogram for the suffix *-ment*, displaying the distribution of the data.

Table 6.9 depicts the outcome of the LDA. The prediction is visualized in Figure 6.12. The orange area for the denominal derivatives is clearly smaller than the purple area for deverbal derivatives. A lot of denominal derivatives are wrongly predicted as they are in the deverbal area.

TABLE 6.9: LDA prediction for *-ment*.

	Reference	
Prediction	N	V
N	8	21
V	1	272
Accuracy	0.9272	
95% CI	(0.8918, 0.9538)	
No Information Rate	0.9702	
P-Value [Acc > NIR]	1	
Kappa	0.3935	
Mcnemar's Test P-Value	5.104e-05	
Sensitivity	0.88889	
Specificity	0.92833	
Pos Pred Value	0.27586	
Neg Pred Value	0.99634	
Prevalence	0.02980	
Detection Rate	0.02649	
Detection Prevalence	0.09603	
Balanced Accuracy	0.90861	
'Positive' Class	N	

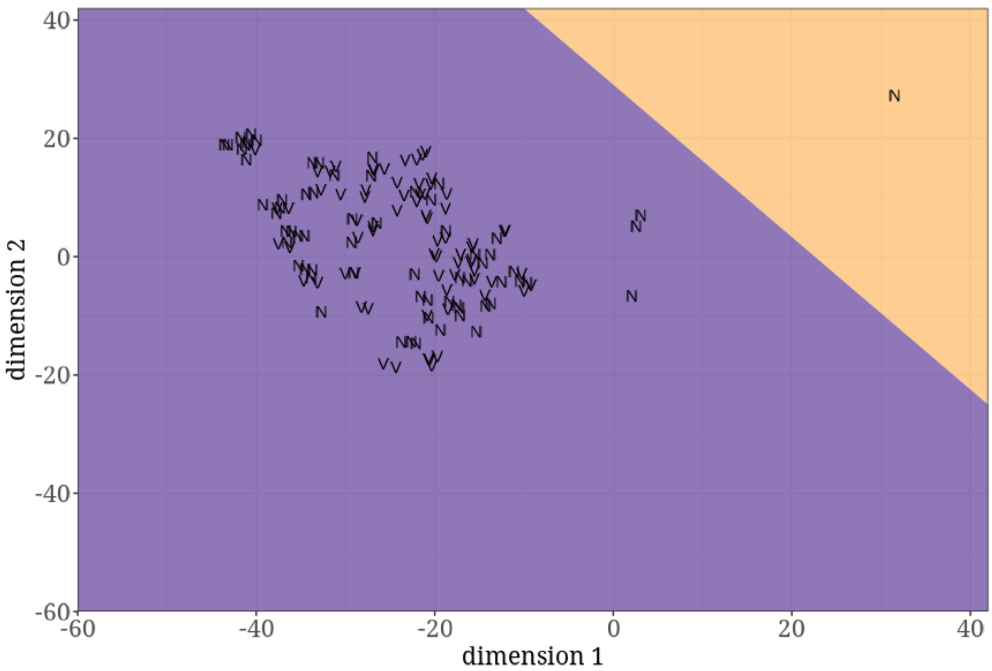

FIGURE 6.12: LDA prediction for the suffix *-ment*.

The *t*-SNE analysis shows a small cluster of denominal derivatives. The performed LDA illustrates that the denominal derivatives in the cluster are correctly predicted as denominal. All other denominal derivatives, which are not part of the cluster, are wrongly predicted as being deverbal. Only one deverbal derivative is wrongly predicted as denominal. The distinction of the derivatives is made due to a semantic effect, i.e, the cluster consists of attitudinal nouns (cf. Chapter 4). Furthermore, the absence of further clusters in the analysis shows that the word-formation process with the suffix *-ment* creates semantically related nominalizations. The results show clearly that the semantics of derivatives are a more influential factor than the word class of the base, which is in line to previous findings (cf. Plag 2004).

6.2.2.3 Results *-ation*

Figure 6.13 shows the *t*-SNE analysis of denominal and deverbal nominalizations with *-ation*. In the upper left corner, as well as in the right half of the plot, denominal clusters are visible.

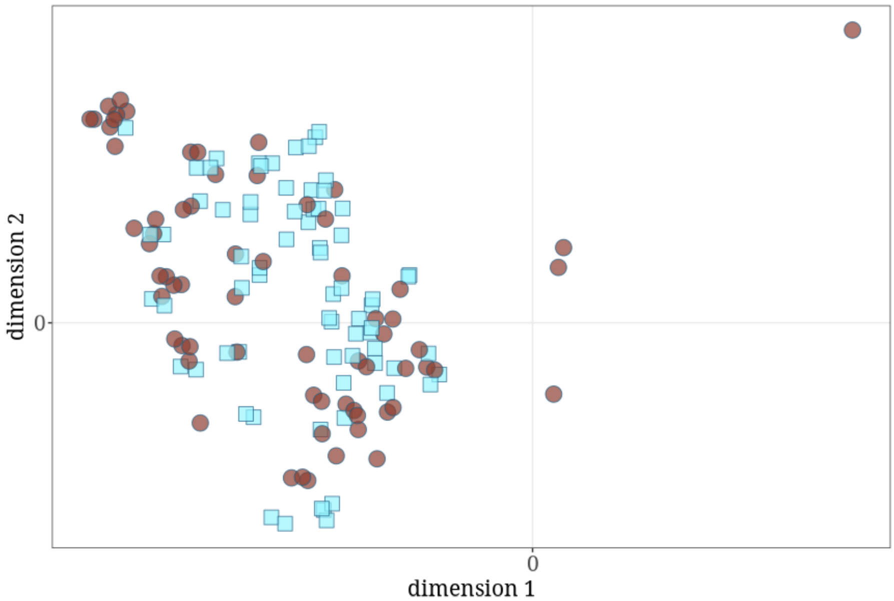

FIGURE 6.13: *t*-SNE for the suffix *-ation*. Red circles are denominal derivatives and blue squares are deverbal derivatives.

Figure 6.14 shows the words in the visible clusters. For convenience, some examples of the denominal cluster in the left corner are given in (2).

(2) a. denominal *-ation*
myristylation, acetylation, glycosylation
b. deverbal *-ation*
extrication, fantastication, lignification

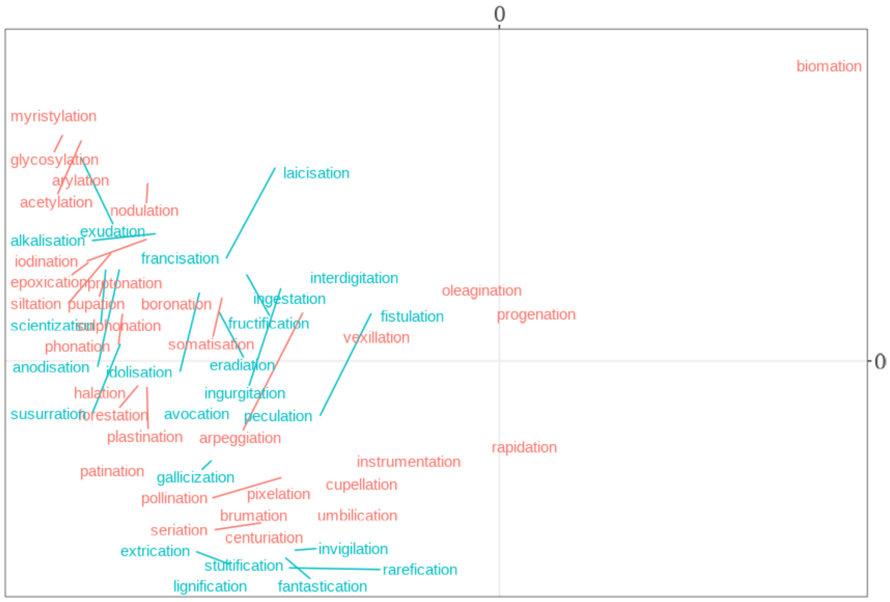

FIGURE 6.14: t-SNE for the suffix -*ation*. Red words are denominal derivatives and blue words are deverbal derivatives.

The histogram of the LDA shows a similar picture. In Figure 6.15 some denominal data points are further to the left, as well as further to the right, than the deverbal data.

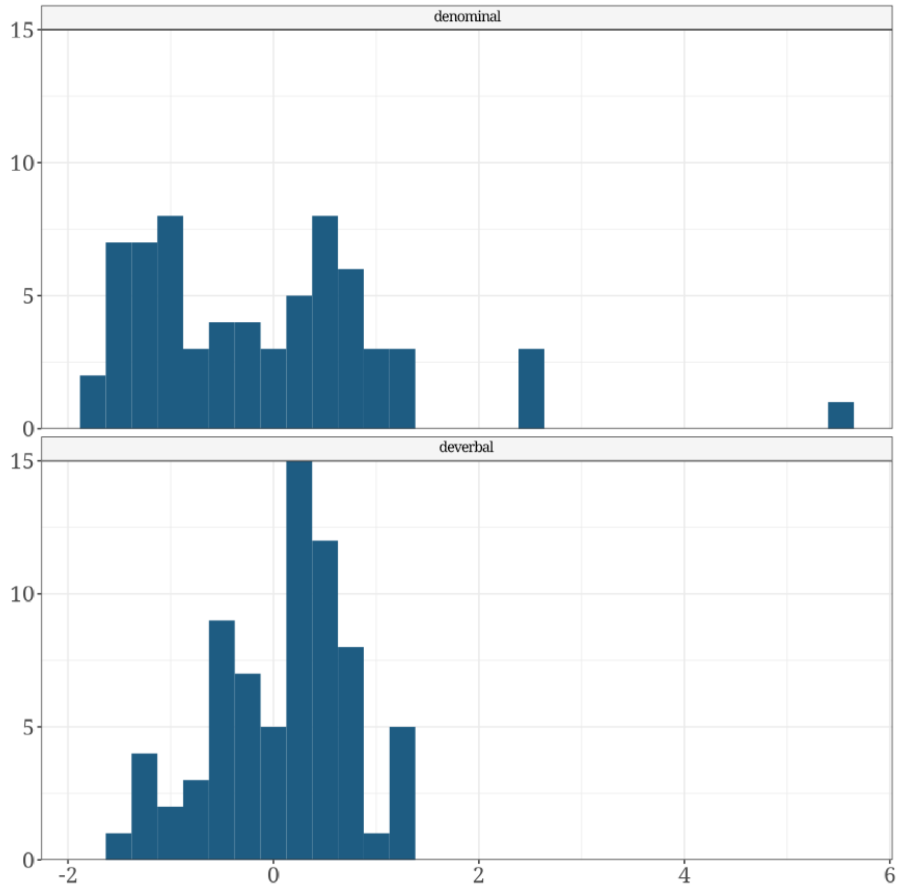

FIGURE 6.15: LDA histogram for the suffix *-ation*, displaying the distribution of the data.

The prediction of the LDA is given in Table 6.10. Figure 6.16 visualizes the prediction. The purple area depicts the space for the deverbal nominalizations and the orange area represents the space for the denominal nominalizations. Deverbal as well as denominal derivatives are wrongly predicted and in the wrong area in the plot.

TABLE 6.10: LDA prediction for *-ation*.

	Reference	
Prediction	N	V
N	31	36
V	18	54
Accuracy	0.6115	
95% CI	(0.5252, 0.6929)	
No Information Rate	0.6475	
P-Value [Acc > NIR]	0.8357	
Kappa	0.2147	
Mcnemar's Test P-Value	0.0207	
Sensitivity	0.6327	
Specificity	0.6000	
Pos Pred Value	0.4627	
Neg Pred Value	0.7500	
Prevalence	0.3525	
Detection Rate	0.2230	
Detection Prevalence	0.4820	
Balanced Accuracy	0.6163	
'Positive' Class	N	

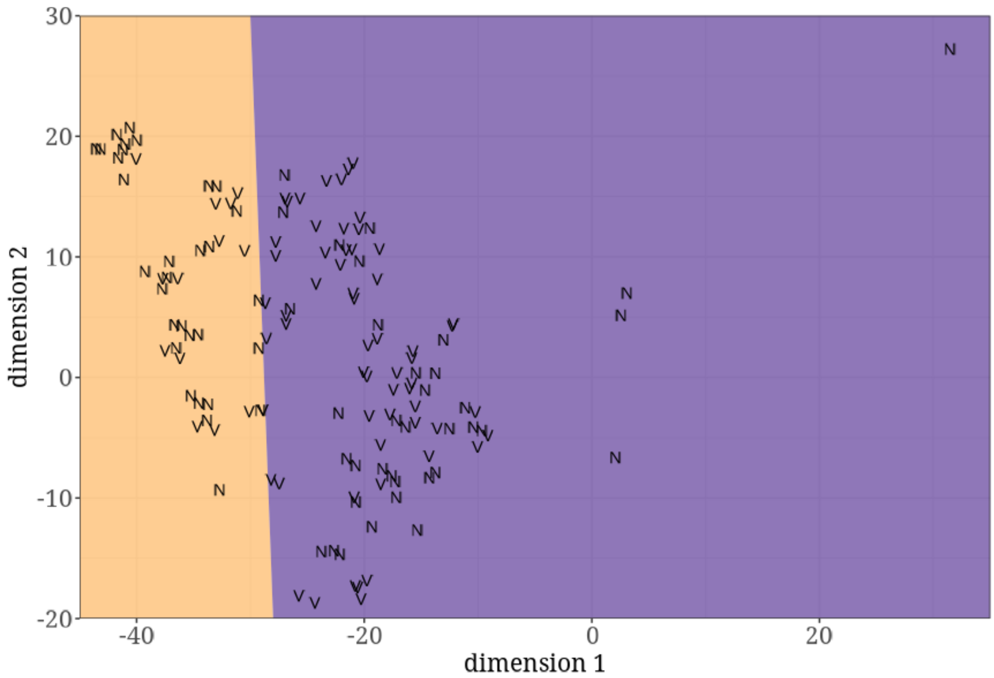

FIGURE 6.16: LDA prediction for the suffix -*ation*.

The results of the t-SNE and the LDA show that for nominalizations with the suffix -*ation* the word class of the base can be a distinguishing factor. One cluster clearly demonstrates chemical -*ation* -derivatives with a nominal base. No other clusters were visible and the LDA still predicts a lot of derivatives wrongly as deverbal. This result hints again to the direction that the semantics of the derivatives and their bases are more important for an account of the word-formation process than the word class of the base.

6.3 Discussion

The previous sections gave insight into the eventuality-related nominalization processes with the suffixes *-ee*, *-ment*, and *-ation* from a distributional semantic perspective. The vectors were taken from a *fastText* implementation using a pre-trained data set and character *n*-grams for the computation of the semantic word vectors (Bojanowski et al. 2016; Mikolov et al. 2018). Cosine similarities were used to measure the similarity of derivatives and their bases. The variable of interest in the analyses was WORD CLASS OF BASE, and the variables RELATIVE FREQUENCY, and BASE POLYSEMY were included as control variables. Their expected effects on the cosine similarity of derivative and base are repeated in Table 6.11.

TABLE 6.11: Dependent variable, variable of interest, control variables and their expected effects.

dependent variable	expectation
COSINE SIMILARITY	The cosine similarity of derivatives and bases indicates how similar one derivative is to its base.
variable of interest	expectation
WORD CLASS OF BASE	Verbal bases are more similar to their derivatives due to their denotation of eventualities.
control variables	
RELATIVE FREQUENCY OF BASE/DERIVATIVE	A higher relative frequency goes together with a higher segmentability and more transparency and thus an increase of cosine similarity.
BASE POLYSEMY	A higher polysemy of the base goes together with a decrease of cosine similarity as the derivative is (often) based on one reading of the base.

The variable of interest is WORD CLASS OF BASE, i.e., if the base of a derivative is nominal or verbal. Due to the ontology of verbs (for more on ontological categories, see, e.g., Van Valin & LaPolla 1997; Haspelmath 2001; Szabó 2015; Moltmann 2019), deverbal derivatives were expected to be more similar to their verbal bases than denominal derivatives to their nominal bases. The found effects of the word class of the base go into different directions for different suffixes. For derivatives in *-ee*, denominal derivatives are more similar to their nominal bases compared to the results for deverbal derivatives and verbal bases. The data sets for *-ment* and *-ation* show the opposite picture: Deverbal derivatives and their verbal bases are more similar to each other compared to denominal derivatives and their nominal bases. However, WORD CLASS OF BASE did not reach significance in the beta regression model for *-ation*.

The differences in cosine similarities regarding WORD CLASS OF BASE for the data sets might be explained by the different readings created by the suffixes. Derivatives with the suffix *-ee* create a participant reading (see, e.g., Barker 1998; Plag 2004; Bauer et al. 2013; Plag 2018; Chapter 4). Derivatives with *-ation* describe mostly processes (see, e.g., Bauer et al. 2013; Plag 2018). The suffix *-ment* can create all different sorts of readings with the exception that it cannot take an animate participant to refer to (Kawaletz 2023). Assuming that participants are usually represented as nouns, and processes and other eventualities are usually denoted by verbs, the differences in the cosine similarities for the denominal and deverbal data can be explained: The significantly higher similarity of denominal *-ee*-formations lies in the semantic nature of the denotation of participants by nouns. The readings created with the suffixes *-ation* and *-ment*, in turn, are semantically more linked to the denotation of verbs, i.e., eventualities, and show a lower similarity of nominal bases and denominal derivatives. The findings for *-ation* and *-ment*, as well as the finding that eventuality-related nominalizations with the suffix *-ee* show the highest similarity of denominal derivatives and their nominal bases, were expected.

However, for all the suffixes in this study, verbs as bases are far more productive. Hence, for all three suffixes, the expectation was that deverbal nominalizations are more similar to their verbal bases than denominal nominalizations to their nominal bases. The results of the data with the suffixes *-ation* and *-ment* are in line with this expectation. The word-formations with *-ee*, contrarily, show the opposite picture: Nominalizations with nominal bases are more similar to their bases than deverbal nominalizations and their verbal bases. This result shows that not only the word class of the base and its semantic structure play a role in the word formation process but also the readings created by adding the suffix.

An issue with the study in this chapter might be the difference in sample size. The sample size itself is problematic for a distributional analysis, as, in an ideal world, all data sets should have a near to equal number and, within the sets, a near to equal sample of nouns and verbs. This is not the case due to two reasons: First, eventuality-related denominal derivatives are clearly rarer than deverbal derivatives (see Chapter 3). Second, the samples for the deverbal derivatives were extracted randomly from several corpora (see Section 6.1). This unequality in the distribution of denominal and deverbal nominalizations in the data may be considered problematic because the results of the comparisons are based on small samples. However, the results show clearly pronounced pictures, and a bigger sample of denominal derivatives would probably not change the overall results observed in this study.

The fundamental assumption for the present distributional semantic study, which was based on the findings for deverbal derivatives in the literature (e.g., Kawaletz 2023) and the analysis of denominal derivatives in Chapter 4, is that bases and derivatives are semantically similar in the first place. More precisely, the base word and the nominalization operate on the same semantic structures and are thus semantically related. The cosine similarities show that a semantic relation is measurable.

In the t-SNE and LDA analyses of the denominal and deverbal nominalizations, no clear effect of the word class of the base is found. Some clusters are visible for the data in *-ment* and *-ation*, but no clusters for *-ee* were found. The found clusters speak in favor of a semantic distinction. For the word class of the base the results indicate that it is not the most influential factor for the word-formation processes, but rather semantic categories like attitudinal nouns are.

The beta regression analysis of eventuality-related nominalizations with nominal and verbal bases showed that the word class of the base seems to play a role in the word-formation process at least for the nominalizations in *-ee* and *-ment*. This is in line with the ontology of words (see, e.g., Van Valin & LaPolla 1997; Haspelmath 2001; Szabó 2015; Moltmann 2019). Verbs are eventive in nature, and the nominalizations in *-ation* and *-ment* clearly show a higher similarity to verbal bases than to nominal bases. However, the word class of the base does not seem to be the most influential factor. Especially the missing clusters of a clear division of denominal and deverbal nominalizations in the t-SNE and LDA analyses raise doubt whether the word class of the base is a strong distinctive variable for word-formation processes. Moreover, the observation of the deverbal derivatives being more similar to their verbal bases than denominal derivatives to their nominal bases is only found for two of the three suffixes. Nominalizations in *-ee* behave differently, i.e., denominal derivatives are semantically more similar to their nominal bases than deverbal derivatives to their verbal bases. Considering the different findings for the importance of the word class as well as the different effects, the question of which factors influence the word-formation process arises. The distributional analysis clearly points out that the suffix is an influential factor. This is unsurprising because several suffixes are used to create different readings, sometimes also based on the same base (e.g., *employer, employee, employment* ♦ *employ*).

To summarize, the *fastText* implementation showed that the meaning of denominal eventuality-related nominalizations is similar to their bases. Furthermore, the comparison of denominal and deverbal derivatives indicates that different suffixes show different results regarding the influence of the word class of the base. The next chapter presents a linear discriminative learning implementation. The implementation is directly comparable to the implementation by *fastText* presented in the current chapter. The computational model of LDL sees bases and derivatives as parts with meaning, i.e., n-grams. The difference between the *fastText* and the LDL approach is the underlying computational method: *fastText* is a deep learning model, whereas discriminative learning is based on cognitive psychology. The implementation of LDL is made to see whether the results from the *fastText* implementation are reproducible by a cognitive based approach which is mathematically simpler and thus gives more transparent results.

Chapter 7

Linear Discriminative Learning

In order to see whether a different computational approach leads to similar results and may strengthen the findings of the *fastText* implementation, Linear Discriminative Learning (LDL) is used.[26] Discriminative learning is a cognitive-based approach which aims to simulate the learning process of a human being. This is of interest for the present investigation due to two reasons. First, many derivatives investigated in this dissertation are low in frequency. Thus, a computational implementation which is able to deal with low frequent words, like *fastText*, is required. The low frequency of the derivatives as well as the fact that many derivatives are not found in pre-trained models makes it reasonable to treat them as pseudowords, i.e., words that newly enter the mental lexicon in a learning process. LDL has been shown to be successful, among other things, for the analysis of pseudowords (cf. Chuang et al. 2021; Schmitz et al. 2021; Schmitz 2022). Second, the LDL learning process represents the resonance of words with the whole lexicon. This resonance can give further insights into the semantics of the derivatives.

[26]The data as well as the scripts used for the individual analyses are available here: `https://osf.io/kaqsv/`.

7.1 Methodology

7.1.1 LDL implementation

The LDL implementation used for the analysis in this dissertation is implemented with the R-package WpmWithLdl (Baayen et al. 2019a). To conceptualize the semantics of words and derivational and inflectional functions, three different approaches can be applied in LDL (cf. Stein 2023 for more information on the different conceptualizations of LDL implementations). The analysis in this dissertation follows a whole word approach. That is, simplex as well as complex words are represented by their pertinent vectors directly taken from a semantic vector space. For example, *banana* is represented as vector $\overrightarrow{banana}$ and *bananas* is represented as vector $\overrightarrow{bananas}$. The LDL implementation using whole words resembles the make-up of the *fastText* implementation (Chapter 6) as the LDL implementation only knows of whole words, i.e., a whole word approach.

In LDL, the mapping of form and meaning is estimated using the linear algebra of multivariate regression (for more details on the mathematics behind the linear mapping of matrices, see, e.g., Baayen et al. 2019b). The present study uses trigrams, i.e., strings of three graphemes, for the form matrix. Trigrams were chosen due to the successful implementation of earlier LDL implementations with strings of three elements, i.e., triphones and trigrams (see, e.g., Milin et al. 2017; Baayen et al. 2019b; Chuang et al. 2021; Schmitz & Esser 2021; Stein & Plag 2021; Schmitz 2022; Stein 2023). Trigrams were chosen instead of triphones to ensure that the studies in Chapter 6 and Chapter 7 are comparable to each other (because *fastText* makes use of n-grams).[27] The presence of a trigram is marked with 1 in the form vector $\vec{c}$ for each word, whereas the absence of a trigram is marked with 0. All trigram occurrences for every word in the LDL lexicon build the form matrix C. An example C matrix is given in Table 7.1. A real matrix has more dimensions than the three trigrams shown here.

[27] The used n-grams are character n-grams as described in Chapter 6.

TABLE 7.1: Example of trigrams in the C matrix.

	#bi	bio	iog	ogr	gra	rap	aph	phy	hy#	ee#
biography	1	1	1	1	1	1	1	1	1	0
biographee	1	1	1	1	1	1	1	0	0	1
biology	1	1	0	0	0	0	0	0	0	0

The other component needed is the meaning of words which is introduced by word vectors. The word vectors can be computed in different ways (see Section 2.2.2). The LDL implementation in this dissertation uses vectors computed by NDL (adopted from Baayen et al. 2019b, using TASA). As the data in TASA provides too much input for the computation of the LDL implementation in R, a subset of the data was chosen, determined by the MALD corpus (Tucker et al. 2019). The subset which was used for the LDL implementation contained all words which are present in both corpora, TASA and MALD. The vectors for the derivatives and their bases were removed to ensure that all the words are newly learned by the computation of the LDL implementation. This procedure was chosen as not all derivatives and bases were in the NDL vector data. Due to the lack of word vectors, an imbalance in the training of the LDL implementation could have been a problem if it knew some derivatives already and saw some target words as new. The word vectors constitute the S matrix, where each word is represented as a semantic vector $\vec{s}$.

The C matrix, consisting of trigrams, and the S matrix, consisting of semantic word vectors, are used to implement the linear discriminative learning process. That is, the forms in C are used to predict the semantics in S as illustrated in Figure 7.1. In order to transform the C matrix into the S matrix, a transformation matrix F is needed. This matrix is received by multiplying the Moore-Penrose generalized inverse of C (for more information on the inverse of C, see, e.g., Moore 1920; Penrose 1955, for the calculation of the inverse of C the R package MASS, Venables & Ripley 2002) with the S matrix.

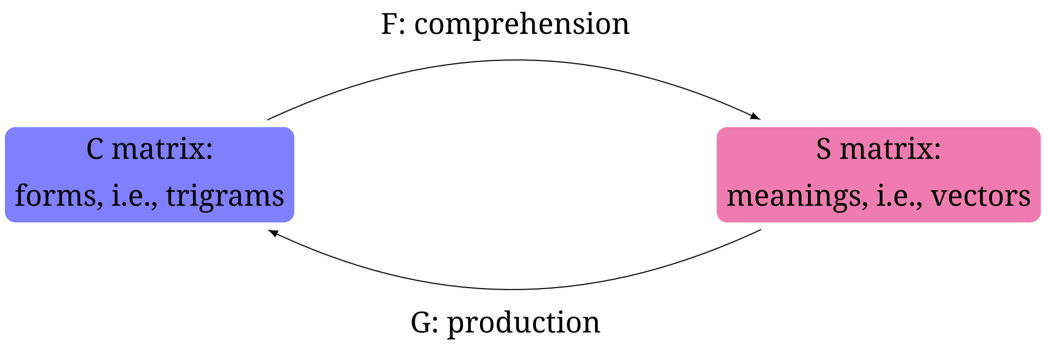

FIGURE 7.1: Schematic illustration of mapping between C and S matrix via F, i.e., comprehension, and S and C matrix via G, i.e., production (cf. Schmitz 2022).

The result of the transformation from the C matrix to the S matrix via the transformation matrix F is the predicted semantic matrix $\hat{S}$. Only estimated semantics are reached due to the mathematical operations on high-dimensional spaces, i.e., matrix multiplication. The comprehension process is successful when the outcome of the computation in $\hat{S}$ shows the highest correlation with its observed vector in S (Baayen et al. 2019b).

Figure 7.1 also shows the reverse process with the transformation matrix G for production. In production, the model tries to produce words, i.e., forms, by using semantic information. For the calculation of the G matrix, the inverse of S is needed instead of the inverse of C. The S matrix is then multiplied with G to reach $\hat{C}$. Both processes, comprehension with F and production with G, work similarly as they linearly map matrices onto each other to either comprehend the meanings of forms or to produce forms from meanings.

The same derivatives and bases for the suffixes *-ee*, *-ment*, and *-ation* were used, ajar to the used words in the *fastText* implementation in Chapter 6. For the retrieval of the denominal as well as the deverbal derivatives and bases see Chapter 3. Note that a few words were lost due to a computational issue during the implementation of LDL: Target words containing a trigram which is not contained in the trained corpus. Thus, two data points were lost resulting in the numbers of types in Table 7.2.

TABLE 7.2: The data of all three suffixes ordered by base type, denominal/deverbal. The numbers in brackets are the original numbers for the data sets, the numbers without brackets are the numbers of derivatives which entered the implementation.

	denominal	deverbal
-ation	66 (67)	72
-ee	46	311 (312)
-ment	29	273
total	141 (142)	656 (657)

The overall idea of LDL is that one can map form onto meaning and meaning onto form. This mapping is performed by matrix multiplication which creates an estimated matrix of the original input. More precisely, the output of the learning process performed by LDL is an estimation of the original input which illustrates how successful the learning process was. If the estimated vectors, either of form or meaning, are close to the original input, the learning process was successful. For the mathematical process, first, two matrices are required to implement LDL. The S matrix is the semantic part of the lexicon, the C matrix is the form part.

The first semantic matrix required for the implementation is the S_{train} matrix based on the vectors from NDL. All target words, the derivatives and bases used in this study, were removed if they were available in the subset of TASA. The S_{train} matrix consists of 3766 lexical words in 5487 dimensions.

In order to perform matrix multiplication for predicting the meanings of forms in LDL, a form matrix is also required (C matrix). The study at hand requires more than one form matrix due to the removal of derivatives and bases under investigation from the S_{train} matrix to implement all derivatives and their bases the same way as in the first learning process. The details of the first learning process will be described later in this section. Overall, three C matrices are created as given in Table 7.3:

TABLE 7.3: Three required C matrices for the LDL implementation.

C_{train}	contains the trigrams of all retained MALD words
C_{target}	contains the trigrams of all target words
$C_{combined}$	contains the trigrams of C_{train} and $C_{combined}$

As pointed out earlier, the transformation matrix F makes the linear mapping of form (C matrix) and semantics (S matrix) possible. The F matrix is created using S_{train} and C_{train}. As it is assumed that

$$C_{train}F_{train} = S_{train}$$

it is also assumed that

$$F_{train} = C'_{train}S_{train}$$

where C'_{train} is the Moore-Penrose generalized inverse of C_{train} (for more information on the inverse of C). In order to create the semantics for the target words, i.e., the S_{target} matrix, F_{train} and C_{target} are used. More precisely, using F_{train}, the semantics of the target words using C_{target} can now be estimated:

$$C_{target}F_{train} = S_{target}$$

For the computation of the LDL implementation and its measures, the entire lexicon has to be modeled. For this implementation, the form matrix $C_{combined}$ and a combined semantic matrix $S_{combined}$ are used. $S_{combined}$ can be obtained by simply combining S_{train} and S_{target}.

Using the resulting bigger $S_{combined}$ matrix and the $C_{combined}$ matrix, a second F matrix, $F_{combined}$, can be computed:

$$F_{combined} = C'_{combined}S_{combined}$$

The estimated semantic matrix $\hat{S}_{combined}$ can now be computed by using $F_{combined}$:

$$F_{combined}C_{combined} = \hat{S}_{combined}$$

The cosine similarities of derivatives and their bases are computed from $\hat{S}_{combined}$. With $\hat{S}_{combined}$ and $S_{combined}$, the accuracy of the implementation as well as pertinent semantic LDL measures can be computed. The accuracy of an implementation can be computed by taking a vector from a word from S and $\hat{S}$ and compute their correlation. Then the same vectors from $\hat{S}$ are taken and its correlation with all other vectors in S is computed. If the correlation coefficient of the word vectors in S and $\hat{S}$ is higher than the correlation of the vector of the word coefficient in $\hat{S}$ with all other vectors in S, then the semantics of a word was comprehended correctly. This is done for all words, i.e., all lexical entries and all rows. The percentage of correctly comprehended words is then computed and a statement about the accuracy of the implementation is available. The accuracy of the implementation is about 70%.

7.1.2 Statistical data analysis

The statistical analysis is twofold. First, beta regression models are fit to see how predictive the variable WORD CLASS OF BASE is for the cosine similarities of derivatives and bases (Section 7.2.1). The beta regression models contain control variables to control for possible influences on the variable of interest. Second, a t-distributed Stochastic Neighbor Embedding (t-SNE) and a linear discriminant analysis (LDA) are performed (Section 7.2.2). These analyses are performed with the vectors computed for the derivatives. The aim is to see whether patterns in the data can be found as well as to see whether denominal and deverbal derivatives are distinguishable from a discriminant view.

7.1.2.1 Beta regression

For the beta regression analyses of the semantic similarity of the derivatives and bases, cosine similarity was used as measure. The explanation of beta regression in general can be found in Chapter 6. The cosine similarities of the word vectors representing the derivatives and their bases from the LDL implementation were computed in R (gdsm package, Schmitz & Schneider 2022).

In a next step, the cosine similarity values entered beta regression models in R (R Core Team 2025) using the betareg package (Cribari-Neto & Zeileis 2010) to find out which factors influence the similarity between base and derivative. Beta regression was chosen as the statistical tool of choice as the cosine similarities in this study were mostly in the interval of [0,1]. Some values were negative, where negative values indicate opposite meanings. Some negative cosine similarities were found in the analysis, but all numbers were nearly equal to zero. Thus, negative values were transformed into positive values as a negative value nearly equal to

0 should be as relevant to any regression model as a positive value nearly equal to 0. The dependent variable for the models is the cosine similarity of derivative and base.

The variable of interest which was used as predictor in the beta regression models, as well as the control variables, are listed in Table 7.4. The first three variables are the same as in the study with the *fastText* implementation in Chapter 6, RELATIVE FREQUENCY, BASE POLYSEMY and WORD CLASS OF BASE. A more detailed description of these three variables is to be found in Chapter 6, Section 6.1.2. Two additional control variables were added to the prediction of the beta regression models, namely, CO-ACTIVATION and NEIGHBORHOOD DENSITY, which are measures extracted from the LDL implementation.

The variable CO-ACTIVATION is the square root of the sum of the squared values of a given word's predicted vector, i.e., its Euclidean distance from the origin. It measures the semantic co-activation, i.e., how much other semantic material is activated in the mental lexicon when activating the reading of the derivative. A higher CO-ACTIVATION is expected to lead to a higher cosine similarity of derivative and base as the possibility of the co-activation of the base is higher.

The variable NEIGHBORHOOD DENSITY is computed by the correlation values of a word's predicted semantic vector $\hat{v}$ and its eight nearest neighbors' semantic vectors $s_{n1}...s_{n8}$ are taken into consideration. The mean of these eight correlation values describes density, i.e., how dense the semantic neighborhood of the derivative is. The closer the semantics of other words in the neighborhood is, the higher is the density. A higher neighborhood density is expected to decrease the cosine similarity of derivative and base because the denser the neighborhood, the more other words have nearly the same semantics (cf. Chuang et al. 2021; Schmitz et al. 2021; Stein & Plag 2021; Schmitz 2022; Stein 2023).

TABLE 7.4: Dependent variable, variable of interest, control variables and their expected effects.

dependent variable	expectation
COSINE SIMILARITY	The cosine similarity of derivatives and bases indicates how similar one derivative is to its base.
variable of interest	**expectation**
WORD CLASS OF BASE	Verbal bases are more similar to their derivatives due to their denotation of eventualities.
control variables	
RELATIVE FREQUENCY OF BASE/DERIVATIVE	A higher relative frequency goes together with a higher segmentability and more transparency and thus an increase of cosine similarity.
BASE POLYSEMY	A higher polysemy of the base goes together with a decrease of cosine similarity as the derivative is (often) based on one reading of the base.
CO-ACTIVATION	The activation of more semantic material goes together with an increase of cosine similarity.
NEIGHBORHOOD DENSITY	A denser neighborhood has more other words which have similar semantics and goes together with a decrease of cosine similarity.

A beta regression model for each suffix, *-ee*, *-ment*, and *-ation*, was fitted. The beta regression formula for the individual suffixes may include precision phi components (cf. Ferrari & Cribari-Neto 2004; Plag et al. 2017). The precision phi component is interpreted as follows: "a low precision coefficient means that the beta regression model estimates the values of this predictor to be more dispersed around the coefficient's mean than in the case of a predictor with a high precision coeffic[i]ent." (Plag et al. 2017: 202). The precision component of a variable was included if the model with the precision reached a significantly lower AIC value (i.e., at least a difference of 2 points) and a higher log-likelihood according to standard procedures (cf. Baayen 2008: ch. 6; Winter 2019: ch. 15).

All variables were tested for problematic correlation coefficients, i.e., $|rho| \geq 0.5$. Correlations might bring collinearity into the model which might in turn lead to unreliable results (Tomaschek et al. 2018). To check for collinearity issues in the fitted models, their VIF-values (variance inflation factors) were computed (Zuur et al. 2010; Fox & Weisberg 2019). Overall, multiple beta regression models were fitted including and excluding variables and phi components. More precisely, in order to find the best model for each suffix, I fitted the models with step-wise exclusion of first phi components and then control variables to determine the best model fit. I chose this procedure of testing which variables and phi components to retain to ensure that the reported models contain as much information as needed but not more than required. The final models reported in this dissertation have the best fit regarding the log-likelihood and the AIC values as well as no issues of collinearity as indicated by the VIF-values (cf. Plag et al. 2017; Zuur et al. 2010; Fox & Weisberg 2019). The detailed formulas for the best fitting beta regression models are reported in the pertinent subsections.

7.1.2.2 t-SNE and LDA analyses

In order to see whether the vectors for the derivatives themselves show patterns, for example, whether the derivatives show distinct clusters for denominal and deverbal nominalizations, t-distributed stochastic neighborhood embedding analyses (t-SNE, for more information on t-SNE analyses, see, e.g., van der Maaten & Hinton 2008; van der Maaten 2014; Krijthe 2015; Shafaei-Bajestan et al. 2022b,a; Schmitz et al. 2023; Schäfer 2025) for the estimated semantic vectors of the derivatives is performed. A t-SNE analysis reduces the dimensions of the vectors, 5487 dimensions in the LDL implementation in this dissertation, to fewer dimensions. For the t-SNE analysis in this dissertation, two dimensions are retained for the analysis (Rtsne package, Krijthe 2015; gdsm package, Schmitz & Schneider 2022). The choice for two dimensions was made due to two reasons: First, two dimensions are the standard parameter in the pertinent literature. Second, the visualization of a two dimensional space is far more easy to interpret than the output of a three (or more) dimensional space.

The output of the analyses shows the clustering behavior of all derivatives in the data set. The data is sorted into denominal and deverbal derivatives. A t-SNE analysis for the data of each individual suffix, *-ee*, *-ment*, and *-ation*, is performed. The visible clusters of the analyses are then inspected further.

To ensure whether the visible clusters of the t-SNE analyses are meaningful, a linear discriminant analysis (LDA, for more information on LDA, see, Xanthopoulos et al. 2013; Trendafilov & Gallo 2021; Shafaei-Bajestan et al. 2022a; Schäfer 2025) is performed. The LDA in this dissertation is based on the two retained t-SNE dimensions. The predictions of the LDAs indicate whether a derivative in the data set is denominal or deverbal. If clusters are visible in the t-SNE visualization, the prediction of the LDA should resemble these clusters. More precisely, LDA predicts whether there is a difference between denominal and deverbal derivatives. Taking the results of the t-SNE visualization as well as the LDA, one may or may not assume the existence of differences between deverbal and denominal derivatives.

First, the results of the beta regression analyses for the three suffixes will be reported in Section 7.2.1. Section 7.2.2 will show the results from the t-SNE and LDA analyses.

7.2 Results

7.2.1 Beta regression analyses

First, the cosine similarities for all three suffixes are summarized. The density of the raw cosine similarities for all suffixes ($n_{\text{denominal}} = 141$, $n_{\text{deverbal}} = 656$) sorted by the word class of the base is shown in the left panels in Figure 7.2. The cosine similarities of denominal derivatives and their nominal bases are in the upper panel, and the density of the cosine similarities of deverbal derivatives and their verbal bases are displayed in the lower panel. The right panels illustrate the cosine similarities in boxplots. The data for denominal *-ation* shows the most variance, but the data for all three suffixes is widely spread in the denominal data, less in the deverbal data.

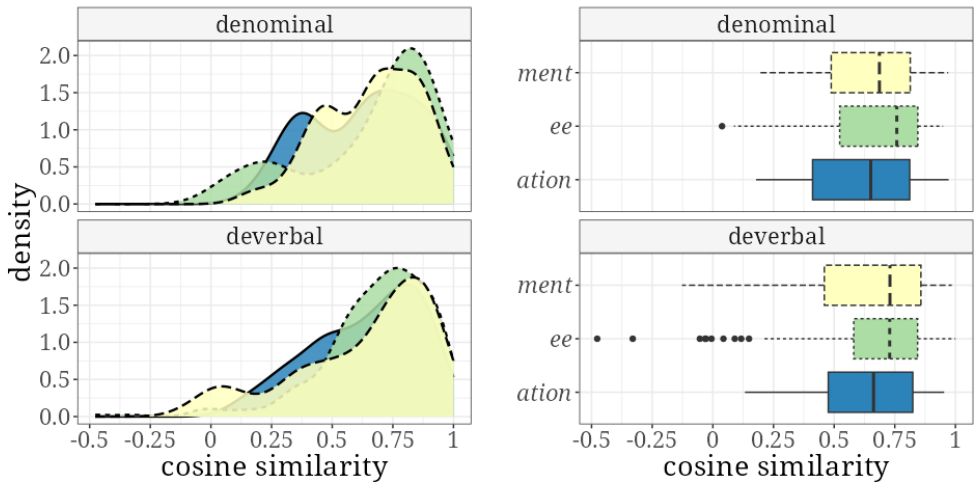

FIGURE 7.2: Density of the cosine similarities of denominal derivatives and nominal bases at the top, and deverbal derivatives and verbal bases at the bottom. The color coding by suffix: -*ation* in blue, -*ee* in green and -*ment* in yellow. The suffix -*ation* is represented by a solid line, -*ee* by a densely dashed line, and -*ment* by a dashed line.

The denominal sets in the upper panel of both the density and the box plot do not look strikingly different. As two comparisons are made per suffix for denominal and deverbal data, i.e., denominal and deverbal -*ee* is compared to -*ation* and -*ment*, -*ation* to -*ee* and -*ment*, and -*ment* to -*ee* and -*ation*, a Bonferroni correction of the comparisons is needed for the denominal and the deverbal data (cf. Baayen 2008: 114; Winter 2019: 176). The p-values of the Wilcoxon-tests are divided by 2 as two comparisons for each suffix are made. A look at the p-values from a Bonferroni corrected Wilcoxon-test are given in Table 7.5 and shows that the differences between the data sets is, in fact, not significant.

TABLE 7.5: Comparison of denominal sets by p-values retrieved from Wilcoxon-tests.

	-*ee*	-*ment*
-*ation*	0.18	0.30
-*ee*		0.19

The lower panel shows the deverbal sets. The density of the data with the three suffixes does not look strikingly different. A look at the p-values from a Bonferroni corrected Wilcoxon-test are given in Table 7.6 and shows that the differences between the data sets is, in fact, not significant.

TABLE 7.6: Comparison of deverbal sets by p-values retrieved from Wilcoxon-tests.

	-ee	*-ment*
-ation	0.06	0.26
-ee		0.16

7.2.1.1 Results *-ee*

The density of the raw cosine similarities for the data in *-ee* ($n_{\text{denominal}} = 46$, $n_{\text{deverbal}} = 311$) sorted by the word class of the base is shown in Figure 7.3. The orange area represents the density for the cosine similarities of denominal derivatives in *-ee* and their nominal bases. The purple area depicts the density of the cosine similarities of deverbal derivatives in *-ee* and their verbal bases. Both curves overlap to a large extent. The orange curve for the denominal pairs has more data in the lower range but also more data in the higher range of the cosine similarities of derivatives and bases. A diptest ($p = 0.62$, diptest package, Maechler 2020) shows that the curve for the denominal data is not bimodally distributed. The denominal pairs have a median cosine similarity of about 0.75 and the deverbal pairs of 0.7. The cosine similarities for denominal derivatives and their nominal bases show a larger box, i.e., more variation compared to the deverbal comparison. The difference in the means between denominal and deverbal pairs is not significant (Wilcoxon test, $p = 0.99$).

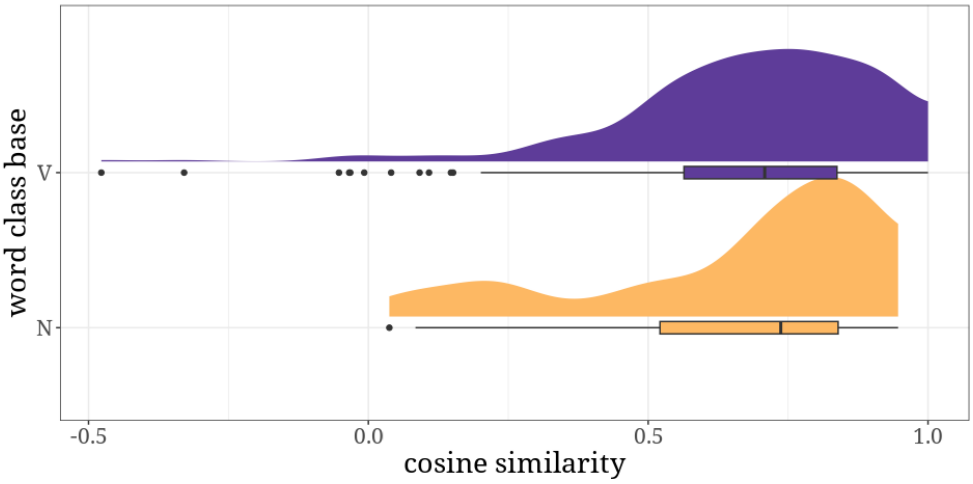

FIGURE 7.3: Density of the cosine similarities of derivatives and bases for the suffix *-ee*. The cosine similarities of the deverbal data are given in purple and the cosine similarities for the denominal data are given in orange.

Before a beta regression model was fitted, the correlation coefficients of independent variables were checked (SfL package, Schmitz & Esser 2021) and no problematic values were found. Several different models were fitted by the procedure explained above until a model with the best performance was found. The best fit was determined via AIC-value and log-likelihood of the models (cf. Baayen 2008: ch. 6; Winter 2019: ch. 15). The VIF-values of the final model confirm that the final model has no collinearity issues (cf. Zuur et al. 2010; Fox & Weisberg 2019). The formula of the final model is given in 7.1. The model includes no precision specifications as the inclusion of the precision estimates did not increase the fit of the model.

$$\begin{aligned}&\text{COSINE SIMILARITY} \sim \\ &\qquad \text{BASE POLYSEMY} + \text{RELATIVE FREQUENCY} + \\ &\qquad \text{WORD CLASS OF BASE} + \text{CO-ACTIVATION} + \text{NEIGHBORHOOD DENSITY} \end{aligned} \tag{7.1}$$

The estimates of the final model and their p-values are given in Table 7.7. The variables RELATIVE FREQUENCY and NEIGHBORHOOD DENSITY do not reach significance. The variables WORD CLASS OF BASE, BASE POLYSEMY and CO-ACTIVATION reach significance in the model.

TABLE 7.7: Fixed-effect coefficients and p-values as computed by the beta regression model for the *-ee* data fitted to the cosine similarity values. Significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.19	0.40	-0.47	0.64	
WORD CLASS OF BASE V	0.32	0.15	2.07	0.04	*
RELATIVE FREQUENCY	0.00	0.02	0.01	0.99	
BASE POLYSEMY	-0.02	0.01	-3.10	0.00	**
CO-ACTIVATION	25.68	3.48	7.38	1.54e-13	***
NEIGHBORHOOD DENSITY	-0.37	0.55	-0.67	0.50	

Figure 7.4 shows the significant effect of the variable WORD CLASS OF BASE. The effect is as expected as the cosine similarities of deverbal derivatives and their bases are higher than for denominal derivatives. The effects of the control variables are as expected: a higher BASE POLYSEMY goes together with a lower cosine similarity, a higher CO-ACTIVATION goes together with a higher cosine similarity.

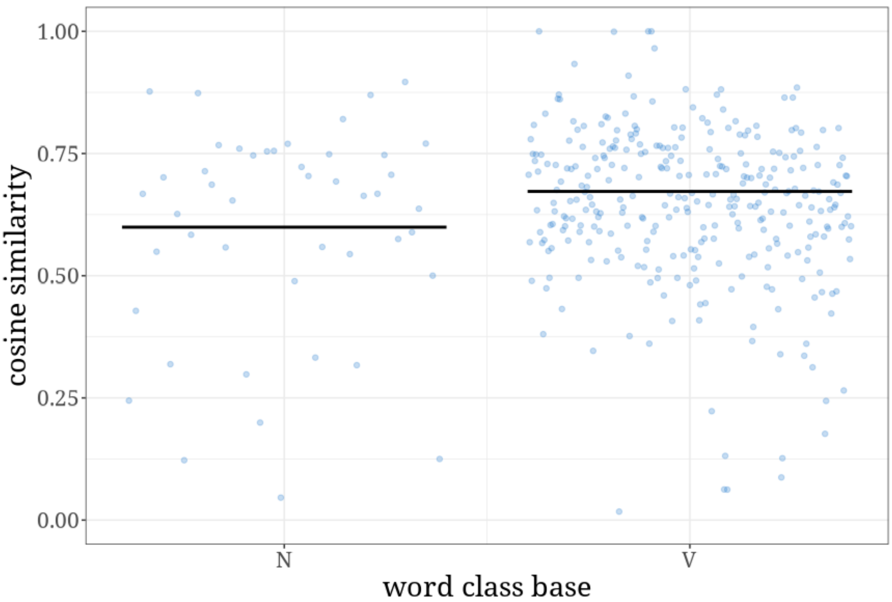

FIGURE 7.4: Significant effect of WORD CLASS OF BASE in the beta regression model with the suffix *-ee*.

The effect of the variable of interest, WORD CLASS OF BASE, behaves as expected. More precisely, due to the eventive nature of verbs (e.g., Van Valin & LaPolla 1997; Haspelmath 2001; Van Valin & LaPolla 2002; Szabó

2015; Moltmann 2019), the higher cosine similarity for deverbal nominalizations and their verbal bases is expected. The found effect is in line with the expectation that deverbal nominalizations are more clearly interpretable compared to denominal nominalizations.

7.2.1.2 Results *-ment*

The density of the raw cosine similarities for the data in *-ment* ($n_{\text{denominal}} = 29$, $n_{\text{deverbal}} = 273$) sorted by the word class of the base is shown in Figure 7.5. The orange area shows the cosine similarity of denominal derivatives with *-ment* and their nominal bases and the purple area illustrates the cosine similarity of deverbal derivatives with *-ment* and their verbal bases. The distribution of cosine similarities of the deverbal derivatives in orange starts earlier as compared to the denominal pairs. At the same time, the deverbal derivatives and their bases are spread wider. The cosine similarity of the denominal derivatives and their bases are differently and more densely distributed. The denominal pairs have a median cosine similarity of about 0.7 and the deverbal pairs of about 0.75. A Wilcoxon test shows that the difference in the means between the two types of word classes of the base is not significant (Wilcoxon test, $p = 0.78$).

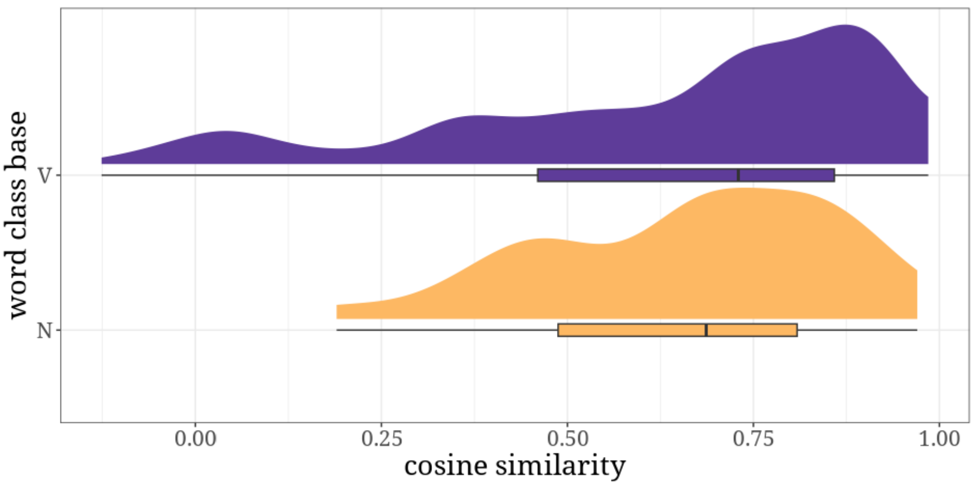

Figure 7.5: Density of the cosine similarities of derivatives and bases for the suffix *-ment*. The cosine similarities of the deverbal data are given in purple and the cosine similarities for the denominal data are given in orange.

The data were then first checked for correlations of independent variables. The correlation test (SfL package, Schmitz & Esser 2021) showed a correlation of the variables RELATIVE FREQUENCY and WORD CLASS OF BASE ($rho = -0.5$). This correlation is exactly at the threshold of when a correlation is seen to be problematic ($|rho| \geq 0.5$, see Section 7.1.2). However, the VIF-values of the final model indicate that the correlation does not result in a collineartiy issue in the model (cf. Zuur et al. 2010; Fox & Weisberg 2019). Thus, the variables were included in the model without any transformation. The best fit was determined via AIC-value and log-likelihood of the models (cf. Baayen 2008: ch. 6; Winter 2019: ch. 15). The formula of the final model is given in 7.2. The model includes precision specifications of all variables except for CO-ACTIVATION (cf. Ferrari & Cribari-Neto 2004; Plag et al. 2017).

$$\begin{aligned} &\text{COSINE SIMILARITY} \sim \\ &\qquad \text{BASE POLYSEMY} + \text{RELATIVE FREQUENCY} + \\ &\text{WORD CLASS OF BASE} + \text{CO-ACTIVATION} + \text{NEIGHBORHOOD DENSITY} \quad | \\ &\qquad \text{BASE POLYSEMY} + \text{RELATIVE FREQUENCY} + \\ &\qquad \text{WORD CLASS OF BASE} + \text{NEIGHBORHOOD DENSITY} \end{aligned} \tag{7.2}$$

The estimates of the final beta regression model and their p-values are given in Table 7.8. The variables BASE POLYSEMY, CO-ACTIVATION, and NEIGHBORHOOD DENSITY, reach significance in the model.

TABLE 7.8: Fixed-effect coefficients and p-values as computed by the beta regression model for the *-ment* data fitted to the cosine similarity values. Significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	1.44	0.42	3.44	0.00	***
WORD CLASS OF BASE V	-0.22	0.26	-0.86	0.39	
RELATIVE FREQUENCY	-0.03	0.03	-0.98	0.33	
BASE POLYSEMY	0.10	0.03	3.79	0.00	***
CO-ACTIVATION	16.57	3.74	4.43	9.46e-06	***
NEIGHBORHOOD DENSITY	-2.65	0.55	-4.78	1.77e-06	***
Phi coefficients	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	2.06	0.55	3.78	0.00	***
WORD CLASS OF BASE V	-0.89	0.36	-2.49	0.01	*
RELATIVE FREQUENCY	-0.01	0.03	-0.52	0.60	
BASE POLYSEMY	0.10	0.03	3.69	0.00	***
NEIGHBORHOOD DENSITY	-0.76	0.62	-1.22	0.22	

The variable WORD CLASS OF BASE does not show a significant effect. The effects of two of the control variables are as expected: A higher CO-ACTIVATION goes together with an increase of cosine similarities, and a higher NEIGHBORHOOD DENSITY goes together with a decrease of cosine similarity. The effect of BASE POLYSEMY is unexpected as a higher value goes together with an increase of cosine similarities.

The unexpected finding of the non-significant effect of the variable of interest, WORD CLASS OF BASE, is interesting for two reasons. First, the expectation that deverbal derivatives are more similar to their bases compared to denominal derivatives originates from the underlying assumption that eventuality-related nominalizations operate on an eventuality denoted by the base. As verbs are ontologically eventive and nouns are not (e.g., Haspelmath 2001; Van Valin & LaPolla 2002; Szabó 2015; Moltmann 2019), the reading of a deverbal derivative should be more clearly observable, thus, depicting a higher similarity of base and derivative. Second, the non-existence of a significant difference between denominal and deverbal derivatives points toward the assumption that the word class of a base is not the most important factor for the word-formation process in general. Plag (2004) already found that semantic factors seem to be more important for word-formation processes than syntactic categories, such as nouns and verbs. The present findings for derivatives in *-ment* point towards this idea.

7.2.1.3 Results *-ation*

The density of the raw cosine similarities for the data in *-ation* ($n_{\text{denominal}} = 66$, $n_{\text{deverbal}} = 72$) sorted by the word class of the base is shown in Figure 7.6. The orange area shows the density of the cosine similarities for denominal derivatives and their bases. The purple area illustrates the density of the cosine similarities for deverbal derivatives and their verbal bases. A diptest ($p = 0.07$, diptest package, Maechler 2020) for the orange area shows that the denominal data is not bimodally distributed. The cosine similarities for denominal and deverbal derivatives and their bases are similarly distributed. The median of the denominal pairs is, as well as the median of the deverbal pairs, about 0.65. The difference in the means is not significant (Wilcoxon test, $p = 0.71$).

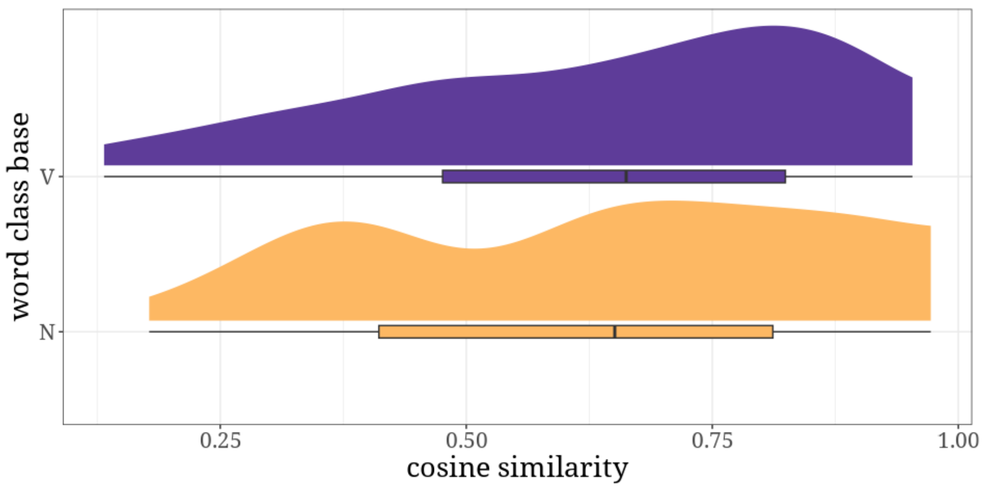

FIGURE 7.6: Density of the cosine similarities of derivatives and bases for the suffix *-ation*. The cosine similarities of the deverbal data are given in purple and the cosine similarities for the denominal data are given in orange.

The correlation test (SfL package, Schmitz & Esser 2021) showed a problematic correlation of three variables, namely, BASE POLYSEMY, RELATIVE FREQUENCY and WORD CLASS OF BASE, in the data for *-ation*. The strengths of the correlations are given in Table 7.9.

TABLE 7.9: Correlations observed in the *-ation* data. The threshold for problematic correlation coefficient values was set at $|rho| \geq 0.5$.

	BASE POLYSEMY	RELATIVE FREQUENCY
WORD CLASS OF BASE	$rho = -0.27$	$rho = -0.65$
BASE POLYSEMY		$rho = 0.58$

The beta regression model was fitted following the procedure explained above. The VIF-values indicate that there is no collinearity issue in the model although some variables correlate (cf. Zuur et al. 2010; Fox & Weisberg 2019). Thus, the variables were included in the model without any transformation. The best fit was determined using the AIC-value and the log-likelihood of the models (cf. Baayen 2008: ch. 6; Winter 2019: ch. 15). The formula of the best beta regression model is given in 7.3. No precision estimates are included as the model does not reach a better fit with their inclusion.

$$\text{COSINE SIMILARITY} \sim \text{BASE POLYSEMY} + \text{RELATIVE FREQUENCY} + \text{WORD CLASS OF BASE} + \text{CO-ACTIVATION} + \text{NEIGHBORHOOD DENSITY} \quad (7.3)$$

The estimates of the final beta regression model and their p-values are given in Table 7.10. The variables RELATIVE FREQUENCY, CO-ACTIVATION and NEIGHBORHOOD DENSITY reach significance in the model.

TABLE 7.10: Fixed-effect coefficients and p-values as computed by the beta regression model for the *-ation* data fitted to the cosine similarity values. Significance codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	2.25	0.38	5.93	3.10e-09	***
WORD CLASS OF BASE V	-0.14	0.18	-0.79	0.43	
RELATIVE FREQUENCY	-0.08	0.03	-2.65	0.01	**
BASE POLYSEMY	0.06	0.05	1.21	0.23	
CO-ACTIVATION	22.45	3.69	6.09	1.14e-09	***
NEIGHBORHOOD DENSITY	-4.44	0.64	-6.96	3.38e-12	***

The variable of interest, WORD CLASS OF BASE, shows no significant effect. The effects of the control variables CO-ACTIVATION and NEIGHBORHOOD DENSITY are as expected: A higher CO-ACTIVATION goes together with an increase of cosine similarity and a higher NEIGHBORHOOD DENSITY with a decrease. The variable RELATIVE FREQUENCY behaves contra the expectation as a higher value goes together with a decrease of cosine similarity.

Similar to the findings for nominalizations with the suffix *-ment*, the non-existent effect of the variable WORD CLASS OF BASE for nominalizations in *-ation* raises two important points. First, the assumption based on ontological considerations that deverbal derivatives should be more similar to their bases than denominal derivatives to their denominal bases cannot be confirmed. The idea of this assumption is that the interpretation of eventuality-related nominalizations is based on eventive structures of the base. Second, the observations for nominalizations in *-ation* contribute to the idea that the word class of the base is not the most important factor for a word-formation process but rather their semantics.

7.2.2 t-SNE and LDA analyses

In order to see whether the vectors for the derivatives themselves show a pattern, a t-distributed stochastic neighbor embedding (t-SNE) analysis was performed. Such an analysis reduces the dimensions of the vectors, 5487 dimensions in this LDL implementation, to fewer dimensions. Two dimensions are retained for the analysis (Rtsne package, Krijthe 2015; gdsm package, Schmitz & Schneider 2022). This analysis was included to investigate if there is a pattern to be found which distinguishes denominal and deverbal derivatives in general. In order to see whether the found clusters have significance, an LDA was performed. The findings for the data of the distribution for the derivatives with each suffix suggest that the deverbal/denominal distinction does not lead to distinctive patterns.

7.2.2.1 Results *-ee*

Figure 7.7 illustrates the distribution of the derivatives for *-ee* in a two-dimensional semantic t-SNE space. No cluster is visible. Denominal and deverbal derivatives are spread all over the two-dimensional space.

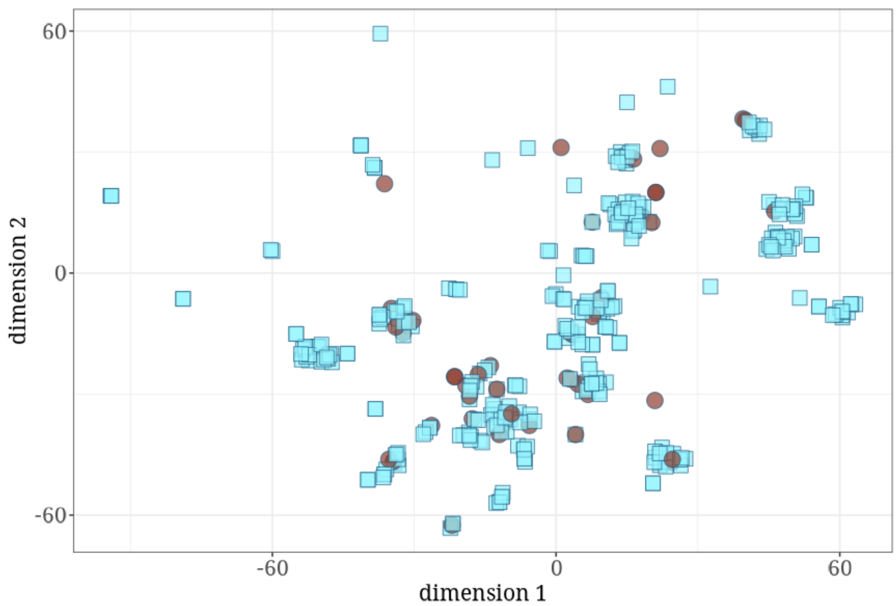

FIGURE 7.7: t-SNE for the suffix *-ee*. Red circles are denominal derivatives and blue squares are deverbal derivatives.

The first step of the LDA is depicted in Figure 7.8. The t-SNE dimensions are used for the prediction. The histograms for nouns and verbs look strikingly similar, and most of the data overlap. This is not surprising as the *t*-SNE analysis does not show any clusters.

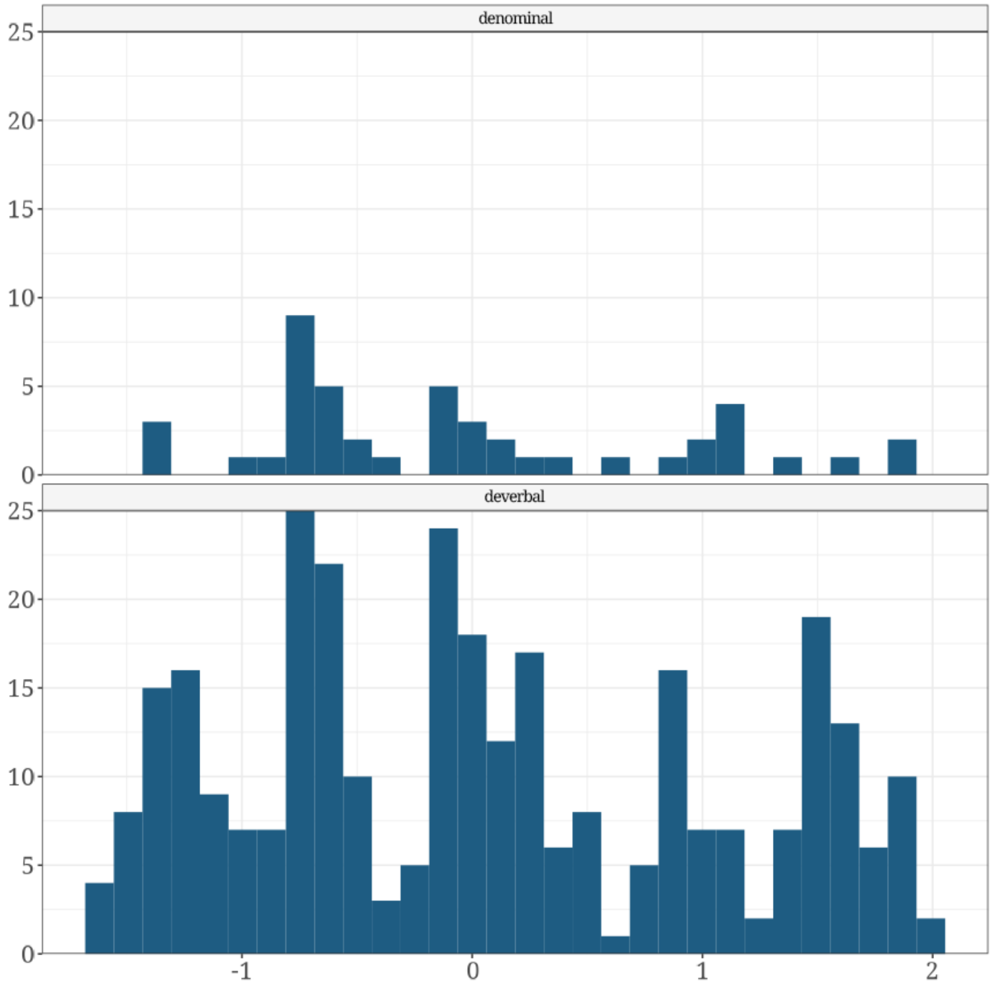

FIGURE 7.8: LDA histogram for the suffix *-ee*, displaying the distribution of the data.

Table 7.11 depicts the prediction of the LDA. All words are predicted to be deverbal. The wrong prediction of denominal derivatives as being deverbal indicates that the LDA cannot discriminate between denominal and deverbal derivatives. This is contra the results from the beta regression analysis for *-ee* which showed a significant effect for the two word classes. The results of the *t*-SNE analysis and the LDA find no effect of the variable WORD CLASS OF BASE when looking at the derivatives only. The fact that there is no clear distinction between denominal and deverbal derivatives shows that the word-formation process is not dependent on the word class of the base. Contrarily, the findings speak in favor of an account of word-formation processes which does not have the syntactic category of the base as the main factor, but rather an account in which the semantics of base and derivative are incorporated.

Table 7.11: LDA prediction for *-ee*.

	Reference	
Prediction	N	V
N	0	46
V	0	311
Accuracy	0.8711	
95% CI	(0.8319 , 0.9041)	
No Information Rate	1	
P-Value [Acc > NIR]	1	
Kappa	0	
Mcnemar's Test P-Value	3.247e-11	
Sensitivity	NA	
Specificity	0.8711	
Pos Pred Value	NA	
Neg Pred Value	NA	
Prevalence	0.0000	
Detection Rate	0.0000	
Detection Prevalence	0.1289	
Balanced Accuracy	NA	
'Positive' Class	N	

7.2.2.2 Results *-ment*

Figure 7.9 illustrates the t-SNE distribution of the *-ment* nominalizations in a two dimensional semantic space. A small cluster in the right bottom corner of the plot is visible. It contains denominal and deverbal derivatives. The rest of the data is spread over the two dimensions, with deverbal *-ment*-derivatives spread further to the right, but without a visible cluster for deverbal or denominal derivatives.

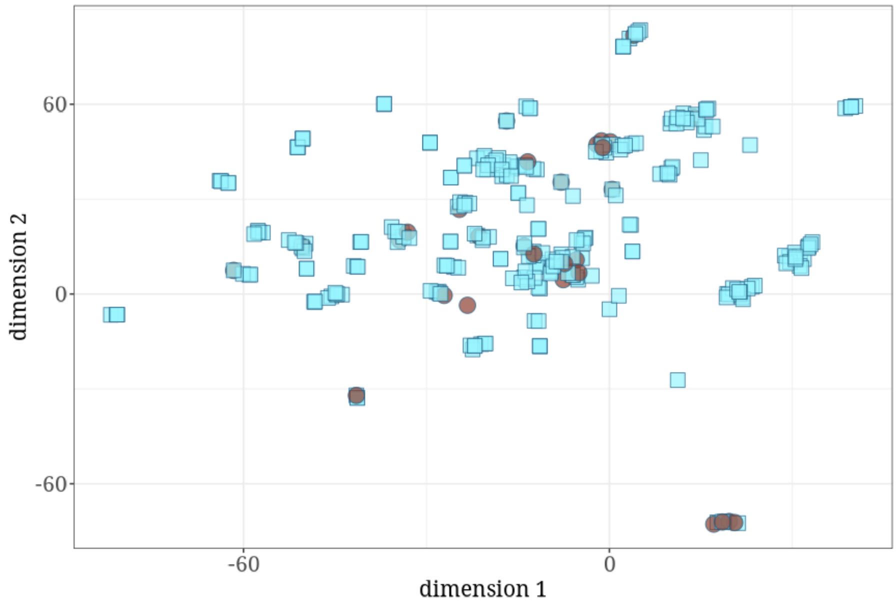

FIGURE 7.9: t-SNE for the suffix *-ment*. Red circles are denominal derivatives and blue squares are deverbal derivatives.

Figure 7.10 shows the words in the visible cluster. The cluster in the right bottom corner contains psych nouns, denominal and deverbal ones. For convenience, some examples are given in (1). The cluster consists of an equal number of denominal and deverbal derivatives.

(1) a. denominal *-ment*
confusionment, illusionment, cantonment
b. deverbal *-ment*
abandonment, appotionment, disillusionment

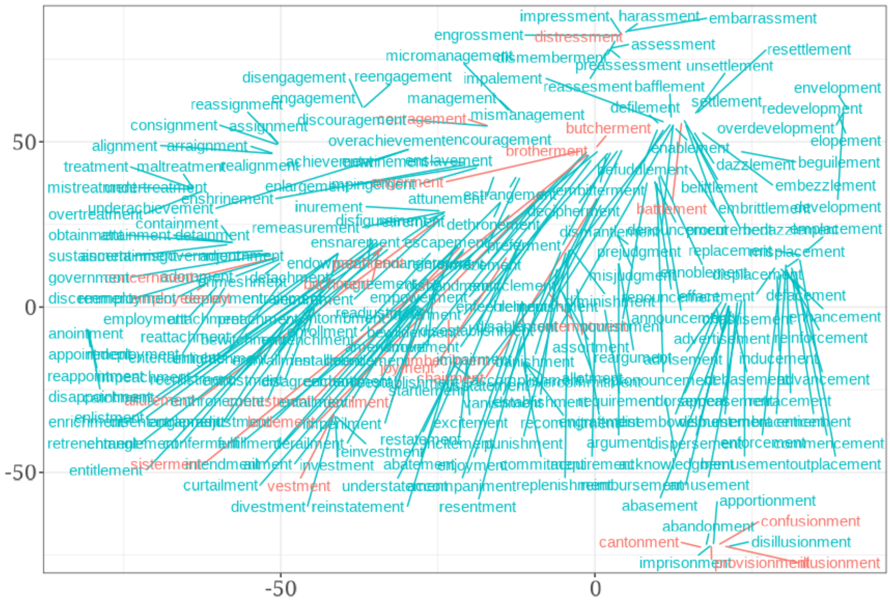

FIGURE 7.10: t-SNE for the suffix *-ment*. Red words are denominal derivatives and blue words are deverbal derivatives.

The histogram of the data in Figure 7.11 reinforces the results from the t-SNE analysis as the denominal and deverbal data visibly overlap. There is no clear distinction of denominal and deverbal derivatives.

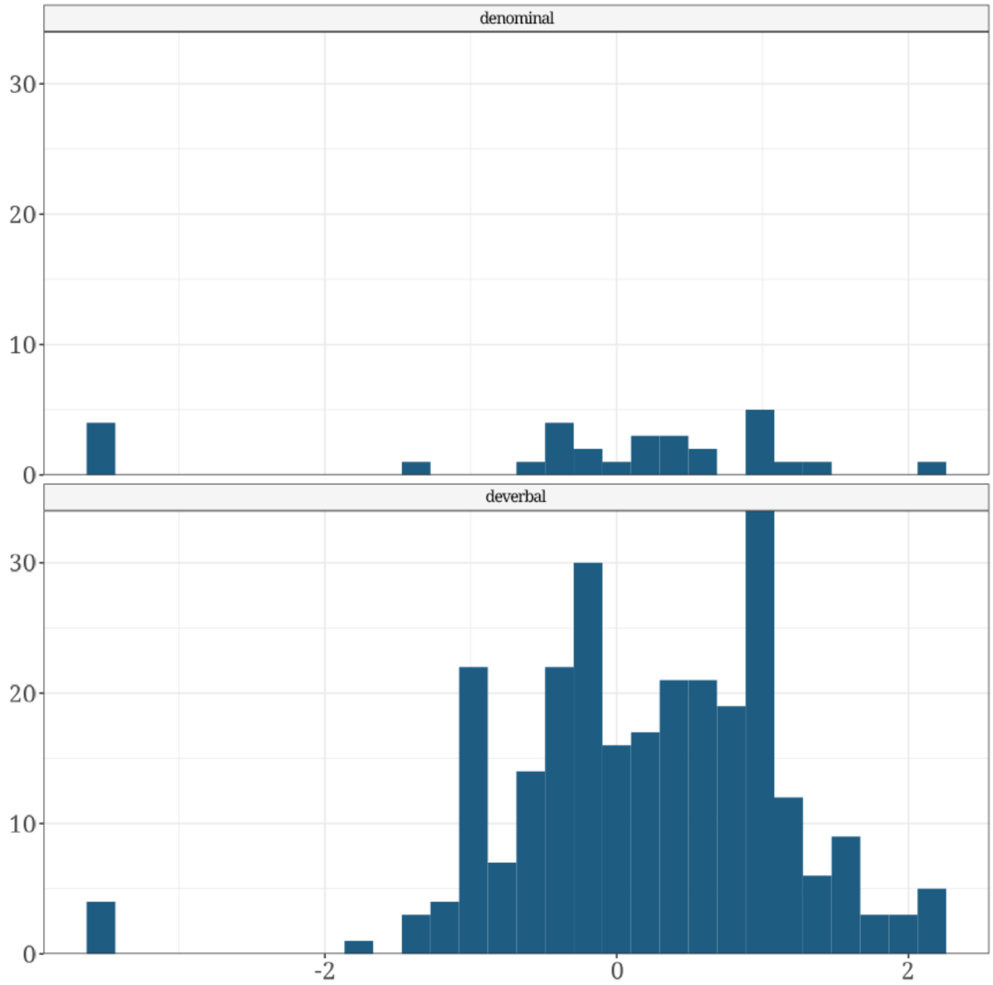

FIGURE 7.11: LDA histogram for the suffix *-ment*, displaying the distribution of the data.

Table 7.12 depicts the prediction of the LDA. All derivatives are predicted as deverbal. The absence of clear visible clusters for denominal and deverbal derivatives in the *t*-SNE plot as well as the prediction from the LDA show that the word class of the base is not the only important factor for a word-formation process. Broadly speaking, the syntactic category of the base cannot account for a complete picture of the word-formation process for eventuality-related nominalizations.

TABLE 7.12: LDA prediction for *-ment*.

	Reference	
Prediction	N	V
N	0	29
V	0	273
Accuracy	0.904	
95% CI	(0.865, 0.9347)	
No Information Rate	1	
P-Value [Acc > NIR]	1	
Kappa	0	
Mcnemar's Test P-Value	1.999e-07	
Sensitivity	NA	
Specificity	0.90397	
Pos Pred Value	NA	
Neg Pred Value	NA	
Prevalence	0.00000	
Detection Rate	0.00000	
Detection Prevalence	0.09603	
Balanced Accuracy	NA	
'Positive' Class	N	

7.2.2.3 Results *-ation*

Figure 7.12 shows the *t*-SNE distribution of the nominalizations with *-ation* in a two dimensional semantic space. Two clusters are visible: one denominal cluster in the top and one deverbal in the bottom left corner of the plot. A closer investigation of the two visible clusters shows that the denominal cloud consists of chemical *-ation*-derivatives with a nominal base. Similarly, the deverbal cloud consists of chemical *-ation*-derivatives with a verbal base.

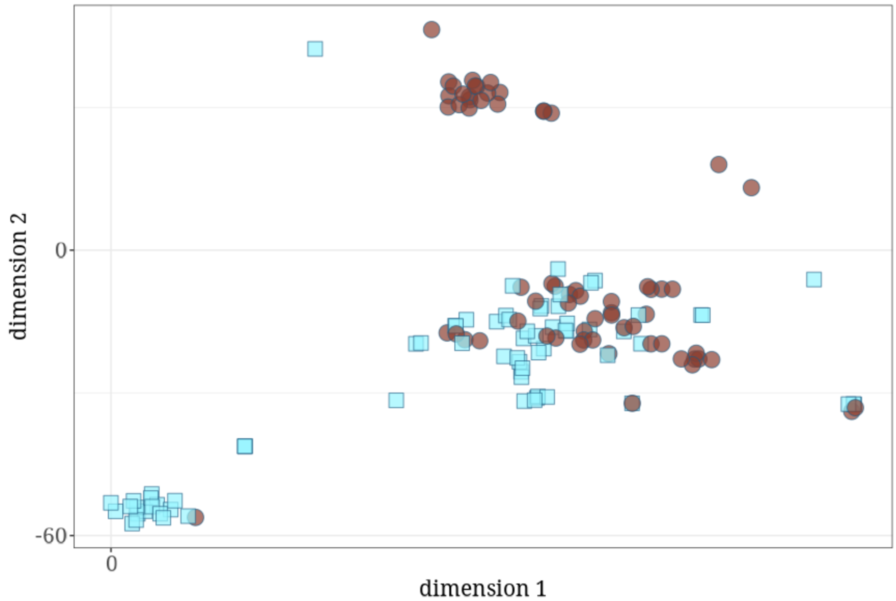

FIGURE 7.12: *t*-SNE for the suffix *-ation*. Red circles are denominal derivatives and blue squares are deverbal derivatives.

Figure 7.13 shows the words in the visible clusters. For convenience, some examples are given in (2).

(2) a. denominal *-ation*
carboxylation, glycosylation, ozonation
b. deverbal *-ation*
akalization, exudation, silanilation

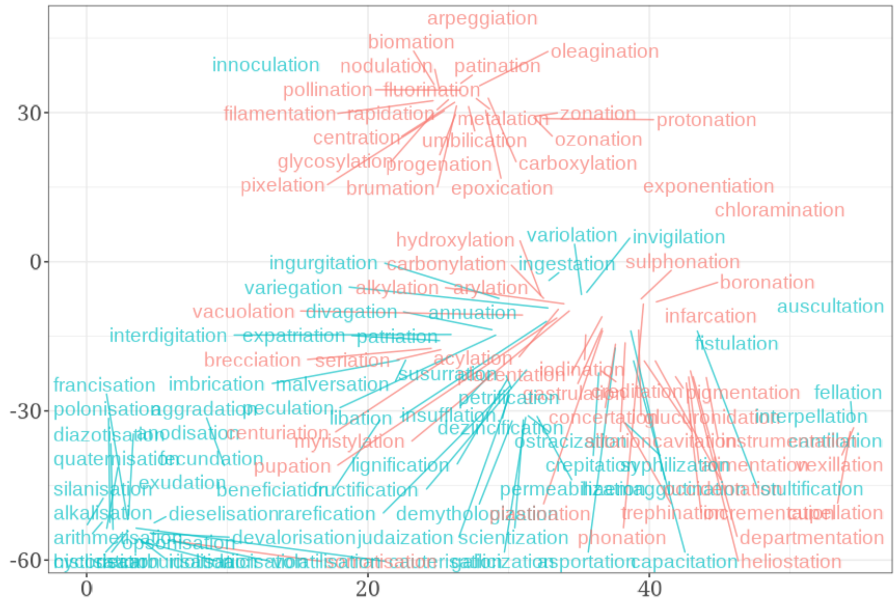

FIGURE 7.13: t-SNE for the suffix -*ation*. Red words are denominal derivatives and blue words are deverbal derivatives.

The histogram of the nominalizations in -*ation* in the LDA also shows that denominal and deverbal derivatives are distinct from each other to some extent. The denominal derivatives in the upper panel have some data on the left side of the x-axis whereas the deverbal derivatives show a few data points on the right side of the x-axis.

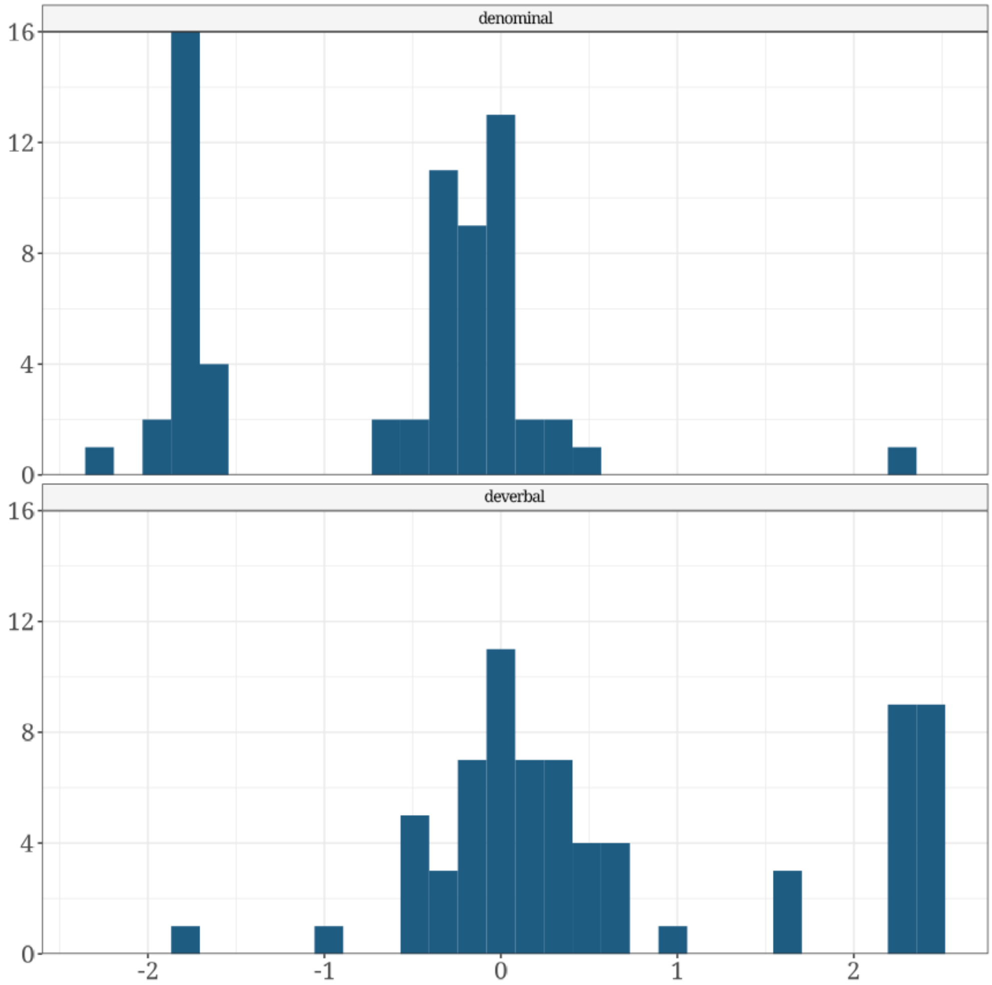

FIGURE 7.14: LDA histogram for the suffix *-ation*, displaying the distribution of the data.

Table 7.13 shows the LDA prediction. It predicted denominal and deverbal derivatives, contrary to the analysis of the data of the other two suffixes which predicted all derivatives to be deverbal.

TABLE 7.13: LDA prediction for *-ation* .

	Reference	
Prediction	N	V
N	48	18
V	17	55
Accuracy	0.7464	
95% CI	(0.6653, 0.8165)	
No Information Rate	0.529	
P-Value [Acc > NIR]	1.263e-07	
Kappa	0.4915	
Mcnemar's Test P-Value	1	
Sensitivity	0.7385	
Specificity	0.7534	
Pos Pred Value	0.7273	
Neg Pred Value	0.7639	
Prevalence	0.4710	
Detection Rate	0.3478	
Detection Prevalence	0.4783	
Balanced Accuracy	0.7459	
'Positive' Class	N	

Figure 7.15 has two colors: The purple area indicates the range for the deverbal data and the orange area the range for denominal data. The verbs in the orange area as well as the nouns in the purple area are wrongly predicted derivatives by the model.

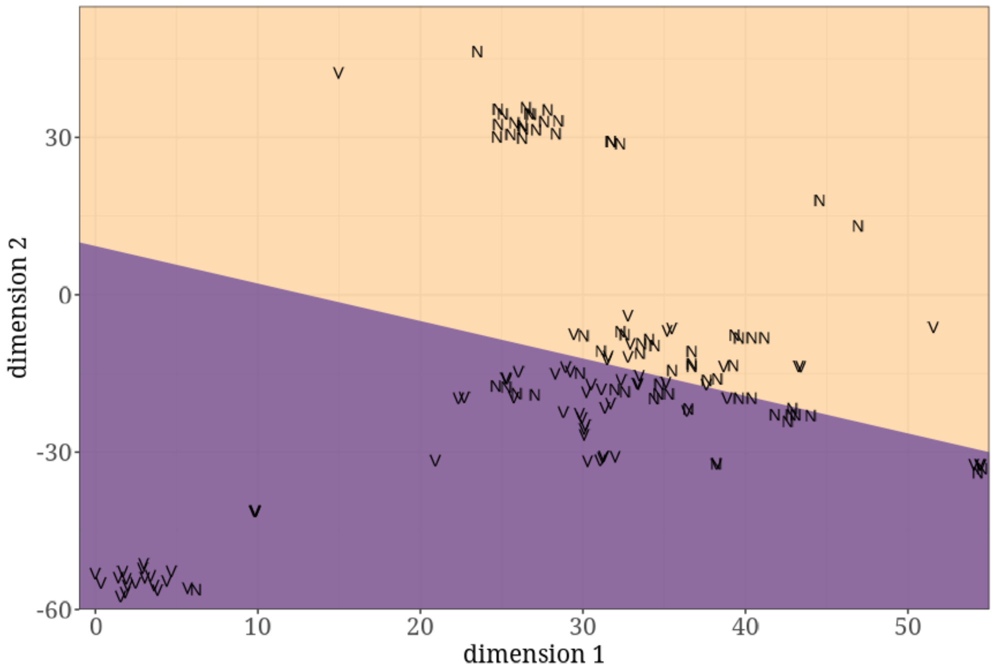

FIGURE 7.15: LDA prediction for the suffix *-ation*.

The results for eventuality-related nominalizations with the suffix *-ation* show that some clusters can arise. The *t*-SNE analysis as well as the LDA can distinguish between chemical denominal and deverbal derivatives. Nonetheless, no clear cluster or prediction for all denominal and deverbal derivatives is visible. The two distinguished clusters show a word class distinction but this distinction can possibly be better explained by semantics than by syntactic features. Chapter 4 shows that chemical substances are non-trivial in their semantic representation. Many of the chemical substances give rise to different readings. The two clusters might thus not only reflect a difference in the word class of the base, but also an important semantic distinction. For instance, it is unclear how the reading process/result/ornative are distributed in denominal and deverbal formations with chemical substances/chemical verbs.

7.3 Discussion

This chapter dealt with an LDL implementation of the data for three suffixes, *-ee*, *-ment*, and *-ation*. The data underwent a beta regression analysis as well as a t-SNE analysis and an LDA. First, the results of the beta regression will be discussed, followed by a discussion of the other two analyses. The variables included in the beta regression model and their expected effects are repeated in Table 7.14.

TABLE 7.14: Dependent variable, variable of interest, control variables and their expected effects.

dependent variable	expectation
COSINE SIMILARITY	The cosine similarity of derivatives and bases indicates how similar one derivative is to its base.
variable of interest	**expectation**
WORD CLASS OF BASE	Verbal bases are more similar to their derivatives due to their denotation of eventualities.
control variables	
RELATIVE FREQUENCY OF BASE/DERIVATIVE	A higher relative frequency goes together with a higher segmentability and more transparency and thus an increase of cosine similarity.
BASE POLYSEMY	A higher polysemy of the base goes together with a decrease of cosine similarity as the derivative is (often) based on one reading of the base.
CO-ACTIVATION	The activation of more semantic material goes together with an increase of cosine similarity.
NEIGHBORHOOD DENSITY	A denser neighborhood has more other words which have similar semantics and goes together with a decrease of cosine similarity.

Starting with the variable of interest WORD CLASS OF BASE: The variable reaches significance for the beta regression model with the suffix *-ee* only. In the data with the suffix *-ee*, the cosine similarity of derivative and base increases if the base is verbal. For the data with the suffixes *-ation* and *-ment*, the variable WORD CLASS OF BASE did not reach significance in the models. The effect found for the data in *-ee* is expected as the derivational process is assumed to be easier interpretable with verbs. This assumption is built on the ontology of verbs in general, as they denote eventualities and eventuality-related nominalizations operate on eventualities (see, e.g., Van Valin & LaPolla 1997; Haspelmath 2001; Szabó 2015; Moltmann 2019 on the ontology of words, and, e.g., Plag et al. 2018; Kawaletz 2023 and Chapter 4 on the eventuality-relatedness of nominalizations). The non-significant effects of the variable WORD CLASS OF BASE for the data with two of the three suffixes implies that the word class of the base itself is not a crucial factor for the interpretation of eventuality-related nominalizations.

The included control variables BASE POLYSEMY, RELATIVE FREQUENCY, CO-ACTIVATION, and NEIGHBORHOOD DENSITY showed different significant effects in the beta regression models. The overall finding is that the variable of interest WORD CLASS OF BASE is stable and not dependent on the outcomes of the control variables.

The *t*-SNE analysis showed only visible clusters for the data with the suffix *-ation*. The two clusters divide deverbal and denominal derivatives based on chemical substances from each other and from the other remaining derivatives. The *t*-SNE analyses for *-ee* and *-ment* do not show clear clusters. Thus, the findings of the *t*-SNE analyses as well as the findings of the LDAs for the suffixes show similar results: LDA only predicts a difference between with the denominal and deverbal *-ation* data. The LDAs of *-ee* and *-ment* wrongly predict every derivative to be deverbal. This finding is expected as already no visible clusters were to be found in the *t*-SNE analysis. The absence of clusters for denominal and deverbal data in general shows that the word class of the base seems to not play a crucial role for the derivational process with the suffixes under investigation, but the semantics of the base is the important factor for a successful word-formation process. This assumption is in line with the findings of the beta regression models where the variable WORD CLASS OF BASE is not a significant predictor for the cosine similarity of derivatives and bases as well as to the other findings in this thesis.

Chapter 8

Discussion: Computational methods

Two computational methods, *fastText* (Chapter 6) and linear discriminative learning (Chapter 7) were used to gain more insight into the semantics of eventuality-related nominalizations. Table 8.1 summarizes the outcome of the beta regression models for the variable of interest WORD CLASS OF BASE for both computational implementations. A check mark represents a significant effect and a cross the absence of a significant effect. The directions of the arrows indicate whether the found effect goes together with an increase (↑) or a decrease (↓) of the cosine similarity of derivatives and bases.

TABLE 8.1: Summary of effects of the variable WORD CLASS OF BASE from the beta regression models for the implementations, *fastText* and LDL. A check mark represents a significant effect and a cross the absence of a significant effect. The direction of the arrows indicate the direction of the found effects.

	fastText			LDL		
	-ation	*-ee*	*-ment*	*-ation*	*-ee*	*-ment*
WORD CLASS OF BASE	✗	✓↓ V	✓↑ V	✗	✓↑ V	✗

The expectation for the effect of WORD CLASS OF BASE was a higher cosine similarity of derivatives and bases when the base is a verb. This expectation is based on the fact that verbs denote eventualities (see, e.g., Van Valin & LaPolla 1997; Haspelmath 2001; Van Valin & LaPolla 2002; Szabó 2015; Moltmann 2019) and eventuality-related nominalizations denote eventive elements, for example, participants or eventualities. The results show that the variable behaves as expected for the data in *-ment* in the *fastText* implementation and the data in *-ee* in the LDL implementation. The effect shows the different direction for the data with *-ee* in the *fastText* implementation. This finding might be due to the fact that *-ee* denotes participants contrary to the other two suffixes which denote processes or states. In other words, the denotation of nominalizations with the suffix *-ee* is overall more noun-like and has ontologically more resemblance to nouns than verbs. No significant effects of the word class of the base was found for *-ation* in both implementations and for *-ment* in the LDL implementation.

The t-SNE and LDA analyses (Chapters 6, 7 and 7, Sections 6.2.2 and 7.2.2) for both computational implementations contribute to the overall observation that the effect of the word class of a base is rather a side effect than the most influential factor. The t-SNE distribution showed clusters for *-ment* and *-ation* in the *fastText* vectors and clusters for *-ation* in the LDL implementation. Nonetheless, the found clusters are not clearly dividing denominal and deverbal derivatives but only for sub-categories like chemical substances as bases for *-ation*, or attitudinal nouns for *-ment*. The finding that there are no clusters for derivatives in *-ee* at all adds to the lack of a clear distinction of semantic vectors for denominal and deverbal derivatives.

The LDAs demonstrate that the classification of whether a nominalization is denominal or deverbal is not a trivial task. The prediction for the data, which showed clusters in the t-SNE analyses, illustrates variation in the prediction because some nominalizations were predicted to be denominal or deverbal. However, the prediction contains many wrongly predicted data points. In cases where no clusters were found, LDA predicted all nominalizations to be deverbal. This analysis contributes to the overall picture about the role of the word class of the base.

A non-trivial difference between the *fastText* vectors (Bojanowski et al. 2016; Mikolov et al. 2018) and the LDL vectors (e.g., Baayen et al. 2019b) is that the LDL vectors are created via comprehension. That is, with LDL vectors we are not dealing with raw vectors as in *fastText* but with vectors that underwent a matrix-multiplication task to simulate comprehension. For the study in this dissertation, the LDL model was trained with raw vectors from NDL, which is a cognitive-based approach (cf. Baayen et al. 2011). It might be interesting to implement a similar LDL model on the basis of the *fastText* vectors.

The suitable use of the two different computational implementations, both based on n-grams, leads to the assumption that smaller units of information are more important for the semantics of a word than morphological rules. This observation is in line with previous findings (cf. Baayen et al. 2011, 2019b; Bojanowski et al. 2016; Arndt-Lappe et al. 2022). Meaning seems to be associated to small form units in the language. However, the small data size for the denominal nominalizations in general as well as the fact that most of these forms are highly infrequent and have therefore been treated as pseudowords[28] in the implementations might bias the models to the finding that form is more important than morphology. That is, the estimation of word vectors in forms of whole words does not specifically carry the information that a derivational process is underlying the change in form from base to derivative.

[28]This term is chosen as the word vectors for the derivatives are created in *fastText* and LDL via character n-grams is comparable to Schmitz et al. (2021).

Chapter 9

General discussion

This dissertation contributes to the understanding of the word-formation process of eventuality-related nominalizations. The analysis of nominalizations with the suffixes *-ee*, *-ment*, and *-ation* serve as illustrative examples for eventuality-related nominalizations. More precisely, such denominal nominalizations have not yet been analyzed in detail. Intuitively, this is not a big surprise as the interpretation of eventuality-related nominalizations normally rests on the presence of the eventuality in the base. Eventualities are ontologically speaking denoted by verbs (e.g., Haspelmath 2001; Van Valin & LaPolla 2002; Szabó 2015; Moltmann 2019). The literature on nominalizations states that nominal and adjectival bases are possible. Nonetheless, the derivational process for denominal nominalizations is unclear:

> The verbal relation is implied by context or can be inferred from the nature of the non-verbal base. [...] Such interpretations follow from the sort of activities that the base nouns could conceivably be involved in. (Bauer et al. 2013: 233)

The studies conducted in this dissertation show the role of the nominal base in the derivational process. In order to do so, first, frame semantic analyses (e.g., Petersen 2007; Löbner 2013, 2014; Petersen & Gamerschlag 2014; Andreou 2017b; Kotowski 2020; see also Chapter 4) were performed. That is, the semantic structure of the nominal bases was decomposed. This semantic decomposition revealed that nouns, although most nouns are non-eventive in nature, do indeed bear eventive structures in their semantics which can be used by derivational processes. In other words, even non-eventive nouns contain an eventive structure accessible for the derivational process with different nominalizing suffixes.

The reference shifting approach, applied in this dissertation for the decomposition of semantic structures, was demonstrated to be already successful in the analysis of deverbal *-ment*-derivatives (Kawaletz 2023). The present findings illustrate that the word-formation process with the suffixes under investigation operates on the semantic structure provided by the base. More precisely, the denoted referent of the expression shifts from the denotation of the base word to the reading of the derivative by shifting the reference to a semantic element in the semantic structure of the base. Thus, the reference shifting approach can also be transferred to eventuality-related nominalizations with nominal bases.

The second part of this dissertation analyzed eventuality-related nominalizations using two different computational methods. Distributional methods, such as *fastText* (Bojanowski et al. 2016; Mikolov et al. 2018), showed to be successful for analyzing language, particularly derivation (e.g., Boleda & Herbelot 2016; Wauquier et al. 2018; Boleda 2020; Wauquier 2020; Huyghe & Wauquier 2020; Wauquier 2022; Kotowski & Schäfer 2023). The computational analyses were performed in order to compare denominal derivatives to deverbal derivatives. The purpose of this comparison was to investigate whether denominal derivatives and their bases show the same semantic similarity as deverbal derivatives and their bases. As previous studies on deverbal nominalizations indicated, the readings of the derivatives depend on the semantic structure of the base. The base and derivative were assumed to also exhibit a similarity regarding computed cosine similarities. The findings in this dissertation from the frame semantic analysis revealed that denominal nominalizations also rely on the semantic structure of the base. Thus, the comparison of denominal and deverbal derivatives and their bases should reflect the similarity of derivative and base operating on the same semantic structures.

The analyses of the *fastText* vectors demonstrated that denominal nominalizations are similar to their nominal bases. This is in line with the expectation as the reading of the derivative relies on the semantic structures provided by the base word. The comparison of deverbal nominalizations and their verbal bases indicates a higher similarity of derivatives and bases for the suffixes *-ation* and *-ment*. Contrarily, the data for the suffix *-ee* shows an overall higher cosine similarity for denominal derivatives and their bases compared to deverbal derivatives and their verbal bases.

The other computational approach was discriminative learning (NDL; e.g., Baayen et al. 2011, 2016; Schmitz et al. 2023, LDL; e.g., Baayen et al. 2019b; Schmitz 2022; Stein 2023). The LDL implementation in this dissertation is applied to the entire words and splits them into trigrams for the information of forms that enter the computation. This implementation is similar to *fastText*. The results of the LDL implementation differ from the *fastText* results. The word class of the base is only an influential factor for the analysis of the derivatives with the suffix *-ee*. Interestingly, the results of the LDL implementation meet the expectation that deverbal derivatives are more similar to their verbal bases than denominal derivatives to their nominal bases. This is contradictory to the findings of the *fastText* analysis.

The *t*-SNE and LDA analyses illustrated that some clusters of derivatives are visible in a two-dimensional semantic space. For the *fastText* vectors, clusters for derivatives in *-ment* and *-ation* were observed. For the LDL vectors, only clusters for the data in *-ation* were found, distinguishing deverbal and denominal chemical expressions. No visible clusters for derivatives in *-ee* were found. While predicting denominal and deverbal derivatives, the LDA revealed that a lot of false predictions of denominal derivatives were made. The lack of a clear cluster in the *t*-SNE analyses, distinguishing denominal and deverbal derivatives as well as the wrong predictions of the LDAs, support the results of the frame analyses and the beta regression models which indicate that the word class of the base is not the main factor in derivational processes in general.

The findings of the computational implementations give rise to two major assumptions: First, the derivational process is unsurprisingly dependent on the suffix. Second, the derivational process shows a clearer connection of derivatives with the semantics of verbal bases than nominal ones as the cosine similarities of deverbal derivatives and their verbal bases are overall higher. Broadly speaking, the difference in the cosine similarities of denominal and deverbal bases and their derivatives is based on ontology. Participants, as in nominalizations in *-ee*, are usually denoted by nouns, whereas eventualities are usually denoted by verbs (e.g., Haspelmath 2001; Van Valin & LaPolla 2002; Szabó 2015; Moltmann 2019). Hence, the different finding of the cosine similarities for denominal derivatives with *-ee* arises because the suffix creates participant readings which are usually denoted by nouns. Contrarily, derivatives with *-ment* and *-ation* create readings based on eventive structures in the base which are more verb-like than participant readings. The ontology of word classes is also involved in the findings of the second assumption: Due to the eventuality-denoting nature of verbs, the derivational process is clearer interpretable with verbs than with non-eventive nouns, because the eventive structure is exposed by the denotation of eventualities of verbs. The high degree of decomposition needed for nominal bases to actually reveal the eventive structures in the nominal base makes the derivational process more difficult to analyze. Nonetheless, the required eventive elements to successfully create an eventuality-related nominalization are provided by nominal bases, too.

Combining the results from the frame semantic analyses with the results from the computational models, novel insights into the derivational process of eventuality-related nominalizations with nominal bases are gained. The derivational process with nominal bases resembles the process with verbal bases. That is, the reference shifts from one component in the semantic structure of a base to a different element in the semantic structure. The node to which the reference shifts is then denoted by the derivative, creating its interpretation.

The reference shift from one element to another element in the same semantic structures of the base word results in similarities of derivative and bases in the computational models. If a derivative is similar to its base, it is likely that the same semantic structure is accessed. Thus, a higher cosine similarity of derivative and base speaks in favor of a semantic similarity. The findings of the computational models confirm the findings of the frame semantic analyses as a similarity of bases and derivatives are observable. This is not surprising as derivative and base do not only share the eventuality on which they create their reading, but they also share similarities in their written (and phonological) form.

The similarity of derivatives and their bases is also dependent on the individual suffix. The frame semantic analyses showed that different suffixes shift the reading from the denotation of the base to the reading of the derivative. Most importantly, the word-formation processes with different suffixes shift the reference to different semantic elements provided by the semantic structure of the base. That is, nominalizations with *-ee* refer to sentient participants (e.g., Barker 1998; Mühleisen 2010), nominalizations with *-ment* can refer to several different semantic elements (e.g., Plag et al. 2018; Kawaletz 2023), and nominalizations with *-ation* refer to processes or result-states provided by the semantic structure of the base (e.g., Plag 2018; Stein 2023; see also Chapter 4). This suffix-specific behavior is partly reflected in the results from the computational analyses: *-ment* and *-ation* do behave differently compared to *-ee* because the former show a overall higher cosine similarity for deverbal derivatives and their bases, whereas *-ee* displays a higher similarity of denominal derivatives and their bases in the *fastText* analysis. The question arises whether the suffix determines which elements are required to be present in the semantic structure of the base. The outcome of a derivational process with a specific suffix requires different semantic elements in the semantic structure of the base.

The effects of the word class of the base in the beta regression models in the computational implementations do not consistently reach significance. Similarly, the t-SNE and LDA analyses do not show an overall clearly distinct cluster for denominal and deverbal nominalizations. Generally speaking, these results speak in favor of a non-crucial effect of the word class of the base. The results from the frame based analyses point to the same direction: For example, psych nouns work similarly to psych verbs. For all eventuality-related nominalizations more closely inspected in this dissertation, the reference shift works similarly to the reference shifting approach for nominalizations with verbal bases (e.g., Kawaletz 2023). This approach can be extended to other nominalizing suffixes, as well as to different types of bases.

The analyses performed in this dissertation contribute to the understanding of nominalization processes in general. First, the analyses showed how word-formation processes work with nominal bases. Second, the syntactic category of the word class has shown to be non-crucial for the interpretation of nominalizations. Plag (2004) already suggested that the word class of the base does not play the most important role for word-formation processes:

> The claim that the word-class specification of the input does not play a crucial role, or even no role at all, in derivational morphology, has serious implications for morphological theory. (Plag 2004: 216)

The analyses in this dissertation contribute to this statement. The word class of the base is not as important as the semantic structure of the base. More precisely, the success of a word-formation process is dependent on the semantic elements provided by the base. The the word-formation process with a specific suffix determines which eventive elements are required in the semantic structure of the base in order to perform a reference shift from the base to the derivative. For morphological theory in general, this result means that a look at the semantic structure of a base is crucial for the interpretation of a nominalization. Contrarily, the syntactic category of the base is of secondary importance. That is, the word class of the base determines how deeply a semantic decomposition of a base is required to reveal the semantic elements tackled by the word-formation process.

Summarizing, the findings of the analyses I conducted in this dissertation answer the question where the required event for eventuality-related nominalizations is to be found. The decomposed semantic structure of the base provides all the eventive elements required for the word-formation process. This is in line with the suggestion of Bauer et al. that the interpretation of such an eventuality-related nominalization can "follow from the sort of activities that the base nouns could conceivable be involved in" (Bauer et al. 2013: 233). Furthermore, the computational analyses illustrated that denominal and deverbal derivatives are related to their bases to a similar degree, as suggested by the frame analyses. In other words, denominal and deverbal eventuality-related nominalizations do not behave strikingly different as both types of nominalizations operate on the semantics of the base and thus show a semantic similarity to their base. Following this line of thought, the relevance of the word class of the base and the prototypical behavior of nouns and verbs (e.g., Haspelmath 2001; Van Valin & LaPolla 2002; Szabó 2015; Moltmann 2019) are suspended as long as the base provides the required eventive elements for the word-formation process in its semantic structure.

Chapter 10

Conclusion

This dissertation provided novel insights into the analysis of eventuality-related nominalizations. It is the first work to systematically investigate non-deverbal eventuality-related nominalizations, focusing on nominalizations with nouns as bases. This investigation was done using two complementary approaches: A semantic reference shifting approach modeled in semantic frames and two different computational approaches, *fastText* and linear discriminative learning.

The frame semantic approach (Barsalou 1992b, and others) was used to decompose the semantic structure of the base noun and its derivative. This semantic decomposition reveals that the semantic material to induce the reference shift from a noun to a nominalization is indeed already available in the nominal base. The novelty of the results of semantic decomposition of nouns as bases for eventuality-related nominalizations illustrates that even non-eventive nouns bear semantic structures which are required for a reference shift from base to derivative. In other words, the analyses with data for the suffixes *-ee*, *-ment*, and *-ation* showed that denominal derivatives behave similarly to deverbal derivatives. That is, the semantics of the nominal base provides the eventuality needed for the reference shift. Nonetheless, one crucial difference is to be found: The semantic structure of nominal bases shows that the eventuality needed for the derivational process is more deeply embedded in its semantic structure. Thus, the nominal base needs a different level of decomposition in contrast to verbal bases as verbs denote eventualities and nouns usually do not.

The computational approaches, one using *fastText* (Bojanowski et al. 2016; Mikolov et al. 2018), and one LDL (Baayen et al. 2019b, and others), were chosen to investigate whether denominal derivatives are semantically as closely related to their base like deverbal derivatives to their verbal bases. I observed differences in the similarities of denominal and deverbal derivatives in several beta regression models. Some differences depend on the suffix and are due to ontological reasons: *-ee* creates participant readings which are associated with nouns, whereas *-ation* and *-ment* create readings which are more closely related to verbs, for example, the denotation of a whole eventuality (e.g., Van Valin & LaPolla 1997 and others). That is, broadly speaking, *-ee* derivatives are due to their noun-likeness closer to nouns, and thus their nominal bases, compared to the suffixes *-ation* and *-ment*.

The t-SNE analyses of all vectors of the derivatives created by the three suffixes showed that in the vector space, denominal and deverbal derivatives are equally distributed without a distinct pattern for distinguishing denominal and deverbal nominalizations. More precisely, some patterns were found, but no clear and general tendency for denominal or deverbal derivatives in a specific place in the two-dimensional semantic space was found. In other words, the analysis indicates that denominal and deverbal derivatives are not completely distinguishable by the word class of the base. The LDA showed issues in the prediction of verbal or nominal bases, too. These findings, as well as the findings of the reference shifting approach that nouns do indeed provide the required eventive structures for the derivational process, show that the word class of the base is not the most crucial factor for the success of the derivational process. More precisely, the word class of the base does not restrict the probability of a reference shift. Rather, the availability of a fitting semantic element in the semantic structure of the base is the most important factor for a successful word-formation process.

Word-formation processes do not pose a trivial problem. This dissertation shed light on the use of nouns as bases for eventuality-related nominalizations and its novel findings contribute to the overall understanding of word-formation processes. The main result – that the word class of the base does not play a central role for derivational processes with the suffixes *-ee*, *-ment* and *-ation* – shows that theoretical approaches are in need of revisions. The semantics of derivatives and their bases must be included and given a greater role in theory on word-formation.

The proposed analyses lead to further interests in this area of research. First, other suffixes that create eventuality-related readings, for example *-age*, *-ance*, *-er*, and *-ure*, are of interest. I assume that these analyses will show similar results, especially with regard to the role of the word class of the base in derivational processes. Second, an exploration of prefixes (e.g., Kotowski 2020; Kotowski & Schäfer 2023) with both methods, frame semantics and computational semantics, is also an interesting point in completing the picture of derivational processes. The analysis at hand leads to the assumption that the semantics of base and derivative are more important than the word class of the base in derivational processes. Thus, the question arises if prefixation behaves like suffixation with regard to the word class of the base. Third, using different vectors for computational analyses, for example, using the *fastText* vectors for an LDL implementation, or using different word embedding algorithms like, e.g., FRACSS (Marelli & Baroni 2015) is work for the future. The results from the analyses in this dissertation suggests that other influencing factors, i.e., not the word class of the base, for the similarity of derivatives and their bases are to be found. With the rise of large language models and the ever-increasing usage of computational methods in linguistic research, I assume that more insights regarding the interrelations of semantics and derivational processes will be gained.

References

Alexiadou, Artemis. 2001. *Functional structure in nominals: Nominalization and ergativity*, vol. 42 Linguistik aktuell. Philadelphia: John Benjamins Publishing Company. doi:https://doi.org/10.1075/la.42.

Andreou, Marios. 2017a. Pius ten hacken (ed.): The semantics of compounding. *Morphology* 27(4). 721–725. doi:https://doi.org/10.1007/s11525-017-9311-1.

Andreou, Marios. 2017b. Stereotype negation in frame semantics. *Glossa: a journal of general linguistics* 2(1). doi:https://doi.org/10.5334/gjgl.293.

Arndt-Lappe, Sabine, Robin Schrecklinger & Fabian Tomaschek. 2022. Stratification effects without morphological strata, syllable counting effects without counts – modelling English stress assignment with naive discriminative learning. *Morphology* 33. 433–457. doi:https://doi.org/10.1007/s11525-022-09399-9.

Arregi, Karlos. 2012. *Morphotactics: Basque auxiliaries and the structure of spellout*, vol. 86 SpringerLink Bücher. Dordrecht: Springer. doi:https://doi.org/10.1007/978-94-007-3889-8.

Baayen, R. Harald. 2001. *Word frequency distributions*, vol. 18 Text, speech and language technology. Dordrecht: Kluwer. doi:https://doi.org/10.1007/978-94-010-0844-0.

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. doi:https://doi.org/10.1017/CBO9780511801686.

Baayen, R. Harald, Yu-Ying Chuang & Maria Heitmeier. 2019a. Wpmwithldl: Implementation of word and paradigm morphology with linear discriminative learning. `http://www.sfs.uni-tuebingen.de/ hbaayen/publications/WpmWithLdl_1.0.tar.gz`.

Baayen, R. Harald, Yu-Ying Chuang, Elnaz Shafaei-Bajestan & James P. Blevins. 2019b. The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity* 2019. 4895891. doi:https://doi.org/10.1155/2019/4895891.

Baayen, R. Harald, Petar Milin, Dusica Filipović Đurđević, Peter Hendrix & Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118(3). 438–481. doi:https://doi.org/10.1037/a0023851.

Baayen, R. Harald, Cyrus Shaoul, Jon Willits & Michael Ramscar. 2016. Comprehension without segmentation: A proof of concept with naive discriminative learning. *Language, Cognition and Neuroscience* 31(1). 106–128. doi:https://doi.org/10.1080/23273798.2015.1065336.

Bach, Emmon. 1986. The algebra of events. *Linguistics and Philosophy* 9(1). 5–16. doi:https://doi.org/10.1007/BF00627432.

Baeskow, Heike. 2021. Noun-verb conversion as a metonymic metamorphosis. *Skase Journal of Theoretical Linguistics* 18(1). 2–34.

Balteiro, Isabel. 2007. *The directionality of conversion in English: A dia-synchronic study*, vol. 59 Linguistic insights. Bern and New York: Lang.

Barbu Mititelu, Verginica, Gianina Iordăchioaia, Svetlozara Leseva & Ivelina Stoyanova. 2023. The meaning of zero nouns and zero verbs. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: Theory, methods, evidence*, 63–102. De Gruyter. doi:https://doi.org/10.1515/9783111074917-004.

Barker, Chris. 1998. Episodic -ee in English: A thematic role constraint on new word formation. *Language* 74(4). 695–727. doi:https://doi.org/10.2307/417000.

Baroni, Marco, Georgiana Dinu & Germán Kruszewski. 2014. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1: Long papers)*, 238–247. Baltimore, Maryland: Association for Computational Linguistics. doi:10.3115/v1/P14-1023.

Baroni, Marco & Alessandro Lenci. 2010. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics* 36(4). 673–721. doi:https://doi.org/10.1162/coli_a_00016.

Barsalou, Lawrence W. 1992a. *Cognitive psychology. An overview for cognitive sciences*. Hillsdale: NJ: Erlbaum. doi:https://doi.org/10.4324/9781315807485.

Barsalou, Lawrence W. 1992b. Frames, concepts, and conceptual fields. In Adrienne Lehrer & Eva Feder Kittay (eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization*, 21–74. Hillsdale, N.J.: Lawrence Erlbaum Associates. doi:https://doi.org/10.4324/9780203062821.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford Univ. Press 1st edn. doi:https://doi.org/10.1017/S0022226714000292.

Belletti, Adriana & Luigi Rizzi. 1988. Psych-verbs and θ-theory. *Natural Language & Linguistic Theory* 291–352. doi:https://doi.org/10.1007/BF00133902.

Bobaljik, Jonathan David. 2012. *Universals in comparative morphology: Suppletion, superlatives, and the structure of words*. MIT Press. doi:https://doi.org/10.1515/lingty-2016-0038.

Bobaljik, Jonathan David. 2013. Distributed morphology. In Mark Aronoff (ed.), *Linguistics*, New York: Oxford University Press. doi:https://doi.org/10.1093/acrefore/9780199384655.013.131.

Bobaljik, Jonathan David, Kleanthes K. Grohmann & Caro Struijke. 2000. The ins and outs of contextual allomorphy. In Kleanthes K. Grohmann & Caro Struijke (eds.), *University of maryland working papers in linguistics*, vol. 10, 35–71.

Bojanowski, Piotr, Edouard Grave, Armand Joulin & Tomas Mikolov. 2016. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606* .

Boleda, Gemma. 2020. Distributional semantics and linguistic theory. *Annual Review of Linguistics* 6(1). 213–234. doi:https://doi.org/10.1146/annurev-linguistics-011619-030303.

Boleda, Gemma & Aurélie Herbelot. 2016. Formal distributional semantics: Introduction to the special issue. *Computational Linguistics* 42(4). 619–635. doi:https://doi.org/10.1162/COLI_a_00261.

Bonami, Olivier & Berthold Crysmann. 2016. The role of morphology in constraint-based lexicalist grammars. In Hippisley, Andrew, Stump & Gregory T (eds.), *Cambridge handbook of morphology*, 609–656. Cambridge University Press.

Bonami, Olivier & Matías Guzmán Naranjo. 2023. Distributional evidence for derivational paradigms. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: Theory, methods, evidence*, 219–258. De Gruyter. doi:https://doi.org/10.1515/9783111074917-008.

Bonet, Eulàlia & Daniel Harbour. 2012. Contextual allomorphy. In Jochen Trommer (ed.), *The morphology and phonology of exponence* Oxford linguistics, 195–235. Oxford: Oxford Univ. Press. doi:https://doi.org/10.1093/acprof:oso/9780199573721.003.0007.

Booij, Geert E. 2010. *Construction morphology* Oxford linguistics. Oxford: Oxford Univ. Press.

Borer, Hagit. 2013. *Structuring sense*, vol. III: Taking form. Oxford: Oxford Univ. Press 1st edn. doi:https://doi.org/10.1017/S0022226714000310.

Bram, Barli. 2011. *Major total conversion in English.* Wellington: PhD dissertation, Victoria University.

Burzio, Luigi. 1994. *Principles of English stress*. Cambridge University Press. doi:https://doi.org/10.1017/CBO9780511519741.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in English transformational grammar*, 184–221. Boston: Ginn.

Chuang, Yu-Ying, Marie Lenka Vollmer, Elnaz Shafaei-Bajestan, Susanne Gahl, Peter Hendrix & R. Harald Baayen. 2021. The processing of pseudoword form and meaning in production and comprehension: A computational modeling approach using linear discriminative learning. *Behavior Research Methods* 53(3). 945–976. doi:https://doi.org/10.3758/s13428-020-01356-w.

Cribari-Neto, Francisco & Achim Zeileis. 2010. Beta regression in R. *Journal of Statistical Software* 34(2). 1–24. doi:https://doi.org/10.18637/jss.v034.i02.

Davies, Mark. 2004-. British National Corpus (BNC) (from Oxford University Press). `https://www.english-corpora.org/bnc/`.

Davies, Mark. 2008-. The Corpus of Contemporary American English (COCA): One billion words, 1990-2019. `https://www.english-corpora.org/coca/`.

Davies, Mark. 2010-. The Corpus of Historical American English (COHA): 400 million words, 1810-2009. `https://www.english-corpora.org/coha/`.

Davies, Mark. 2016-. *Corpus of News on the Web (NOW)*. `https://www.english-corpora.org/now/`.

Davies, Mark. 2018-. : The 14 billion word iweb corpus. `https://www.english-corpora.org/iWeb/`.

Desmets, Marianne & Florence Villoing. 2009. French VN lexemes: morphological compounding in hpsg. In Stefan Müller (ed.), *Proceedings of the 16th international conference on head-driven phrase structure grammar, georg-august-universität göttingen, germany*, 89–109. Stanford, CA: CSLI Publications.

Embick, David. 2003. Locality, listedness, and morphological identity. *Studia Linguistica* 57(3). 143–169. doi:https://doi.org/10.1111/j.0039-3193.2003.00102.x.

Embick, David. 2010. *Localism versus globalism in morphology and phonology*, vol. 60 Linguistic inquiry monograph. Cambridge, Mass.: MIT Press.

Evert, Stefan & Marco Baroni. 2007. zipfR: word frequency distributions in R. In *Proceedings of the 45th annual meeting of the acl on interactive poster and demonstration sessions* ACL '07, 29–32. USA: Association for Computational Linguistics. doi:https://doi.org/10.3115/1557769.1557780.

Fábregas, Antonio. 2024. *The fine-grained structure of the lexical area*. Philadelphia, PA: John Benjamins Publishing Company. doi:10.1075/ihll.39.

Fellbaum, Christiane (ed.). 1998. *Wordnet: An electronic lexical database* Language, speech, and communication. Cambridge, Mass.: MIT Press.

Ferrari, Silvia & Francisco Cribari-Neto. 2004. Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31(7). 799–815. doi:https://doi.org/10.1080/0266476042000214501.

Fox, John & Sanford Weisberg. 2019. *An R companion to applied regression*. SAGE Publications, Ltd.

Gamerschlag, Thomas, Doris Gerland, Rainer Osswald & Wiebke Petersen (eds.). 2014. *Frames and concept types*. Dordrecht: Springer International Publishing. doi:https://doi.org/10.1007/978-3-319-01541-5.

Gamerschlag, Thomas, Wiebke Petersen & Liane Ströbel. 2013. Sitting, standing, and lying in frames: A frame-based approach to posture verbs. In Guram Bezhanishvili, Sebastian Löbner, Vincenzo Marra & Frank Richter (eds.), *Logic, language, and computation*, 73–93. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:https://doi.org/10.1007/978-3-642-36976-6_7.

Grimshaw, Jane Barbara. 1990. *Argument structure*, vol. 18 Linguistic inquiry monographs. Cambridge, Mass: MIT Press.

Harbour, Daniel. 2007. *Morphosemantic number: From Kiowa noun classes to UG number features* Studies in Natural Language and Linguistic Theory. Springer: Netherlands. doi:http://dx.doi.org/10.1007/978-1-4020-5038-1.

Harris, Zellig S. 1954. Distributional structure. *WORD* 10(2-3). 146–162. doi:https://doi.org/10.1080/00437956.1954.11659520.

Haspelmath, Martin. 2001. Word classes and parts of speech. In Neil J. Smelser & Paul B. Baltes (eds.), *International encyclopedia of the social & behavioral sciences*, 16538–16545. Amsterdam: Elsevier. doi:https://doi.org/10.1016/B0-08-043076-7/02959-4.

Hay, Jennifer & Harald Baayen. 2003. Phonotactics, parsing and productivity. *Italian Journal of Linguistics* 1. 99–130.

Huyghe, Richard & Marine Wauquier. 2020. What's in an agent? *Morphology* 30(3). 185–218. doi:https://doi.org/10.1007/s11525-020-09366-2.

Jackendoff, Ray. 1991. Parts and boundaries. *Cognition* 41(1). 9–45. doi:https://doi.org/10.1016/0010-0277(91)90031-X.

Jackendoff, Ray S. 1990. *Semantic structures*. Cambridge: MIT Press.

Jones, Michael N. & Douglas J. K. Mewhort. 2007. Representing word meaning and order information in a composite holographic lexicon. *Psychological Review* 114(1). 1–37. doi:https://doi.org/10.1037/0033-295X.114.1.1.

Kawaletz, Lea. 2023. *The semantics of English -ment nominalizations* (Empirically Oriented Theoretical Morphology and Syntax 12). Berlin: Language Science Press. doi:https://10.5281/zenodo.7915801.

Kawaletz, Lea & Ingo Plag. 2015. Predicting the semantics of English nominalizations: A frame-based analysis of –ment suffixation. In Laurie Bauer, Lívia Körtvélyessy & Pavol Štekauer (eds.), *semantics of complex words*, 289–319. Dordrecht: Springer.

Kennedy, Christopher & Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81(2). 345–381.

Koenig, Jean-Pierre. 1999. *Lexical relations* Stanford monographs in linguistics. Stanford, Calif.: CSLI Publ. Center for the Study of Language and Information.

Kotowski, Sven. 2020. The semantics of English out -prefixation: a corpus-based investigation. *English Language and Linguistics* 1–29. doi:https://doi.org/10.1017/S1360674319000443.

Kotowski, Sven. 2023. Modeling locative prefix semantics. A formal account of the English verbal prefix out-. *Morphology* 115–152. doi: https://doi.org/10.1007/s11525-023-09405-8.

Kotowski, Sven & Ingo Plag. 2023. The semantics of derivational morphology: Introduction. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: Theory, methods, evidence*, 1–14. De Gruyter. doi:https://doi.org/10.1515/9783111074917-001.

Kotowski, Sven & Martin Schäfer. 2023. Quantifying semantic relatedness across base verbs and derivatives: English out-prefixation. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: Theory, methods, evidence*, 177–218. Berlin, Boston: De Gruyter. doi: https://doi.org/10.1515/9783111074917-007.

Kotowski, Sven, Viktoria Schneider & Lea Kawaletz. 2023. Eventualities in nominalisation semantics: The case of denominal -ment formations. *Roczniki Humanistyczne* 71. 149–173. doi:https://doi.org/10.18290/rh237111-6s.

Krijthe, Jesse H. 2015. Rtsne: T-distributed stochastic neighbor embedding using barnes-hut implementation. `https://github.com/jkrijthe/Rtsne`.

Lakoff, George. 1966. Stative adjectives and verbs in English. *Mathematical Linguistics and Automatic Translation Report NSF-17*. 1–16.

Landauer, Thomas K. & Susan T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104(2). 211–240. doi:https://doi.org/10.1037/0033-295X.104.2.211.

Lapesa, Gabriella, Lea Kawaletz, Ingo Plag, Marios Andreou, Max Kisselew & Sebastian Padó. 2018. Disambiguation of newly derived nominalizations in context: A distributional semantics approach. *Word Structure* 11(3). 277–312. doi:https://doi.org/10.3366/word.2018.0131.

Larson, Richard K. 1998. Events and modification in nominals. *Semantics and Linguistic Theory* 8. 145–168. doi:http://dx.doi.org/10.3765/salt.v8i0.2803.

Lieber, Rochelle. 2004. *Morphology and lexical semantics*. Cambridge: Cambridge University Press. doi:https://doi.org/10.1017/CBO9780511486296.

Lieber, Rochelle. 2016. *English nouns: The ecology of nominalization*, vol. 150 Cambridge studies in linguistics. Cambridge: Cambridge University Press. doi:https://doi.org/10.1017/CBO9781316676288.

Lieber, Rochelle. 2017. Nominalization: General overview and theoretical issues. In Mark Aronoff (ed.), *Oxford research encyclopedias*, Oxford: Oxford University Press. doi:https://doi.org/10.1093/acrefore/9780199384655.013.248.

Löbner, Sebastian. 2013. *Understanding semantics* Understanding language series. London: Routledge 2nd edn. doi:https://doi.org/10.4324/9780203528334.

Löbner, Sebastian. 2014. Evidence for frames from human language. In Thomas Gamerschlag (ed.), *Frames and concept types*, vol. 94 Studies in Linguistics and Philosophy, 23–67. Cham: Springer. doi:https://doi.org/10.1007/978-3-319-01541-5_2.

Löbner, Sebastian. 2017. Frame theory with first-order comparators: Modeling the lexical meaning of punctual verbs of change with frames. In Helle Hvid Hansen, Sarah E. Murray, Mehrnoosh Sadrzadeh & Henk Zeevat (eds.), *Logic, language, and computation* Lecture Notes in Computer Science, 98–117. Berlin, Heidelberg and s.l.: Springer Berlin Heidelberg. doi:https://doi.org/10.1007/978-3-662-54332-0_7.

Löbner, Sebastian. 2018. Barsalou-frames in Wort- und Satzsemantik. In Stefan Engelberg, Henning Lobin, Kathrin Steyer & Sascha Wolfer (eds.), *Wortschätze* Jahrbuch des Instituts für Deutsche Sprache, 189–212. Berlin and Boston: DE GRUYTER MOUTON. doi:https://doi.org/10.1515/9783110579963-011.

van der Maaten, L.J.P. 2014. Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research* 15. 3221–3245.

Maechler, Martin. 2020. diptest: Hartigan's dip test statistic for unimodality - corrected. `https://github.com/mmaechler/diptest`.

Maienborn, Claudia. 2003. *Die logische Form von Kopula-Sätzen* Studia grammatica. Berlin: De Gruyter. doi:https://doi.org/10.1524/9783050082271.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In A. Dimitriadis, L. Siegel & al (eds.), *Proceedings of the 21st Annual Penn Linguistics Colloquium* University of Pennsylvania Working Papers in Linguistics, 201–225.

Marelli, Marco & Marco Baroni. 2015. Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review* 122(3). 485–515. doi:https://doi.org/10.1037/a0039267.

Melloni, Chiara. 2007. *Polysemy in word formation: The case of deverbal nominals*. Verona: Università degli studi di Verona dissertation. `https://iris.univr.it/retrieve/handle/11562/337830/3289/TESI.pdf`.

Metzger, Sarah, Sebastian Bücking, Frauke Buscher, Natascha Elxnath, Johanna Herdtfelder, Julia Lukassek, Anna Prysłopska, Sarah Zobel & Claudia Maienborn. 2019. Basisontologie zur semantischen Annotation von Nomina – Manual (version: 07. november 2019) manuscript.

Mikolov, Tomas, Kai Chen, Greg Corrado & Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space `https://arxiv.org/pdf/1301.3781`.

Mikolov, Tomas, Edouard Grave, Piotr Bojanowski, Christian Puhrsch & Armand Joulin. 2018. Advances in pre-training distributed word representations. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis & Takenobu Tokunaga (eds.), *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*, Miyazaki, Japan: European Language Resources Association (ELRA).

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado & Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani &

K.Q. Weinberger (eds.), *Advances in neural information processing systems*, vol. 26, Curran Associates, Inc.

Milin, Petar, Laurie Beth Feldman, Michael Ramscar, Peter Hendrix, R. Harald Baayen & Hedderik van Rijn. 2017. Discrimination in lexical decision. *PLoS ONE* 12(2). e0171935. doi:https://doi.org/10.1371/journal.pone.0171935.

Moltmann, Friederike. 2019. Natural language and its ontology. In Alvin I. Goldman & Brian P. McLaughlin (eds.), *Metaphysics and cognitive science*, 206–232. New York, NY, USA: Oxford University Press. doi:https://doi.org/10.1093/oso/9780190639679.003.0009.

Moore, E. Hastings. 1920. On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society* 26. 394–395.

Morzycki, Marcin. 2009. Degree modification of gradable nouns: Size adjectives and adnominal degree morphemes. *Natural Language Semantics* 17(2). 175–203. doi:https://doi.org/10.1007/s11050-009-9045-7.

Moskal, Beata. 2015a. Domains on the border: Between morphology and phonology. *Doctoral Dissertations. University of Connecticut* `https://opencommons.uconn.edu/dissertations/892`.

Moskal, Beata. 2015b. Limits on allomorphy: A case study in nominal suppletion. *Linguistic Inquiry* 46(2). 363–376.

Mühleisen, Susanne. 2010. *Heterogeneity in word-formation patterns: A corpus-based analysis of suffixation with -ee and its productivity in English*, vol. 118 Studies in Language Companion Series. Amsterdam and Philadelphia: John Benjamins. doi:https://doi.org/10.4000/lexis.1791.

Müller, Stefan. 2015. HPSG – a synopsis. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – theory and analysis: An international handbook*, vol. 2 Handbücher zur Sprach- und Kommunikationswissenschaft, 937–973. Berlin: Walter de Gruyter Verlag.

Naumann, Ralf. 2013. An outline of a dynamic theory of frames. In Guram Bezhanishvili (ed.), *Logic, language, and computation*, vol. 7758 FoLLI publications on logic, language and information, 115–137. Berlin and Heidelberg: Springer. doi:https://doi.org/10.1007/978-3-642-36976-6_9.

Nieder, Jessica, Fabian Tomaschek, Enum Cohrs & Ruben de van Vijver. 2022. Modelling maltese noun plural classes without morphemes. *Language, Cognition and Neuroscience* 37(3). 381–402. doi:https://doi.org/10.1080/23273798.2021.1977835.

OED. 2025. *Oxford English Dictionary online*. Oxford University Press. `https://www.oed.com/`.

Paradis, Carita. 2008. Configurations, construals and change: expressions of degree. *English Language and Linguistics* 12(2). 317–343. doi:https://doi.org/10.1017/S1360674308002645.

Penrose, Roger. 1955. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society* 51(3). 406–413. doi:https://doi.org/10.1017/S0305004100030401.

Pesetsky, David Michael. 1996. *Zero syntax: Experiencers and cascades*, vol. 27 Current studies in linguistics. Cambridge, Mass.: MIT Press 1st edn.

Petersen, Wiebke. 2007. Representation of concepts as frames. In Jurgis Skilters, Fiorenza Toccafondi & Gerhard Stemberger (eds.), *Complex cognition and qualitative science: A legacy of oswald kulpe (the baltic international yearbook of cognition, logic and communication 2)*, vol. 2, 151–170. Riga: University of Latvia.

Petersen, Wiebke & Thomas Gamerschlag. 2014. Why chocolate eggs can taste old but not oval: A frame-theoretic analysis of inferential evidentials. In Thomas Gamerschlag (ed.), *Frames and concept types*, vol. 94 Studies in Linguistics and Philosophy, 199–218. Cham: Springer. doi:https://doi.org/10.1007/978-3-319-01541-5_9.

Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation*, vol. 28 Topics in English Linguistics /TiEL]. Berlin: De Gruyter. doi:https://doi.org/10.1515/9783110802863.

Plag, Ingo. 2004. Syntactic category information and the semantics of derivational morphological rules. *Folia Linguistica* 38(3-4). 193–225. doi:https://doi.org/10.1515/flin.2004.38.3-4.193.

Plag, Ingo. 2018. *Word-formation in English* Cambridge textbooks in linguistics. Cambridge, United Kingdom and New York, NY: Cambridge University Press 2nd edn. doi:https://doi.org/10.4000/lexis.4532.

Plag, Ingo, Marios Andreou & Lea Kawaletz. 2018. A frame-semantic approach to polysemy in affixation. In Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer (eds.), *The lexeme in descriptive and theoretical morphology*, 467–486. Berlin: Language Science Press.

Plag, Ingo, Julia Homann & Gero Kunter. 2017. Homophony and morphology: The acoustics of word-final s in English. *Journal of Linguistics* 53(1). 181–216. doi:https://doi.org/10.1017/S0022226715000183.

Pollard, Carl Jesse & Ivan A. Sag. 1994. *Head-driven phrase structure grammar* Studies in contemporary linguistics. Stanford: Center for the Study of Language and Information. doi:https://doi.org/10.5281/zenodo.5543318.

Princeton University. 2010. About wordnet. `https://wordnet.princeton.edu/`.

Pustejovsky, James. 1996. *The generative lexicon* A Bradford book. Cambridge, Mass.: MIT Press 2nd edn.

R Core Team. 2025. R: A language and environment for statistical computing. `https://www.r-project.org/`.

Raffelsiefen, Renate. 1998. Phonological constraints on English word formation. In *Yearbook of morphology 1998*, 225–287. Springer.

Rescorla, Robert A. & Allan R. Wagner. 1972. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Abraham H. Black & William F. Prokasy (eds.), *Classical conditioning II: Current research and theory*, 64–99. New York: Appleton-Century-Crofts.

Riehemann, Susanne Z. 1998. Type-based derivational morphology. *Journal of Comparative Germanic Linguistics* 2(1). 49–77. doi:https://doi.org/10.1023/A:1009746617055.

Sag, Ivan A. 2012. Sign-based construction grammar: An informal synopsis. In Hans Christian Boas & Ivan A. Sag (eds.), *Sign-based construction grammar* CSLI lecture notes, 69–202. Stanford, Calif.: CSLI Publ. Center for the Study of Language and Information.

Sag, Ivan A. & Thomas Wasow. 1999. *Syntactic theory: A formal introduction* CSLI lecture notes. Stanford, CA: Center for the Study of Language and Information.

Schäfer, Martin. 2025. The role of meaning in the rivalry of -ity and -ness: evidence from distributional semantics. *English Language and Linguistics* 1–46. doi:https://10.1017/S1360674324000443.

Schmid, Hans-Jörg. 1999. Towards a functional-cognitive lexicology of nouns. In Wolfgang Falkner & Hans-Jörg Schmid (eds.), *Words, lexemes, concepts - approaches to the lexicon. studies in honour of leonhard lipka,*, 213–226. Tübingen: Gunter Narr Verlag.

Schmitz, Dominic. 2022. *Production, perception, and comprehension of subphonemic detail.* Language Science Press. doi:https://doi.org/10.5281/zenodo.7267830.

Schmitz, Dominic. 2023. In German, all professors are male. In *Ingo 6.0. the proceedings. new empirical insights about language, presented on a great day out in september*, doi:https://doi.org/10.31234/osf.io/yjuhc.

Schmitz, Dominic & Janina Esser. 2021. Sfl: Statistics for linguistics. `https://github.com/dosc91/SfL`.

Schmitz, Dominic, Ingo Plag, Dinah Baer-Henney & Simon David Stein. 2021. Durational differences of word-final /s/ emerge from the lexicon: Modelling morpho-phonetic effects in pseudowords with linear discriminative learning. *Frontiers in Psychology* 12. doi:https://doi.org/10.3389/fpsyg.2021.680889.

Schmitz, Dominic & Viktoria Schneider. 2022. gdsm: General functions for distributional semantics. `https://github.com/dosc91/gdsm`.

Schmitz, Dominic, Viktoria Schneider & Janina Esser. 2023. No genericity in sight: An exploration of the semantics of masculine generics in German. *Glossa Psycholinguistics* 2(1). 1–33. doi:https://doi.org/10.31234/osf.io/c27r9.

Schneider, Viktoria. 2023. Eventualities in the semantics of denominal nominalizations. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: Theory, methods, evidence*, 37–62. De Gruyter. doi:https://doi.org/10.1515/9783111074917-003.

Schneider, Viktoria. forthcoming. The semantics of -ee and -ation: a distributional semantic approach. *INGO 6.0. The Proceedings. New empirical Insights about laNguage, presented on a Great day Out in September* doi: https://doi.org/10.31234/osf.io/fre9v_v1.

Schulzek, Daniel. 2014. A frame approach to metonymical processes in some common types of German word formation. In Thomas Gamerschlag, Doris Gerland, Rainer Osswald & Wiebke Petersen (eds.), *Frames and concept types*, vol. 94, 221–242. Dordrecht: Springer International Publishing. doi:https://doi.org/10.1007/978-3-319-01541-5_10.

Schulzek, Daniel. 2019. *A frame approach to German nominal word formation*. Düsseldorf: PhD Dissertation, Heinrich-Heine-Universität Düsseldorf. `https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=51530`.

Shafaei-Bajestan, Elnaz, Masoumeh Moradipour-Tari, Peter Uhrig & R. Harald Baayen. 2022a. Semantic properties of English nominal pluralization: Insights from word embeddings. doi:https://doi.org/10.48550/ARXIV.2203.15424.

Shafaei-Bajestan, Elnaz, Peter Uhrig & R. Harald Baayen. 2022b. Making sense of spoken plurals. *The Mental Lexicon* 17(3). 337–367. doi:https://doi.org/10.1075/ml.22011.sha.

Shaoul, Cyrus & Chris Westbury. 2010. Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods* 42(2). 393–413. doi:https://doi.org/10.3758/BRM.42.2.393.

Sitikhu, Pinky, Kritish Pahi, Pujan Thapa & Subarna Shakya. 2019. A comparison of semantic similarity methods for maximum human interpretability. doi:https://doi.org/10.48550/arXiv.1910.09129.

Solt, Stephanie. 2015. Measurement scales in natural language. *Language and Linguistics Compass* 9(1). 14–32. doi:https://doi.org/10.1111/lnc3.12101.

Stein, Simon David. 2023. *The phonetics of derived words in English: Tracing morphology in speech production*, vol. 585 Linguistische Arbeiten. Berlin and Boston: De Gruyter. doi:https://doi.org/10.1515/9783111025476.

Stein, Simon David & Ingo Plag. 2021. Morpho-phonetic effects in speech production: Modeling the acoustic duration of English derived words with linear discriminative learning. *Frontiers in Psychology* 12. 678712. doi:https://doi.org/10.3389/fpsyg.2021.678712.

Steriade, Donca. 1999. Lexical conservatism in French adjectival liaison. In *Formal perspectives on romance linguistics*, 243–270. John Benjamins Publishing Company.

Steriade, Donca. 2000. Lexical conservatism in French adjectival liaison. In *Papers in laboratory phonology V: Acquisition and the lexicon*, 313–334. Cambridge University Press.

Szabó, Zoltán Gendler. 2015. Major parts of speech. *Erkenntnis* 80(S1). 3–29. doi:https://doi.org/10.1007/s10670-014-9658-1.

Tomaschek, Fabian, Peter Hendrix & R. Harald Baayen. 2018. Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics* 71. 249–267. doi:https://doi.org/10.1016/j.wocn.2018.09.004.

Tomaschek, Fabian, Ingo Plag, Mirjam Ernestus & R. Harald Baayen. 2021. Phonetic effects of morphology and context: Modeling the duration of word-final s in English with naïve discriminative learning. *Journal of Linguistics* 57(1). 123–161. doi:https://doi.org/10.1017/S0022226719000203.

Trendafilov, Nickolay & Michele Gallo. 2021. *Linear discriminant analysis (lda)* 229–268. Cham: Springer International Publishing. doi:10.1007/978-3-030-76974-1_7.

Tribout, Delphine. 2010. How many conversions from verb to noun are there in French? In Stefan Müller (ed.), *Proceedings of the 17th international conference on head-driven phrase structure grammar, université Paris Diderot, Paris 7, France*, 341–357. Stanford, CA: CSLI Publications.

Tucker, Benjamin V., Daniel Brenner, D. Kyle Danielson, Matthew C. Kelley, Filip Nenadić & Michelle Sims. 2019. The massive auditory lexical decision (MALD) database. *Behavior Research Methods* 51(3). 1187–1204. doi:https://doi.org/10.3758/s13428-018-1056-1.

van der Maaten, Laurens J.P. & Geoffrey E. Hinton. 2008. Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research L.J.P. van der Maaten. Accelerating t-SNE using Tree-Based Algorithms. Journal of Machine Learning Research 15(Oct):3221-3245, 2014.* 9. 2579–2605.

van Rossum, Guido & Fred L. Drake. 2009. *Python 3 reference manual*. Scotts Valley, CA: CreateSpace.

Van Valin, Robert D. & Randy J. LaPolla. 1997. *Syntax: Structure, meaning and function* Cambridge textbooks in linguistics. Cambridge: Cambridge Univ. Press.

Van Valin, Robert D. & Randy J. LaPolla. 2002. *Syntax: Structure, meaning and function* Cambridge textbooks in linguistics. Cambridge: Cambridge Univ. Press reprinted. edn.

Venables, William N. & Brian D. Ripley. 2002. *Modern applied statistics with s*. New York: Springer. doi:https://doi.org/10.1007/978-0-387-21706-2.

Wagner, Allan R. & Robert A. Rescorla. 1972. Inhibition in pavlovian conditioning: Application of a theory. In R. A. Boakes & M. S. Halliday (eds.), *Inhibition and learning*, 301–334. London: Academic Press Inc.

Warren, Paul. 2013. *Introducing psycholinguistics* Cambridge introductions to language and linguistics. Cambridge: Cambridge Univ. Press 1st edn. doi:https://doi.org/10.1017/CBO9780511978531.

Wauquier, Marine. 2020. *Confrontation des procédés dérivationnels et des catégories sémantiques dans les modèles distributionnels*: Université Toulouse le Mirail - Toulouse II Theses. https://tel.archives-ouvertes.fr/tel-03543115.

Wauquier, Marine. 2022. Apports de la sémantique distributionnelle pour la morphologie dérivationnelle. *Corpus [En ligne]* 23. doi:https://doi.

org/10.4000/corpus.6303. Mis en ligne le 02 mars 2022, consulté le 30 mars 2025.

Wauquier, Marine, Cécile Fabre & et Nabil Hathout. 2018. Différenciation sémantique de dérivés morphologiques à l'aide de critères distributionnels. *SHS Web of Conferences* 46. 08006. doi:https://doi.org/10.1051/shsconf/20184608006.

Winter, Bodo. 2019. *Statistics for linguists: An introduction using R*. Routledge. `https://doi.org/10.4324/9781315165547`.

Winter, Yoad & Joost Zwarts. 2012. On the event semantics of nominals and adjectives -the one-argument hypothesis. *Proceedings of Sinn und Bedeutung* 16(2). 639–647.

Xanthopoulos, Petros, Panos M. Pardalos & Theodore B. Trafalis. 2013. *Linear discriminant analysis* 27–33. New York, NY: Springer New York. doi:10.1007/978-1-4419-9878-1_4.

Zuur, Alain F., Elena N. Ieno & Chris S. Elphick. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1. 3–14. doi:https://doi.org/10.1111/j.2041-210X.2009.00001.x.