PREFERENTIAL ATTACHMENT AND LANGUAGE CHANGE: *WERDEN* AND THE EMERGENCE OF THE PASSIVE AND THE FUTURE CONSTRUCTIONS

by

Valentina Concu

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Languages and Cultures

West Lafayette, Indiana

May 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. John Sundquist, Chair
School of Languages and Cultures

Dr. Elaine Francis
School of Interdisciplinary Studies

Dr. Colleen Neary-Sundquist
School of Languages and Cultures

Dr. Lynne Miles-Morillo
Indipendent Scholar

Approved by:
Dr. Jennifer William

I dedicate this dissertation to my parents and my sister for always nursing me with love and affection

ACKNOWLEDGMENTS

Throughout the writing of this dissertation, I have received a great deal of support. First, I would like to thank the School of Languages and Cultures and the School of Interdisciplinary Studies for giving me the opportunity to conduct my graduate studies at Purdue University.

I would also like to thank my advisor, Professor John Sundquist, and all the members of my committee for their feedback, help, and invaluable suggestions.

In addition, I would like to thank my parents and my sister for their wise guidance and sympathetic ear. You are always there for me. Finally, I could not have completed this dissertation without the support of my friends and colleagues. A particular thank goes also to Babak Ravandi for his insighful teachings on Network Science.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This study explores historical syntactic changes within a complex network framework focusing on the development of the German verb *werden* (to become) and the emergence of the related passive and future periphrases. The data are collected from a corpus of Middle and Early New High German texts and the analysis of the instances is carried out in two different stages. The first stage focuses on the frequency of the verb *werden* and the elements that co-occurred with it throughout Middle and Early New High German. The second stage investigates the same instances through a complex network framework by applying descriptive statistics to uncover the features of the Middle and Early New High German networks that have been created with the occurrences of *werden* found in the corpus.

The results of the analysis show that *werden* experienced an increase in the type of connections it was able to establish throughout the centuries. Such a process is known in the literature as preferential attachment. This suggests that linguistic networks, and specifically, syntactic networks, are also subjected to processes that are common among non-linguistic networks.

CHAPTER 1. INTRODUCTION

Recent discoveries in network science (Barabási, 2013) have brought forth broad interest in the application of a complex network framework for the study of languages. As argued by Briscoe (1998); Steels (2000), Beckner, Blythe, Bybee, Christiansen, Croft, Ellis, Nick, Holland, Ke, Larsen-Freeman, & Schoenemann (2009), and Solè, Mustra, Valverde, and Steels (2010); and languages are complex adaptive systems that show striking similarities with other complex systems in other disciplines, such as ecological webs, genomes, brain networks, and the internet (Solè et al., 2010, p. 1). Solè et al. (2010) list two main features that are shared by "most complex networks, both natural and artificial" (p. 1). The first one is the "small world" structure, which indicates that the nodes in a specific network are well connected to each other. This means that only a small series of jumps are needed to move from one node to another (p. 3). The second one refers to the distribution of the nodes, which follows a power law, also known as Zipf's law. According to this law, the majority of the nodes in the network have a low number of connections, whereas only a few of them have a high number of links. The power law degree of distribution is defined as one of "the key components of web complexity" (p. 3). Languages are "clearly an example of a complex dynamical system" because they exhibit highly intricate network structures at different levels (phonetic, lexical, syntactic, semantic) that were shaped by speakers over extended periods of time "as they adapt and change them to their needs as part of ongoing local interactions" (p. 3). For this reason, different scholars have claimed that languages are an "evolving word web" (Dorogovtsev & Mendes, 2001) and can be, therefore, "represented as a complex network in [their] several levels of complexity" (Amancio, 2014, p. 1). In their study, Solè et al. (2010) conclude that:

> There are statistical universals in language networks, which are similar to the features found in other 'scale-free' networks arising in physics, biology, and the social sciences. This observation is very exciting from two points of view. First, it points to new types of universal features of languages, which do not focus on properties of the elements in language inventories as in the traditional study of universals (e.g., phoneme inventories or word order patterns in sentences) but rather on statistical properties. Second, the pervasive nature of these network features suggests that language must be subject to the same sort of self-organization dynamics than other natural and social systems, and so it makes sense to investigate whether the general laws governing complex dynamical systems apply to language as well and what aspects of language they can explain (p. 6).

Because "it makes sense" to question whether languages are subjected to the same dynamics of other complex systems, this study aims to capture the development of the verb *werden* from the 11th to the 17th centuries using a complex network approach. I will also try to find evidence for one of the most common processes involved in the growth of complex networks, namely preferential attachment, or the Rich-Get-Richer phenomenon (Barabási & Albert, 1999). This process indicates the tendency by which new nodes in the networks are more likely to be connected to already highly connected ones. The Rich-Get-Richer phenomenon was later reformulated by Bianconi and Barabási (2001) when they noticed that a node's growth rate wasn't determined only by its degree, but also by some other intrinsic properties. Such properties were defined as "fitness" values, and the phenomenon was renamed as "Fitter-Get-Richer" (Barabási, 2006). This value is a measure of the inherent competitive ability that the nodes have in the networks or to their capacity to attract more links (Bell, Perera, Piraveenan, Bliemer, Latty & Reid, 2017).

Other than the reasons stated by Solè et al. (2005), there are two additional reasons for the application of a complex network framework for the study of language change. First, to date, there are not many detailed studies combining historical linguistics and complex networks. Second, the studies available do not focus on languages other than English. This study will try to fill this gap in the literature in both senses while attempting to provide further evidence for the existence of "statistical universals in language networks" (Solè et al., 2005).

The main approach used here follows a Complexity Theory framework. As I will elaborate in the next chapter, I consider languages to be complex adaptive systems with a grammar that can be considered "temporal" and "emergent" (Hopper, 1987). This approach considerably differs from the generativist syntactic theory, according to which syntactic changes are associated with child language acquisition and the transition from one generation to the other (Hock, 1991). I argue that historical changes are better explained within a framework that recognizes the dynamic nature of languages and sees diachronic changes and synchronic variations as closely interrelated to each other. Hence, I propose an analysis in which languages are viewed as complex systems, and the developments of specific verbs and the related emergence of periphrastic constructions as a process similar to the growth of connections in complex networks such as the World Wide Web (Barabási, & Réka 1999).

There are two central research questions which were formulated considering languages as complex adaptive systems:

1. Since "behind each complex system there is an intricate network that encodes the interactions between the system's component" (Barabási, 2016, p.24), are linguistic networks subjected to the same dynamics that are common among non-linguistics networks?

2. If so, what can a complex network approach tell us about historical syntactic linguistic changes?

To answer these two research questions, I will focus on the historical development of the verb *werden* (to become) which can be found in combinations with a large variety of elements. These include nouns, as shown in example (1), adjectives, as shown in example (2), present participles, as shown in example (3), and in the passive and future periphrastic constructions, as shown in examples (4) and (5).

(1) Es **wird** Nacht
It becomes-PRS night-NOM
'Night falls'

(2) Sie **werden** alt
They become-COP old-ADJ
'They are getting old'

(3) Er **wird** wütend
He becomes-COP angry-PTCP.PRS
'He gets angry'

(4) Das Haus **wird** gebaut
The house becomes-AUX built-PTCP.PST
'The house is being built"

(5) Lukas **wird** ein Haus **kaufen**
Lukas will-AUX a house buy-INF
'Lukas will buy a house'

The examples from (1) to (5) show how versatile this verb is today in Modern German, functioning as both a full verb and an auxiliary. From a historical perspective, *werden* acquired the capacity of carrying out such syntactic functions throughout the process of *Desemantisierung*

(desemantization) that took place starting in the Old High German period (9th and 10th centuries) and culminating in the Early New High German period (half of the 14th century - 17th century) (Kotin, 2003, p. 17). Such processes caused, for instance, the increase of the use of *werden* in the passive constructions. This process is also directly involved in the emergence of *werden* as a future marker. This use emerged probably only in the last decades of the 14th century in the East Middle German and Upper High German dialect areas" (Diewald & Wischer, 2013, p. 197), after rarely occurring with an infinitive verb in the Old High German period (p. 211).

The singularities of the development of *werden* make it, I believe, an exceptional candidate to verify if linguistic networks are subject to the same dynamics of extralinguistic networks and to gather evidence for preferential attachment. In this study, I will first describe the frequency and the types of elements found in combination with *werden* in a wide range of texts throughout eight centuries, tracking its growth from a historical linguistic perspective. Although such growth has already been extensively described in the literature (Valentin, 1987; Kotin, 2003 for the passive; Bogner, 1989 for the future), my analysis aims to capture it using the tools offered by network science. After building the syntactic networks for both Middle High and Early New High German, I will analyze the "life" of *werden* using so-called "centrality measures", which are among the most relevant means in network science for determining the role of a given node in a network. These measurements enable tracking both the frequency and the quality of the in-and out degree connections of the nodes (Matas, Martincic-Ipšic, Meštrovic, 2017). Specifically, I will use degree centrality, betweenness centrality, eigenvector centrality, and eccentricity centrality.

One may argue that, since the history of *werden* has already been extensively addressed in the literature, the aim of this work could be seen of secondary importance. This dissertation, however, connects network science tools and measurements to linguistic theory with the aim of showing how historical linguistic changes can be concretely captured using the tools of a discipline that are usually used outside of linguistics. Hence, this study must be perceived as a preliminary step in this direction. The successful implementation of a complex network framework will, hopefully, open doors to further research and bring new perspectives and discoveries to the field of historical linguistic.

The structure of this study is as follows. The second chapter offers a review of the literature and is divided into two sections. In the first one, I offer a survey of the different approaches and attitudes) of the scientific community towards syntactic change. It starts from the nineteenth

century and goes to the generative approaches and Labovian sociolinguistics, and ends with Emergent Grammar, Complexity Theory, Cognitive Grammar and Network Science. In the second section, I provide a more detailed description of the different theories from which I lay the theoretical foundation of this work, namely Complexity Theory, and Network Science. The third chapter offers a detailed literature review on *werden*, with particular attention to the passive and the future constructions in German. This section extensively covers the theories brought forth by the scientific community on the emergence of both these constructions until today. In the fourth chapter, I discuss the corpora used for both Middle and Early New High German and address the methodology used to analyze the core data. I also address the methodological framework I use to approach them within a complex network framework. The fifth and the sixth chapters deal with the analysis of the instances from the corpora of Middle and Early New High German, which I present in various charts and describe in detail. The seventh chapter deals with the representation of the development of *werden* within a complex network framework. The eighth chapter provides a summary of the results and includes a discussion on the implications of this study and some directions for further research.

CHAPTER 2. THEORY

The current chapter discusses the chronological progression of several theoretical approaches towards syntactic change. First, this survey begins with those who considered language change as a sign of linguistic decadence. Then, it moves to those who favored an autonomous view of syntax in its synchronic dimension and disregarded any historical data. Finally, it concludes with those who perceive diachrony and synchrony as closely related and essential for a deep understanding of the meaning of linguistic structures. The goal of this chapter is twofold: 1) to demonstrate how a framework such as Complexity Theory better complements the analyses of syntactic changes in general and how it functions as a theoretical bridge between traditional linguistics and network science; 2) to describe the emergence of periphrastic constructions such as the passive and the future in more specific terms.

2.1 Syntactic Changes through the Centuries

Campbell (1998) noticed that, especially in the last decades, "the study of syntactic changes is an extremely active area of historical linguist" (p. 283). Nevertheless, through the centuries, different scholars have looked at the evolution of languages in various and, sometimes, divergent ways. In this chapter, I will offer a brief account of the major approaches to syntactic changes, starting with the Neogrammarians and going to generative accounts, before transitioning to the more recent functionalist and usage-based approaches.

According to Hock (1991), the idea of language change as "decay" and "corruption" was highly accepted among the early scholars of historical linguistic in the nineteenth century (p. 629). The access to written testimonies showed how languages had slowly lost part of their complexity on different levels, morphologically and syntactically, and how these losses were perceived as a decline and deterioration. For instance, German scholars considered Early New High German, which was spoken between 1350 and 1650, a corrupted and oversimplified version of the more prestigious Middle High German (Hartweg & Wegera, 2005, p. 22), dated between 1050 and 1350 and used in the lyrics of the *Minnesang* and heroic poems such as the *Nibelungenlied*.

It was when the so-called Neogrammarians started to gain credibility among scholars that the idea of language change as decay was finally set aside (Hock, 1991, p. 630) The

Neogrammarians (Ger. *Junggrammatiker)* were a group of linguists from Leipzig and became the most distinguished "and dominant force in linguistics science" in the nineteenth century (Bergs & Brinton, 2012, p. 255). Some believe that this group founded modern linguistics as we know it today, while others claimed that their approach did not differ much from that of their precursors (Einhauser, 2001, p. 1338). The Neogrammarian approach was mainly focused on sound changes, which were believed to be regular, gradual, imperceptible, and to affect the language in an equal manner without any exception (p. 1341). The other linguistic changes, such as those that affected the morphology and the syntax, were, on the contrary, believed to be irregular and abrupt. In 1879, Karl Brugmann and Hermann Osthoff, who were two of the main figures of the Neogrammarian group, listed methods and goals of their new approach to languages in the introduction to the first volume of *Morphologischen Untersuchungen* (Morphological Studies). Notably, Brugmann and Osthoff put emphasis on the need to focus on every-day language instead of considering only written samples. They also highlighted the importance of colloquial language and produced relevant scholarship on the dialectological landscape of Germany at the time (Auer & Schmidt, 2010).

Although many believe that the Prague School was formed as a reaction to the work of the Neogrammarians (Akamatsu, 2001, p. 1769), this group, which came into existence in the first years of the twentieth century, shares with the *Junggrammatiker* the attention to spoken and colloquial language. The scholars of the Prague School, who held their first official meeting on the 6th October 1926 (Ehlers, 2005, p. 1), can be seen as some of the main precursors of structural linguistics. This group was "mainly concerned with examining the question of why language structure is the way it is and with finding the explanation in language use" (Bischoff & Jany, 2013, p.1). The Prague School was largely influenced by some of the teachings of Ferdinand de Saussure, from which they took the view of the language as a system consisting of other linked sub-systems (Akamatsu, 2001, p. 1768). According to Vachek (1966), the Prague school was also functionalist in the sense that they believed that any element of the language solely exists to fulfill a specific function, which is mostly communicative in nature — hence, the large number of diachronic studies that this group produced during the years in which it was active (p. 7).

The work of Saussure, and especially his ideas of "arbitrariness" of the linguistic sign and of "detachment" of the diachronic to the synchronic dimensions of the language, also represented the theoretical framework from which structuralism drew its principles at the beginning of the

twentieth century (Givón, 2013, p. 12). It rose as a reaction to the functionalist approaches that had dominated the study of language until the moment and shifted the attention of the scientific community to the synchronic aspects of a language. According to the structuralist approach (in which we can include the American School of Bloomfield), although irregularities can be explained through historical processes, the language that exists in any given moment has to be considered as a system formed by the collective knowledge of the speaking community. Speakers, although not aware of the linguistic forms used in earlier periods, master the language anyway. Therefore, dealing with the history of the language is "irrelevant" to understanding the language itself (Matthews 2001, p. 14). As a result, the structuralists put the study of linguistic changes aside, focusing solely on the synchronic dimension of variation.

The generative grammar account, formulated by Noam Chomsky in his work "Syntactic Structure" and published in 1957, also neglects to consider the diachronic dimension of the language. However, Chomsky's contribution to the field of linguistics is still today one of the most significant. According to Allen (2017), "it is not an exaggeration to say that this little book [Syntactic Structure] caused a revolution among linguistics who took a structuralists' approach to language" (p. 42). According to the generativists, language changes "consist of the addition of optional rules at the end of the grammar" and "these optional rules reflect changes in linguistic fashion comparable to change in clothing (etc.)" (Hock, 1991, p. 642). Generativists also apply to sound changes specific syntactic rules under the notion of rule addition (p. 642). On the other hand, they associate syntactic changes with child language acquisition and with the transition from an older generation to a younger one. After learning the grammar as children, adults may later add other rules which make the syntactic structure of the language no longer optimal. Their children then, hearing this modified grammar, restore it making it more optimal (Campbell, 1998, p. 289). The most relevant work on diachronic generative syntax within this view of diachronic syntactic changes was compiled by Lightfoot in 1979, who viewed syntactic variations as a "special kind of reanalysis, or rather new analysis." According to Lightfoot, language learners perform a particular type of "parsing" during the process of language acquisition. When what learners acquire doesn't converge with the grammatical standard of a given speaking community, the discontinuity of transmission between generations is what generates diachronic linguistic changes. For this reason, change is considered abrupt instead of gradual (Madariaga, 2017, para. 3). Lightfoot's work, "Principles of diachronic syntax", laid the foundation for a productive body of research "within

formal linguistic studies," giving rise to numerous works and initiatives in the linguistic community, including collective reference handbooks and conferences (Madariaga, 2017, para. 3).

A couple of decades after Chomsky and Lightfoot 's works, the scientific community experienced another small revolution as William Labov published his well-known study of Martha's Vineyard. In this work, which brought forth renewed interest in linguistic change, he recognized that language variation is not random, but rather patterned (Gordon, 2013, p. 163). Labov also elaborated his "apparent-time" hypothesis, which states that comparing speakers of different areas can be used as a tool to examine specific linguistic communities within a diachronic perspective. In the Martha's Vineyard' study, Labov was also able to show how, firstly, sound changes originate in a reduced number of lexical items, and secondly, that changes on different levels, including phonetical, phonological, and morphological variations, are remarkably correlated to social factors (Hock, 1991, p. 647). His research highlights the key role of social attitude and social motivation for language change, fostering a new way to understand how speakers contribute to linguistic variations over time.

The strong correlation between languages and speakers has also been formulated by Hermann Paul, whom many consider the father of usage-based linguistics (Bybee & Hopper, 2001; Bybee, 2010). This approach towards grammar, in which speakers actively shape languages as they use it, laid the fundamentals for Emergent Grammar, a term coined by Paul Hopper in his famous presentation of the same title at the 1987 conference of the Berkeley Linguistic Society. Hopper borrowed the term from the anthropologist James Clifford, according to whom culture is temporal and emergent. In the same way, grammar is also temporal and emergent (Hopper, 1987, p. 3). In his paper, Hopper affirms that "the notion of Emergent Grammar is meant to suggest that structure, or regularity, comes out of discourse and it is shaped by discourse as much as it shapes discourse in an ongoing process" (p. 3). He also points out how "grammar is hence not to be understood as a pre-requisite for discourse" and how "its forms are not fixed templates but are negotiable in face-to-face interaction in ways that reflect the individual experience of these forms" (p. 3). The Emergent Grammar framework sees grammatical structures as a byproduct of communication and fosters an understanding of languages as dynamic entities subjected to continuous changes. Thus, speakers create new linguistic structures to communicate and express their intent and ideas, and they do so with linguistic material already available in the language. Emergent Grammar emphasizes the productive participation of speakers in understanding,

inferring, and producing intelligible new structures that will eventually grammaticalize. Hopper demonstrates how "speakers borrow heavily from their previous experiences of communication in similar circumstances, on similar topics, and with similar interlocutors" (1998, p. 158). In borrowing from past experiences, speakers adjust, adapt their previous knowledge to the situation, and engage in continuous cognitive efforts in recalling and applying learned strategies to new communicative contexts. Thus, the Emergent Grammar approach supports a view of languages that sees both languages and speakers as a whole and not as separate entities. Speakers and languages continuously interact with each other, triggering changes in different levels of the linguistic system. Emergent Grammar draws attention to the strict correlation between modern language use and language change, which, to the contrary of what structuralists and generativists assume, have to be considered not as separate, but rather, closely related to each other.

The eighties were also the years during which two other revolutionary linguistic approaches came into existence. On the one hand, the renewed attention to morphosyntactic changes from a diachronic perspective led to the rise of studies on grammaticalization (Heine, 2003). On the other hand, the works of scholars such as George Lakoff, Ron Langacker, and Len Talmy laid the foundations for Cognitive Linguistics. The studies of the first type investigate grammaticalization as a linguistic change that refers to the process whereby lexical items and linguistic constructions acquire new grammatical meanings and functions, and how new grammatical functions arise from already grammaticalized forms (Hopper & Traugott, 2003). According to Heine (2003), languages are historical products and should, therefore, be analyzed with a strong focus on the historical processes that are responsible for their current structure. Hence, findings based on the process of grammaticalization offer more comprehensive explanations than findings confined to a mere synchronic analysis. Furthermore, the development of grammatical categories is unidirectional, from concrete/lexical to abstract/grammatical meaning (p. 577). Grammaticalization can, for instance, be found in the emergence of the future marker "will" in English. This verb, which had the original meaning of "want" and "desire", acquired a new function over time that indicated future temporal references (Bybee & Pagliuca, 1987, p. 112).

Cognitive Linguistics considers languages a tool for organizing, processing, and communicating information (Geeraets & Cuyckens, 2007, p. 3). This framework looks at grammar as an "essential aspect of the conceptual apparatus through which we apprehend and engage the world" (Langacker, 2008, p. 4) and as the conceptualization of the human mind (Croft & Cruse,

2004, p. 1). Speakers produce new lexical and grammatical structures to give "material" forms to their thoughts and their intentions. According to this framework, the cognitive processes behind language use are essentially the same processes that govern other nonlinguistic cognitive abilities (p. 2). Although Cognitive Grammar was first applied to synchronic data, the first diachronic works within this framework can also be traced back to the early eighties. From these studies emerged relevant theories such as prototype theories and categorization around prototypes, which are still dominant today in research in Cognitive Grammar (Winters et al. 2010, p. 7).

The idea of grammar as "emergent" and the renewed interest in the connections between synchronic and diachronic analyses is shared also among those scholars who directed their attention to the role of "cultural transmission" in language evolution. Cultural transmission is "the process of language adaptation in a community via communication of individuals from the same or different generations" (Gong & Shuai, 2016, p. 237). Recent works within this approach (Christiansen & Kirby, 2003; Labov, 2001; Mufwene, 2001) have suggested that interactions between speakers of different age, sex, cultural background, social status, and generations may have a significant influence in shaping particular grammatical structures and creating new vocabulary. Such a generational aspect of language change that addresses "interaction", "transmission", and "shaping of linguistic structures" emphasizes a view of language change as a dynamic process in which speakers play a highly significant role in determining the evolutionary path of a given language and its syntactic structures.

A remarkably similar view on language change is also shared by Complexity Theory. Bybee (2010) states that "when a linguistic structure is viewed as emergent from the repeated application of underlying process, rather than given a priori or by design, then language can be seen as a complex adaptive system" (p. 2). Complexity Theory offers new perspectives to apply to theoretical, applied, and historical linguistic research, triggering a change in the way human languages are conceived. Like Emergent Grammar, Complexity Theory sees languages as continuously evolving systems that develop through usage and repetition. According to Larsen-Freeman and Cameron, "from a Complexity Theory perspective, a language, at any point in time, is the way it is because of the way it has been used" (2008, p. 80). Furthermore, Complexity Theory also accentuates the dynamic nature of human languages and considers linguistic patterns as "epiphenomena of interaction," emphasizing the essential roles of the agents and their interactions with each other. Speakers and communication are thus the leading forces of language change and

evolution. In this regard, and in line with the Cognitive Grammar approach, scholars who assume the tenants of Complexity Theory highlight that "the history of a language reflects the behavior of its speakers" (Larsen-Freeman & Cameron, 2008, p. 91). Nettle (1999) points out that the structure of language has emerged from the kind of message speakers wish to convey and the kind of cognitive, perceptual, and articulatory mechanisms they have to express them, either by biological evolution, cultural evolution, or more likely by some combination of the two (p. 13). Thus, Complexity Theory draws attention to the strong connection between speakers and languages and how the first influences the second, and vice versa. Larsen-Freeman and Cameron note that "language emerges upwards in the sense that language-using patterns arise from individuals using the language interactively, adapting to another's resources. However, there is reciprocal causality, in that the language-using patterns themselves, downwardly entrain period patterns" (2008, p. 80).

A Complexity Theory approach combines synchronic analyses with diachronic ones because "language change is not just a peripheral phenomenon that can be tacked on to a synchronic theory; synchrony and diachrony have to be viewed as an integrated whole" (Bybee, 2010, p. 105). Similarly, Beckner et al. (2009) emphasize the strict connection between language change and communication, noting how "language has a fundamentally social function. Processes of human interaction along with domain-general cognitive processes shape the structure and knowledge of language" (p. 1). Languages, therefore, "emerge from the verbal interaction among humans" (Lee, Mikesell, Joaquin, Mates & Schumann, 2009, p. 3) and their most fundamental features are biological adaptations for cooperative social interaction in general" (Tomasello, 1999, p. xi).

Because languages are emergent complex adaptive systems (Bybee, 2003, 2006, 2010; De Bot, 2009; Larsen-Freeman & Cameron, 2009; Beckner et al., 2009) and because they can be "represented as a complex network in [their] several levels of complexity" (Amancio, 2014, p. 1), different scholars have started analyzing various aspects of languages using a complex network account, such as phonetics, phonology, morphology, lexicon (Choudhury & Mukherjee, 2009), syntax (Cancho & Solé, 2001), semantics (Steyvers & Tenenbaum, 2005; Vitevitch, 2008), and stylistic variations (Cong & Liu, 2014; Ke, 2007). According to Beckage and Colunga (2016), the strength of a complex network approach to the study of languages "lies in the fact that one can formally study a system of objects (nodes) and relations (edges)" (p. 4). Adversely to linguists' traditional focus on languages as an object, this framework allows "for the interactions of objects

to be as important as the objects themselves, allowing the system, as well as the constituents of the system, to be studied" (p. 4).

On the one hand, these studies show the value of the applications of a complex network account for the analysis of languages, but on the other hand, they mainly focus on the synchronic representation of linguistic features. In fact, very few scholars address historical linguistic change within this framework. One of the few is the study by Ke et al. (2008), in which they apply computational modeling to simulate the spread of innovations in a speaking community using different network typologies to represent the population structure. Their findings show how the rate of diffusion of linguistic innovations changes depends on the network type. In random and scale-free networks, the spread of innovations is indeed lower than in regular and small-world networks and follows an S-curve trajectory (Hickei, 2003). Another relevant study is the one by Perc (2012). The focus of his study is on the most common English words since the beginning of the 16th century. The analysis shows how these words retained their top rank over extensive periods of time despite new entries in the lexicon. Perc also looks at lexical co-occurrence and demonstrates how these words have also increased the number of elements with which they could be combined. Concu (2015) also looks at historical changes using a complex network framework, providing preliminary evidence for preferential attachment in the historical development of the German *Perfekt* (present perfect). She compares the instances of this structure in the *Lay of the Nibelungs* (ca. 1200), a long heroic poem in Middle High German, and *The Sorrow of Young Werther* (1774), one of the most famous novels of Wolfgang von Goethe. She observed that, although the overall use of the *Perfekt* was very similar in both texts in terms of frequency, the types of verbs used in Goethe's novel that appeared in this form were significantly greater than the ones in the Middle High German poem. Furthermore, this study shows that the growth of the German *Perfekt's* "connectivity" reduced the number of instances of the past participle, which was widely used as an independent unit in Old High German (Kotin, 2009). Bentz and Buttery (2014) also focus on linguistic changes using computer modeling to simulate the process of grammaticalization, and specifically, how some word forms, when they frequently co-occur, tend to merge in a single unit over time. Their findings show how it is also possible to apply a mathematical algorithm to represent historical linguistic change. Another study worth mentioning is the one by Thurner, Hanel, Liu, & Corominas-Mutra (2015). They propose a sample-space-reducing model to understand the existence of Zipf's law in word frequency, analyzing the

sentence formations from ten different English books written between the 17th and 19th centuries. In this study, they claim that "the usage of specific words in a sentence highly restricts the usage for consecutive words, leading to a nesting (or sample-space reducing) process" (p. 2). Zipf's law then can be seen as a direct result of this "nesting" selective process. The idea of a restrictive- and semantic-motivated space from which speakers choose while conveying information can account for the frequency distribution of certain nouns and verbs. However, it fails to acknowledge that, in a language's history, processes such as grammaticalization (and the related semantic bleaching), have also very often contributed to the current frequency patterns but in the exact opposite way. The weakening of semantic features of the verb "can" in the Middle English period, for instance, allowed its use with a continuously increasing number of verbs. This change amplified its "sample-space" and made it one of the most frequent verbs in Modern English (Bybee, 2003).

The findings of the aforementioned studies suggest that the application of a network framework can be very fruitful and worthy of more extensive research. More specifically, Ke et al. (2008) indicate that scale-free networks, in particular, are suitable to represent the dynamics of the spreading of innovations. Perc (2012) and Concu (2015) provide evidence for preferential attachment in regard to the lexicon and the emergence of new grammatical structures. Thurner et al. (2015) propose an interesting framework for the understanding of the frequency patterns of certain categories of words in modern English

2.2 Summary

The previous section has offered a brief account of the main approaches to syntactic changes. The generalist approach, as discussed in the previous pages, disregards any attempt to connect modern languages to their historical development and misses the importance of the speaker's role in the creation of linguistic forms. Although one cannot diminish how the scholarship produced within this account had and still has a huge impact within the linguistic scientific community, its sole focus on synchrony inevitably fails to produce works that go beyond mere descriptions. That is not to say that their contribution has to be completely put aside, and this dissertation will use dependency grammar to build the syntactic networks used in the network-related part of the analysis. However, the narrow scope of generative grammar cannot account, I believe, for the complexity of language variation, even within a synchronic perspective. Conversely, Emergent Grammar and Complexity Theory consider both the diachronic and the

synchronic aspects of languages as they interact with each other. Complexity Theory, in particular, acknowledges the dynamic nature of language, restores the significance of historical analyses and provides a new framework to examine at languages. Complexity Theory functions as a theoretical bridge between linguistics and the science of complex networks. These frameworks can, indeed, address both the dynamics and complexity of languages, possibly shedding new light on the nature of historical linguistic changes.

Until this point, I have highlighted the suitability of Complexity Theory and network science for the research carried out in this work but, what exactly are complex systems and how are they related to complex networks? In the next section, I attempt to answer these questions, establishing the theoretical foundations of these approaches. This section begins with an outline of the general features of complex adaptive systems and a discussion of the terminology used in Complexity Theory. Next, I provide a framework for the definition of "language" as a complex system and define the targeted grammatical constructions I analyze in the current work within the same framework. Then, I describe complex networks and how they can be employed for the study of language evolution in general, and to the rise of the passive and future periphrases in German in more specific terms.

2.3 The Science of Complexity

Complexity Theory studies the behavior of unique systems in which a "large number of relatively simple entities organizes themselves, without any benefit of any central controller, into a collective whole that creates patterns, uses information, and in some cases, evolves and learns" (Mitchell, 2009, p. 15). Examples of such systems are insect colonies, the brain, the immune system, economies, and the World Wide Web (p. 16). The internet, for instance, is formed by a significant number of pages that are highly interconnected to each other with no controlling entity that guides their actions at any given time. Hence, the current structure of the internet can be considered emergent as the result of self-organization. Such systems are also called "complex" but, according to Larsen-Freeman and Cameron (2008), the word "complex" doesn't mean merely complicated. Instead, "complex" refers to the fact that "a system's behavior emerged from the interactions of its components" (p. 2). As such, a complex system is defined as "a system in which large networks of components with no central control and simple rules of operation give rise to

complex collective behavior, sophisticated information processing, and adaptation via learning or evolution" (Mitchell, 2009, p. 25).

Complexity Theory originated in biology, mathematics, and physics, but it can be traced back to the time of Aristotle (384 - 322 B.C.E.) and his theory of motion (Mitchell, 2009, p. 27). The interest in dynamics and motion was also shared by another two figures who significantly contributed to the emergence of the study of complexity: Galileo Galilei and Isaac Newton (p. 29). However, the first example of a complex system was given by the French mathematician Henri Poincaré. He discovered sensitive dependence on initial conditions when he was trying to predict the motion of a hurricane using Newton's dynamic theory (p. 32). This discovery brought to the formulation of one of the main principles of the science of complex systems: the principle of "nonlinearity", according to which a system can be described by the maximum that the whole is greater than the sum of the parts (p. 33). Non-linearity refers to an emergent behavior that is often "disproportionate to its causal factors" (Larsen-Freeman & Cameron, 2008, p. 2). A major change in the system could be, in fact, the result of the slightest pressure; on the other hand, a large amount of energy could provoke little or no changes at all to the system.

Another key feature of complex systems is self-organization, a process whereby "order is created out of disorder" (Mitchell, 2009, p. 40). The states of order and disorder in Complexity Theory are measured through "information" and "computation" (Mitchell, 2009, p. 51). The first concept refers to the message that any unit can send (Mitchell, 2009, p. 64). Computation is the process that indicates both the analysis and the storage of information (Mitchell, 2009, p. 66). Another two key notions are "evolution" and "adaptation," which are strictly related to each other, as the first can be a direct consequence of the second and vice versa.

Larsen-Freeman and Cameron (2008) list further key features that are common in complex systems. One of those features is "heterogeneity" which refers to the elements, agents, or processes which are integral parts of a given system. These can be, indeed, of different types, share similar functions or carry out completely different duties and, often, can be complex systems themselves, or "subsystems of the bigger system" (p. 28). Cities, for instance, consist of different subsystems such as roads, city planning, and schools. Another relevant feature of complex adaptive systems is "openness", which is understood as the fact that changes in the system can come from forces both inside and outside of it. Such a characteristic guarantees both a system's structural adaptability and stability, although stability here has to be understood as "stability in motion" or as "dynamic

stability" (p. 32). The context in which complex systems operate must also be taken into account, since it is able to influence the system but can also be influenced in turn by the latter.

According to Mitchell (2009), changes in complex systems can be of different kinds (smooth or radical, for instance), and can bring the system to a different state or account for different behaviors patterns (p. 44). These shifts happen in the so-called "state space" of the system, which is the landscape of the possible states toward which a system can move (p. 47). A preferred state in the landscape is called an "attractor". This is responsible for the (dynamic) stability and variability of a complex system. Attractors can be fixed, cyclic, and chaotic. While a fixed attractor represents a preferred spot in the state space in which the system tends to stay, a cyclic attractor refers to a state toward which the system moves periodically, changing after a variable amount of time. The last type of attractor causes an unstable behavior within the system and increases its sensitivity to even the smallest change (Larsen-Freeman and Cameron, 2008, p. 57). Independent from which type of attractor has caused the system to shift from a state to another in its landscape, the new shape it acquires is the direct result of internal self-organizing processes that can affect the structure of the system in two different directions: upwards and downwards. The changes at a lower level cause further changes upwards at higher levels of the system. In the same way, changes in the upper level push downwards for further shifts. The new state is substantially different from the previous one and it can, therefore, be considered emergent (p. 59). As such, emergence through self-organization is one of the most peculiar features of complex systems that distinguishes them from other types of systems. The description of complex systems as presented here raises a relevant question for the aim of this study: how can languages fit into this description of complex adaptive systems?

Some of the aspects that enable us to consider language as complex systems are relatively easy to identify, such as the agents, which are represented by the speakers, and the operational environment of the system, which overlaps with the speech community in which a given language is spoken. Others, such as the system's heterogeneity, its internal organization, and adaptive mechanisms, and its emergent behaviors, are less straightforward. To better understand how languages fit the description of complex systems, it is useful to look at what other scholars have discussed in this regard. As noted in Table 1 below, Larsen-Freeman and Cameron (2009) offer some useful examples of complex systems in applied linguistics, namely 'spoken interactions' and 'classroom language learning':

Table 1: Examples of complex systems in applied linguistics (Larsen-Freeman and Cameron, 2008, p. 37)

Field	Spoken Interaction	Classroom language learning
Agents	Speakers, their language resources	Student, teacher, languages
Heterogeneity	Speaker background, style, discourse topics	Abilities, personalities, learning demands
Organization	Dyads, speech communities	Class, group, curricula, grammars
Adaptation	Shared semantics, pragmatics	Imitation, memorizing, classroom behaviors
Dynamics	Conversation dynamics, negotiation of understanding	Classroom discourse, tasks, participation patterns
Emergent behavior	Discourse events, idioms, specific languages, e.g. "English"	Language learning, class/group behavior, linguae francae

The two examples in the table share a common range of components that clearly illustrates the complexity and the compositional nature of each system. The description of languages as complex adaptive systems used in this work builds on such representations and shares the same definition of the agents as the speakers of a given community and their language resources. For what concerns the categories "heterogeneity", "organization", "adaptation", "dynamics", and "emergent behavior", however, the view of the language as a complex adaptive system entails a larger number of components and takes into account every type of interaction or communicative event that takes place in written and multimodal forms. The agents of the system share not only background, style, and discourse topics, but also similar abilities to understand and interpret different types of information and communicate using means other than spoken language.

With "organization" it is understood that language systems are organized around social spaces, such as families, schools, and workplaces, and associative groups (political, religious, sport-related, etc.). With "adaptation", we are referring to those means shared in and recognized by a given speech community not only for oral interactions but also for written and multimodal interactions. These means are, for instance, the group of symbols that constitute the alphabet and the platforms that account for the transmission of knowledge and information, including books, articles, newspapers, magazines, blogs, and posts on the internet. As far as the category 'dynamics' goes, conversations and negotiations of meaning in oral interactions need to be placed side by side with those that develop when the agents exchange information and communicate in forms other than oral speech —for instance, the exchange of text messages between two people trying to find a time to meet or the exchange of documents between two business partners finalizing a transaction.

In this category, we also find types of communication in which speakers are not directly involved in the process and in which they are passive recipients of the information. Typical examples of such communicative activities are advertisements, movies, and books, in which information is presented following specific dynamics. Next, the concept of "emergent behavior" involves the outcome of the various processes of interactions and adaptations within the system that lead to relevant linguistic changes over time, such as the rise of new grammatical constructions and vocabulary. Larsen-Freeman and Cameron (2009) define such outcomes as the rise of "form-meaning-use dynamic patterns of language use" because such term allows both to capture their variability and stability and to emphasize their unpredictability and their compositional nature (p. 82).

The following chart illustrates the theoretical model of "language as a complex adaptive system" applied in the current study. As the aim of this study is to capture the development of *werden* from Middle to Early New High German, I first frame the view of the language as a complex adaptive system in general terms, and secondly, I define the targeted research object of the current work.

Table 2: The representation of language as a complex adaptive system

Field	Language
Agents	Speakers, their language resources
Heterogeneity	Speaker background, style, discourse topics, similar abilities to understand information, and communicate
Organization	Speech communities, social entities, and associations
Adaptation	Shared alphabet, semantics, pragmatics, platforms for communication, interaction, and spread of knowledge
Dynamics	Discourse dynamics in oral, written, and multimodal communication
Emergent behavior	Form-meaning-use dynamic patterns of language use

Table 2 displays the components of the complex adaptive system "language" as intended in the current work. This research aims to explore those "form-meaning-use dynamic patterns of language use" which had *werden* as a central element. The analysis will be carried out within specific forms of written prosaic texts (mostly religious) produced during a time period that can be defined as a part of the medieval and Early German society. The speakers shared language resources, symbolic means, and abilities for the production and the reception of such texts. In order to capture such emergence, this study uses two separate sources, one for Middle and one for Early

New High German. The tools of complex networks are used to address historical syntactic networks because, as claimed by Barabási, "behind each complex system, there is an intricate network that encodes the interactions between the system's components" (2016, chap. 1). Thus, the implementation of a complexity theory account for the exploration of languages and language evolution shapes part of the methodological choices in the current work. However, there is another factor that further motivates the implementation of network science to explore syntactic changes: the power law degree of distribution.

2.4 Power Laws and Complex Networks

Power-laws are "a signature of complex systems" (Larsen-Freeman & Cameron, 2008, p. 111). Such laws were first observed by the American linguist George Kingsley Zipf, who explored the word frequency distribution in English, comparing the most frequent words in 'Alice's Adventures in Wonderland' and in the 'Chemical Pathology of the Liver.' Although different, both works showed very similar patterns: throughout the text, a small number of words were used with a high frequency, while the majority of them were used only once or twice. Because of these findings, Zipf formulated the famous law which still bears his name: the occurrence of a word in a given text is inversely proportional to its rank in the frequency table (Li, 1992). Around the same time as Zipf's discovery, similar frequency distributions were also observed in areas other than linguistics (Barabási, 2002). For instance, Vilfredo Pareto, an influential Italian economist, investigated Italy's land distribution and noticed that 80 percent of the territory was owned by only 20 percent of the population. Pareto's rule was later turned into the Murphy's Law of management when scientists noticed striking similar distributions in the world of business: 80 percent of profits were produced by only 20 percent of the employees, 80 percent of customer service problems were created by only 20 percent of the customers, and 80 percent of decisions were taken during 20 percent of the meeting time (p. 66). Watts (2003) also found analogous patterns examining book sales, actors, and movie popularity: "… for every Harry Potter and Blair Witch Project that explodes out of nowhere to capture the public's attention, there are thousands of books, movies, authors and actors who live their entire inconspicuous lives beneath the featureless sea of noise that is modern popular culture" (p. 207). Power laws can thus be considered very common distributional parameters not only in languages but also in many areas of social and natural science. According to Sihna and Pan (2006), power laws also account for city sizes, scientific paper

citations, language speakers, companies' dimensions, and internet traffic; they are also attested in bibliometrics, informetrics, scientometrics, and library science (Saichev, Malevergne, & Sornette, 2009). Barabási and Albert (1999) discovered the same degree of distribution looking at the organization of the pages on the internet. Similar to Zipf's findings, Barabási and Albert noticed that just a small number of pages had a larger number of connections, while the rest (also the majority) had only a few.

The discovery of power-law structures in both languages and large networks triggered the rise of a new interdisciplinary approach within the complex system research community, namely, network science. As the name suggests, this discipline studies a wide range of networked systems from different fields, including networks encoding the interactions between genes and proteins, networks capturing the connections between neurons in the brain, networks formed by social relationships, and networks that describe how communication devices interact with each other (Barabási, p. 24). Although very different, these networks have significantly similar structures and are governed by the same organizing principles (p. 25). This discovery enabled the use of network science for the study of different phenomena, which led to major achievements and new insights on relevant issues that traditional approaches weren't able to offer. The representation of the population's structure as a network, for example, has enabled the identification of influential nodes and hubs inside communities. These discoveries turned out to be of strategical importance for more effective immunizations and a more efficient control of the spread of diseases (Ke, 2007, p. 4). The fight against terrorism has also benefited from the implementation of tools of network science, with more targeted and decisive responses to terrorist activities (Barabási, p. 31).

Viewing languages as complex adaptive systems has also allowed researchers to apply network approaches to understand different linguistic phenomena. Among the linguistic networks that scholars have developed for the analysis of linguistic properties, (for instance, word association networks, word co-occurrence networks, and semantic networks), syntactic relationships have been explored within syntactic dependency networks. In these networks, which rely on the formalism of dependency grammar, the links between nodes represent syntactic dependencies (Ferrer I Cancho, 2005). As most of the links go from a head to a modifier, these are usually direct, which means that they often go in only one direction. In terms of their properties, Ferrer I Cancho notices the presence of small world phenomena: despite the large number of nodes, the distance between them is notably small (p. 65). Syntactic dependency networks also display a

degree distribution that takes the form of a power law, as the majority of the vertices have a few connections, while a small number of them has a significantly high number of links (p. 66). In the present study, the development of *werden* will be analyzed using syntactic dependency networks which are constructed using a set of rules that I describe in detail in the methodology section.

As syntactic dependency networks share some of the most peculiar properties with other complex networks, such as small world structures and a degree of distribution which follows a power law, it is also plausible that they will be subject to similar (if not the same) patterns of growth. Complex networks such as the World Wide Web increase their size when new nodes join the system. In the same way, a language network expands when new words make their way into the system. However, it has been demonstrated that, for instance, new nodes of the internet are not attached randomly. They exhibit preferential attachment, such that the likelihood of connecting to a node depends on the node's degree (Barabási & Albert, 1999). For instance, a new web page will be more likely to be connected with already largely popular sites, such as Google, YouTube, and Yahoo. Consequently, popular pages like these tend not only to maintain their high connectivity status but also to increase their connection's degree over time. For this reason, preferential attachment was also called the Rich-Get-Richer Phenomenon, in the sense that already "rich" pages will constantly become "richer" and connect to a continuously increasing number of other pages. Easley and Kleinberg (2010) suggest that the connectivity or popularity of a given node "should grow according to the same rule that governs the growth of bacterial colonies and compound interest: a page's popularity grows at a rate proportional to its current value, and hence exponentially with time" (p. 548). The discovery of this phenomenon was a turning point in the study of complex networks and their evolution. This work aims to gather evidence of preferential attachment in the growth of vertices in syntactic networks and focuses on a particular node, *werden*, and its "connectivity" increase over a specific time frame, namely from Middle High (from 1050 ca. to 1350 ca.) to Early New High German (1350 ca. to 1650 ca.).

Implementing a complex network account for the analysis of linguistic properties enables the use of measurements that were originally developed for the study of systems outside of linguistics. Centrality measures are among the most relevant tools in network science for determining the role of a given node in a particular network and enable tracking the "quality" of the in- and out-degree connections of the nodes (Matas, Martincic-Ipšic, Meštrovic, 2017). In the next section, I define the measurements that are used to gather evidence of preferential attachment

in linguistic terms, namely, degree centrality, betweenness centrality, eigenvector centrality, and eccentricity centrality. These measurements will be applied to all the nodes in the networks.

2.5 Centrality Measures

2.5.1 Degree Centrality

In the present work, degree centrality, which measures the number of links in a node (Matas, Martincic-Ipšic, Meštrovic, 2017), refers to the "valency" property of a specific verb. This property refers to the range of "complementation patterns that different verbs can be found with as well as these patterns which do not seem to be available" (Faulhaber, 2011, p. 1). A valency pattern can be defined as the simultaneous choice of one or more complements that are used with and determined by a verb functioning as a valency carrier (p. 6). Consider sentence (5):

(5) Anna liest oft Bücher in ihrem Zimmer
Anna reads-PRS often books-ACC in her room-DAT
'Anna often read books in her room'

The verb *lesen* (to read) is connected to three different complements: the temporal adverb *oft* (often), the direct object *Bücher* (books), and the prepositional phrase *in ihrem Zimmer* (in her room). The meaning of the verb "to read" governs the selection of such complements and directly influences its valency properties. If we substitute "to read" with a verb with different valency properties such as *schlafen* (to sleep), we notice that, while the adverbial and prepositional complements are still available, a direct object is not:

(6) *Anna schläft oft Bücher in ihrem Zimmer
Anna sleeps-PRS often books-ACC in her room-DAT
Anna often sleeps books in her room

Hence, "to read" and "to sleep" have different valency properties, since they have the capacity to appear with distinct complements. Some of those complements are not available for a particular group of verbs, such as a direct object for the verb "to sleep". Hence, the verb *lesen* will display a higher degree than the verb *schlafen*. In this work, this measurement will refer to the number of links that *werden* is able to establish during the centuries. Since the syntactic networks

are direct graphs, the means that the connections only go in one direction, the focus will be on the out-degree, meaning the connection going from *werden* to other nodes.

2.5.2 Betweenness Centrality

The second measurement used in this work is called 'betweenness centrality' which measures how often a node is on the shortest path in a network. It has been shown how vertices with high betweenness centrality may have considerable influence within a network by virtue of their control over information passing between other nodes (Matas, Martincic-Ipšic, Meštrovic, 2017). Consider again the following sentence:

(7) Anna liest oft Bücher in ihrem Zimmer
Anna reads-PRS often books-ACC in her room-DAT
'Anna often read books in her room'

According to *Gephi,* [1] which is one of the two software for network analysis and visualization used in this dissertation, the nodes with the highest betweenness centrality values are the noun *Zimmer* and the preposition *in*. This means that, when visualized as a network, these nodes appear most frequently in the shortest path among all the other nodes in the sentence. For instance, the preposition *in* appears in the shortest paths from *Zimmer* and *ihrem* to the other nodes of the sentence. In the syntactic networks used in this dissertation, betweenness centrality refers to how many times a specific node has served as the shortest path in the network among elements such as nouns, pronouns, prepositions, adjectives, adverbs, and other verbs.

2.5.3 Eigenvector centrality

Eigenvector centrality considers not only the number of the connections of a particular node (degree centrality) but also the "quality" of those connections, that is if a node is connected to other important nodes. It indicates a node's "influence" within a network and gives higher scores to those nodes that are central and are connected to other nodes with high degree centrality. This measurement is based on the assumption that, in a network, the connections established by a particular node are not equal. Consider again the following sentence:

[1] https://gephi.org.

(8) Anna liest oft Bücher in ihrem Zimmer
Anna reads-PRS often books-ACC in her room-DAT
'Anna often read books in her room'

Gephi shows that in this sentence, when visualized as a network, the elements with the highest eigenvector centrality values are the noun *Zimmer* and the preposition *in*. While degree centrality considers all the connections, and it is a good indicator of the role of nodes in the network, eigenvector centrality also looks at the type of connections and discriminates between links. This measurement is particularly useful for the scope of this research because it allows us to gain a better understanding of the development of the connections in the syntactic networks over time. Specifically, this measurement will address the type of connections that are the most influential in Middle and Early New High German.

2.5.4 Eccentricity Centrality

Eccentricity centrality is another common measure to analyze the nodes in a network. In a biological network, it can be described as how easy is for a protein to be reached by all other proteins in the network. For this reason, a protein with high eccentricity could be more easily influenced by the activity of other proteins or could easily influence the other proteins (Scaldoni & Laudanna, 2012). In a syntax network, this measurement indicates which grammatical elements are the most easily reachable by the others, thus, easy to establish connections with. Consider again the following sentence:

(9) Anna liest oft Bücher in ihrem Zimmer
Anna reads-PRS often books-ACC in her room-DAT
'Anna often read books in her room'

Gephi assigns the highest eccentricity values to the verb *lesen*, meaning that it is the easiest node to be reached by all the other nodes in the sentence. This verb is attached indeed to the noun *Bücher*, the adverb *oft*, and the prepositional phrase *in ihrem Zimmer*. Including this measurement in the analysis, will illustrate if *werden* is among those nodes in the network that are the easiest to establish links with. As the goal of this dissertation is to find evidence for preferential attachment, this measurement can be useful to understand if this "easiness" in establishing connection plays a role too in the growth of connections of a particular node.

2.6 More on Preferential Attachment

As previously mentioned, preferential attachment is one of the most common processes in the growth of networks. Its discovery represented a small revolution in the field of complex systems and enabled the scientific community to gain more insightful information for the study of the evolution of networks. However, this finding raised significant issues concerning the nature of such growth. As preferential attachment refers to the process whereby a small number of highly connected nodes increase their connectivity, it is conceivable to investigate how such nodes gain their status in the first place. It is also useful to inquire about how these nodes continue acquiring new links over time. For instance, how could Google become so popular, although it joined the internet relatively late in comparison to other search engines, such as Alta Vista and Inktomi (Barabási, 2006, p. 94)? And how does it steadily acquire new links? Very similar questions could be raised, for instance, about the emergence of the periphrastic passive and future constructions in German. The review of the literature will address how the passive started increasing its lexical base in the last centuries of the Old High German period, constantly growing the number of verbs that could appear in this form. It will also address how *werden* as a future marker had to compete with the very frequent modal constructions with *wollen* (want) and *sollen* (should) for the expression of future time references, eventually becoming the only means to carry out such a function. The affinities between the success of Google, the growth of the passive, and the rise of the future periphrasis could suggest that these apparently unrelated phenomena respond to similar underlining processes.

The emergence of preferential attachment has caught the attention of different scholars in various disciplines. Easley and Kleinberg (2010) relate the rise of this phenomenon to the influence of so-called information cascades. Information cascades refer to the process by which, in a given community, people are "aware of earlier decisions made between two alternatives (e.g. accepting an idea or rejecting it)" (p. 550) and make their decision accordingly. From this perspective, popular choices are what drives the emergence of preferential attachment. However, this approach does not address why particular elements were chosen over others in the first place. In the year 2000, for instance, Yahoo suddenly fired Inktomi as its search engine and decided to replace it with Google, which was only two years old at the time (Barabási, 2014, p. 93). Bianconi and Barabási (2001) used quantum mechanics and the Bose-Einstein condensation models to

understand why latecomers such as Google turned into the biggest and most influential nodes in the World Wide Web, despite their late arrival.

They also noticed that such capacity is not limited to the internet. Some people are, indeed, better than others in turning a random meeting into a very successful social event. Some papers published only recently acquired a high number of citations in a very short time, and new websites can sometimes take over older ones (p. 441). This capacity must be an intrinsic property of that particular paper, person, or website, which contributes to its success in the network. For this reason, Bianconi and Barabási (2001) decided to assign to every node a specific fitness value, which refers to the ability to "make friends" (to gain more connections) in comparison to everybody else. The assignment of different values to each vertex implies, firstly, that a complex network is a competitive environment and, secondly, that not all the nodes in a network are equal. "Fitter" nodes will thus be able to overtake older ones, even those that are very well connected but less fit.

The uncovering of the Fitter–Get-Richer phenomenon by Bianconi and Barabási sheds light on the very nature of complex scale-free networks such as the World Wide Web and provides insight into the way they grow. One of the most interesting aspects of the Fitter–Get-Richer model is that “fitness is not assigned by any individual but reflects the network’s collective perception of a node’s importance relative to the other nodes. We can, therefore, determine a node’s fitness by comparing its time evolution to the time evolution of other nodes in the network” (Barabási, 2016, p. 208). The cognitive dimension behind the assignment of the fitness value to a specific node and the relevance of its historical development highlights the members’ active participation in the internal organization of the network. In the view of a language as a complex adaptive system, the concept of the fitness value could be also explored when dealing with syntactic changes. It is very common in the history of languages, indeed, that some grammatical forms become used more frequently and become the main means by which to convey a particular meaning or the fulfillment of a particular function, despite their late emergence. As previously mentioned, many scholars have pointed out how *werden* as a future marker had to compete with the modal verbs *wollen* and *sollen* to express future time references. Eventually, *werden* became the only future marker in German.

Implementing a network account for the analysis of the rise of the passive and the future in German could provide some insight into the existence of a fitness value in syntactic dependency networks and provide some directions for further research on this topic. Nonetheless, any valid

attempt to describe what the fitness value is in linguistic terms must also take into account the cognitive processes behind the collective behavior of the speakers. This leads to the perception of a node's importance in relation to the other nodes. Among the different approaches, cognitive linguistics (henceforth CL) deals with the cognitive dimension of languages, and it is, I believe, the most suitable for this purpose.

2.7 The Cognitive Dimension of Languages

Cognitive Linguistics is a part of the functionalist approach that was described in the first part of this chapter. It views grammar as a linguistic symbol in which meaning is paired with form. Lexicon and grammar are not strictly separated, since both are the results of symbolic assemblies that "have achieved the status of conventional units" (Langacker, 2008, p. 22). Grammar, specifically, rather than being "autonomous resides in schematized patterns of conceptual structuring and symbolization" (p. 27). The meaning of such linguistic expressions can be found in both discourse and social interactions, as well as in the internal conceptualizations in the brain (p. 29). Since grammar is symbolic, grammatical constructions can be considered as an "assembly" of such symbols, and the elements that constitute them are put together following constructional schemas (p. 167). The meaning of such composition is not the mere sum of the parts, but it is an "integrated structure" in which the elements are related to each other in particular ways (p. 168). When a constructional schema is used to form a particular expression, it relates to that expression through a categorizing relationship. Usually, an expression conforms to a specific schema through conventional ways that determine its grammaticality. When an expression does not follow such conventional constructional schema, then it is ungrammatical (p. 170).

Cognitive Linguistics also supports a view of semantics as conceptual instead of truth-conditional. In this sense, the conceptual patterns of a language that are activated can be considered semantic prototypes. A prototype is the best example of a category, although it does not necessarily display all the defining features of that particular group (Arnett, 2004, p. 32-33). Coleman and Kay (1981) define the prototype schema as containing a finite list of properties that are gradient. In general, and this is also true for extra-linguistic categories, the existence of prototypes can be explained through "categorization", which takes place by comparing a given entity to the prototype (Arnett, 2004, p. 33). In the case of linguistic prototypes, categorization is better described as the "process by which people, in using language, are categorizing their experience of the world" (Tyler,

2003, p. 6). Prototypical categories are in this sense "experiential" (Lakoff, 1987) since they further indicate how "the epistemological relationship between concepts and the world rather than the structural characteristics of those concepts." Prototype experiential categories are flexible because there are not rigidly defined boundaries and this uncertainty allows the incorporation of members which do not possess all the definitional feature of a given category (Geeraerts, 2006, p. 147). That is why membership can be extended to less prototypical instances (Arnett, 2004, p. 33). Speakers assign linguistic elements to a specific category based on their features through a process that reassembles what Barabási claims about the attribution of fitness values to the nodes in a given network. Hence, the process of "categorization", together with the notions of "linguistic symbols", grammatical constructions as an "assembly", and the concept of linguistic "prototypes" are, I believe, relevant tools to understand how speakers perceive specific linguistic structures. From Complexity Theory and network science perspectives, this is comparable to how "fitness" values are assigned to specific nodes.

In the current study, I won't attempt to define the fitness value in linguistic terms, nor is it the focus of this research. However, in an attempt to use network science to capture and analyze historical linguistic change, I will address the issue of fitness throughout this study, together with its possible application for further research.

2.8 Summary

The purpose of the previous section was to establish the theoretical framework of the current study. In this work, languages are viewed as complex adaptive systems. Specifically, this work focuses on *werden* and on those emergent behaviors or, as Larsen-Freeman (2009) defined them, on those "form-meaning-use dynamic patterns of language use", which resulted from the continuous interactions of the speakers. The exploration of these syntactic changes within a complexity theory account creates the necessary theoretical background for applying the tools of network science to address this phenomenon. In particular, I attempt to capture the use of *werden* throughout the centuries. Further, it will investigate if the growth of *werden* over time resembles preferential attachment, according to which relatively highly connected nodes will keep increasing their number of links. In order to do that, this work makes use of syntactic dependency networks. The analysis will focus on a specific node, namely the verb *werden*, and it will explore the elements with which it connects during a time frame of ca. eight centuries (the Middle and Early New High

German periods). This study will also provide an opportunity to further explore the existence of a fitness value in syntactic dependency networks, with consideration of some of the cognitive factors that could be involved in determining the importance of a specific node in a network.

CHAPTER 3. REVIEW OF THE LITERATURE

In the previous chapter, I outlined a brief chronological overview of how the linguistic community handles syntactic change. It was discussed how some aspects of linguistic theory often failed to take into account the importance of the diachronic dimension of syntactic change to explain contemporary synchronic variation. As mentioned earlier, approaches such as Emergent Grammar and Complexity Theory provide a good fit for the research goal of the current work, which focuses on the development of the verb *werden* and the related emergence of the periphrastic passive and future. The last chapter also laid the theoretical foundations of the current study, highlighting how Complexity Theory in particular delivers the necessary and most suitable framework for the application of network science in the field of historical linguistics. The present study takes on the new possibilities offered by both the aforementioned accounts and applies those to the study of syntactic changes in the history of German to gather further evidence of preferential attachment in language evolution, as shown in Perc (2012) and Concu (2015). It also provides a valuable opportunity to join the community of scholars that have convincingly demonstrated that languages can be "represented as a complex network in [their] several levels of complexity" (Amancio, 2014, p. 1). However, first, I will offer a detailed review of the literature on previous scholarship addressing *werden* and the periphrastic constructions in which it appeared throughout the centuries.

3.1 *Werden* as a Full Verb and Copula

While the rise of *werden* periphrastic constructions with the past participles and infinitive verbs has been addressed frequently in the literature, comprehensive accounts on the evolution of *werden* as a full and copula verb are much more limited. Kotin (2003) presents the only study that puts it alongside the emergence of the *werden* passive and future. Kotin (2003) reports that *werden* originated from the Proto-Indo-European verbal root **uert(h)*, which meant 'to turn' or "to twist" and it has been also found in languages outside the Germanic family, such as in Sanskrit (*vártatê* – 'tu turn') and Latin (*vertere*- 'to turn'). In Early Germanic, this verb indicated a development and a change of state instead of a movement, thereby losing its original meaning in Proto-Indo-European (p. 29). *Werden* is found later in different Germanic languages: *wairpan* in Gothic,

uuerdan in Old High German, *werdhan* in Old Saxon, *weorphan* in Anglo-Saxon, and *verða* in Old Icelandic. At this stage, it meant "happen", "occur", and "to come into being" (Kotin, 2003, p. 30). Consider the following example from the *Tatian*:

(6) **Uuard** thô thaz **arfuorun** fon in thie engila in the himil
happened-PST then that rose-PST from them the angels-NOM into the sky
'it happened then that the angels rose from them into the sky'
(*Tatian*, 6.4)

Kotin (2003) outlines the development of *werden* from Germanic to Old High German as follows: **uuerth* ('to turn' with ambiguous semantic) → *werdan* atomic[2]-mutative[3] → non-additive[4] (combined with perfective or telic verbs denotes rapid changes) → durative[5] (with the neutralization of the opposition between fast and rapid changes) → non-atomic-mutative, additive → non-mutative (p.110). The loss of its mutative component, which took place during the 10th century, favored the increase of instances of *werden* as a copula verb, as shown in (7).

(7) dir **wirt** von mir ein **zwic**
to you-DAT becomes-PRS from me an oppression-NOM
'I (will) give you oppression/I will oppress you'
(*Nibelungenlied*, 18,33)

Werden as a full verb, however, lost the capacity to express meanings such as "happen" and "occur", while the meanings of "to come into being" and "to be given something" (*zuteilt werden*) were still very common at this stage.

Starting in the Early New High German period, the use of *werden* as a full verb slowly decreases, so much so that such use of *werden* has become rare today in Modern German (Kotin, 2003, p. 211). While its function as a copula remains stable in terms of frequency, the decline of the instances with *werden* functioning as a full verb corresponds to the strengthening of its status as auxiliary. This change brought with it an increase in the occurrences of *werden* in different periphrastic constructions, especially those with past participles and infinitive verbs. Although

[2] Atomic verbs denote the existence of an event as a whole (to cough).
[3] Mutative verbs denotes a change of place of position, or a change of state or condition.
[4] Additive verbs are a particular category of verbs indicating the action of adding something to something else already existing. Additive are verbs such as "beifügen" (to add); "beilegen" (to attach), and "hinzufügen" (to add).
[5] Durative verbs indicate an ongoing action without indicaty any temporal boundaries (to sleep).

very frequent in Old High and Middle High German, the combinations with present participles gradually disappeared in this period (p. 211).

In the next sections, I will offer a review of scholarship on the *werden* periphrastic constructions with the present and past participles and with infinitive verbs.

3.2 *Werden* with Present Participles

The *werden* periphrases with present participles have received little attention from the scholarly community compared to the constructions with the past participles and infinitive verbs. According to Kotin (2003, p. 152), the present participle in Old High German had a strong 'adjective-like nature" -which is still present in Modern German when it is used in both attributive (*das schlafende Kind* – 'the sleeping child') and predicative positions (*sie ist wütend* – 'she is angry') - and indicated a long-lasting state. The function of the combinations with *werden* was to mark the entering into a particular state or the beginning of an action. Thus, the present participles were particularly compatible with durative verbs. Consider the following examples:

(8) thô **uuard** mund sîner sâr **sprechntér**
Then became-PST mouth his very speaking-PTCP-PRS
'Then he started speaking a lot '
(*Otfrid*, I, 9,27)

(9) Inti nu **uuridist** thu **suigen-ti** inti ni math sprehhan
And now become-PRS you seeing-PTCS.PRS and not may speak
'And now you start to be silent and may not speak'
(*Tatian*, Lk, 1,20)

The meanings of 'starting to speak' in (1) and 'starting to be silent' in (2) are realized through the combination of *werden* and the durative verbs 'speaking' and 'being silent' in the present participle.

In Middle High German, the meaning and function of this periphrasis remained unchanged (p. 163). In Early New High German, however, this periphrasis gradually vanished. Kotin (2003, p. 165) links this phenomenon to the simultaneous disappearance of another similar periphrasis: *sîn* (to be) and present participles. This construction indicated non-mutative and non-terminative actions, conveying a meaning that was close to the semantic of any given verb in its finite form:

habund bist vs. *hâst* or *redened war* vs. *redete* – 'you are having' vs. 'you have' or 'you were speaking' vs. 'you spoke'. According to Kotin (2003, p. 166), it is exactly this redundancy in meaning that caused the decline of this periphrasis, which also prompted the disappearance of the constructions of *werden* plus present participle.

3.3 *Werden* with Past Participles

3.3.1 The Earliest Accounts: Valentin (1986) and Eroms (1998)

One of the earliest accounts on the emergence of the German passive is by French scholar Paul Valentin (1986). In his article, *Geschichte zum Deutschen Passiv* ('History of the German Passive'), he explores the development of the passive from Old High to Early New High German. In Old High German, the verb *werden* was used in combination with the past participle. But, while the combinations with the simple past were similar to the current use in Modern German, the use of the present tense had a future reading that referred to the moment in which the subject was entering into a particular state (p. 9). Consider the following example:

(10) endi **uuirdit** siin namo **chinemnit** uuadarliih
and becomes-PRS his name called-PTCS-PST glorious
'and his name will be glorious'
(*Valentin*, 1986, p. 8)

At this stage, *werden* was not yet fully grammaticalized as an auxiliary, and the combinations of this verb plus a participle were not yet considered to be full passive constructions (p. 10). In the later part of the Old High German period, *werden* loses its future and ingressive reading and starts to express an ongoing process defined by the past participle (p. 11). Between the 14th and 15th centuries, the development of *werden* takes a significant turn as this verb starts to behave like the modern passive. The passive construction then, especially in the 16th and 17th centuries, starts to appear in the present perfect form as a result of its increased frequency (p. 12). Finally, in the middle of the 17th century, the combinations of *werden* plus past participle seem to have reached a full grammaticalized status. Valentin (1986) offers two possible explanations for the emergence of the *werden* passive. On the one hand, he claims that the semantic development of this verb from a future to a more progressive reading in the present tense allowed the passive

combinations to grammaticalize and acquire the meaning that they have today. On the other hand, the emergence of the present perfect made the use of the opposition between the *sein* and *werden* passives as an aspectual marker no longer necessary. This enabled a broader use of *werden*, which consequently could grammaticalize as the modern passive (p. 14).

Eroms (1992) offers another relevant study that is among the earliest accounts on the German passive. He opens his analysis of the history of the German passive by describing this construction in Gothic, which was realized through the verbs *wairthan* ('to become') and *wisan* ('to be'). The passive could express the beginning of an event with *wairthan* or the repetition of an action with *wisan* (p. 231). In Old High German, the passive could also be formed with *werdan* and *wesan*/*sin*. However, contrary to what Valentin (1986) claimed, these verbs in the present tense could not be ascribed to any specific aspectual meaning. The passive forms with *wesan*/*sin*, for instance, could also convey an ongoing process like the passive with *werdan* (p. 233). The *werdan* passive was also used to express forthcoming actions. Thus, it was primarily used to translate the Latin future tense. In the past tense, *werdan* referred to the entering into a specific state or passing from one state to another, whereas *wesan*/*sin* indicated the result of an action (p. 235). In Middle High German, *werden* slowly loses its ingressive meaning but, in combination with the past participle, keeps its future reading. It is not until the 14th century when *werden* reaches the auxiliary status it has today in Modern German. At that point, it becomes the preferred way to form the passive, while the use of *sein* slowly decreases and is restricted to atelic transformative verbs (p. 239). In the Early New High German period, Eroms (1992) notes, there is an emergence of the dative passive, whose first instances were found in Luther's bible translation (p. 241).

3.3.2 The Passive as the Result of "Paradigmatization"

Michael Kotin offers two of the most significant contributions to the study of the German passive. In his first work from 1998, he traces the development of the passive from Old High to Early New High German. In the instances from Old High German, he claims the combinations of *werden* plus the past participle were not yet grammaticalized as passive. Both *werden* and the past particle could be considered at this stage autonomous parts of a grammatical combination, in which *werden* had the function of determining the aspectual nuance (processual) of the construction (p. 82). In the data from the tenth and eleventh centuries, he notes that the number of instances in which the participle had an adjectival ending drastically decreased, a possible sign of the beginning

of the grammaticalization process of the passive periphrasis (p. 83). Like Valentin (1986) and Eroms (1992), Kotin (1998) observes that *werden* used in the present tense plus past particle had a future reading, while its use in the past tense resembles the modern passive (p. 88).

In Middle High German, the classes of the verbs that appear in the passive increased in number, due to the semantic weakening of both *werden* and *sein*. Specifically, *werden* started to be combined not only with a higher number of transitive verbs but also with intransitive verbs that required the genitive case and had prepositional objects, as well as with verbs that could be used with both transitive and intransitive meanings (p. 113). In this period, Kotin also finds the first instances of both the passive constructions without an overt subject and the use of the singular neutral pronoun *es* as a "place holder" in the position that was traditionally occupied by the subject (p. 115). In the twelfth and thirteenth centuries, the *werden* passive was still often used, like in Old High German, to express a forthcoming process, whereas the *sein* passive denoted the result of a state (p. 124). However, as Kotin (1998) points out, such opposition was not always realized through the choice between *werden* and *sein*. Consequently, and as also indicated by Eroms (1992), it is impossible to ascribe a specific aspectual component to any of these two verbs. In the instances from this period, resultative states could also be expressed with *werden*, while *sein* could refer to a forthcoming process (p. 126). Hence, the situation of the *werden* passive in Middle High German can be described as follows: *werden*, when used in the present tense, expressed an upcoming and expected result; when used in the past tense, it described the entering into a specific state or achieving a particular result (p. 127). In Middle High German, it was also possible to find *werden* combined with both the weak and strong forms of the past participle of the same verb. The first form denoted a completed action, whereas the second one did not provide any specific information about the duration of the event described. Such contrasts strongly suggest that the aspectual nuances of an action were still mainly determined by the past participle and not by the use of *werden* (p. 130).

In Early New High German, *werden* is completely deprived of its original lexical meaning as a result of its grammaticalization process. At this stage, the *werden* passive acquires its status as *Vorganspassiv* and is able to denote an ongoing process in the present without any future reference. This shift is also shown in the appearance of the passive future, in which speakers added an additional *werden* to the constructions with the past participle periphrases in order to convey future references (p. 145).

Kotin (2003) offers an even more extensive study on the evolution of the *werden* passive. However, *Die werden-Perspektive und die werden-Periphasen im Deutschen* cannot simply be considered a longer version of his previous research. Instead of focusing only on the evolution of the dichotomy between active vs. passive in German, Kotin (2003) analyzes the development of all the constructions with *werden* (p. 9). Hence, his analysis considers not only the combinations with the past participle, but also those with the present participle, the infinitive, and the constructions with *werden* in the subjunctive mode. Furthermore, his study takes into consideration the particular status of the past participle as an autonomous entity.[6]. He proposes an analysis of *werden* that represents a relevant departure from his previous work. It becomes clear, in fact, how not only the grammaticalization of *werden* but also the evolution of the past participle contributed to the emergence of the German passive as it is today.

Kotin (2003) claims that the combinations of this verb with the past participle in the 8th and 9th centuries could be considered as the sum of two autonomous units in which both elements behaved independently but kept their original semantics. He notes that the past participle in these combinations, although traditionally treated as an adjective, had a rather ambiguous reading that makes it difficult to regard such combinations as simple "copula + predicate" constructions, as still held today in the literature (Valentin, 1986; Eroms, 1992; Fleischer & Oliver, 2011). Hence, he argues that such combinations in the earliest texts from the 8th and 9th centuries could be considered both "copula + predicate" and a combination of two independent verbs at the same time. *Werden* was back then in a sort of a middle stage between its status as a full verb and as an auxiliary and kept its processual meaning in all the combinations with nouns, adjectives, present, and past participles. Similar claims are made about the past participle, which at this point, was neither a predicate nor a full verb (p. 92). In the combinations with *werden* (and in those with *sein* and *haben* – to be and to have), it could also appear with or without adjective endings with any significant differences between the declined forms and those that were not declined. This behavior supports both Kotin's (2003) claim about the ambiguous nature of the past participles and his choice to name the combinations with *werden* "nominal-verbal predicates" (p. 93). The ambiguity of the past participles in Old High German can also be seen in combinations that included *werden* with intransitive verbs such as *queman* (to come) and *uuerdan* (with the meaning of "to become"), a phenomenon that can be ascribed only to this early stage of German, but also in Old Saxon and

[6] See also Kotin (2000) for a deeper analysis of the past participle in Old High German.

Old English (p. 125). Such combinations were possible because of the affinity between the meaning of *werden* and the resultative reading of intransitive verbs such as *queman* and *uuerdan*. However, such constructions slowly disappeared, and their usage is limited to the first centuries of the Old High German period. In Notker (late 9th and 10th centuries), for instance, the intransitive verbs appeared only with *uusan* (p. 134). This suggests that *werden* was at that point already losing its original meaning as a result of the beginning of the grammaticalization process (p. 134).

The status of *werden* plus past participles in the late part of the Old High German period is summarized by Kotin as follows: "es handelt sich seit dem späten Althochdeutschen um eine biverbiale Fügung mit semantisch diffusem Finitum und einem Partizip, bei dem die verbialen Eigenschaften stärker ausgeprägt sind als die nominalen" (p. 96) [Starting from the late period of Old High German, we find a bi-verbal combination of a finite verb with a vague meaning and a participle, in which the verbal features have become stronger than the adverbial ones]. During this period, *werden* plus past participle could also be found with a dative object (considerably earlier than what claimed is by Eroms, who dates the dative passive to the Early New High German period), in impersonal constructions with *man* (with intransitive verbs) and with no subject, in which the participles referred back to an unspecified recipient. According to Kotin (2003), in the instances without a subject or *man*, the omission of the subject emphasized the verbal feature of the past participles, in which *es* (it) was implied but did not have to appear (p. 114). The use of *werden* was then an alternative to the *man*-constructions and favored the shift from an unknown entity to an "indefinite impersonality". These types of combinations, which were limited to dative objects at first, were extended later to genitive and prepositional objects (p. 115).

In Middle High German, with the weakening of its mutative meaning, *werden* is used less frequently as a full verb, while its status as an auxiliary gained further strength. At this stage, the combinations with the past participle looked more like a "function verb + full verb", as in Modern German constructions with modal verbs plus an infinitive (p. 144). *Werden*, as a result of this newly started process, weakened its "concrete" meaning in favor of a more abstract one. According to Watts (2001), this is the first step of grammaticalization (p. 120).

One of the main consequences of this shift is the expansion of the lexical base which favored the combination with a larger number of transitive and intransitive verbs. During this period, the passive started to appear in the newly grammaticalized present perfect, and the instances with genitive, dative, and prepositional objects increased in number. In these instances,

according to Kotin (2003), the focus is on the process expressed by the verb (p. 149). As already mentioned in the previous pages, many scholars have described the status of the passive in Middle High German as formed by the aspectual opposition between *sin* and *werden*, in which the first one indicated a state and the latter marked a change or a process. Kotin (2003) argues against such a position, reporting that it was the past participle that carried the aspectual meaning of these periphrases. The loss of the mutative feature of *werden*, which started during the later years of Old High German and the early part of the Middle High German period, made it possible that verbs could indicate an ongoing process that had been incompatible with *werden* in the earlier periods. However, in these combinations, the information regarding the lexical aspect was still mainly conveyed by past participles. In the occurrences with verbs such as *finden* and *suchen*, for instance, the aspectual readings of these actions were solely expressed by the intrinsic semantics of these two verbs and not by *werden*, which kept its "aspectual ambiguity" (Kotin, 2003, p. 193). The Middle High German period is then, according to Kotin, the stage during which the original mutative semantics of *werden* was dismantled, and the periphrases with the past participles started to look like the modern passive (p. 193).

The Early New High German period marks the most relevant grammaticalization phase of the constructions with *werden* plus past participles, a process which had already started near the end of the Old High German period. Interestingly, Kotin (2003) claims that, in this period, we witness both the end of the grammaticalization of the verb *werden*, which caused the loss of its original meaning, and the "paradigmatization" of the periphrases with the past participles (p. 204). Kotin, in fact, distinguishes between "grammaticalization" and "paradigmatization": while the first takes place as a result of the adaptation of a new grammatical construction to an existing archetype, the second one is an independent and spontaneous development that can be considered the result of analogical associations (Smirnova, 2008, p. 50). Kotin (2003) defines the last process as the inclusion of already grammaticalized forms (here *werden* and the past participles) into the passive verbal paradigm through analogy, which could now encompass a higher number of verbs in Early New High German in comparison to the Old High and Middle High German periods.

3.3.3 The Latest Accounts

Arnett (2004) addresses the development of *werden* and *sein* passive periphrasis, applying a cognitive linguistic approach to the description of these constructions. According to her study,

werden and the past participle had a future meaning in Old High German. However, when *werden* was combined with a participle denoting an action that included two participants and an asymmetrical event, the focus was shifted to the process instead of the result of the action (p. 56-57). In Middle High German, Arnett observes that the passive slowly loses its capacity to refer to future events and starts behaving more like the modern passive. Later on, *werden* denoted only ongoing processes. Hence, and as also reported by Eroms (1992), the future passive was realized with a second *werden* added to the combination of *werden* plus participle (p. 58).

Salmons (2012) also touches briefly on the development of the periphrastic *werden* passive from Old High to New High German. For Old High German, he cites the divergent positions of Lockwood (1968) and Schrodt (2004). While Lockwood (1968) argues that such constructions had a future, ingressive, and inchoative meaning (p. 158), Schrodt (2004) argues that the future reading was solely realized using the present tense plus the past participles of transitive verbs (p. 159). In Middle High German, Salmons reports an increased usage of periphrastic constructions in general. He describes the periphrases with *werden*, *sein*, and *haben* as in a middle stage between Old High and contemporary German (p. 207). In the Early New High German period, the *werden* passive behaves more and more like its modern counterpart, while *werden* alone starts to appear as an infinitive to express the future tense (p. 259).

Petre (2013) gives an entirely different account on the development of the *werden* passive, claiming that this verbal periphrasis emerged as a result of the grammaticalization of the verb-second position in narrative contexts. In narrative contexts, time adverbials usually occupied the first position, and *werden* normally appeared in the second position. The author analyzes the relative frequency of the use of the combinations of time adverbials and *werden* in two major works in Old High German, namely the *Tatian* and the *Evangelienbuch* (the Gospel Book). According to his analysis, *werden* started to become grammaticalized during this period, and as a result of this process, it loses the capacity to express a change of state. The loss of its original semantics led to an increased use of this verb, which appeared at that point with verbs that were normally combined with *wesan* (p. 90). The high frequency of *werden* plus temporal adverbs in main clauses in which the word order was inverted set up the environment for the grammaticalization of *werden* as the "default" passive auxiliary. The emergence of *werden* as an auxiliary is, according to Petre (2013), to be considered strictly as a development of word order in German. Hence, examples with an adverb (temporal or spatial) in the first position with inverted

word order represented the prototypical environment in which the *werden* passive started its grammaticalization path (p. 97).

3.4 Summary

The overview of the scholarship on *werden* passive that is offered in the previous pages has shown the diverse and, sometimes, divergent positions on the process by which the German passive emerged. The study by Kotin (2003) offers the most extensive account on this topic, including an analysis of the role of the verb *werden* and the contribution of the past participle to the emergence of this construction. Nonetheless, each of the works included here repeatedly reports the expansion of the lexical base of the verb *werden* as soon as this verb started to become grammaticalized.

3.5 *Werden* with Infinitive Verbs

The scholarship on the periphrastic future in German has brought forth an equally diverse and divergent number of hypotheses about its possible sources. There remains little consensus about its origin today (Fleischer & Schallert, 2011). Moreover, there is little agreement on the period during which this periphrasis started to be used with future references. Some scholars argue that it emerged in the last years of the Middle High German period, while others include the rare instances in the Old High German texts and claim that this periphrasis existed as early as the 9th century. In this section, I will offer a review of this scholarship, addressing hypotheses and theories about this puzzling construction in detail.

3.5.1 The Periphrases of *werden* with Present Participles as the Source for the Emergence of the *werden* Future

The hypotheses on the emergence of the future as related to the periphrases of *werden* with the present participles were formulated first at the beginning of the 20th century (Bechs 1901; Kleiner, 1925) and supported later on by numerous scholars (Weinhold, 1883; Behaghel, 1924; Ebert, 1978; Betten, 1987, von Polenz, 1991). According to these scholars, the periphrases of *werden* plus infinitive have originated from the combination of *werden* and the present participle. The loss of the ending of the participle nullified the difference with the infinitive, which,

consequently, starts to be used in combination with *werden* to indicate future events. Meanwhile, the constructions of werden with the present participles disappear. This process started allegedly in the 13th century in the Low German area, reaching the southern speaking areas shortly after that. Kleiner's 1925 claim diverges a little from this position, arguing that the periphrases with werden plus infinitive have originated instead in the Alemannic dialect. These are then the result of the analogy (and "confusion") between the combinations with the present participle and the dative infinitive that, in this dialect, was often used without the inflectional ending -de.[7]

3.5.2 The Future Marker as a Result of Analogy with the Modal Verbs

Many scholars have looked at the German modal verbs as the source for the emergence of the future periphrasis with *werden*. The first account of this type is by Bogner (1989). He arrives at this conclusion by analyzing the occurrences of *werden* plus infinitive in the *Bonner Frühneuhochdeutschkorpus* and by comparing the instances he finds with those with *wollen* (will), *sollen* (shall), and *müssen* (must). He also looks at when these verbs, *werden* included, were used with temporal or modal readings. He notices that the occurrences of *werden* with temporal reading were an average of fifty in the first two centuries of the Early New High German period, but that those instances were more than 200 two centuries later, especially in the Upper Saxon dialect area. At the same time, the use of the modal verbs with temporal readings significantly declined (Bogner 1989, p. 77). Bogner (1989) bases his claim only on the instances in Early New High German and identifies the modal verbs with future readings as the model for *werden* plus infinitive periphrastic constructions.

Fritz (1997, p. 83) claims that, together with the analogy with the modal verbs, the emergence of the periphrastic future was also influenced by the "perfective" semantics of the verb *werden*, sharing the same assumptions of Valentin (1987) about the existence of an aspectual opposition between *uuerdan* and *sin* in Old High German. Fritz claims that, when used alone, *werden* referred to complete actions and established the temporal frame in which a given event

[7] Saltveit (1961) is one of first scholars who argued against this hypothesis, claiming that the combination of *werden* plus an infinitive, which was found already in Old High German, did not have a temporal reading at the beginning but rather a modal one. Later on, it acquired the additional function to refer to future events, slowly replacing modal verbs such as *sollen* (shall) and *wollen* (will) that, until that point, had been used to express both modal and future meanings. Other scholars have criticized such theory (Leiss, 1985; Schmid, 2000; Harm, 2001) but the "exact details of this development are still a matter of dispute" (Diewald and Smirnova, 2010, p. 234).

took place. The feature *Grenzbezogenheit* (relatedness to a boundary) was central to the semantics of this verb in its earliest occurrences. Starting from Notker' time and at the beginning of the Middle High German period, *werden* loses part of its semantics. Consequently, the instances of *werden* plus past participle start to behave like the modern *Vorgangspassiv*, acquiring the ability to also describe "slow" changes, while *werden* alone loses part of its perfectivity. This process comes to an end in Early New High German, when speakers started to use *werden* together with the infinitive with no aspectual but future meaning, following the model of the modal verbs (Fritz 1997, p. 88).

Similar to Bogner's claim, Schmid (2000) considers modal verbs as the primary source for the emergence of the periphrastic future. He analyzes a large number of texts from the upper, middle, and lower German dialect areas written between the twelfth and fifteenth centuries and claims that the combinations of *werden* and infinitives are the outcome of linguistic contamination with the constructions with the modal verbs. Citing Paul (1920), he defines contamination as the process by which two semantically related elements are slowly perceived as being equal and, consequently, start to be used in the same way. Schmidt bases his claim on the instances of modal verbs combined with *werden* that he finds throughout the corpus. He argues that the communicative intent was to collocate the narration in the future (Schmid, 2000, p. 13).

Like Bogner (1989) and Schmid (2000), Harm (2001, p. 300) also suggests that the emergence of the *werden* periphrastic future was the outcome of analogic associations with the modal verbs that had future readings. Thus, the replacement of past participles with an infinitive was triggered by the frequent use of modal verbs with a future reading. Furthermore, the frequent future and ingressive readings associated with the passive combinations with *werden* in the present tense and the past participle elicited another association in which *werden* was perceived as the main carrier of the future meaning (Harm, 2001, p. 304).

3.5.3 The German Future as a Product of Language Contact

Other scholars strongly depart from the aforementioned hypotheses, arguing that the *werden* future constructions in German emerged as a result of language contact. Leiss (1985, p. 295) identifies the instances of language contact between German and Czech during the so-called *Ostkolonisierung* (east colonization) between the twelfth and the fourteenth centuries as the main source for the emergence of the *werden* periphrasis. Specifically, the construction with *budo* (to

be) plus infinitive, which was used in Czech to indicate forthcoming events, functioned as the model for the formation of the future periphrasis in German. Leiss based her hypothesis based on three different factors. First, the construction in Czech, she argues, was older than its German counterpart with *werden*. Second, *werden* plus infinitive originated in the eastern Middle German dialect, which was spoken in an area close to where Czech was spoken. Third, several episodes of language contact between Czech and German triggered the formation of the *werden* periphrases based on the budo plus infinitive combinations.[8]

Diewald and Habermann (2005) offer their account on the emergence of the periphrastic future and suggest that "werden erst durch den Sprachkontakt mit dem Lateinischen zur systematischen Futurmarkierung des Deutschen wurde" [*werden* became the systematic future marker in German firstly through contact with Latin) (241). According to Diewald and Habermann 2005, this construction evolved between the last years of the Middle High German and the beginning of the Early New German periods. Although they acknowledge the existence of these constructions in Old High German, they identify the time frame between the fourteenth and the sixteenth centuries as the relevant stage in which this structure gained its status as a future tense marker. They differentiate between two separate steps in the grammaticalization process of this structure. The first one is characterized by the succession of internal linguistic processes such as analogy, which they define citing Traugott 2005, as a process that "involves attraction of extant forms to already existing structures, hence generalization". Mentioning also Bybee et al. 1994, they point out that the periphrastic future constructions have normally emerged from deontic modals, whereas the emergence from lexeme with a perfective semantics such as *werden* is, on the contrary, very rare. The meanings of desire, wish or obligations are traditionally recognized sources for the development of future tense forms (Diewald and Habermann 2005: 278). Nonetheless, the authors list different factors that favored the emergence of *werden* as the auxiliary to realize future temporal references. First, the intrinsic semantics of *werden* itself made it a suitable candidate for this purpose. Second, *werden* frequently appeared in different types of constructions, showing a large "konstrutionelle Variabilität" (constructional variability). Third, werden plus infinitive could strengthen its future meaning through the analogic integration of these

[8] Leiss's theory has been questioned by many scholars (Fleischer and Schallert 2011), although there have also been some occasional attempts to provide further support to her claims (see, for instance, Masarik 1994).

constructions into the small group of ingressive verbs. However, these factors alone do not suffice to fully explain the spread of *werden* plus infinitive as a future marker. According to Diewald and Habermann (2005) is especially the language contact with Latin that played a key role in the emergence of the periphrastic future (239).

3.5.4 The *werden* Future Constructions as the Product of Analogy with *stantan* and *biginnan*

Some scholars have argued that the source for the *werden* future can be found in the analogy with verbs other than the modals. Kotin (2003) addresses the constructions with *werden* plus infinitive, considering also the instances from the latest years of the Old High German period. According to him, the instances in Old High German, which appeared with *werden* in the past tense, had the function of marking the beginning of an action in the past. Additionally, the constructions were modeled after the constructions with *duginnan* (to begin) in Gothic and *stantan* (to stand) in Old High German which were used to mark the moment in which an action began. In Middle High German, *werden* plus infinitive starts to appear in the present tense and had a similar status to that of the periphrases with *werden* plus the past participle: they could be considered bi-verbal constructions in which both elements still had a relatively large degree of autonomy. *Werden* plus infinitive in the present tense competed at this stage with the constructions with modal verbs, since they could also convey a future-related meaning (Kotin, 2003, p. 167). In Middle High German, *werden* continues to lose its mutative reading, which is, according to Kotin, one of the prerequisites for its grammaticalization as a future auxiliary (Kotin, 2003, p. 172). At the end of the 15th century, Kotin reports relevant changes in the verbal system: the periphrasis with *werden* plus present participle and those in which *werden* was used in the past tense and combined with the infinitive slowly disappeared. *Werden* in the present tense plus infinitive becomes the only construction used to express future related meanings. At the same time, the verbs such as *sollen* and *wollen* begin to be used only in their modal meaning. In Early New High German, the future periphrases with *werden* in the present tense plus infinitive complete their grammaticalization process and make their way into language use (Kotin, 2003, p. 202).

Another work that sees analogy as the process responsible for the emergence of the *werden* future auxiliary is Krämer (2005). She addresses the emergence of the periphrastic future and distinguishes between the grammaticalization of *werden* plus infinitive with a future reading and

werden plus infinitive with a modal reading. The first process started with the combination of *werden* and the present participle in which the first behaved as a copula verb. Although such claims are similar to those argued by Bech (1901), the reasons Krämer lists as accountable for such a phenomenon are significantly different. According to Krämer (2005), the grammaticalization of the future periphrases occurs in two different steps. During the first step, the infinitive replaces the present participle through analogic association with the Old High German constructions *biginnan* (to begin) plus an infinitive. Such a process was motivated by both constructions sharing a clear ingressive reading. At the beginning of this process, the periphrasis with the present participle and the infinitive coexist and *werden* still behaves as a copula. In the second step, the periphrases with the infinitive are reanalyzed and *werden* acquires its status as an auxiliary for the expression of the future tense. Once the future constructions were fully established, the modal reading arises as a result of a second step in the grammaticalization process. At the same time, the combinations of *werden* plus present participle completely disappear (Krämer, 2005, p. 92). Krämer summarizes the development of *werden* as follows:

> Full verb → copula → auxiliary with the function to indicate forthcoming events
> → auxiliary with the ability to express modality.

The most recent account to date on the future periphrases with *werden* is by Diewald and Wischer (2013). In their study, they compare the future auxiliary "candidates" in Old English and Old High German. Other than *werdan*, Old High German candidates were verbs such as *wellan* and *schulan*. These last two verbs, although occurring frequently and very often used with future implications, were never suitable candidates for such a role because of their semantic components that involved volition or obligation. Conversely, *werdan* was able to be grammaticalized as such because of its dominant ingressive meaning which made it compatible with predicates of any kind. Additionally, "throughout the history of German, *werden* has displayed a high constructional variability. It has always been used simultaneously in a range of syntactic functions spanning from full verb via copula to auxiliary" (Diewald & Wischer, 2013, p. 209). Diewald and Wischer (2013, p. 215) also report some instances of *werden* in the present tense combined with an infinitive from the Gospel Book by Otfrid, contradicting in part what is claimed by Kotin (2003, p. 135), who stated that such combination in Old High German could appear only with *werden* in the past tense.

3.6 Summary

This overview of the scholarship on the periphrastic future has shown the large variety of theories and hypotheses on both the meaning and the origin of this construction put forward by various scholars. However, and similarly to what was reported for the passive periphrases, the various accounts included in this literary review all report significant changes in the verbal system of German during the time in which the periphrastic future emerged. Such changes include, for instance, the disappearance of combinations such as *werden* in the past tense plus infinitive, *werden* with the present participle, and the expansion of the lexical base of the periphrastic future as a result of its grammaticalization.

The analysis that will be conducted in the next pages will also help shed light on the origin of the future periphrases since it focuses on a significant number of instances from a period of time that goes from Middle to Early New High German. The results will provide support for one of the theories presented in the previous pages. Among these, the claims made by Kotin (2003) and seems to be the most convincing one. Kotin (2003) takes into consideration the occurrences with the infinitive verbs with *werden* in both *Präsens* and *Präteritum* found in Old High German. He demonstrates how the combinations of *werden,* and the infinitive could be considered bi-verbal constructions in which both elements still had a relatively large degree of autonomy. That is the reason why, in Old High German and in the first centuries of the Middle High German period, it was possible to find the infinitive verbs together with *werden* in the *Präteritum*. These attestations, on the one hand, reinforce the claim that the combinations of *werden* plus the present participles were not the source of the future periphrases since they both appear for the first time in this period. On the other hand, they also disprove theories such as those that claim that the future periphrases were the results of language contact with Czech (Leiss, 1985) or Latin (Diewald & Habermann, 2005).

In the next section, I discuss the methodology of the present work. The first part addresses the issues involved in the collection of the data. The second part describes the methodology used for the application of the tools of network science for the analysis of the development of *werden* and the related emergence of the passive and the future periphrastic constructions.

CHAPTER 4. METHODOLOGY

This first part of this section discusses the methodological issues involved in the selection and data collection, which consist of the instances of *werden* found in a total of fifty-one texts from Middle and Early New High German. I begin with a description of the corpora, and I continue describing how I analyze the instances I find in the selected texts. The second part of this section describes the methodology used to build the networks from the targeted sentences with *werden* and how the measurements of degree centrality, betweenness centrality, and eigenvector centrality are used to gather evidence for preferential attachment.

4.1 The Corpora

The texts from which I have collected the data come from two different databases: the database of Middle High German and the database of Early New High German.

The corpus for Middle High German consists of 30 different prose texts from the *Referenzkorpus Mittlehochdeutsch* (Klein, T., Wegera, K., Dipper, S., Wich-Reif, C., 2016), which consists of three different sub-corpora, including the *Kölner Korpus hessisch-thüringischer Texte, the Bonner Korpus mitteldeutscher Texte, and the Bochumer Mittelhochdeutschkorpus* (BoMiKo).

The text selection for this study covers the entire Middle High German period (11th, 12th–13th, and early 14th century) and five different dialectal areas (West Middle German, East Middle German; West Upper German, East Upper German, and Nord Upper German). These texts offer a way to cover a vast geographical area from the 11th to the 15th centuries. This corpus contains an average number of two texts for each century and dialect. The selected writing comprises mostly religious and legal texts, which allows this project to concentrate on texts with similar topics. Further, poetic texts are excluded. The total word count for Middle High German is 202,385. The titles of the texts and the exact date of composition can be found in Appendix.

The database for Early New High German includes twenty texts from the *Bonner Frühneuhochdeutschkorpus*, which contains texts from the late 14th century to the 17th century. The corpus encompasses religious as well as legal prose texts, together with other text types. Like the database for Middle High German, this corpus covers five different dialect areas: West Middle German, East Middle German, West Upper German, East Upper German, and Nord Upper

German. The total amount of words for Early New High German is 253,559. The titles of the texts and the exact date of composition can be also found in Appendix.

The data were collected from the annotated online corpora (Middle High German: **https://www.linguistics.rub.de/rem/index.html**; Early New High German: **https://korpora.zim.uni-duisburg-essen.de/FnhdC/index.html**) and all the instances of *werden* were compiled in two different databases, one for Middle High and Early New High German. Each database contains all the tokens of [WERDEN + past participle], [WERDEN + present participle], [WERDEN + adjective], [WERDEN + infinitives], [WERDEN + nouns[9]], and [modal verb + WERDEN]. The data also show in which tenses *werden* was used (present, simple past, present, or past perfect) independently of the mood (indicative or subjunctive). The analysis did not include sentences with *werden* in which the annotations of one of the elements was missing, as shown in example (11). The corpora display missing annotation as [!] in the Middle High German corpus, and as "*unbekannt*" (unknown) in the Early New High German corpus. Instances of *werden* with some missing annotations were included when these elements were not directly connected with *werden* (12).

(11) unte dero sententia **uuart** reprobate
and of their [!] became-PST [!]
'..and of their…..was….'
(*Hoheliedkommentar*, 11v,05)

(12) ze vile [!] **worden** **sint** die [!]
too many [!] became-PTCP.PST are-AUX.PRS the [!]
'too many had been…the'
(*Rheinfränkische Interlinearversion der Psalmen*, 3v,4)

In (11), the verb *werden* is used in combination with the word *rebrobate*, but no annotations are provided by the corpus about this word, and it is not possible to establish how the target verb has been used in this instance. In (12), there are also missing annotations, but those refer to elements not directly connected with *werden*. This instance can, therefore, be included in the data.

[9] This category can also include pronouns such as personal and demonstrative pronouns.

4.2 The Network Analysis

In this section, I will describe what is commonly known as "network mapping", which is "the creation of a networked-based visualization of a complex system" (Kolaczyk, 2009, p. 50). This process requires three specific steps: (1) collecting the data; (2) creating a representation as networks from the data; and (3) creating a visual image that represents that networks. (Kolaczyk, 2009, p. 50). The corpus analysis described in the first portion of this chapter, which refers to the collection of all the instances of *werden* in the Middle and Early New High German corpora, represents step (1). Here, I will describe the second and the third steps of the "network mapping". From the data collected, I have created syntactic dependency networks for each sentence included in the database. The rules used to build such networks are inspired by the principles of dependency grammar (Jurafsky & Martin, 2017). This approach is based on the notion of "grammatical relation" according to which the relationship between a head and a dependent is established in a given sentence (Jurafsky & Martin, 2017, p. 2). The head is "the central organizing word of a larger constituent" (e.g., the primary noun in a noun phrase, or verb in a verbal phrase). The remaining words in the constituent are either directly, or indirectly, dependent on their head (p. 2). Furthermore, "the head-dependent relationship is made explicit by directly linking heads to the words that are immediately dependent on them" (p. 2). The relationships established between the head and its dependent "allows us to further classify the kind of grammatical relations, or grammatical functions, in terms of the role that the dependent plays with respect to its head (p. 3). Dependency grammar uses "dependency trees" to visualize the relationship between a head and its dependents. Consider the following figure:

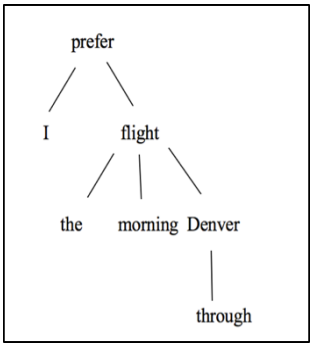


Figure 1: A dependency-style parse for "I prefer the morning flight through Denver" (Jurafsky & Martin, 2017, p. 2).

Figure 1 shows a dependency-style parse for “I prefer the morning flight through Denver”. The governing element in this sentence is the verb “prefer” on which both the pronoun “I” and the noun “flight” depend. At the same time, the noun “flight” functions as the head with three dependents: the determinative article “the”, and the nouns “morning” and “Denver”. “Denver” has also another dependent, namely the preposition “through”. According to Jurafsky and Martin (2017, p. 4), dependency structures as the one shown in Figure 1 are direct graphs that respect the following constraints:

1. There is a single designated root node that has no incoming arcs.
2. With the exception of the root node, each vertex has exactly one incoming arc.
3. There is a unique path from the root node to each vertex in V.

As previously mentioned, the networks were built using rules which were inspired by the formalism of dependency grammar and by the aforementioned constraints. There is a total of three rules in the current study, and these reflect the syntactic relations between the elements in a sentence. A rule is classified as a nominal phrase (NP) if it has a noun or a pronoun as the head. A rule is classified as a verbal phrase (VP) if it has a verb as the head. A rule is classified as a prepositional phrase (PP) if it has a preposition as the head. To be included in the database and have lemmas and rules assigned, a sentence must contain the target verb *werden* plus elements such as present and past participles, infinitive verbs, nouns, and adjectives, as shown in example (12):

(12) sô siu swanger **werden** sole
so sie pregnant-ADJ become-COP.INF shall-AUX.PRS
‘she shall become pregnant so’
(*Älterer Physiologus*, 32v,33)

The sentence in example (12) satisfies all the requisites previously mentioned. It contains the target verb *werden* (here as a copula verb) in the infinitive form in combination with the modal verb *sole* (shall) and the adjective *swanger* (pregnant). The function of *werden* is perfectly recognizable, and this sentence is, therefore, included in the database. When complements were found close or in-between the elements mentioned above, or two different instances of *werden*

were connected to the same subject, as shown in example (13), these elements were also included in the database.

(13) der mense **gestedigit** **wirt** an rehtime lebene vnde
the human anchored-PTCS.PST becomes-AUX.PRS in the right life and
gereinigit-PTCS.PST **wirt**-AUX.PRS
cleaned becomes
'the human being gets anchored in the joust life and gets cleaned'
(*Salomon Haus*, 109a, 14)

In (13), the verb *gestedigt* (anchored), which is the participle connected to the first *werden*, has a prepositional object, *rehtime lebene* (the right life), which is introduced by the preposition *an* (in). The propositional object is situated in between the two instances of *werden*. For this reason, it is also included in the database.

After the selection of the target sentences and the creation of the database of both Middle and Early New High German, I have listed the sentence networks in Excel spreadsheets. Every excel sheet has three different columns:

1. Source: the nodes from which an outgoing connection starts;
2. Target: the nodes that receive an incoming connection;
3. Interactions: the syntactic relationships between elements (NP; VP; PP).

All the source and target nodes are listed in their lemma form taken from the annotated online corpora. Each node in the network represents the lemma of a given word, which corresponds to the canonical word form (for instance, the singular form of a noun is the lemma for a plural form, or the infinitive form of a verb is the lemma of an inflected form of a verb).

I have also assigned to every node in the source and target columns a denomination according to its grammatical role:

AD	Adverb	PR	Preposition
AJ	Adjective	PP	Personal Pronoun
AR	Article	PS	Possessive Pronoun
AX	Auxiliary	PcPr	Present Participle
CJ	Coordinating Conjunction	PcPs	Past Participle
DM	Demonstrative Pronoun	RX	Reflexive Pronoun

IV	Infinite Verb	RP	Relative Pronoun
MV	Modal Verb	SC	Subordinating Conjunction
N	Noun	V	Verb
PK	Particle		

Consider the sentences in examples (14) and (15):

(14) **uuiret** er ofto an heligero geſcrifte **genamit**
becomes-AUX.PRS he often in the holy writings named-PTCP.PST
'He is often mentioned in the holy writings'
(*Älterer Physiologus*, 31ra 3)

(15) si irteilt **wart**
she condemned-PTCP.PST became-AUX.PST
'She was condemned'
(*Rheinauer Gebete*, 1va,13)

I have assigned the grammatical roles of the elements of (14) and (15) as shown in Table 3 and Table 4 respectively:

Table 3: The creation for the network for the sentence "uuiret er ofto an heligero geſcrifte genamit"

Source	Target	Interaction
AX warden	PP er	VP
AX warden	PcPs namen	VP
PcPs namen	PR an	VP
PcPs namen	AD ane	VP
PR an	N geschrifte	PP

Table 4: The creation for the network for the sentence "si irteilt wart"

Source	Target	Interaction
AX werden	PP er	VP
AX werden	PcPs erteilen	VP

The verb *werden* can get assigned four different denominations, depending on the form and the role it has in a given sentence: when used as an auxiliary in combination with present and past participle, and infinitive verbs, it is listed as AX (auxiliary). When it is combined with a noun

or an adjective, it is listed as V (full verb). When it is connected to a modal verb, it is listed as IV (infinitive verb), regardless of the function it has in the sentence (auxiliary, full verb, or copula). When it appears in the present and past perfect constructions, it is listed as PcPs (past participle), regardless of the function it has in the sentence (auxiliary, full verb, or copula). Such differentiation permits to run the centrality measures on *werden* in these four different forms. After assigning to each element its denomination and to each connection between source and target nodes their syntactic relationship, the Excel sheets were imported in the following free online software for visualization: *Gephi* and *Cytoscape*.[10]

The network visualizations for (13) and (14) are as shown in Figures 2 and 3:

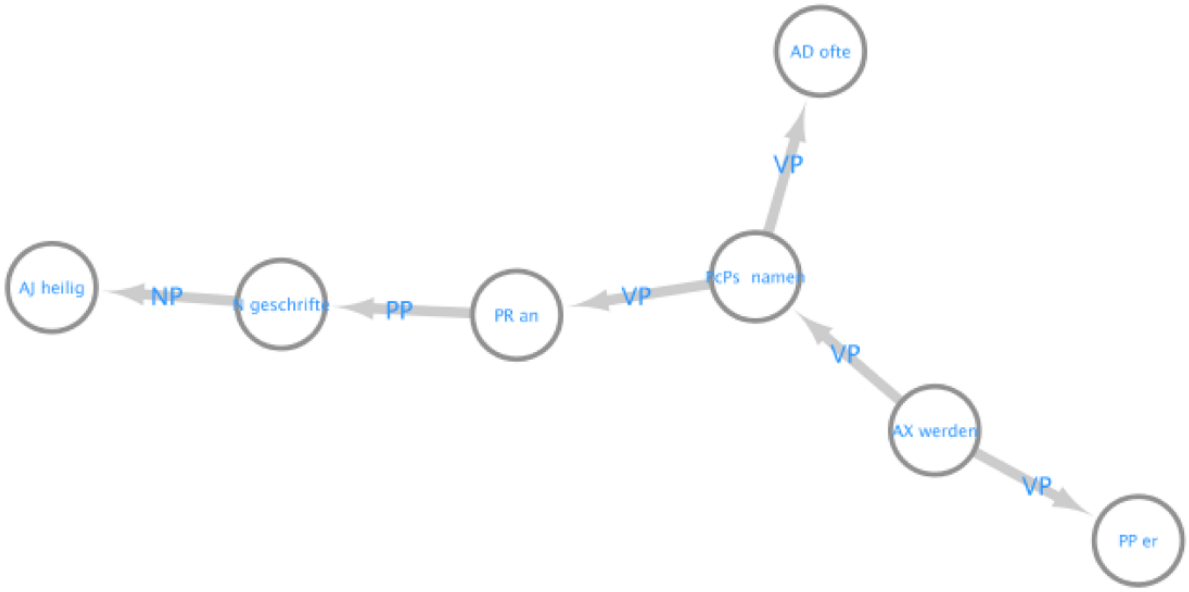


Figure 2: The network visualization for "uuiret er ofto an heligero geſcrifte genamit"

The visualization of this network shows the grammatical roles given to every node and the name assigned to every rule the connects these nodes. The auxiliary (AX) *werden* is the head of the verbal phrase (VP) with the personal pronoun (PP) *er* (he) and the past participle (PcPs) of the

[10] **https://cytoscape.org**.

verb *namen* (to name/to mention). The verbal phrase governed by *namen* is formed by the adverb *ofte* (often) and the preposition *an* (here 'in'). *An* governs the prepositional phrase composed by the noun *geschrifte* (writings) which, at the same time, governs the nominal phrase with the adjective *heilig* (holy). This network follows the constraints listed by Jurafsky and Martin (2017, p. 4): the auxiliary verb *werden* does not have any incoming arcs, since it is the root. Each vertex has one incoming edge, and there is a unique path from the root node to each other node.

Figure 3 shows the network of the sentence "si irteilt wart."

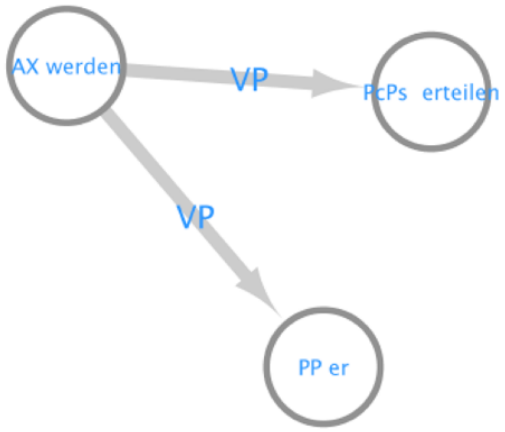


Figure 3: The network visualization for "si irteilt wart"

This network visualization also shows the syntactic relationships which connect the different elements in the sentence. The auxiliary (AX) *werden* governs the verbal phrase composed by the past participle (PcPs) of the verb *erteilen* (to condemn) and the personal pronoun (PP) *si* (she – here in the lemma form *er*). As observed for example (12), this network also follows the constraints listed by Jurafsky and Martin (2017, p. 4). As the auxiliary verb *werden* does not have any incoming arcs, each vertex has one incoming edge, and there is a unique path from the root node to each node.

After these networks were created, they were merged to build overall networks for both Middle and Early New High German. Each network is then accurately analyzed in *Gephy* using descriptive statistics that address the characteristics of the nodes and the edges. The first of these statistics is the *degree distribution*, which provides a summary of the connectivity in the graph (Kolaczyk, 2009, p. 81). The second one is the average degree, which is the number of edges

divided by the number of nodes and is a good indicator of the connectivity of a network (Jiang, Yu, & Liu, 2019) followed by the cluster coefficient, which measures the fraction of pairs of neighbors of a given center node that are connected (Yin, Benson & Leskovek, 2019). The last of these statistics is the average path length, which refers to the shortest path between nodes in a graph (Downey, 2018). To explore if these networks have small-world properties, for each century, I will create a random network with the same degree using *Cytoscape*. This software will be also used to determine if out-degree fits a power law.

After the implementation of the aforementioned statistical measurements, the networks were analyzed with the centrality measures described in the second part of chapter 2.

4.3 Summary

This chapter outlined the methodology that deals with aspects of this research project that relates to historical linguistics and network science. The first part of the methodology describes the data collection from the Middle High German and Early New High German corpora. The corpora contain religious and secular prose texts written between the 11th and the 17th centuries. All the instances of *werden* found in these corpora and the elements with which it was found are analyzed and discussed in separate tables. Raw and normalized frequencies for all the instances in the different corpora are then calculated. The results for both Middle High German and Early New High German are then compared.

The second part of the methodology illustrates the creation of the digital corpora from which I build dependency networks. Firstly, I assign every element in the target sentences a specific denomination that refers to its grammatical role. Secondly, I connect these elements through a set of rules that were constructed following the principles of dependency grammar (Jurafsky & Martin, 2017) which reflect the syntactic relationships between nodes. The networks are then accurately described using descriptive statistics that address the characteristics of the nodes and the edges. Lastly, centrality measures are implemented to gather evidence of preferential attachment.

In the next chapter, I offer a detailed account of the attestations of the verb *werden* found in the Middle High German corpus.

CHAPTER 5. MIDDLE HIGH GERMAN

This chapter focuses on how the attestations of *werden* were parsed, using texts from the *Referenzkorpus Mittlehochdeutsch* listed in the first section of the previous chapter. Firstly, I present the results for each century. Secondly, I discuss the data as a whole for Middle High German. The attestations were found using the lemma search option in the online corpus. As seen in the review of the literature, *werden* was used with elements such as modal verbs, participles, adjectives, infinitives, and nouns. The attestations included here are therefore centered on all the instances of *werden*, independently of the elements with which it is combined. For each attestation, I have indicated the tense (present, simple past, present- and past perfect). In this analysis, I keep the instances of *werden* used with modal verbs separated from the others in order to have a clear distinction between the modal constructions and the use of *werden* as the main verb.

5.1 11th Century

5.1.1 Analysis

A total of 288 instances for the 11th century was analyzed. Of these, 259 were included in the database. No forms of *werden* as a full verb were found. 29 attestations had missing annotations (indicated with the symbol [!] in the corpus) and were therefore excluded.

The following figure (Figure 4) shows the number of instances of *werden* in the present (*Präsens*), simple past (*Präteritum*), present perfect (*Perfekt*), and past perfect (*Plusquamperfekt*) tenses. Figure 4 does not include the instances of *werden* with modal verbs which are presented in a separate chart Figure 5.

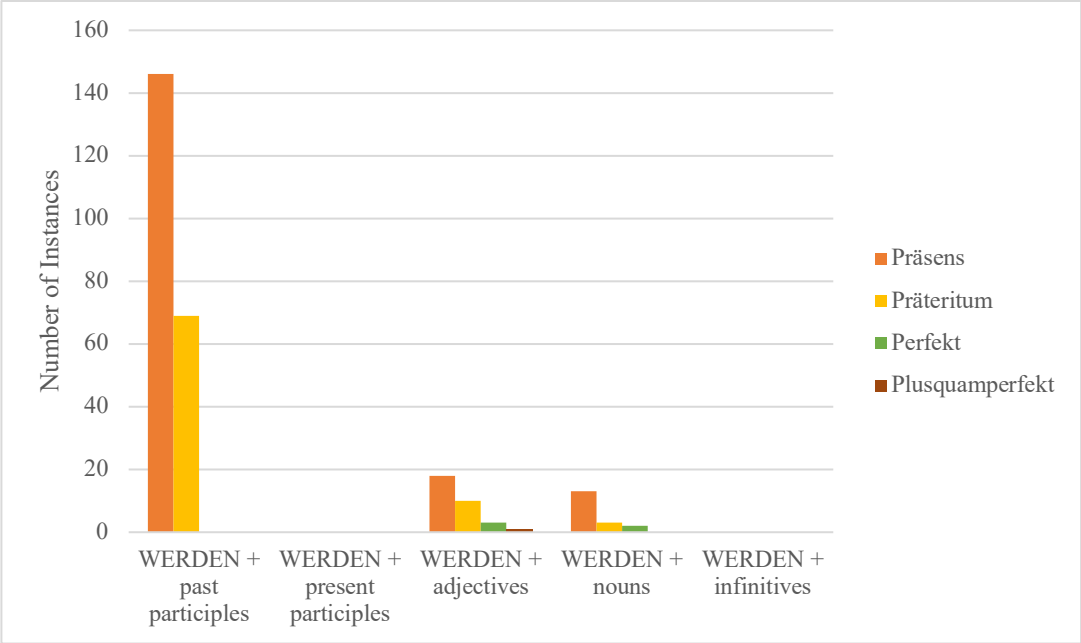


Figure 4: *werden* in the 11th century with past and present participles, adjectives, nouns, and infinitive verbs

W*erden* was found in the present, simple past, present, and past perfect tenses. In the majority of the instances, *werden* was used as an auxiliary for the passive, whereas a lower number of attestations referred to its use as a copula with adjectives and nouns. In these texts, no instances of *werden* with present participles and verbs in the infinitive forms were found. The highest number of occurrences of *werden* was in the present tense, and it was found combined with past participles (146 instances), as in example (16), with adjectives (18 instances), as in example (17), and nouns (13 instances), as in example (18):

(16) So da **geborin** **werdint** die sundere
So there be born-PTCS.PST become-AUX.PRS the sins
'so there are born the sins'
(*Rheinfränkische Interlinearversion der Psalmen*, 3va,7)

(17) **werde** ich ze **pulvere**
Become-COP.PRS I to dust-NOUN
'I become dust'
(*Wiener Notker*, 29,10)

(18) sô **wirdet** ez sâ **müede**
So becomes-COP.PRS er so tired-ADJ
'He got so tired'
(*Älterer Physiologus*, 27,107)

Werden in *Präteritum* was also found for the most part with past participles (69 instances), as shown in example (19). It was also found as a copula verb with adjectives (10 instances), as in example (20), and with nouns (3 instances), as in example (21):

(19) er vone Judae **verrâten** **wart**
He by the Jews betrayed-PTCS.PST became-AUX.PST
'He was betrayed by the Jews'
(*Wessobrunner Glaube und Beichte*, 104ra,10)

(20) **wurden** **unkreftic** mîne krefte
Became-COP.PST weak-ADJ my strenghts
'I became weak'
(*Wiener Notker*, 31,3)

(21) Daz ne **uuart** **alliz** nieth sines
That not become-COP.PST all-DEM not his
'That became not all his'
(Hoheliedkommentar, 17, ra23)

The corpus from the 11th century also shows a total of 6 instances of *werden* in the *Perfekt* tense with adjectives (4), and with nouns (1), as shown in examples (22) and (23).

(22) mîn ouge **ist** **trüebe** **worden** in zorne
my eye is-AUX.PRS bleak-ADJ became-COP.PTCP.PST in rage
'My eye has turned bleak because of rage'
(*Wiener Notker*, 30,10)

(23) Mîn uuíne **iſt** mir **uuórdan**
My wine is-AUX.PRS to me became-COP.PTCP.PST
édele **uuîntrûbo** uóne cýpro
noble sediment-NOUN from Cyprus
'My wine has become to me noble sediment from Cyprus'
(*Hoheliedkommentar*, 17, ra23)

There was an instance in the *Plusquamperfekt* in the subjective mood in which *werden* was combined with the interrogative pronoun "uvie" (how) that probably implies the presence of an adjective, as shown in the example (24):

(24) uvie daz obaz in der talaslathe **uvordan**
how-PRON the fruit in the gorge became-COP.PTCP.PST
uvare
would have-AUX.SBJV.PST
'how the fruit in the gorge would have become'
(*Hoheliedkommentar*, 17, ra23)

The next figure (Figure 5) shows the occurrence of *werden* with modal verbs:

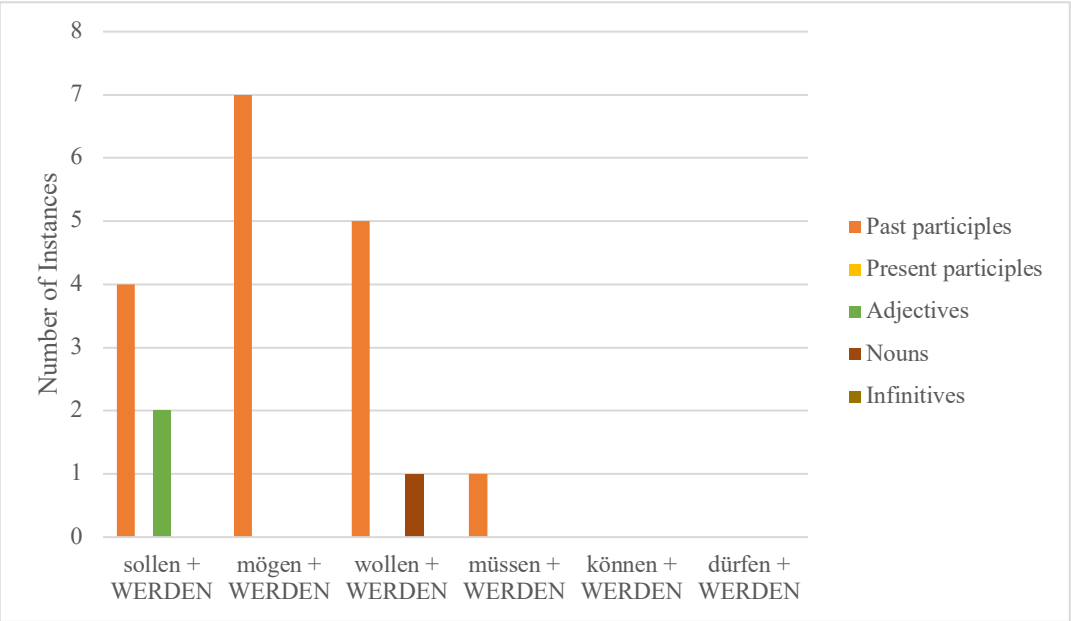


Figure 5: *werden* with modal verbs in the 11th century

A total of 20 instances were found with *werden* with the modal verbs *sollen* (should), *mögen* (to like), *müssen* (must), and *wollen* (want). Of these instances, 17 had *werden* with a past participle used as an auxiliary for the passive, as shown in example (25). In the remaining occurrences, *werden* was used as a copula with nouns and adjectives, as shown in example (26):

(25) dv bigrabin **woltost** **werden**
You buried- PTCP.PST wanted-AUX.PST become-AUX.INF
'You wanted to be buried'
(*Rheinauer Gebete*, 04v,14)

(26) sô siu swanger **werden** **sole**
so sie pregnant-ADJ become-COP.INF shall-AUX.PRS
'She shall become pregnant so'
(*Älterer Physiologus*, 32v,33)

The data from this century does not contain any instances of *werden* with the modal verbs *können* (can) and *dürfen* (to be allowed to).

5.1.2 Discussion

The data analyzed in this section indicated the use of *werden* in the texts of the 11th century that were included in this study. *Werden* was mainly used as an auxiliary for the passive in the *Präsens* and *Präteritum* tenses. However, when it was used as a copula with nouns and adjectives, it was also found in the *Perfekt* and Plusquamperfekt tenses. These instances are interesting for two reasons. On one hand, the passive (at least in this century), seems to be incompatible with the present and the past perfect, as no instances were found. On the other hand, in these instances, the past participles of *werden* did not add the *ge-* prefix. In Modern German, the past participle loses its prefix *ge-* only in the passive, while it keeps it when used as a copula verb (*es ist gemacht* ***worden*** vs. *es ist kalt* ***geworden*** – it has been done vs. it has gotten cold). According to Kotin (2000), in both Old and Middle High German, the *gi-/ge-* prefix was added to mark the completion of a given action (p. 328). The absence of this prefix in the instances from these periods has been considered as directly related to the intrinsic semantics of the specific verbs. For instance, verbs such as *bringen* (to bring) and *finden* (to find) did not add this prefix when used in their past participle forms in Old High and Middle High German (p. 239). The participles of *werden* without the *gi/ge*-prefix could suggest that *werden*, when used as a copula, was perceived as semantically close to these verbs. Such instances could also suggest that *werden*, at this stage, retained part of the original atomic-mutative meaning that Kotin (2003) observed in Old High German (p. 144).

The occurrences of *werden* with the modal verbs show an equal distribution between the use of this verb as an auxiliary and its use as a copula with adjectives and nouns. The absence of

instances with the modal verbs *können* (can) and *dürfen* (may) can be related to the fact that these verbs, during this period, were not fully grammaticalized. According to Diewald (2012), in Middle High German, *können* and *dürfen* were still subjected to some semantic and syntactic restrictions that limited significantly their use in the language (p. 299). These restrictions are probably the reason why the data of the 11th century did not contain any attestation of these modal verbs with *werden*.

5.1.3 Summary

The texts from the 11th century analyzed in this section contained a total of 259 instances of *werden*. The majority of them were in the present tense, but there were instances in the simple past, present, and past perfect. *Werden* was mostly used as the auxiliary to build the passive in *Präsens* and *Präteritum*, but it was used as a copula with adjectives and nouns also in *Perfekt* and *Plusquamperfekt*. When werden was found in these two tenses, the past participle did not add the *ge-* prefix.

The texts from the 11th century analyzed in this section also contained a total of 20 instances of *werden* with the modal verbs. These instances were with the modal verbs *sollen*, *mögen*, *müssen*, and *wollen*. The majority of the instances found were with *werden* used as an auxiliary for the passive. Further, no instances of *werden* with *können* and *dürfen* were found.

5.2 12th Century

5.2.1 Analysis

A total of 259 instances for the 12th century was analyzed. Of these, 254 were included in the database, whereas five of them had to be excluded due to missing annotation (indicated with the symbol [!] in the online corpora). As observed for the 11th century, no forms of *werden* used as a full verb were found.

The following figure (Figure 6) shows the number of instances of *werden* in the present (*Präsens*), simple past (*Präteritum*), present perfect (*Perfekt*), and past perfect (*Plusquamperfekt*) tenses. Figure 6 does not include the instances of *werden* with modal verbs which are presented in a separate chart (Figure 7).

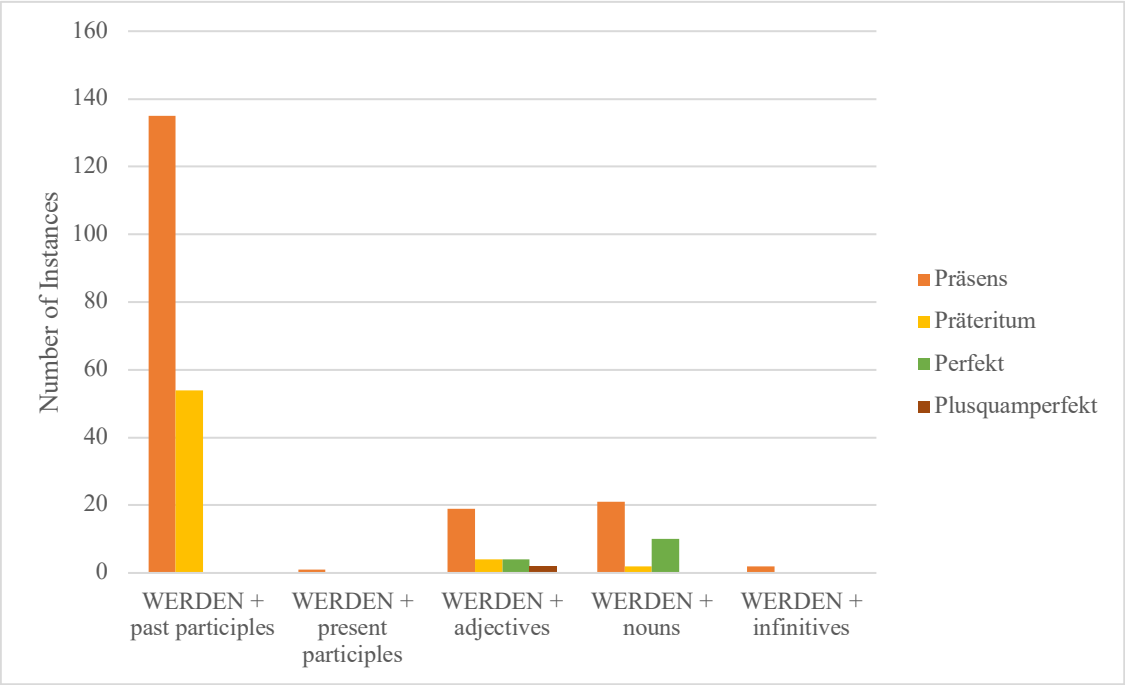


Figure 6: *werden* in the 12th with past and present participles, adjectives, nouns, and infinitive verbs

W*erden* was found in the present, simple past, present- and past perfect tenses. There was a total of 178 instances of *werden* used in the passive periphrases. These constructions were found in both the *Präsens* (135 instances) and the *Präteritum* (54 instances). The data also contained instances of *werden* with adjectives (29 in total) and with nouns (33 in total) in the *Perfekt* and *Plusquamperfekt* tenses. As observed in the instances analyzed for the 11th century, the instances in the present- and past perfect did not add the *ge-* prefix, as shown in examples (27) and (28):

(27) wande er hivte ein **mènnische** **worden** **ist**
why he today a human being-NOUN become- COP.PTCP.PST is-AUX.PRS
'why he has today become a human being'
(*Krakauer Fragmente*, 12r6)

(28) ob si an dem libe dem chinde **gelih** **worden**
if they at the body of the child similar-ADJ become-COP.PTCP.PST
waeren
would have been-AUX. SBJV.PST
'if they wound have become the same to the body of the child'
(*Krakauer Fragmente*, 12r6)

The texts of the 12th century also show examples of *werden* combined with a present participle, as shown in example (29), and twice with verbs in the infinitive forms, as shown in examples (30) and (31):

(29) so **wirt** er geiunget unte **gesehente**
so becomes-AUX.PRS he to get younger and seeing-PTCP.PRS
'so he becomes younger and able to see'
(*Wiener Physiologus*, 151r,0)

(30) sô **werde** ouch uns **vergeben**
so becomes-AUX.PRS also to us forgive-INF
'and to us they will forgive'
(*Alkuins Traktat*, 48r,3)

(31) so du **werdost** **ermanon**
so you become-AUX.PRS warn-INF
'so it will warn you'
(Bamberger Glaube, G142,15)

The following figure (Figure 7) shows the occurrence of *werden* with modal verbs:

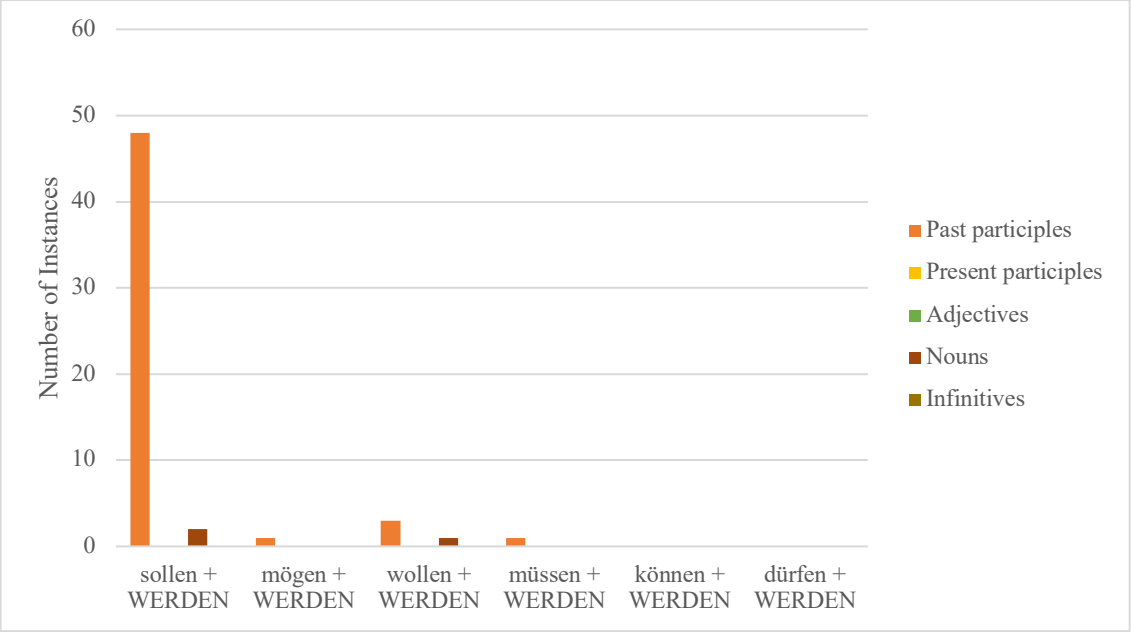


Figure 7: *werden* with modal verbs in the 12th century

A total of 56 instances were found with *werden* combined with the modal verbs *sollen* (should), *mögen* (to like), *müssen* (must), and *wollen* (want). Of these instances, 53 had *werden* with a past participle, while the remaining occurrences displayed *werden* as a copula in combinations with nouns and adjectives. *Sollen* appears to be the modal verb used most frequently, as shown in examples (32) and (33). Similar to what was found for the 11th century, no instances with the modal verbs *können* and *dürfen* were found.

(32) **solen** **bekêrt** **werden** vînde mîne hinderwert
Shall-AUX.PRS converted-PTCP.PST become-AUX.INF enemies mine back
'My enemies should be converted back'
(*Trierer Interlinearversion zum Psalter*, 16ra17)

(33) ob er gemartert **scolt** **werden**
if he martyrized -PTCP.PST should-AUX.PST become-AUX.INF
'..if he should be martyrize'
(*Krakauer Fragmente*, 2ra,14)

5.2.2 Discussion

The attestations from the 12th century do not differ much from what was found in the 11th century. The highest number of instances were with *werden* used as an auxiliary for the passive periphrases in the *Präsens* and the *Präteritum* tenses. The instances in which *werden* was used as a copula were additionally found in the present- and past perfect. As observed in the data of the 11th century, in these instances, the past participles of *werden* did not add the *ge*-prefix. Additionally, no forms of passive were found in these two tenses. On one hand, this suggests that, in this century, the passive still maintains the "apparent" incompatibility with both the *Perfekt* and *Plusquamperfekt* tenses. On the other hand, it also indicates that the meaning of *werden* (atomic-mutative) remained mostly unchanged from the previous century.

The texts from the 12th century also contain a single combination of *werden* with a present participle and two with infinitive verbs, which the texts of the 11th century did not include. These instances are extremely interesting, especially if framed within the discussion on the emergence of the German periphrastic future. As discussed in the review of the literature, occurrences of both *werden* plus present participles and *werden* plus infinitive verbs had been attested already in Old High German (Diewald and Wischer, 2013). The simultaneous presence of both these forms in the 12th century could indicate that the combinations with the present participles were probably not the source from which the periphrastic future emerged. Scholars such as Bechs (1901) have claimed that the development of this construction took place only in the 13th century with the simultaneous disappearance of the combinations of *werden* with the present participles. The instances found in this section seem to disprove such a theory while supporting the claims of scholars such as Kotin (2003). He argued that both the constructions with *werden* and the present participle and *werden* with the infinitive coexisted for a long time, and specifically from the moment in which they emerged in Old High German until the disappearance of the constructions with the present participles in Early New High German (p. 165). These instances could also suggest that the periphrastic future was already relatively well established in the language before the Early New High German period. This further supports the probable existence of the combination of *werden* plus infinitive with future time references in a much earlier period than the 15th or the 16th centuries, as also claimed by scholars such as Fritz (1997), Bogner (1989), and Kotin (2003).

The instances of *werden* with the modal verbs show a large use of *sollen* for the passive periphrases, while the absence of instances with *können* and *dürfen* suggests that these verbs were

still subjected to the same semantic and syntactic restrictions that have been observed in the 11th century.

5.2.3 Summary

The texts from the 12th century analyzed in this section contained a total of 254 instances of *werden*. The highest number of them were in the present tense, but there were instances in the simple past, present- and past perfect. *Werden* was used as the auxiliary to build the passive in *Präsens* and *Präteritum*, but it was also used as a copula with adjectives and nouns in *Perfekt* and *Pluquamperfekt*. When *werden* was found in these two tenses, the past participle did not add the *ge-* prefix. The data of the 12th century also contained one instance of *werden* with a present participle and two instances of *werden* with a verb in the infinitive form.

The texts from the 12th century also contained a total of 56 instances of *werden* and modal verbs. The majority of these instances were with *sollen*, whereas a few attestations have been found with *mögen*, *müssen*, and *wollen*. As observed for the 11th century, no instances of *werden* with the modals *können* and *dürfen* were found.

5.3 13th Century

5.3.1 Analysis

A total of 679 instances for the 13th century was analyzed. Of these, 675 were included in the database, whereas four of them had been excluded due to missing annotation (indicated with the symbol [!] in the online corpora). There was also an instance of *werden* used as a full verb.

The following figure (Figure 8) shows the number of instances of *werden* in the present (*Präsens*), simple past (*Präteritum*), present perfect (*Perfekt*), and past perfect (*Plusquamperfekt*) tenses. Figure 8 does not include the instances of *werden* with modal verbs which are presented instead in a separate chart (Figure 9).

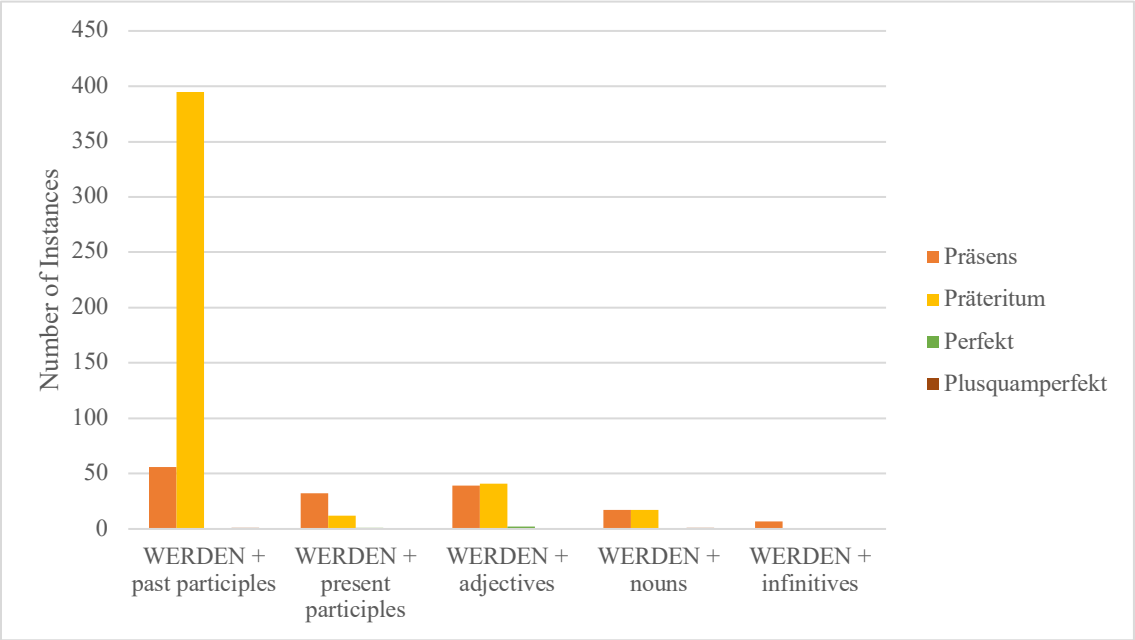


Figure 8: *werden* in the 13th with past and present participles, adjectives, nouns, and infinitive verbs

The verb *werden* was found in all four tenses: present, simple past, present, and past perfect. The data from the 13th century also confirms the trend observed in the 11th and 12th centuries. The majority of forms of *werden* were used in passive periphrases (452 forms in total), although in this century, the tense with the highest number of occurrences was the simple past (395 in total) instead of in the present (56 in total). There is also a modest number of forms of *werden* as a copula verb with adjectives and nouns (a total of 82 and 35 instances respectively), which were also found in the in *Perfekt* and *Plusquamperfekt* tenses. These instances still show past participles of *werden* without the *ge*-prefix, as shown in example (34). For the first time, the *Plusquamperfekt* was found in combination with a passive form, as shown in example (35):

(34) **ist** ez erger **worden**
Is-AUX.PRS it more evil-ADJ become- COP.PTCP.PST
'it has become more evil'
(*Schwabenspiegel*, 121rb,12)

(35) vnd alle die darab **geloubit** **warin** **wordin**
and all the there believed- PTCP.PST were-AUX.PST become-AUX.PTCP.PST
'and all those who have been believed'
(*Jenaer Martyrologium*, 58va1)

The data contained only one instance with *werden* used as a full verb, in which *werden* expresses the meaning of "come into being", as shown in example (36):

(36) da **wirt** daz da gescribin ist
There becomes-PRS what there written is
'There it comes into being what was written there'
(*Mitteldeutsche Predigten*, c3ra, 20)

There are several instances of *werden* combined with the present participles, some of them with the auxiliary verb in the *Präteritum*, as shown in examples (37) and (38):

(37) Ioseph **wart** **weinent** vnd sprach
Joseph became-AUX.PST craying- PTCP.PRS and talked
'Joseph started crying and talked'
(*Buch der Könige*, 3va 28)

(38) vnd **wart** si ser **drvkent**
and became-AUX.PST they very pressuring- PTCP.PRS
'and (he) started pressuring them'
(*Buch der Könige*, 03vb,23)

The data also show some instances of *werden* combined with verbs in infinitive forms, as shown in examples (39):

(39) ir dc nu **werdent** **sehende** dc mich die iuden
you the now become-AUX.PRS see-INF the me the jews
werden **vahende**
become-AUX.PRS catch-INF
'you will see that the Jews will catch me'
(*Schwarzwälder Predigten*, 009,va10)

The following figure (Figure 9) shows the occurrence of *werden* with modal verbs:

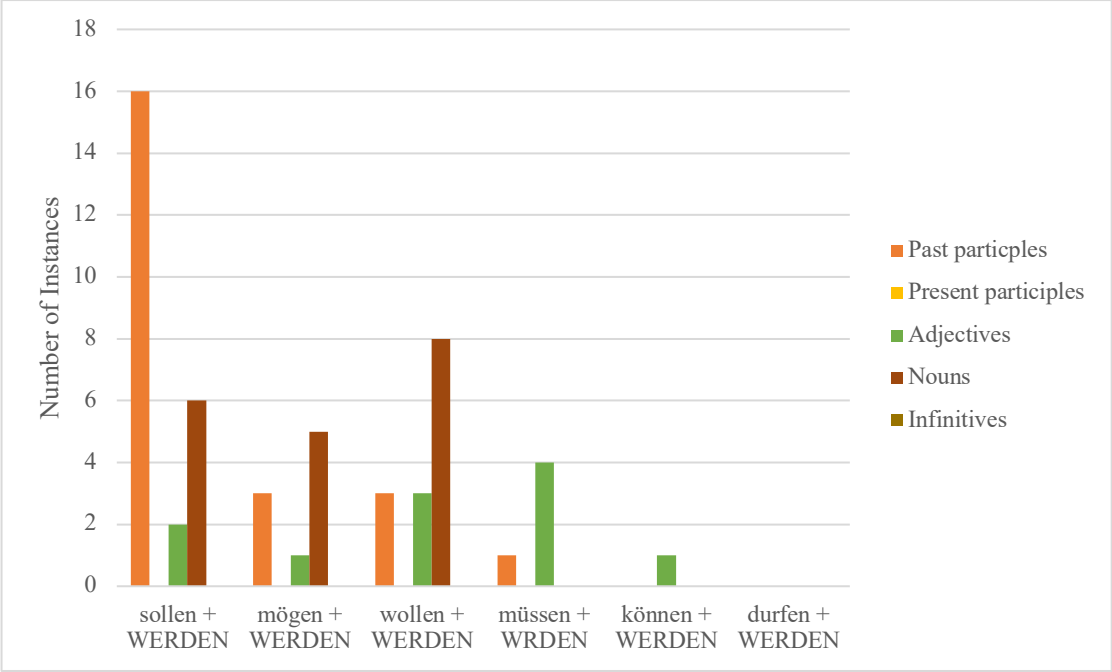


Figure 9: *werden* with modal verbs in the 13th century

A total of 53 instances of *werden* used with the modal verbs *sollen* (should), *mögen* (to like), *müssen* (must), and *wollen* (want) were found in the texts from the 13th century. *Sollen* is the modal verb used more frequently for the passive (16 instances), as observed also for the 11th and 12th centuries. The texts analyzed did not contain any forms of *werden* combined with *dürfen* but had an instance of *werden* copula and an adjective with the modal verb *können*, as shown in example (40):

(40) daz niemer zan **kan** **werden**
That no more submissive-ADJ can-AUX.PRS become-COP.INF
'That can't become submissive anymore'
(*Schwabenspiegel*, 115va,03)

5.3.2 Discussion

The raw frequency of the forms of *werden* in the 13th century is higher than the number of attestations found in the 11th and the 12th centuries. If the normalized frequency confirms such a trend (and this will be shown in the last section of this chapter), this, on one hand, could be related

to the higher number of texts that were available for this century. On the other hand, and as reported in the review of the literature, this can be also considered as evidence for the grammaticalization process of *werden* (and the related frequency increase) that took place in the latest stages of the Middle High German period (Kotin, 2003).

As seen in Figure (8), the data shows a high number of forms of *werde*n with present participles. In these attestations, *werden* was combined with verbs such as *sehen* (to see), *hören* (to hear), *wundern* (wonder), *bieten* (to prey), *sprechen* (to speak), *weinen* (to cry), *herrschen* (to rule), and *streiten* (to fight), which fall under the category of atelic verbs. These verbs differ from telic verbs such as *finden* and *bringen*, as they do not specify any temporal boundaries (Schumacher, 2005, p. 151). The high number of attestations in the present participle of atelic verbs in the data of the 13th century could be seen as a sign of the compatibility of these verbs with the atelic-affine semantics of the present participle (Kotin, 2000, p. 322). A comparison with the instances of *werden* and verbs in the infinitive forms reveals a similar distribution. The verbs found in the infinitive forms are also verbs such as *sehen*, *führen* (to lead), and *töten* (to kill). Consider again the examples (41) and (42):

(41) ier **werdent** mich ain clain wil niht **sehente**
you become-AUX.PRS me a short time not seeing-PTCP.PRS
und dar nach so **werdent** ier mich ain clain wil aber **sehente**
and then so become AUX.PRS you me a short time but seeing-PTCP.PRS
'You won't be seeing me for a while and then you will be seeing me shortly again'
(*Schwarzwälder Predigten*, 010,ra15)

(42) ir dc nu **werdent** **sehende** dc mich die iuden
you the now become-AUX.PRS see-INF the me the jews
werden **vahende**
become-AUX.PRS catch-INF
'you will see the Jews will catch me'
(*Schwarzwälder Predigten*, 009,va10)

The occurrences of *werden* in (41), in which it is combined twice with the present participles of *sehen*, and the ones in (42), in which it is found with the infinitive forms of the verbs *sehen* and *vahen*, indicate that the same or similar verbs could appear in both these periphrases. Kotin (2003) states that the present participle had a strong 'adjective-like nature" and indicated a long-lasting event or state (p. 152). When combined with *werden*, the present participle indicated

the entering into that particular event or state for an indefinite amount of time and independently of the type of verb (durative of resultative) (p. 152). Conversely, the periphrasis with the infinitives, which at the beginning could appear with *werden* in the *Präteritum*, were more similar to those with the past participles in that all the elements of these combinations had a significant "verbal-like nature" (p. 160). The infinitive did not specify any temporal boundaries, which were solely expressed by the semantics of the verb itself. Durative verbs kept their temporal indefiniteness (as happened with the present participle), but resultative verbs did not lose their telic "aspectual" features. The difference between these two periphrases lies therefore in the semantics of both the present participle and the infinitive. Hence, Kotin (2003) refers to these combinations as "twin constructions" that originated from different sources and coexisted in the language for long time (p. 165). The instances of both these periphrases found in the 12th and 13th centuries seem to support such claims.

The instances of *werden* with the modal verbs show a large use of *sollen* for the passive periphrases. While there are not instances of *dürfen,* as observed in the 11th and 12th centuries, the data shows the very first combination of *werden* with the modal verb *können*. *Werden* was used in this instance in combination with an adjective, thus used as a copula verb. The semantic and syntactic restrictions that have limited the use of *können* in the language until now (Diewald, 2012, p. 299) may be starting to become weaker, allowing the use of this modal verb in more communicative contexts.

5.3.3 Summary

The texts from the 13th century analyzed in this section contained a total of 674 instances of *werden*. The highest number of them were in the *Präteritum*, but there were instances in the present, and present and past perfect. *Werden* was used as the auxiliary to build the passive in *Präsens* and *Präteritum*, but it was also used as a copula with adjectives and nouns in *Perfekt* and *Plusquamperfekt*. As observed for the 11th and the 12th centuries, when *werden* was found in the present and past perfect, the past participle didn't have the *ge-* prefix. The data of the 13th century also contained a high number of *werden* combined with a present participle and some instances of *werden* with a verb in the infinitive form.

The texts from the 13th century analyzed in this section contained a total of 53 instances of *werden*. The majority of the instances were with *sollen*, whereas a few examples have been

found with *mögen*, *müssen*, and *wollen*. Only one instance was found with *können*, whereas no attestations of *werden* with the modal verb *dürfen* were found.

5.4 14th Century (First Half)

5.4.1 Analysis

A total of 502 instances were found in the texts from the first half of the 14th century. Of these, 500 were included in the database. There were two attestations of *werden* used a full verb, whereas two attestations had missing annotations (indicated with the symbol [!] in the online corpora) and were therefore excluded.

The following figure (Figure 10) shows the number of instances of *werden* in the present (*Präsens*), simple past (*Präteritum)*, present perfect (*Perfekt*), and past perfect (*Plusquamperfekt*) tenses. Figure 10 does not include the instances with modal verbs which are presented in a separate figure (Figure 11).

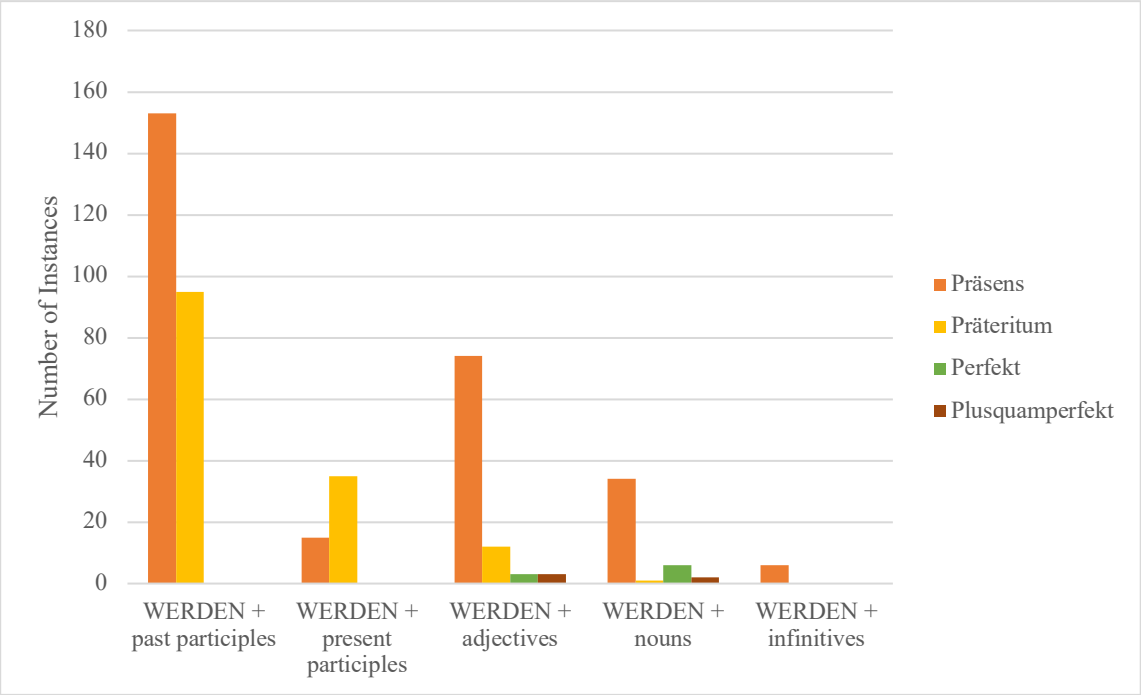


Figure 10: *werden* in the first half of the 14th century with past and present participles, adjectives, nouns, and infinitive verbs

As observed for the 11th, 12th, and 13th centuries, the verb *werden* was found in all four tenses: present, simple past, present- and past perfect. Overall, the data from the first half of the 14 century is similar to the instances found in the previous centuries. The main use of *werden* is for the passive constructions, both in the *Präsens* (153 instances) and in the *Präteritum* (95 instances). A high number of instances was found with *werden* used as a copula in combination with adjectives and nouns (92 and 43 instances respectively). The data contain numerous instances of *werden*, both in the *Präsens* and in the *Präteritum* (15 and 35 respectively), combined with the present participle and some instances of *werden* with verbs in the infinitive form (6 instances). The constructions with the present participles found in this century convey the meaning of entering into the particular event or state expressed by the verb, as shown in examples (43) and (44):

(43) an dem dritten tage **wirt** er **vf stande** von dem tovde
On the third day becomes-AUX.PRS he **rising**-PTCP.PRS from the death
'On the third day, he will rise from death'
(*Nikolaus von Straßburg: Predigten*, 37ra,17)

(44) Du an irre sucheden me **minende** inde me
You at her the same more loving-PTCP. PRS and more
begerende **wart**
desiring-PTCP.PRS became-AUX.PST
'You started loving and desiring more like her'
(*Rede von den 15 Graden*, 089r,15)

The instances with the infinitive verbs express a comparable meaning to the same type of constructions in Modern German, as shown in examples (45) and (46):

(45) daz man z doch von im **wirt** **vber haben**
that one it anyway from him will-AUX.PRS have-INF
'that on will have from him anyway'
(*Würzburger Polizeisätze*, 250va,15)

(46) mit dem **wirdet,** **tragen** ûf daz hûs, daz zeichene
with that becomes-AUX.INF carry-INF on the house the sign
'(he-she) will carry the sign on the house with that'
(*Nürnberger Stadtbuch*, 10ra,14)

The instances in the *Perfekt* and the *Plusquamperfekt* tenses appear also in this century combined only with *werden* used as a copula with nouns and adjectives without the *ge*-prefix, as shown in examples (47) and (48):

(47) die seluen die drunken **worden** **sint**
the same that drunk-ADJ become-PTCP.PST are-AUX.PRS
'…the same the got drunk'
(*Rede von den 15 Graden*, 081v,09)

(48) Vnde do (Jesus) **was** **worden** zwelf iar **alt**
And the (Jesus) was-AUX.PST become-COP.PST twlve years old-ADJ
'and then Jesus had become twelve years'
(*Evangelienbuch des Matthias von Beheim*, 139va,5)

Among the occurrences of *werden* used as a copula verb, there was also one instance of *werden* in the infinitive form combined with the verb *machen* (to do) that functions as a sort of auxiliary, as shown in example (49):

(49) kumit nach mir vnd ich **mache** uch **werden**
come to me and I do-AUX.PRS you become-COP.INF
fisschre der lute
fisher-NOUN of the people
'Come to me and I will make you a fisher of people'
(*Evangelienbuch des Matthias von Beheim*, 058r,18)

This particular combination was used in Middle High German to connote the action as voluntarily initiated by the subject and it served also to express the lexical aspect of a verb (Schwerdt, 2008, p. 73). In this example, the agency of the subject is emphasized by the verb *machen*, which implies that a voluntary action is taking place.

In the same text, two instances of *werden* used as a full verb were found, as shown in examples (50) and (51):

(50) Din wille der **werde** alse in dem himele
Your will that becomes.PRS also in the sky
'And your will become true also in the sky'
(*Evangelienbuch des Matthias von Beheim*, 061r,13)

(51) Wan wo din schatz ist. Do **wirt** ouch din herze
When where your treasure is there become-PRS also your heart
'Where your treasure is, there will be also your heart'
(*Evangelienbuch des Matthias von Beheim*, 061v,10)

In these two examples, *werden* is used again in the sense of "coming into being".

The following figure (Figure 11) shows the number of instances of *werden* with the modal verbs:

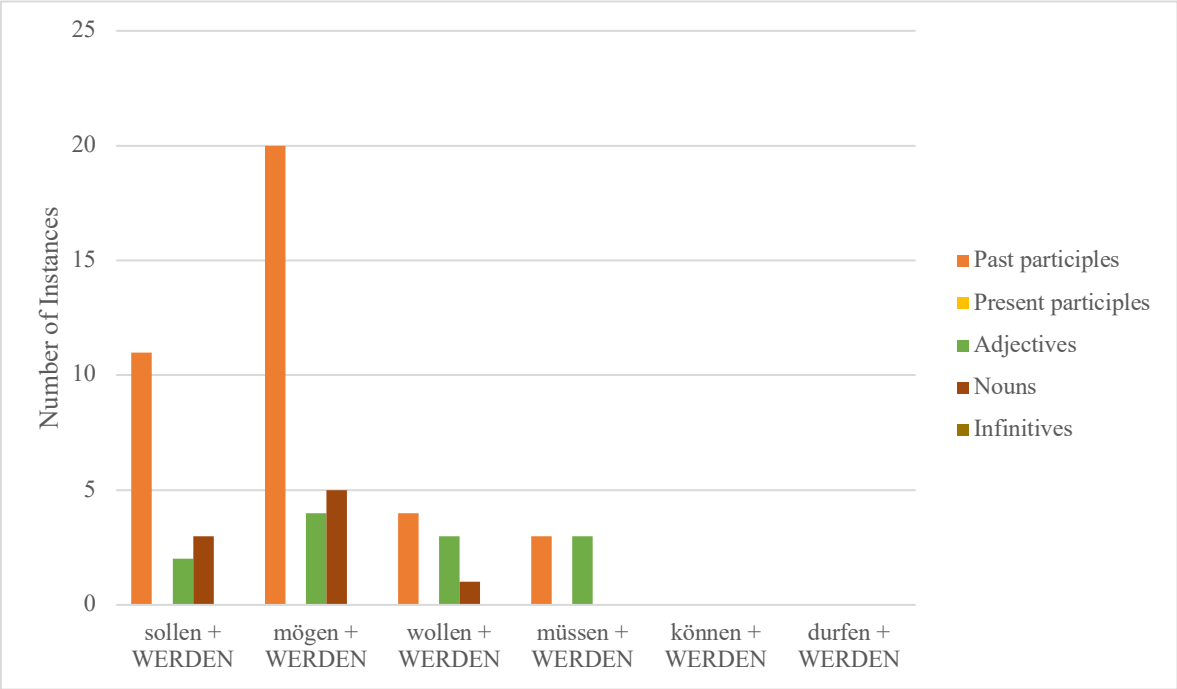


Figure 11: *werden* with modal verbs in the first half of the 14th century

A total of 39 instances were found with *werden* combined with the modal verbs *sollen* (should), *mögen* (to like), *müssen* (must), and *wollen* (want). Similar to the data of the 11th, 12th, and 13th centuries, the highest number of occurrences in the modal constructions is with *werden* used as an auxiliary for the passive periphrases. In this century, however, *mögen* is the most frequent verb used with *werden* and past participles, as shown in example (52), with nouns, as shown in example (53), and with adjectives, as shown in example (54). Further, no forms of *werden* were found in combination with *können* and *dürfen*.

(52) he **mac** mit werken der minnen **behalten** **werden**
he likes-AUX.PRS with deeds of the love kept-PTC.PST become-AUX.INF
'He may be kept with the deeds of the love'
(*Rede von den 15 Graden*, 84va,14)

(53) vnd **mag** im niemer me **werden** wesliches **lones**.
and may-AUX.PRS to him no more become-COP.PRS a true reward-NOUN
'and it may not be for him a true reward anymore'
(*Nikolaus von Straßburg*, 46ra,06)

(54) Dero **mugen** vins **tusent** **werden**
Of it may-AUX.PRS to us thousand-ADJ become-COP. INF
'We may become a thousand of them'
(*Nikolaus von Straßburg*, 46rb, 21)

5.4.2 Discussion

The data from the first half of the 14th century present a higher number of attestations in comparison to the 11th and the 12th centuries. This number is, however, smaller when compared to the instances found in the 13th century. This could be related to a difference in the type of texts analyzed for this century and to different styles of the authors of these writings.

Figure 10 shows a significant number of instances of *werden* with the present participle and some of them are with the verbs in the infinitive forms. The verbs in the present participle were verbs such as *sehen* (to see) and *beten* (to pray), *hassen* (to hate), *kreuzigen* (to crucify), *verdammen* (to condemn), *fangen* (to catch), and *achten* (to respect/to pay attention to something). *Werden* plus infinitive was found with verbs such as *hassen* (to hate), *kaufen* (to buy), *erkennen* to (recognize), and *lachen* (to laugh). The overlap of the types of verbs used in both the periphrases of *werden* with the present participle and those of *werden* with the infinitive verbs could be seen as further evidence for Kotin's (2003) claims about the existence of "twin constructions."

The instances of *werden* with the modal verbs show the highest number of instances with the verb *mögen* used in the passive periphrases and with *werden* as a copula verb. There are also attestations with the modals *sollen*, *müssen*, and *wollen*. The data show, however, no instances of *werden* combined with *dürfen* or *können*. This confirms the trend observed in the 11th, the 12th, and the 13th centuries.

5.4.3 Summary

The texts from the first half of the 14th century analyzed in this section contained a total of 500 instances of *werden*, and in two of them, *werden* was used as a full verb. The highest number of attestations were in the *Präsens* tense, but there were instances in the simple past, and present- and past perfect. *Werden* was mostly used as the auxiliary to build the passive in the present and the simple past, but it was also used as a copula with adjectives and nouns in the *Perfekt* and *Plusquamperfekt* tenses. As observed for the 11th, 12th, and the 13th centuries, when *werden* was found in the present and past perfect, the past participle did not add the *ge-* prefix. The data of the first half of the 14th century also contained a high number of *werden* combined with a present participle and some instances of *werden* with a verb in the infinitive form.

The texts from the first half of the 14th century analyzed in this section also contained a total of 39 instances of *werden* combined with the modal verbs. The majority of these instances were with *mögen*, whereas some occurrences have been found with *sollen*, *müssen*, and *wollen*. No instances of *werden* with *können* or *dürfen* were found.

5.5 *Werden* in Middle High German

5.5.1 Analysis

A total of 1,851 forms of *werden* were analyzed in the texts from Middle High German. Among those, there were also three attestations of *werden* used as a full verb (one found in the 13th century, and two found in the first half of the 14th century). The following figure (Figure 12) shows the raw and normalized[11] frequencies of the instances of *werden* with the elements it was combined for each century analyzed in this chapter.

[11] In this chapter, the word count has been normalized per 10,000 words.

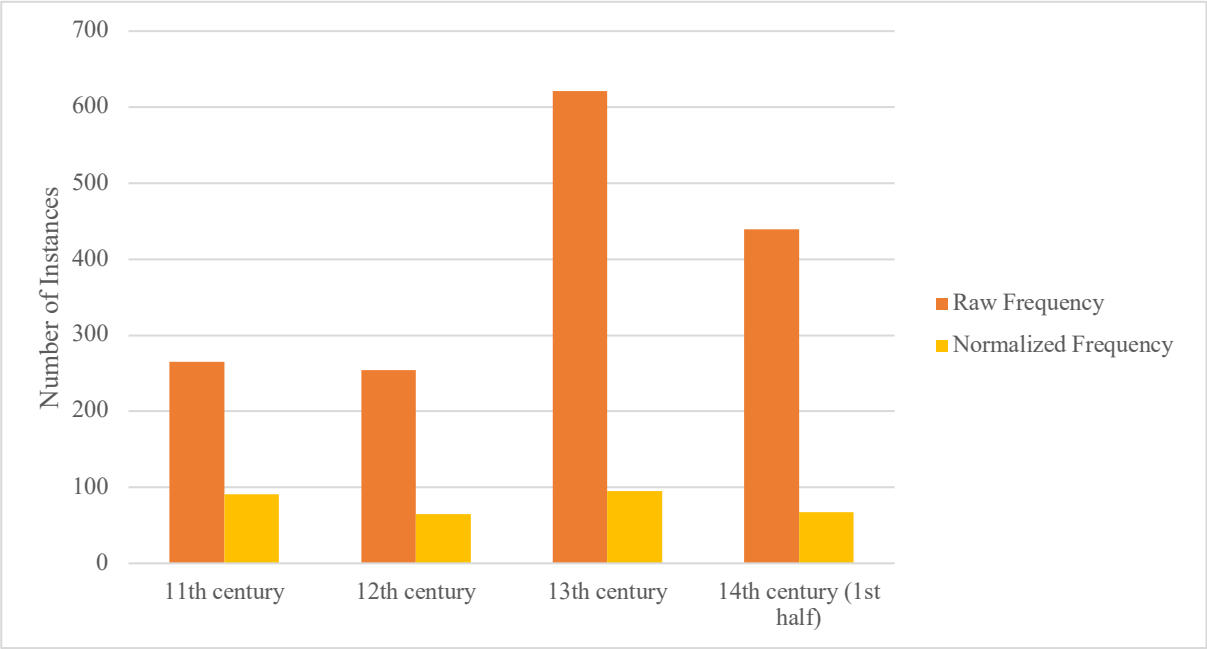


Figure 12: Raw and normalized frequencies of the instances of *werden* in Middle High German with past and present participles, adjectives, nouns, and infinitive verbs

Figure 12 indicates that instances of *werden* were found in all four centuries analyzed in this chapter and, except for the 11th century from which no texts were available for East Middle German, in all dialect areas included in this study. The normalized frequency shows that the centuries in which *werden* was found with most frequency were the 11th and the 13th centuries. The highest number of attestations in these two centuries could depend on the type of texts included in the corpus and it does not necessarily imply that the passive *werden* went through phases of greater and lesser popularity among the Middle High German writers. Further, the number of attestations in both the 12th and 14th centuries is very similar, meaning that the use of *werden* for the passive constructions and as a copula was similar even in those centuries in which the number of attestations is not as high as in the other two centuries.

Figure 13 shows the raw and normalized frequencies of *werden* and the elements with which it was combined in the corpus of Middle High German:

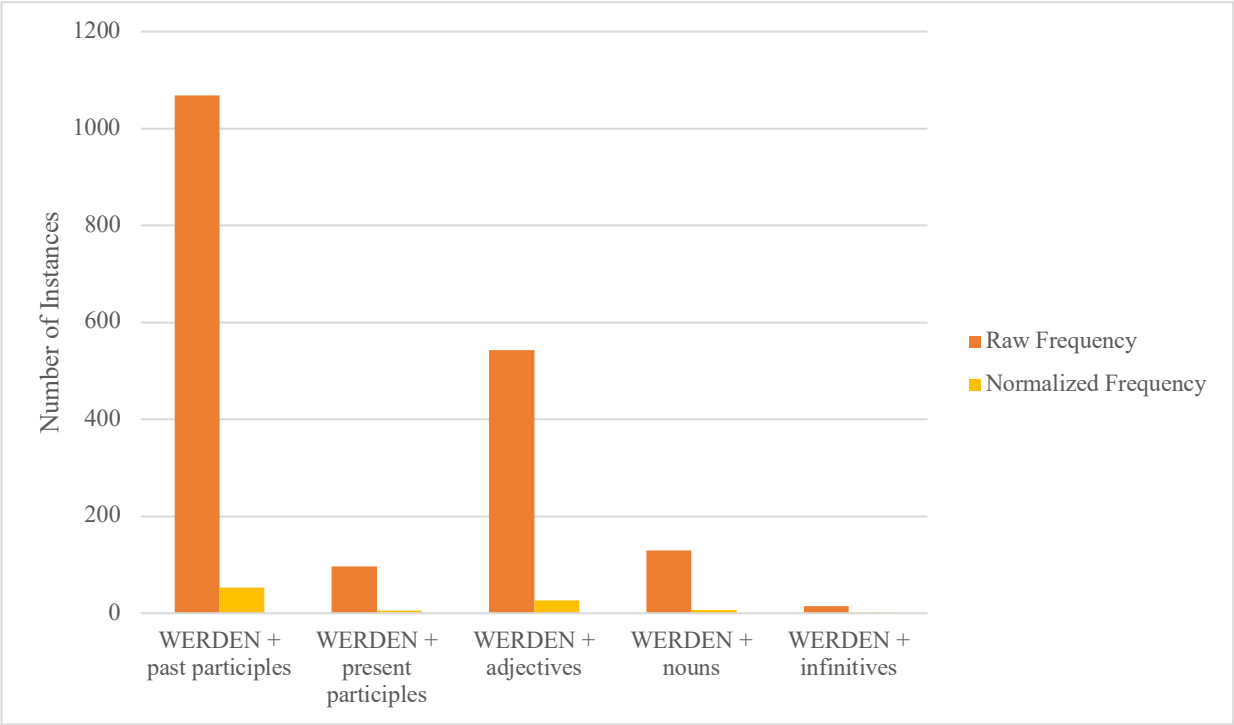


Figure 13: Raw and normalized frequencies of the instances of *werden* with past and present participles, adjectives, nouns, and infinitive verbs

Figure 13 shows that the main use of *werden* in Middle High German was for the passive constructions, followed by *werden* used as a copula in combination with adjectives and nouns. The lowest number of instances was with *werden* used as auxiliary with the present participles and the verbs in the infinitive forms.

The next figures (Figure 14 and Figure 15) show the raw (14) and normalized (15) frequencies of all the elements found with *werden* for each century:

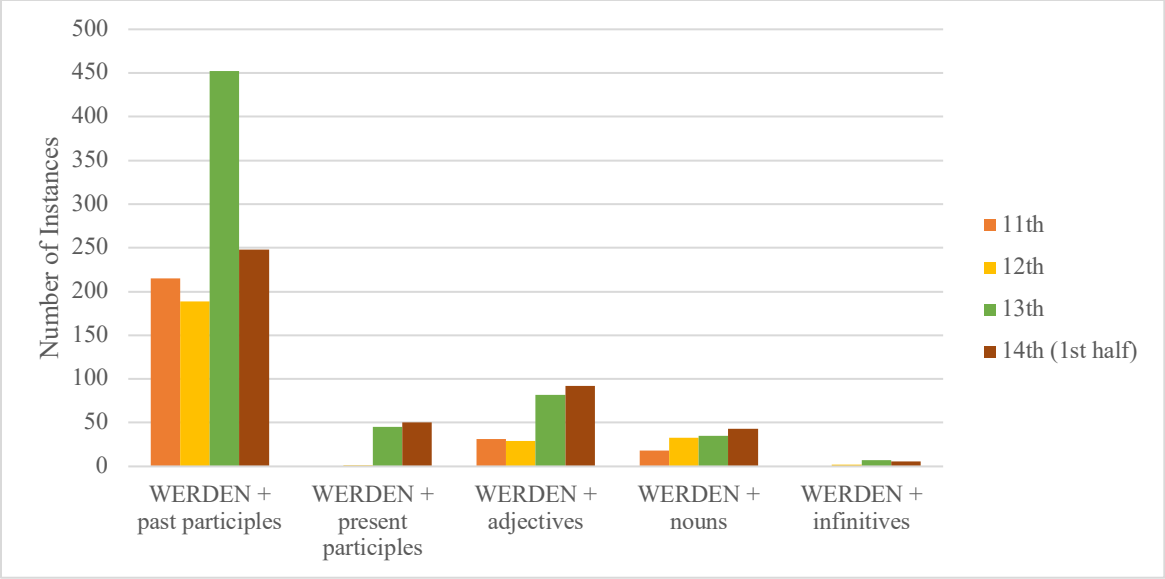


Figure 14: Raw frequency per century of the instances of *werden* with past and present participles, adjectives, nouns, and infinitive verbs

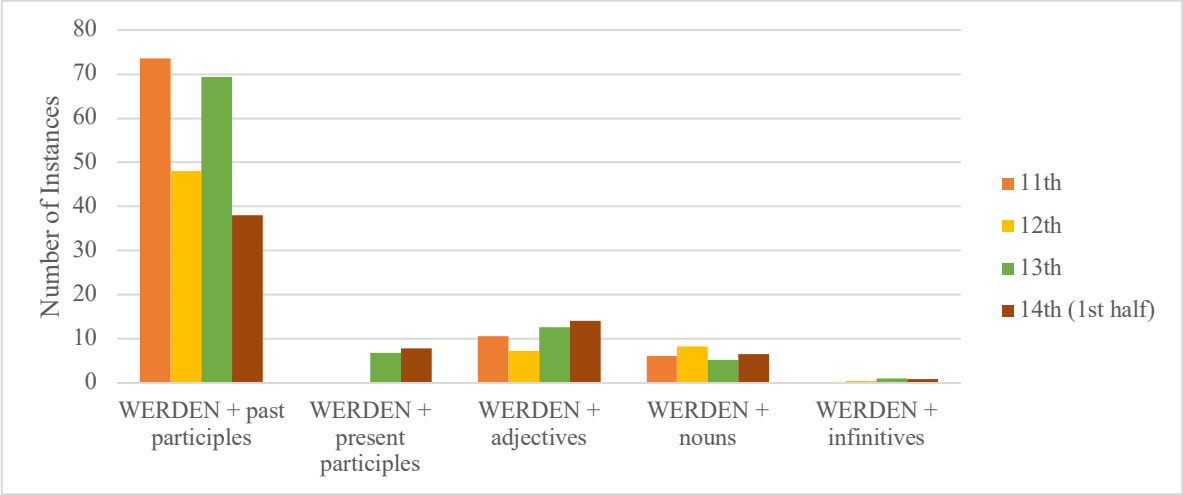


Figure 15: Normalized frequency per century of the instances of *werden* with past and present participles, adjectives, nouns, and infinitive verbs

The comparison between the raw and normalized data indicates that the century in which the passive is used with most frequency is the 11th, which is the century that has the lowest number of texts (and words) available. The first half of the 14h century had the highest number of attestations of *werden* used as an auxiliary with the present participle and as a copula in combination with adjectives and nouns. This century, together with the 13th century, also had the

highest number of instances of the use of *werden* as an auxiliary combined with the verbs in the infinitive forms.

The following figure (Figure 16) shows the raw and normalized frequencies with the *werden* attestations with a focus on the tenses in which these instances were found:

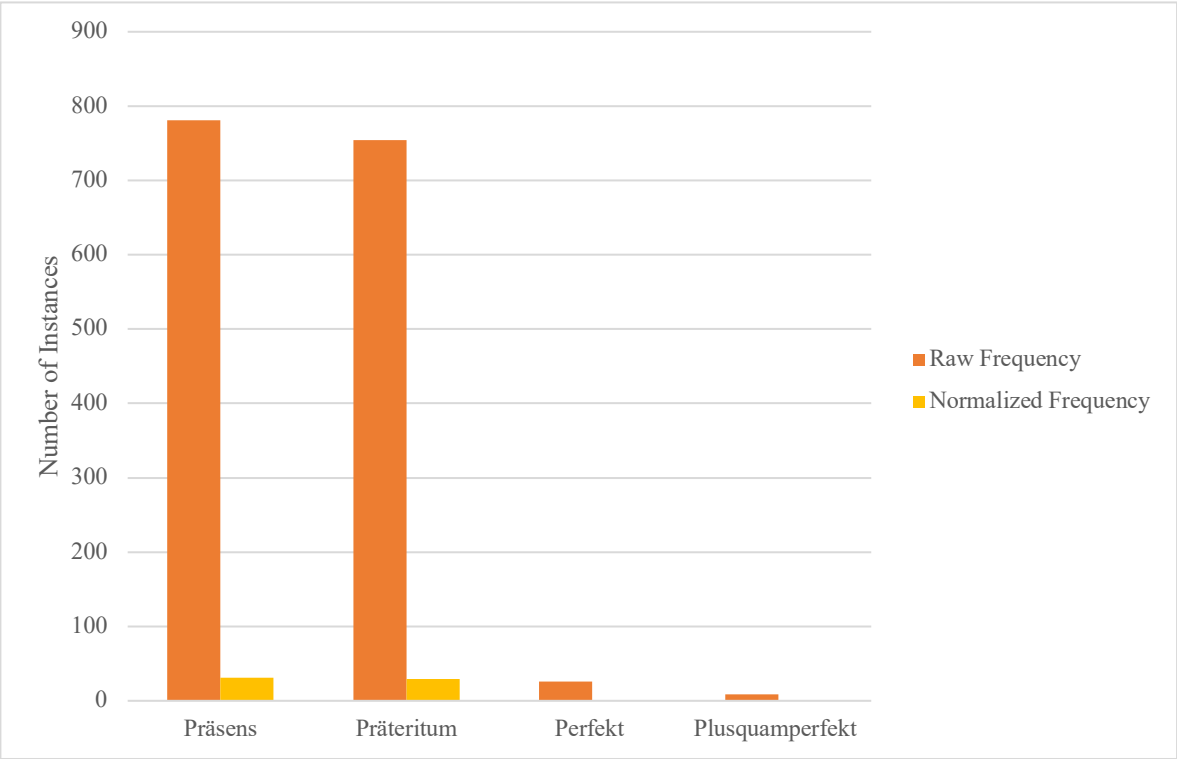


Figure 16: Raw and normalized frequency of the instances of *werden* according to the tenses

The raw and normalized frequency data displays a comparable number of attestations of *werden* found in the *Präsens* and the *Präteritum*. The *Perfekt* and the *Plusquamperfekt* are still less commonly used in this corpus, a probable indicator of the fact that these tenses were still in the early stages of their grammaticalization process.

The following figures (Figure 17 and Figure 18) display the raw and normalized frequencies of the attestations of *werden* in the present (*Präsens*), simple past (*Präteritum)*, present perfect (*Perfekt*), and past perfect (*Plusquamperfekt*) tenses with a focus on the centuries in which they were found:

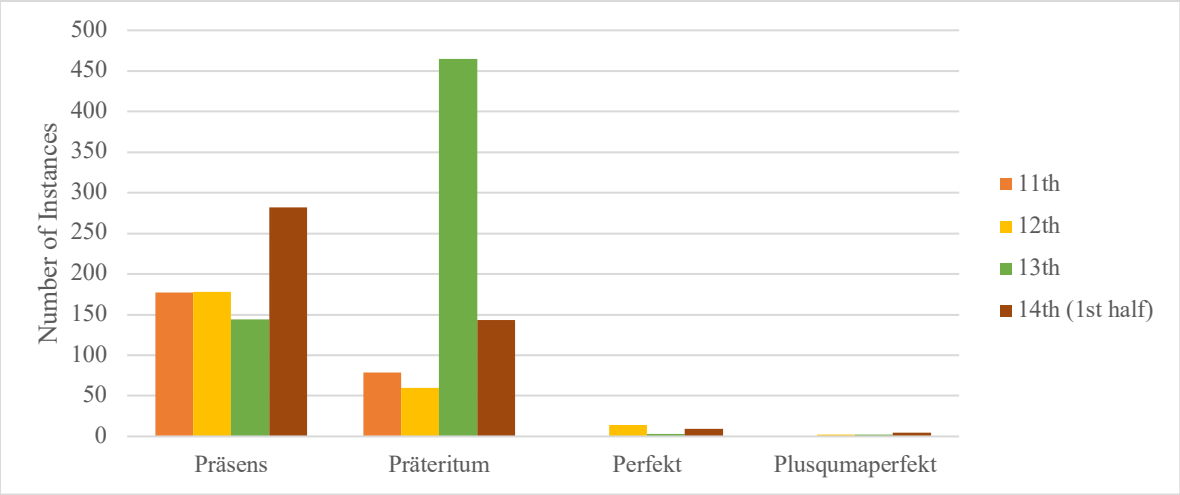


Figure 17: Raw frequency per century of the instances of *werden* according to the tenses

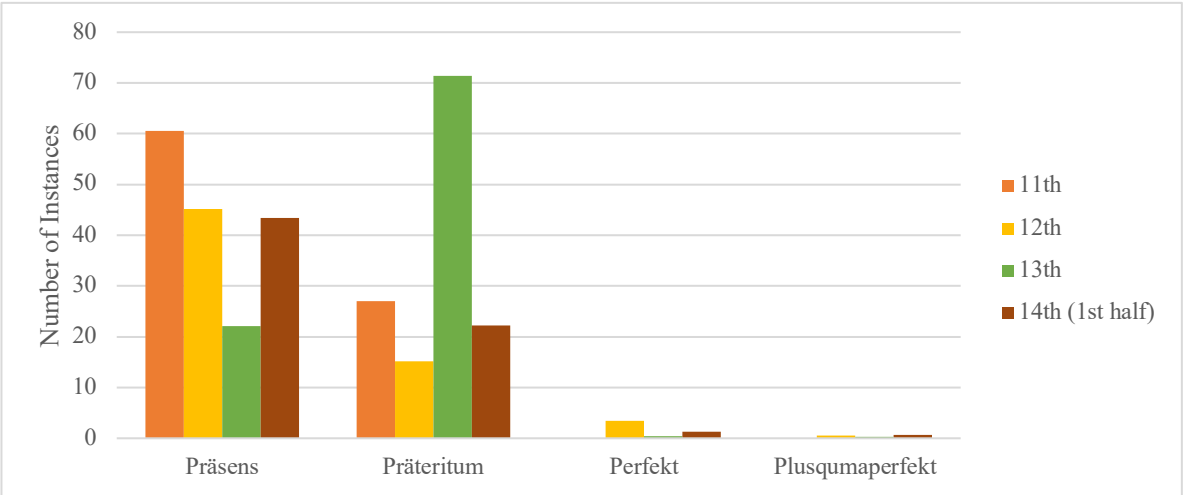


Figure 18: Normalized frequency per century of the instances of *werden* according to the tenses

The raw and normalized frequency data show that, in this corpus, *werden* in the *Präsens* was most frequently used in the 11th century, while the *Präteritum* was found with the highest

frequency in the 13th century. *Perfekt* and the *Plusquamperfekt* are the tenses with the lowest number of instances in all four centuries.

The next figures (Figure 19 and Figure 20) display the raw and normalized frequencies of the elements combined with *werden* with focus on the tenses:

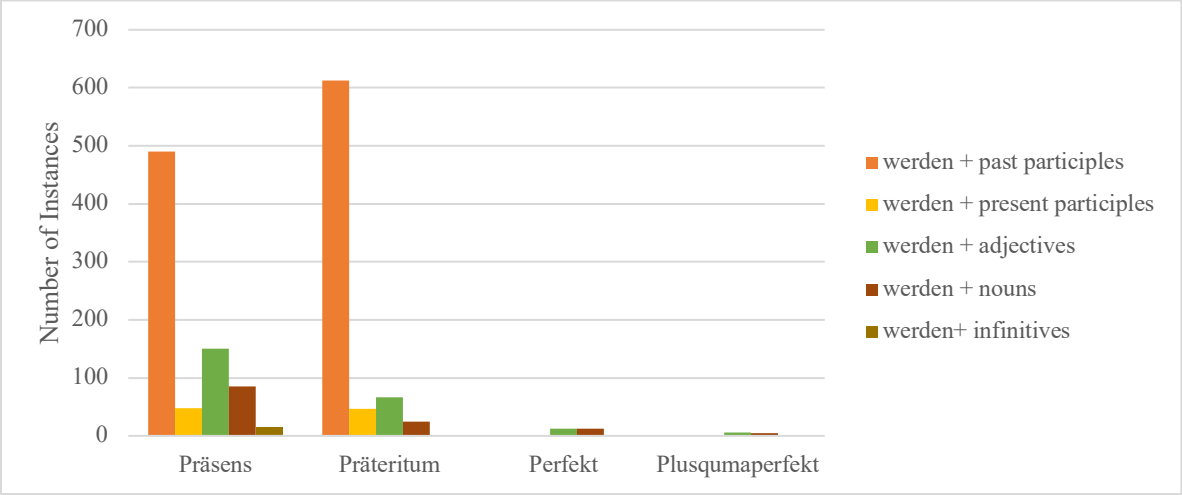


Figure 19: Raw frequency of the elements combined with *werden* according to the tenses

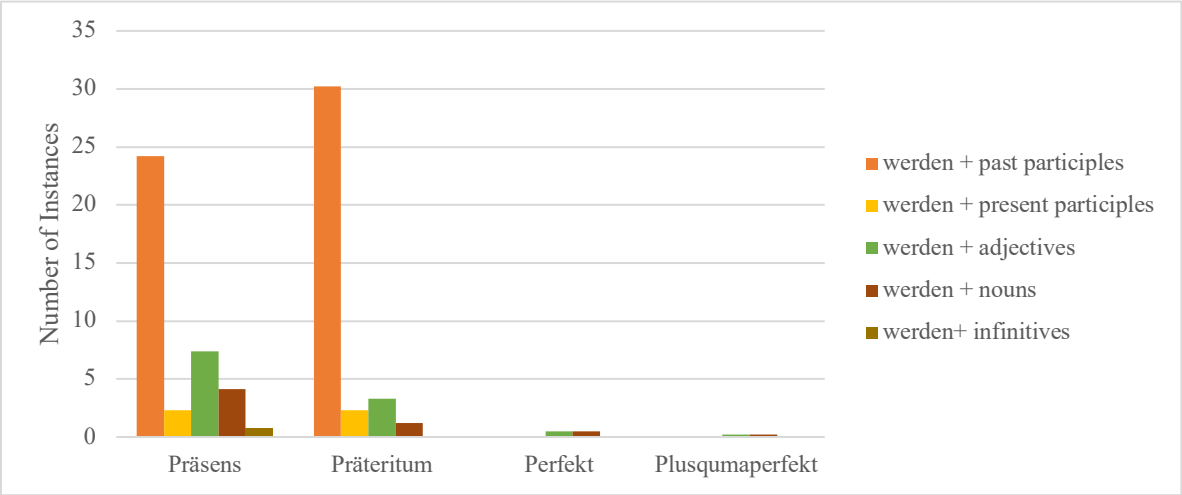


Figure 20: Normalized frequency per century of the instances of *werden* according to the tenses

Both figures show that the passive constructions were used with almost the same frequency in both the *Präteritum* and the *Präsens* tenses, and the same can be argued for *werden* used in combination with the present participles. The preferred tense for the use of *werden* as a copula

verb with nouns and adjectives and for the combinations with the verbs in the infinitive forms was the *Präsens*. The *Perfekt* and *Plusquamperfekt* are the tenses in which nouns and adjectives appear with the most frequency, as only one instance of the *werden* passive was found in the past perfect.

A total of 186 modal verbs combined with *werden* have been found in the corpus of Middle High German. The next figure (Figure 21) shows the raw and normalized frequencies of the combinations of *werden* with the modal verbs.

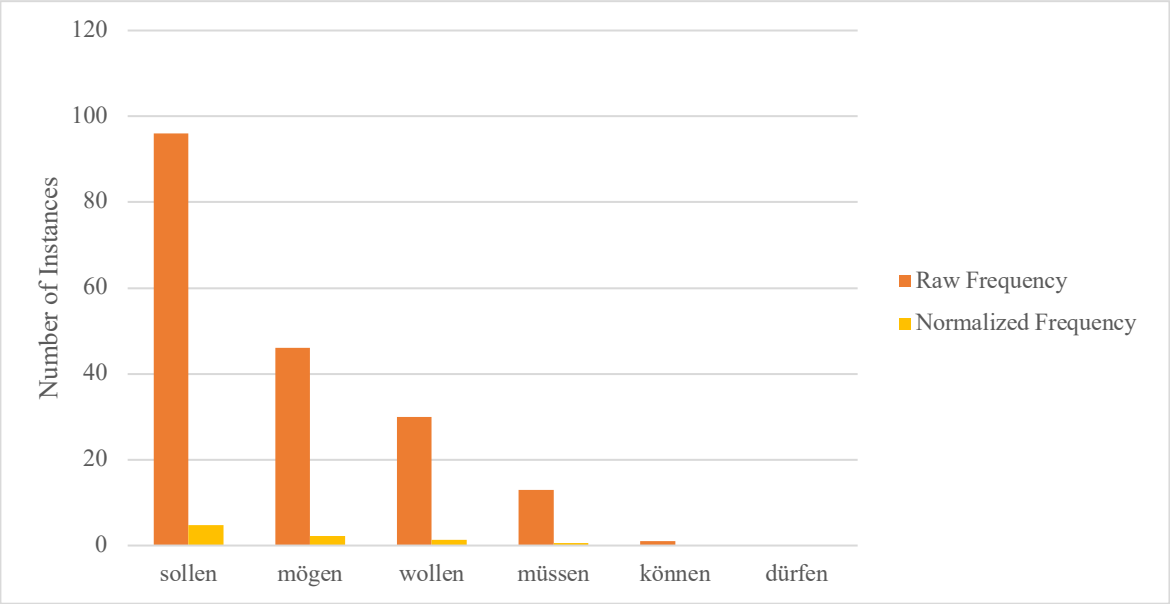


Figure 21: Raw and normalized frequency of the instances of *werden* with modal verbs

Figure 21 shows that the modal verb used with most frequency is *sollen*, followed by *mögen*, *wollen*, and *müssen*. Only one form of *können* was attested, while no forms of *dürfen* were found.

The next figures (Figure 22 and Figure 23) show the raw and normalized frequencies of the combinations of modal verbs and *werden* used as an auxiliary or copula:

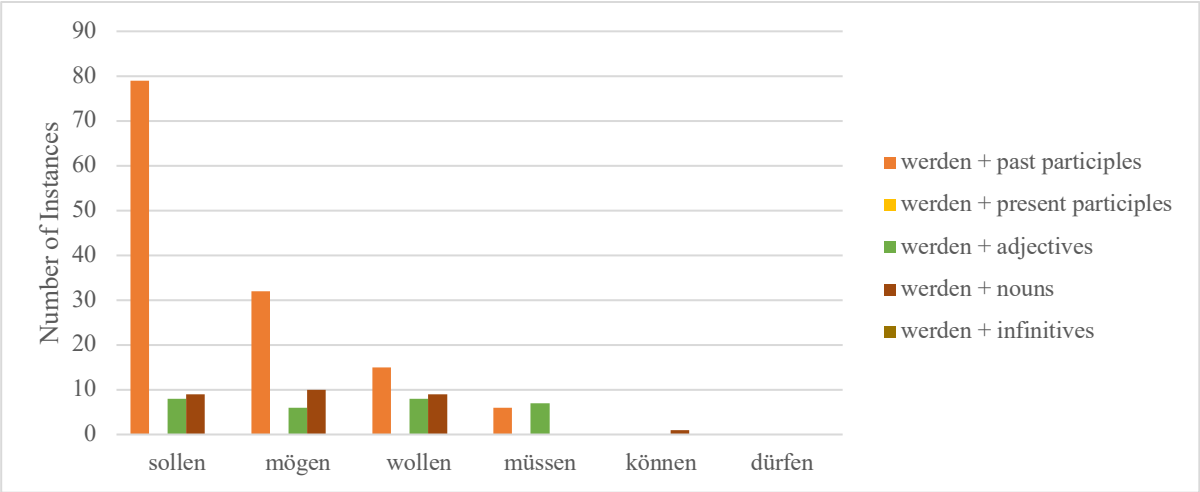


Figure 22: Raw frequency per century of the instances of *werden* according to the tenses

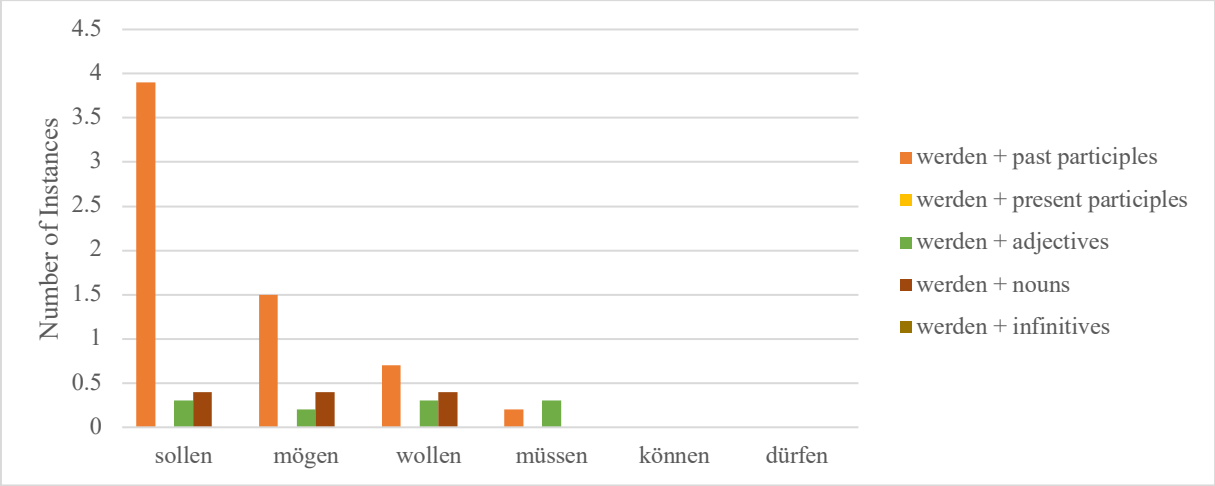


Figure 23: Normalized frequency per century of the instances of *werden* according to the tenses

The raw and normalized frequencies of the instances of *werden* in the four centuries analyzed in this chapter indicate that, on one hand, the 12th century is the time frame with the highest number of attestations of *werden* combined with modal verbs. On the other hand, the 13th century is the time frame in which *werden* was found with the highest number of different modal verbs (*sollen*, *mögen*, *wollen*, *müssen*, and *können*).

5.5.2 Discussion

The instances of *werden* from the Middle High German texts offer valuable indications about the development of *werden* and the elements with which it was combined, such as the present participles, the past participles, and the infinitives forms. Further, they also give a deeper insight into some of the most relevant syntactic changes in the history of the German language happening between the 11th and the first half of the 14th century. Overall, *werden* seems to be regularly used in the language, especially in the constructions with the past participles. As reported in the review of the literature, this periphrasis was already common in Old High German. The number of occurrences found in the Middle High German corpus indicates that the *werden* passive was frequently used in this period and homogeneously spread in all five dialect areas considered in this study. The majority of these attestations had *werden* in the *Präsens* and the *Präteritum* tenses, with only one attestation found in the *Plusquamperfekt* in the 13th century. In this instance, the past participle of *werden* did not add the *ge-* prefix. The absence of the *ge-* prefix was also observed in the attestations in the *Perfekt* and *Plusquamperfekt* tenses in which *werden* was used as a copula verb with adjectives and nouns. As discussed throughout this chapter, the missing prefix could be a result of analogic processes with verbs of a similar semantics, such as *bringen* (to bring) and *finden* (to find), which also did not add the prefix when used in their past participle forms. The meaning that *werden* still had in Middle High German ("to come into being", "to happen") was indeed comparable to those of *finden* and *bringen*, which denoted an action with a clear endpoint. This semantic affinity could have influenced the way *werden* was used in its past participle form and the consequent drop of the *ge-* prefix. Further, and as discussed in the review of the literature, the passive, the present perfect, and past perfect were still in the early stages of the grammaticalization process, and both *werden* and the past participles still had part of the original autonomy that they possessed in Old High German (Kotin, 2003). The low number of attestations in the present and the past perfect encountered in this corpus could also be explained by the status of these constructions in Middle High German.

The data of Middle High German also provided, on the one hand, insightful information on the use of the constructions with *werden* with present participles and, on the other hand, indications about the development of *werden* combined with infinitive verbs. As reported in the review of the literature and throughout this chapter, some scholars, such as Bechs (1901) and Kleiner (1925), have argued that the *werden* with the present participles was the source from which

the German periphrastic future emerged in the 13th century (Leuschner, Mortelmans, & Groodt, 2011, p. 237). However, the instances found in this corpus have suggested that this was not the case. Firstly, these periphrases are attested in the 12th century. Secondly, in the texts from 13th and 14th centuries, these constructions are used with the same or similar types of verbs and in analogous contexts. These findings seem to support the claims of scholars such as Kotin (2003), who claimed that the combinations of *werden* and the present participles and those with *werden* and the verbs in the infinitive forms were "twin constructions" that originated from different sources. These coexisted for a long time until significant changes in the language, such as the disappearance of the periphrases with *sîn* (to be) and the present participles, brought the constructions with *werden* and the present participle to extinction (Kotin, 2003, p. 166).

The attestations of *werden* with the modal verbs showed a high number of instances with the modal verbs *sollen* and *mögen*. These modal verbs were already frequently used in Old High German (Diewald, 2012), and the number of attestations found in Middle High German is probably related to the use of this verb in the previous period. The texts analyzed in this chapter also showed only one instance of *werden* combined with *können*, while they did not contain any occurrences of *werden* with the modal verb *dürfen*. The absence of attestations with these two modal verbs is related to the semantic and syntactic restrictions that limited their use in Middle High German (Diewald, 2012, p. 299).

5.5.3 Summary

The texts from the Middle High German corpus contained a total of 1,851 instances of *werden* with present and past participles, adjectives, nouns, and verbs in the infinitive forms and 186 instances of modal verbs used in combination with this verb. The majority of the attestations of *werden* were in the passive constructions, but *werden* was also used frequently as a copula verb in combination with nouns and adjectives. These attestations were found in the *Präsens*, the *Präteritum*, the *Perfekt*, and the *Plusquamperfekt* tenses. When used in the present- and past perfect tenses, the past participles of *werden* did not include the *ge-* prefix. The constructions with the present participles and the verbs in the infinitive forms were less frequent. Further, only three instances of *werden* used as a full verb were found in the corpus.

The highest number of instances with the modal verbs were in passive constructions, while a modest number of attestations was found with *werden* used as a copula with nouns and adjectives.

The corpus contained only one attestation of *können*, whith no instances of *werden* combined with *dürfen* were found.

Overall, the findings in this chapter provided support for the claims of scholars such as Kotin (2003), in regard to the rise of the periphrastic future. The attestations of *werden* and present participles and *werden* with the infinitives found in Middle High German, on the one hand, contradicted the claims of scholars such as Bechs (1901) and Kleiner (1925), who have argued that the *werden* with the present participles was the source from which the German periphrastic future emerged in the 13th century (Leuschner, Mortelmans, & Groodt, 2011, p. 237). On the other hand, they supported the claim about the existence of "twin constructions" that originated from different sources, meaning that the combinations of *werden* with the present participles are not the sources of the periphrastic future (Kotin, 2003).

CHAPTER 6. EARLY NEW HIGH GERMAN

This chapter focuses on how the attestations of *werden* were parsed, using texts from the *Bonner Frühneuhochdeutschkorpus*, listed in the first section of the methodology chapter. First, I present the results for each century. Second, I discuss the data as a whole for Early New High German. The attestations were found using the lemma search option in the online corpus. As seen in the review of the literature, *werden* was used with elements such as modal verbs, participles, adjectives, infinitives, and nouns. The attestations included here are centered on all the instances of *werden*, independently of the elements with which it is combined. For each attestation, I have indicated the tense (present, simple past, present, and past perfect). As done for Middle High German, in the analysis, I keep the instances of *werden* used with modal verbs separated from the others in order to have a clear distinction between the modal constructions and the use of *werden* as the main verb.

6.1 14th Century (Second Half)

6.1.1 Analysis

A total of 546 instances for the second half of the 14th century was analyzed. Of these, 528 were included in the database. Eighteen attestations had missing annotations (marked as *unbekannt* [unknown] in the corpus) and were therefore excluded. Only two forms of *werden* used as a full verb were found.

The following figure (Figure 24) shows the number of instances of *werden* in the present (*Präsens*), simple past (*Präteritum*), present perfect (*Perfekt*), and past perfect (*Plusquamperfekt*)

tenses. Figure (24) does not include the instances of *werden* with modal verbs which are presented in a separate chart (Figure 25).

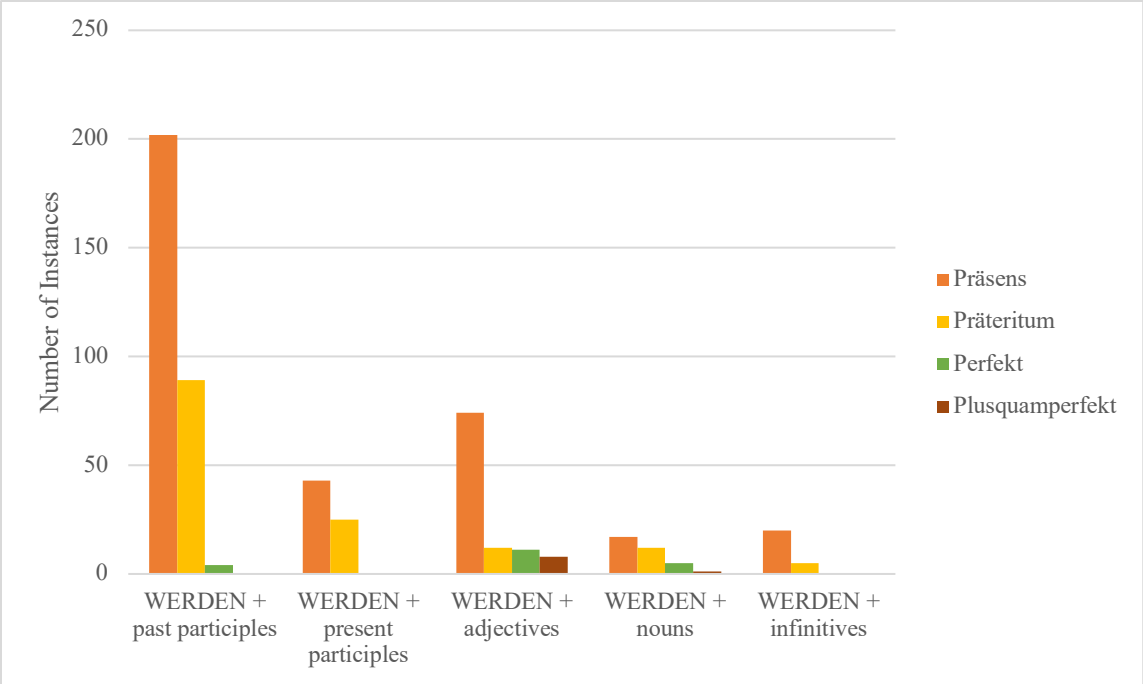


Figure 24: *werden* in the second half of the 14th century with past and present participles, adjectives, nouns, and infinitive verbs

The verb *werden* was found in all four tenses: present, simple past, present, and past perfect. The data from the second half of the 14h century does not differ much from the first half of the same century, except for a higher number of instances of *werden* with infinitive verbs (a total of 25), not only in the present tense, as shown in example (55), but also in the past tense (4 instances). Two of these instances had the auxiliary in the subjunctive, as shown in example (56), while three of them had *werden* in the indicative *Präteritum*, as shown in example (57):

(55) di **wirt** noch mer **hungern**
that become-AUX.PRS even more be hungry-INF
'They will become even more hungry'
(*Altdeutsche Predigten*, 6, 10)

(56) ein ygeliche mensche unser herren got vil
a [unknown] human our lord god a lot
innercliche **würde** **bitten**
interiorly would become-AUX. SUBJ.PAST beg-INF
'A man would interiorly beg our lord god a lot'
(*Altdeutsche Predigten*, 10, 33)

(57) ich do **wart** gelosen **sehen**
I there became-AUX.PAST calm see-INF
'I was seen calm there'
(*Mannen*, 19, 11)

Other than these examples, the majority of the attestations were found in combination with past participles in passive constructions. For the first time, such attestations were found not only in the *Präsens* and in *Präteritum* tenses (202 and 89 respectively), but also in the *Perfekt* tense (4 instances in total) with *werden* without the *ge-* prefix, as shown in examples (58) and (59):

(58) czu der zeit der czwelifpoten **ist** mess
to the time of the herald is-AUX.PRS mass
gesungen **worden**
sung-PTCP.PST become-AUX.PTCP.PST
'To the time of the herald mess was sang'
(*Rationale*, 9.16)

(59) Die **sint** **zubrochin** und verwazentlich
They are-AUX.PRS broken-PTCP.PST and doubtful
worden in iren gelusten
become-AUX.PTCP.PST in their desires
'They have become broken and doubtful in their desires'
(*Benediktinerregel Oxford*, 10.26)

The texts analyzed for this part of the 14th century also contained a high number of attestations of *werden* used as an auxiliary with the present participle (both in the present and past tenses), and as a copula in combination with adjectives and nouns. The combinations of *werden* with the present participles appear to have kept a meaning similar to the same type of constructions from the 13th and first half of the 14th centuries, as shown in examples (60) and (61):

(60) ir **wert** her nach **weinent** vurege trehen
You become-AUX.PRS after crying-PTCP.PRS fiery tiers
'You will cry after burning tears'
(*Altdeutsche Predigten*, 8, 14)

(61) ich minen lichamen also gar vbele **hassende** **wart**
I my body therefore totally badly hating-PTCP.PRS became-AUX.PRS
'I started hating my body therefore so badly'
(*Mannen*, 6, 15)

The next figure (Figure 25) shows the instances of *werden* in combination with the modal verbs.

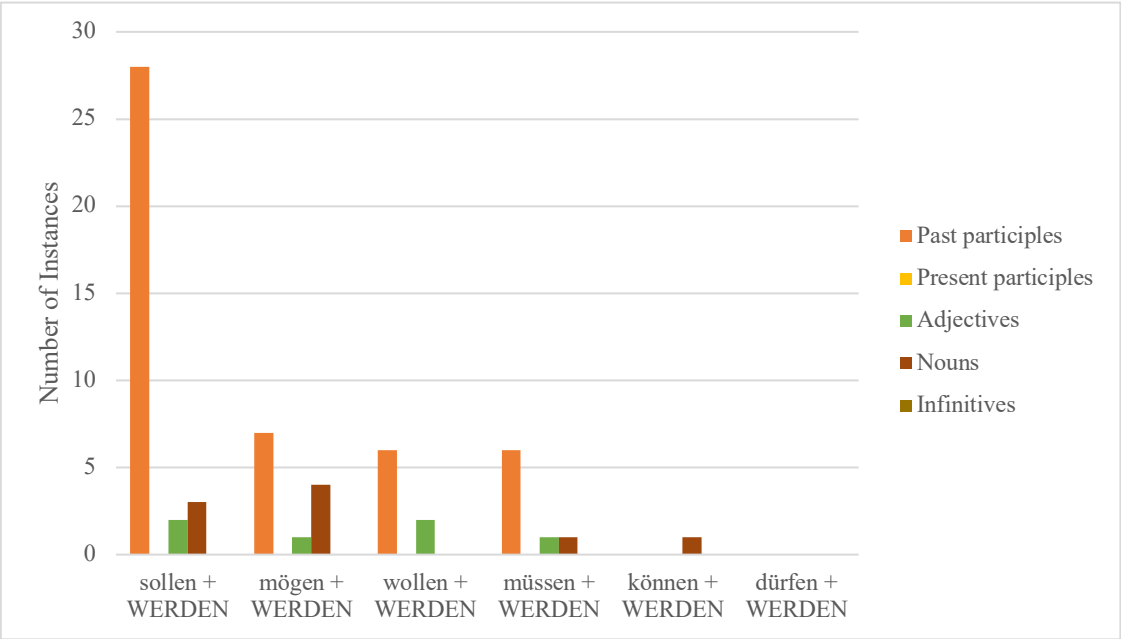


Figure 25: *werden* with modal verbs in the second half of the 14th century

A total of 62 instances were found with *werden* and the modal verbs *sollen* (should), *mögen* (to like), *müssen* (must), *wollen* (want), and *können*. The highest number of attestations was with *werden* used as an auxiliary for the passive constructions. The majority of these combinations were with *sollen* (28), while a comparable number of instances was found with *mögen*, *wollen*, and *müssen*. Instances of *werden* as a copula verb were found with all four modal verbs. Here, I offer instances with *sollen*, as shown in example (62), with *mögen* as shown in example (63), and with

wollen, as shown in example (64). The data also contained an instance of *können*, as showed in example (63). No forms of *dürfen* with *werden* were found.

(62) **solt** dv schiere **gewar** **werden**
should-AUX.PRS you fast true-ADJ become-AUX. INF
'You should become fast and true'
(*Mannen*, 35, 21)

(63) herzo nye **brief** noch **gesetze** weder **werden**
Heart no letter-NOUN no law-NOUN neither become-AUX.INF
en **kunde** noch en **mochte**
not could-AUX.PST nor not wanted-AUX.PST
'The heart could not and did not want to become neither a letter nor a low'
(*Buch Köln*, 423, 5)

(64) es **welle** **gut** **werden**
it wanted-AUX.SUBJ good-ADJ become
'It wanted to become good'
(*Mannen*, 31, 23)

6.1.2 Discussion

The data from the second half of the 14th century does not differ much from the data of the first half of the same century. *Werden* was used mostly as an auxiliary for the passive constructions and in combinations with the present participle. It also appeared quite often as a copula verb with adjectives and nouns. The data also show an increased number of instances of *werden* and verbs in the infinitive form. Further, among these attestations, there were cases in which the auxiliary was also found in the *Präteritum* tense. Consider again example (57), reported here at 65:

(65) ich do **wart** gelosen **sehen**
I there became-AUX.PAST calm see-INF
'I was seen calm there'
(*Mannen*, 19, 11)

Example (65) is not an isolated occurrence of *werden* in the simple past combined with a verb in the infinitive form. Other instances were found, as shown in example (66):

(66) **wart** aber in dem blicke gelosen **sehen**
became but in the look calm see-INF
vnd **bevinden** solliche wunder
und find-INF such wonders
'[I] was seeing with a calm gaze and I was finding such wonders'
(*Mannen*, 19, 8)

Other than in the two instances reported above, the use of *sehen* in the infinitive after the adjective *gelosen* also occurring in combination with the auxiliary *haben* in *"het gelosen sehen*" ('has calm seen' – *Mannen*, 13, 30) in what seems to be a verb in the *Perfekt* tense. This last example could raise some doubts about the true form of *sehen* in the examples (65) and (66). It has been reported that in Old High German and in Middle High German, some verbs such as *bringen* (to bring) and *finden* (to find), because of their intrinsic perfective semantics, did not add the prefix *ge-* when used in their past participle forms (Kotin, p. 239). Nonetheless, *sehen* does not fall into the category of perfective verbs, since it denotes an ongoing process. Further, the presence of the second verb in the infinitive *bevinden* (to find), which is also connected to *werden* in the simple past, suggests that *sehen* is indeed in the infinitive form. This could indicate that these instances are among the attestations of *werden* in the *Präteritum* tense with a verb in the infinitive form. According to Smirnova (2006), in Old High German these constructions had a prevalent inchoative meaning, while in Middle and Early New High German they also added a mutative component in their semantics, denoting a cause-result relationship (p. 248). This means that the verbs combined with *werden* in the simple past denoted actions that resulted from previous events mentioned earlier on in the text.

The corpus of the second half of the 14th century also contained the first attestations of *werden* in the subjunctive past combined with verbs in the infinitive forms. Although these periphrases occur for the first time in the corpus in the second half of the 14th century, they had emerged already in Middle High German (Smirnova, 2006, p. 269). These constructions were used to express a desire or to indicate a hypothetical situation in subordinating clauses with independent clauses in the past tense (Smirnova, 2006, p 271).

For the first time, the data also show some instances of the passive constructions in the *Perfekt* tense. In previous centuries, the only attestations found were with *werden* as a copula combined with nouns and adjectives. The presence of this occurrence could indicate two different circumstances. On one side, the "paradigmatization" (Kotin, 2003, p. 204) of *werden* plus past

participles could have advanced, allowing this periphrasis to also appear in the present tense. It could be assumed that processes of analogical associations with the instances of *werden* as a copula were involved in this syntactic innovation. One aspect that would support this hypothesis is the fact that all the instances did not have the *ge-* prefix, as observed in the attestations with *werden* as a copula verb. On the other hand, the passive in the *Perfekt* tense could also hint to the ongoing grammaticalization of the *Perfekt* itself. In Concu (2016), I reported a drastic frequency increase in the instances of this tense around the end of the 14th and the beginning of 15th century. This boost in frequency could have also influenced the use of the passive in this tense, suggesting an extension of the types of verbs that could appear in the *Perfekt* tense. A third scenario could indicate a combination of both the further "paradigmatization" of the passive constructions and the ongoing grammaticalization of the present perfect in the early centuries of the Early New High German period.

Regarding *werden* with the modal verbs, the only significant difference with the first half of the same century (and also of the 11th, 12th, and 13th centuries) is *werden* combined with the modal verb *können*. *Werden* was used here as a copula in combination with a noun. This attestation could imply a possible weakening of those semantic and syntactic restrictions that had significantly limited the use of this modal verb in the previous centuries (Diewald, 2012, p. 299). Similar restrictions, however, seem to be present for the modal verb *dürfen*, as no instances of this verb were found in the texts of this century.

6.1.3 Summary

The texts from the second half of the 14th century contained a total of 546 instances of *werden*. The highest number was found in the present tense, but instances were also found in the simple past, in the present- and past perfect. The main use of the verb *werden* was as an auxiliary in the passive constructions, which appeared for the first time also in the *Perfekt* tense. *Werden* was also found in combination with present participles and with verbs in the infinitive forms. Among the attestations of *werden* and infinitive verbs, there were some instances of *werden* in the simple past. In this corpus, it was also used as a copula with nouns and adjectives.

The texts from the second half of the 14th century contained a total of 62 instances of *werden* combined with the modal verbs. The majority of these instances were with *sollen*, both in combination with the passive periphrasis and with *werden* used as a copula. For the first time, the

data contained an instance of *können* with *werden*, while no forms of the modal verb *dürfen* were found.

6.2. 15th Century

6.2.1 Analysis

A total of 479 instances for the 15th century was analyzed. Of these, 474 were included in the database. A number of three attestations had missing annotations (marked as *unbekannt* [unknown] in the corpus) and were therefore excluded. Two forms of *werden* used as a full verb were found.

The following figure (Figure 26) shows the number of instances of *werden* in the present (*Präsens*), simple past (*Präteritum*), present perfect (*Perfekt*), and past perfect (*Plusquamperfekt*) tenses. Figure 26 does not include the instances of *werden* with modal verbs which are presented in a separate chart (Figure 27).

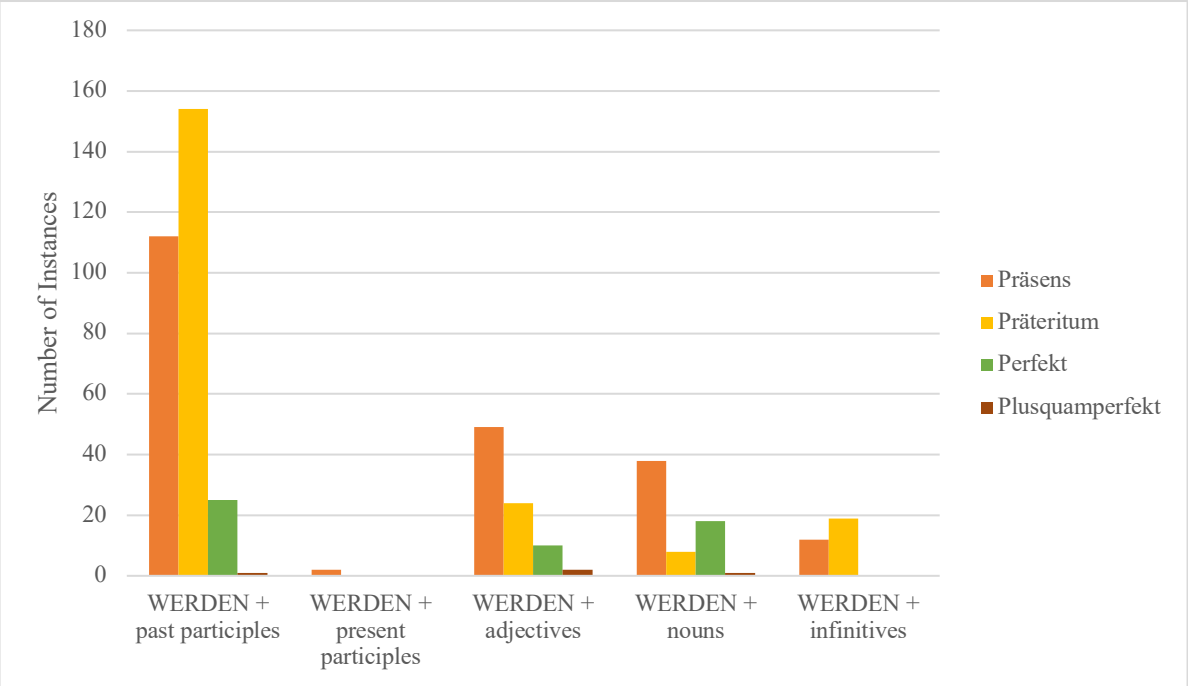


Figure 26: *werden* in the 15th century with past and present participles, adjectives, nouns, and infinitive verbs

Werden was found in the present, simple past, present- and past perfect tenses. Further, the data from the 15th century shows some of the trends observed in the previous centuries. The majority of the instances were found with *werden* used in the passive constructions (a total of 292) and as a copula in combination with adjectives (a total of 85) and nouns (a total of 65). Surprisingly, only two attestations of *werden* plus present participles were found, as shown in example (67), whereas a modest number of attestations with *werden* combined with infinitive verbs were found, as shown in examples (68) and (69):

(67) die selbigen werck **werden** vil **lustiger**
the same deeds become-AUX.PRS more happily
smeckende vnd messende
tasting-PTCS.PRS and measuring-PTCS.PRS
'The same deeds become way tastier and more appropriate'
(*Pillenreuth Mystik*, 191, 3)

(68) ir **wert** morgen nicht gen Prespurgk **faren**
you become-AUX.PRS tomorrow not to Prespurgk go-INF
'You all won't go tomorrow to Prespurgk'
(*Denkwürdigkeiten*, 19, 23)

(69) so **wirt** er dich tailhafftig **machen** seiner pitter marter
so become-AUX.PRS he you partially do-INF his bitter martyrdom
'he will partially make you his bitter martyrdom'
(*Pillenreuth Mystik*, 167, 18)

Although the majority of the instances with *werden* in the past combined with the infinitive verbs were in the subjective mood, two forms were still in the indicative, as shown in (70):

(70) das Liecht was vmbgefalen vnd **ward** **prýnnen**
the light was fallen and became-AUX.PST burn-INF
'the light had fallen and started to burn'
(*Denkwürdigkeiten*, 11, 9)

Among the attestations of *werden* used in the passive constructions, there was a higher number of instances in the *Perfekt* (a total of 25) in comparison to the second half of the 14th century. In all the attestations, *werden* did not add the *ge-* prefix, as shown in (71) and (72):

(71) zw einmal **pin** ich **gestaint** **worden**
to one time **am**-AUX.PRS I stoned-PTCP.PST become-AUX.PTCP.PST
'once I have been stoned'
(*Pillenreuth Mystik*, 188, 8)

(72) Als in dem begyne des eyrsten alders **is**
when in the beginning of the first time is-AUX.PRS
die werlt **gemacht** **worden**
the world made-PTCP.PST become-AUX.PTCP.PST
'When at the beginning of time the world was made'
(*Johannes Rothe:Chronik*, 15, 17)

In the data, there were also two instances of *werden* used as a full verb, in which this verb expressed the meaning of "coming into being" and "happening", as shown in example (73):

(73) Nu **werde** eyne vestickeit mitten yn den wassern
Now would become-SBJV.PST a firmness middle in the waters
'It would happen a firmness in the waters now'
(*Johannes Rothe:Chronik*, 11, 26)

The next figure (Figure 27) shows the instances of *werden* in combination with modal verbs:

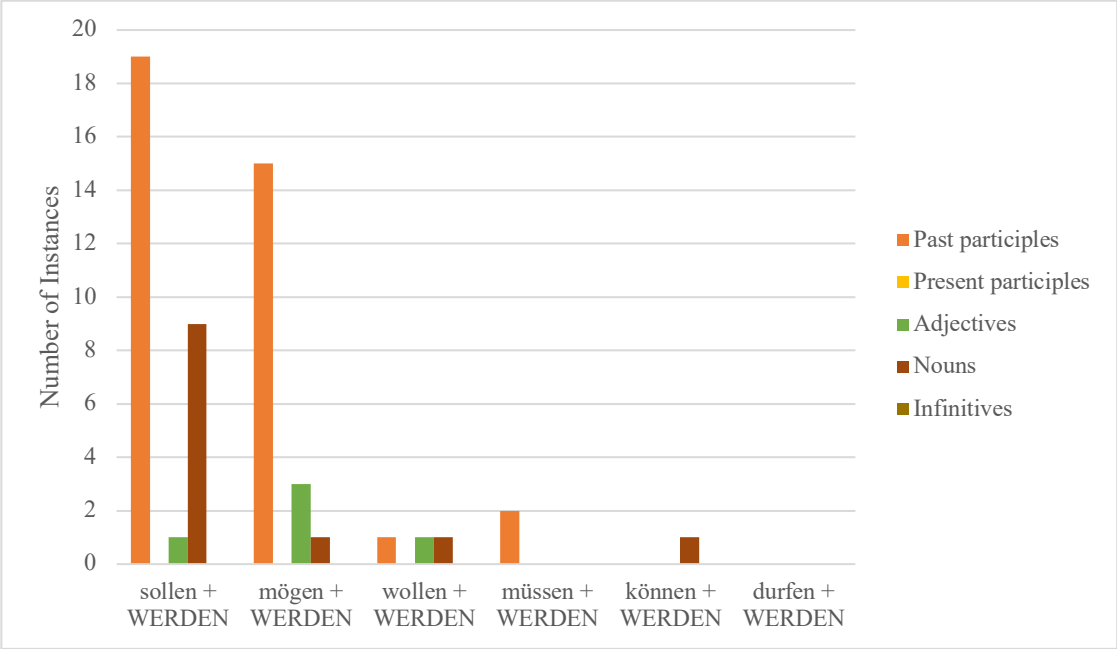


Figure 27: *werden* with modal verbs in the 15th century

A total of 39 instances were found with *werden* combined with the modal verbs *sollen* (should), *mögen* (to like), *müssen* (must), and *wollen* (want). The most used modal verbs in the 15th century are *sollen* and *mögen*, which were found combined with passive constructions, as shown in examples (74) and (75) and with *werden* used as a copula, as shown in examples (76) and (77):

(74) ich **solt** darumb **verdampt** **werden**
I shall-AUX.PRS therefore doomed-PTCP.PST become-AUX.INF
'I shall therefore be doomed'
(*Denkwürdigkeiten*, 16, 10)

(75) dar durch eyn mynsch vnder wijst ind
that through a human being wise become-PTCP.PST and
getroest **mach** **werden**
comforted-PTCP.PST like-AUX.PRS become-AUX.INF
'through that a person will be wise and comforted'
(*Johann Koelhoff: Chronik*, 3r, 12)

(76) Ich **solt** auch iren gnaden **gevëtrInn** **werden**
I should-AUX.PRS also her mercy godmother-NOUN become-AUX.INF
'I should become in her mercy her godmother'
(*Denkwürdigkeiten*, 20, 36)

(77) an den gelauben nymant **hailwertig** **werden** **mag**
by the faith no one waiting-ADJ become-AUX.INF like-AUX.PRS
'No one likes to wait on faith'
(*Pillenreuth Mystik*, 156, 13)

There was only one form of *werden* as a copula combined with *können*, as shown in example (78)[12] whereas no forms of *werden* plus *dürfen* were found.

(78) **kan** ir doch nicht **werden** in dieser zceit
Can-AUX.PRS to her not become-COP.INF in this time
'It can become to her in this time'
(*Pillenreuth Mystik*, 224, 8)

[12] Example (72) has been classified as *werden* plus noun because of the sentence in which it was found. This instance of *werden* referred to the noun *Vernunft* (reason) to be found in line 6.

6.2.2 Discussion

The data from the 15th century is quite similar to the data found in the second half of the 14th century. *Werden* is mainly used as an auxiliary verb for passive constructions, but for the first time in this study, it appears not only in the *Präsens* and the *Präteritum* tenses, but also in the present and past perfect. The increased number of instances in these two tenses can be related to the higher frequency with which *Perfekt* and *Plusquamperfekt* started to be used in the 15th century (Concu, 2016). Further, in all these attestations, *werden* did not add any *ge-* prefix. The absence of this prefix suggests that part of the original mutative meaning that *werden* had in Old and in Middle High German still persists in this century. The presence of instances in which *werden* is in the indicative *Präteritum* with verbs in the infinitive form seems to further support this claim. *Werden* was, indeed, still relatively autonomous at this stage, and, in the simple past, could be combined with verbs in the infinitive forms. In these cases, this periphrasis had the capacity to convey the entering into a new state that was the result of previous actions (Smirnova, 2006, p. 246). Consider again example (79):

(79) das Liecht was vmbgefalen vnd **ward** **prýnnen**
the light was fallen and became-AUX.PST burn-INF
'the light had fallen and started to burn'
(*Denkwürdigkeiten*, 11, 9)

In this example, the entering of the subject into the new state of "burning" seems to be the direct consequence of the previous action of "falling." The use of the infinitive combined with *werden* in the simple past collocates this entering into the new state in the past because the previous action also takes place in the past.

Together with the combinations of *werden* in the indicative *Präteritum* plus verbs in the infinitive form, the texts from this century contain also a modest number of instances with *werden* in the subjective *Präteritum*. In the first centuries of the Early New High German period, these constructions had to compete with the constructions with the modal verb *mögen* in the indicative and the modal verbs *sollen* and *wollen* in the simple past to express the potential taking place of a specific action (Smirnova, 2006, p. 272). The construction with *werden* in the subjunctive emerges later in Modern German as the preferred way to convey such meaning (p. 272).

Another notable change in comparison to the previous century is the near absence of the periphrases of *werden* and the present participles. There is an increase in the number of instances with *werden* in the indicative present plus infinitive verbs.

The instances with the modal verbs did not differ much from those analyzed in the second half of the 14th century. The highest number of combinations was with the *werden* passive, followed by a modest number of instances of werden used as a copula verb. Among these instances, there was also only one attestation of *können*. No attestations of *dürfen* were found. This could suggest that the semantic and syntactic restrictions that affected *dürfen* until now were still in place in the 15th century.

6.2.3 Summary

The texts from the 15th century contained a total of 474 instances of *werden*. The highest number was found in the present tense, but instances were found also in the simple past, present- and past perfect. For the first time, the passive periphrases also appeared in the *Perfekt* tense. This corpus contained a high number of attestations of *werden* and infinitive verbs. Among these instances, there are also cases in which *werden* was used in the simple past. Further, there was a modest number of attestations of *werden* used as a copula with adjectives and nouns.

The texts from the 15th century contained a total of 39 instances of *werden* with modal verbs. The majority of those were with *sollen* and *mögen*, both in combination with the passive periphrases and with *werden* as a copula. For the second time in this study, there was an instance of *können*, while there were still no forms of the modal verb *dürfen*.

6.3 16th Century

6.3.1 Analysis

A total of 569 instances for the 16th century was analyzed. Of these, 557 were included in the database. Nine attestations had missing annotations (marked as *unbekannt* [unknown] in the corpus) and were therefore excluded. Only two forms of *werden* used as a full verb were found.

The following figure (Figure 28) shows the number of instances of *werden* in the present (*Präsens*), simple past (*Präteritum*), present perfect (*Perfekt*), and past perfect (*Plusquamperfekt*)

tenses. Figure 28 does not include the instances of *werden* with modal verbs which are presented instead in a separate chart (Figure 29).

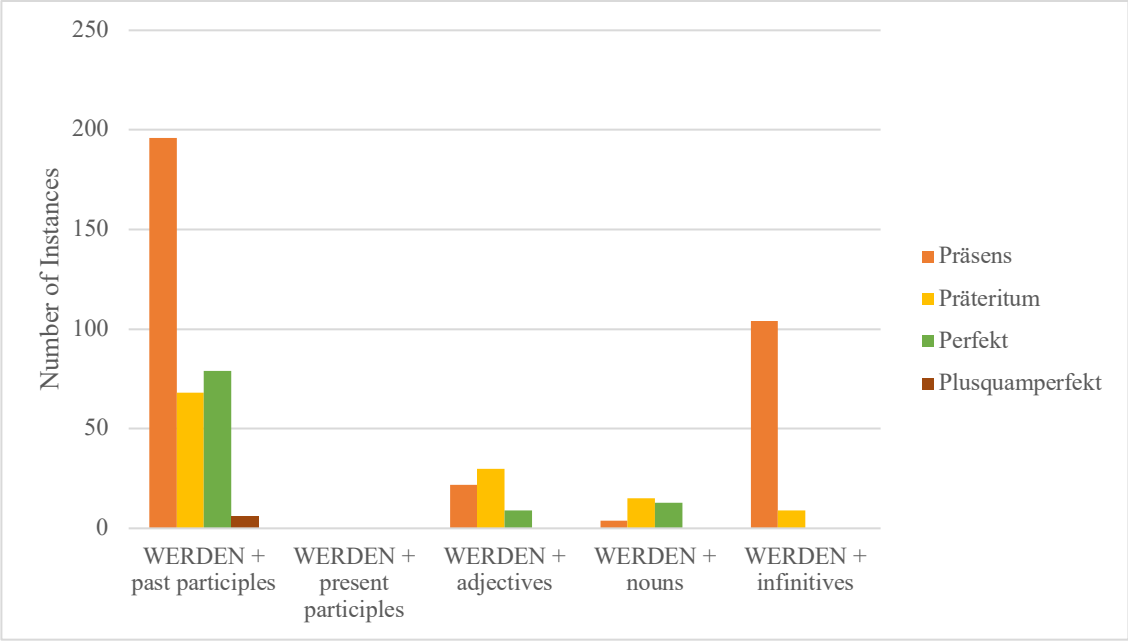


Figure 28: *werden* in the 16th century with past and present participles, adjectives, nouns, and infinitive verbs

The data from the 16th century show a different scenario from what has been observed in previous centuries. *Werden* is still used in the majority of the cases for the passive constructions (a total of 349), but the number of instances of the passive in the *Perfekt* tense has significantly increased in comparison to the number that was found in the 15th century. The majority of the forms were in the indicative, while four of these were in the subjunctive. However, and independently of the mood in which these forms were found, in all the instances in the *Perfekt* tense, *werden* didn't add the *ge-* prefix in the past participle form, as shown in examples (80) and (81). Further, among those instances, seventeen attestations found in subordinating clauses can be classified as afinite constructions, as shown in examples (82) and (83):

(80) Es **ist** mir auch von etlichen Hispaniern fur eine gewisse
It is-AUX.PRS to me also from several Hispanics for a certain
Warheit **gesagt** **worden**
Truth said-PTCP.PST become-AUX.PTCP.PST
'It has been said to me a certain truth by several Hispanics'
(*Amerika*, 6, 10)

(81) am letste tag Meyens sind **verbrennt** **worden**
on the last day of May are-AUX.PRS burnt-PTCP.PST become
'on the last day of May (they) have been burnt'
(*Gespenster*, 24v,17)

(82) so dann des Gedenim geschlachts **gedacht** **worden**
so then of Gedenim lineage thought-PTCP.PST become-AUX.PTCP.PST
'Then the lineage of Gedenim (has) been considered'
(*Mosconia*, E1r, 3)

(83) dergleichen zuvorn nicht **geseheen** **worden**
similar ones before not seen-PTCP.PST become-AUX.PTCP.PST
'Similar ones (have) never been seen before'
(*Amerika*, 6, 43)

While the corpus did not contain any combinations with the present participles, a remarkable number of *werden* plus verbs in the infinitive (113 in total) was found. While all the forms in the *Präteritum* (9) were in the subjunctive mood, among those in the *Präsens*, seventy-nine of were in the indicative mood, as shown in examples (84) and (85), and twenty-five in the subjunctive mood, as shown (86) respectively:

(84) der heilig Geist **wird** **kommen**
the holy spirit becomes-AUX.PRS come-INF
'The holy spirit will come'
(*Summaria*, 18v, 19)

(85) An jenem tag **wird** es **offenbar** **werden**
on a day becomes-AUX.PRS it obvious-ADJ become-COP.INF
'One day it will become obvious'
(*Summaria*, 29, 34)

(86) so **werde** sie **gesund** **werde**
so would-AUX.SBJV.PRS she healthy-ADJ become-COP.INF
'So she would be healthy'
(*Summaria*, 18v, 28)

The number of instances of *werden* used as a copula with adjectives and nouns does not differ much from the numbers found in the previous century and, in the attestations in which *werden* is used as full verb (one in the indicative mood and one in the subjunctive mood), it still

retains its original meaning of "coming into being" and "to happen", as shown in examples (87) and (88):

(87) Anno. 81 **Wurde** Titus.
Year 81 became-PST Titus
'In the 81st year Titus was born'
(*Johann Bange:Chronik*,14r, 14)

(88) die wunden **wurde** jm am karfrytag
The wounds became-PST to him on Good Friday
'He became wounded on Good Friday'
(*Gespenster*, 26v, 28)

The following figure (Figure 29) shows the number of instances that were found in the constructions with the modal verbs.

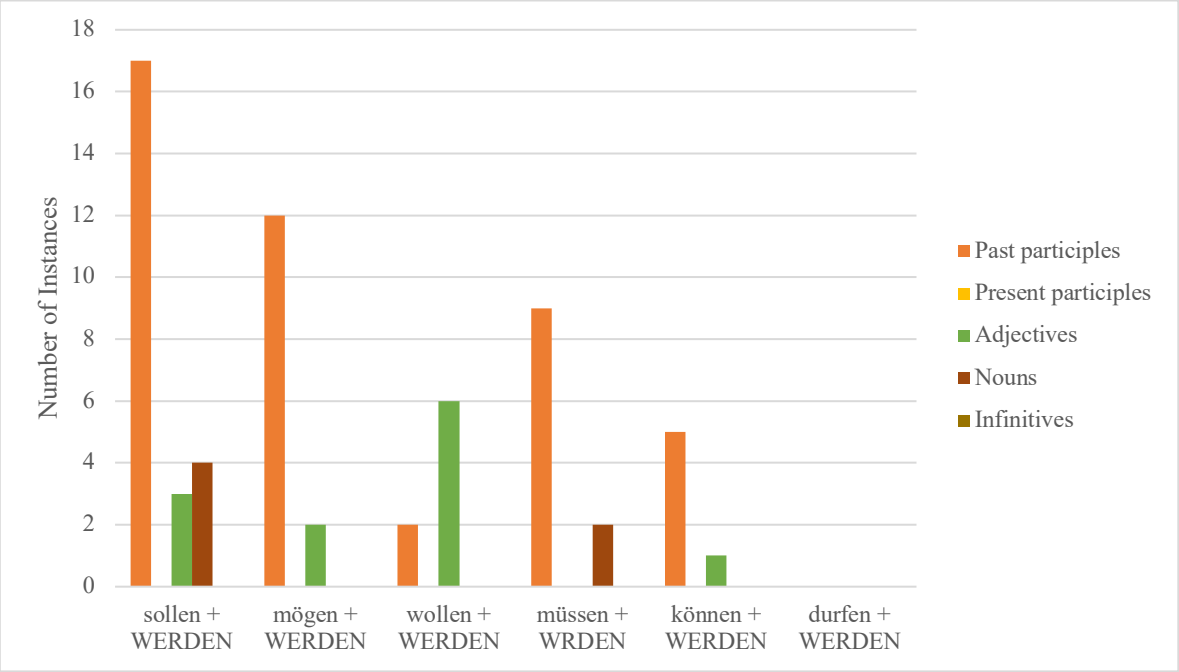


Figure 29: *werden* with modal verbs in the 16th century

A total of 63 instances of *werden* used with the modal verbs *sollen* (should), *mögen* (to like), *müssen* (must), and *wollen* (want) were found in the texts from the 16th century. As observed

in the 15th century, *sollen* und *mögen* are again the most used modal verbs, with the passive constructions, as shown in examples (89) and (90):

(89) wo jemandt anders darzu **moechte** **erwehlet**
Where someone else to that would like chosen-PTCP.PST
vnd **gezogen** **werden**
and pulled-PTCP.PST become-AUX.INF
'where someone would like to be chosen and pulled'
(*Amerika*, 5, 26)

(90) die arm seel **mochte** **erloßt** **werden**
the poor soul wanted redeemed become
'The poor souls wanted to be redeemed'
(*Gespenster*, 25v, 14)

The corpus of the 16th century does not contain any form of *dürfen,* but it has some instances of *können* combined with *werden* used for the passive constructions and with *werden* as a copula. Further, it was found in both the indicative, as shown in example (91), and in the subjunctive moods (92):

(91) Guiana durch diesen Weg nicht **entdeckt**
Guiana through this Way not discovered-PTCP.PST
kan **werden**
can-AUX.PRS become-AUX.INF
'Guiana cannot be discovered through this way'
(*Amerika*, 11, 21)

(92) alle Schiff der Welt damit **kondten**
all ships of the world with that could-AUX.PTS.SBJV
beladen **werden**
loaded-PTCP.PST become-AUX.INF
'All the ships in the world could be loaded with that'
(*Amerika*, 2, 6)

6.3.2 Discussion

The data from the 16th century reflect many of the changes happening to the German language in this period. First and foremost, the passive now appears in all four tenses, and the

number of forms found in the *Perfekt* is even higher than those found in the *Präteritum*. The boost in frequency that this tense had experienced in the previous century continues through this century as well (Concu, 2016). Along with the increased frequency of forms in the present perfect, the corpus analyzed for this century also shows the first examples of another interesting syntactic phenomenon: seventeen of the instances in the *Perfekt* can be classified as afinite constructions. What distinguishes these instances from others is the "ellipsis of the finite auxiliary (hence 'afinite') from a periphrastic verbal construction (like perfect or passive) in embedded clauses" (Breithbarth, 2005, p. 1). Other elements that can be dropped are the *sein* auxiliary in the *sein* passive and, as observed in the data, the *sein* auxiliary in the *werden* passive. Consider the following examples:

(93) so dann des Gedenim geschlachts **gedacht** **worden**
so then of Gedenim lineage thought-PTCP.PST become-AUX.PTCP.PST
'Then the lineage of Gedenim (has) been considered'
(*Mosconia*, E1r, 3)

(94) dergleichen zuvorn nicht **geseheen** **worden**
similar ones before not seen-PTCP.PST become-AUX.PTCP.PST
'Similar ones (have) never been seen before'
(*Amerika*, 6, 43)

Both these examples show the ellipsis of the auxiliary *sein* for the *werden* passive. Breitbarth (2005) offers two different accounts to explain both the rise of these constructions and their disappearance. Firstly, the emergence of the afinite constructions has to be considered as the result of the "fixation of the sentence-final verb placement as an unambiguous marker of embedding," (p. 16) as the textual complexity advanced in this period. This fixation created the right environment for the ellipsis of the auxiliaries of the periphrastic constructions in subordinating clauses (p. 16). Secondly, their disappearance has to be related to "stylistic changes resulting in a reduction to text complexity and the written language use turned back to a more 'natural' language" (p. 16). The afinite constructions found in this corpus must also reflect the increased textual complexity observed throughout this century.

The second main change found in the data is a drastic increase in the periphrases of *werden* and the verbs in the infinitive forms. Consider examples (95) and (96):

(95) der heilig Geist **wird** **kommen**
the holy spirit becomes-AUX.PRS come-INF
'The holy spirit will come'
(*Summaria*, 18v, 19)

(96) An jenem tag **wird** es **offenbar** **werden**
on a day becomes-AUX.PRS it obvious-ADJ become-COP.INF
'One day it will become obvious'
(*Summaria*, 29, 34)

Both these examples give some relevant information about the status of this periphrasis. Example (95) shows that this construction, in Early New High German, is already very similar to the *werden* future in Modern German. Example (96), with the use of two forms of *werden* (one as an auxiliary and one as a copula verb), shows that this verb is probably losing part of its original meaning and autonomy. According to Bybee (2003):

With repetition, sequences of units that were previously independent come to be processed as a single unit or chunk. This repackaging has two consequences: the identity of the component units is gradually lost, and the whole chunk begins to reduce in form. (p. 7)

The presence of the combinations of *werden* as a copula verb with nouns and pronouns, on one hand, and the increased frequency of the constructions of *werden* plus the infinitives that express future meaning, on the other hand, are presumably influencing the way these verbs are perceived in the language, (i.e., single units rather than a combination of separated elements). The direct consequence of this process is the gradual loss of the original autonomy and meaning that *werden* had in Old and Middle High German.

The instances with the modal verbs also offer some insights into the syntactic changes affecting the language in this period. For the first time in the corpus, there are instances with the *werden* passive combined with the modal verb *können*. One can assume that the syntactic restrictions that have limited the use of this verb in the language and in passive constructions (Diewald, 2012) are slowly disappearing during this century. Such a change did not affect the modal verb *dürfen*, as attested in the corpus in this study. No forms of this modal verb with *werden* were found.

6.3.3 Summary

The texts from the 16th century contained a total of 557 instances of *werden*. The highest number was found in the present tense, but instances were also found in the simple past, present- and past perfect. The attestations in the *Perfekt* were higher than those in the *Präteritum* this time. The number of attestations of *werden* plus infinitive was also very high, while no combinations of *werden* and present participles were found. Further, this corpus showed some attestations of *werden* used as a copula with adjectives and nouns.

The texts from the 16th century contain a total of 63 instances of *werden* with modal verbs. The majority of those were with *sollen* and *mögen*, both in combination with the passive and with *werden* as a copula. There were also five instances of *können* combined with the passive, while there were still no forms of the modal verb *dürfen*.

6.4 17th Century

6.4.1 Analysis

A total of 579 instances for the 17th century was analyzed. Of these, 568 were included in the database. Eleven attestations had missing annotations (marked as *unbekannt* [unknown] in the corpus) and were therefore excluded. No forms of *werden* as a full verb were found.

The following figure (Figure 30) shows the number of instances of *werden* in the present (*Präsens*), simple past (*Präteritum*), present perfect (*Perfekt*), and past perfect (*Plusquamperfekt*) tenses. Figure (30) does not include the instances of *werden* with modal verbs which are presented in a separate chart (Figure 31).

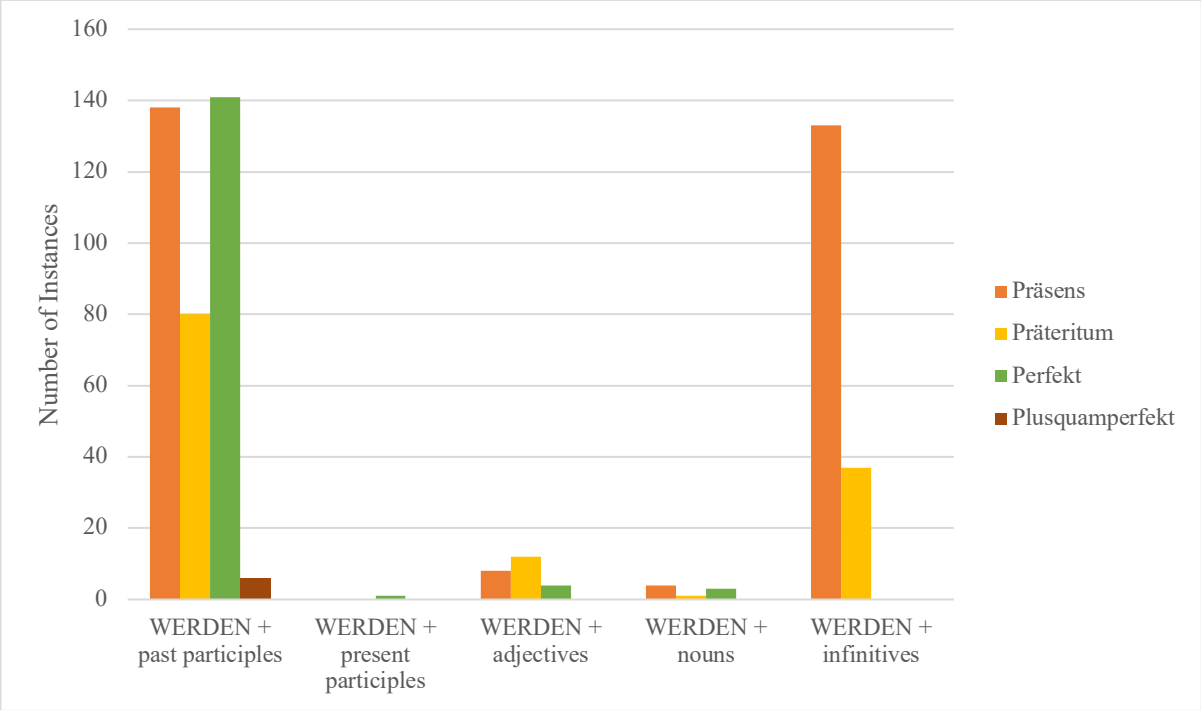


Figure 30: *werden* in the 17th century with past and present participles, adjectives, nouns, and infinitive verbs

The corpus from the 17th century confirms some of the trends observed in the 16th century. The main use of *werden* was in the passive constructions (365 in total), but for the first time, the tense used in the majority of the attestations was in the *Perfekt* (with 3 instances in the subjunctive form). In all the attestations, *werden* did not add the *ge-* prefix, as shown in examples (97) and (98), and there were 68 instances of afinite constructions, such as the one shown in example (99):

(97) nach drey Tagen **ist** der gebenedeyte JEsus
after three days is-AUX.PRS the blessed Jesus
gefunden **worden** im Tempel
found-PTCP.PST become-AUX.PTCP in the temple
'After three days the blessed Jesus was found again in the temple'
(*Deo Gratias*, 20, 21)

(98) Warum **ist** mein Rath allezeit
Why is-AUX.PRS my advice always
verachtet **worden**?
disdained-PTCP.PTS become-AUX. PTCP.PTS
'Why has been my advice always disdained?'
(*Jugendlust*, 79, 24)

(99) sein Corper **verbrandt** **worden**
his body burnt-PTCP.PST become-AUX.PTCP.PST
'His body has been burned'
(*Schaubühne*, 3940 Sp. B, 16)

The afinite constructions in the *Perfekt* were also found in the attestations of *werden* used as a copula with adjectives and nouns, as shown in examples (100) and (101):

(100) … daß es niemand **gewahr worden**?
…that is no one true-ADJ become-AUX.PTCP.PST
'… that nobody knew about it?'
(*Schaubühne*, 3738 Sp. D, 34)

(101) Daß Esther aus einem gemeinen Magdlein eine
That Estehr from a mean girl a
Kȯnigin **worden**
queen-NOUN become-AUX.PTCP.PST
'That Ester went from being a mean girl to a queen'
(*Deo Gratias* 36,16)

The data shows one instance of *werden* combined with a present participle in the *Perfekt* tense, also in an afinite construction, as shown in example (102):

(102) zwey oder drey Monat alte Kinder auf den Armen
two or three months old children in the arms
ihrer Mutter **redend** **worden**
of their mothers talking-PTCP.PRS become-AUX.PTCP.PST
'Two or three years old children started talking in the arms of their mother'
(*Deo Gratias*, 25, 5)

The data contained also the highest number found so far of the combinations of *werden* plus a verb in the infinitive form (170 in total). While the majority of the instances with *werden* in *Präsens* were in the indicative mood, as shown in example (103), only three instances were in the subjunctive mood, such as the one shown in example (104). Further, all the attestations in *Präteritum* tense had *werden* in the subjective, as shown in example (105).

(103) so **werde** ich endlich **sprechen**
so become-AUX.SUBJV.PRS I finally speak-INF
'So I will finally speak'
(*Jugendlust*, 89, 9)

(104) Jch halte nicht darvor daß Huet **sagen werde**
I think not of it that Huet say-INF become-AUX.SBJV.PRS
'I don't think much about what Heut would say'
(*Mythoscopia*, 25, 15)

(105) es **wurde** mir nichts im Wege **stehen**
it would-AUX.SBJV.PST to me nothing on the Way stand-INF
'Nothing would stand on my way'
(*Jugendlust*, 125, 5)

The following figure (Figure 31) shows the number of instances that were found in the constructions with the modal verbs.

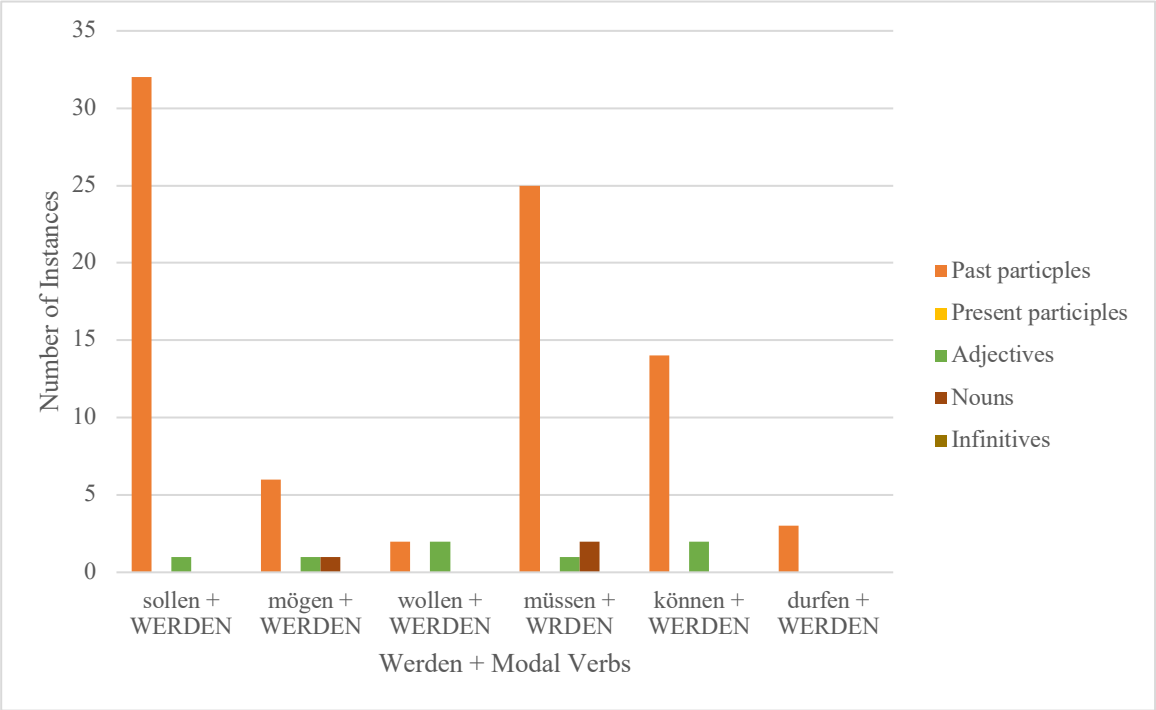


Figure 31: *werden* with modal verbs in the 17th century

A total of 92 instances of *werden* used with the modal verbs *sollen* (should), *mögen* (to like), *müssen* (must), and *wollen* (want) were found in the texts from the 17th century. This corpus shows a slightly different distribution of the combinations of the modal verbs in comparison to the centuries previously analyzed. While *sollen* is still the most used modal verb for the passive (32 instances in total), as shown in the example (106), *müssen* is also frequently used for this periphrasis (25 instances in total), as shown in example (107):

(106) mit Waffen nicht **sollen** **vertheidigt** **werden**
with weapons not shall-AUX.PRS defended-PTCP.PST become-AUX.INF
'[they] shall not be defended with weapons'
(*Spiegel,* 72 Sp. A, 4)

(107) der Hafen zur See **gesperret** **werden** **muste**
the port to the sea blocked-PTCP.PST become-AUX.INF must-AUX.SBJV.PST
'the port to the see should be blocked'
(*Schaubühne*, 6566 Sp. A, 39)

Among the attestations of *müssen* used with *werden* as a copula verb, there was also an instance with *werden* in the past participle instead of in the infinitive combined with the auxiliary *sein*, as shown in example (108):

(108) und davon **schwanger** **worden**
and from that pregnant-ADJ become-AUX.PTCP.PST
seyn müste
be-AUX.INF should-AUX.SBJV.PST
'And she should become pregnant from that'
(*Schaubühne*, 5152 Sp. B, 41)

Further, the data contained the highest number of combinations with *können* found so far, which was combined with both the passive (14 instances), as shown in example (109), and with *werden* as a copula combined with adjectives (2 instances), as shown in example (110):

(109) solches zum Lose-Gelde der theuren Sicherheit
such thing as lose money of the expensive safety
kan **angewendet** **werden**
can-AUX.PRS utilized-PTCP.PST become.AUX.INF
'Such (things) can be used to pay for the expensive safety'
(*Jugendlust*, 109, 15)

(110) kan es auß etlich **wenig** Tropfen **inne werden**.
can-AUX.PRS it from many few-ADJ drops become-AUX.INF
'From many drops it can turn into a few'
(*Mythoscopia*,13,10)

Finally, and for the first time in the data, there are the first attestations of the *werden* passive combined with the modal verb *dürfen*, as shown in example (111):

(111) Die Vorsorge vor krancke Personen **darff** nicht
The provisions in front of sick people must-AUX.PRS not
also **ausgeleget** **werden**
then put out-PTCP.PRS become-AUX.INF
'The provisions must not be put out in front of sick people'
(*Jugendlust*, 153, 4)

6.4.2 Discussion

The data of the 17th century is comparable to what was found in the 16th century. The passive periphrasis is still the main use of *werden*, although the highest number of instances was in the *Perfekt* tense. According to Bybee (2006):

A grammaticalizing construction's frequency of use increases dramatically as it develops. One source of the increased frequency is an increase in the types of contexts in which the new construction is possible. (p. 2)

The number of instances in this tense (not only in the passive but also with *werden* used as a copula) reflects the advanced grammaticalized status of the present perfect in the 17th century.

Among the instances in the *Perfekt* tense, there are also examples of afinite constructions. According to Breitbarth (2005), these constructions experienced an increase in frequency starting from the last two centuries of the Early New High German period, before disappearing after the 17th century. The data of this century reflect this frequency increase, with a total of sixty-eight instances of these constructions.

This century displays also the highest number of attestations of *werden* combined with the verbs in the infinitive forms so far, and the ones with the auxiliary in the past tense were all in the subjunctive mood. The increased frequency of this periphrasis and the absence of instances with the auxiliary in the indicative *Präteritum* indicate that this construction has reached the end of its grammaticalization process as a future marker. This is because it went from the old aspectual meaning to the new temporal one (Kotin, 2003, p. 169). Once this development is completed, the use of the auxiliary in the indicative *Präteritum* is incompatible with this newly acquired function. As a consequence, the combinations with *werden* in the simple past and the verbs in the infinitive forms slowly disappear from the language (Smirnova, 2006, p. 261).

Other disappearing constructions in this period are the combinations of *werden* with the present participles. No instances of these periphrases were found in the previous century, and only one was found in the 17th century. As discussed in the review of the literature, Kotin (2003) attributes this disappearance to the almost simultaneous decay of the periphrases of the present participles with the verb *sîn* (to be). Kotin (2003) discusses a semantic opposition between the non-mutative and non-terminative meaning expressed by the *sîn* constructions and the prevalent ingressive meaning of those with *werden*. With the decline of the periphrases with *sîn*, there is no need to have a semantically contraposing periphrasis, and the constructions with *werden* and the present participle disappear as a result of such syntactic change (p. 166).

The instances with the modal verbs also display also some attestations of the *werden* passive combined with the modal verb *dürfen*. Such findings indicate an advance of the grammaticalization process of this modal verb, which started already in Old High German and culminates precisely in the 17th century (Diewald, 2012).

Among the instances of *müssen* and *werden* used as a copula verb, there was one attestation with *werden* in the past participle. Consider example (112):

(112) und davon **schwanger worden**
and from that pregnant-ADJ become-AUX.PTC.PST
seyn müste
be-AUX.INF should-AUX.SBJV.PST
'And she should become pregnant from that'
(*Schaubühne*, 5152 Sp. B, 41)

In Modern German, such constructions contain *werden* in the infinitive form and the only finite form is the modal verb. The presence of *werden* in the past participle could suggest that, in this particular case, *worden* and the adjective *schwanger* were perceived as a unit rather than a combination of two separate elements. As has been discussed for the previous century, the frequency with which a given combination is used can influence the way such combination is regarded in the language (Bybee, 2003, p. 7). The attestations of *werden* used as a copula verb in the *Perfekt* and *Plusquamperfekt* tenses have been found with regularity throughout the texts of the Middle and Early New High German corpora. Such frequency could be the underlining cause for the presence of instances such as the one shown in examples (96).

6.4.3 Summary

The texts from the 17th century contained a total of 579 instances of *werden*. Although the highest number of forms overall was in the present tense, the major part of the instances of the *werden* passive was in the present perfect. Among those, there were also occurrences of afinite constructions. *Werden* as a copula verb was found in all four tenses and there was an increased number of instances of *werden* with verbs in the infinitive forms. All those attestations with the auxiliary in the simple past were in the subjunctive mood. Only one form of *werden* plus present participles was found. No forms of *werden* used as a full verb were found.

Texts from the 17th century contained a total of 92 instances of *werden* with modal verbs. The majority of those were with *sollen* and *müssen*. The data contained also some instances of the *werden* passive and *werden* as a copula and three instances of *dürfen* combined with the *werden* passive.

6.5 *Werden* in Early New High German

6.5.1 Analysis

A total of 2,129 forms of *werden* were analyzed in texts from Early New High German. Among those, there was also a total of five attestations of *werden* used as a full verb (one found in

the second half of the 14th century, two in the 15th century, and another two in the 16th century). The following figure (Figure 32) shows the raw and normalized[13] frequencies of the instances of *werden* for each century analyzed in this chapter.

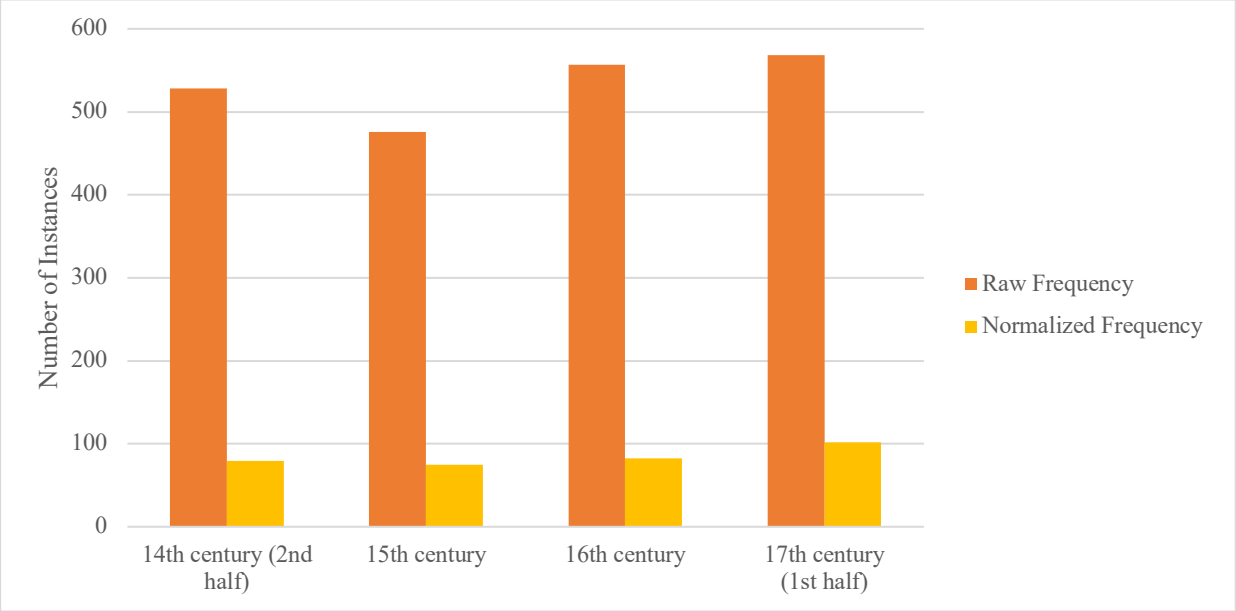


Figure 32: Raw and normalized frequencies of the instances of *werden* in Early New High German with past and present participles, adjectives, nouns, and infinitive verbs

Figure (32) shows a stable use of *werden* throughout the Early New High German period. The difference between raw and normalized frequencies between centuries is minimal and without any significant discrepancies between the four centuries analyzed.

Figure (33) shows the raw and normalized frequencies of *werden* and the elements with which it was combined in the corpus used for Early New High German:

[13] In this chapter, the word count has been normalized per 10,000 words.

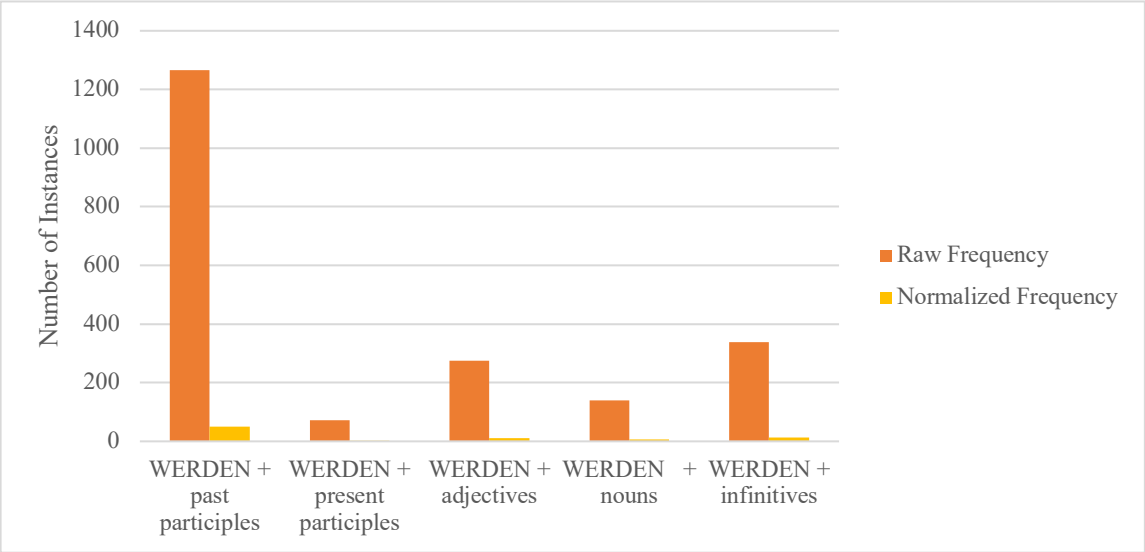


Figure 33: Raw and normalized frequencies of the instances of *werden* with past and present participles, adjectives, nouns, and infinitive verbs

Figure 33 shows that the main use of *werden* in Early New High German was for the passive constructions, followed by *werden* used as an auxiliary in combination with verbs in the infinitive form. The use of *werden* as a copula with adjectives and nouns is also common, while the lowest number of instances was that of the combinations of *werden* with the present participles.

The next figures (Figure 34 and Figure 35) show the raw (35) and normalized (36) frequencies of all elements found with *werden* in each century.

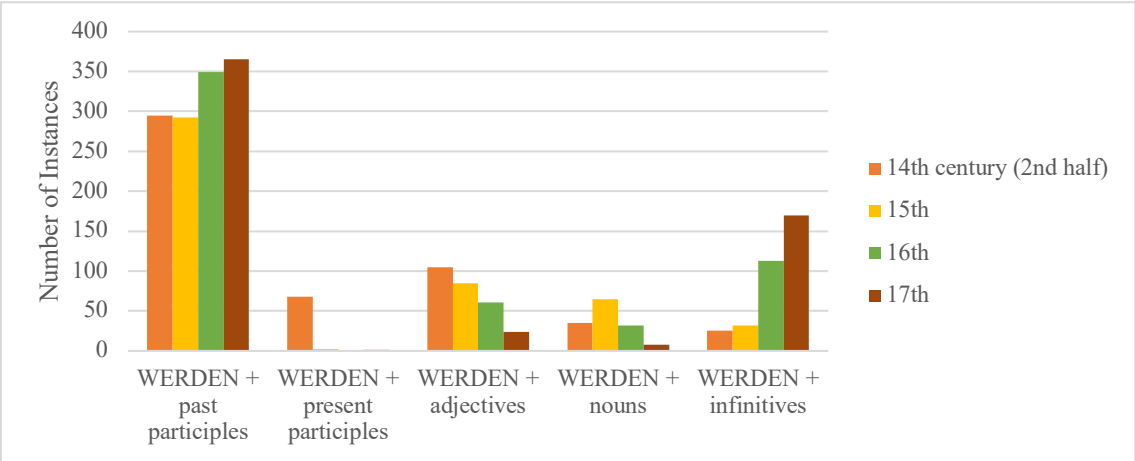


Figure 34: Raw frequency per century of the instances of *werden* with past and present participles, adjectives, nouns, and infinitive verbs

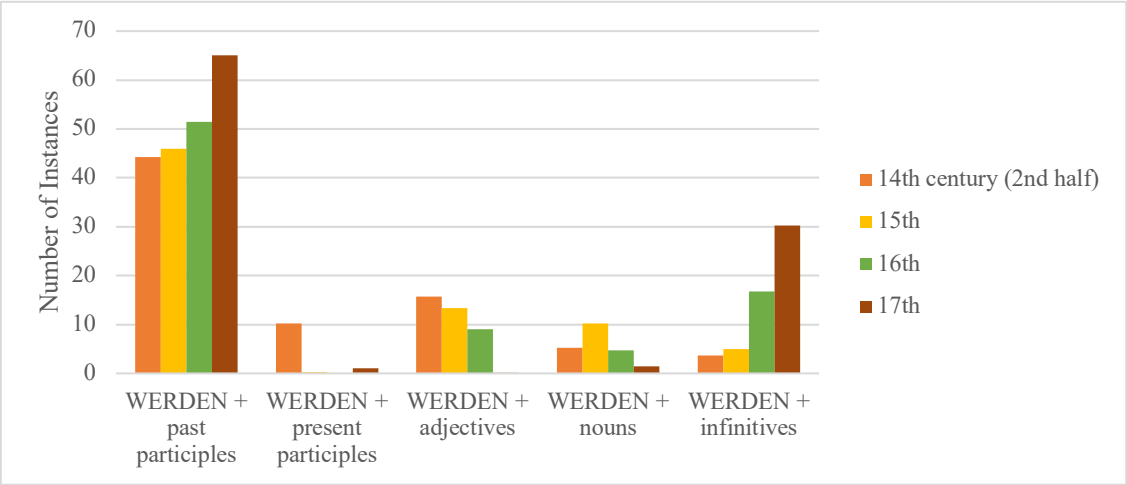


Figure 35: Normalized frequency per century of the instances of *werden* with past and present participles, adjectives, nouns, and infinitive verbs

The raw and normalized data indicate that the period in which the passive was used with most frequency is the 17th century. This is also the century with the highest number of combinations of *werden* with infinitive verbs and *werden* in the *Perfekt* tense. The 17th century is, together with the 16th century, also the period with the lowest number of instances with *werden* with the present participles and *werden* used as a copula with adjectives and nouns. In Early New High German, the highest number of attestations of *werden* plus present participles were found in the second half of the 14th century.

The following figure (Figure 36) shows the raw and normalized frequencies with the *werden* attestations with a focus on the tenses in which these instances were found:

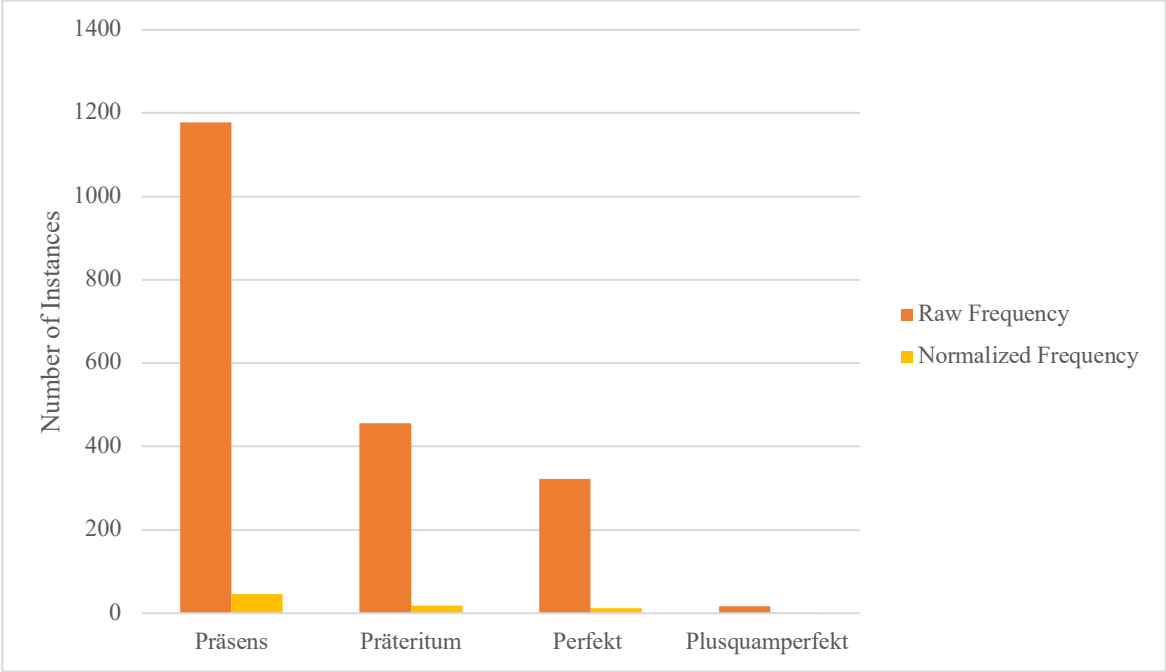


Figure 36: Raw and normalized frequencies of the instances of *werden* according to the verbal tenses

The raw and normalized frequency data display a clear preference for *werden* used in the *Präsens* tense, followed by the *Präteritum* and the *Perfekt*. The *Plusquamperfekt* is the least used tense in this corpus.

The following figures (Figure 37 and Figure 38) display the raw (37) and normalized (38) frequencies of the attestations of *werden* in the present (*Präsens*), simple past (*Präteritum)*, present perfect (*Perfekt*), and past perfect (*Plusquamperfekt*) tenses with focus on the centuries in which they were found:

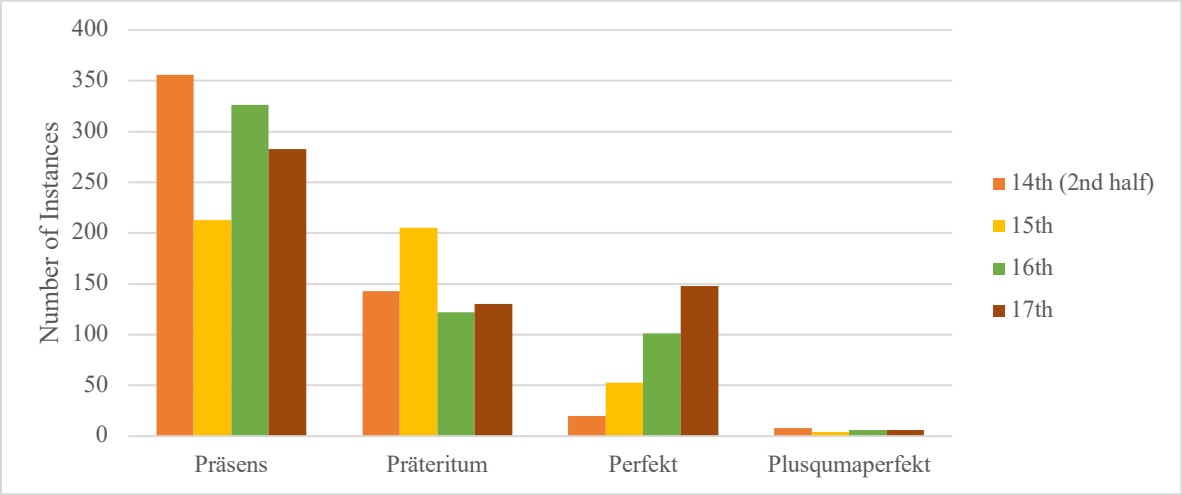


Figure 37: Raw frequency per century of the instances of *werden* according to the tenses

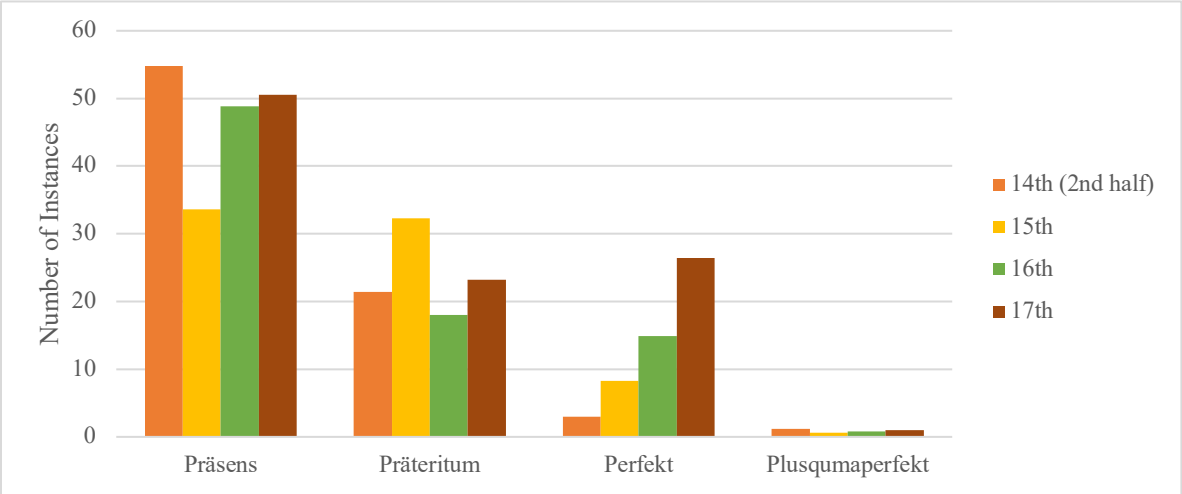


Figure 38: Normalized frequency per century of the instances of *werden* according to the tenses

The raw and normalized frequency data indicate that the *Präsens* is the main tense used in all four centuries of the Early New High German period. While the number of instances in the *Präteritum* does not change dramatically between centuries, both figures show a slow increase in the use of *Perfekt* starting from the 15th century. The low frequency of the *Plusquamperfekt* can be found in all four centuries without any significant difference in the number of instances contained in the texts.

The next figures (Figure 39 and Figure 40) display the raw and normalized frequencies of the elements combined with *werden* according to the tenses in which *werden* was found:

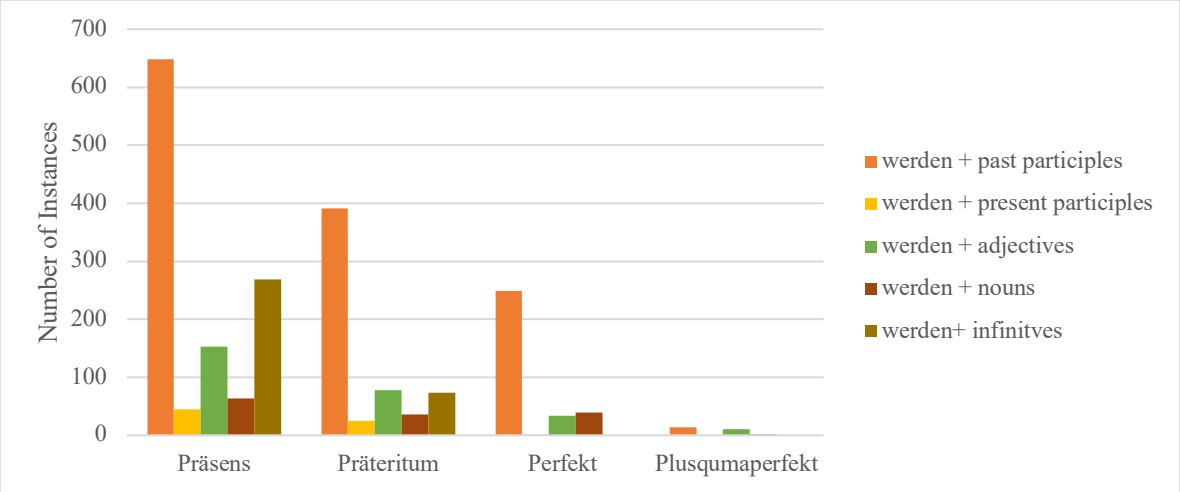


Figure 39: Raw frequency of the elements combined with *werden* according to the tenses

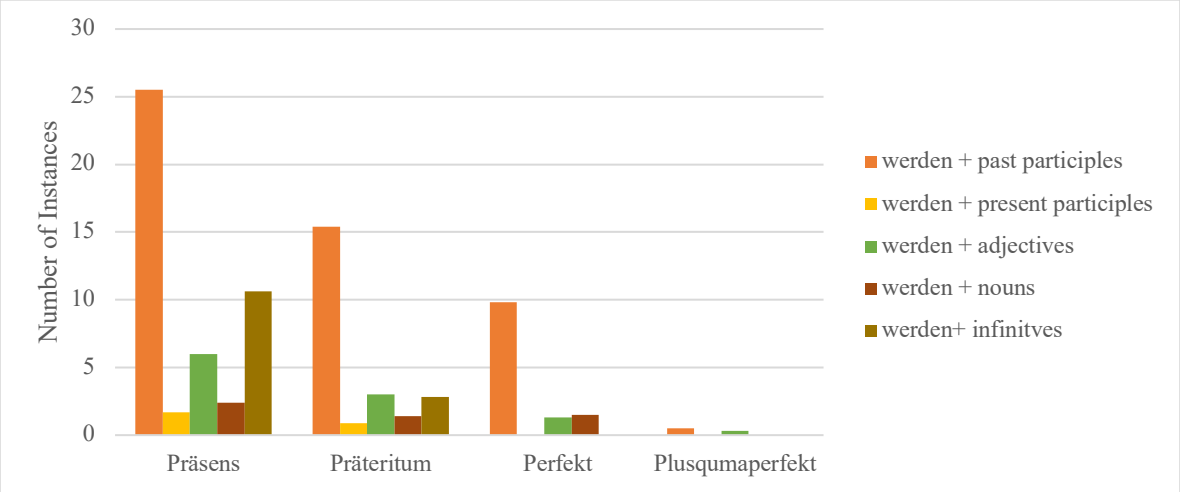


Figure 40: Normalized frequency of the instances of *werden* according to the tenses

Both figures indicate that the passive constructions were mostly used in the *Präsens*, followed by the *Präteritum*, the *Perfekt*, and the *Plusquamperfekt*. Similar tense preferences were found for *werden* used as a copula verb with adjectives and nouns. Present participles and verbs in the infinitive forms only appear in the present and the simple past.

A total of 270 modal verbs combined with *werden* have been found in the corpus of Early New High German. The next figure (Figure 41) shows the raw and normalized frequencies of the combinations of *werden* with the modal verbs.

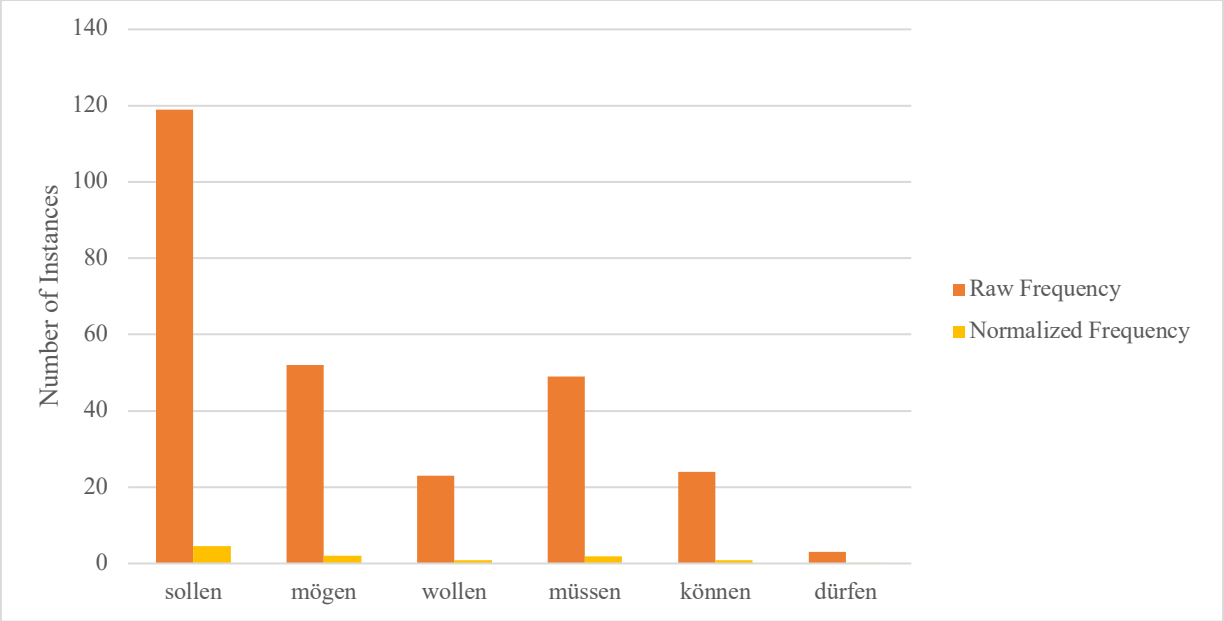


Figure 41: Raw and normalized frequency of the instances of *werden* with modal verbs

Figure 41 shows that instances of all modal verbs were found in this corpus. Specifically, it shows that *sollen* was used most frequently, followed by *mögen*, *wollen*, and *müssen*. The modal verbs used less frequently in this corpus were *können* and *dürfen*.

The next figures (Figure 42 and Figure 43) show the raw and normalized frequencies of combinations of modal verbs and *werden* used as an auxiliary or copula:

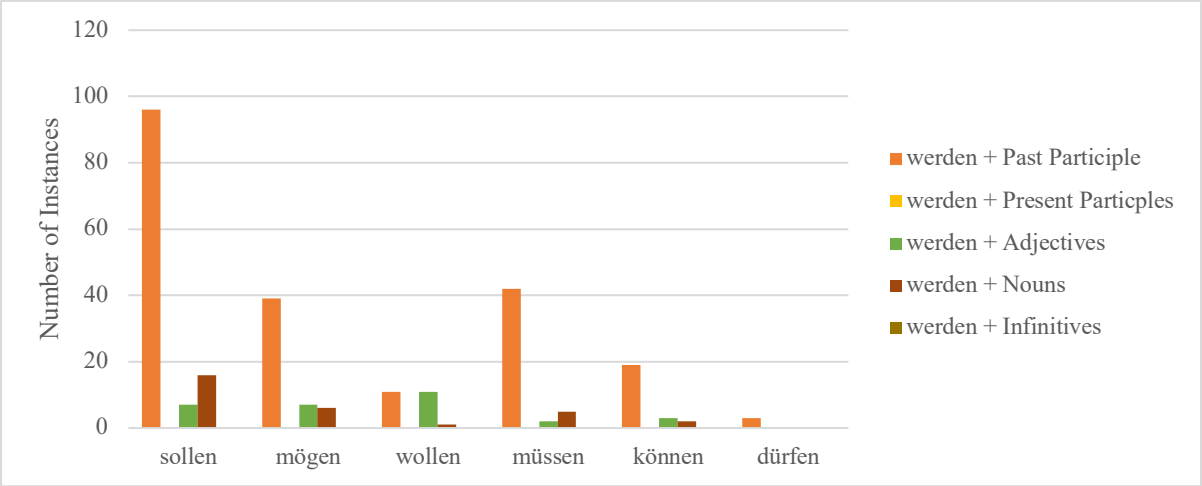


Figure 42: Raw frequency of the instances of *werden* according to the tenses

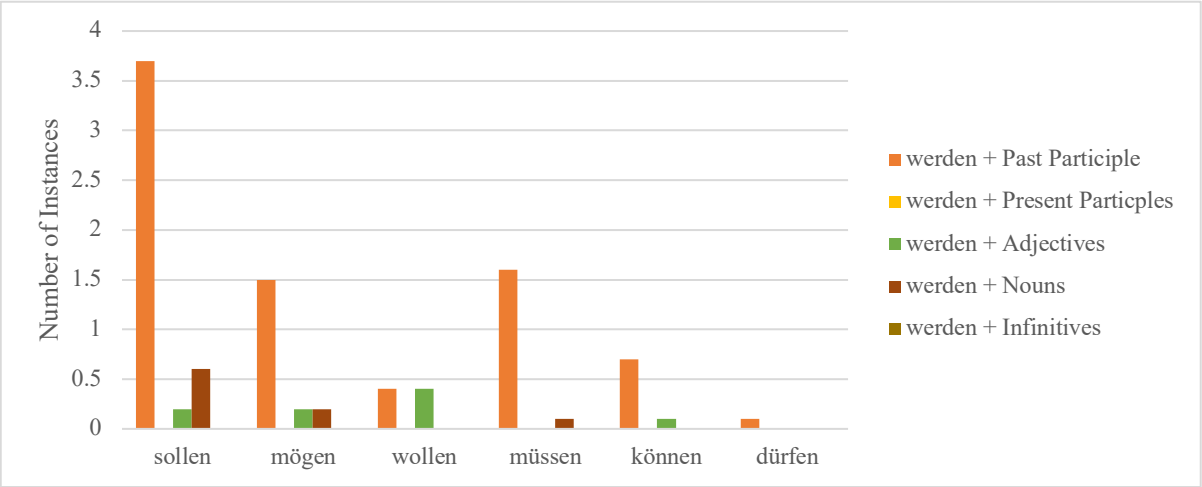


Figure 43: Normalized frequency of the instances of *werden* according to the tenses

The data from the two figures shows that, in the majority of the instances, all six modal verbs were used with the highest frequency in combination with *werden* in the passive constructions. The instances in which *werden* was used as a copula with adjectives and nouns are also quite common. Among the modal verbs, the most frequent one was *sollen* (for both passive and *werden* copula), followed by *müssen*, *mögen*, *können*, *wollen*, and *dürfen.*

The following figures (Figure 44 and Figure 45) display the raw and normalized frequencies of the instances of *werden* and modal verbs:

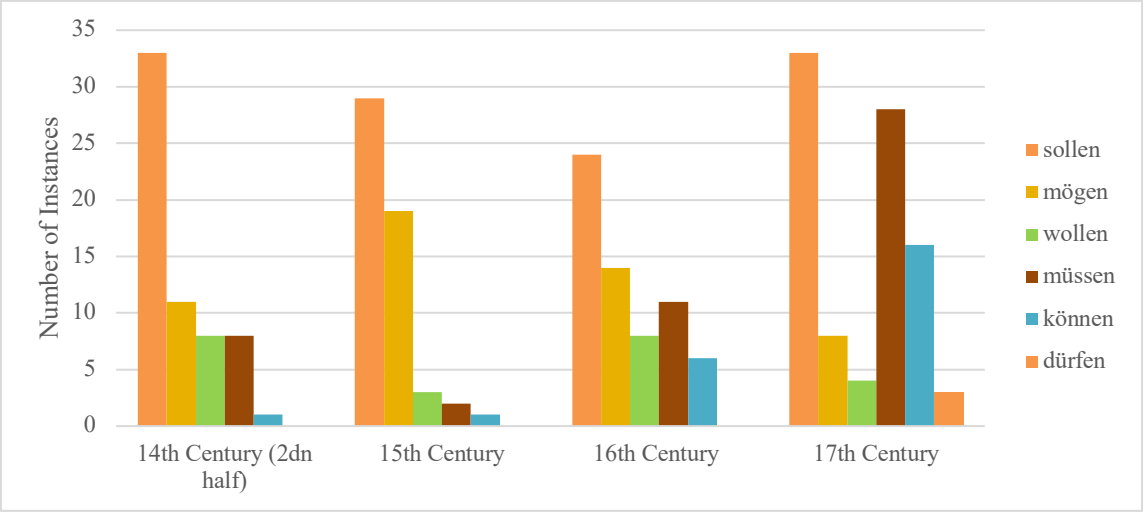


Figure 44: Raw frequency of the instances of *werden* divided per centuries

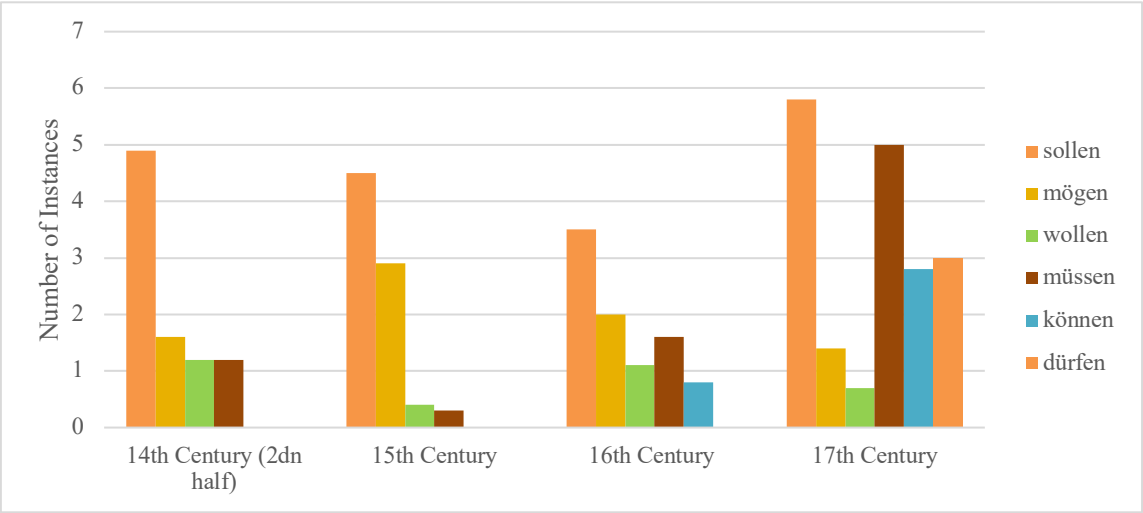


Figure 45: Normalized frequency of the instances of *werden* divided per centuries

The raw and normalized frequencies of the instances of *werden* in the four centuries analyzed in this chapter indicate that the 17th century is both the century with the highest number of instances overall and the century in which *werden* was found with all the six modal verbs *sollen*,

mögen, *wollen*, *müssen*, *können*, and *dürfen*. The century with the second-highest number of instances overall is the second half of the 14th century followed by the 15th and the 16th centuries.

6.5.2 Discussion

The instances analyzed in this chapter offered some valuable insights into the historical development of *werden* and the elements with which it was combined, such as the present and past participles, and verbs in infinitival form. Further, the attestations from this corpus reflected some of the most important syntactic changes that happened in German during the Early New High German period. Overall, and as observed also for Middle High German, the most frequent use of *werden* was as an auxiliary in the passive constructions. Such periphrases were found in all the texts from all five dialect areas included in this chapter. The *werden* passive was found mostly in the *Präsens* tense, followed by the *Präteritum*. This corpus, however, contained a significant number of instances in the *Perfekt* tense (and a few in the *Plusquamperfekt* tense). This increase in the number of occurrences in the present perfect can be linked to the fact that the Early New High German period is the stage in which this tense reaches the final steps of its grammaticalization process which had started in Old High German (Kuroda, 1999; Concu, 2016).

It has been already discussed how "a grammaticalizing construction's frequency of use increases dramatically as it develops" and how this corresponds to a broadening "of the contexts in which the new construction is possible" (Bybee, 2006, p. 2). The data show that this boost in frequency started in the 15th century and continued in the 16th and 17th centuries. Further, in Middle High German, the attestations in the *Perfekt* tense were limited to the combinations of *werden* copula plus nouns and adjectives. In Early New High German, the *werden* passive appeared also in the present perfect, suggesting that this tense had started expanding the communicative contexts in which it could be used exactly in this period. Such expansion relates also to the advanced grammaticalization status of this tense in these centuries. Further, in all the attestations found in the *Perfekt* and in the *Plusquamperfekt* tenses (both the ones in which *werden* is used as an auxiliary for the passive and as a copula verb with adjectives and nouns), the past participle form of *werden* did not add the *ge-* prefix. In the previous chapter, it has been discussed that the absence of this prefix could be rooted in the original semantics of *werden* of "to come into being" and "happen." Further, such meanings have presumably enhanced processes of analogical association with verbs with a similar semantics such as *finden* and *bringen*. These verbs also did

not add the *ge*-prefix in their past participle forms when used in the *Perfekt* and in the *Plusquamperfekt* tenses. The Early New High German data hints to a similar situation, implying the possibility that the meaning of *werden* had remained almost unchanged from the Middle High German period.

Many of the attestations in the present perfect tense found in this corpus, and specifically those in the texts from the 16th and 17th centuries, have been classified as afinite constructions. This denomination refers to the ellipsis of the conjugated auxiliary in subordinating clauses. Breitbarth (2005) claims that the rise of these constructions was caused by two different factors strictly linked to each other. On the one hand, in Early Middle High German, she reports an increase in textual complexity with a consequent intensification of the use of embedded clauses. On the other hand, the verb in sentence-final position started to be considered a marker of subordination. The combinations of these two shifts created the right environment for the ellipsis of the conjugated verb and the rise of the afinite constructions (Breitbarth, 2005, p. 16). All the occurrences of afinite constructions from this corpus have indeed been found only in subordinating clauses, further supporting Breitbarth's claims.

The Early New High German data also show the rise of the *werden* future and the simultaneous slow decline of combinations of *werden* with the present participle. Although some scholars have argued that the periphrases of *werden* + infinitives originated from the ones with *werden* + participles, the instances analyzed in the centuries from the 11th to the first half of the 14th century had already indicated an alternative scenario. The attestations found in this chapter, especially those found in the 16th and 17th centuries, confirmed the trends observed in Middle High German, supporting the claims of scholars such as Kotin (2003) and Krämer (2005). According to them, both these periphrases originated from different but related sources and coexisted as "twin" constructions until the periphrases of *werden* plus present participles slowly disappeared (Kotin, 2003, p. 165). Their decline is not to be linked to the rise of the *werden* future but to the disappearance of the periphrases of *sîn* and the present participles (p. 165). The gradual decrease of the number of instances of *werden* and the present participles observed here suggests that both these types of periphrases (*sîn* and *werden* with present participles) were slowly declining in Early New High German. The rise of the *werden* future is to be interpreted as a result of the ongoing grammaticalization process of this construction as a whole, and to the related semantic changes that affected *werden* at this stage.

The instances of *werden* with modal verbs have shown a prevalence of combinations with the verbs *sollen*, *müssen*, and *mögen*. However, the four centuries of the Early New High German period contained attestations with *wollen*, *können*, and *dürfen*. In Middle High German, these two modal verbs did not appear combined with the passive (*können*) or not at all (*dürfen*). According to Diewald (2012), their use was initially limited by semantic and syntactic restrictions which explained the low number of occurrences of these verbs in any kind of combination with *werden*. In Early New High German, the data could indicate that these restrictions are gradually becoming weak and that the number of contexts in which *können* and *dürfen* are allowed to be used is slowly increasing.

6.5.3 Summary

The texts from the Early New High German corpus contained a total of 2,129 instances of *werden* and a total of 270 attestations of modal verbs combined with this verb. The majority of the instances were of *werden* used in the passive periphrases. The auxiliary was mostly found in the *Präsens* and in the *Präteritum*, but some attestations were also found in the *Perfekt* tense. All the instances in this tense and in the *Plusquamperfekt* did not add the prefix *ge-* to *werden* in the past participle form. This corpus also contained a modest number of attestations of *werden* used as a copula verb in combination with adjectives and nouns, and instances of *werden* with verbs in the infinitive forms. The combinations of *werden* and the present participles were less common, especially if compared with the data of the Middle High German corpus.

The corpus of Early New High German contained instances with all modal verbs, including *können* and *dürfen*. The highest number of combinations was with *werden* used as an auxiliary with the passive periphrases, while a lower number of attestations had *werden* as a copula used in combinations with nouns and adjectives.

In the next section, I will discuss and compare the attestations of both the Middle High German and the Early New High German corpora.

6.6 *Werden* in Middle and Early New High German

In this section, I will compare the instances of *werden* found in the texts from the Middle High German and Early New High German corpora. Firstly, I will display both raw and normalized

frequencies of all the attestations of *werden* with past and present participles, adjectives, nouns, and verbs in the infinitive forms. Secondly, I will show the combinations of *werden* with the modal verbs.

Figure (46) shows the raw and normalized frequencies of the total number of attestations of *werden*.

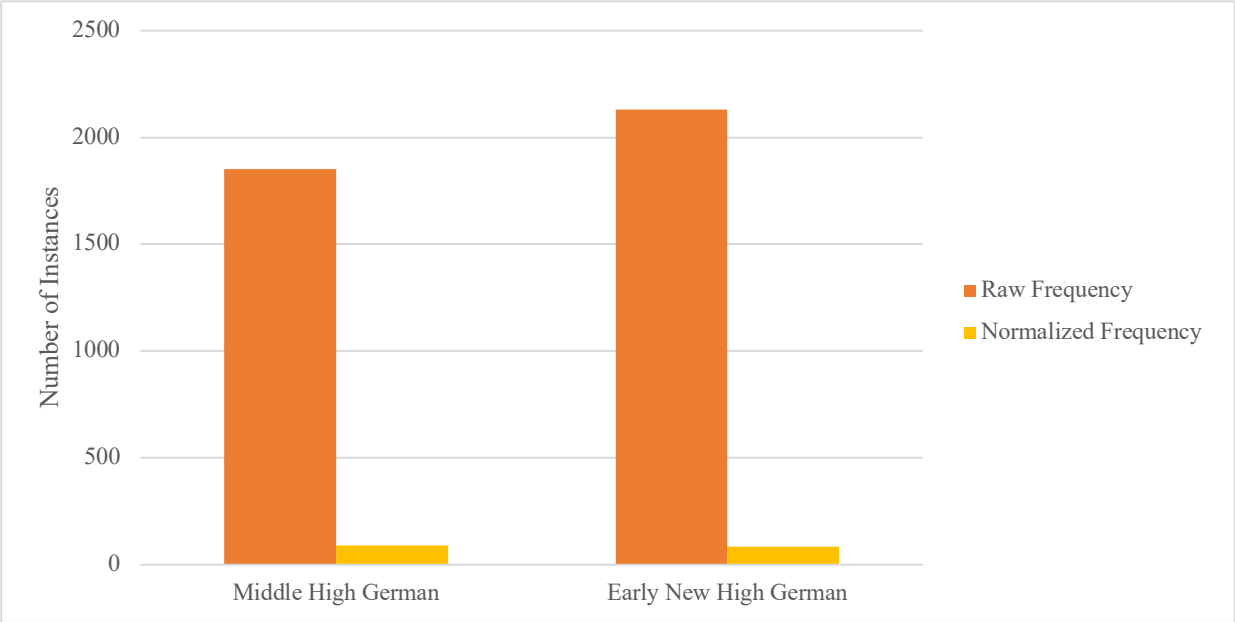


Figure 46: Raw and normalized frequencies of all the instances of *werden*

The data of both the raw and normalized frequencies show that *werden* was slightly more frequent in the corpus of Middle High German than in the corpus of Early New High German. However, the numbers of instances do not differ much, and this probably indicates that the use of *werden* in the written language was equally distributed in both periods.

The next figures (figure 47 and figure 48) display the raw and normalized frequencies of the elements with which *werden* was found:

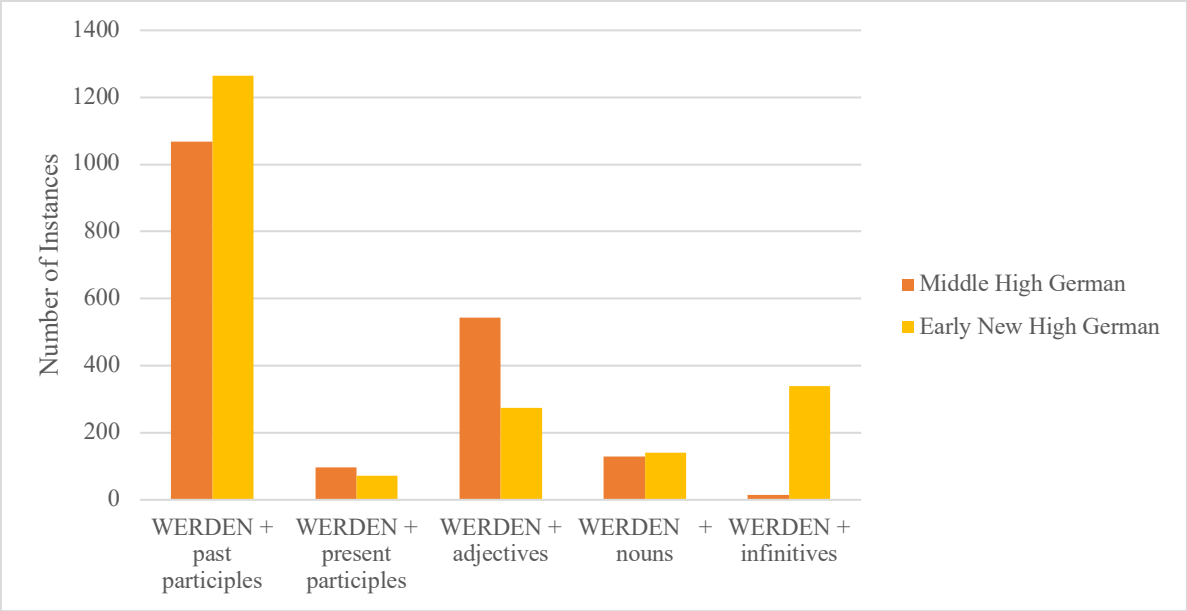


Figure 47: Raw frequency of *werden* with past and present participles, adjectives, nouns, and infinitive verbs

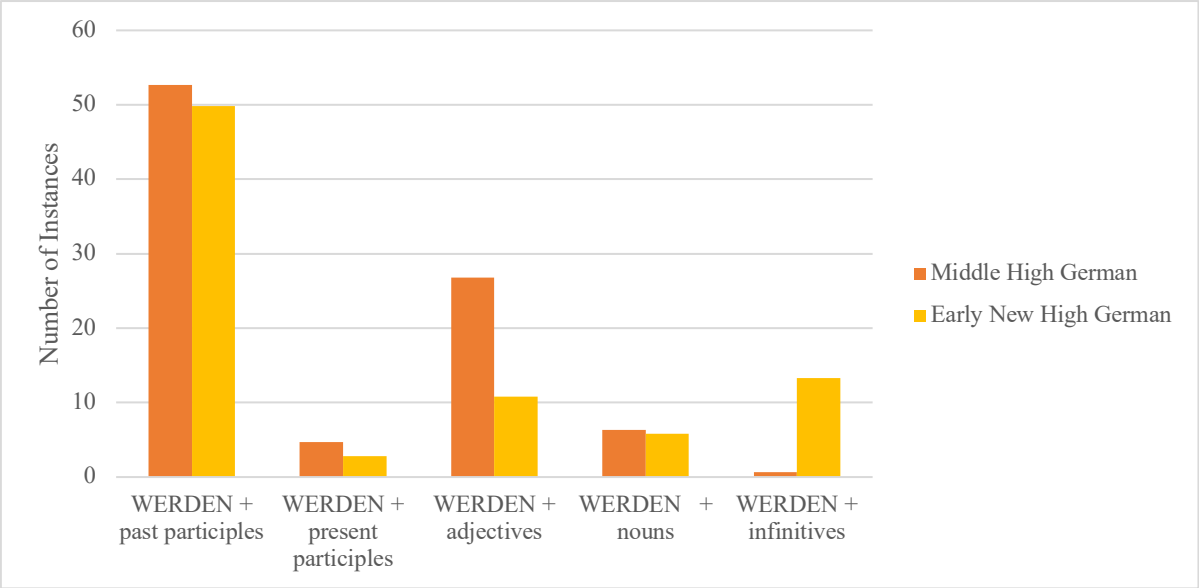


Figure 48: Normalized frequency of *werden* with past and present participles, adjectives, nouns, and infinitive verbs

The raw and normalized frequencies show that *werden* passive periphrases were used in almost equal frequency in both Middle and Early New High German. The same can be argued for

the instances in which *werden* was used in combination with present participles (although almost all the instances in Early New High German were from the texts of the second half of the 14th century) and with *werden* used as a copula with nouns. *Werden* with adjectives was used most frequently in Middle High German, whereas the majority of the instances of *werden* with verbs in the infinitive forms were found in Early New High German.

The next figures (figure 49 and figure 50) show the raw and normalized frequencies of the attestations of *werden* in *Präsens*, *Präteritum*, *Perfekt*, and *Plusquamperfekt*:

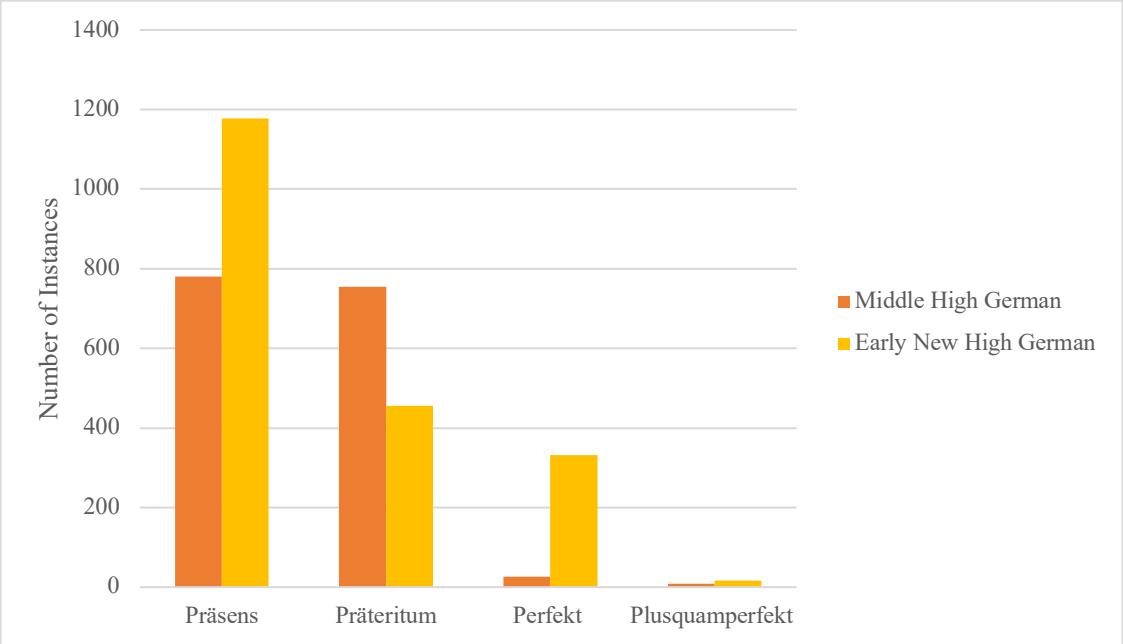


Figure 49: Raw frequency of the attestations of *werden* in *Präsens*, *Präteritum*, *Perfekt*, and *Plusquamperfekt*

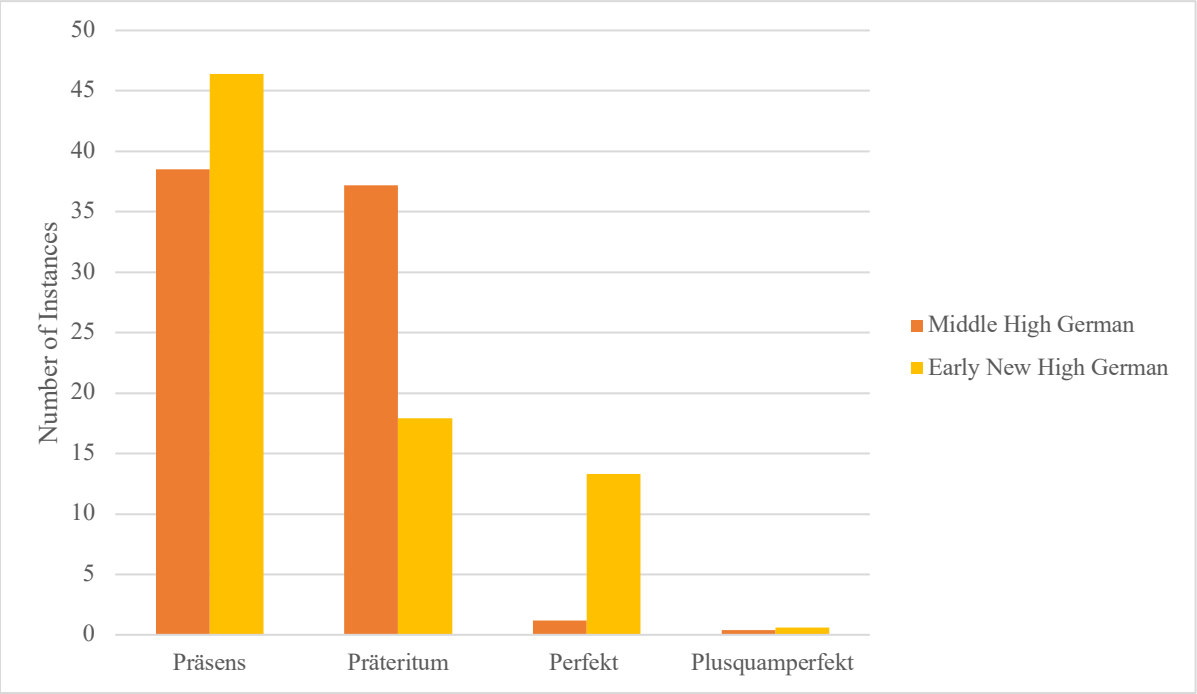


Figure 50: Normalized frequency of the attestations of *werden* in *Präsens*, *Präteritum*, *Perfekt*, and *Plusquamperfekt*

The raw and normalized frequencies in the figures (49) and (50) indicate that, in Middle High German, the *Präsens* and the *Präteritum* were the tenses used most frequently. In Early High German, however, the data show a different situation. The increase of the frequency of the *Perfekt*, which was grammaticalizing as a past tense in these centuries (Kuroda, 1999, Concu, 2016), probably overlaps with a slow decrease in the use of the *Präteritum* as a unique tense to refer to past events. Although the simple past is still frequent, speakers in Early New High German switched gradually to the use of the present perfect in some of the communicative contexts in which it was normally used in Old and Middle High German (Concu, 2016). The *Plusquamperfekt*, although it increases its frequency in Early New High German, still remains limited to a few attestations in both periods.

The next figures (figure 51 and figure 52) show the raw and normalized frequencies of the total number of instances of *werden* combined with modal verbs.

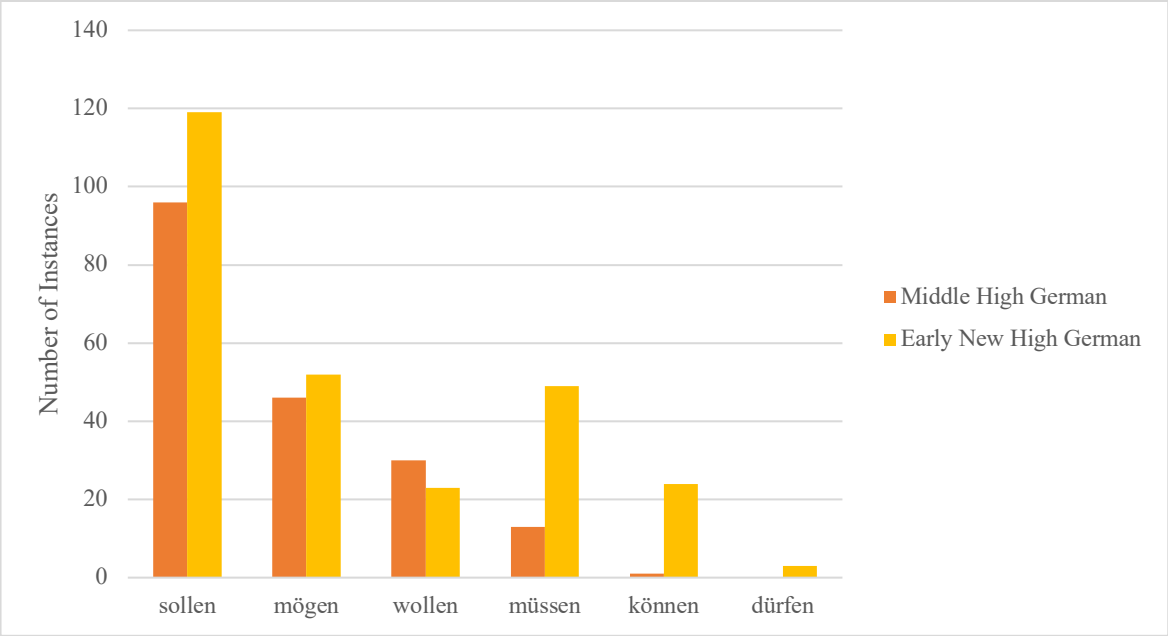


Figure 51: Raw frequency of the instances of *werden* with modal verbs

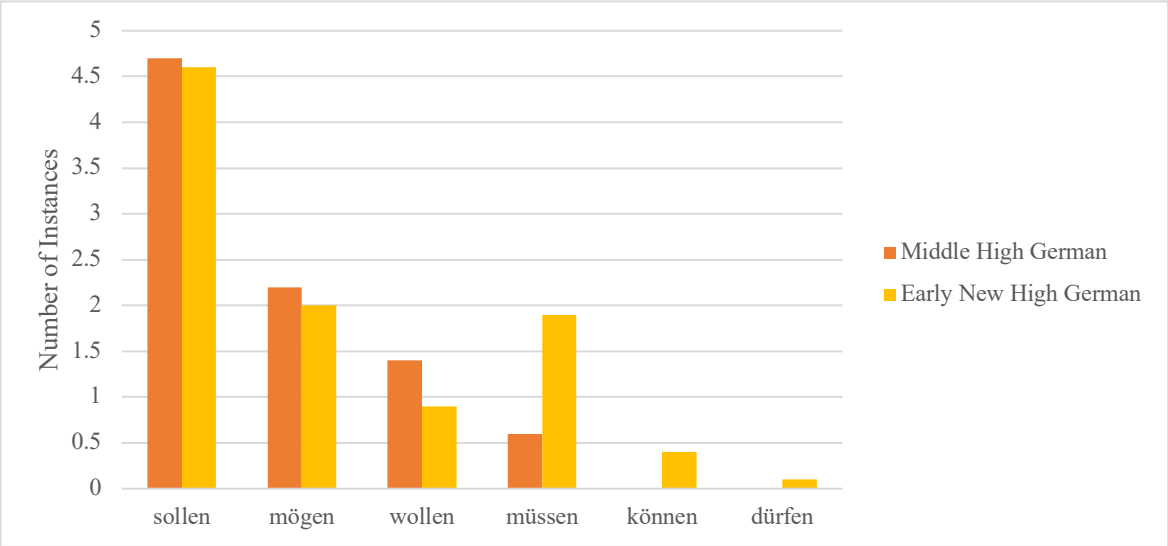


Figure 52: Normalized frequency of the instances of *werden* with modal verbs

The data of the raw and normalized frequencies shows that the use of the modal verbs *sollen*, *mögen*, *wollen*, and *müssen* was quite similar in both Middle and Early New High German. The differences between these two periods are visible when considering the attestations of *können*

and *dürfen*. In Middle High German, only one instance of *können* was found, whereas the corpus did not contain any form of *werden* (as an auxiliary in the passive periphrases or as a copula verb with adjectives and nouns) combined with *dürfen*. In Early New High German, instances were found with *werden* combined with both *können* and *dürfen*, indicating a different grammatical status of modal verbs in this period.

6.6.1 Summary

The comparison between raw and the normalized frequencies of the attestations from the Middle and Early High German corpora shows that *werden* was used in the passive periphrases and as a copula verb with nouns and adjectives in comparable ways in both these periods. The discrepancies between the two corpora are in the attestations of *werden* combined with verbs in the infinitive forms and in the number of instances found in the *Perfekt* tense. The frequency of both these types of attestations increases drastically in Early New High German, starting specifically in the 15th century.

The data on the modal verbs also show some disparity between the two corpora. While the combinations of *werden* with *sollen*, *mögen*, *wollen*, and *müssen* are similar in both Middle and Early New High German, the differences are in the use of *werden* combined with *können* and *dürfen*. Only one instance of *können* and no instances of *dürfen* were found in Middle High German, while the Early New High German corpus contained attestations of *werden* with both these modal verbs.

In the next section, I will offer the analysis of the attestations of *werden* in Middle and Early High New High German within a complex network framework.

CHAPTER 7. THE NETWORK ANALYSIS

This chapter focuses on the analysis of the networks and the attestations of *werden* from Middle and Early New High German. First, I present the results for the networks created for each century and discuss their features. Second, I present the results obtained after applying the analysis of centrality measures to the attestations of *werden*. After that, I discuss the results as a whole and compare the results of both periods.

7.1 Middle High German

7.1.1 Analysis

In this section, I will describe the features of the networks created for the Middle High German period (11th – first half of the 14th century), concentrating the analysis first on the number of nodes and edges, the cluster coefficient, and the average path length for each of the four networks. The next Figure 53 displays the number of nodes and edges:

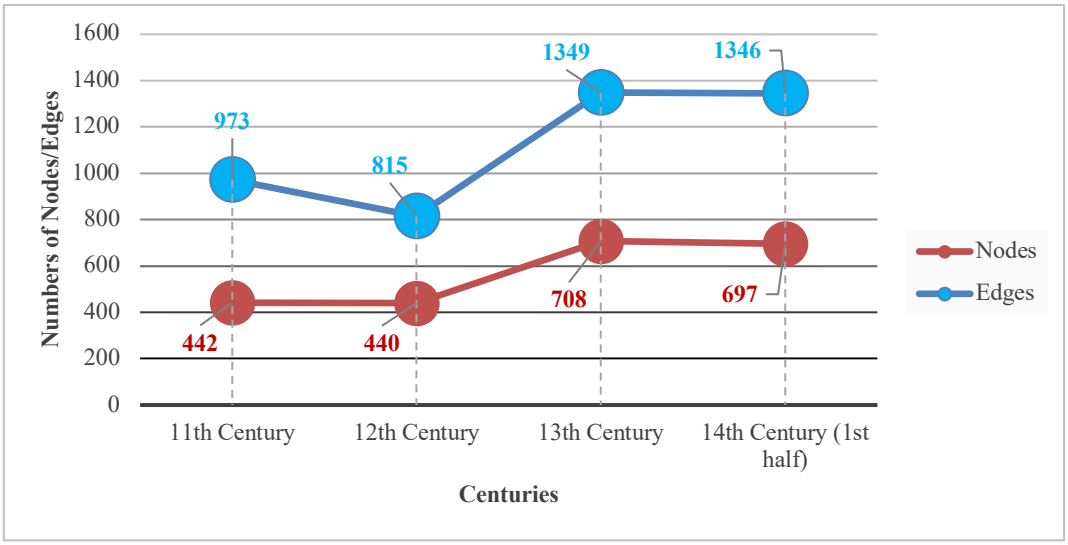


Figure 53: Number of nodes and edges of the Middle High German networks

Figure 53 shows the gradual growth of the number of both nodes and edges from the 11th to the first half of the 14th century. After a brief decline in the 12th century, the numbers in the

last two centuries show a positive correlation between the increase of the number of nodes and the expansion of the edges.

As discussed in the introduction, one of the common features of complex networks is related to the so-called small-world properties. These networks have a high cluster coefficient (the nodes are well connected to each other and the neighbors of a node are also linked together) and short average path length (only a few jumps in the network are required to move from one node to any given other node). In order to find out if the Middle High German networks also exhibit these features, four random networks[14] with the same degree have been created using the network randomizer application in Cytoscape. The results for both cluster coefficient and average path length and are shown in the following figures:

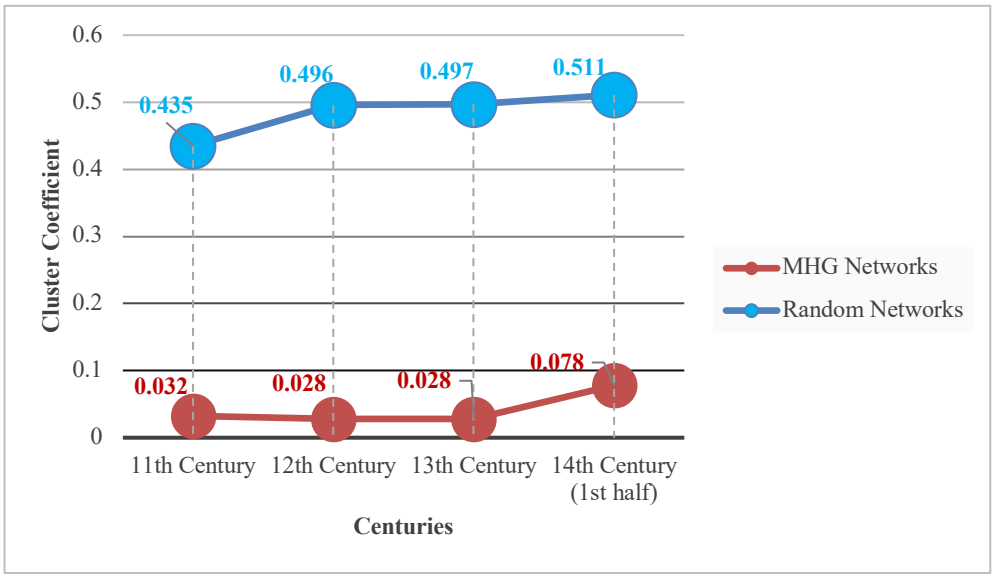


Figure 54: Cluster Coefficient in the Middle High German networks and in their random counterparts

[14] In this type of networks, links between nodes are established randomly, without any rules governing such connections.

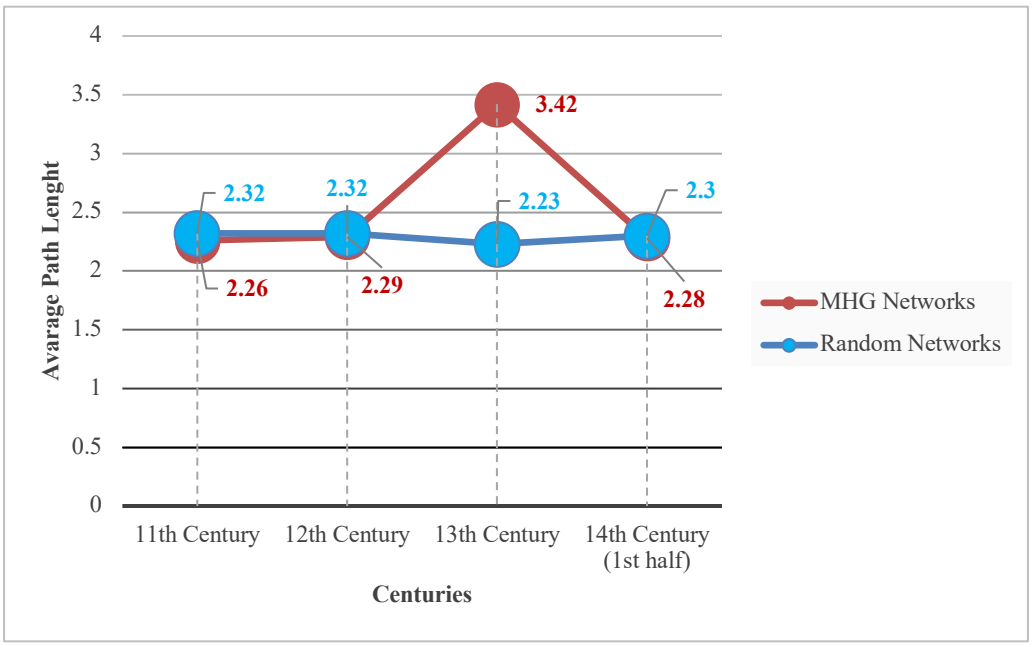


Figure 55: Average Path Length in the Middle High German networks and in their random counterparts

The comparison of the real syntactic networks with their random counterparts shows interesting results: As shown in Figure 54, the Middle High German networks have a significantly lower cluster coefficient than the random ones. The cluster coefficient refers to how well connected the neighboring nodes of a specific node are - in a social network, a high cluster coefficient means that "my friends" all know each other. The nodes in the Middle High German networks do not tend to cluster together in any of the four centuries. However, as shown in Figure 55, and with the exception of the 13th century, the values for the average path length of the real syntactic networks are similar if not slightly lower than those of their random counterparts. This means that the average number of steps along the shortest paths do not diverge much between the reals and the random networks (real vs. random).

Another common feature among complex networks is a degree distribution that follows a power law. In these networks (also called scale-free), only a few nodes possess the majority of the links, while the rest of the nodes have fewer connections. To establish if a network is scale-free, one must look at the degree distribution of the in- and out-degrees and plot these degrees on a double logarithmic scale (log-log plot). According to Barabási, "a quick estimate of the degree exponent can be obtained by fitting a straight line to pk [average degree) on a log-log plot" (2016,

Section 4.5). Cytoscape lets users fit a power law of the form of $y=axb$ in which y is the estimated degree exponent. The scatter plot is able to identify the presence of a long tail in the degree distribution, which is characteristic of scale-free networks. The results for the out-degree in Middle High German are shown in the next figures, in which the x-axis represents the out-degree, the y-axis signifies the number of nodes, and the red line is the fitted power law.

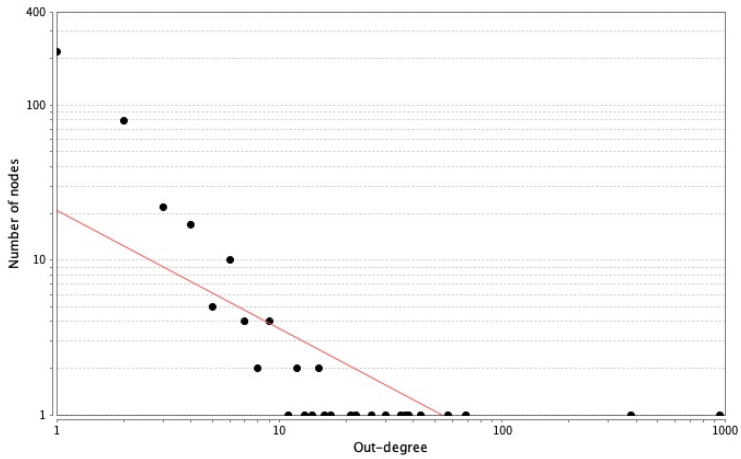


Figure 56: Out-degree 11th Century

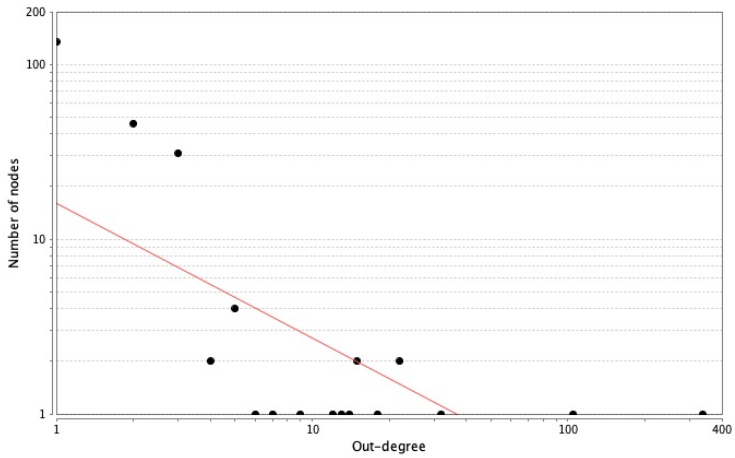


Figure 57: Out-degree for the 12th Century

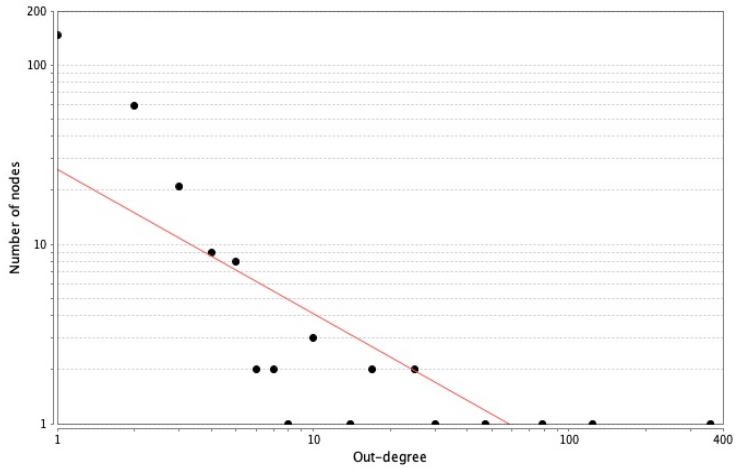


Figure 58: Out-degree for the 13th Century

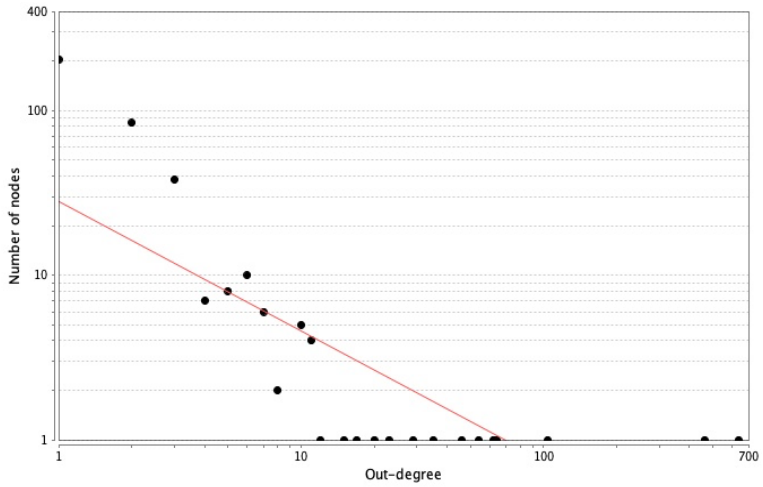


Figure 59: Out-degree for the 14th Century (first half)

The scatter plots of the Middle High German networks reveal the presence of a distribution in which only a few nodes have high out-degree values and those are visible on the low right corner of the x-axis. Moving from the right to the left of the x-axis, we can observe that the number of nodes that have lower out-degree values increases, until it reaches the highest part of the left corner of the y-axis. Although the presence of a distribution that follows a power law must be proven using advanced statistics, by observing the scatter plot it can be said that these networks are probably scale-free or, at least, that their out-degree distribution shows patterns that are similar to those found in scale-free networks.

The next figures show the raw and normalized frequencies of the types of interactions (verbal phrase; nominal phrase, and prepositional phrase) in the networks.

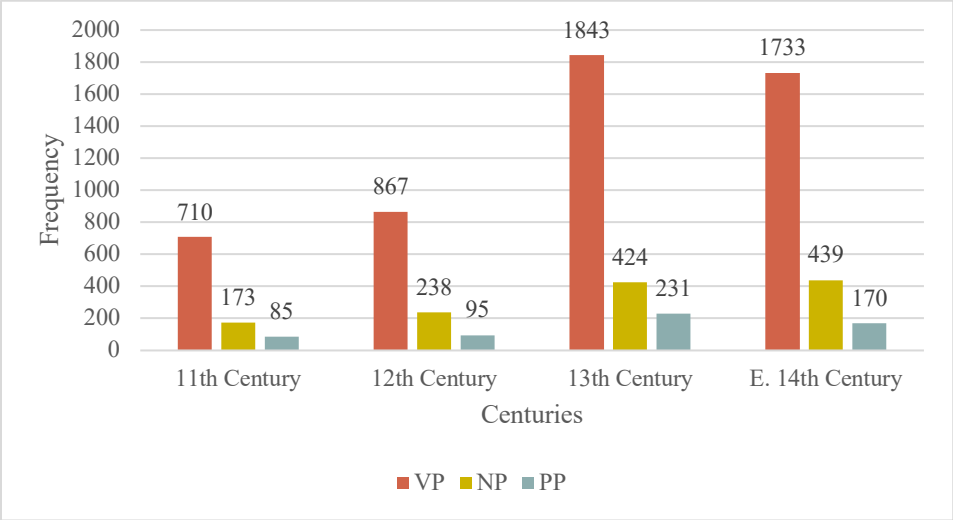


Figure 60: Raw frequency per century of the interactions between nodes

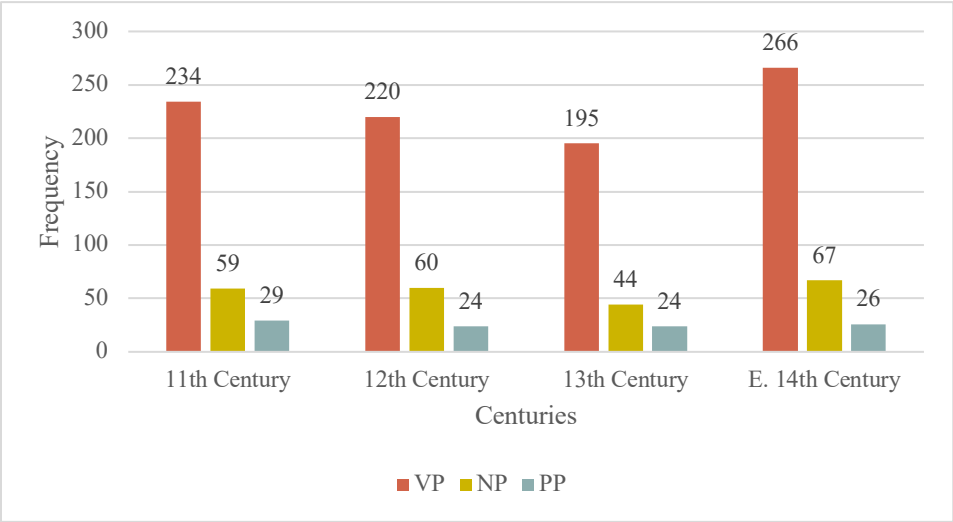


Figure 61: Normalized frequency per century of the interactions between nodes

The raw and normalized frequencies show that the types of interactions between nodes are very similar in Middle High German. The majority of the interactions are under verbal phrases, followed by nominal and prepositional phrases.

7.1.2 Discussion

As discussed in the previous section, the networks from Middle High German had a lower cluster coefficient than their random counterpart and the reason behind that may lay in the way these networks were built. The syntactic networks were created using a set of rules that were inspired by dependency grammar. This means that the nodes are linked together through asymmetrical relationships and that not all the nodes can be directly connected to each other in an equal manner. For instance, it would never be possible to have a direct link between a noun and an adverb, or between a verb and an adjective. Thus, the wiring probability of the neighborhood nodes of a specific node is lower than in a random network.

The comparison between real and random networks, however, also showed that the average path lengths in the real networks were similar if not slightly lower than those in their random counterparts (with the only exception of the 13th century). These results may reflect the syntactic inventory of German at this stage. For instance, analytic periphrastic constructions such as the Present and Past Perfect, which are formed by combining the auxiliary *haben* (to have) and *sein* (to be) and a past participle, were still not fully grammaticalized in this period and were, therefore, rarely used (Kuroda, 1999; Concu, 2016). In such constructions, the steps required from the head to all the other components of the verbal phrase are between three and four (assuming that there is at least one prepositional object attached to the verb). Speakers in Middle High German often used the synthetic simple past to refer to past events instead. In this case, the steps required from the head to reach the other component of the verbal phrase are only two (assuming again here that there is at least one prepositional object attached to the verb). Consider the following sentences again:

(19) er vone Judae **verrâten** **wart**
He by the Jews betrayed-PTCS.PST became-AUX.PST
'He was betrayed by the Jews'
(*Wessobrunner Glaube und Beichte*, 104ra,10)

(71) zw einmal **pin** ich **gestaint** **worden**
to one time **am**-AUX.PRS I stoned-PTCP.PST became-AUX.PTCP.PAST
'I have been stoned once'
(*Pillenreuth Mystik*, 188, 8)

The next figures display the difference between these two sentences in terms of average path length. The figure shows the nodes as words in their lemma forms and the interactions as verbal, nominal, or prepositional phrases:

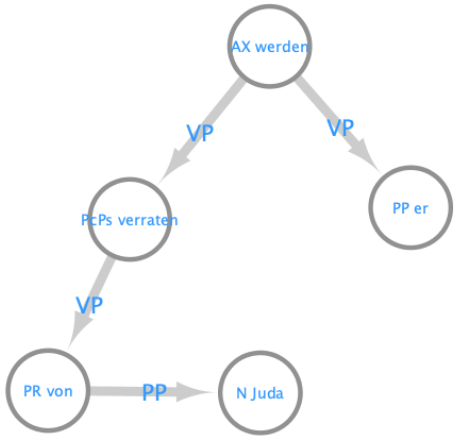


Figure 62: Syntactic network for “er vone Judae verrâten wart”

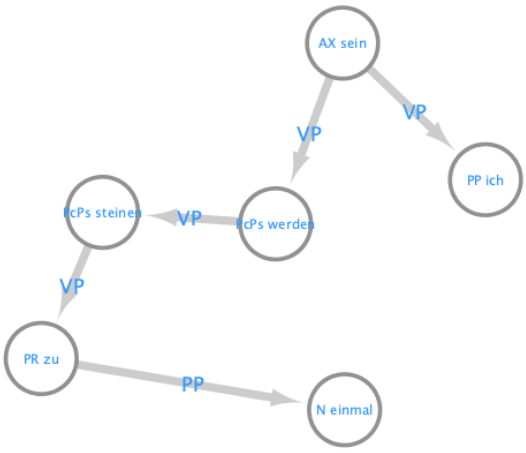


Figure 63: Syntactic network for “zw einmal pin ich gestaint worden”

The average path length in the first sentence with the passive in the simple past form is 1,571. The average path length in the sentence in which the passive appears in the present perfect on the right is 1,909. The fact that periphrastic constructions such as the present and past perfect

were not fully grammaticalized at this stage and not often used translates into few instances in which the verb *werden* can appear in such constructions. This can be then considered as one of the reasons for the small-world-like values of the average path length in the Middle High German networks.

Murray et al. (2016) claim that in a small-world network, "the clustering parameter is much larger than that of a random network while the average path length is similar" (p. 32). Thus, it can be argued that the Middle High German networks possess some of the properties of small-world networks. As discussed in the introduction, these properties are considered one of the main features of complex networks (Jiang et al., 2019) and they have also been observed in linguistic networks. For instance, Corominas and Murtra (2009) found that small-world properties already emerge in L1 syntactic networks at the age of 24 months, and therefore, the emergence of syntax implies a shift from a non-small-world language structure to a small-world language network. Small-words properties have been found also for L2 syntactic development, although these features were already present in the networks (no shift like in the L1 studies), as learners could rely from the beginning on the syntactic properties of their native language (Jiang et al., 2019). The analysis of the small-world properties of the Middle High German networks overlap in part with what was found in the literature.

Another property of complex networks is a power law degree of distribution, according to which, in a network, only a minority of nodes have a significantly high degree, while the majorities have a relatively low degree. This degree of distribution is a feature of scale-free networks. According to Cong and Liu (2014), in a syntactic network, the scale-free properties indicate that the ability of nodes to syntactically combine with other nodes are not equal but differ from one node to the other. As discussed in the analysis section, out-degrees of all four Middle High German networks show some similarities with the degree distribution that is normally found in scale-free networks. As for the small-world properties, these results can be seen as a reflection of the networks' structure, since the capacity of the nodes to establish a connection with another node depends on their grammatical function and are, therefore, not equal.

As the main goal of this study is to capture if the ability of *werden* to connect with specific nodes in the network changes over time, the next section of this chapter focuses on *werden* and analyzes its behavior in each of the four centuries of the Middle High German period. In order to do that, I will use the following centrality measures: betweenness, eigenvector, and eccentricity

centralities. The results for each century will be displayed in different charts and discussed in detail. As discussed in the methodology section, the in-degree refers to the incoming links that a node has. In the case of *werden*, such incoming connections come from modal verbs or the auxiliary *sein* in the perfect constructions. The out-degree refers to the outgoing links from the verb *werden* and these can be present and past participles, infinitive verbs, nouns, and adjectives. Betweenness centrality measures how often a node is on the shortest path in a network. In the case of *werden*, this refers to how many elements are directly connected to this verb (participles, nouns, adjectives) and how frequently it appears in the shortest path between these elements and the other nodes of the sentences. Eigenvector centrality indicates a node's "influence" within a network and gives higher scores to those nodes that are central and are connected to other nodes with high degree centrality. In the case of *werden*, this measure indicates if the elements with which it is connected have also high out-degree values. The last centrality measure, eccentricity centrality, indicates which elements in a network are the most easily reachable by the others, thus, easy to establish connections with. For *werden*, this measure shows if *werden* displays a valency that allows connections with a large variety of elements in the same sentence.

7.2 11th Century

The following chart displays in- and out-degrees, the betweenness, eigenvector, and eccentricity centralities for each form of *werden* - auxiliary (AX werden), full verb (V werden), infinitive verb (IV werden), and past participle (PcPs werden) - found in the network of the 11th century.

Table 5: Centrality measures for *werden* in the 11th Century

	In-degree	Out-degree	Betweenness C.	Eigenvector C.	Eccentricity
AX werden	0	184	0.0	0.0	5.0
V werden	0	74	0.0	0.0	4.0
IV werden	4	25	521.5	0.014	5.0
PcPs werden	1	14	63.5	0.003	4.0

The chart displays some variation in the values obtained using the centrality measures for the four grammatical functions in which *werden* was used. The values obtained for the in-degree imply that *werden* as an auxiliary and *werden* used as a full verb can never get an incoming link,

since their function as the head of their respective verbal phrases. *Werden* in its infinitive and past participle forms get incoming links but the low values for this centrality measure indicate that the equally low frequency of the instances of *werden* in combination with modal verbs and in the perfect constructions. The greatest discrepancies among all the *werden* forms can be found for the betweenness centrality, for which *werden* as an auxiliary and used as a full verb have a score equal to zero. The eigenvector centrality shows significantly low values for all the forms, especially for *werden* used as an auxiliary and as a full verb. Eccentricity is the only centrality that displays similar results for all the forms of *werden*.

The table below shows the first ten nodes ranked for the in-and out-degrees (with numeric values) and centrality measures in the 11th century. The forms of *werden* are in bold.

Table 6: First 10 nodes of the 13th century ranked for in- and out-degrees, betweenness, eigenvector, and eccentricity centralities

	In-Degree		Out-Degree		Betweenness C.	Eigenvector C.	Eccentricity
1	AR der	52	**AX werden**	184	PR von	AR der	MV sollen
2	PR von	25	**V werden**	74	PR in	PP er	MV wollen
3	PP er	22	PR von	26	**IV werden**	PR von	MV mögen
4	PR in	19	PR in	19	PR mit	AJ mein	MV müssen
5	AJ mein	14	**IV werden**	14	**V werden**	AJ sein	**AX werden**
6	AD so	12	**PcPs werden**	14	N sund	DM der	**IV werden**
7	PP ich	11	PR mit	13	PcPs heilen	PP ich	AX sin
8	AJ sein	11	AX sin	10	N got	PR in	**V werden**
9	CJ unte	11	MV sollen	7	PR an	AJ dein	**PcPs werden**
10	PR mit	10	PT an	6	N ouge	N sund	PcPs heilen

The results obtained through the application of the aforementioned centrality measures give relevant information about nodes and edges in the network for the 11th century.

The in-degree column shows the elements in the network that have the highest number of incoming connections. According to this measure, the determinative article *der* is ranked first with 52 incoming links. Articles are used in combinations with nouns, therefore there is a high probability that their ranking for in-degree reflects the corresponding high frequency of nouns in this network. Other elements that rank high for this measure are prepositions and personal pronouns, and such ranking also reflects the frequency with which they are used in the network.

The out-degree column shows which elements in the networks have the highest number of outcoming links. Since these networks are built around the verb *werden*, it is not surprising to find all the forms of *werden* occupying the first positions for this centrality measure. Other elements with high out-degree values are some prepositions, the auxiliary verb *sein* (to be), and the modal verb (*sollen*).

The values for betweenness centrality indicate that the elements that often appear on the shortest path in the network for the 11th century are mostly prepositions. As mentioned in the theory section, vertices with high betweenness centrality could have substantial influence within a network by virtue of their control over information passing between other nodes (Matas, Martincic-Ipšic, Meštrovic, 2017). That fact that prepositions are among the elements with the highest betweenness centrality values is, from a network perspective, not surprising, since they specify the relationship between the verb and the nouns attached to them (Feigenbaum, 2002, p. 235). Further, their position between a verb and a prepositional phrase makes it possible for them to often appear on the shortest paths between the elements of the PP and the other nodes in the sentences. Both *werden* as an infinitive and *werden* used as a full verb are also highly ranked by this centrality measure, and such results could also be explained by the fact that their position (at the top or in the middle of a VP) in the sentence puts them often on the shortest paths in the network.

The column for the eigenvector centrality displays central nodes that are connected to other nodes with high degree centrality. The resemblances between the elements ranked in the first positions in this column and those ranked similarly by the in-degree centrality confirm the results obtained for this centrality measure and indicate that the nodes with the highest eigenvector centralities are often connected with those with high out-degree centrality values. For instance, the determinative article *der* occupies the first position for both in-degree and eigenvector centrality, indicating again that the elements with which it is connected, nouns, have high out-degree values.

The last centrality measure, eccentricity centrality, indicates which nodes are the most easily reachable by the others, thus, easy to establish connections with. According to this measure, modal verbs are the easiest to be reached, followed by all the four forms of *werden* and the auxiliary *sein*.

7.3 12th Century

The following chart displays in- and out-degrees, the betweenness, eigenvector, and eccentricity centralities for each form of *werden* - auxiliary (AX werden), full verb (V werden), infinitive verb (IV werden), and past participle (PcPs werden) - found in the network of the 12th century.

Table 7: Centrality measures for *werden* in the 12. Century

	In-degree	Out-degree	Betweenness C.	Eigenvector C.	Eccentricity
AX werden	0	189	0.0	0.0	4.0
V werden	0	69	0.0	0.0	3.0
IV werden	5	36	617.83	0.017	4.0
PcPs werden	1	15	72.66	0.003	4.0

As discussed for the previous century, both *werden* as an auxiliary and *werden* used as a full verb do not get any incoming link, since they always function as the head of the verbal phrase. Hence, the in-degrees for these two forms are equal to zero. *Werden* in its infinitive and past participle forms get incoming links but these numbers have not increased much in this century in comparison to the previous one. The scores of the other centrality measures, out-degree (outgoing links), betweenness (nodes that often appear on the shortest path), eigenvector (central nodes connected to other nodes with high degree centrality), and eccentricity (nodes that are the easiest to reach by other nodes) centralities displays also comparable results to the 11th century, giving an overall impression that the status of *werden* in the network (and the related syntactic structure) for this century did not change much from the previous one.

The table below shows the first ten nodes ranked for the in-and out-degree (with numeric values) and centrality measures in the 12th century. The forms of *werden* are in bold.

Table 8: First 10 nodes of the 12th century ranked for in- and out-degrees, betweenness, eigenvector, and eccentricity centralities

	In-Degree		Out-Degree		Betweenness C.	Eigenvector C.	Eccentricity
1	AR der	70	**AX werden**	189	PR in	AR der	AX sein
2	AJ sein	24	**V werden**	69	PR von	AJ sein	MV müssen
3	PR von	23	**IV werden**	36	**IV werden**	PP er	MV mögen
4	PR in	18	MV sollen	27	PR zu	N got	MV wollen
5	AD nicht	17	PR in	23	PR mit	N welt	PcPs legen
6	PP er	13	PR von	20	PR an	PR in	PcPs kehren
7	AR ein	12	**PcPs werden**	15	N kind	N erde	PcPs erteilen
8	N got	11	PR mit	13	**V werden**	PR zu	PcPs wissen
9	PR mit	11	PR zu	11	N sund	PR von	PcPs herren
10	AJ dein	11	AX sin	10	N got	AJ all	PcPs laden

The results for the 12th century for the five centrality measures do not differ much from the results obtained in the previous one. For instance, both in-degree (incoming links) and eigenvector (finds nodes that are connected to other nodes with high degree centrality) centralities rank again the determinative article *der* as the element with the highest values, implying that in this network as well, nouns are frequently used. Prepositions and personal pronouns are also highly ranked, confirming the trends observed in the previous century.

The out-degree (outcoming links) ranks again all the forms of *werden* among the nodes with the majority of the outcoming links, and such an outcome is related to the fact that the networks in this work focus on the verb *werden*.

Betweenness centrality (gives high values to the nodes that often appear on the shortest paths) also shows comparable scores to the 11th century, ranking prepositions and *werden* in its infinitive and past participle forms among the nodes with the highest values.

The last centrality measure, eccentricity (finds the nodes that are the most easily reachable by the others), is the only one that shows some differences when compared with the results for the 11th century: while modal verbs retain their rank, none of the forms of *werden* got assigned high scores.

7.4 13th Century

The following chart displays in- and out-degrees, the betweenness, eigenvector, and eccentricity centralities for each form of *werden* - auxiliary (AX werden), full verb (V werden), infinitive verb (IV werden), and past participle (PcPs werden) - found in the network of the 13th century.

Table 9: Centrality measures for *werden* in the 13th Century

	In-degree	Out-degree	Betweenness C.	Eigenvector C.	Eccentricity C.
AX werden	0	323	0.0	0.0	4.0
V werden	0	174	0.0	0.0	3.0
IV werden	5	45	1322	0.011	4.0
PcPs werden	1	9	163	0.002	7.0

The results obtained in this century do not diverge much from those of the previous two centuries. The in-degree shows again no incoming links for *werden* as an auxiliary and *werden* used as a full verb, and a low number of those links for both *werden* in the infinitive and past participle forms. The out-degree shows some increase in comparison to the 11th and 12th centuries, but such a change is related to the larger number of texts available for this century (as also shown by the normalized frequency of the attestation of *werden* displayed in figure 15, section 5.5.1). The scores for the other centrality measures, out-degree, betweenness, eigenvector, and eccentricity centralities displays similar results to the two previous centuries, and such an outcome suggests that the status of *werden* in this particular network did not change much from the prior two centuries.

The table below shows the first ten nodes ranked for the in-and out-degree (with numeric values) and centrality measures in the 13th century. The forms of *werden* are in bold.

Table 10: First 10 nodes of the 13th century ranked for in- and out-degrees, betweenness, eigenvector, and eccentricity centralities

	In-Degree		Out-Degree		Betweenness C.	Eigenvector C.	Eccentricity
1	AR der	98	**AX werden**	323	PR von	AR der	PcPs manen
2	PR von	44	**V werden**	174	PR mit	AR ein	PcPs führen
3	PP er	31	PR von	49	PR zu	PP er	PcPs legen
4	PR mit	27	**IV werden**	45	N venustianus	AJ sein	PcPs kaufen
5	AR ein	23	PR mit	34	N hand	N got	PcPs stehlen
6	AJ sein	21	PR zu	30	PR in	PR von	PR in
7	AJ sanctu	19	PR in	20	N marter	AJ sanctu	AX sein
8	PR an	18	PR an	19	N betehus	PR mit	N hand
9	PR zu	16	PcPs bringen	13	**IV werden**	N richter	N marter
10	CJ unte	16	PR durch	12	PR an	N minne	N betehus

The table displays similar results to those obtained in the two previous centuries. The in-degree and eigenvector centralities rank again the determinative article as the node with the highest scores, together with prepositions and some adjectives.

The out-degree ranks all the forms of *werden* highly, and the same can be said for the prepositions. The only form of *werden* that makes it among the nodes with the highest values for betweenness centrality is *werden* in its infinitive form.

The last centrality measure shows some differences in comparison to the 11th and 12th centuries since no modal verbs and none of the forms of *werden* are ranked among the 10 nodes with the highest scores.

7.5 14th century (First Half)

The following chart displays in- and out-degrees, the betweenness, eigenvector, and eccentricity centralities for each form of *werden* - auxiliary (AX werden), full verb (V werden), infinitive verb (IV werden), and past participle (PcPs werden) - found in the network of the early 14th century.

Table 11: Centrality measures for *werden* in the 14th Century (first half)

	In-degree	Out-degree	Betweenness C.	Eigenvector C.	Eccentricity
AX werden	0	275	0.0	0.0	4.0
V werden	0	192	0.0	0.0	4.0
IV werden	6	75	1425.73	0.013	4.0
PcPs werden	2	15	108.0	0.004	5.0

The scores obtained for the first half of the 14th century do not differ much from those of the prior centuries. The in-degree values are not increasing, and the out-degree scores are also stable without any significant growth. The same trends can be observed for the other centrality measures, which show comparable results to the 11th and 12th centuries. This outcome suggests again that the status of *werden* in the network did not undergo significant changes during this century.

The table below shows the first ten nodes ranked for the in-and out-degree (with numeric values) and centrality measures in the early 14th century. The forms of *werden* are in bold.

Table 12: First 10 nodes of the early 14th century ranked for in- and out-degrees, betweenness, eigenvector, and eccentricity centralities

	In-Degree		Out-Degree		Betweenness C.	Eigenvector C.	Eccentricity
1	AR der	143	**AX werden**	274	PR von	AR der	MV sollen
2	PR von	36	**V werden**	192	**IV werden**	N wind	MV mögen
3	PP er	29	**IV werden**	75	PR in	PP er	AX sein
4	PR in	23	PR von	36	PR mit	PR von	MV wollen
5	AD nicht	17	PR in	32	PR an	PP du	MV müssen
6	AJ sein	15	AX sein	19	PP er	CJ und	MV können
7	CJ und	15	**PcPs werden**	15	AR der	N got	AX machen
8	PR mit	13	MV sollen	15	PR zu	PR in	**PcPs werden**
9	AD auch	13	MV mögen	15	N grad	N rat	**V werden**
10	AJ dies	13	PR mit	14	PcPs geben	AJ sein	**AX werden**

Table (12) displays a similar scenario to the ones observed in the previous three centuries. The determinative article *der*, together with prepositions and pronouns are among the nodes with the highest scores for both the in-degree and eigenvector centralities. All the forms of *werden*

continue to have the highest number of outgoing links, and such a trend can be related again to the fact that the networks are built around this verb.

The verb *werden* in its infinitive form is again the only one ranked highly in the betweenness centrality column, while *werden* as an auxiliary, used as a full verb, and in its past participle form get ranked among the 10 nodes with the highest scores according to the eccentricity centrality.

7.6 Discussion

7.6.1 Out-degree

The next figures show the results obtained for the out-degree centrality for all the 4 forms of *werden*:

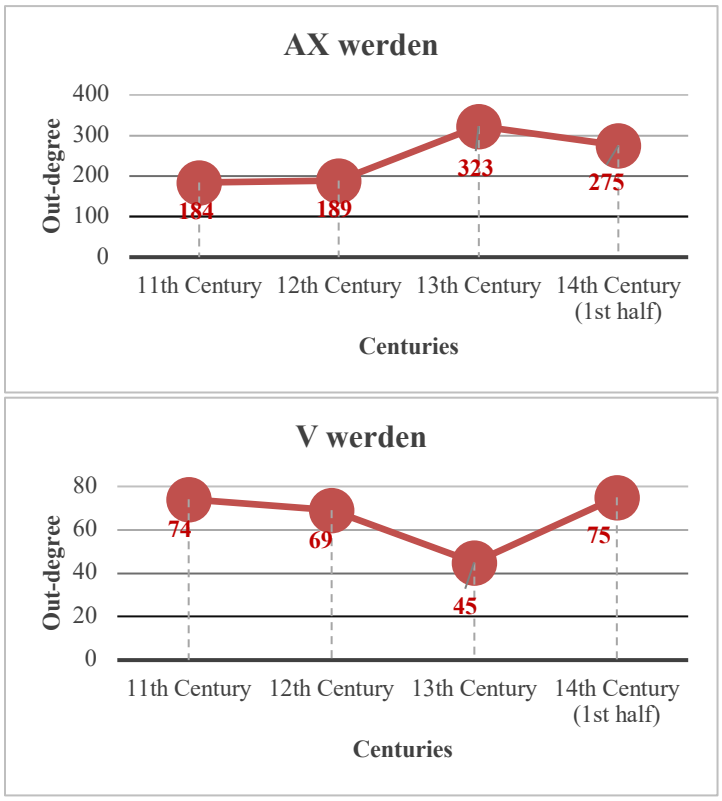


Figure 64: Out-degree for *werden* as an auxiliary and as a full verb

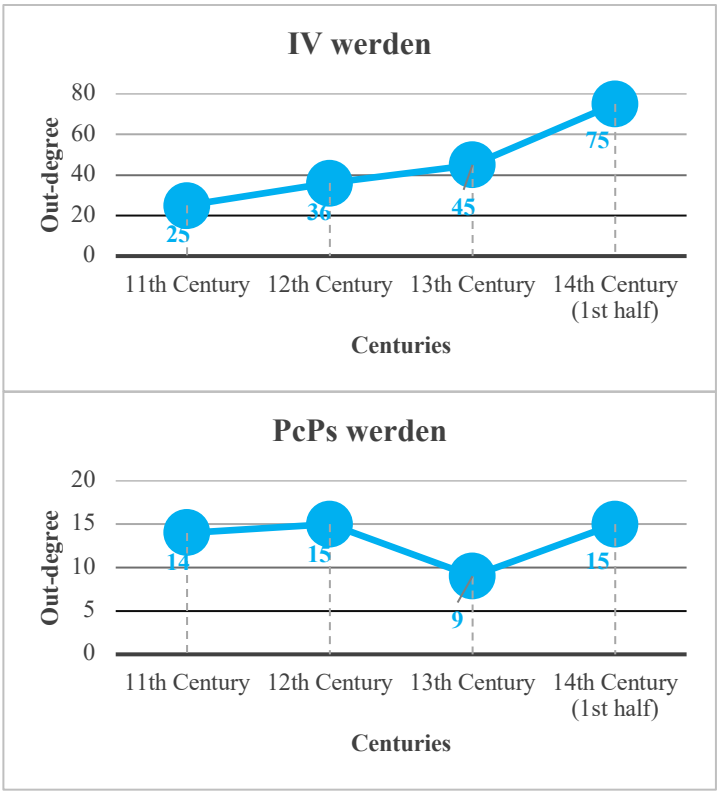


Figure 65: Out-degree for *werden* in its infinitive and past participle forms

The out-degree centrality counts the number of outgoing links from a given node. Among all the measurements used in the network related part of this analysis, this centrality is the one that better reflects the data collected in the historical linguistic analysis part of this dissertation. As discussed in chapter 5, in Middle High German, *werden* when used as an auxiliary had the highest number of instances, and this is visible in the results obtained in this section, as shown in figure (63). The same can be said about the use of *werden* as a full verb that appeared in combination with nouns and adjectives. The lower out-degree values of both *werden* in the infinitive form and in the past participle indicates, for the first one, the relatively low number of instances found of *werden* in combination with modal verbs. For the second one, the low out-degree finds its explanation in the relatively young status of the present and past perfect constructions in Middle High German (Kuroda, 1999; Concu, 2016).

7.6.2 Betweenness Centrality

The next figure shows the results obtained for the out-degree centrality for *werden* in its infinitive and past participle forms, since *werden* as an auxiliary and as a full verb got scores equal to zero throughout the entire Middle High German period.

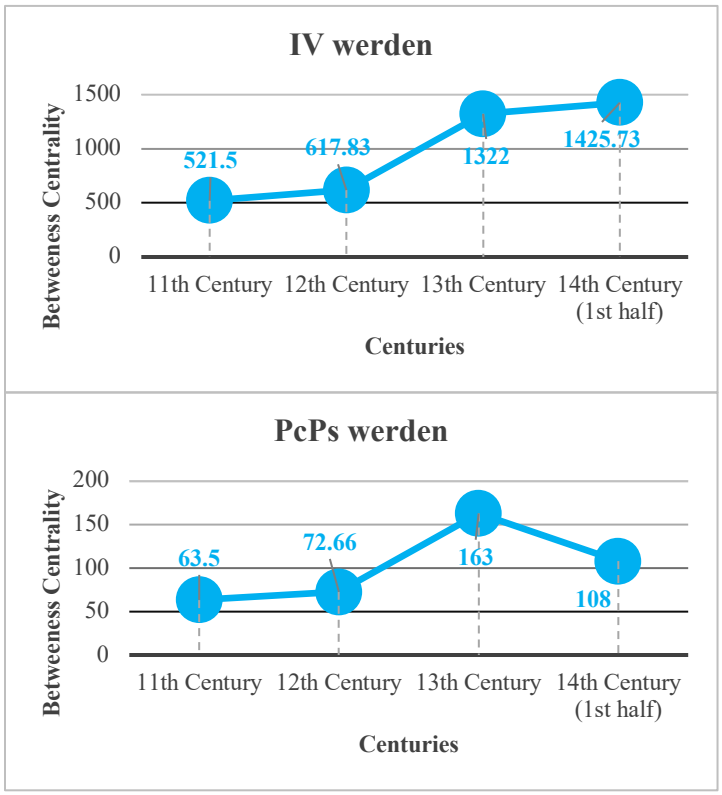


Figure 66: Betweenness centrality for *werden* in its infinitive and past participle forms

Betweenness centrality displayed significantly different scores for all the forms of *werden*. Both *werden* as an auxiliary and *werden* when used as a full verb got scores equal to zero throughout the Middle High German period. This means that *werden* in these two forms never served as the shortest path among the other elements in the sentences in these four centuries. Opposite to what happened to *werden* as an auxiliary and *werden* as a full verb, both *werden* used in the infinitive form and *werden* used as a past participle increase their betweenness centrality values in the 13th century. The increment for the infinitive, which continues also in the first half of the 14th century, can be explained by looking at the betweenness centrality scores for this form throughout the four centuries. *Werden* as an infinitive remains quite stable among the 10th highest

in Middle High German, and while its value grows, so do also the values of the other elements that also ranked high. This means that there are no substantial changes for *werden* when combined with modal verbs in these centuries since it continuously remains among the elements that serve as the shortest paths among the other elements in the sentences. *Werden*, in its past participle form, never gets ranked among the 10th highest nodes for this centrality measure but the data shows an increase that starts also in the 13th century.

7.6.3 Eigenvector Centrality

The next figure shows the results obtained for the out-degree centrality for *werden* in its infinitive and past participle forms since *werden* as an auxiliary and as a full verb got scores equal to zero throughout the entire Middle High German period.

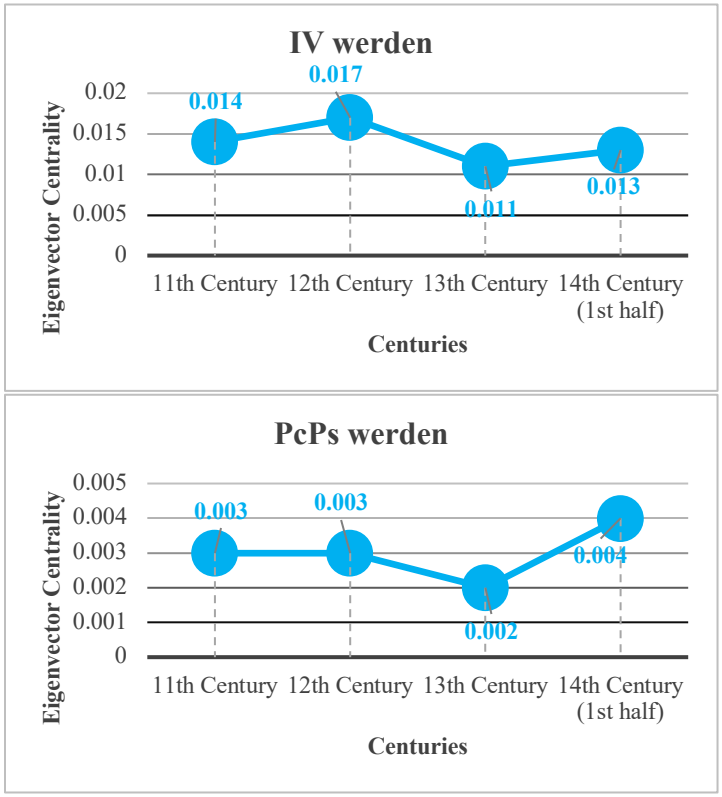


Figure 67: Eigenvector centrality for *werden* in its infinitive and past participle forms

Eigenvector centrality discriminates between the links in a network and assigns high scores to those nodes connected with other nodes with high degree centrality. The data shows that the

elements with the highest values are articles, nouns, pronouns, and prepositions. Further, the data indicate that the nodes with which these elements are linked are the most influential. Because of the nature of the networks in which every sentence contains a form of *werden*, one can assume that, indirectly, the influence of *werden* is also increasing. The articles connect with nouns that can function as subjects of a form of *werden* or a complement of a verb which is used in combination with *werden* as an auxiliary. The personal pronouns can do the same and the prepositions can be connected to verbs that come in combination with *werden* as an auxiliary. As the networks do not capture the functions of nouns or pronouns (subjects or objects of a verb), it is not possible to exact capture which connections are the most significant for the eigenvector centrality. Additionally, the scores for all the forms of *werden* remain constantly low with no drastic changes throughout the centuries, meaning that the influence of the elements to which they are connected is also not changing. The results of this centrality measure could be read as an indirect indication of the influence of the forms of *werden* in Middle High German.

7.6.4 Eccentricity Centrality

The next figures show the results obtained for the eccentricity centrality for all the 4 forms of *werden*:

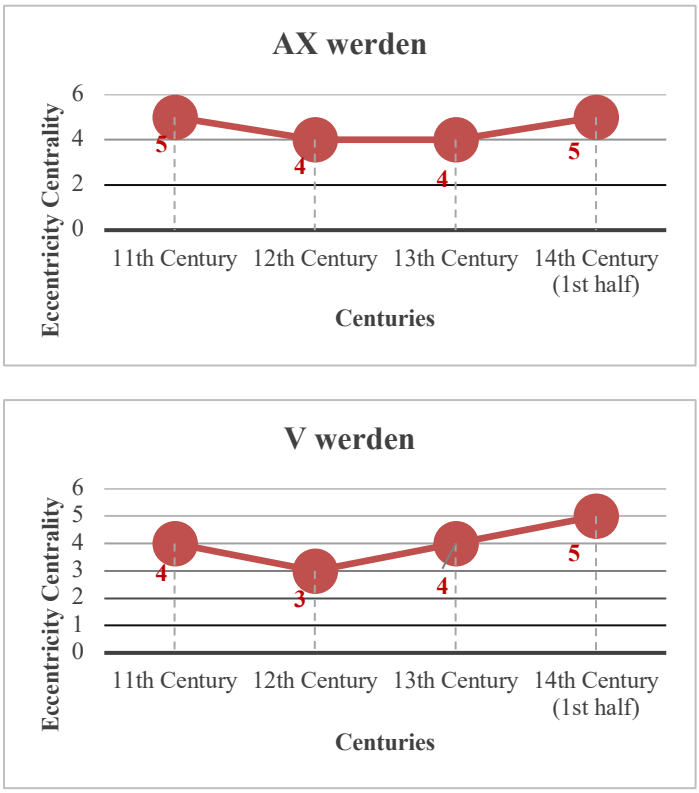


Figure 68: Out-degree for *werden* as an auxiliary and as a full verb

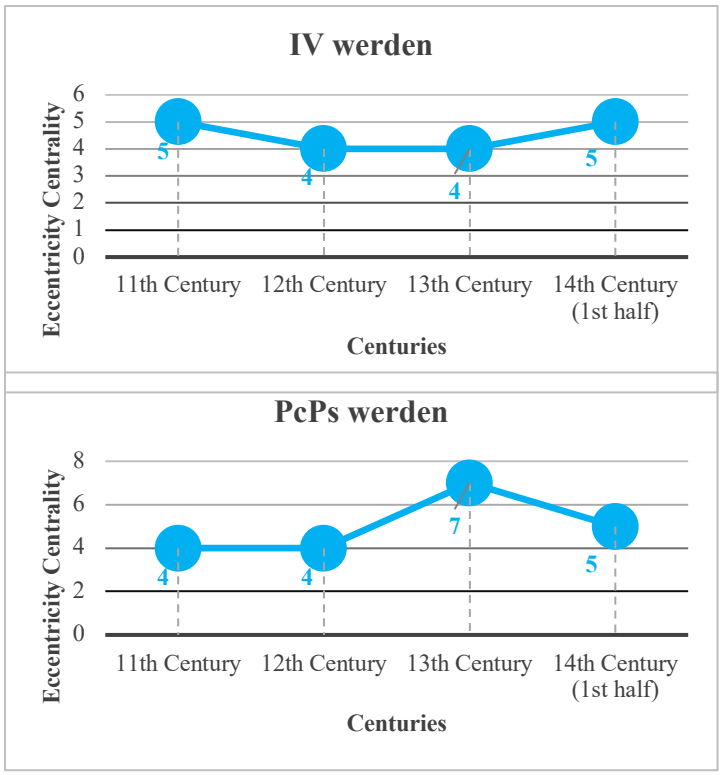


Figure 69: Out-degree for *werden* in its infinitive and past participle forms

This measure ranked all the forms of *werden* among the 10 nodes with the highest scores only in the 11th century. In the first half of the 14 centuries, this happens to *werden* as an auxiliary, *werden* as a full verb, and *werden* used as a past participle, while *werden* in its infinitive form does not rank as highly as the other forms. These results indicate that *werden* was one of the easiest elements to reach by other connections in the network in the first and last centuries of the Middle High German period. In the other two centuries (the 12th and the 13th), this centrality ranks high nodes such as modal verbs, the auxiliary *sein*, some past participles, and nouns. One of the aspects that this centrality shows is that a node with a high out-degree is not always the easiest to be reached by the other nodes in the network. Because all the networks created for this research are made with sentences that contain at least one form of *werden*, this node has always the highest scores for the out-degree. Being the most common verb in the network, one would assume that is

also the easiest node with which other nodes can establish connections. The results obtained with this centrality point out that this is not always the case, especially for the 12th and 13th centuries.

7.7 Summary

The analysis of the verb *werden* in Middle High German and the application of the centrality measures give some relevant information about the status of the verb *werden* in these centuries. The out-degree centrality confirms the results obtained in the historical linguistic analysis carried out in chapter 5: the instances of *werden* slowly increased and such an increase corresponds to the growth of the number of the outgoing connections that *werden* could establish in Middle High German. Betweenness centrality assigned scores equal to zero to *werden* as an auxiliary and used as a full verb, meaning that these forms never appeared on of the shortest paths between nodes in the network's sentences. The opposite was observed for both *werden* used in the infinitive form and as a past participle since their scores slowly rise. As previously discussed, the results obtained through this centrality measure could be related to the position that *werden* occupied in the sentence when combined with modal verbs and in the perfect constructions. In these cases, *werden* is not the head of the verbal phrase but stays in-between the head and the other components of the VP, increasing its chance to appear on the shortest path among the nodes in the sentences.

The eigenvector centrality did not assign any high scores to any of the forms of *werden* but to elements such as articles, pronouns, and adjectives. This can indicate that the nodes with which those elements are connected are the most important in the networks. Because the networks were built around the verb *werden*, it can be assumed that all the forms of this verb that are directly or indirectly connected to the elements with the highest eigenvector centrality values are the most influential in the networks.

The results from eccentricity centrality showed the *werden* was one of the easiest nodes to be reached in the network in the first and last centuries of the Middle High German period, although its out-degree scores have constantly been the highest one of all the other nodes in the networks. As previously discussed, this centrality measure could be directly related to the valency property of werden, which reflects its capacity to be used with different grammatical elements in the same sentence. The historical analysis (chapter 5) shows an increase in the elements with which *werden* could be combined in the first half of the 14th century. The high ranking for *werden* in the

same time frame could be connected to this increase and the eccentricity centrality was able to capture such change.

In the next section, I will provide the network-related analysis of the instances of *werden* in Early New High German.

7.8 Early New High German

7.8.1 Analysis

In this section, I will describe the features of the networks created for the Early New High German period (second half of the 14th century- 17th century), concentrating the analysis first on the number of nodes and edges, the cluster coefficient, and the average path length for each of the four networks. The next figure displays the number of nodes and edges:

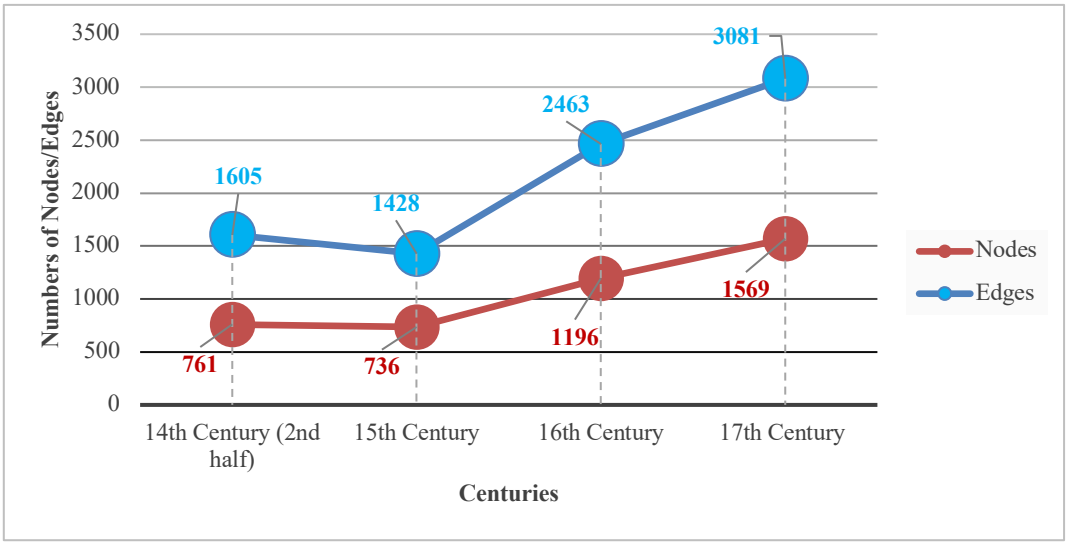


Figure 70: Number of nodes and edges of the Early New High German networks

Figure 70 shows the gradual growth of the number of both nodes and edges from the second half of the 14th to the 17 centuries. After a brief decline in the 15th century, the numbers in the last two centuries show a positive correlation between the increase of the number of nodes and the expansion of the edges.

As previously discussed, one of the common features of complex networks is related to the so-called small-world properties. These networks have a high cluster coefficient (the nodes are

well connected to each other and the neighbors of a node are also linked together) and short average path length (only a few jumps in the network are required to move from one node to any given other node). In order to find out if the Early New High German networks also exhibit these features, four random networks[15] with the same degree have been created using the network randomizer application in Cytoscape. The results for both cluster coefficient and average path length and are shown in the following figures:

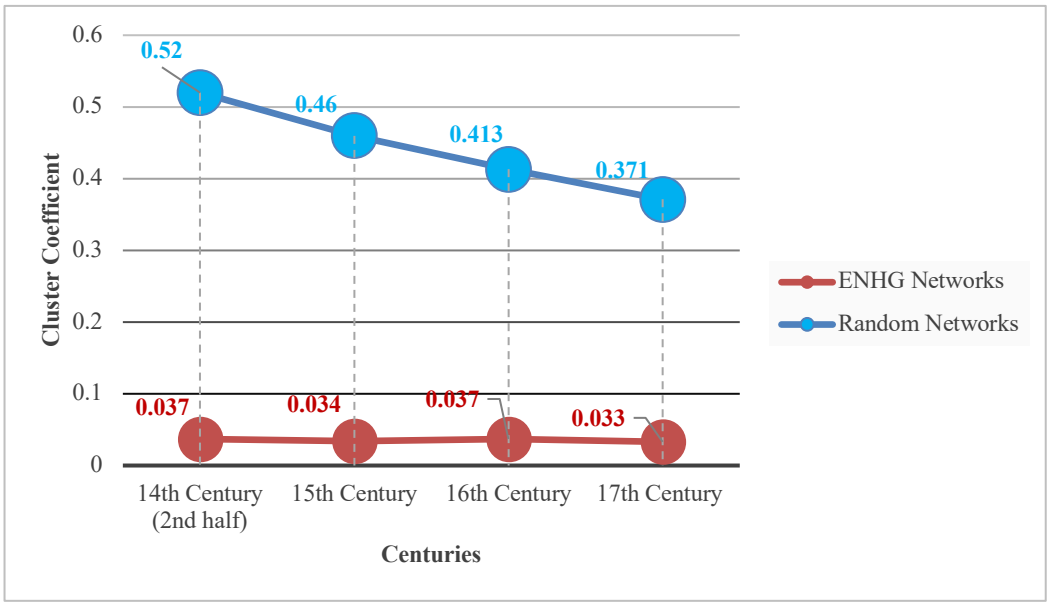


Figure 71: Cluster Coefficient in the Early New High German networks and in their random counterparts

[15] In this type of networks, links between nodes are established randomly, without any rules governing such connections.

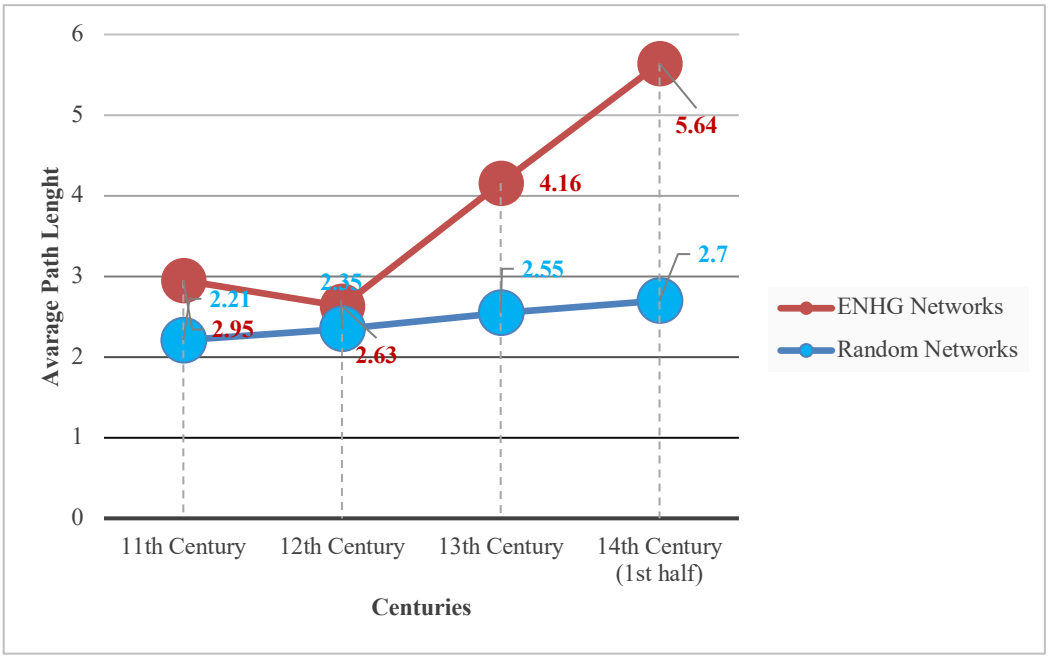


Figure 72: Average Path Length in the Early New High German networks and in their random counterparts

The results obtained here are contrasting compared to that of the Middle High German networks. While in Middle High German the networks show only in part small-world properties (lower cluster coefficients than the random networks but similar average path length), the Early New High German ones do not possess any small-world property at all. All four networks have, in fact, a lower cluster coefficient than their random counterparts, meaning that the nodes in the Early New High German networks do not tend to cluster together. The average path length also increases its value starting in the 16th century, meaning that the average number of steps along the shortest paths more than double in the last two centuries. On the contrary, the same parameter stays almost unchanged in the random networks throughout the four centuries.

As done for the networks in Middle High German, Cytoscape was used to fit a power law of the form of $y=axb$ in which y is the estimated degree exponent. The scatter plot is able to identify the presence of a long tail in the degree distribution, which is characteristic of scale-free networks. The results for the out-degree in Early New High German are shown in the next figures, in which the x-axis represents the out-degree, the y-axis signifies the number of nodes, and the red line is the fitted power law.

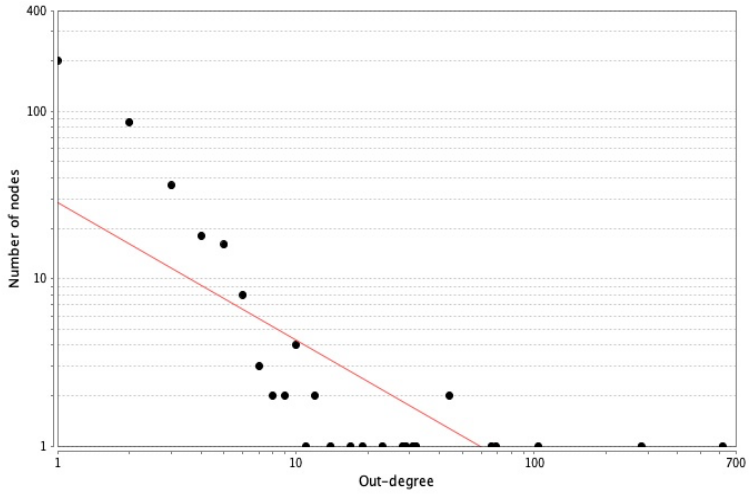


Figure 73: Out-degree 14th Century (second half)

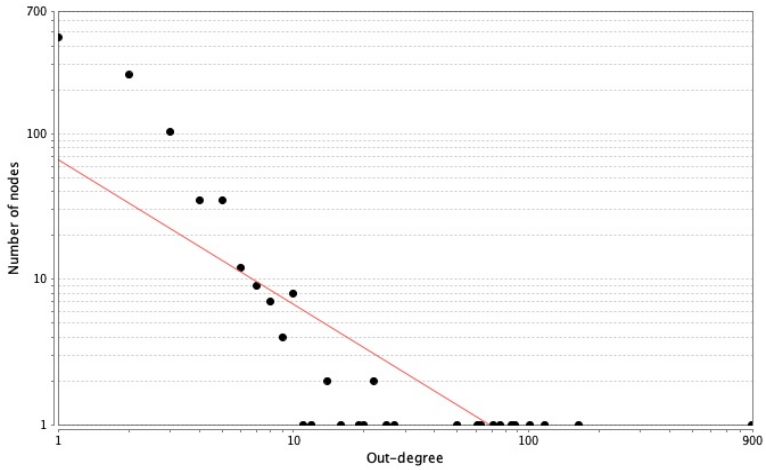


Figure 74: Out-degree 15th Century

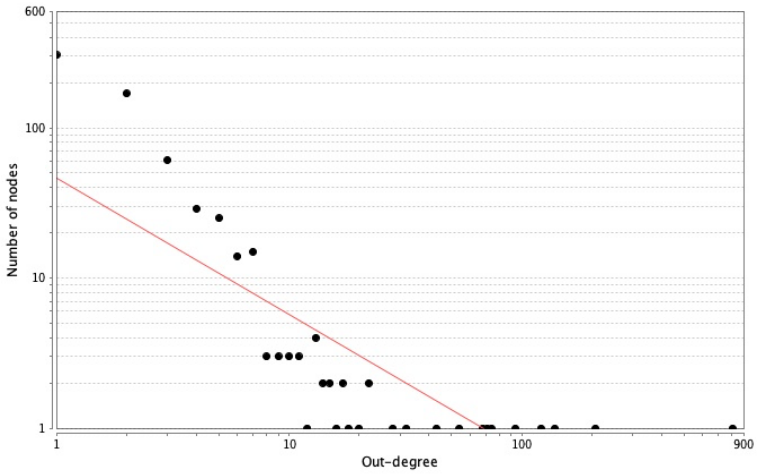


Figure 75: Out-degree 16th Century

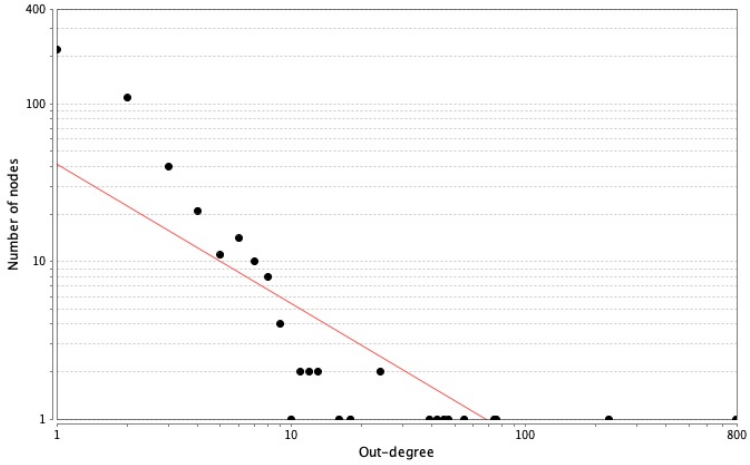


Figure 76: Out-degree 17th Century

The scatter plots of the Ealy New High German networks reveal here too the presence of a distribution in which only a few nodes have high out-degree values and those are visible on the low right corner of the x-axis. Moving from the right to the left of the x-axis, the number of nodes that have lower out-degree values increases, until it reaches the highest part of the left corner of the y-axis. Although the presence of a distribution that follows a power law must be proven using advanced statistics, by observing the scatter plot it can be said that these networks are probably scale-free or, at least, that their out-degree distribution shows patterns that are similar to those found in scale-free networks.

The next figures show the raw and normalized frequencies of the types of interactions (verbal phrase; nominal phrase, and prepositional phrase) in the networks.

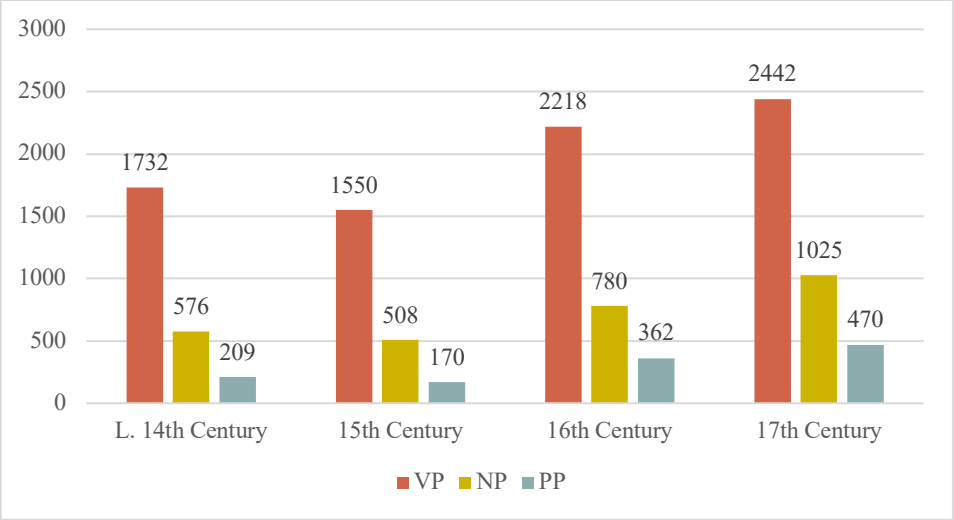


Figure 77: Raw frequency per century of the interactions between nodes

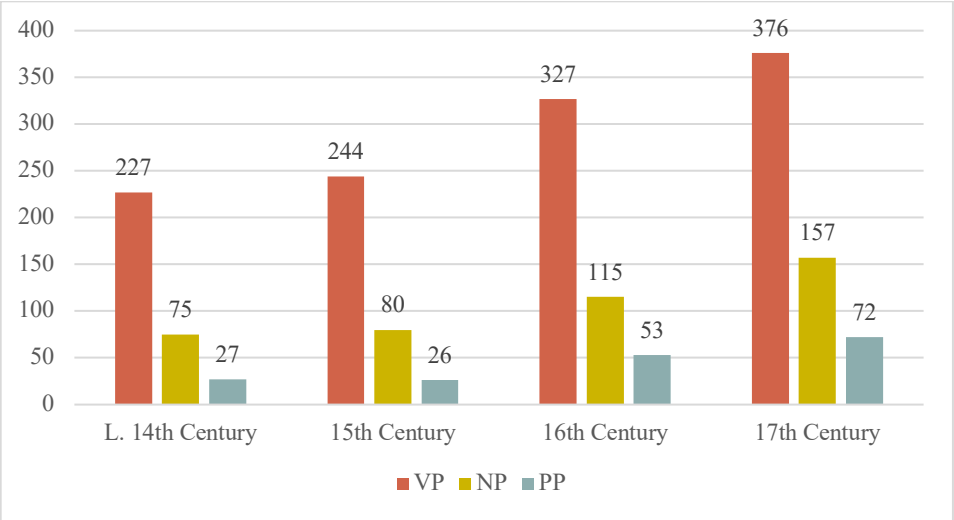


Figure 78: Normalized frequency per century of the interactions between nodes

The figures show the distribution of the types of interactions in the Early New High German corpus. The first figure displays an increase in nominal and prepositional phrases, especially in the last centuries.

7.8.2 Discussion

The analysis showed that none of the Early New High German networks possessed small-world properties. As observed for Middle High German, the cluster coefficients in the real

syntactic networks were lower than the coefficients in the random ones. The networks were built using the same set of rules used for Middle High German, which means that the connections between nodes were established through asymmetrical relationships and that not all the nodes could be linked to each other in an equal manner. Hence, both the Middle and Early New High German networks had lower clustering coefficients than their random counterparts. However, the Early New High German networks showed some different results in regard to the average path lengths. The comparison between real and random networks, in fact, showed that the values for this parameter in the real networks were, especially in the 16th and 17th centuries, significantly higher than those in their random counterparts. As discussed for Middle High German, these results may also be an indicator of the syntactic inventory of German at this stage. In the Early New High German period, the grammaticalization process of the present and past perfect reached its final stage. One of the most common effects of grammaticalization is the increase in the frequency of that particular construction. Kuroda (1999) reports an increase in the present and past perfect exactly in Early New High German and such a frequency increase is probably reflected in the higher average path length. Further, this can be also related to the rise of the *werden* future periphrastic constructions in which speakers started adding an additional *werden* to express future events as a result of the semantic bleaching through which this verb because of its grammaticalization (or desemantization). As observed for the present and past perfect, the average path length would be higher than in the constructions found with only one form of *werden*. Consider the following sentences from the corpus:

(30) sô **werde** ouch uns **vergeben**
so becomes-AUX.PRS also to us forgive-INF
'and to us they will forgive'
(*Alkuins Traktat*, 48r,3)

(85) An jenem tag **wird** es **offenbar** **werden**
on a day becomes-AUX.PRS it obvious-ADJ become-COP.INF
'One day it will become obvious'
(*Summaria*, 29, 34)

The next figures display the difference between these two sentences in terms of average path length:

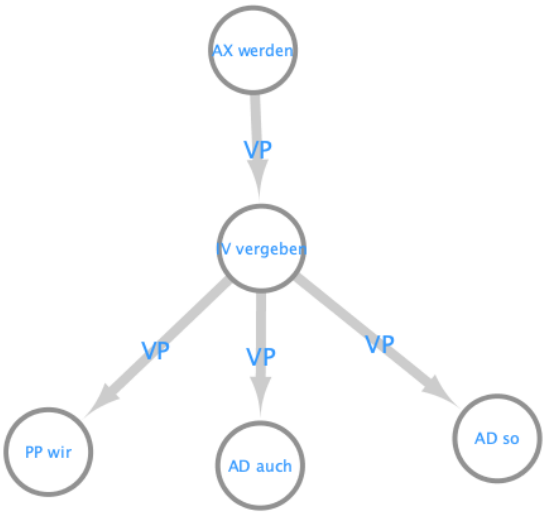


Figure 79: The syntactic network for “sô werde ouch uns vergeben”

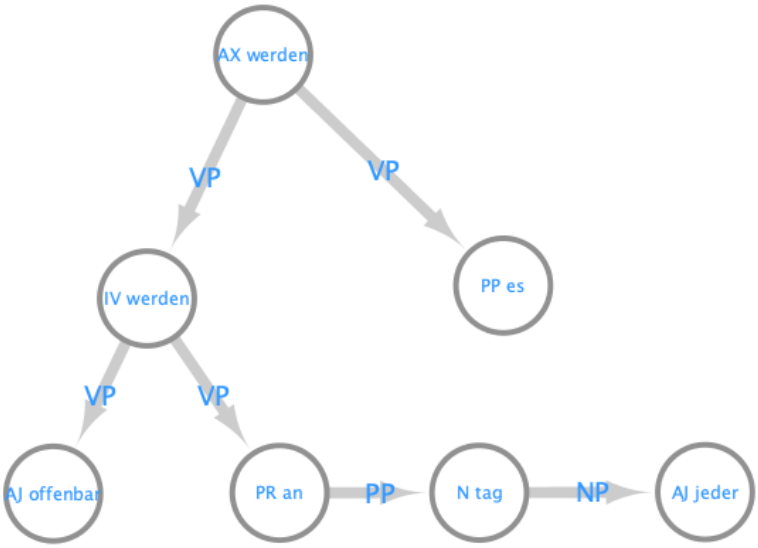


Figure 80: The syntactic network for “an jenem tag wird es offenbar werden”

The average path length in the sentence with only a form of *werden* is 1,429. The average path length in the sentence in which there are two forms of *werden* is 1,846. The appearance of a second form of *werden* in communicative contexts that in Middle High German only required a form of *werden* may also be related to the increased path length.

It has previously been discussed how small-world properties are one of the main features of complex networks (Jiang et al., 2019) and how these properties have been found in syntactic networks (Corominas & Murtra, 2009; Jiang et al., 2019). Because of the low cluster coefficients and the high average path lengths, the Early New High German networks cannot be considered small-world networks. Further, the analysis of the frequency of the types of interactions showed an increase in the nominal and prepositional phrases in the 16th and 17th centuries. Such a change could also be one of the causes behind the increment of the path lengths in these two last centuries.

The Early New High German networks, although they cannot be considered small-world, have an out-degree distribution that shows some similarity with the degree distribution that is normally found in scale-free networks. As argued for the Middle High German networks, these results can be seen as a reflection of the networks' structure, since the capacity of the nodes to establish a connection with another node depends on their grammatical function and are, therefore, not equal.

In the next section, I will analyze the networks for each century. As done for Middle High German, here too I will address the results for all the four forms of *werden* (auxiliary, full verb, infinitive, and past participle forms) obtained through the following centrality measures: in- and out-degrees (incoming and outgoing connections), betweenness (nodes frequently on the shortest path between nodes), eigenvector (central nodes connected to nodes with high degree centrality), and eccentricity (nodes that are the easiest to reach) centralities.

7.9 14th Century (Second Half)

The following chart displays in- and out-degrees, the betweenness, eigenvector, and eccentricity centralities for each form of *werden* - auxiliary (AX werden), full verb (V werden), infinitive verb (IV werden), and past participle (PcPs werden) - found in the network of the second half of the 14th century.

Table 13: Centrality measures for *werden* in the 14th Century (Second Half)

	In-degree	Out-degree	Betweenness C.	Eigenvector C.	Eccentricity
AX werden	0	329	0.0	0.0	4.0
V werden	0	123	0.0	0.0	4.0
IV werden	6	64	1606	0.012	5.0
PcPs werden	2	38	241.83	0.002	5.0

The results showed in this chart display a similar outcome to the one observed for the first half of the 14th century. Since they function as the head of the verbal phrases, *werden* used as an auxiliary and a full verb are never the target nodes of any incoming links. On the contrary, *werden* in its infinitive and past participle forms do receive incoming links when are used with modal verbs and in the perfect constructions. In comparison to the previous centuries, there is an increase of the outgoing links for the past participle of *werden*, which probably reflects the ongoing grammaticalization of the perfect constructions and the related frequency increase of these periphrases. The other centrality measures do not show any significant variations in the scores when compared with the same scores from the prior centuries, indicating that the status of *werden* did not change much in terms of its significance in the network.

The table below shows the first ten nodes ranked for the in-and out-degree (with numeric values) and centrality measures in the late 14th century. The forms of *werden* are in bold.

Table 14: First 10 nodes of the early 14th century ranked for in- and out-degree, betweenness, eigenvector, and eccentricity centrality

	In-Degree		Out-Degree		Betweenness C.	Eigenvector C.	Eccentricity
1	AR der	170	**AX werden**	329	PR von	AR der	PR um
2	PR von	36	**V werden**	123	PR mit	N wort	N schuld
3	PR in	33	**IV werden**	64	AR der	AJ all	PR auf
4	PR mit	31	PR von	43	N got	AJ dies	PR durch
5	CJ und	20	**PcPs werden**	38	N wort	PP er	PR von
6	PP er	20	PR mit	34	PR in	N got	PR nach
7	AJ all	19	PR in	33	N mensch	AJ mein	PcPs gehen
8	AJ mein	18	PR zu	21	**IV werden**	CJ und	PcPs rufen
9	AD nicht	18	MV sollen	20	PR zu	PP ich	PcPs dienen
10	PP ich	16	PR an	14	PR an	AR ein	PR anstatt

The scores in Table 14 display a similar situation to the ones observed in Middle High German. The determinative article *der*, prepositions, and pronouns are again among the nodes with the highest scores for both the in-degree and eigenvector centralities. All the four forms of *werden* (auxiliary, full verbs, infinitive and past participles) remain in the highest-ranking positions for the number of outgoing links.

The verb *werden* in its infinitive form keeps its high betweenness centrality scores, while none of the forms of *werden* get assigned high scores by the eccentricity centrality.

7.10 15th Century

The following chart displays in- and out-degrees, the betweenness, eigenvector, and eccentricity centralities for each form of *werden* (auxiliary, full verb, infinitive verb, past participle) found in the network of the 15th century.

Table 15: Centrality measures for *werden* in the 15th Century

	In-degree	Out-degree	Betweenness C.	Eigenvector C.	Eccentricity
AX werden	0	301	0.0	0.0	5.0
V werden	0	147	0.0	0.0	5.0
IV werden	6	56	1291	0.014	7.0
PcPs werden	1	52	100.2	0.002	5.0

The scores displayed in Table 15 do not differ much from those of the second half of the 14th century, except for the higher out-degree values for *werden* used in its past participle forms. The growth of the out-going links is related to the frequency increase of this form in the perfect constructions. As discussed in chapter 6, after an apparent incompatibility in Middle High German, in Early New High German, the number of instances of the passive in the present perfect tense started to be found with more frequency in the texts. Such an outcome could be connected to the further "paradigmatization" of the passive constructions and the ongoing grammaticalization of the present perfect reported in the early centuries of the Early New High German period. Betweenness, eigenvector, and eccentricity centralities do not display any relevant alteration when compared to the previous centuries. This indicates that the status of *werden* in the network remains mostly unchanged.

The table below shows the first ten nodes ranked for the in-and out-degree (with numeric values) and centrality measures in the 15th century. The forms of *werden* are in bold.

Table 16: First 10 nodes of the 15h century ranked for in- and out-degree, betweenness, eigenvector, and eccentricity centrality

	In-Degree		Out-Degree		Betweenness C.	Eigenvector C.	Eccentricity
1	AR der	122	**AX werden**	301	PR von	AR der	MV sollen
2	CJ und	44	**V werden**	147	PR in	AR ein	MV wollen
3	PR von	33	**IV werden**	56	PR zu	CJ und	MV mögen
4	AR ein	28	**PcPs werden**	52	**IV werden**	N got	MV müssen
5	PR in	22	PR von	37	N könig	PR von	MV können
6	AD nicht	21	AX sein	30	PR mit	PP er	**IV werden**
7	PR zu	19	PR in	29	N land	PR zu	PcPs halten
8	PR mit	16	PR zu	24	N mensch	N tag	PcPs setzten
9	PP er	15	MV sollen	17	PP er	N könig	PcPs führen
10	AJ sein	15	MV mögen	16	N tag	N mensch	PcPs teilen

The values showed in the table confirm the trends observed both in Middle High German and in the second half of the 14th century. Articles, pronouns, and nouns keep their high ranking for both in-degree and eigenvector centralities, whereas all forms of *werden* continue to occupy the first positions in the out-degree column. *Werden* in its infinitive form is in this century as well the only form that is highly ranked for betweenness centrality. Further, it appears also in the eccentricity centrality column, indicating that, in this period, it was one of the nodes that were the easiest to reach by other nodes. This outcome is probably related to the fact that modal verbs are among the highest ranked nodes according to this centrality, positively influencing the rank for *werden* in its infinitive form as well.

7.11 16th Century

The following chart displays in- and out-degrees, the betweenness, eigenvector, and eccentricity centralities for each form of *werden* (auxiliary, full verb, infinitive verb, past participle) found in the network of the 16th century.

Table 17: Centrality measures for *werden* in the 16th Century

	In-degree	Out-degree	Betweenness C.	Eigenvector C.	Eccentricity
AX werden	0	394	0.0	0.0	5.0
V werden	0	135	0.0	0.0	4.0
IV werden	6	74	4653	0.017	5.0
PcPs werden	1	109	986.53	0.004	5.0

The scores obtained for the 16th century do not vary much for *werden* as an auxiliary and when used as a full verb. The most substantial changes can be found in the betweenness centrality values for *werden* used in its infinitive and past participle forms. Such variations can be connected first and foremost to the increase of their out-degree links, as also reported in the chart. Further, as shown in section 7.8, the 16th century displays a drastic growth in the average path length. Such an outcome has been related to the frequency increase of the perfect constructions (*Perfekt* and *Plusquamperfekt*) and the further grammaticalization of the *werden* future. This latter process is visible in the instances in which speakers start to add an extra form of *werden* in the future periphrases. The increased between centrality values captured such changes in the syntactic structure of the networks, assigning to *werden* in both its infinitive and past participle forms significantly higher scores than in the prior centuries. These scores mean that these two forms appeared with more frequency on the shortest paths among the sentences 'nodes.

The table below shows the first ten nodes ranked for the in-and out-degree (with numeric values) and centrality measures in the 16th century. The forms of *werden* are in bold.

Table 18: First 10 nodes of the 16h century ranked for in- and out-degree, betweenness, eigenvector, and eccentricity centrality

	In-Degree		Out-Degree		Betweenness C.	Eigenvector C.	Eccentricity
1	AR der	185	**AX werden**	394	PR von	AR der	PcPs reichen
2	CJ und	73	**V werden**	135	PR zu	CJ und	PcPs erneuen
3	PR von	57	**PcPs werden**	109	PR mit	AR ein	IV richten
4	PR in	47	**IV werden**	74	PR in	PP er	N fülle
5	AR ein	46	PR von	61	IV nehmen	AJ all	PcPs anstecken
6	PP er	40	PR in	61	N könig	AJ sein	IV sterben
7	PR zu	37	PcPs nennen	60	N leute	PR von	IV wissen
8	AD nicht	30	PR zu	45	PR aus	PR zu	IV ärgern
9	PR mit	29	AX sein	34	N rat	PR in	IV strafen
10	AJ sein	23	PR mit	29	**IV werden**	DM der	PR an

The data showed in the table do not indicates any substantial change for both in-degree and eigenvector centralities, since articles, prepositions, and pronouns are again among the nodes with the highest scores. The out-degree also shows a stable trend for all the forms of *werden* and such an outcome is connected to the fact that these networks were built around this verb. The changes that were visible in the table previous to this one are detectable here as well. In the eccentricity centrality column, which ranks the nodes that are the easiest to reach and thus, the easiest with which nodes can establish a connection, the nodes with the highest ranks are past participle and infinitive verbs. The variations in the syntactic structures of the network priorly discussed are captured by this centrality as well and relate to the frequency increase of the perfect constructions and the further grammaticalization of the *werden* future.

7.12 17th Century

The following chart displays in- and out-degrees, the betweenness, eigenvector, and eccentricity centralities for each form of *werden* (auxiliary, full verb, infinitive verb, past participle) found in the network of the 17th century.

Table 19: Centrality measures for *werden* in the 17th Century

	In-degree	Out-degree	Betweenness C.	Eigenvector C.	Eccentricity
AX werden	0	506	0.0	0.0	6.0
V werden	0	78	0.0	0.0	8.0
IV werden	6	98	148723	0.045	6.0
PcPs werden	1	136	44964	0.005	6.0

As observed for the previous century, the scores for both *werden* as an auxiliary and when used as a full verb do not change much in comparison to the 16th Century. The most notable variations are again in relation to *werden* in its infinitive and past participle forms. The betweenness centrality values for these two forms are substantially higher than in the previous century, and so are also their scores for the outgoing links. In section 7.8, it has been reported than the average path length for this period is also higher than in any other century. All these factors can be again connected with the syntactic changes that are happening in the language in the latest century of the Early New High German period. As already mentioned, among these changes, there

are the further development of the perfect constructions and the further grammaticalization of the *werden* future. The increase in the frequency of these two structures boosts the chances for *werden* in the infinite and past participle forms to appear in the shortest path among nodes due to their positions in the sentences' networks. Betweenness centrality is able again to capture such changes and assigns high scores to both these forms of *werden*.

The table below shows the first ten nodes ranked for in-and out-degree (with numeric values) and centrality measures in the 17th century. The forms of *werden* are in bold.

Table 20: First 10 nodes of the 17h century ranked for in- and out-degree, betweenness, eigenvector, and eccentricity centrality

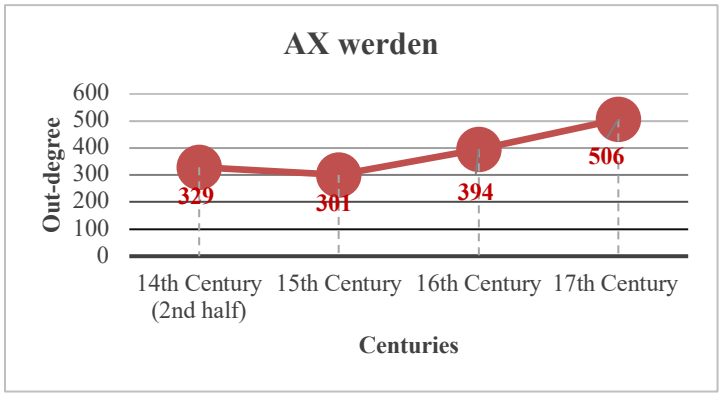
	In-Degree		Out-Degree		Betweenness C.	Eigenvector C.	Eccentricity
1	AR der	259	**AX werden**	506	PR zu	AR der	PcPs sehen
2	AR ein	77	**PcPs werden**	136	**IV werden**	AR ein	IV mahen
3	PR von	73	**IV werden**	98	AJ viel	CJ und	N bischoff
4	PR zu	58	**V werden**	78	PR von	PR von	PR um
5	CJ und	56	PR von	76	N zeit	PR zu	N gut
6	PR in	56	PR in	60	PR mit	PR in	N reich
7	PR mit	50	PR mit	60	PR in	PP er	PR wegen
8	AD nicht	45	PR zu	59	IV sein	AD nicht	PcPs bezahlen
9	PP er	32	IV sein	41	**PcPs werden**	PR mit	PcPs loben
10	AJ sein	25	AX sein	32	PR durch	N stadt	PcPs liefern

As observed for the entire Middle High German period and for the first three centuries in Early New High German, the scores for in-degree and eigenvector centralities have remained mostly unchanged. The out-degree shows also all the forms of *werden* in the highest-ranking positions, indicating that also according to this centrality, the status of *werden* did not change much through the century. The presence of *werden* among the nodes with the highest betweenness centrality score can be related here to the changes that have been previously discussed. The occurrences of past participles and infinitive verbs in the eccentricity centrality columns also reflect that changes captured by the out-degree and betweenness centralities.

7.13 Discussion

7.13.1 Out-degree

The next figures show the results obtained for the out-degree centrality for all the 4 forms of *werden*:



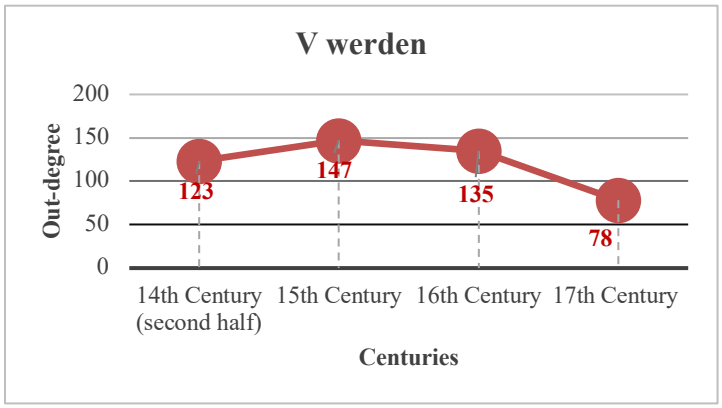


Figure 81: Out-degree for *werden* as an auxiliary and as a full verb

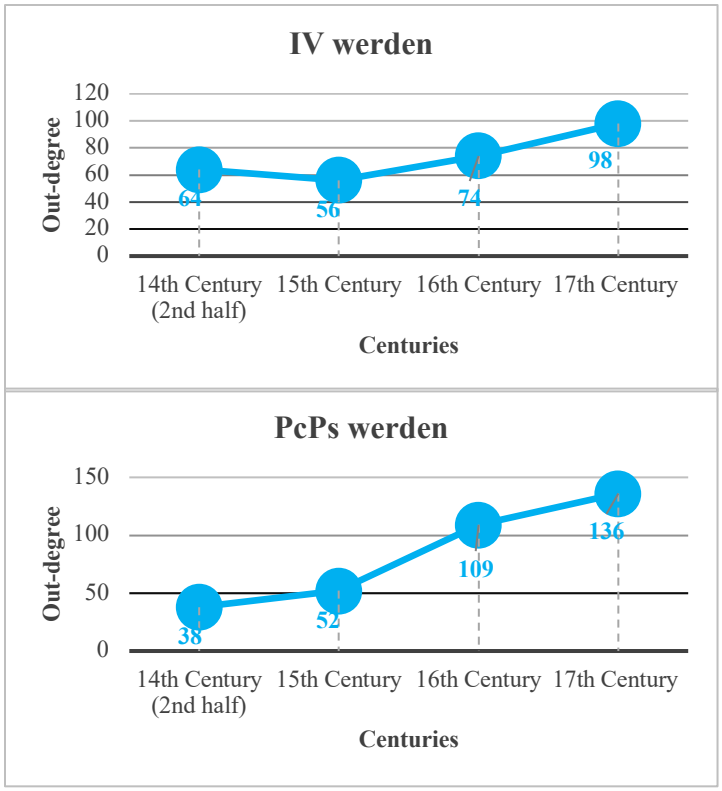


Figure 82: Out-degree for *werden* in its infinitive

The out-degree centrality measures the outgoing link from a particular node. Except for *werden* used as a full verb, all the other forms of *werden* display an increase of their out-degree scores. Such a growth, especially for the infinitive and past participle forms of *werden*, was reported in chapter 6. In this chapter, it has been shown that the Early New High German corpus has a larger number of instances of the passive in the perfect constructions and with *werden* used with modal verbs. The values of the out-degree centrality confirm the results of the historical linguistic analysis for all the forms of *werden*.

7.13.2. Betweenness Centrality

The next figure shows the results obtained for the out-degree centrality for *werden* in its infinitive and past participle forms since *werden* as an auxiliary and as a full verb got scores equal to zero throughout the entire Early New High German period.

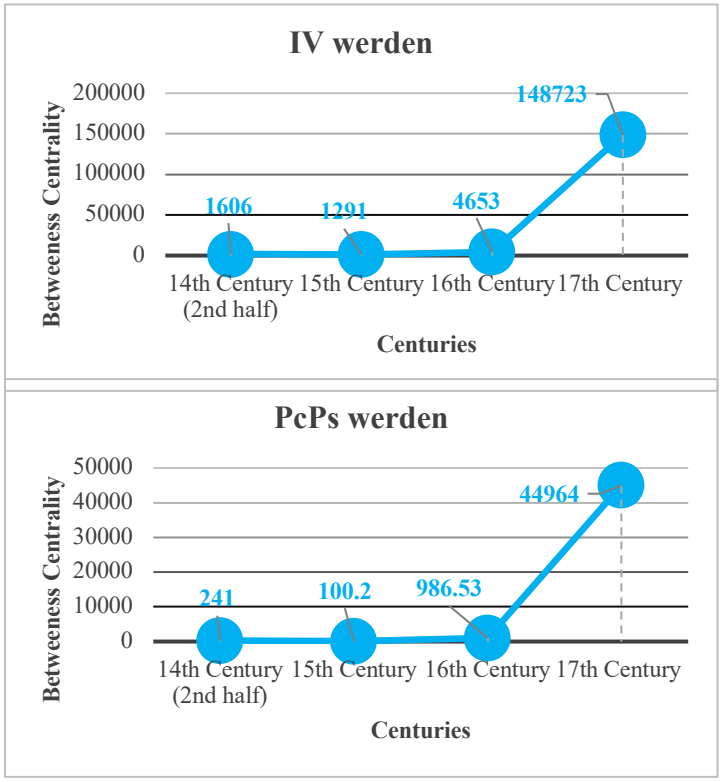


Figure 83: Betweenness centrality for *werden* in its infinitive and past participle forms

As observed in Middle High German, this centrality assigned considerably different scores for all the forms of *werden*. Both *werden* as an auxiliary and *werden* when used as a full verb got scores equal to zero throughout the Early New High German period as well. This means that *werden* in these two forms never served as the shortest path among the other elements in the sentences in these four centuries either. *Werden* in its infinitive and past participle forms, on the contrary, increase their score substantially, especially in the last two centuries. Such variations in the scores have been attributed to the further development of the future periphrases, the grammaticalization of the perfect constructions, and to the frequency increase of the combination of *werden* with modal verbs. These processes have significantly increased the instances in which *werden* appears on the shortest paths among the sentences' nodes, due to its intermediate position between the head and the other elements of the verbal phrases in which it is used.

7.13.3 Eigenvector Centrality

The next figure shows the results obtained for the out-degree centrality for *werden* in its infinitive and past participle forms since *werden* as an auxiliary and as a full verb got scores equal to zero throughout the entire Early New High German period.

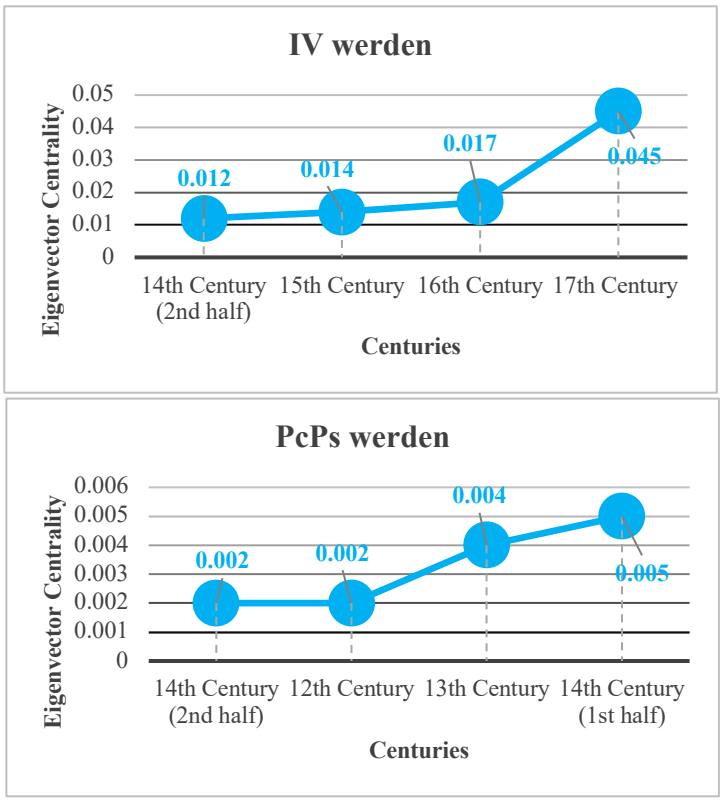


Figure 84: Eigenvector centrality for werden in its infinitive and past participle forms

Eigenvector centrality assigns high scores to those nodes that are connected with other nodes that have high degree centrality scores. The results for Early New High German showed an analog situation to the one observed for Middle High German: the elements in the highest-ranking positions were articles, nouns, pronouns, and prepositions. These high scores also implied that the nodes with which these elements are linked are the most influential in the networks. Because every network's sentence in this study contains a form of *werden*, this trend indirectly indicates that *werden* is a particular influent node. As previously explained, articles connect with nouns that can function as subjects of a form of *werden* or a complement of a verb which is used in combination with *werden* as an auxiliary. The personal pronouns can do the same and the prepositions can be

connected to verbs that come in combination with *werden* as an auxiliary. As these networks do not specify the functions of nouns or pronouns (subjects or objects of a verb), it is not possible to capture which connections are the most significant for the eigenvector centrality. One notable change for this centrality is the score increase for *werden* in its infinitive form in the 17th century. The descriptive statistics and the other centrality measures (especially betweenness centrality) have already shown how the network for this century is characterized by significant changes in its structure. The slight increase in the score for this centrality probably reflects these changes as well, even those the values for this centrality stay relatively low in both Middle and Early New High German.

7.13.4 Eccentricity Centrality

The next figures show the results obtained for the eccentricity centralities for of *werden*:

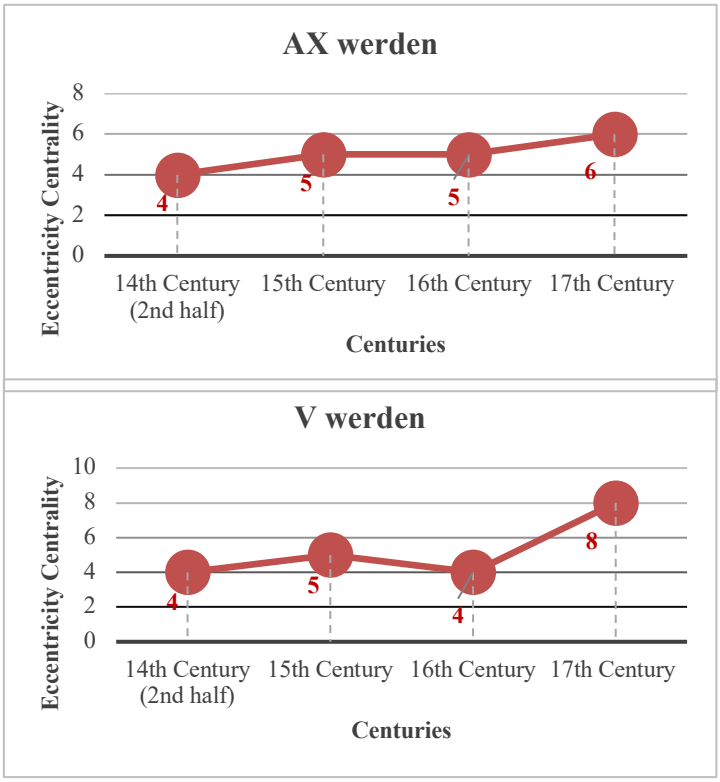


Figure 85: Out-degree for *werden* as an auxiliary and as a full verb

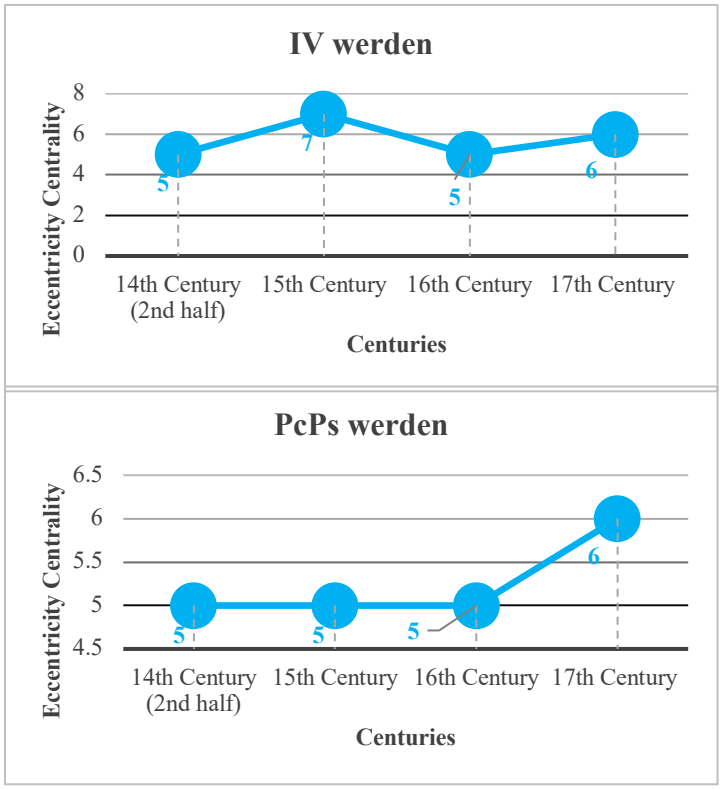


Figure 86: Out-degree for *werden* in its infinitive and past participle forms

This centrality measure indicates which elements in a network are the most easily reachable by the others, thus, easy to establish connections with. Only *werden* in its infinitive form is ranked among the 10 nodes with the highest scores but only in the 15th century. The nodes that are usually ranked highly by this measure are modal verbs, but also past participles and infinitive verbs. It has been discussed how the elements present in this column are a reflection of the syntactic changes that have been observed in Early New High German, and especially in the 17th century. The increase of the scores for almost all the forms of *werden* in this exact century could be related to these changes as well, as new structures emerge or further grammaticalized.

7.14 Summary

The analysis of the verb *werden* in Early New High German and the use of the centrality measures offer relevant information about the status of the verb *werden* in the time frame that goes from the second half of the 14th to the 17th centuries. The out-degree centrality confirms the trend

observed in Middle High German and the results obtained in the historical linguistic analysis carried out in chapter 6: the instances of *werden* keep increasing and such an outcome parallels the growth of the outgoing connections that *werden* could establish in Early New High German. Betweenness centrality assigned scores equal to zero to *werden* as an auxiliary and used as a full verb in these centuries as well. This implies that these forms never appeared on of the shortest paths between nodes in the network's sentences probably because they function as the head of the verbal phrases in which they are used. A reverse trend was observed for both *werden* used in the infinitive form and as a past participle since their scores slowly rise in the first two centuries and grown exponentially in the last two. As previously discussed, the results obtained through this centrality measure have to be related to the many syntactic changes happening in this period, such as the further development of the periphrastic feature and the grammaticalization of the perfect constructions.

The eigenvector centrality did not assign any high scores to any of the forms of *werden* in Early New High German as well. The nodes that constantly occupy the first positions were articles, pronouns, and adjectives. This scenario indicates that the nodes with which those elements are connected are still the most important in the networks. Because the networks were built around the verb *werden*, it can be assumed that all the forms of this verb are directly or indirectly connected to the elements with the highest eigenvector centrality values and that *werden* is one of the most influential nodes in the networks.

Eccentricity centrality, which finds the nodes that are the easiest with which one node can establish a connection, showed the *werden* in its infinitive form was one of these nodes in the 15th century only. The elements that are ranked highly by this centrality measure are modal verbs, past participles, and infinitive verbs. These results are also probably related to the many changes that happen to the networks in Early New High German, especially in the last two centuries. The increase of frequency of the passive, the instances of these periphrases in the perfect constructions, and the further development of the analytic future boost the accessibility to past participles and infinitive verbs, among others. This centrality measure was able to capture these changes throughout the four centuries, even though only indirectly.

7.15 *Werden* in Middle and Early New High German

The following charts summarize the results obtained through the descriptive statistics of the networks of Middle and Early New High German. The first chart displays the values for the cluster coefficient, which indicates how well connected the neighboring nodes of a specific node are. The second chart shows the values of the average path length, which measures the shortest path between nodes in a graph.

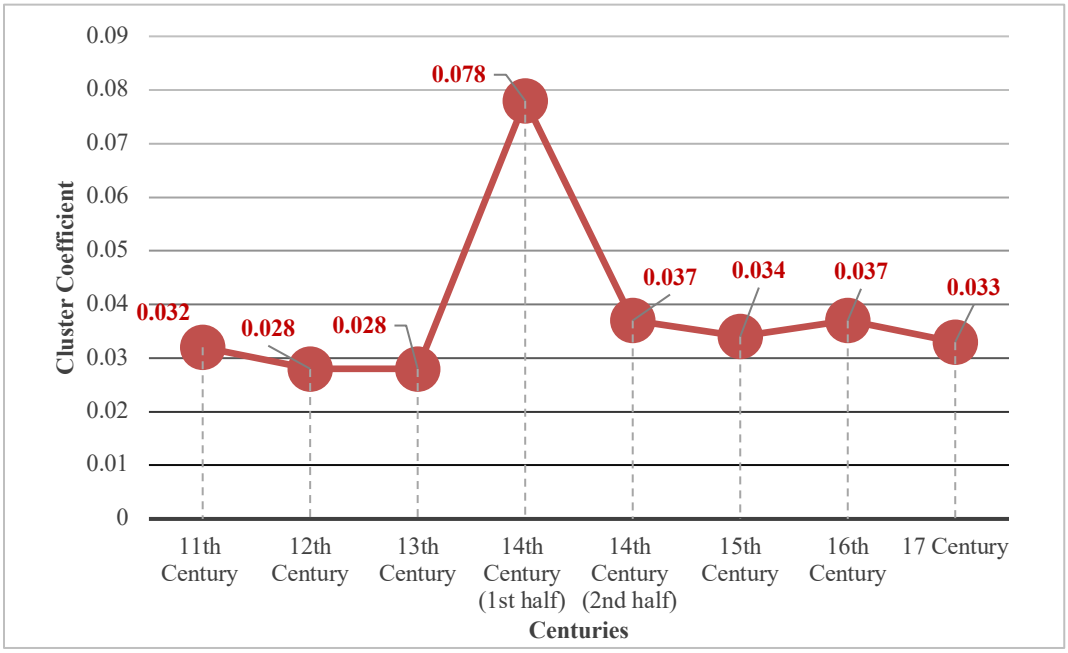


Figure 87: Cluster Coefficient in Middle and Early New High German

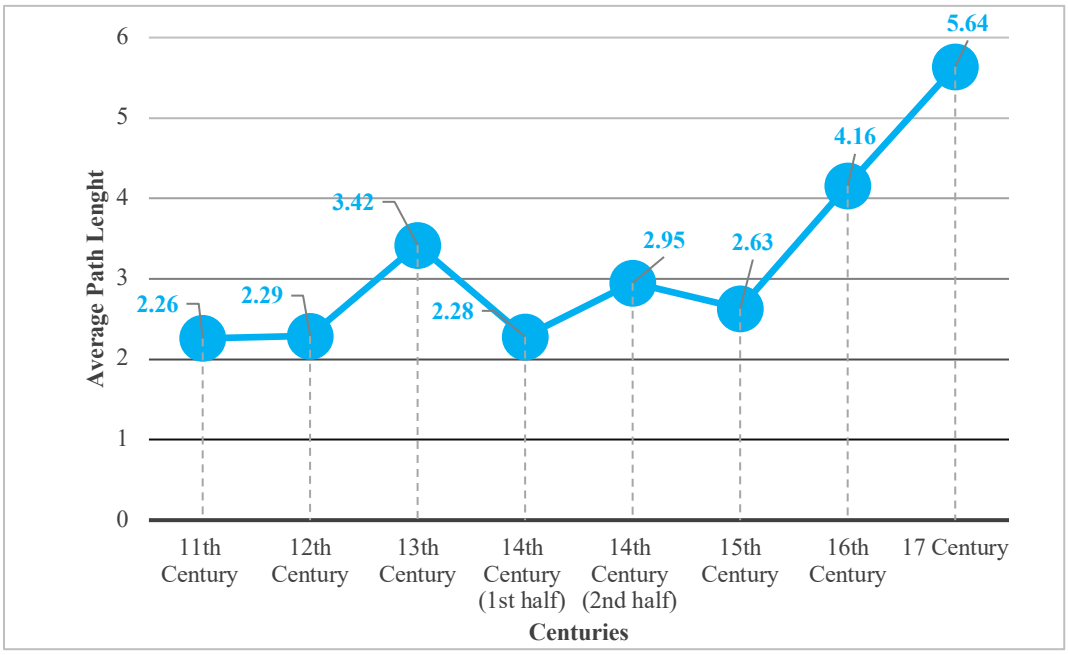


Figure 88: Average Path Length in Middle and Early New High German

The first chart shows how the cluster coefficient remains stable for the majority of the centuries (except for the early 14th century). This outcome is probably related to the fact that the nodes are connected together through asymmetrical relationships. This means that not every node can be directly connected to the other nodes in the network in an equal manner. For instance, it is not possible to have a direct link between a noun and an adverb, or between a verb and an adjective. Hence, the wiring probability of the neighborhood nodes of any given node is lower than in a random network, in which every node can be connected to any other one randomly. In light of these results, and as also showed with the comparison with the random counterparts, the networks of both Middle and Early New High German cannot be considered small-world.

The second chart shows how the average path length remains stable in Middle High German and in the first two centuries in Early New High German. However, after this period of stability, it rises significantly in the 16th century and continues to grow in the 17th century. The changes for this parameter have been attributed to the different syntactic changes that have been observed in these last two centuries. One of these is related to the grammaticalization process of the present and past perfect, which reaches its final stage exactly in this time frame (Concu, 2016). As already discussed, one of the most common effects of grammaticalization is the increase in the

frequency of that particular construction. Kuroda (1999) reports an increase of the present and past perfect exactly in Early New High German. Another significant change is the further development of the *werden* future periphrastic constructions. Such a development was attested in constructions in which an additional form of *werden* was found, that was attached to the already frequent combinations of *werden* and infinitive verbs (Kotin, 2003). These syntactic changes were echoed in the higher average path length of the networks of the 16th and 17th centuries. These results suggest that, while the average path length in Middle High German and in the first two centuries of the Early New High German period displayed small-world-like patterns, the networks of the 16th and 17th centuries shifted to non-small-world structure.

The following figures summarize the results for both Middle and Early New High German for all the four forms of *werden* (auxiliary, full verbs, infinitive and past participle). The first sets of figures display the scores for the out-degree, which measures the outgoing links. The second set refers to betweenness centrality, which indicates how often a node appears in the shortest path between the nodes in the networks' sentences. Since both *werden* as an auxiliary and as a full verb had scores equal to zero, only the values for *werden* in its infinitive and past participle forms will be displayed. The third set displays the values for the eigenvector centrality, which finds central nodes that are connected to those nodes in the networks with high out-degree scores. For this centrality as well only the results for *werden* in its infinitive and past participle forms will be displayed. The last set of figures shows the values for the eccentricity centrality, which finds the nodes in the networks that are the easiest to reach by the other nodes in the network.

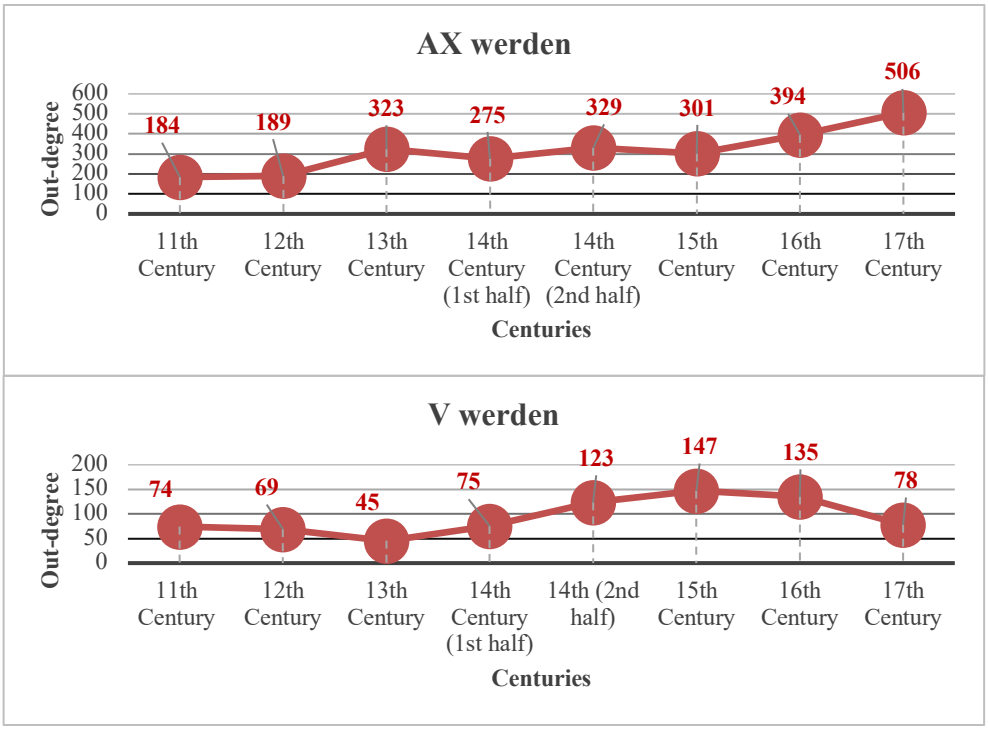


Figure 89: Out-degree for *werden* as an auxiliary and as a full verb

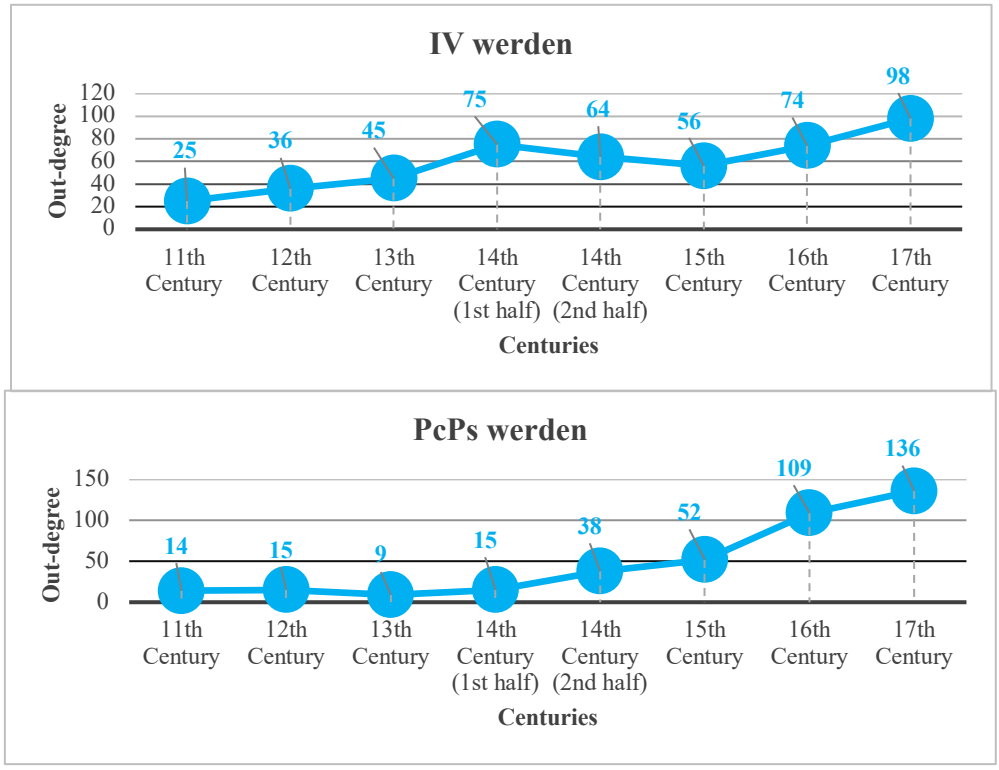


Figure 90: Out-degree for *werden* in its infinitive and past participle forms

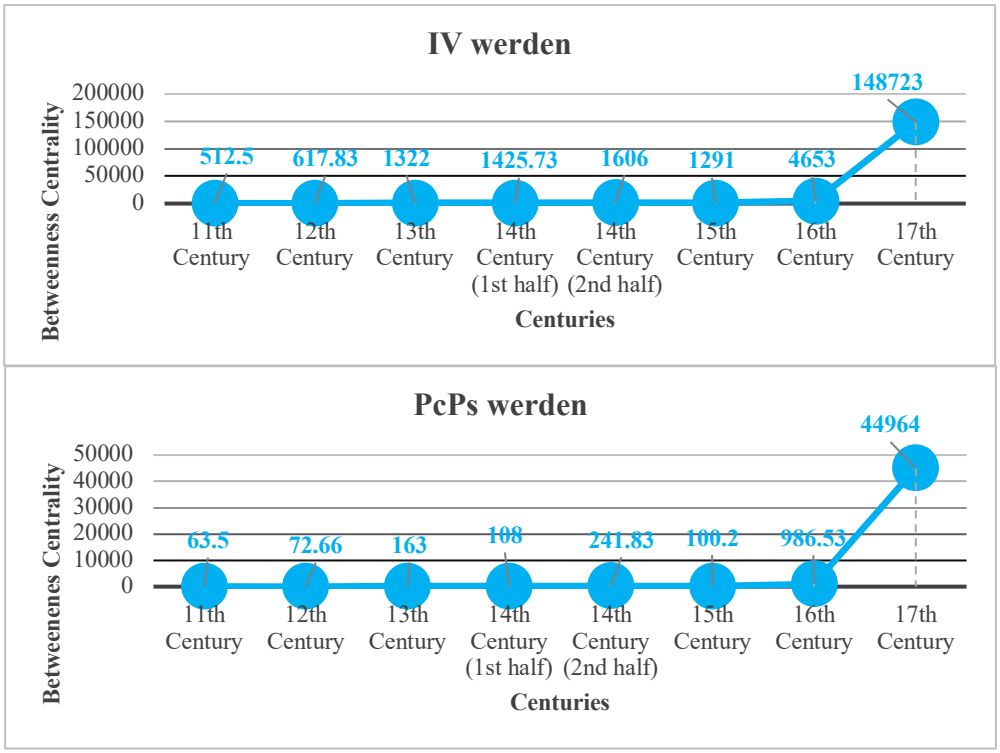


Figure 91: Betweenness Centrality for *werden* in its infinitive and past participle forms

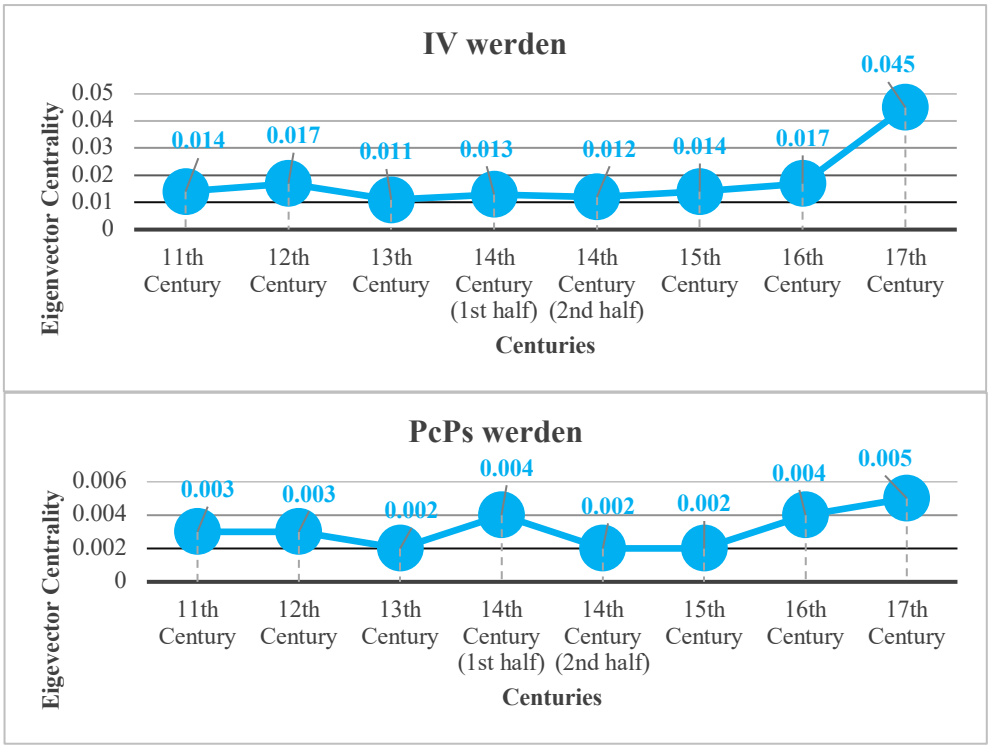


Figure 92: Eigenvector Centrality for *werden* in its infinitive and past participle forms

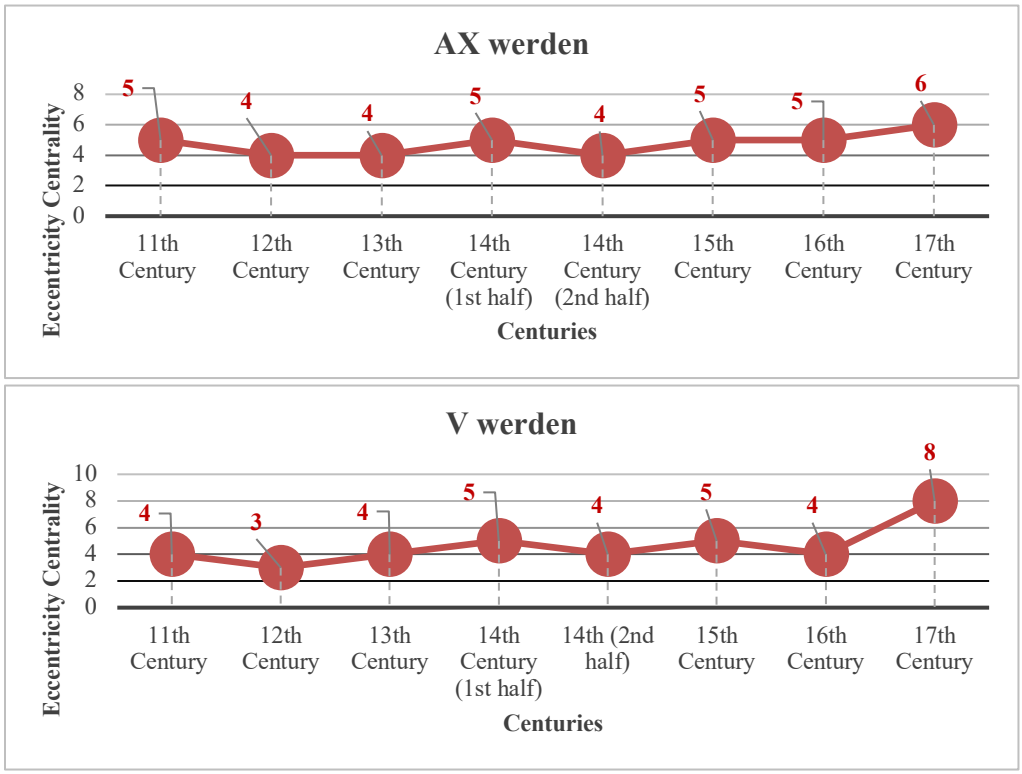


Figure 93: Eccentricity Centrality for *werden* as an auxiliary and as a full verb

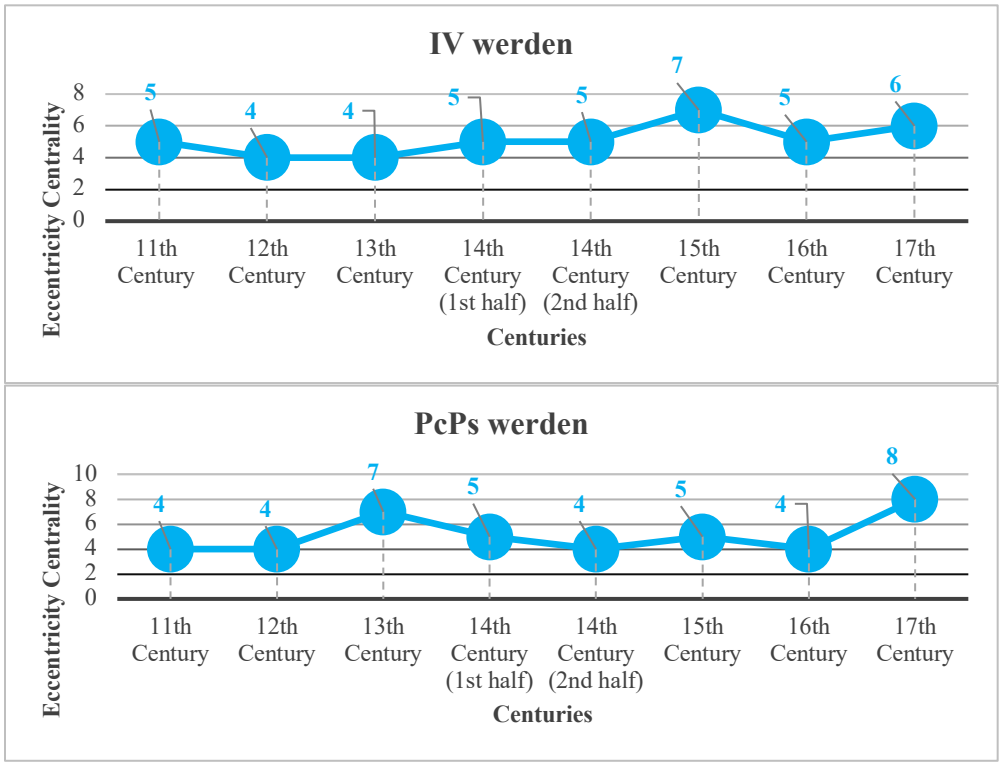


Figure 94: Eccentricity Centrality for *werden* in its infinitive and past participle form

The charts display the behavior of *werden* in Middle and Early New High German. The out-degree is the measure that shows a similar trend for almost all the forms of *werden*. In fact, all the forms, except for *werden* when used as a full verb, constantly increase their values in the time frame that goes from the 11th to the 17th centuries. The growth of the out-going links echoes the results obtained in the historical linguistic analysis carried out in chapters 5 and 6. Due to processes of desemantization, *werden* is able to establish links with a constantly growing number of elements, and such a trend could be captured by the out-degree centrality.

Betweenness centrality assigned scores equal to zero to both *werden* as an auxiliary and when used as a full verb. *Werden* in its infinitive and past participle forms, on the contrary, see an increase in their values in the last two centuries of the Early New High German period. The results obtained through this centrality measure suggest, first and foremost, that betweenness centrality is related to the in-degree of the nodes. Since both *werden* as an auxiliary and when used as a full verb always function as the head of the respective verbal phrases, they are never the target of incoming links. Quite the opposite, *werden* in its infinitive and past participle forms do receive incoming links since they are used in combinations with modal verbs and in the perfect constructions. In these occurrences, *werden* is positioned between the head and the other elements of the verbal phrases, and such a position boosts its chances to appear on the shortest path among nodes in the networks' sentences. The higher scores of *werden* in its infinitive and past participle forms in the 16th and 17th centuries reflect the increased frequency of the periphrastic future with an additional form of *werden*, the increase of the combinations with the modal verbs, and the growing of the instances of the passive used in the perfect constructions.

Eigenvector centrality also assigns to *werden* as an auxiliary and when used as a full verb scores equal to zero and relatively low values to *werden* in its infinitive and past participle forms. As already discussed, the nodes ranked high for this centrality are articles, pronouns, and prepositions. Such an outcome is linked to the fact that these elements have direct connections to the most influential nodes. Since all the networks were built around the verb *werden*, the nodes with the highest scores for this centrality are always directly and indirectly connected to *werden*, which is, therefore, one of the most influential nodes in the networks.

The last centrality measure, eccentricity centrality, attributes similar scores to all the forms of *werden* with an overall increase in the last century. However, *werden* rarely appears among the 10 nodes with the highest scores. High values are often assigned to modal verbs, past participles,

and infinitive verbs, and such trends probably reflect the syntactic changes affecting the networks throughout the centuries. Such transformations are influencing the degree by which nodes can be reached by the other nodes in the network at any given time. Interestingly enough, this centrality seems not to be correlated to the out-degree centrality, meaning that the easiness in establishing connections is not always determined by the growth of outgoing links.

Overall, the application of the centrality measures to the forms of *werden* and to the nodes connected to it has revealed how syntactic changes that affect a particular verb or structure have also repercussions on those elements that are directly or indirectly connected to it and the network as a whole. Traditional linguistic analyses tend to focus only on changes happening within a particular structure without taking into consideration the effects of such changes to the language system. Network analyses allow us to focus on particular structures and, at the same time, to observe changes that are taking place on a more 'global' scale, giving more exhaustive information about the overall state of the language at any given time. The possibility to look at syntactic changes on both 'local' and 'global' scales, as shown in this dissertation, highlights the advantages of implementing an interdisciplinary approach such as network analysis for the study of historical changes and languages as a whole. In this dissertation, I have shown how the changes affecting the verb *werden* have affects not only on the periphrastic structures in which it is used, but also on the other elements (such as articles, pronouns, and prepositions) that are used around those structures. The centrality measures were able to capture that and to give valuable information about the status of the language in Middle and Early New High German.

CHAPTER 8. CONCLUSIONS

The goal of this dissertation was to track the historical development of the verb *werden* in Middle High and Early New High German combining a corpus linguistic approach with the tools of network science. For this reason, the analysis was carried out in two different stages. In the first stage, I focused on the frequency of the verb *werden* and the elements that co-occurred with it throughout Middle and Early New High German. This part of the analysis showed, on the one hand, that *werden*, when appearing in passive periphrases and as a copula verb, was used in similar ways in both Middle and Early New High German. On the other hand, the data displayed an increase of the attestations of *werden* combined with verbs in infinitive forms and a simultaneous decline of the combinations of *werden* with present participles in Early New High German. These results were particularly insightful in regard to the origin of the future periphrases. The occurrences found in the corpus provided, in fact, support for the claims of scholars such as Kotin (2003) and Krämer (2005). According to them, these periphrases originated from different but related sources and coexisted as "twin" constructions until the periphrases of *werden* plus present participles slowly disappeared (Kotin, 2003, p. 165). The data also showed a significant increase in the instances in which the passive appeared in *Perfekt* and *Plusquamperfekt* in Early New High German. While in Middle High German *werden* used as a copula was the only form that could appear in these two tenses, in Early New High German, the data showed instances in the passive as well. These results can be related to the grammaticalization of the present and past perfect constructions, which took place around the same time as the data showed the first instances of the passive in *Perfekt* and *Plusquamperfekt*. Further, the instances of the modal verbs indicated that the combinations of *werden* with *sollen*, *mögen*, *wollen*, and *müssen* were used in a comparable way in both Middle and Early New High German, while instances of *werden* combined with *können* and *dürfen* could only be found in Early New High German.

In the second stage of this study, I investigated the same data through a complex network framework. Descriptive statistics were applied to uncover the features of the networks that have been created for each of the centuries included in this research. The results showed some differences between the Middle and Early New High German networks. The main discrepancies were related to the average path length. These characteristics were discovered by comparing the real syntactic networks with their random counterparts which had the same in- and out-degrees. In

regard to the cluster coefficients, the networks of both Middle and Early New High German had lower values than the random networks, meaning that, the real networks cannot be classified as being small-world. Further, the average path lengths were similar to their random counterparts only in Middle High German. The values for the Early New High German networks were, indeed, higher than those of the random networks. These results indicated that the structure of the networks shifted to a non-small-world structure between these two periods and such a shift has been related to the changes that have been observed especially in the 16th and 17th centuries. These findings do not align with the literature, since many scholars have claimed that language networks are always small-world (Liu & Hu, 2008). This discrepancy could be explained either through the specific features of the Middle and especially Early New High German networks, or the way in which the networks were built. However, it is worth mentioning that the vast majority of the literature focuses on modern languages, while this dissertation addressed the German language between the 11th and the 17th centuries. Hence, it is not possible to know if Modern German developed small-world features after the Early New High German period. Research on Modern German needs to be conducted in order to find out if the properties of the Middle and Early New High German networks are specific to these periods or if German represents an exception overall.

In regard to the degree distribution, the scatter plots of the Middle and Early New High German networks showed that only a few nodes had high out-degree values, which were visible on the low right corner of the x-axis. Moving from the right to the left of the x-axis, it was also noticeable that the number of nodes that have lower out-degree values increased until it reaches the highest part of the left corner of the y-axis. The data distribution on the scatter plots in both Middle and Early New High suggested that the networks were probably scale-free or, at least, that their out-degree distribution shows patterns that are similar to those found in scale-free networks.

After the analysis of the networks through descriptive statistics, I tracked the development of *werden* in Middle and Early New High German using the centrality measures that have been described in chapter 2. These measures were in- and out-degree, betweenness, eigenvector, and eccentricity centralities. The out-degree was the measure that was able to capture the increase of connections that was observed in the historical linguistics part of this dissertation. As previously mentioned, the data showed a growth in the number of combinations in which *werden* appeared with the verbs in the infinitive form starting from the 15th century. In Early New High German, *werden* also started to appear in the modal constructions with *können* and *dürfen*. The out-degree

registered an increase in connections for all the forms in which *werden* was used except for *werden* when used as a full verb, whose out-degree values slightly decrease in the last centuries of the Early New High German period.

From a complex network perspective, it can be said that the node *werden* increased its capacity to establish connections to a broader type of nodes with a different grammatical status between the 11th and 17th centuries. As discussed in the theory chapter, the process by which a node that is already "rich" becomes "richer" and connects to a continuously increasing number of other nodes is called preferential attachment. The data obtained in this research indicate that *werden* was able to become "richer" throughout the centuries, providing evidence for preferential attachment in these syntactic networks.

In chapter 2, I also discussed how the "rich-get-richer" phenomenon was revisited and renamed the "fit-get-richer" as Barábasi discovered how, despite their late arrival in the networks, some nodes were able to overtake older and well-connected ones (Bianconi & Barabási, 2001). They forged the concept of "fitness value", explaining also how nodes in a network are not equal and the network itself is a competitive environment. In the review of the literature, it has been reported how scholars such as Bogner (1989), Fritz (1997), Schmidt (2000), and Harm (2001) noticed how *werden* had to compete with the constructions with modal verbs such as *wollen* and *sollen* to express future time references. The node *werden* was, however, able to acquire the status of unique auxiliary for the future and such a capacity became stronger especially in the 16th and 17th centuries. Throughout the centuries, *werden* must have acquired (or lost) specific features that made it the best option to carry on such a function. The centrality measures used in this research helped shed light on the behavior of *werden* in Middle and Early New High German and can be used to address this topic.

Betweenness centrality, which finds the nodes in the network that serve as the shortest path between the other nodes, showed that both *werden* as an auxiliary and *werden* used as a full verb did not play a big role in this sense. *Werden* in the infinitive form and used in its past participle form, on the contrary, saw their scores increasing over time. This can be explained by looking at the position that these forms occupied in the sentences. In these networks, *werden* used in its infinitive form is usually connected to a modal verb, and in the last centuries of the Early New High German period, to another *werden* that functions as an auxiliary in the future periphrases. As for *werden* in the past participle, this form appears when the passive is used in the present and past

perfect tenses. In these occurrences, *werden* is located between the head and the other elements of the verbal phrases. The growth of instances of *werden* in its infinitive and past participle forms boosted the chances of this verb to appear on the shortest path of the nodes in the sentences' networks, as showed by the betweenness centrality scores. These results imply that betweenness centrality is influenced by the syntactic position that a word occupies in any given sentence. Further, it can be also said that it is not necessary for a node to serve as the shortest path between the other nodes in order for it to be able to acquire more and more connections through time. In social networks, nodes with high betweenness are critical because they control the flow of information among the other nodes (Golbeck, 2015). In a direct network, however, betweenness centrality can mean different things: it can mean that a particular node is connected to nodes that are not connected to its neighbors, or that it connects nodes that would be otherwise distant from each other (Golbeck, 2013, p. 30). The data obtained in this dissertation suggest that in a syntactic network, betweenness centrality refers to a particular position in a sentence. Further, this centrality is not a good predictor of the capacity of a node to acquire more links, since only *werden* in the infinitive form and as past participle had high betweenness values.

Eigenvector centrality, which measures the importance of a node considering also the importance of its neighbors (Golbeck, 2013, p. 30) only indirectly indicated that *werden* was an influential node. In both Middle and Early New High German, the elements that were constantly ranked among the ones with the highest scores were articles, prepositions, and personal pronouns. Every node in the networks was directly or indirectly connected to *werden*, meaning that this verb was one of the most influential. Opposite to what was observed for betweenness centrality, this measure could be a good indicator of which nodes in the networks are the most likely to gain the highest number of connections based on their influence.

The last centrality used, eccentricity centrality, distinguishes the nodes that are the easiest ones to be reached by other nodes. The results obtained through this centrality were similar to the ones obtained with eigenvector centrality. The forms of *werden* were among the 10 elements with the highest eccentricity values only for a few centuries. The data obtained through this centrality indicate that, first, preferential attachment is not connected to how easy a node can be reached by other connections, and, second, this measure, is not the best predictor of which nodes will be able to acquire more connections over time. The "fitness-value" in a syntactic network, or at least in the syntactic networks created in this dissertation, is, therefore, not strictly related to the features

highlighted by betweenness and eccentricity centralities. The out-degree and, to some extends, the eigenvector centrality, gave more insightful information on this topic. As reported in chapter 2, one of the most fascinating aspects of the Fitter-Get-Richer model was that "fitness is not assigned by any individual but reflects the network's collective perception of a node's importance relative to the other nodes. We can, therefore, determine a node's fitness by comparing its time evolution to the time evolution of other nodes in the network" (Barabási, 2016, p. 208). The out-degree was the measure that captured the evolution of *werden* over time and it is then a good indicator of how "fit" *werden* got throughout the centuries. Eigenvector centrality also indicated the *werden* was one of the most influential nodes in the network and such a characteristic may also be a contributor to its fitness values.

As already mentioned, the goal of this dissertation was not to find a definition for a fitness-values in linguistics. However, the data allowed to address this issue briefly in what can be considered a first step in the determination of such a notion when analyzing syntactic networks. The cognitive mechanisms behind how nodes are perceived in a network are definitely worthy of further investigations, and the centrality measures, as shown in this dissertation, could be implemented as a useful tool to track the evolution of the nodes in a network.

The results obtained through the historical linguistic and network-related part of this dissertation can be used to answer the two central research questions stated in the introduction:

1. Since "behind each complex system there is an intricate network that encodes the interactions between the system's component" (Barabási, 2016, p. 24), are linguistic networks subjected to the same dynamics that are common among non-linguistic networks?

The data from Middle and Early New High German displayed an increase of the type of connections that the node *werden* was able to establish throughout the centuries. Such a process is known in the literature as preferential attachment and the evidence obtained in this dissertation indicate that linguistic networks, and specifically, syntactic networks, are not exempt from processes such as the rich-the-richer (or fit-get-richer) phenomenon that are common among non-linguistic networks. The findings also suggest that syntactic networks are a competitive environment in which the nodes are not equal and compete against each other to gain more

connections. In a syntactic network such as those analyzed in this dissertation, more links could indicate the ability of a specific node to carry out new functions and the consequent rise of new periphrastic constructions in the language.

2. If so, what can a complex network approach tell us about historical syntactic linguistic changes?

The results obtained in this dissertation provide evidence for preferential attachment in syntactic networks. The application of a complex network approach showed that some of the historical linguistic changes that are common in language evolution (such as the rise of the periphrastic future tense that has been addressed in this dissertation) resemble the growth dynamics that scholars have observed in other non-linguistic complex networks. The node *werden* acquired new connections over time and was able to broaden the types of elements with which it established new links, showing that a phenomenon such as preferential attachment can be also found in syntactic networks. The network-related analysis of this dissertation captured not only the changes related to the status of the node *werden*, but it was also able to track other significant changes in the language. In particular, the increase of the average path length in the networks of the last centuries of the Early New High German period showed how syntactic changes affected the entire structure of the networks, causing a shift to a non-small-world structure. In light of these results, it can be further claimed that the application of a complex network approach highlighted the fact that historical linguistic changes affect the language systems as a whole instead of affecting only a particular group of words or constructions. The use of descriptive statistics, such as the ones utilized in this dissertation, was useful in this sense and reinforced the idea of languages as complex adaptive systems, in which a small change can affect the system as a whole. The analysis carried out in this study also showed how languages are dynamic and how they evolve over time as they respond to the need of speakers to address old and new communicative needs. For this reason, this dissertation also supports the view of grammar as emergent and as a byproduct of communication.

Overall, the results of this dissertation have shown that the application of a complex network approach for the study of historical linguistic changes is a compelling way to analyze how languages change over time. This dissertation has only scratched the surface of the many

possibilities that this approach offers, since some of the questions that were raised are left unanswered. For instance, it was impossible to determine what makes a node fitter than another one, and what the average path length means in terms of the cognitive effort from the perspective of the speakers. These questions are worth investigating, as a complex network approach is constantly revealing itself to be a useful tool to uncover many of the mechanisms behind the complexity and structure of many aspects of society.

APPENDIX

Middle High German

11th Century

Dialect Area	Text	Year	Word Count
West Middle German	Rheinfränkische Interlinearversion der Psalmen	N/A	514
East Middle German	N/A	N/A	N/A
West Upper German	Rheinauer Gebete	1150	1,484
	Älterer Physiologus	1070	1,613
	Wessobrunner (Ahd.) Predigtsammlung A	1100	1678
	Wessobrunner (Ahd.) Predigtsammlung B	1100	1354
	Wessobrunner (Ahd.) Predigtsammlung C	1100	841
East Upper German	Wiener Notker	1100	4,501
	Wessobrunner Glaube u. Beichte I	1100	1,687
	Otlohs Gebet	1067	889
Nord Upper German	Hoheliedkommentar	1065	14,601
Total			**29,189**

12th Century

Dialect Area	Text	Year	Word Count
West Middle German	Trierer Psalmen	1200	15,206
East Middle German	Schleizer Psalmen	1200	2,454
	Wiggertsche Psalmen	1200	1,350
West Upper German	Züricher Arzneibuch	1200	2,714
	Alkuins Traktat	N/A	1,229
	Engelberger Gebete	1200	1,092
East Upper German	Wiener Physiologus	1120	5,983
	Krakauer Fragmente	N/A	4,962
	Seckauer Gebete A	1200	1,083
Nord Upper German	Bamberger Glaube u. Beichte	N/A	3,330
Total			**39,358**

13th Century

Dialect Area	Text	Year	Word count
West Middle German	Mitteldeutsche Predigten	N/A	10,329
East Middle German	Jenaer Martyrologium	1300	13,879
West Upper German	Schwabenspiegel	1275	12,721
East Upper German	Buch der Könige	1280	12,356
Nord Upper German	Salomon Haus	1278	14,534
Total			**63,819**

14th Century (first half)

Dialect Area	Text	Year	Word Count
West Middle German	Rede von den 15 Graden	1350	14,108
East Middle German	Evangelienbuch des Matthias von Beheim	1343	16,362
West Upper German	Nikolaus von Straßburg: Predigten (C)	1330	21,434
East Upper German	Oberaltaicher Evangelistar	N/A	13,012
Nord Upper German	Nürnberger Stadtbuch	1302	4,920
	Würzburger Polizeisätze	1342	8,534
Total			**64,939**

Early New High German

14th Century (second half)

Dialect Area	Text	Year	Word Count
West Middle German	Benediktinerregel Oxford,	N/A	14,977
East Middle German	Altdeutsche Predigten I	N/A	14,278
West Upper German	Mannen	1370	12,664
East Upper German	Rationale	1384	13,847
Nord Upper German	Namen	N/A	10,740
Total			**76,075**

15th Century

Dialect Area	Text	Year	Word Count
West Middle German	Johann Koelhoff: Chronik	1499	12,020
East Middle German	Johannes Rothe: Chronik	1485	12,809
West Upper German	Gerold Edlibach: Chronik	1485	11,165
East Upper German	Denkwürdigkeiten	1452	14,547
Nord Upper German	Pillenreuth Mystik	1463	12,839
Total			**63,380**

16th Century

Dialect Area	Text	Year	Word Count
West Middle German	Amerika	1599	13,662
East Middle German	Johann Bange: Chronik	1599	13,529
West Upper German	Gespenster	1578	12,519
East Upper German	Moscouia	1557	14,737
Nord Upper German	Summaria	1578	13,211
Total			**67,658**

17th Century

Dialect Area	Text	Year	World Count
West Middle German	Schaubühne	1699	11,734
East Middle German	Jugendlust	1648	13,089
West Upper German	Mythoscopia	1698	11,329
East Upper German	Deo Gratias	1680	6781
Nord Upper German	Spiegel	1668	12,058
Total			**64,936**

REFERENCES

Akamatsu, T. (n.d.). The development of functionalism from the Prague school to the present. In S. Auroux (Ed.), *Geschichte der Sprachwissenschaften* (Vol. 2). Walter de Gruyter.

Amancio, D. (2014). A perspective on the advancement of natural language processing tasks via topological analysis of complex networks. *Physics of Life Reviews*, *11*, 641–643. https://doi.org/10.1016/j.plrev.2014.07.010

Arnett, C. (2000). A Cognitive Approach to the Old Saxon Processual Passive. *American Journal of Germanic Linguistics and Literature*, *12*(1), 81–99.

Arnett, C. (2004). A Cognitive Approach to the Semantic of the German Passive. The Edwin Mellen Press.

Auer, P., & Schmidt, J. E. (2010). *Theories and Methods*. Walter de Gruyter.

Barabási, L. (2013). Network Science. Science, Proceedings of the National Academy of Sciences, 110, 1–6.

Barabási, L. (2014). *Linked*. Persus Publishing.

Barabási, L. (2016). *Network Science*. Cambridge University Press.

Barabási, L., & Réka, A. (1999). Emergence of Scaling in Random Networks. *Science*, *286*, 509–512.

Barceló-Coblijn, L.; Corominas-Murtra, B.; Gomila, A. (2012). Syntactic trees and small-world networks: syntactic development as a dynamical process. *Adaptive Behavior*, 20.

Bechs, F. (1901). Beispiele von der Abschleifung des deutschen Participium Präsentis und von seinem Ersatz durch den Infinitiv. *ZDWF*, *1*, 81–109.

Beckage, M., & Colunga, E. (2016). Language Networks as Models of Cognition: Understanding Cognition through Language. In *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks.* Springer Verlag.

Becker, C.; Ellis, N.; Blythe, R.; Holland, J.; Bybee, J.; Ke, Jinyun.; Christiansen, M.; Larsen-Freeman, D.; Croft, W.; Schoenemann, T. (2009). Language Is a Complex Adaptive System: Position Paper. *Language Learning*, 59, 1, 1-26.

Bell, M., Perera, S., Piraveenan, M., Bliemer, M., Latty, T., & Reid, C. (2017). Network growth models: A behavioral basis for attachment proportional to fitness. *Scientific Report*, *7*. https://doi.org/10.1038/srep42431

Bentz, C., & Buttery, P. (2014). Towards a computational model of grammaticalization and lexical diversity. *Proc. of 5th Workshop on Cognitive Aspects of Computational Language Learning*, 38–42.

Bianconi, G., & Barabási, L. (2001). Competition and multiscaling in evolving networks. *Europhysics Letters*, *54*, 436–442.

Bergs, A., & Brinton, L. J. (Eds.). (2012). *English Historical Linguistics. An International Handbook* (Vol. 1). De Gruyter Mouton.

Bischoff, S., & Jany, C. (Eds.). (2013). *Functional Approaches to Language*. Walter d Gruyter.

Bogner, I. (1989). Zur Entwicklung der periphrastischen Futurformen im Frühneuhochdeutschen. *Zeitschrift Für Deutsche Philologie*, *108*, 56–85.

Borgatti, S. P. (2006). Identifying Sets of Key Players in a Network. 12.

Breithbarth, A. (2006). *The short life of the afinite construction.* Retrieved from https://pdfs.semanticscholar.org/d1cb/7d110852f034439414d343d6af0f819bc822.pdf

Brinton, L. J. (Ed.). (2017). Generative Approaches. In *English Historical Linguistics: Approaches and Perspectives.* Cambridge University Press.

Briscoe, E. (1998). Language as a Complex Adaptive System: Coevolution of Language and of the Language Acquisition Device. *Proceedings of Eighth Computational Linguistics in the Netherlands Conference*.

Bybee, J., & Hopper, P. (2001). *Frequency and the Emergence of Linguistic Structure*. John Benjamins Publishing Company.

Bybee, J. (2003). Cognitive processes in grammaticalization. In M. Tomasello (Ed.), *The New Psychology of Language*. Lawrence Erbaum Associates.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge University Press.

Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The Evolution of Grammar. Tense, Aspect, and Modality in the Languages of the World.* Chicago London: The University Chicago Press.

Campbell, L. (1998). Historical Linguistics: An Introduction. MIT Press.

Chondhury, M.; Mukherjee, A. (). The Structure and Dynamics of Linguistic Networks. In: Ganguly N., Deutsch A., Mukherjee A. (eds) Dynamics On and Of Complex Networks. Modeling and Simulation in Science, Engineering and Technology. Birkhäuser Boston. https://doi.org/10.1007/978-0-8176-4751-3_9

Christiansen, M. H., & Kirby, S. (2003b). Language Evolution: Consensus and Controversies. Cog. Sci., 7(7), 300–307.

Clark, E. (1990). On the pragmatics of contrast. Journal of Child Language, 17, 417–431.

Concu, V. (2015). The Rich-Get-Richer Phenomenon and Language Evolution: The case of the German Present Perfect. *Working Papers of the Linguistics Circle*. 25 (2), 2015.

Concu, V. (2016). Toward the German Present Perfect as an Emergent Structure. *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis*. 21(1), 2016.

Cong, J., and Liu, H. (2014). Approaching human language with complex networks. *Phys. Life Rev*. 11, 598–618. doi: 10.1016/j.plrev.2014.04.004

Corominas-Murtra, B., Valverde, S., and Solé, R. V. (2010). Emergence of scale-free syntax networks. In S. Nolfi Editor & M. Mirolli Editor (Eds.). *Evolution of Communication and Language in Embodied Agents* (83–101). Springer.

Croft, W., & Cruse, A. (2004). *Cognitive Linguistics*. Oxford University Press.

Diewald, G. (2012). Die Modalverben im Deutschen: Grammatikalisierung und Polyfunktionalität. Walter de Gruyter.

Diewald, G., & Hebermann, M. (2005). Die Entwicklung von Werden + Infinitive. In *Grammatikalisierung im Deutschen*. Walter de Gruyter.

Diewald, G., & Wischer, I. (2013). Markers of Futurity in Old High German and Old English. A Comparative Corpus-Based Study. In G. Diewald, L. Kahlas_Tarkka, & I. Wischer (Eds.), *Comparative Studies in Early Germanic Languages: With a focus on verbal categories.* (pp. 201–222). John Benjamins Publishing Company.

Dorogovtsev, S., & Mendes, J. F. F. (2002). Evolution of networks. *Adv. Phys.*, *51*(1079). https://doi.org/10.1080/00018730110112519

Downey, A. (2018). *Think Complexity*. Newton: O'Reilly Media.

Ehlers, K.-H. (2005). Strukturalismus in der deutschen Sprachwissenschaft: die Rezeption der Prager Schule zwischen 1926 und 1945. Walter d Gruyter.

Einhauser, E. (2001). Die Entstehung und frühe Entwicklung des junggrammatischen Forschungsprogramms. In S. Auroux (Ed.), *Geschichte der Sprachwissenschaften* (Vol. 22). Walter de Gruyter.

Engelberg, S. (2000). Verben, Ereignisse und das Lexicon. Walter de Gruyter.

Eroms, H.-W. (1992). Das deutsche Passiv in historicher Sicht. Deutsche Syntax: Ansichten Und Aussichten, 225–249.

Feigenbaum, S. (2002). Prepositions in Their Syntactic, Semantic, and Pragmatic Context. Netherlands: J. Benjamins Pub.

Ferrer i Cancho, R.; Solé, R. (2010). The small world of human language. *Proceedings of the Royal Society of London*, 268.

Fleischer, J., & Schallter, O. (2011). Historische Syntax des Deutschen. Eine Einführung. Narr Verlag.

Friz, G. (1997). Historische Semantik der Modalverben. Problemskizze – exemplarische Analysen – Forschungsübersicht. In Frizt, G, and Gloning, T. (eds). *Untersuchung zur semantischen Entwicklungsgeschichte der Modalverben im Deutschen*. Niemeyer.

Geeraerts, D., & Cuyckens, H. (Eds.). (2007). *The Oxford Handbook of Cognitive Linguistics.* Oxford University Press.

Givón, T. (2013). On the Intellectual Roots of Functionalism in Linguistics. In *Functional Approaches to Language*. Walter de Gruyter.

Golbeck, J. (2015). Introduction to Social Media Investigation. Elsevier.

Golbeck, J. (2013). *Analyzing the Social Web*. Elsevier.

Gordon, M. (2013). *Labov: A Guide for the Perplexed*. Bloomsbury Academic.

Gong, T., & Shuai, L. (2016). Simulating the Effects of Cross-Generational Cultural Transmission on Language Change. In *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*. Springer Verlag.

Harm, V. (n.d.). Zur Herausbildung der deutschen Futurumschreibung mit werden + Infinitv. *Zeitschrift Für Dialektologie Und Linguistik*, *68*, 288–307.

Hartweg, F., & Wegera, Kl.-P. (2005). Frühneuhochdeutsch: Eine Einführung in die deutsche Sprache des Spätmittelalters und der frühen Neuzeit. Niemeyer.

Hilpert, M. (2017). Psycholinguistic Perspectives. In L. J. Brinton (Ed.), *English Historical Linguistics: Approaches and Perspectives.* Cambridge University Press.

Hock, H. H. (1991). *Principles of Historical Linguistics*. De Gruyter Mouton.

Hopper, P. (1987). Emergent Grammar. Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society, 13, 139–157.

Hopper, P. (1998). Emergent Grammar. In M. Tomasello (Ed.), *The New Psychology of Language.* (pp. 155–175). Lawrence Erlbaum Associates Publishers.

Jiang, J., Yu, W., & Liu, H. (2019). Does Scale-Free Syntactic Network Emerge in Second Language Learning? *Front. Psychol*, 10- 925.

Jurafsky, D; Martin, J. (2019). *Speech and Language Processing*. Retrieved from: https://web.stanford.edu/~jurafsky/slp3/15.pdf

Ke, J., Gong, T. & Wang, W. S.-Y. (2008), *Language change and social networks*. Communications in Computational Physics, *3*, 4, 935–949.

Klein, T., Wegera, K., Dipper, S., Wich-Reif, C. (2016). *Referenzkorpus Mittlehochdeutsch*. https://www.linguistics.rub.de/rem/index.html.

Kleiner, M. (1925). Zur Entwicklung der Futur-Umschreibung werden mit dem Infinitiv. *University of California Publications in Modern Philology*, *12*(1).

Kolaczyk, E. D. (2009). Statistical analysis of network data methods and models. Springer.

Kotin, M. (1998). Die Herausbildung der grammatischen Kategorie des Genus verbi im Deutschen: eine historische Studie zu den Vorstufen und zur Entstehung des deutschen Passiv-Paradigmas. Buske Verlag.

Kotin, M. (2003). Die werden-Perspektive und die werden-Periphrasen im Deutschen: Historische Entwicklung und Funktionen in der Gegenwartsprache. Peter Lang.

Kotin, M. (2009). Das Partizip II in hochdeutschen periphrastischen Verbalfügungen im 9. – 15. Jh. Zur Ausbildung des analytischen Sprachbaus. *Zeitschrift Für Germanistische Linguistik*, *28*(3), 319–345. https://doiorg.ezproxy.lib.purdue.edu/10.1515/zfgl.2000.28.3.319

Krämer, S. (2005). Synchrone Analyse als Fenster zur Diachronie. Die Grammatikalisierung von "werden" + Infinitiv. (n.d.). LIMCOM.

Kuroda, S. (1999). *Perfektkonstruktionen im Deutschen*. Buske Verlag.

Labov, W. (2001). Principles of Linguistic Change: Social Factors. Basil Blackwell.

Langacker, R. (2008). *Cognitive Grammar*. Oxford University Press.

Larsen-Freeman, D., & Cameron, L. (2008). *Complex System and Applied Linguistics*. Oxford University Press.

Lee, N.; Mikesell, L.; Joaquin, A.; Mates, A.; Schumann, J. (2009). *The Interactional Instinct: The Evolution and Acquisition of Language*. Oxford: Oxford University Press.

Leiss, E. (1985). Zur Entstehung des neuhochdeutschen analystischen Futurs. Retrieved from https://epub.ub.uni-muenchen.de/5839/1/5839.pdf

Leuschner, T. Mortelmans, T., Groodt, S. (2011). *Grammatikalisierung im Deutschen*. Walter de Gruyter.

Lightfoot, D. (1979). *Principles of Diachronic Syntax*. Cambridge: Cambridge University Press.

Liu, H.; & Hu, F. (2008) What role does syntax play in a language network? *Europhysics Letters*, 83, doi:10.1209/0295-5075/83/18002.

Madairaga, N. (2017). Understanding Grammars through Diachronic Change. *Frontiers in Psychology*, *8* (1226). https://doi.org/10.3389/fpsyg.2017.01226

Masarik, Z. (1994). Bemerkungen zum analystischen Futur im Deutschen. *Brünner Beiträge Zur Germanistik Und Nordistik*, *16*, 61–67. Matthews, P. H. (2001). *A Short History of Structural Linguistics*. Cambridge University Press.

Matas, N., Martincic-Ipšic, S., & Meštrovic, A. (2017). Comparing Network Centrality Measures as Tools for Identifying Key Concepts in Complex Networks: A Case of Wikipedia. *Journal of Digital Information Management*, *15*(4).

Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.

Mufwene, S. S. (2001). *The Ecology of Language Evolution*. Cambridge University Press.

Murray, J.; Wettin, P.; Pande, P.P.; & Shirazi, B. (2016). Sustainable Wireless Network-on-Chip Architectures. Elsevier.

Nettle, D. (1999). *Linguistic Diversity*. Oxford University Press.

Perc, M. (2012). Evolution of the most common English words and phrases over the centuries. Journal of the Royal Society, Interface, 9, 77, 3323–3328. Retrieved from: doi:10.1098/rsif.2012.0491

Saichev, A.; Malevergne, Y.; Sornette, D. (2009). *Theory of Zipf's Law and Beyond*. Berlin: Springer.

Salmons, J. (2018). *A history of German: What the past reveals about today's language*. Oxford University Press.

Salveit, L. (1961). Studien zum deutschen Futur. Die Fügungen werden mit dem Partizip des Präsens und werden mit dem Infinitiv in ihren heutigen Funktionen und in ihrer geschichtlichen Entwicklung. *Acta Universitatis Bergensis, Series Humaniorum Litterum*, *2*, 195.

Scardoni, G., & Laudanna, C. (2012). Centralities Based Analysis of Complex Networks. *New Frontiers in Graph Theory*, ISBN: 978-953-51-0115-4.

Schmidt, H. (200AD). Die Ausbildung des werden-Futurs. *Zeitschrift Für Dialektologie Und Linguistik*, *67*(1), 6–27.

Schwerdt, J. (2008). Morphosemantik der schwachen Verben im Ostgermanischen und Kontinentalwestgermanischen. Peter Lang.

Smirnova, E. (2008). Die Entwicklung der Konstruktion würde + Infinitiv im Deutschen: Eine funktional-semantische Analyse unter besonderer Berücksichtigung sprachhistorischer Aspekte. Walter de Gruyter.

Solé, R., Crominas-Murtra, B., Valverde, S., & Steels, L. (2010). Language Networks: Their Structure, Function, and Evolution. *WileyPeriodicals,Inc*, *15*(6). https://doi.org/10.1002/cplx.20305

Steels, L. (2000). Language as a Complex Adaptive System. *Proceedings of PPSN VI, Lecture Notes in Computer Science*, 17–26.

Steyvers, M.; Tanenbaum, J. (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. Cognitive Science, 29, 41- 47.

Thurner, S.; Hanel, R.; Liu, B.; Corominas-Murtra, B. (2015). Understanding Zipf's law of word frequencies through sample-space collapse in sentence formation. *Journal of the Royal Society, Interface*, 12, 108, https://doi.org/10.1098/rsif.2015.0330

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge: Harvard University Press.

Traugott, E., & Dasher, R. (2002). *Regularity in Semantic Change.* Cambridge University Press.

Vachek, J. (1966). *The Linguistic School of Prague.* Bloomington London: Indiana University Press.

Valentin, P. (1987). Zur Geschichte des Deutschen Passivs. In Das Passiv im Deutschen: Akten des Kolloquiums über das Passiv im Deutschen, Nizza 1986. Walter de Gruyter.

Vater, H. (Ed.). (1997). Zur Grammatikalisierung der zusammengesetzten Verbformen mit werden—Werden und die Modalverben im früheren Deutsch und heute. In Zu Tempus und Modus im Deutschen. WVT Wissenschaflicher Verlag Trier.

Vitevitch, M. (2008). What Can Graph Theory Tell Us About Word Learning and Lexical Retrieval? *J Speech Lang Hear Res,* 51, 2, 408-422.

Von Wright, G. (1977). Handlung, Norm und Intention: Untersuchungen zur deontischen Logik. Walter de Gruyter.

Winters, M., Tissari, H., & Allan, K. (Eds.). (2010). *Historical Cognitive Linguistics*. Berlin New York: Walter de Gruyter.

Yin, H., Benson, A., & Leskovec (2019), The Local Closure Coefficient: A New Perspective On Network Clustering. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM ’19). Association for Computing Machinery, New York, NY, USA, 303–311.

Zipf, G. (1935). The Psychobiology of Language: An Introduction to Dynamic Philology. MIT Press.