THE UNIVERSITY *of* EDINBURGH

A cross-linguistic investigation of the *way*-construction in English, Dutch, and German

Dan McColm

Submitted for the degree of

Doctor of Philosophy

The School of Philosophy, Psychology and Language Sciences

The University of Edinburgh

2019

Declaration

I declare that this thesis was composed by myself and the material contained therein has not been submitted for any other degree or qualification. The work reported is my own, except where explicitly stated otherwise in the text.

(Dan McColm)

Abstract

This thesis presents a large-scale corpus investigation into the *way*-construction (found in sentences such as *Peter made his way to the front door*) in English, Dutch, and German in a Construction Grammar framework. A cross-linguistic investigation of the *way*-construction on this scale has never been carried out; this thesis fills a gap in the literature by chronicling the development and synchronic state of the construction in each of the three languages, and presenting a cross-linguistic comparison. In Chapter 2, I present a justification for employing a Construction Grammar framework for this investigation, and also outline the diachrony and synchrony of the construction in each of the three languages. In this chapter, I also show that all three of these languages have (at least) one other construction similar in form and function to the *way*-construction, and that multiple sources have played a role in the development of the construction in each language (cf. the papers in van de Velde et al. 2013 on multiple source constructions). Chapter 3 outlines the methodology of the study and describes the corpora and statistical analysis techniques used in this study. Chapters 4 and 5 concern the role of reanalysis and analogy in the development of the *way*-construction in the three languages. In these chapters I refine some of the principles of reanalysis and analogy in light of my data on the *way*-construction, and to contribute to the debate as to whether reanalysis or analogy (or neither) is the primary mechanism of language change. I show in these chapters that reanalysis and analogy have worked in tandem (cf. Fischer 2007); the reanalysis of *way* and its Dutch and German equivalent *weg* as a non-referential object in the three languages has facilitated a long chain of analogical extensions (cf. Israel 1996). Chapter 6 deals with frequency effects and exemplar representations in the development of the *way*-construction. In this chapter I add to the growing body of work which shows that frequency effects are abundant in language, and that part of the development of the *way*-construction in the three languages can be attributed to frequency effects, and that the verbs occurring in the *way*-construction in the three languages can be grouped into exemplar clouds of semantically similar items. The role of language contact and borrowing in the development of the Dutch and German *way*-constructions is discussed in Chapter 7. I show that these concepts can be incorporated into a Construction Grammar framework (as e.g. Höder 2012 has done), and that the productivity and schematicity of the Dutch and German *way*-

constructions has increased considerably as a result of contact with English. Chapter 8 concludes the thesis.

Lay Summary

This thesis investigates the *way*-construction (found in sentences such as *Peter made his way to the front door*) in English, Dutch, and German. This investigation will be carried out in a Construction Grammar framework, a theory of language which states that language is made up of constructions, which are defined as pairings of form and meaning. This thesis chronicles the development of the *way*-construction in each of the three languages, and describes the form and function of the construction in each language today. To achieve this, this thesis makes use of corpora – huge databases of naturally-occurring sentences, to which I apply a range of statistical analyses. This thesis also explores the role of reanalysis and analogy in the development of the *way*-construction in the three languages. The term 'reanalysis' refers to the phenomenon whereby a hearer may interpret an item differently from the interpretation intended by the speaker, e.g. understanding *is going to* in *He is going to see his sister* to be a marker of future events rather than a verb of motion. The term 'analogy' is designed to capture the fact that speakers may perceive a pattern linking two items, and apply this pattern to new items which were not previously subject to that pattern, e.g. extending the plural *-s* pattern found in *dog* > *dogs* to *brother* > *brothers*, with *brothers* replacing the older plural *brethren*. In these chapters I show that both reanalysis and analogy have played an important part in how the *way*-construction has changed over time. I go on to assess the role that a change in frequency of an item plays in language change; in the case of the *way*-construction, I show that frequency does play a role. Finally, I discuss the phenomena of language contact (contact between speakers of two or more languages) and borrowing (borrowing linguistic items from a language into another language, e.g. *sushi* being borrowed into English) and show that the Dutch and German *way*-constructions changed considerably due to influence from English.

Acknowledgements

Firstly, I would like to thank the Scottish Graduate School for Arts and Humanities, without whose generous financial assistance this thesis would not have been possible. I would also like to thank Dr Graeme Trousdale and Prof Nik Gisborne for their supervision and for their valuable feedback on previous versions of this thesis. I would also like to thank Dr Timothy Colleman for his assistance and supervision during my research visit to Ghent, and to Britta Mondorf and her colleagues for their supervision during my research visit to Mainz. Last but not least, thanks also go to my friends, girlfriend, and family, especially my parents Adrian and Andrea, for their financial and emotional support while I completed this thesis.

Table of Contents

List of tables and figures

1. Introduction

1.1 Aims of the Thesis

This thesis presents a large-scale corpus investigation of the *way*-construction in English, Dutch, and German. Jackendoff (1990: 223) calls the construction a "fairly outrageous mismatch between the syntax and semantics"; the thesis aims to answer why this is the case. This investigation is novel in both scope and method; the total size of the corpora used in this investigation is over 8 billion words, and this investigation also uses a combination of statistical analyses that have not been applied to the *way*-construction in all three of these languages. An example of the construction in each of the three languages is given in (1) to (3) below.

(1) Babe Ruth homered his way into the hearts of America (Jackendoff 1992: 219, ex. 24a).

(2) Zo blufte zij zich een weg uit Auschwitz[1]
Thus bluffed she REFL a way out Auschwitz
'That's how she bluffed her way out of Auschwitz' (Verhagen 2003a: 33, ex. 10).

(3) Er grub sich seinen Weg aus dem Gefängnis
He dug REFL POSS way out the prison
'He dug his way out of prison' (Ludwig 2005: 10, ex. 27a)

This thesis aims to answer the following research questions:

a) What does the *way*-construction look like in the three languages today, and what was the diachronic development of the construction in each language?
b) What can corpora and statistical analysis of corpus data tell us about the development of the English, Dutch, and German *way*-construction?
c) What led to the constructionalizaiton of the *way*-construction in each language, and what post-constructionalization constructional changes took place in each language?

[1] In this thesis, I gloss all examples apart from those in Modern English (i.e. 16th Century and later). Unless otherwise stated, all translations and glosses are my own.

d) What was the role of analogization and analogical extension in the development of the *way*-construction in the three languages?
e) Is there evidence for frequency effects and exemplar representation of language in the *way*-construction in any of the three languages?
f) What is the role of borrowing in the Dutch and German *way*-construction, and how can a constructional theory of language accommodate this borrowing?

1.2 Constructionist approaches to language

The analysis in this thesis will be couched in a Construction Grammar framework. Although there are a number of different theories of language which can be grouped under a broad 'Construction Grammar' umbrella, such as Berkeley Construction Grammar (see e.g. Fillmore 1988), Sign-Based Construction Grammar (see e.g. Boas & Sag 2012), Fluid Construction Grammar (see e.g. Steels 2011), Embodied Construction Grammar (see e.g. Bergen & Chang 2005), Cognitive Grammar (Langacker 1986), Radical Construction Grammar (Croft 2001), and Cognitive Construction Grammar (Goldberg 1995, 2006), these approaches share several common tenets.

Firstly, constructionist approaches to language all assume that the entirety of a speaker's knowledge of language consists of constructions, which are defined as learned pairings of form and function (Goldberg 2013: 15). The formal pole of the construction contains phonological, morphological, and syntactic information, while the functional pole contains information about the semantics, pragmatics, and discourse properties of the construction (see e.g. Croft 2001). While earlier definitions stated that non-compositionality was criterial to constructionhood (i.e. that the meaning of a construction cannot be predicted from the sum of its parts (Goldberg 1995: 4)), fully compositional items are now considered constructions provided they occur sufficiently frequently (Goldberg 2006: 5).

Constructions differ in their complexity and schematicity. The term 'complexity' refers to the degree to which a construction can be split into smaller meaningful parts. Thus *cat* is a minimally complex construction, because it cannot be broken down into any further meaningful units, while *I saw John in town yesterday* is more complex as it contains (at least) the following smaller constructions: *I*, *saw*, *John*, *in*, *town*, and

yesterday. The term 'schematicity' refers to the degree to which a slot in a constructional schema can recruit items (Croft & Cruse 2004: 255). A construction such as *by and large* is not at all schematic, because no other items may occupy any of the slots (e.g. **by and big*), while the ditransitive construction NP_i V NP_j NP_k is maximally schematic, because no slot in the schema is filled phonologically.

Secondly, all constructionist approaches to language have in common that they posit no transformational or derivational rules which turn one structure into another (as set out for example in Chomsky 1957 and subsequent works). In constructionist approaches to language, the formal and functional poles of a construction are directly linked (Goldberg 2013: 15). Thirdly, all constructionist theories of language assume that each construction constitutes a node in the mental constructional network of the speaker (this network is sometimes known as the 'constructicon' in constructionist literature). Nodes in the constructional network are linked by inheritance links; for instance, the P N construction (a preposition followed by a bare noun, e.g. *to bed*) inherits its word order from the more general prepositional phrase construction (e.g. *to the shop*). These nodes may change over time; when this occurs, 'constructional changes' are said to have taken place (Traugott & Trousdale 2013: 1). Constructional changes can affect all dimensions of a construction, as well as its collocational preferences (ibid.). As well as affecting existing nodes, change may also lead to new nodes being created in the network; this is known as constructionalization, i.e. the creation of a $form_{new}$-$meaning_{new}$ pair (ibid.).

Fourthly, constructionist approaches seek to explain the variability within and between languages in terms of domain-general cognitive processes (see e.g. Croft 2001, Evans & Levinson 2009), rather than positing some kind of innate grammar module in the mind (as in e.g. all versions of Chomskyan theory).

1.3 On the importance of using frequency data and corpora

This thesis makes extensive use of corpora[2], frequency data, and statistical analyses applied to these data. The diachronic development of the *way*-construction in the three languages can only be ascertained by using corpora. To this end, I use a

[2] The corpora used in this study are described in more detail in Chapter 3.

combination of corpora covering a wide temporal range and diversity of genres; in English, the corpora used are ARCHER (1650-2000), BNC (late 20th century), CLMET3.1 (1710-1920), COCA (1990-2017), COHA (1810-2009), and Helsinki (8th century-1710). Because the ARCHER and Helsinki corpora returned relatively few tokens of the *way*-construction, these corpora were used for qualitative analysis only. These corpora cover a wide range of genres; from more informal, spoken texts in COHA to translations of the Bible in the Helsinki corpus.

The Dutch dataset consists of material from three corpora: *De Gids* (1837-1936), whose texts are taken from the literary periodical of the same name, SoNaR (1954-2011), a corpus of Netherlandic and Belgian Dutch covering a range of both informal and formal written genres, and NLCOW14A – a Dutch web-based corpus containing very informal texts. This combination of corpora therefore ensures genre diversity and wide time span.

The German corpora used cover the period from the late 15th century to the present day, as well as a variety of genres. The *Deutsches Textarchiv* is a large diachronic corpus whose texts were written between 1473 and 1927 and are typically highly formal. The *Berliner Zeitung* (1994-2005) is a newspaper corpus, and the DECOW16A-NANO corpus is a very modern web-based corpus containing informal texts.

For much of the twentieth century, information about the frequency of words was considered irrelevant to the study of linguistic structure (Bybee 2007: 5). However, there is abundant experimental evidence that speakers know the relative frequencies of the words, constructions, collocations, and all the other elements of their language (see, among many others, Ellis 2002; Diessel 2007; Robinson & Ellis 2008; Arnon & Snider 2010; Divjak & Caldwell-Harris 2015).

The importance of using frequency data has been amply demonstrated in the literature. Frequency data have been used to answer research questions in a wide range of linguistic fields, from phonology and morphology to first language acquisition and word recognition. Indeed, Hilpert and Diessel (2016: 5) find frequency effects in virtually all aspects of language, including a) the emergence of collocations and syntactic constituents, (b) the interaction between lexemes and constructions, (c) the

productivity of linguistic schemas, (d) the ability of language users to assess the grammaticality of novel linguistic forms (Lapata et al. 1999; though see also Newmeyer 2003, who denies any correlation between grammaticality and frequency), (e) the occurrence of phonetic reduction and coalescence in language change, (f) the segmentation of the speech stream (see also Aslin & Newport 2012); (g) the extraction of syntactic categories in L1 acquisition (see also Redington et al. 1998), (h) the maintenance of frequent linguistic strings under pressure from analogy, (i) the choice between alternative structures in language production, (j) the processing of the unfolding sentence in language comprehension (see also Pierrehumbert 2006; Diessel 2007), and (k) the flagging or marking of infrequent forms. This suggests that relying on frequency information from corpora can offer considerable insight into virtually all aspects of language, which would be impossible solely by relying on introspective acceptability judgements or experimental data.

This thesis builds on previous accounts of the *way*-construction by making use of frequency data; frequency data is often missing from studies on the *way*-construction (e.g. Israel 1996; Traugott & Trousdale 2013; Fanego 2017, 2018). Both type and token frequency are considered in this study. The token frequency of a construction is the number of times it occurs, and the type frequency of a construction refers to the number of distinct items that may occupy a given slot of this construction (Divjak & Caldwell-Harris 2015: 55). This study uses token and type frequency data to answer questions on the connection between these two variables and entrenchment, prototypes and exemplars, and the role of analogy in the development of the *way*-construction in the three languages. In particular, I examine whether there is a correlation between token frequency and entrenchment, whether the prototypical item in a construction is always the most frequent, and whether the likelihood of analogical extension taking place is dependent on the token frequency of previously experienced similar items (cf. Israel 1996), or dependent on the correlation coefficient between the token frequency of similar items in the item's semantic domain and the type frequency of novel verbs in that domain (cf. Zeschel 2010). I show that the relationship between token frequency and entrenchment is not as crude as is sometimes assumed (e.g. Gries 2014b), that the prototypical item in a construction is not always the most frequent, as is often claimed (cf. Gilquin 2006), and that there is a strong correlation between the likelihood of analogical extension of an item and the token frequency of similar items in its semantic domain.

1.4 Structure of the remainder of the thesis

The remainder of the thesis is structured as follows. In Chapter 2, I outline the synchronic state of the *way*-construction in the three languages, and present a justification for analysing the *way*-construction through a constructionist lens. I also give an account of the diachronic development of the construction in each language. Chapter 3 describes the methodology of the thesis, listing the corpora used, presenting a justification for using this combination of corpora, and explaining the statistical methodology behind the quantitative analysis employed in this thesis. Chapter 4 concerns the role of reanalysis in the development of the *way*-construction, and discusses whether traditional accounts of reanalysis provide an adequate explanation of changes to the *way*-construction in each language. In Chapter 5, I discuss the role of analogy in the development of the *way*-construction, and attempt to refine some of the principles of analogy in order to give a constructionist account of the development of the *way*-construction. Chapter 6 deals with the role of frequency and exemplar representations in the *way*-construction. In this chapter, I determine whether part of the development of the *way*-construction in English, Dutch, and German can be explained in terms of frequency effects, or whether other variables such as genre may play a role. I also test Bybee's (2013) hypothesis that the items that can occur in a construction can be grouped into two or more exemplar clouds of semantically similar items. Chapter 7 discusses the role of language contact and borrowing in the development of the Dutch and German *way*-constructions. I also discuss how these phenomena can be integrated into a constructionist framework, focusing on Diasystematic Construction Grammar (see especially Höder 2012). In this chapter, I test the hypothesis that the 'incidental activity' reading (where the verb encodes an activity not causally related to the subject's motion) of the Dutch and German *way*-constructions is a borrowing from English. Chapter 8 concludes the thesis by answering the research questions posed in Section 1.1 and outlining possibilities for further research on the *way*-construction and related phenomena.

2. Synchrony and diachrony of the *way*-construction in English, Dutch, and German

2.1 Introduction

2.1.1 Aims of the chapter

This chapter of the thesis is concerned with the first research question posed in Section 1.1: what does the *way*-construction look like in the three languages today, and what was the diachronic development of the construction in each language? An example of the construction in each language was given in the introduction; these are repeated here for convenience as (4) to (6).

(4) Babe Ruth homered his way into the hearts of America (Jackendoff 1992: 219, ex. 24a).

(5) Zo blufte zij zich een weg uit Auschwitz
Thus bluffed she REFL a way out Auschwitz
'That's how she bluffed her way out of Auschwitz' (Verhagen 2003a: 33, ex. 10).

(6) Er grub sich seinen Weg aus dem Gefängnis
He dug REFL POSS way out the prison
'He dug his way out of prison' (Ludwig 2005: 10, ex. 27a)

In this chapter, I summarise the similarities and differences of the construction in each of the three languages, and explain its synchronic behaviour. The remainder of the chapter is structured as follows. In Section 2.1.2, I present a justification for treating the English, Dutch, and German *way*-constructions as constructions in the Construction Grammar sense (see e.g. Goldberg 1995, 2006). Section 2.2 concerns the English *way*-construction. Section 2.2.1 clarifies the terminology relating to the three readings of the *way*-construction; the construction is often divided into the 'means', 'manner', and 'incidental activity' readings (e.g. Goldberg 1995; Israel 1996), but these terms are used inconsistently. The diachrony and synchrony of the English

way-construction are outlined in Sections 2.2.2 and 2.2.3 respectively. In Section 2.2.3 I also outline the formal schema of the construction and discuss its syntactic and semantic constraints, and discuss different sub-constructions of the *way*-construction, with a focus on Israel's (1996) three sub-constructions and Fanego's (2018) five sub-constructions. In Section 2.2.4 I discuss constructions that are formally and/or functionally similar to the *way*-construction, in particular the so-called fake reflexive resultative construction (henceforth FRR; cf. Simpson 1983). Section 2.2.5 presents the *way*-construction as a polysemy network in light of the ideas discussed in Section 2.2.3 and 2.2.4. A description of the Dutch *way*-construction is presented in Section 2.3, including its diachronic development (Section 2.3.1), its formal schema and constraints (Section 2.3.2), and the similarities and differences to the English *way*-construction (Section 2.3.3). In Section 2.3.4 I discuss functionally and formally similar constructions in Dutch. Section 2.4 deals with the German *way*-construction. In Section 2.4.2 I outline the form and function of the synchronic German *way*-construction, and describe the similarities and differences to the Dutch and English *way*-constructions. Related constructions in German are discussed in Section 2.4.3. Section 2.5 concludes by enumerating the similarities and differences between the construction in all three languages. In this conclusion, I show that all three languages have (at least) two constructions which are very similar in form and function to the *way*-construction.

2.1.2 Previous theoretical approaches to the *way*-construction

The *way*-construction presented a serious challenge to generative theories of language because, as Jackendoff (1990: 223) puts it, there is a "fairly outrageous mismatch between the syntax and semantics" of the construction. Jackendoff attempted to solve this problem by positing that the construction consists of a simple lexical entry, differing only in the specification of a possessive pronoun, the noun *way*, and an open head. Israel (1996) points out that the head is not necessarily open, because not all verbs are permissible in the construction, as will be shown in later sections. Marantz (1992: 181) approached the problem by calling the *way*-construction a set of transitivizing operations, but according to this analysis, intransitive motion verbs that appear in the construction, such as *go* and *wend*, have a transitive sense that occurs only in the *way*-construction, which is implausible, because the phrase headed by the possessive pronoun and *way* is not a true object

of the verb, as will be shown in later in this chapter; this casts doubt on the transitivity of *go* and *wend* and the other intransitive verbs of motion which can appear in the construction.

Instead, as mentioned in the introduction, I will investigate the *way*-construction through a constructionist lens. The *way*-construction was arguably the first set of historical changes to be investigated from a Construction Grammar perspective (Israel 1996), and it has since been the focus of a considerable amount of literature in constructionist frameworks (see also, among many others, Gisborne & Patten 2011; Goldberg 1995, 2006; Mondorf 2011; Traugott & Trousdale 2013; Fanego 2017, 2018; Perek 2018).

Fully compositional items are now considered constructions as long as they occur with sufficient frequency (Goldberg 2006: 5). More restrictive views of the term 'construction' suggest that, in order to account for the need of a particular construction, it must be shown that independently motivated principles fail to predict all of the facts about the use, i.e. that it is non-compositional (Michaelis 2013: 140). In the following paragraphs I will show that, even according to this more restrictive definition, the *way*-construction in English, Dutch, and German exhibits idiosyncratic behaviour which suggests non-compositionality, and therefore that the *way*-construction is an independent construction in each of the three languages.

The first piece of evidence for the non-compositionality of the construction is that in many cases, the post-verbal argument cannot plausibly be licensed by the verb, as the following examples show.

(7) We talked our way into the VIP area (Perek 2018: 66, ex. 3)

(8) De mens gokt zich een weg door 't leven
The person gambles REFL a way through the life
'This person gambles his way through life'[3]

[3] https://forum.politics.be/archive/index.php?t-113523.html, last accessed 19th September 2018.

(9) Mann pinkelte sich den Weg aus Lavine
Man peed REFL the way out avalanche
A man peed his way out of an avalanche' (Ludwig 2005: 11, ex. 33).

In addition, some of the verbs which occur in the construction cannot occur with both PP and NP complements, as the following examples show. This suggests that, in the *way*-construction, the VP's complement structure is determined by the composite effects of the verb and the construction, rather than by the verb alone (Jackendoff 2002: 176; Goldberg & Jackendoff 2004: 534; cf. also Michaelis 2003 on coercion[4]).

(10) Bill belched his way out of the restaurant (adapted from Perek 2018: 66, ex.5a).

(11) *Bill belched (a belch) out of the restaurant (adapted from Perek 2018: 66, ex. 5a).

(12) Eet en drink jezelf een weg door de lekkerste
Eat and drink REFL a way through the most.delicious
culinaire bestemmingen
culinary destinations
'Eat and drink your way through the most delicious culinary destinations' (NLCOW, date unknown)

(13) *Hij drinkt (bier) zich naar Amsterdam.
He drinks (beer) REFL to Amsterdam.
'*He drinks (beer) to Amsterdam'

(14) Sue pfiff sich einen Weg durch den Tunnel
Sue whistled REFL a way through the tunnel
'Sue whistled her way through the tunnel' (adapted from Ludwig 2005: 14, ex. 43)

(15) ??Sue pfiff (eine Melodie) sich durch den Tunnel
Sue whistled (a melody) REFL through the tunnel
'??Sue whistled (a melody) through the tunnel' (adapted from Ludwig 2005: 14, ex. 43)

[4] Defined as the process whereby a meaning is attributed to a construction that is absent from the syntax.

(15) is at best highly unusual; it is completely ungrammatical unless whistling is the means by which Sue moves through the tunnel, which is unlikely. The variant with the NP complement is also ungrammatical, even if the NP complement appears after the reflexive pronoun.

Further, the motion interpretation of the construction cannot be attributed to any of its individual components. Goldberg (1995: 199) shows that the motion interpretation of the English *way*-construction is available regardless of whether the verb encodes motion. The motion interpretation of the construction cannot be attributed to the noun *way* in isolation, because (16) and (17) differ only in the choice of determiner and (16) encodes motion while (17) does not. The motion interpretation of the English construction is also not available when the noun *way* is replaced by other semantically similar nouns, as (18) and (19) show.

(16) *Frank found his way to New York, but he hasn't gone yet. (Perek 2018: 66, ex. 6a).
(17) Frank found a way to New York, but he hasn't gone yet (Perek 2018: 66, ex. 6b).
(18) *Frank dug his way out of prison, but he hasn't gone yet (Perek 2018: 67, ex. 7a).
(19) Frank dug his escape route out of prison, but he hasn't gone yet (Perek 2018: 67, ex. 7b).

A similar constraint exists on the Dutch and German constructions. The Dutch and German constructions entail motion; however, both languages also have a benefactive construction with *weg* 'way' and *banen*/*ebnen* 'pave' that differs in the choice of determiner and which does not entail motion, because the beneficiary does not necessarily traverse the path (cf. Verhagen 2003a, 2003b), e.g. *jemandem den Weg ebnen* in German 'to pave the way for somebody', and its Dutch equivalent *de weg voor iemand banen*.

Further, the phrase headed by *way*/*weg* is not a semantic argument of the verb, despite appearing in direct object position (Jackendoff 2002: 174). This phrase cannot be pronominalized (20-22), ellipted (23-25), questioned (26-28), or clefted (29-31).

These sentences also show that this phrase is discursively inert (cf. König 1999), because it cannot be referred back to by pronouns.

(20) *I elbowed my way through the crowd and then pushed it to the bar.

(21) Hij baande zich een weg naar Utrecht
He made REFL a way to Utrecht
en dan baande het naar Breda.
and then made it to Breda
‘*He made his way to Utrecht and then made it to Breda.'

(22) *Er bahnte sich einen weg nach München
He made REFL a way to Munich
und dann bahnte ihn nach Augsburg.
and then made it to Augsburg.
‘*He made his way to Munich and then made it to Augsburg.'

(23) *I elbowed my way through the crowd and she whistled hers to the pub.

(24) *Hij baande zich een weg naar de deur
He made REFL a way to the door
en ik baande me een naar de slaapkamer
and ik made REFL a to the bedroom
'*He made his way to the door and I made mine to the bedroom'

(25) *Ich bahnte mir einen Weg zum Haus
I made REFL a way to.the house
und er bahnte sich einen in die
and he made REFL a into the
'*I made my way to the house and he made his to town'

(26) *What did he elbow through the crowd? His way.

(27) *Wat heeft hij door de menigte gebaand? Zijn weg
What has he through the crowd made his way
'*What did he make through the crowd? His way.'

(28) *Was hat er durch die Menge gebahnt? Seinen Weg.
What has he through the crowd made his way
'*What did he make through the crowd? His way.'

(29) *It was his way that he made through the crowd.

(30) *Het was mijn weg die ik baande
It was my way that I made
'*It was my way that I made'

(31) *Es war der Weg, der er sich bahnte
It was the way that he REFL made
'*It was his way that he made'

This contrasts with fully compositional sentences of English such as *John kicked the ball into the goal*, and their Dutch and German equivalents, where the phrase in direct object position is referential. Unlike the phrase in the *way*-construction, such phrases can be pronominalized, ellipted (32-34), questioned (35-37), and clefted (38-40).

(32) John kicked the ball into the goal and the goalkeeper picked it out of the net.
(33) John schopde de bal in het doel en de keeper pakte het op uit het net.
(34) John schoss den Ball ins Tor und der Torwart holte ihn aus dem Netz.
(35) What did he kick into the goal? The ball.
(36) Wat schopde hij in het doel? De bal.
(37) Was schoss er ins Tor? Den Ball.
(38) It was the ball that John kicked into the goal.
(39) Het was de bal, die John in het doel schopde.
(40) Es war der Ball, den John ins Tor schoss.

Further, the *way*-phrase cannot be passivised in any of the three languages, as (41) and its Dutch and German equivalents in (42) and (43) show, while phrases in direct object position in fully compositional sentences can, as the same sentence in the three languages shown in (44) to (46) proves. (Jackendoff 1990: 216, 1997: 546). These properties suggest that the noun *way*/*weg* is a meaningless syntactic marker of the construction (cf. Jackendoff 1992: 167), rather than naming a path, as Marantz (1992: 180) claims.

(41) *His way is being made through the crowd.
(42) *Zijn weg werd door de menigte gebaand.
(43) *Sein Weg wird durch die Menge gebahnt.
(44) The ball was kicked into the goal by John.

(45) De bal werd door John in het doel geschopd.

(46) Der Ball wurde von John ins Tor geschossen.

This view of *way* as a meaningless syntactic marker is challenged, however, by Goldberg (1996: 38), who points to examples where *way* is modified by an adjective, arguing that being meaningful is a prerequisite for modification. However, in sentences such as (47), the adjective *miserable* describes Bill or the entire verbal subevent, rather than the path itself (Jackendoff 1992: 167). Sentences such as (48), where the adjective does describe the path, are only very rarely attested in my dataset[5], and have all but fallen out of use except as an archaism.

(47) Bill belched his miserable way out of the restaurant.

(48) The Roman emperor explored his perilous way through the Black Sea (CLMET3.1, 1776).

In sum, this subsection has shown that the English, Dutch, and German *way*-constructions all exhibit non-compositional properties and can therefore be considered constructions, even under the more restrictive definition of the term (as put forward by e.g. Michaelis 2013). The following section provides a partial answer to the first research question posed in my introduction by describing the synchrony and diachrony of the English *way*-construction. In this section, I divide the construction into three distinct threads: the means reading, the manner reading, and the incidental activity reading. I show that the English *way*-construction as it is found today emerged as a blend of two precursor constructions (cf. Traugott & Trousdale 2013). I also discuss in this section some of the semantic and syntactic constraints on the *way*-construction in English.

[5] The composition of my dataset is outlined in Chapter 3.

2.2 The English *way*-construction

2.2.1 Means, manner, incidental activity: some necessary terminological clarifications

The present-day English *way*-construction (i.e. a verb followed by a possessive pronoun, the noun *way*, and a PP describing a path) is sometimes divided into three readings: 'means', 'manner', and 'incidental activity' (e.g. Israel 1996; Perek 2018). However, the terms 'means' and 'manner' have been used inconsistently in the literature. Goldberg (1995, 1997), Goldberg and Jackendoff (2004), and Szczesniak (2013) use the term 'means' for verbs such as *walk*, *float*, and *crawl*, because "most conflation patterns involving 'manner' verbs imply that the particular manner is the means of motion [...] *The bottle rolled down the hill* entails not only that 'the bottle moved down the hill while rolling' but also that 'the bottle moved down the hill by rolling'" (Goldberg 1995: 232, fn. 20). Goldberg uses the term 'manner' for sentences such as *I knitted my way to fame and fortune*, because these verbs encode the manner, not the means of motion. Israel (1996), on the other hand, uses these terms in exactly the opposite way. He uses the term 'means' for verbs such as *furrow out* and *plough*, and 'manner' for verbs such as *walk* (cf. also Fanego 2017: 7, fn. 3). In the following subsections, I adopt Israel's terminology, using the term 'means' for sentences such as (49), 'manner' for sentences such as (50), and 'incidental activity' for sentences such as (51). This is because digging is the means by which Rasselas exits the valley in (49), and *limped* in (50) describes the manner in which the soldiers move. *Giggled* in (51) is clearly an incidental activity, because it cannot be the means by which the subject moves up the stairs, nor can it describe the manner of motion.

(49) Rasselas dug his way out of the happy valley.
(50) The wounded soldiers limped their way across the field.
(51) Convulsed with laughter, she giggled her way up the stairs.
(Israel 1996: 218, exx. 1-3).

This tripartite division of the *way*-construction has been criticised by Fanego (2018: 3-4), who argues that it cannot account for the complexity of the construction's development. Fanego (2018) therefore follows Goldberg (1997) and establishes a finer-grained division, splitting the *way*-construction into five distinct threads;

elaboration (basic motion verbs), means reading, manner reading, sound emission verbs (where the sound is emitted as a result of the motion), and the incidental activity reading. Fanego is no doubt correct that the construction is very complex, but there is in principle no end to the number of threads the construction could be divided into. The incidental activity reading alone contains several further semantic categories, including verbs of ingestion, verbs of performance, and verbs encoding bodily functions (Perek 2018: 78). As Fanego herself concedes (2018: 24), verbs of elaboration are now very rare in the construction, as are verbs of sound emission. For these reasons, and to facilitate cross-linguistic comparison, I follow the three threads established by Israel (1996) (as followed by, among others, Goldberg 1995 and Perek 2018), as verbs of elaboration are very rare in the Dutch and German constructions, with the exception of *banen*/*bahnen*, and verbs of sound emission (in Fanego's sense, i.e. the sound is the result of the motion) are not attested at all in the German construction.

2.2.2 The diachrony of the English *way*-construction

2.2.2.1 The emergence of the English *way*-construction

As mentioned above, Israel (1996: 221) identifies three independent threads of the *way*-construction. He dates these from the 14th century to the present day. He further states that all three of these threads were motivated by the lexical semantics of *way*, and each underwent a series of local analogical extensions.[6] Goldberg (1995: 218), on the other hand, argues that the *way*-construction is a 'conventionalized amalgam of two constructions: the creation construction and the intransitive motion construction'. Traugott and Trousdale (2013: 79-80) also identify two precursor constructions. Both of these contained the noun *way*, though one was transitive and the other was intransitive.

Fanego (2018: 12) points out that these precursor constructions occur as early as Old English, and that they served to translate the Latin *via* 'road, path', for example *viam ferro patefacere* 'to make one's way with the sword'. In the intransitive construction, the noun *wei* and phrases such as *on wei* could combine with a wide range of verbs

[6] The analogical extensions of the *way*-construction in the three languages are discussed in more detail in Chapter 5.

denoting locomotion and forward progress (Traugott & Trousdale 2013: 79-80). These verbs include *go*, *wend*, *fare*, *flee*, and *ride* (ibid.). In the transitive construction, the verb used was usually a verb of acquisition such as *niman* 'take', although Fanego (2018: 13) finds that *aredian* 'reach', *geceosan* 'choose', *don* 'do', *ryman* 'clear', and *(ge)wyrcan* 'make, do, perform' were also used in this precursor construction. Fanego (2012) also finds that nouns semantically similar to *way* were used in this transitive precursor construction, such as *pas* 'step' (cf. Traugott & Trousdale 2013: 82, ex. 36a), and argues that this transitive construction was the template for the modern-day *way*-construction.

The noun *way*, and other semantically similar nouns, function as referential objects in both the transitive and intransitive precursor constructions. Traugott and Trousdale (2013: 80-82) point out that *way* could occur with an article, a demonstrative, in an object inversion construction, with a preposition, with the noun *way* in the plural, and undergo adjectival modification. They therefore hypothesise (2013: 82) that this transitive acquisition construction with *way* or *pas* are subschemas of the more general transitive construction, in which they function as a spatial object. There was therefore no *way*-construction as it is found today by this point, only a transitive motion construction that involved the noun *way*.

By the 16th century, an unergative construction emerges with basic motion verbs such as *go* and *come*. In this construction, there was a possessive pronoun rather than a definite article, no preceding preposition attached to *way*, and pluralisation of *way* had fallen out of use by this point; a directional argument was also possible in this construction (Traugott & Trousdale 2013: 83). These facts suggest that a subschema of the intransitive construction was emerging, consisting of a basic motion verb, a possessive pronoun, and the noun *way*, meaning 'go along a path' (ibid.).

Israel (1996: 221) dates the modern *way*-construction to the late Middle English period, arguing that it was in use as early as the 15th century and that it was well established by the 17th century. However, he appears to conflate the precursor constructions with the *way*-construction itself; the example he gives here does not contain a directional argument, which is now obligatory in the construction.

(52) The moving legions speed their headlong way (Israel 1996: 227, ex. 12).

Traugott and Trousdale (2013: 84), on the other hand, date the modern *way*-construction to the end of the 17th century. They argue that, by this point, a *way*-construction with transitive verbs had emerged, with the formal schema [[$SUBJ_i$ $V_{TRcausative}$ $POSS_i$ *way* DIR]] and the meaning 'SEM_i cause to traverse a path'. They go on to state that this construction was independent of the more general transitive construction, but still closely related to it because the verbs sanctioned by both constructions were transitive. By this time, the noun *way* no longer functioned as a fully referential object, but was a fixed part of a constructional schema with a directional argument (Traugott & Trousdale 2013: 84). Fanego (2018: 15), however, points out that the obligatorification of the directional argument had taken place by Middle English, and that Traugott and Trousdale's dating is therefore far too late.

2.2.2.2 The means reading

The means reading ('path-creation sense', in Perek's (2018) terminology) of the construction encodes motion, even though there may not be a verb of motion in the construction (cf. Jackendoff 1990; Goldberg 2006). Israel (1996: 223) dates the means reading to the end of the 16th century. He shows that verbs such as *pave* and *smooth* emerged by 1650, with verbs of path clearing and path creation such as *cut*, *furrow out*, and *eat out* emerging soon after. However, as the examples below show, Israel is again conflating the precursor constructions with the *way*-construction itself; only (55) contains a directional argument.

(53) Like as a fearefull Dove, which through the raine Of the wide ayre her way does cut amaine[7] (1590).
(54) Arminius paved his way first by aspersing and sugillating the fame and authority of Calvin (1647).
(55) Bacon was one of those that smoothed his way to a full ripeness by liquorish and pleasing passages (1653).
(Israel 1996: 224, exx. 20-22).

However, Fanego (2017: 16) argues that the means reading emerged much earlier than this, citing the following example from ca. 1325; but *make* in this example is a

[7] Amain(e) = 'with all one's strength'.

highly general verb, and is not the equivalent of *struck his way*, which would be a more typical means verb.

(56) Corineus [...] harde smot [...] & made is way bi eiþer side
Corineus hard struck & made his way by every side
Corineus struck hard and made his way in every direction' (Fanego 2017: 16, ex. 21a)

In the mid-17th century, more verbs of path building and path creation begin to be used in the means reading. These verbs include *bridge*, *chalk out*, and verbs encoding cutting or clearing a path such as *plough* and *dig* (Perek 2018: 78). By 1770, a cluster of verbs encoding fighting emerges, possibly by analogical extension of *force*.[8] This cluster of verbs increases in productivity by 1875, with *shoot*, *push*, and *struggle* attested by this date (ibid.). From around 1880, verbs emerge related to bodily functions such as ingestion (*eat*, *drink*, *nibble*), olfaction (*scent*, *smell*), actions involving the mouth (*bite*, *chew*, *gnaw*), as well as verbs relating to commerce and finance (*borrow*, *buy*, *export*) (ibid.). Around the same period, verbs describing various types of attack, coercion and misconduct begin to emerge, such as *bribe*, *bully*, *cheat*, *conspire* and *kill* (ibid.). Increasingly indirect ways of reaching a goal also emerge in the mid-19th century, as (57) shows (Israel 1996: 224-225). Notably, the activity encoded by the verb does not necessarily entail the endpoint.

(57) Cattermole...now prostitutes his talent...and blots his way to emolument and oblivion (1844; Israel 1996: 225, ex. 29).

Goldberg (1996: 35) considers the means reading to be the more central reading of the *way*-construction. In the means reading, the subject creates a path for himself and moves along this path. This can involve the removal of literal obstacles, or merely enabling some kind of metaphorical motion, as in (57) above. The means reading largely consists of verbs encoding forceful actions leading to the creation of a path despite an obstacle, although as pointed out above, an increasing number of verbs are used which do not typically describe the creation of a path, at least not a literal one. When a metaphorical path or metaphorical motion is encoded, difficulty or an

[8] This hypothesis is tested in more detail in Chapter 5.

obstacle must be present, as the difference in grammaticality between (58) and (59) shows.

(58) ??Sally drank her way through a glass of lemonade (Goldberg 1996: 36, ex. 15a).
(59) Sally drank her way through a case of vodka (Goldberg 1996: 36, ex. 15b).

2.2.2.3 The manner reading

Israel (1996) claims that the manner reading originated from a *go-your-path* construction with simple verbs of motion such as *go*, *pass*, and *fly* (cf. Israel 1996: 220, exx. 4-7). However, these verbs of motion mostly occurred with no directional argument, which is now obligatory; this leads Traugott and Trousdale (2013: 89) to suggest that such verbs were marginal to the construction. Verbs found in the manner reading today encode the manner in which the subject moves, and this motion must be through a literal or metaphorical self-created path (Fanego 2017: 15). These verbs typically encode the speed of motion (*inch*, *run*), aspects of the terrain in which the motion takes place (*swim*, *wade*), the shape of the path (*thread*, *wend*), or laborious motion (*stagger*, *trudge*) (Perek 2018: 68). Verbs of body-internal motion such as *flap* are also attested (Rappaport Hovav & Levin 1995: 226-227).

Goldberg (1996: 35) points out that basic verbs of motion such as *walk* and *run* are now unacceptable in the construction because the subject must move despite some external difficulty.[9] However, the addition of difficulty does not make these basic motion verbs more felicitous, as the following example shows.

(60) ?With extreme difficulty and pain, he walked his way through the stinging nettles (Szczesniak 2013: 178, ex. 12).

Before 1700, manner verbs were not widely attested in the construction; Israel (1996: 222) finds 16 types, although given his conflation of the precursor constructions with the *way*-construction itself, the true number of manner verbs may in fact be smaller than this (cf. also Fanego 2018). After 1700, Fanego (2012) documents a steep

[9] These verbs are very rarely attested in my data, and only when the author is attempting to replicate an archaic style.

increase in the frequency of manner verbs, with 250 innovative manner verbs attested between 1700 and 1900. This is unsurprising; as a satellite-framed language, English characterises motion by using manner verbs, and speakers constantly expand this domain by using increasingly more precise verbs in the manner reading; for example, *walk* involves the subject using his legs to move at a normal pace, but *saunter* encodes the leisurely manner of this motion (Slobin 2004, 2006). Perek (2018) also chronicles a steep increase of verbs describing clumsy or unsteady motion, such as *scramble* and *stumble*. Verbs of body-internal motion such as *wriggle* and *writhe* then join the construction (ibid.). From 1880, verbs which encode motion without difficulty are attested, such as *pace*, *fly* and *ooze*. Verbs of walking (*shuffle*, *walk*, *tiptoe*), verbs of rapid motion (*dash*, *run*, *speed*), verbs of aquatic motion (*swim*, *wade*), verbs of manner of motion relating to a specific type of vehicle (*drive*, *fly*, *glide*), and verbs of fluid motion (*drip*, *ooze, pour*) all join the manner reading of the construction at around this time (ibid.).

2.2.2.4 The incidental activity reading

The incidental activity reading (co-occurrence relation, in Fanego's (2018) terminology) of the *way*-construction is a relatively modern innovation, and is not accepted by all speakers; Israel (1996) considers it marginal (see also Asudeh et al. 2008: 268). In the incidental activity reading, the verb occurs at the same time as the motion, but is not causally related to it (Perek 2018: 68); the *way*-construction imposes the meaning of 'move while V-ing' by coercion (see e.g. Michaelis 2003).

The incidental activity reading has undergone a considerable increase in frequency since 1900. Traugott and Trousdale (2013: 88) suggest that it is the most productive use of the *way*-construction today (see also Szczesniak 2013: 160). Verbs in the incidental activity reading may encode ingestion (*drink*, *eat*), actions involving the mouth (*puff*, *smoke*), performance (*dance*, *play*, *sing*), and bodily functions (*cough*, *sneeze*, *sweat*) (Perek 2018: 78). Towards the middle of the 20th century, clusters of verbs encoding social interaction (*chat*, *chatter*, *grin*, *joke*, *kid*, *laugh*) and some verbs relating to cognition (*think*, *worry*) also emerge at this time (ibid.).

A notable subcategory of verbs in the incidental activity reading is that of sound emission, a trend dated by Israel (1996: 222) to the end of the 19th century. Fanego

(2018) here makes a distinction between the sound-emission schema, where the sound emerges as a result of the motion, to verbs of sound emission in the co-occurrence relation schema, where the sound is not related to the motion at all. However, she concedes (2018: 17) that the former is usually preferred in the Intransitive Motion Construction (IMC) (e.g. *The trolley rumbled through the tunnel*; cf. Fanego 2017 on the IMC).

Verbs of sound emission in the incidental activity reading may encode noise (*bang*, *clatter*, *crunch*; these verbs correspond to those in Fanego's sound-emission schema), speech (*chatter*, *talk*, *whisper*), a loud cry (*bellow*, *roar*, *scream*) or other human sounds (*gasp*, *groan*, *grunt*) (Perek 2018: 72). Perek (2018: 72) hypothesises that the productivity of sound emission verbs in the construction can be attributed to the fact that sound emission is amenable to be performed along with motion. Szczesniak (2013: 174), on the other hand, argues that these verbs are permissible in the *way*-construction because they share the atelic properties of verbs encoding the manner of motion; this is supported by my corpus data, which shows that all verbs of sound emission occurring in the dataset are atelic.

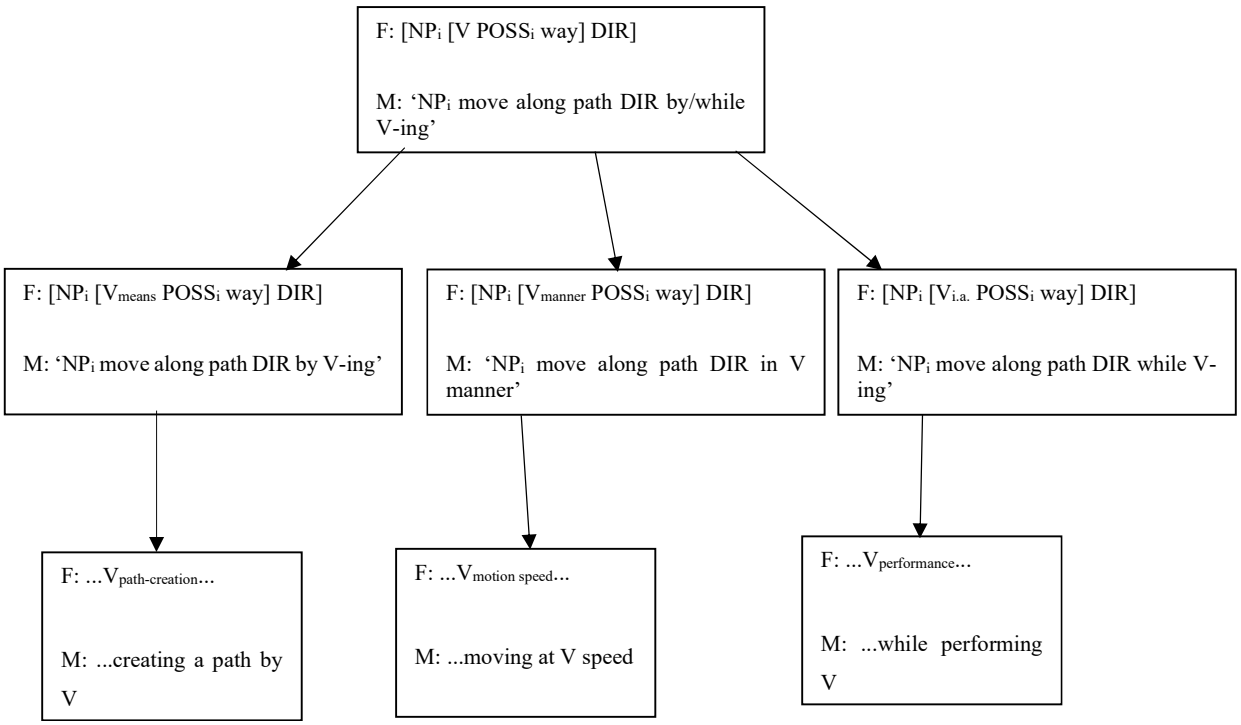
2.2.2.5 The English *way*-construction as a polysemy network

The subsections above have shown that the *way*-construction in English consists of (at least) three distinct subschemas: one with verbs encoding the means of path creation or path traversal; one with verbs encoding the manner in which the subject moves along this path; and one with verbs encoding an activity that is not causally related to the subject's motion along the path, but that merely accompanies this motion. These three subschemas underwent host-class expansion; defined by Himmelmann (2004: 32) as an increase in the range of collocations of an item within a construction. As a result of this host-class expansion, there was a reorganization of expressions with the form [$SUBJ_i$ [V $POSS_i$ *way*] (DIR)] linked to a motion meaning, and they were gathered into the contemporary formal schema [$SUBJ_i$ [V $POSS_i$ *way*] DIR] with the meaning 'SEM_i traverse path PP while/by V-ing' (Traugott & Trousdale 2013: 88). This is an instance of constructionalization, i.e. the creation of a new form-meaning pair in the network. Because the *way*-construction expanded to include intransitive verbs such as *beg* and *worm*, the *way*-subschema of the intransitive construction was analogized to the originally transitive construction and absorbed into

it, causing the *way*-construction to have an intransitive motion subschema (Traugott & Trousdale 2013: 88).

The means, manner, and incidental activity readings are a form of constructional polysemy; the same form is identified with different, but related, senses (Goldberg 1996: 43). Goldberg (1995) suggests viewing the *way*-construction as a polysemy network, though Israel (1996) argues that the notion of a polysemy network does not go far enough, because verbs in all three readings of the construction tend to form clusters of semantically similar verbs, an argument supported by Perek's (2018) distributional semantic study of the construction. Israel (1996) therefore suggests that the *way*-construction should be viewed as a much larger network. In Figure 1 below, I adapt Goldberg's (1995) and Israel's (1996) view of the construction based on Perek's (2018) and my own findings (note: i.a. refers to the incidental activity reading).

Figure 1: The way*-construction as a polysemy network*



The diagram above is of course a simplification; there are many other clusters of verbs in each of the three readings. In this diagram, all three of the readings inherit their

form from the more general *way*-construction schema, and differ from this schema in the type of verb that is specified.

2.2.3 Synchrony of the English *way*-construction

Several formal schemas have been proposed for the contemporary *way*-construction. Christie (2011: 2) states that the *way*-construction has the formal schema [$SUBJ_i$ [V $POSS_{i/j}$ *way*] OBL]]. Asudeh et al. (2008) also propose a $POSS_{i/j}$ argument in their schema because the possessor of the *way*-phrase and the subject do not always co-refer; they cite examples such as *He had bought his son's way into an exclusive military academy*. Though this construction is certainly related to the *way*-construction, it is not an instantiated by it, because the meaning of self-propelled motion along a path is absent. Further, non-coreferential possessors are limited to the idioms *pay*/*buy someone's way*, which I consider separate idioms. This view is supported by Mondorf (2011: 402); she considers only tokens with co-referential subjects and possessive pronouns to be instances of the *way*-construction.

In contrast to Christie (2011: 2), who uses an oblique argument in her formal schema, Salkoff (1988: 49) uses a PP. However, this obscures the fact that directional is crucial to the meaning of the construction, because not all PPs encode direction. For this reason, I follow Traugott and Trousdale (2013: 76) and replace the oblique argument with a directional argument. This directional argument must specify a path; (61) is ungrammatical because only a source is specified.

(61) *He belched his way from Chicago (Marantz 1992: 184, ex. 9a).

The contemporary *way*-construction in English has a fixed syntax whereby the verb occurs with a possessive pronoun, the noun *way* (sometimes modified by an adjective), and an adverbial of direction describing a literal or metaphorical path (Jackendoff 1997: 545). Semantically, the construction describes the traversal of this path (Perek 2018: 66); Jackendoff (2002: 174) suggests that the construction has the meaning *traverse a path while/by doing V*. Goldberg (1995: 207) suggests that the prototypical verb in this construction is *make*[10], and that the prototypical *way*-

[10] This hypothesis is tested in Chapter 6.

construction involves a frame with three components: the creator-theme, the createe-way, and the path.

Jackendoff (1990: 213) observes that the verb in the *way*-construction must denote a process or describe a repeated, bounded event. He adds that the verb can have no syntactic argument of its own apart from the subject, as the difference in grammaticality between the following examples shows.

(62) He drank his way across the country.
(63) *He drank beer his way across the country.
(examples adapted from Jackendoff 2002: 174).

Rappaport Hovav and Levin (1995: 137) state that, while almost all unergative verbs can occur in the *way*-construction, unaccusative verbs cannot[11].

Rappaport Hovav and Levin (1995: 148ff.) propose a number of further constraints on the verb. They state that verbs of inherently directed motion cannot appear in the construction (64). Stative verbs are also ruled out (65), as are verbs of appearance (66). They argue that verbs encoding a change of state are also impermissible in the construction, but this is contradicted by (67), although this is the only such example attested in my data. However, according to Rappaport Hovav and Levin's own account, the ungrammaticality of (64) and (66) can be explained by the fact the verbs are unaccusative.

(64) *I fell my way down the stairs.
(65) *He is knowing his way to a PhD.
(66) *Mariah Carey appeared her way to stardom.
(67) The Seattle air was full of mist and I felt as if I were melting my way through it as I walked along the sidewalk (COHA, 1993).

The following subsection summarises the main findings relating to the synchrony and diachrony of the English *way*-construction, and explains the similarities and difference between my own account and the existing accounts in the literature. Because

[11] Only one counterexample is found in my data: *Sondheim is sort of dying his way into the canon, and it's not pretty* (COHA, 1998); this is clearly a case of coercion.

comparatively very little has been written about the synchrony and diachrony of the Dutch and German constructions, I do not have a similar summary in Section 2.3 and 2.4 respectively.

2.2.4 Summary of the diachrony and synchrony of the English *way*-construction

Israel (1996) identifies three precursor constructions, which he dates to the 14th century. Israel's dating corresponds with my own; the first instance of a precursor construction found in my data is from the Helsinki corpus and dates between 1250-1350 (*He ete and dranc and went his wai*). However, my account differs from Israel's in that I, following Traugott and Trousdale (2013) and Goldberg (1996), identify two precursor constructions: one transitive, the other intransitive. While Israel (1996) dates the modern English *way*-construction to the 15th century, Traugott and Trousdale (2013) date it to the end of the 17th century; my dating corresponds with Traugott and Trousdale's, because the first instance of the modern English *way*-construction in my dataset dates to 1658 (*The then randesvouz of the contrary side fought their way in*). The reliability of this finding is challenged, however, by the very small size of the corpora I used for Old and Middle English.

I showed above that Fanego (2018) is incorrect in dating the means reading of the *way*-construction to 1325, because the example she gives is not a typical means verb. My dating of the emergence of the means reading differs from both Fanego's and Israel's, who dates it to the end of the 16th century; the first example of a means verb in my dataset is the example from 1658 which was given above. As for the manner reading, Israel (1996) claims it has its origins in a "go-your-path" construction; however, he conflates this precursor construction with the *way*-construction itself. This precursor construction did not have an overt path argument (e.g. *He ete and dranc and went his wai*). Therefore, I follow Traugott and Trousdale (2013) in saying that these verbs were marginal to the construction.

Synchronically, I argued that the schema of the contemporary construction must contain a directional argument, as Traugott and Trousdale (2013) posit. Christie (2011) and Salkoff (1988) have an OBL and PP argument after POSS *way* respectively, but this is in error, because it conceals the fact that direction is key to

the meaning of the construction; not all PPs are equally felicitous in the construction (e.g. **He whistled his way in the back garden*). I also argued that idioms such as *He paid his daughter's way through college* are not instantiations of the *way*-construction, as Asudeh et al. (2008) claim.

2.2.5 Related constructions in English

The *way*-construction is part of a larger family of resultative constructions (see e.g. Jackendoff 1990; Goldberg 1995; Mondorf 2011). In a resultative construction, such as *He drank the coffee pot empty*, the main verb encodes an activity and the resultative argument is always predicated of the syntactic object. This is known as the Direct Object Restriction (Rappaport Hovav & Levin 1995). The *way*-construction violates this restriction, because the resultative argument is predicated of the subject (Mezhevich 2003: 166).

The *way*-construction is formally and functionally similar to the fake reflexive resultative construction (FRR; cf. Simpson 1983) and body part resultative constructions (Rappaport Hovav & Levin 1995: 198). Wendt (1911: 197) shows that the *way*-construction can be used in some of the same contexts as the FRR. He points out that up to the beginning of the 17th century, the FRR was preferred, but this trend reversed in the 18th century. Since then, Mondorf (2011: 404) claims that the *way*-construction has continually increased in frequency at the expense of the FRR.[12]

Although the *way*-construction and the FRR have sometimes been considered almost identical in the literature (Jackendoff 1990; Marantz 1992), they exhibit important differences. The means reading of the *way*-construction involves the creation of the path traversed by the subject (Goldberg 1996: 50); this reading is not possible in the FRR, as the difference between (68) and (69) shows.

(68) Bill elbowed his way through the crowd.
(69) *Bill elbowed himself through the crowd.

[12] These hypotheses concerning the relation between the *way*-construction and the FRR will be revisited in Section 5.4.2.

As pointed out in 2.2.1.4, the *way*-construction can be used in an incidental activity reading, where the subject performs an activity which is not the means or manner of motion, but merely accompanies this motion. In the FRR, this is at best extremely unusual; only one such instance is attested in my data.

(70) He whistled himself away, and presently there came along Master Edward Sharpless (COHA, 1900).

There is also a difference in telicity between the two constructions. The *way*-construction can be used to describe atelic events, as in (71). The FRR describes achievements or reaching an endpoint; these events are inherently telic, and so (72) is ungrammatical.

(71) Sarah chatted her way through the party.
(72) *Sarah chatted herself through the party.

As mentioned above, the directional argument in the *way*-construction must denote a path; in the FRR, it must denote the endpoint of the event. Further, Mondorf (2006) argues that intentional actions are more strongly associated with the *way*-construction than the FRR.

The *way*-construction is also similar to the Intransitive Motion Construction (IMC), found in sentences such as *The trolley rumbled through the tunnel* (cf. Fanego 2017, 2018). The two constructions have been strongly intertwined since their emergence (Fanego 2017: 40-41). In Middle English, both the IMC and the *way*-construction could be used to express motion, and the two constructions were in competition during this period. Both constructions are attested with semantically similar or even the same verbs in Middle English, such as *rid(den)* 'clear a path'; in Modern English, these verbs are only attested in the *way*-construction, which suggests that the *way*-construction is encroaching on the former territory of the IMC.

Although the *way*-construction and IMC are very similar both syntactically and semantically, they exhibit some notable differences. The selectional restrictions on verbs of sound emission are far greater in the IMC than in the *way*-construction; all sound-emission verbs can be employed in the *way*-construction, but in the IMC, the emission of sound must result from the motion, cf. (73) and (74) (Fanego 2017: 10).

This was not always the case; the innovative sound verbs in the Modern period are attested in the IMC much earlier than in the *way*-construction. Many of these do not occur in the *way*-construction at all (cf. the tables in Fanego 2017: 30, 36-37).

(73) He whistled his way into the room.
(74) *He whistled into the room.

Jackendoff (1997) describes a similar construction known as the time-*away* construction, found in sentences such as *We're twistin' the night away*. In both constructions, arguments appear which are not always licensed by the verb in isolation, as the following examples show; the meanings of 'spend the whole night sleeping' and 'move down the road while whistling' are imposed by the construction.

(75) He slept the night away.
(76) He whistled his way down the road.

The two constructions differ in that the *way*-construction has a fixed string $POSS_i$ *way* in direct object position and a free directional argument, while the time-*away* construction has a free direct object slot and a lexically fixed oblique argument *away* (Jackendoff 1997: 547).

Although the *way*-construction shares some formal and functional properties with the other constructions described in this section, it differs considerably from these other constructions in that it offers enormous scope for speakers to be creative; as will be shown later in the thesis, the verb slot of the construction is now extremely productive with a high number of hapaxes, and these hapaxes have often triggered a change in the semantics of the construction by analogization.[13] Linguistic creativity is evident in virtually all parts of language (Zawada 2009: 37), and L1 speakers are no exception (Gerrig & Gibbs 1988: 3; Pateman 1997: 227). Speakers use existing words in creative and innovative ways (Aitchison 1987: 143), and this may be motivated by social factors; speakers make use of jokes to reinforce intimacy between group members by means of non-serious social interaction (Gerrig & Gibbs 1988: 8; Zawada 2009: 42). Taylor (1991: 220, fn. 5) observes that "a speaker's linguistic creativity may extend a construction beyond its conventional limits". Thus, in the example below, we

[13] This will be explored in more detail in Chapter 5.

see a novel verb derived by sound symbolism used in the construction, for humorous effect.

(77) Yes. Yeah. Just put on your squeaky shoes and *eee-eee-eee* your way out of my life (*The Big Bang Theory*, series 8, episode 12).

Such creative uses of verbs in the *way*-construction may also be explained in terms of the Maxim of Extravagance (Haspelmath 1999: 105): "Talk in such a way that you are noticed" (see also Keller 1994: 95-107). Eitelmann and Haumann (2019) observe that extravagance may trigger language variation and change. The following section describes the synchrony and diachrony of the Dutch *way*-construction and compares it to that of the English construction. I show that the modern Dutch *way*-construction is very similar to its English counterpart, because it can be divided into the means, manner, and incidental activity readings. There are formal differences between the construction in English and Dutch, however; the Dutch construction contains a reflexive pronoun and indefinite article before *weg*, whereas the English construction has a possessive pronoun before *way*. In this section I also describe related constructions in Dutch.

2.3. Synchrony and diachrony of the Dutch *way*-construction

2.3.1 The diachrony of the Dutch *way*-construction

In the 17th century, the verb *banen* had a meaning equivalent to 'flatten out', and frequently appeared in a construction with *weg* 'way'. The noun *weg* in this early construction could occur with both definite and indefinite articles and could undergo pluralisation, as in the English precursor constructions, and both reflexive and non-reflexive pronouns could be used (Verhagen 2003b: 231). This construction did not necessarily entail the goal being reached or that there was movement along a path, as (78) shows; rather, the construction encoded the creation of a path or the removal of obstacles (ibid.).

(78) wilt ghy hebben een ... gesegent Huwelijck gy en moet
want you have a blessed wedding you and must
u selven daer toe den wegh niet banen
you self there to the way not pave
'If you want to have a blessed marriage, you must not pave the way towards it yourself' (Verhagen 2002: 423, ex. 41).

Kramer (2002) shows that the verbs that first appeared in the Dutch *way*-construction besides *banen* were verbs encoding force such as *openen* 'open'. Verhagen (2002: 424) hypothesises that the development was *weg banen* 'smooth a road/path' to convey the idea of creating a possibility to reach a goal, after which the creation aspect of *banen* became conventionalised due to metaphorical extension. A movement aspect of the construction developed after this construction merged with a reflexive construction. This is not surprising; if somebody creates a path for themselves, it is unlikely that they will not traverse it. The connection between path creation and the reflexive was conventionalised in the 20th century, at which point the reflexive pronoun became obligatory, and the construction was emancipated from the other construction involving *weg* and *banen* (Verhagen 2003b: 232).

2.3.2 The synchrony of the Dutch *way*-construction

The contemporary Dutch *way*-construction has the formal schema [$SUBJ_i$ [V $REFL_i$ *een weg* DIR]] (Verhagen 2003b: 226). The most frequently occurring verb in the construction is *banen*, a highly specific verb found only in this construction. The noun *weg* 'way' denotes a literal or metaphorical path along which the subject moves (van Egmond 2006: 39), and must occur with an indefinite article; the presence of a reflexive beneficiary is incompatible with a definite article before the noun (Verhagen 2003a: 30). The reflexive pronoun is almost always a weak reflexive such as *zich*, which cannot be stressed or topicalised (Geurts 2004), as opposed to its strong equivalent *zichzelf*, and must agree with the subject in person and number (Verhagen 2003a: 31-32). The phrase following *weg* must be a prepositional phrase, and the preposition used is typically *door* 'through' (van Egmond 2006: 34). Formally, the Dutch *way*-construction is very similar to the English one. The only differences are that the Dutch construction has three arguments: a subject, a direct object, and an indirect object, while the English construction only has a subject and direct object.

Further, the possessor of the *way*-phrase is marked by a reflexive pronoun in Dutch, rather than a possessive pronoun as in the English construction (Verhagen 2003b: 227).

Similarly to the English *way*-construction, the Dutch *way*-construction has a means reading which involves the creation of a path and motion along that path despite some difficulty (Verhagen 2003b: 224). Another similarity to the English construction is that verbs can be used that do not necessarily encode motion, such as *boren* 'bore' in (79).

(79) Twee bussen boren zich een weg naar het hart van Istanbul
two buses bore REFL a way to the heart of Istanbul
'Two buses bore their way to the heart of Istanbul' (Verhagen 2003b: 224, ex. 4).

The Dutch *way*-construction also has a manner reading, which is shown in (80). As in the English construction, the verb in the manner reading must be able to be construed as a process, and must denote a volitional and self-propelled action. The verb must be used intransitively, and verbs which inherently require a reflexive, such as *zich voorstellen* 'imagine', are ruled out (van Egmond 2006: 42-47).

(80) Onder blauwe bergen kronkelden we ons een weg
under blue mountains wriggled we REFL a way
'We wriggled our way under blue mountains' (SoNaR, date unknown).

A further similarity to the English construction is that the noun *weg* often functions as a meaningless syntactic marker of the construction. When adjectival modifiers appear with *weg*, as in (81), they often describe the motion along this path or the subject, rather than the path itself, suggesting that the noun *weg* in this construction is not referential (van Egmond 2006: 68).

(81) Telkens opnieuw lezen ze zich een sensuele weg door romans
again again read they REFL a sensual way through novels
'Again and again they read their sensual way through novels' (SoNaR, date unknown).

Verhagen (2003b) claims that the incidental activity reading does not exist in Dutch, citing the following example, which he marks as unusual.

(82) Hij floot zich een weg naar de voordeur
He whistled REFL a way to the front.door
He whistled his way to the front door' (Verhagen 2003b: 228, ex. 13).

He hypothesises (2003a: 38) that the lack of an incidental activity reading is due to a difference in the syntax; the English way-construction is a diachronic blend of two distinct precursors, one with transitive verbs encoding path creation, the other with verbs of movement with the string *V one's way* being used adverbially. In both of these precursor constructions, the collocational range of verbs in this construction expanded by analogization, until there was considerable semantic overlap between the two constructions, such that one construct could be read as being instantiated by either pattern. In Dutch, the semantic difference between *make one's way* and *go one's way* is correlated with a clear syntactic difference; the latter construction has no indirect object. However, the following example from van Egmond (2006) shows that the incidental activity reading is now possible in Dutch, although her questionnaire data shows that the incidental activity reading is not acceptable to all speakers.

(83) Ruzieden we ons een weg over de Laan van Meerdervoort heen
quarrelled we REFL a way over the Lane of Meerdervoort away
We quarrelled our way over the Lane of Meerdervoort' (van Egmond 2006: 55, ex. 32b).

In this example, quarrelling cannot plausibly be the means by which they move over the road; it must therefore be an incidental activity accompanying this motion.[14]

In sum, this subsection has shown that the Dutch *way*-construction is very similar to its English counterpart, and can be split into the same three threads as the English construction: the means reading, the manner reading, and the incidental activity

[14] The incidental activity reading of the Dutch *way*-construction is discussed in more detail in Section 7.5.1.

reading. The following subsection discusses formally and functionally similar constructions in Dutch.

2.3.3 Related constructions in Dutch

Verhagen (2003b: 228) describes a similar construction with *weg* and *banen*. In this construction, the noun *weg* occurs with a definite article and there is no reflexive pronoun, as (84) shows.

(84) Haar strijd baande de weg voor legalisatie van abortus
Her struggle paved the way for legalisation of abortion
'Her struggle paved the way for the legalisation of abortion' (Verhagen 2003b: 228, ex. 15).

This construction differs from the *way*-construction in that only transitive verbs can occur, and the object is compatible with their selection requirements; these verbs indicate the creation or removal of an obstacle, such as *openen* 'open', *vrij maken* 'make free', *laten* 'let', and *blokkeren* 'block'. Unlike the *way*-construction, intransitive verbs of path creation or path traversal cannot be used (Verhagen 2003a: 44).

The non-reflexive construction also differs from the *way*-construction in the possibility of negation. The Dutch *way*-construction cannot always be negated, as (85) shows. When negation is used, this indicates modality, usually involving ability or desire, as in (86).

(85) ??Zij baanden zich geen weg door de menigte
They made REFL no way through the crowd
'They made no way through the crowd' (Verhagen 2003a: 44, ex. 40).

(86) Je kunt je geen weg schieten naar het hart van een volk
You can REFL no way shoot to the heart of a people
'You can't shoot your way into the heart of a people' (Verhagen 2003a: 45, ex. 42).

In the non-reflexive construction, on the other hand, negation can be freely used, as (87) shows.

(87) Hij liet de weg voor onderhandelingen niet open
He let the way for negotiations not open
'He did not leave the way for negotiations open' (Verhagen 2003a: 45, ex. 25').

These facts show that, although this construction is similar to the *way*-construction, the *way*-construction implies that the path has been traversed, whereas the non-reflexive construction with *weg*, *banen*, and a definite article is only concerned with the presence or absence of an obstacle (Verhagen 2003b: 229-230).

Verhagen (2003a) points to another related construction, where the possessor of the *way*-phrase is marked with a possessive pronoun. This construction is exemplified by (88).

(88) In het gebouw kunnen bezoekers met computers hun weg zoeken
In the building can visitors with computers their way seek
'In the building, visitors can find their way around with computers' (Verhagen 2003a: 47, ex. 1).

Banen is not used in this construction; the most frequently occurring verb is *vinden* 'find' (Verhagen 2003a: 48). In this construction, the path is not conceptualised as having been created, hence the definite article before *weg*; the path exists independently of the event (Verhagen 2003a: 48-49). As in the *way*-construction, an oblique phrase is required, but this construction differs in that the oblique phrase does not always specify the path travelled; in (88) above, the oblique phrase only specifies the boundaries of the region in which the path is located (Verhagen 2003a: 49).

The Dutch *way*-construction shares grammatical properties with a reflexive construction lacking the noun *weg*. Some of the verbs used in the *way*-construction are also attested in the reflexive construction, as (89) and (90) show.

(89) Zij worstelde zich een weg door de menigte
She fought REFL a way through the crowd
'She fought her way through the crowd' (Verhagen 2007: 263, ex. 8).

(90) Zij worstelde zich door de menigte
She fought REFL through the crowd
'She fought her way through the crowd' (Verhagen 2007: 263, ex. 9).

The construction exemplified by (90) is known as the Transition to Location Construction (TLC) (van Egmond 2006). The *way*-construction is not an instance of the TLC, because the means reading of the *way*-construction in Dutch connotes the removal or overcoming of obstacles, while the TLC does not, as (91) shows.

(91) Zij bewoog zich rustig naar de uitgang
She moved REFL quiet to the exit
'She quietly moved to the exit' (Verhagen 2007: 263, ex. 10).

Unlike the *way*-construction, TLC does not necessarily involve the traversal of a path. Further, the traversal of the path and the activity encoded by the verb are not necessarily co-extensive; in (92), the act of bluffing can take place before the subject gets out of the awkward situation (Asudeh et al. 2008).

(92) Janneke bluft zich uit de benarde situatie
J. bluffs REFL out the awkward situation
'Janneke bluffs her way out of the awkward situation' (Asudeh et al. 2008: 78, ex. 40).

Verhagen (2003a: 42) describes yet another related construction with *weg* in which the beneficiary is not a bare reflexive pronoun, but a nominal, typically with the preposition *voor* ‘for’, as in (93), although *naar* ‘to’ is also highly frequent.

(93) Sex baande voor hem ook de weg naar de roem
Sex made for them also the way to the fame
'Sex also paved their way to fame' (Verhagen 2003a: 40, ex. 20).

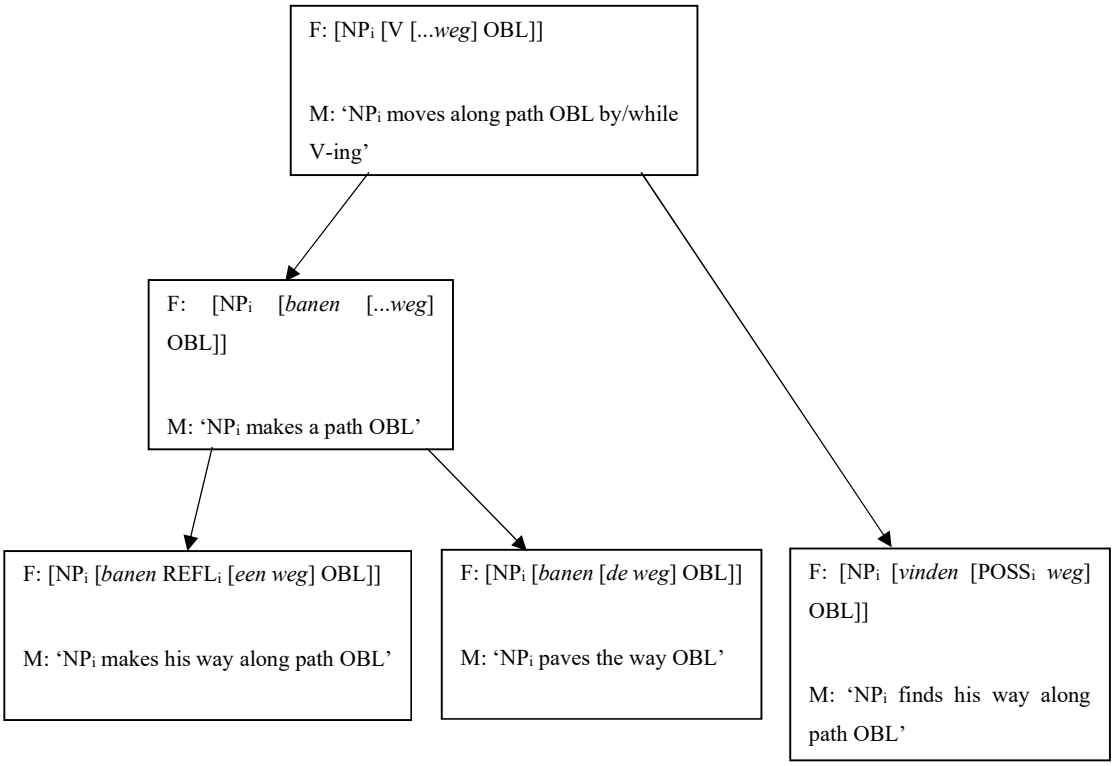
As in the *way*-construction, *banen* is a highly frequent verb in this construction, but this construction differs from the *way*-construction in that the subject and the beneficiary do not always co-refer, as in (93) above. A further difference is that the role of beneficiary is not obligatory, as (94) shows. In the *way*-construction, the role

of beneficiary is obligatory; it is unlikely that someone will create a path for themselves and not traverse it.

(94) Hij liet de weg voor onderhandelingen open
He let the way for negotiations open
'He left the way for negotiations open' (Verhagen 2003a: 42, ex. 25).

The constructions described in this subsection are linked to the *way*-construction in the constructional network, which is illustrated below in Figure 2 (adapted from Figure 5 in Verhagen 2003a: 53).

Figure 2: The network of the Dutch way*-construction and related constructions*



The following section discusses the German *way*-construction, and some functionally and formally similar constructions. I show in this section that, like Dutch, German has a functionally similar reflexive construction, and that, like its English and Dutch counterparts, the German *way*-construction can also be divided into the means,

manner, and incidental activity readings.

2.4. The German *way*-construction

2.4.1 The diachrony of the German *way*-construction

Very little has been written on the German *way*-construction. The few works that have been written on the construction do not deal with the diachrony; I therefore leave an account of the diachrony of the construction to Section 5.2.2, where I document the series of analogical extensions that the construction and its precursors underwent.

2.4.2 The synchrony of the German *way*-construction

Ludwig (2005) shows that German has two closely related constructions which are very similar in meaning to the English *way*-construction; one denoting motion along a path, and the TLC that is also found in Dutch. Van Egmond (2006) suggests that all Germanic languages have these two constructions. Though the two constructions are similar, they differ in that one uses the noun *Weg* 'way' and the other uses a reflexive pronoun. An example of both constructions is given in (95) and (96).

(95) Der Song stampft sich seinen Weg ins Unterbewusstsein
The song stamps REFL POSS way into.the subconscious
'The song stamps its way into the subconscious' (Asudeh et al. 2008: 79, ex. 44).

(96) Er bettelt sich durchs Land
He begs REFL through.the country
'He begs his way through the country' (Asudeh et al. 2008: 79, ex. 45).

The *way*-construction in German generally has the formal schema [NP_i [V [$REFL_i$ $POSS_i$ *Weg*] DIR]]. This schema differs from that of the English and Dutch constructions in that the possessor of the *way*-phrase can be doubly marked by a reflexive and possessive pronoun (Verhagen 2003b). The German *way*-construction is more schematic than the English and Dutch constructions, because the item

occupying the slot before the noun *Weg* is not fixed; an indefinite article, a definite article, and a possessive pronoun can all appear in this slot.

(95) exemplified the means reading of the German *way*-construction. As well as the means reading, the German *way*-construction also contains a manner and incidental activity reading, as the English and Dutch constructions do. The manner and incidental activity reading are shown in (97) and (98) respectively.

(97) Erklimmt euch den Weg in den nördlichen Teil
Climb REFL the way into the northerly part
'Climb your way into the northerly part' (DECOW, date unknown).

(98) Dieser Schmetterling tanzt sich einen Weg durchs Leben
this butterfly dances REFL a way through.the life
'This butterfly dances its way through life' (Ludwig 2005: 11, ex. 31).

As in Dutch, the most frequently occurring verb in the German *way*-construction is *bahnen*, occurring in over 80% of the tokens in Ludwig's (2005) corpus investigation.[15] *Bahnen* is a highly specific verb that only occurs in the *way*-construction; when it is used in the *way*-construction, the meaning is roughly 'make one's way'. Neither *bahnen* nor its English equivalent have paraphrases, as (99) and (100) show.

(99) *He went across the room, making (intended as a paraphrase of *he made his way across the room*).

(100) *Er ging durch die Menge indem er bahnte
He went through the crowd by he made
'*He went through the crowd by making' (Ludwig 2005: 13, ex. 37b).

The restrictions Jackendoff (1992: 162) proposes for verbs in the English *way*-construction also hold of the German construction. The verb must denote a process or describe a repeatable event, and must denote an intentional action on the part of the subject (Ludwig 2005: 11). (101) is ungrammatical because these constraints are violated.

[15] The accuracy of this figure will be tested in Section 6.3.2.

(101) *Er fiel sich seinen Weg aus dem Gefängnis
He fell REFL POSS way out the prison
'*He fell his way out of prison' (Ludwig 2005: 11, ex. 29).

Verhagen (2003b: 233) states that the German *way*-construction can also be used in cases where a path has not been traversed, as in (102), but Verhagen is here conflating the *way*-construction with a related benefactive construction, because of the presence of *und anderen* 'for others'.

(102) Die Heckert-Werke [...] , die sich und anderen den Weg
The Heckert.works that REFL and others the way
zur automatisieren Fabrik der Zukunft ebnen
to.the automated factory of.the future pave
'The Heckert works that pave the way for themselves and others to the automated factory of the future' (adapted from Verhagen 2003b: 233, ex. 31).

As in the English and Dutch constructions, transitive structures cannot occur in the German construction, as the ungrammaticality of (103) shows.

(103) *Er grub sich eine Grube seinen Weg aus dem Gefängnis
He dug REFL a pit POSS way out the prison
'*He dug a pit his way out of prison' (Ludwig 2005: 10, ex. 27b).

The German *way*-construction differs from the English and Dutch constructions in that adjectival modification of *Weg* is not attested in my dataset.

2.4.3 Related constructions in German

Besides the transition to location construction mentioned in 2.4.1, Callies and Szczesniak (2008: 25) point to a construction called the 'manner of obtainment construction' (MOC). This obtainment can be expressed in German by the verbal prefixes *er-* 'obtain by X', *ein-* 'in', *heraus-* 'out', and *ab-* 'away'. The MOC is not always directly equivalent to the *way*-construction, but the two constructions exhibit the same selectional restrictions. Although most meanings conveyed by the MOC can also be expressed by the *way*-construction, the *way*-construction exhibits different syntactic

behaviour. In the MOC, the obtainer is an agent and the obtained is a theme, as in (104); in the *way*-construction, on the other hand, the obtained is expressed as a path and the obtainer is a theme.

(104) als Scholes einen Elfmeter herausholte
as S. a penalty fetched.out
'As Scholes won a penalty' (Callies & Szczesniak 2008: 32, ex. 21b).

As well as the MOC, Ludwig (2005) mentions the Directed Motion Construction (DMC). This construction is found in other Germanic languages, including Swedish (Toivonen 2002) and Norwegian (Seland 2001). Like the *way*-construction, the DMC contains an argument in direct object position that is not a thematic argument of the verb; in (105), Karl is not actually kicking himself.

(105) Karl kickt sich durch die Tür
Karl kicks REFL through the door
'Karl kicks his way through the door' (Ludwig 2005: 14, ex. 41).

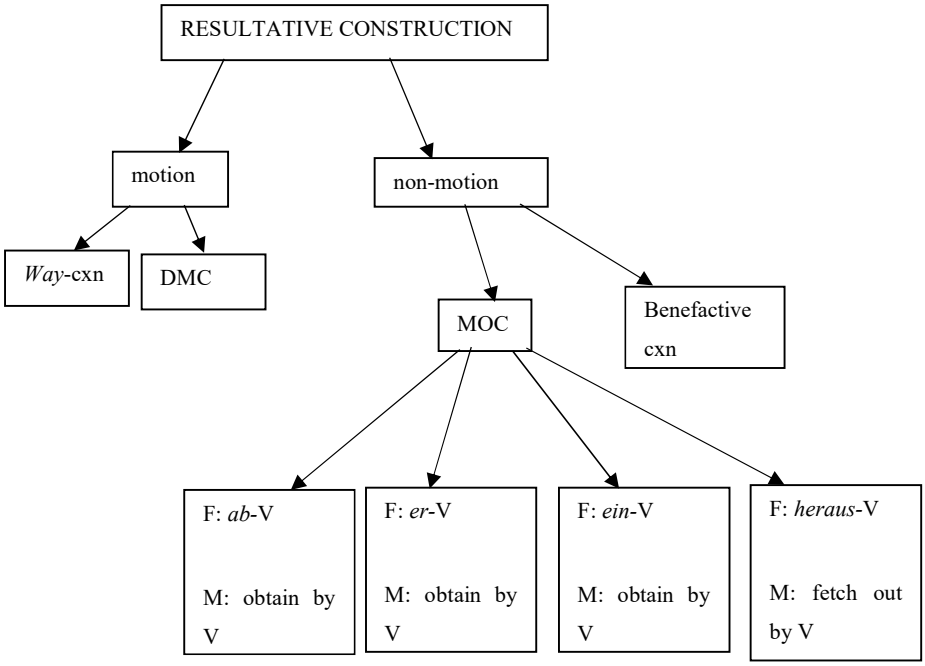
The DMC is similar to the *way*-construction in that both constructions encode motion along a literal or metaphorical path; however, Ludwig (2005: 17) points out that the *way*-construction is more strongly associated with literal paths than metaphorical ones, whereas the DMC is equally felicitous with both kinds of path, a hypothesis I return to in Section 4.2. Her data further show that the *way*-construction is less productive than the DMC (2005: 18). Another difference between the two constructions is that the incidental activity reading cannot be used in the DMC, as (106) shows; (106) is ungrammatical unless Sue's whistling is the means by which she moves through the tunnel, which is unlikely.

(106) ??Sue pfiff sich durch den Tunnel
Sue whistled REFL through the tunnel
'Sue whistled herself through the tunnel' (Ludwig 2005: 14, ex. 43).

As (101) showed, German also has a benefactive construction with *Weg*. This construction is similar in that both constructions contain the noun *Weg* 'way', but it does not instantiate the *way*-construction because the path does not necessarily have to be traversed; in the *way*-construction, the notion of path traversal must be present.

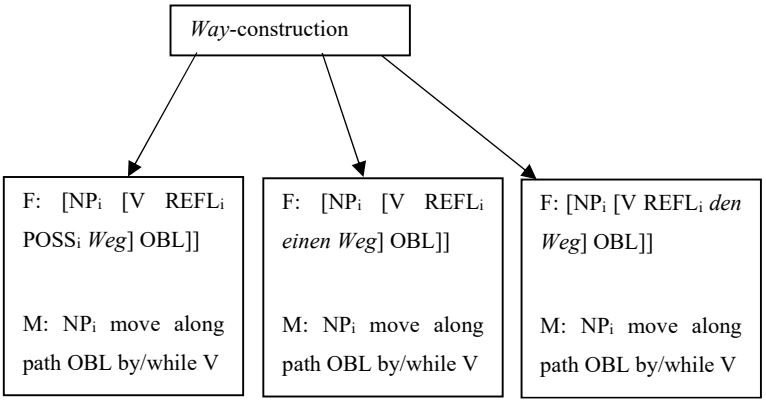
The three readings of the German *way*-construction and the other constructions related to the superordinate *way*-schema can be thought of as being linked in a network, which is illustrated in Figure 3 below.

Figure 3: The network of the German way*-construction and related constructions*



The *way*-construction itself splits into three further nodes, which are illustrated in Figure 4.

Figure 4: Nodes of the German way*-constructional network*



2.5 Conclusion

This chapter of the thesis has presented a justification for a constructional approach to the *way*-construction in English, Dutch, and German. In Section 2.2, I described the synchrony and diachrony of the English *way*-construction, showing that there are three distinct threads: the means reading, the manner reading, and the incidental activity reading. A explanation of the terms 'means' and 'manner' was also presented in this section. I presented a justification for using the term 'means' for verbs which describe the means by which the path is created, and 'manner' for verbs which describe the manner of the subject's motion. It was also shown in this section that the incidental activity reading is not as marginal as is sometimes assumed; it is now extremely productive and has undergone considerable host-class expansion, as I show in more detail in Section 5.2.2. Some constraints on the contemporary English *way*-construction were also outlined in Section 2.2. Section 2.3 concerned the Dutch *way*-construction and its synchrony and diachrony. I showed in this section that the construction emerged via a merger of a reflexive construction and a transitive construction with *banen* and *weg*. Synchronically, I showed that the Dutch construction can also be divided into the same three threads as the English construction, namely the means, manner, and incidental activity readings; this is at odds with Verhagen's (2003b) claim that the incidental activity reading is not permissible in the Dutch *way*-construction (see also Verhagen 2004). Section 2.4 presented a synchronic account of the German *way*-construction. It was shown in this section that the German *way*-construction is more schematic than its English and Dutch equivalents, because the noun *Weg* can be modified by a definite article, indefinite article, or a possessive pronoun, in addition to the obligatory reflexive pronoun. I further showed that all three languages have at least two formally and functionally similar constructions; the FRR and IMC in English, a range of constructions with *weg* in Dutch, and the MOC and DMC in German. Both German and Dutch also have a transition to location construction. The similarities and differences between the English, Dutch, and German constructions are summarised in Table 1 below.

Table 1: Similarities and differences between the way*-construction in English, Dutch, and German*

Property	**English**	**Dutch**	**German**
Possessor of *way*-phrase marked by?	Possessive pronoun	Reflexive pronoun	Reflexive pronoun; sometimes also possessive pronoun
Modification of *way*/*weg* attested?	Yes	Yes	No
Incidental activity reading attested?	Yes	Yes	Yes
Transitive structures attested?	No	No	No

3. Method

3.1 Introduction

This chapter answers the second research question asked in my introduction: what can corpus data, and the statistical analysis of this data, tell us about the diachrony of the *way*-construction in each language? The combination of methods used in the present study is novel for several reasons. Firstly, a cross-linguistic corpus study of the *way*-construction on this scale has never been carried out (though see Pedersen 2013 for a smaller cross-linguistic study of the *way*-construction). Secondly, this study fills a gap by providing a quantitative analysis which has been missing from many studies on the *way*-construction (see e.g. Jackendoff 1992; Marantz 1992; Goldberg 2006; Traugott & Trousdale 2013 on the English construction; Ludwig 2005 on the German construction; Verhagen 2007 on the Dutch construction). In studies which do take a quantitative approach, either a small range of corpora is used or a relatively small dataset is investigated. Mondorf (2011), for example, takes her data from 77 million words of a selection of British prose corpora. Perek (2018) uses a larger dataset, but only from one corpus, COHA (Davies 2012); the easiest documentation of a development is a comparison across two or more corpora that represent different periods of time (Hilpert 2013: 462). Israel (1996) uses the OED database, a very large dataset, but detailed frequency is missing from his study. The problem of small or temporally narrow datasets is also present in studies on the Dutch *way*-construction; Verhagen (2003a) uses only a small CD-ROM corpus from the 1995 *Volkskrant* newspaper; his study is therefore limited in both size, genre, and temporal range. In a study on the German *way*-construction, Ludwig (2005: 2) mentions “several corpora” but does not say what these are, and detailed frequency information and statistical analysis are absent from her study.

Thirdly, many quantitative studies have simply provided frequency information without performing any rigorous statistical analysis (see e.g. Fanego 2012, 2017). Frequency data alone are not informative without reference to whether these data are statistically significant (Gries 2006b: 5 and references cited there). Statistical analysis is especially important in diachronic studies; more rigorous quantification of diachronic data is needed when research questions go beyond simply detecting a change

(Hilpert & Gries 2016: 38). In addition, relying on raw frequencies alone can obscure differences that hold across more than one set of data (Hilpert 2012a: 142). Perek (2018) remedies the lack of statistical analysis in many investigations of the *way*-construction by using a range of modelling and significance testing measures in his study. In this thesis, I follow Perek (2018) in adopting a statistically informed approach, using collostructional analysis techniques[16] and a Variability-Based Neighbour Clustering (VNC) algorithm[17] to periodise my data.

The remainder of this chapter is structured as follows. Section 3.2 argues for the importance of corpus linguistics to the historical linguist more generally, with a focus on its application to diachronic Construction Grammar. The application of corpus linguistics to this study is described in Section 3.3, where I list the corpora used in this study, and evaluate the advantages and disadvantages of choosing this combination of corpora. I argue that the combination of corpora chosen provides a wide scope for analysis thanks to its broad temporal and genre range, although there were some problems, especially in the Dutch and German corpora, because the date of publication of the text was often missing. Section 3.4 outlines the process of collecting and coding the data. Section 3.5 discusses the statistical analysis techniques employed in the study, with a focus on how to measure entrenchment. Langacker (1987: 100) posits a linear relationship between token frequency and entrenchment, whereas Gries (2014b) argues that token frequency is a poor predictor of entrenchment, and that other measures should be used. To test these hypotheses, I perform a correlation analysis on the token frequency of each verb in the *way*-construction and the *p*-value of the Fisher-Yates exact test, which is considered the most accurate approximation of entrenchment (Stefanowitsch & Flach 2016: 105), and show that token frequency may not be as poor a predictor of entrenchment as Gries (2014b) assumes. I also discuss in this section how the calculation of these variables was applied to my dataset. Section 3.5 ends by describing the application of the VNC algorithms to my dataset. Section 3.6 concludes the chapter.

[16] All collostructional analyses were performed using the Coll.Analysis 3.5 script (Gries 2014a).
[17] The VNC analysis was performed using a script in the statistical software package R (RStudio Team, 2015) which was downloaded at http://global.oup.com/us/companion.websites/fdscontent/uscompanion/us/static/companion.websites/nevalainen/Gries-Hilpert_web_final/vncindividual.html, last accessed on 25th September 2018.

3.2 Corpus linguistics

Corpus linguistics existed even before the invention of computers (López-Couso 2016: 127); Gries (2006b: 3) points out that corpora have been used even as far back as the 19th century. The advent of corpus linguistics could in fact be dated even earlier than this; Samuel Johnson arguably created a corpus to produce his dictionary. In the era of computers, the creation of electronic corpora has given us access to a very large amount of data, far more than could be obtained in an experimental setting (Kytö 2012: 1510), and data from these corpora, whether analysed quantitatively or qualitatively, have helped improve our understanding of how, why, and when a language has changed over time (López-Couso 2016: 129). Corpora also provide concrete evidence of which items are acceptable (or not) in a given construction. For example, it has often been claimed that the verb *donate* does not occur in the ditransitive construction; however, this claim has been based on introspective judgements from linguists, rather than on any evidence (Gries 2013: 98). Data from Stefanowitsch (2011) show that *donate* does, in fact, occur in the ditransitive construction.

Although the use of corpora offers significant advantages to researchers interested in historical linguistics, this method is not without its drawbacks (Kytö 2012: 1521). Firstly, there can be large temporal gaps in corpora, especially in early corpora; there is often a considerable amount of time between the date the original text was written and the date of the earliest preserved copy (Kytö 2012: 1522). Another disadvantage is that corpora are often not representative of all groups of society; in the earliest stages of English, for example, there are far more texts written by men than by women; texts written by the lower classes of society are also very rare during this period (ibid.). As a result, historical corpora are seldom balanced (ibid.).

More recently, work in diachronic Construction Grammar has adopted a corpus-based methodology (see, among others, Gries & Hilpert 2009; Hilpert 2012b; Hilpert 2013; Hennemann 2015) in a variety of languages, dialects, and registers (Gries 2013: 97 and references there). This has been made possible in part thanks to POS-tagged corpora. The term ‘POS-tagged’ refers to a corpus in which every item has been annotated for its part of speech. As well as searching for individual lexical items, POS-tagged corpora allow the researcher to search for more schematic constructions, such

as the ditransitive construction NP_i V NP_j NP_k (for a corpus-based study of the ditransitive construction, see Stefanowitsch 2011).

Corpus-based research can be applied in many ways in a diachronic Construction Grammar framework. Considering the importance of usage in most constructionist theories of language (see e.g. Langacker 1987, 2000, 2008; Bybee 2010), it follows that frequency of usage is also important. Changes in frequency are a widely-used measure of constructional change (Hilpert 2013: 461), and questions pertaining to frequency can readily be answered by corpora, from the raw frequency of an item or construction in an entire corpus, to increases or decreases in frequency over a period of time. Subject to further statistical testing, the significance of this frequency trend, if there is one, can be evaluated. This information in turn can be used to assess whether a construction is becoming more or less entrenched, as increasing frequency leads to increasing entrenchment, and vice versa (Langacker 1987: 100).

As well as changes in frequency, corpora have also been used to analyse diachronic variation in a Construction Grammar framework (Hilpert & Gries 2016: 36). A construction may vary in its form and meaning, and both types of variation can be investigated with corpora (Hilpert 2013: 459). Corpora can be used to investigate several processes of constructional change, including morphosyntactic change, change in argument structure, host-class expansion (cf. Himmelmann 2004), metaphor and metonymy, analogical extension, and collocational change (cf. the references in Hilpert 2013).

However, corpus linguists should be cautious about the conclusions they draw when using corpora in a usage-based framework. Corpus linguists have often assumed that the prototypical item in a construction is the one that is most frequently attested in the corpus (Gilquin 2006). For example, experimental data suggests that the prototypical meaning of *give* is one of transfer, but the most frequent use of *give* in the Switchboard corpus is in a light verb construction where the meaning of transfer is absent, e.g. *He gave a smile* (Arppe et al. 2010: 9).

A further challenge to the corpus-based linguist lies in the fact that it can be difficult to find a sufficient amount of data in smaller corpora, especially for mid- or low-frequency phenomena (Kytö 2012: 1516). Similarly, larger corpora can provide far too

much data for the researcher to analyse, especially where high-frequency phenomena are concerned (Kytö 2012: 1517). In this case, researchers have to rely on random sampling techniques to extract a smaller sample of the data (see e.g. Scott 1999; Sinclair 1999; McCarthy & Carter 2003; McColm & Trousdale 2019), or constructing smaller sub-corpora (Tribble 1997; O'Keeffe 2003; Koester 2006). Despite these challenges, corpus linguistics remains an extremely valuable tool to the historical linguist, and can offer significant insights into language change, especially in a diachronic Construction Grammar framework.

The following section lists the corpora used in the study and provides a brief overview of the types of text used in each corpus. I also explain the measures taken to ensure comparability between corpora, and explain the reasoning behind the choice of corpora.

3.3 The corpora

3.3.1 The English corpora

3.3.1.1 ARCHER

Kytö (2012: 1512) distinguishes between multipurpose and specialised corpora. ARCHER (*A Representative Corpus of Historical English Registers*) is an example of the former; covering the period from 1650 to 1990, it contains texts from a wide variety of genres in both British and American English (López-Couso 2016: 131). As its name suggests, ARCHER also contains texts from a wide range of registers, from formal to informal, and from written to more speech-like registers (ibid.). The original version of ARCHER contained less than 2 million words (Biber et al. 1994), but has since been considerably expanded. This study uses the untagged version of ARCHER 3.2[18], which contains roughly 3.8 million words across 1,710 texts. This size is still very small compared to some of the other corpora used in this study, which mitigates the advantage offered by the diversity of registers.

[18] This version of the corpus was accessed via the CQPweb interface hosted by Lancaster University at https://cqpweb.lancs.ac.uk/archer_untagged/.

3.3.1.2 The BNC

The BNC (*British National Corpus*) was completed in 1994, and contains roughly 112 million words of written and spoken British language from a wide range of sources from the late 20th century (Burnard 2007). To ensure balance and representativeness, a wide variety of types of material was sampled, and care was taken to select materials with a wide range of distribution (Aston & Burnard 1998: 28). Although no new texts have been added since its completion, the BNC was revised in 2001 and again in 2007 upon the release of the XML edition[19]. The distribution of the corpus by year and text type is shown in Tables 2 and 3 below, which are adapted from Burnard (2007).

Table 2: Distribution of the BNC by text type

Text Type	No. Texts	No. Words
Spoken demographic	153	4,233,955
Spoken context-governed	755	6,175,896
Written books and periodicals	2,685	79,238,146
Written to be spoken	35	1,278,618
Written miscellaneous	421	7,437,168
Total	**4,049**	**98,363,783**

Table 3: Distribution of the BNC by year

Date	No. Texts	No. Words
Unknown	162	1,831,585
1960-1974	46	1,718,449
1975-1984	169	4,730,889
1985-1993	3,672	90,082,860
Total	**4,049**	**98,363,783**

3.3.1.3 CLMET3.1[20]

The Late Modern English period is a much-neglected period in the history of English (De Smet 2005: 69). CLMET3.1 (*Corpus of Late Modern English Texts*) seeks to fill

[19] The XML edition of the BNC was used in this study. It was accessed via the CQPweb interface hosted by Lancaster University at https://cqpweb.lancs.ac.uk/bncxmlweb.
[20] A copy of this corpus was given to me by Dr Hendrik De Smet.

this gap; it is a tagged corpus covering the period from 1710 to 1920, and is based on texts from Project Gutenberg and the Oxford Text Archive (De Smet 2005: 70). The corpus is divided into three subperiods of 70 years each, and is designed such that the homogeneity within each subperiod is increased and the homogeneity between subperiods is decreased (ibid.). When designing the corpus, De Smet (2005) undertook a number of measures to reduce confounding factors that could arise from studies based on this corpus. To reduce dialect variation, De Smet (2005) sampled texts only from British authors; this facilitates comparison with other corpora of British English such as the BNC (De Smet 2005: 71). De Smet also ensured that each author in the corpus contributed no more than 200,000 words, in order to prevent the idiosyncrasies of one author from skewing the register of the corpus (ibid.).

In order to ensure genre variability, De Smet deliberately favoured non-literary texts over literary ones, and texts from lower rather than higher registers. He also included texts from female authors wherever possible. However, this creates a sociolinguistic bias in the corpus, albeit the opposite bias characteristic of corpora such as Project Gutenberg and the Oxford Text Corpus, whose sources are typically very formal texts written by upper-class men (De Smet 2015: 72). As sociolinguistic analysis is not the goal of this thesis, this bias does not pose a significant problem.

Another potential problem with the corpus is that internet sources often provide no information on which version of a text is the basis for the internet edition. Information on who the editors were and what changes they made to the original texts is also often missing (De Smet 2005: 79). However, because I am not studying phenomena which may attract editorial intervention, such as punctuation and spelling, this did not pose a problem to my analysis.

The CLMET3.1 corpus contains roughly 40 million words, which makes it ideal for studying phenomena for which smaller corpora provide insufficient data (De Smet 2005: 80). The distribution of the corpus by decade and by genre is shown in Tables 4 and 5 below.

Table 4: Distribution of CLMET3.1 by decade

Decade	Word Count	Decade	Word Count
1720s	320,480	1830s	2,557,261
1730s	451,598	1840s	4,754,880
1740s	4,009,364	1850s	1,770,118
1750s	2,111,691	1860s	2,255,218
1760s	2,076,686	1870s	1,542,594
1770s	3,417,896	1880s	2,032,245
1780s	997,374	1890s	3,336,998
1790s	1,371,755	1900s	2,385,824
1800s	381,935	1910s	1,596,469
1810s	1,277,455	1920s	189,550
1820s	2,190,712	**Total**	**41,028,103**

Table 5: Distribution of CLMET3.1 by genre

Genre	Word Count
Treatise	5,287,936
Drama	1,797,876
Other	6,653,239
Narrative fiction	19,122,107
Letters	2,666,863
Narrative non-fiction	5,567,149
Total	**41,095,170**

3.3.1.4 COCA

A major breakthrough in diachronic corpora comes from so-called mega-corpora such as COCA (*Corpus of Contemporary American English*) (López-Couso 2016: 137). COCA is updated every year; as a result of its latest update in December 2017, it now contains 560 million words of text from 1990 to 2017, divided so that there are roughly 20 million words per year (https://corpus.byu.edu/coca/). The corpus is also equally divided in terms of genre and register; spoken, fiction, popular magazines, newspapers, and academic texts are all equally represented (ibid.). Table 6 below shows the distribution in COCA in each year and across genres.

Table 6: Distribution of COCA per year and genre

Year	Spoken	Fiction	Magazines	News	Academic	Total
1990	4,241,820	4,100,296	3,993,642	4,000,927	3,914,328	20,251,013
1991	4,183,317	4,075,428	4,099,198	4,003,173	3,980,425	20,341,541
1992	4,367,946	3,792,255	4,292,672	3,984,942	3,957,009	20,394,824
1993	4,336,787	3,860,406	4,250,973	4,041,673	4,078,421	20,568,260
1994	4,305,046	4,046,747	4,293,745	4,040,013	3,977,781	20,663,332
1995	4,396,172	3,847,142	4,288,730	4,009,933	3,948,436	20,490,413
1996	3,965,565	3,858,640	4,277,667	3,987,828	4,037,870	20,127,570
1997	3,774,994	3,678,700	4,259,465	4,036,195	4,342,502	20,091,856
1998	4,314,807	3,683,747	4,283,190	4,019,406	4,038,454	20,339,604
1999	4,286,305	4,045,331	4,281,338	3,998,758	3,951,864	20,563,596
2000	4,297,830	3,850,344	4,282,437	3,949,191	4,019,668	20,399,470
2001	3,896,284	3,789,875	4,194,943	3,984,202	3,895,326	19,760,630
2002	4,230,138	3,674,168	4,210,790	4,001,474	3,980,495	20,097,065
2003	4,297,895	4,015,842	4,222,326	3,937,025	3,972,378	20,445,466
2004	4,224,432	3,999,217	4,229,015	4,003,463	3,938,459	20,394,586
2005	4,300,773	3,998,572	4,252,853	4,010,857	3,856,046	20,419,101
2006	4,210,862	4,004,822	4,205,020	4,005,230	3,994,522	20,420,456
2007	3,774,535	3,948,324	4,112,852	3,891,029	4,226,689	19,953,429
2008	3,533,287	4,076,895	4,191,580	3,969,842	3,917,939	19,689,543
2009	3,883,612	4,069,557	3,897,508	3,955,928	3,992,413	19,799,018
2010	4,023,555	3,885,982	3,765,169	4,219,629	3,787,581	19,681,916
2011	4,760,687	4,166,029	4,199,378	3,986,321	4,551,005	21,663,420
2012	4,336,058	4,335,155	4,294,190	4,173,813	4,337,823	21,477,039
2013	4,019,619	4,225,162	4,173,336	4,133,917	3,531,695	20,083,729
2014	4,004,868	4,134,220	4,266,683	4,142,500	3,456,761	20,005,032
2015	4,005,894	4,255,674	4,195,487	4,130,818	3,609,226	20,197,099
2016	4,371,199	4,197,883	4,087,037	4,134,560	4,005,824	20,796,503
2017	4,404,291	4,228,709	4,252,889	4,242,760	4,109,588	21,238,237
Total	**116,748,578**	**111,845,122**	**117,354,113**	**112,995,407**	**111,410,528**	**570,353,748**

3.3.1.5 COHA

COHA (*Corpus of Historical American English*) is a 400 million word corpus covering the period 1810-2009. It is 100 times larger than any other structured corpus of historical English (Davies 2012: 121-122). The corpus contains texts from fiction, magazines, newspapers, and non-fiction books, and is balanced by genre from decade to decade. COHA is also balanced across decades for subgenres and domains (Davies 2012: 124). The one exception is the lack of newspaper texts between 1810 and 1850 (Davies 2012: 124, fn. 3). While COCA contains a roughly equal amount of text in each year, COHA is not equally balanced by time period. The distribution of texts in COHA per decade and genre is shown in Table 7 below.

Table 7: Distribution of COHA per year and genre

Decade	Fiction	Magazines	Newspapers	Non-fiction	Total
1810s	641,164	88,316	0	451,542	**1,181,022**
1820s	3,751,204	1,714,789	0	1,461,012	**6,927,005**
1830s	7,590,350	3,145,575	0	3,038,062	**13,773,987**
1840s	8,850,886	3,554,534	0	3,641,434	**16,046,854**
1850s	9,094,346	4,220,558	0	3,178,922	**16,493,826**
1860s	9,450,562	4,437,941	262,198	2,974,401	**17,125,102**
1870s	10,291,968	4,452,192	1,030,560	2,835,440	**18,610,160**
1880s	11,215,065	4,481,568	1,355,456	3,820,766	**20,872,855**
1890s	11,212,219	4,679,486	1,383,948	3,907,730	**21,183,383**
1900s	12,029,439	5,062,650	1,433,576	4,015,567	**22,541,232**
1910s	11,935,701	5,694,710	1,489,942	3,534,899	**22,655,252**
1920s	12,539,681	5,841,678	3,552,699	3,698,353	**25,632,411**
1930s	11,876,996	5,910,095	3,545,527	3,080,629	**24,413,247**
1940s	11,946,743	5,644,216	3,497,509	3,056,010	**24,144,478**
1950s	11,986,437	5,796,823	3,522,545	3,092,375	**24,398,180**
1960s	11,578,880	5,803,276	3,404,244	3,141,582	**23,927,982**
1970s	11,626,911	5,755,537	3,383,924	3,002,933	**23,769,305**
1980s	12,152,603	5,804,320	4,113,254	3,108,775	**25,178,952**
1990s	13,272,162	7,440,305	4,060,570	3,104,303	**27,877,340**
2000s	14,590,078	7,678,830	4,088,704	3,121,839	**29,479,451**
Total	**207,633,395**	**97,207,399**	**40,124,656**	**61,266,574**	**406,232,024**

Using corpora such as COCA and COHA presents a twofold advantage. Firstly, their huge size means that researchers can obtain statistically significant results; this is a problem when using smaller corpora because there is often not enough data to establish statistically significant conclusions (Davies 2012: 151). Secondly, using historical corpora such as COHA is much more informative than a text resource such as the Oxford English Dictionary, which can provide data on the first attestation of a word but not its change in frequency over time (Davies 2012: 126). Given that the temporal range of COCA and COHA overlap partially, care was taken to exclude duplicated tokens of the *way*-construction across these two corpora.

3.3.1.6 The Helsinki Corpus[21]

The Helsinki corpus pioneered the field of corpus linguistics (Hilpert & Gries 2009: 385). The corpus was first published in 1991 by Matti Rissanen, and contains roughly 1.8 million words spanning the 8th century to 1710 (López-Couso 2016: 130). This

[21] The untagged version of the corpus was accessed at https://cqpweb.lancs.ac.uk/helsinki/.

wide time span makes the corpus ideal for investigating long-term developments, though this advantage is mitigated by its relatively small sized compared to other corpora. In order to ensure representativeness, the corpus contains texts from a variety of genres, from the Bible and legal writing to fiction and drama (Kytö 2012: 1512). As well as written works, the corpus also contains spoken texts such as trial proceedings (López-Couso 2016: 130). The corpus contains 400 samples of running texts divided into subperiods of 70 to 100 years; this facilitates the comparison of changes over these subperiods (ibid.).

3.3.2 The Dutch corpora

3.3.2.1 *De Gids*[22]

The *De Gids* corpus is a historical corpus spanning 1837 to 1936. It contains roughly 78 million words of text from the literary periodical of the same name; the number of words per decade is shown in Table 8 below.

Table 8: Word count per decade in the De Gids *corpus*

Decade	Word Count
1830s	6,496,541
1840s	5,486,008
1850s	6,294,627
1860s	8,527,608
1870s	8,920,834
1880s	8,528,949
1890s	8,268,649
1900s	7,916,786
1910s	7,532,321
1920s	6,269,992
1930s	3,932,935
Total	**78,175,250**

Because the corpus contains only material from a literary periodical, it is extremely unbalanced in terms of both genre and register. For this reason, I used other, more balanced, corpora in conjunction with *De Gids* in order to account for confounding genre effects. Another problem with this corpus is that it is not POS-tagged. This made

[22] A copy of this corpus was given to me by Dr Timothy Colleman.

searching for tokens of the Dutch *way-construction* more difficult; the process of data collection in the *De Gids* corpus is discussed in more detail in Section 3.3.4.

3.3.2.2 NLCOW14[23]

The NLCOW14 corpus (Schäfer & Bildhauer 2012; Schäfer 2015) is a web-based corpus of Dutch. The web and web-based corpora can provide a valuable insight into recent and ongoing change in a language (López-Couso 2016: 143; see also Hundt et al. 2007; Mair 2007, 2012), given that language change on the internet typically progresses at a rapid rate (see e.g. Crystal 2001). Another advantage of using web-based corpora is their massive size (López-Couso 2016: 144); the NLCOW14 corpus, for instance, contains almost 6.9 billion words. Further, web-based corpora also take into account a lot of non-standard variation that is missing from other corpora (Schäfer 2015: 28). One obvious disadvantage of using this corpus is its genre imbalance; the texts in this corpus are typically very informal texts such as blog and forum posts. I therefore used Dutch corpora from a combination of genres.

3.3.2.3 SoNaR[24]

SoNaR (*Stevin Nederlandstalig Referentie corpus* 'Stevin Dutch language reference corpus') is a tagged corpus of contemporary written Dutch from 1954 to 2011 containing 500 million words ("SoNaR: Dutch Reference Corpus", 2011). The corpus is balanced in terms of genre and register; the texts range from formal writing such as newspapers to more informal styles such as forum posts and SMS messages (ibid.). The corpus is also balanced according to the number of speakers in Dutch-speaking regions; one third of the texts comes from Flanders, while the other two thirds come from the Netherlands (ibid.).

[23] The tagged version of the NLCOW14 corpus was accessed via the NoSketchEngine interface hosted at www.webcorpora.org.
[24] The SoNaR corpus was accessed via the Opensonar interface hosted at https://portal.clarin.inl.nl.

3.3 The German corpora

3.3.3.1 *Berliner Zeitung* (1994-2005)[25]

The *Berliner Zeitung* (1994-2005) corpus is a tagged corpus containing roughly 237 million words of text taken from the online edition of the newspaper of the same name. Because the corpus contains texts from only one genre over a relatively short period of time, larger and more varied corpora were used alongside the *Berliner Zeitung* corpus.

3.3.3.2 DECOW16A-NANO[26]

The DECOW16A-NANO corpus (Schäfer & Bildhauer 2012; Schäfer 2015) is a web-based corpus of German containing roughly 990 million words (DECOW16A-NANO, 2016). The corpus is almost exclusively made up of texts from Germany; 97.27% of the texts come from Germany, and Austrian and Swiss texts make up 0.94% and 1.79% respectively (ibid.). The NANO version of the corpus was chosen because the full DECOW16A corpus contains almost 20 billion words; in combination with the other corpora used in this study, using the full DECOW16A corpus would yield too many tokens of the *way*-construction to investigate.

3.3.3.3 *Deutsches Textarchiv*[27]

The *Deutsches Textarchiv* is a tagged, relatively large diachronic corpus of about 210 million words covering the period from 1473 to 1927. The corpus is balanced in terms of genre; of the 3190 texts, 759 are drawn from newspapers, 699 come from works of fiction, 817 are taken from scientific works, and 915 come from functional literature, though there are no texts from spoken genres (http://www.deutschestextarchiv.de/doku/textauswahl). Care was also taken to select texts which were in circulation throughout the whole country (ibid.). The size and

[25] The corpus was accessed at https://www.dwds.de/r. The *Berliner Zeitung* (1946-1993) corpus could not be used due to copyright restrictions.
[26] The tagged version of this corpus was accessed via the NoSketchEngine interface hosted at www.webcorpora.org.
[27] The *Deutsches Textarchiv* was accessed at www.dwds.de/r.

balance of the corpus therefore make it an ideal resource for the researcher concerned with diachronic developments in German.

3.3.4 Measuring comparability between and within corpora

Results concerning the same phenomenon may vary between corpora or even within the same corpus, because corpora are not always homogeneous (Gries 2006a: 110). In order to measure variability between and within corpora, word frequencies have often been used (Gries 2006a: 114). Kilgarriff (2001), for example, compares corpora by comparing the observed and expected frequencies in two or more corpora, and performing a chi-squared test to see if the difference between the observed and expected frequencies is statistically significant. A Wilcoxon signed-rank test and Fisher's exact test are suggested as alternatives to the chi-squared test (Kilgarriff 2001: 103-105). Rayson and Garside (2000), on the other hand, use part-of-speech frequencies to compare corpora. However, these approaches are limited in their applicability, because analysis at the lexical level or part-of-speech level is uninformative to researchers who are interested in grammatical phenomena (Gries 2006a: 115-116). In comparing corpus frequency counts, I adopt a simpler approach by comparing frequencies per million words, and only comparing corpora that are varied in genre. In addition, I adopted a strictly qualitative approach when comparing between languages.

Rayson and Garside (2000: 1) raise the issue of corpus homogeneity when comparing two or more corpora. However, as Gries (2006a: 137) points out, the homogeneity of a corpus is irrelevant, because the assessment of a corpus's homogeneity hasn't been performed with respect to a phenomenon that the researcher is interested in; the homogeneity of the corpus with respect to the relevant phenomenon may well be different.

A further issue that arises when comparing corpora is representativeness. To be representative, a corpus should contain samples of all major text types (Leech 1993) and if possible, proportional to their usage in everyday language (Clear 1992). To ensure reliability of the results, the researcher should compare corpora with a similar degree of representativeness. It would not be a methodologically valid comparison to compare the same phenomenon in, for instance, a corpus of newspaper texts to the

same phenomenon in a corpus of informal spoken language. For this reason, I have chosen a wide range of corpora in all three languages, to facilitate comparison within and between each language.

In using the combination of corpora outlined above, care was taken to ensure that I did not count data from the same texts twice. This is relevant when comparing corpora whose time spans overlap; the latter part of CLMET3.1 (1810-1920) coincides with the first century of COHA. Further, COHA and COCA make partial use of the same texts (corpus.byu.edu/coha); duplicate tokens of the *way*-construction were therefore excluded manually.

This section of the thesis has provided an overview of the corpora used in the three languages, showing that a very large dataset has been used from a wide range of sources, ranging from very formal texts from the 8th century to modern web-based data. Some methodological considerations concerning the comparison of different corpora were also discussed. In the following section, I describe in detail the process of collecting and coding the data from the three languages. I also discuss some challenges that arose during this process and the steps I took to resolve them.

3.4 Collecting and coding the data

3.4.1 Collecting and coding the English data

3.4.1.1 What counts as an instance of the English *way*-construction?

To be considered an instance of the *way*-construction, tokens had to conform to the schema [NP$_i$ [V POSS$_i$ (ADJ) *way* DIR]]. The verb in the construction had to denote the means or manner of motion, or an incidental activity accompanying that motion. Following Perek (2018), tokens where the noun *way* is used in an adverbial phrase, e.g. *He came our way*, were excluded. Idioms such as *have/get/lose one's way* were also excluded. All irrelevant or duplicated data were manually excluded.

3.4.1.2 Collecting the data from ARCHER

As the ARCHER corpus is untagged, I searched the corpus for all instances of the noun *way* and its alternative spellings between 1650 and 1900[28]. All concordances were downloaded in KWIC (keyword in context) format as a *.txt* file and then converted to *.csv* format for use in a spreadsheet. 125 tokens of the *way*-construction were found in total[29].

3.4.1.3 Collecting the data from the BNC

As the BNC is a tagged corpus, I was able to search for all instances of a verb followed by a pronoun, an optional adjectival modifier, the noun *way*, and a preposition by using the query *_{VERB} _{PRON} (_{A})? way _{PREP}*. These tokens were downloaded in KWIC format as a *.txt* file and then converted into *.csv* format. In total, 4,624 tokens of the *way*-construction were found.

3.4.1.4 Collecting the data from CLMET3.1

A copy of the CLMET3.1 corpus was uploaded to concordancing software Wordsmith 7.0 (Scott 1999). All instances of a verb followed by a possessive pronoun, an optional adjectival modifier, the noun *way*, and a preposition were searched for, yielding 1,544 results. Concordance lines were downloaded via the Wordsmith interface as a *.csv* file for later use in a spreadsheet.

3.4.1.5 Collecting the data from COCA

As the COCA corpus is very large and I did not have access to an offline version of the corpus for use with Wordsmith, a number of smaller searches had to be carried out in order to prevent an interface timeout; these are shown in Table 9. As the corpus is tagged, I was able to search for all instances of a verb followed by a possessive pronoun, an optional adjectival modifier, the noun *way*, and a preposition.

[28] A list of these can be found at www.oed.com, s.v. *way*.
[29] All frequency counts reported in this section are after the exclusion of irrelevant or duplicate data.

Concordances were copied and pasted directly from the interface into an Excel spreadsheet.

Table 9: COCA search queries and number of results

Query	Hits
_v* my way _i*	1,793
_v* your way _i*	613
_v* his way _i*	7,156
_v* her way _i*	2,386
_v* its way _i*	3,423
_v* our way _i*	1,166
_v* their way _i*	5,334
_v* my _j* way _i*	37
_v* your _j* way _i*	15
_v* his _j* way _i*	168
_v* her _j* way _i*	67
_v* its _j* way _i*	108
_v* our _j* way _i*	39
_v* their _j* way _i*	122
Total	**22,427**

Key

_v* = any verb

_j* = any adjective

_i* = any preposition

3.4.1.6 Collecting the data from COHA

As COHA is a very large corpus, several smaller searches were carried out. Because COCA and COHA share the same interface, the queries shown in Table 9 were also used to search COHA. The number of results from each query is shown in Table 10; the same key applies.

Table 10: COHA search queries and token frequencies

Query	Hits
_v* my way _i*	2,309
_v* your way _i*	1,505
_v* his way _i*	6,342
_v* her way _i*	3,169
_v* its way _i*	4,805
_v* our way _i*	1,794
_v* their way _i*	6,652
_v* my _j* way _i*	31
_v* your _j* way _i*	19
_v* his _j* way _i*	81
_v* her _j* way _i*	63
_v* its _j* way _i*	50
_v* our _j* way _i*	58
_v* their _j* way _i*	86
Total	**26,964**

3.4.1.7 Collecting the data from the Helsinki corpus

As the Helsinki corpus is untagged, I searched the corpus for all instances of the noun *way* and its spelling and dialectal variations as far back as the 8th century[30]. Concordances were downloaded via the CQPweb interface in KWIC format and saved as a *.txt* file before being converted to *.csv* format for later use in a spreadsheet. 16 tokens of the *way*-construction and precursor constructions were found in total.

3.4.1.8 Coding the data from the English corpora

After collection of the data, all tokens of the *way*-construction were coded for lemma (i.e. *He made his way through the crowd* and *He makes his way through the crowd* both were coded as *MAKE*), means, manner or incidental activity reading, literal or metaphorical path, and semantic category (e.g. 'actions involving the mouth', 'performance'). The semantic categories used broadly follow Perek's (2018), though new ones were created where appropriate. The relevance of these semantic categories will be explained in more detail in Section 6.3.2.

[30] A list of these can be found at www.oed.com, s.v. *way*.

3.4.2 Collecting and coding the Dutch data

3.4.2.1 What counts as an instance of the Dutch *way*-construction?

For the purposes of the investigation, I considered an instance of the Dutch *way*-construction any sentence conforming to the schema [NP_i [V $REFL_i$ *een weg* DIR]]. Both strong and weak reflexives were searched for, e.g. *zich* and *zichzelf*. Tokens where the subject and reflexive pronoun did not co-refer were excluded. As in the English data, the verb in the construction had to encode either the means or manner of motion, or an incidental activity accompanying this motion. In addition, I also examined instances of the construction where the noun *weg* was modified, in order to compare the restrictions on modification of the noun in the English, Dutch and German constructions. Irrelevant or duplicated data were excluded manually.

3.4.2.2 Collecting the data from *De Gids*

A copy of the *De Gids* corpus was uploaded into concordancing software Wordsmith 7.0. The corpus was searched for all instances of the indefinite article *een* followed by the noun *weg* 'way'. Concordance lines were downloaded via the software as a *.csv* file. 133 tokens of the *way-construction* were found in total.

3.4.2.3 Collecting the data from NLCOW14

As the NLCOW14 corpus is very large, a number of smaller searches were used; these are shown in Table 11, along with the number of results returned by each search. Concordance lines were downloaded as a *.txt* file from the NoSketchEngine interface and then converted into *.csv* format for later use in a spreadsheet.

Table 11: NLCOW14 search queries and their token frequencies

Query	Hits
me een weg	451
mezelf een weg	28
je een weg	1,140
jezelf een weg	169
u een weg	147
uzelf een weg	5
zich een weg	6,012
zichzelf een weg	134
ons een weg	1,048
onszelf een weg	12
REFL + een + ADJ + weg	486
Total	**9,632**

3.4.2.4 Collecting the data from SoNaR

All instances of a reflexive pronoun followed by an indefinite article, optional adjectival modifier and the noun *weg* were retrieved using the query *[pos="VNW.*] "een" ? "weg"*, where *VNW.** represents a reflexive pronoun and *?* represents a wildcard. Concordances were exported as a *.txt* file via the Opensonar interface and then converted to *.csv* format for use in a spreadsheet. 766 tokens of the construction were found in total.

3.4.2.5 Coding the Dutch data

All tokens were coded for lemma, means, manner or incidental activity reading, literal or metaphorical path, and semantic category of the verb.

3.4.3 Collecting and coding the German data

3.4.3.1 What counts as an instance of the German *way*-construction?

As mentioned in Section 2.4.2, the German *way*-construction is more schematic than its English and Dutch counterparts. Although the construction usually has the formal schema [NP_i [V [$REFL_i$ $POSS_i$ *Weg*] DIR]], the possessive pronoun can be replaced by an indefinite or definite article. Ludwig (2005) mentions a semantically very similar

construction with the schema [NP_i [V $REFL_i$ OBL]] where the reflexive pronoun is not a semantic argument of the verb, but instances of these were impractical to retrieve, because searching for a verb followed by a reflexive pronoun and a preposition would find far too many instances of true reflexive verbs to exclude manually. As such, I did not search for this reflexive construction. Tokens where the subject and reflexive pronoun did not co-refer, e.g. *Sie zeigte mir den Weg* 'she showed me the way' were excluded. As in the English and Dutch data, the verb in the construction had to encode the means or manner of motion, or an incidental activity accompanying this motion. I also searched for modification of the noun *Weg*, although this was not attested in my data. Irrelevant or duplicated data were excluded manually.

3.4.3.2 Collecting the data from the *Berliner Zeitung* (1994-2005) corpus

As the *Berliner Zeitung* corpus is relatively large, a number of smaller queries had to be carried out. These queries and the resulting number of tokens from each one are shown in Table 12. Data were downloaded from the DWDS interface into a *.txt* file and then converted into a *.csv* file.

Table 12: Berliner Zeitung (1994-2005) search queries and their token frequencies

Query	Hits
mir den Weg	8
mir einen Weg	0
mir meinen Weg	0
dir den Weg	1
dir einen Weg	0
dir deinen Weg	0
sich den Weg	32
sich einen Weg	42
sich seinen Weg	57
sich ihren Weg	38
uns den Weg	10
uns einen Weg	3
uns unseren Weg	0
euch den Weg	0
euch einen Weg	0
euch euren Weg	0
modification of Weg	0
Total	**191**

Tokens where *Weg* was modified were searched for by adding *#2* to the search query between the possessive pronoun or article and *Weg*, to denote a wildcard of maximally two words (to retrieve tokens such as *einen sehr langen Weg* 'a very long way').

3.4.3.3 Collecting the data from the DECOW16A-NANO corpus

A number of smaller searches were carried out due to the very large size of the corpus. These smaller searches and the number of tokens from each one are shown in Table 13. Tokens where *Weg* was modified were searched for by inserting the *?* symbol to denote a wildcard.

Table 13: DECOW16A-NANO search queries and their token frequencies

Query	Hits
mir den Weg	293
mir einen Weg	36
mir meinen Weg	35
dir den Weg	153
dir einen Weg	21
dir deinen Weg	16
sich den Weg	312
sich einen Weg	127
sich seinen Weg	188
sich ihren Weg	211
uns den Weg	419
uns einen Weg	51
uns unseren Weg	25
euch den Weg	101
euch einen Weg	19
euch euren Weg	11
modification of Weg	0
Total	**2,018**

Concordance lines were downloaded from the NoSketchEngine interface as a *.txt* file and then converted to a *.csv* format.

3.4.3.4 Collecting the data from *Deutsches Textarchiv*

Table 14 presents the queries used to search the *Deutsches Textarchiv* corpus and the number of tokens retrieved by each query. To retrieve tokens where *Weg* was modified, I added the string *#2* to the search query between the possessive pronoun or article and the noun, to denote a wildcard of maximally two words.

Table 14: Deutsches Textarchiv *search queries and their token frequencies*

Query	Hits
mir den Weg	11
mir einen Weg	0
mir meinen Weg	6
dir den Weg	3
dir einen Weg	0
dir deinen Weg	0
sich den Weg	3
sich einen Weg	3
sich seinen Weg	0
sich ihren Weg	0
uns den Weg	10
uns einen Weg	0
uns unseren Weg	0
euch den Weg	3
euch einen Weg	0
euch euren Weg	0
modification of Weg	0
Total	**39**

These concordances were downloaded from the DWDS interface as a *.txt* file and converted to *.csv* format for use in a spreadsheet.

3.4.3.5 Coding the data from the German corpora

All data were coded for lemma, means, manner or incidental activity reading, literal or metaphorical path, and semantic category of the verb.

The following section discusses measurements of entrenchment. I show that both token frequency and measures such as the *p*-value of the Fisher-Yates exact test are valid measurements of entrenchment. In this section I also discuss collostructional analysis techniques and how these were applied to my data. I outline some of the

methodological flaws in previous applications of collostructional analysis, and explain the measures I undertook in order to correct these flaws. I go onto discuss VNC algorithms and their application to my dataset, and describe the detection of frequency trends in my data.

3.5 Statistical analysis techniques employed in the study

3.5.1 Measuring entrenchment: a literature review

It is generally assumed in the usage-based literature that token frequency correlates with entrenchment. The term 'entrenchment' refers to the degree to which a construction of any degree of complexity or schematicity forms a unit in the mental grammar of a speaker (Stefanowitsch & Flach 2016: 101), and dates to Langacker (1987: 100), who posits a continuous scale of entrenchment: he proposes that when a structure is used, it becomes more entrenched, and periods of disuse lead to a structure becoming less entrenched.

A salient question presents itself at this point: what token frequency is required for an item to be entrenched? In fact, there is no empirical evidence to support a minimum frequency for entrenchment. Alegre and Gordon (1999) propose a threshold of 6 occurrences per million words for inflected forms, but frequency effects have been observed below that threshold (see, for instance, Baayen et al. 1997, 2007; Arnon & Snider 2010; Blumenthal-Drame 2012; Caldwell-Harris et al. 2012). This view is supported by experimental data from Divjak and Caldwell-Harris (2015), which show that entrenchment can occur without any repetition frequency, since robust memories can be formed by single-trial learning. Further, if a single token of an item is not enough to lead to its entrenchment, nothing would ever be entrenched, as the frequency count in the speaker's mind would always remain at 0 (Gurevich et al. 2010). Divjak and Caldwell-Harris (2015) instead propose that a speaker builds evidence from the first exposure to an item, but reliable hypotheses cannot be formulated until the speaker has gathered sufficient evidence. However, this leads to somewhat circular reasoning, because a question now arises as to what constitutes 'sufficient evidence'.

Recent studies indicate that the relationship between frequency and entrenchment is not as simple as Langacker (1987) suggested. Schmid (2010: 125) says that neither the nature of frequency itself nor its relationship to entrenchment have been understood, and we are no closer to describing this relationship in quantitative terms. Stefanowitsch and Flach (2016: 101) further state that counting raw frequency of occurrence can only measure the entrenchment of minimally complex and minimally schematic items; most linguistic units display more varied degrees of complexity and schematicity (Stefanowitsch & Flach 2016: 105). Treating complex units analogously to simple units by measuring the frequency of the complex unit as a whole is problematic, because it ignores the frequency of the individual components (Stefanowitsch & Flach 2016: 110). For example, the collocation *other people* is more entrenched than *Soviet Union* because *other* and *people* are themselves very highly frequent (ibid.). For this reason, as I mentioned in Section 3.5.1, I applied collostructional analysis techniques to my dataset, which takes into account the frequency of the individual components both outside and within the construction.

A further problem with counting raw token frequencies is that different forms from the same stem contribute jointly to the entrenchment of the stem (Stefanowitsch & Flach 2016: 109). Taken together, the forms *year* and *years* are more frequent in the BNC than the form *people*, and therefore the stem *year* can be said to be more entrenched than the stem *people* (ibid.). However, it is controversial whether morphologically complex items are retrieved whole or as complex entities, though it is agreed that some morphologically complex words are processed as whole units, especially if they are highly frequent or semantically or phonologically opaque, such as *important* and *social* (ibid.).

Another problem with relying on frequency counts from corpora is that individuals differ in their linguistic experience. It is unlikely that a lawyer will have the same linguistic experience as a call centre operative, but a corpus cannot take this into account; a corpus can only measure the linguistic experience of the community as a whole, and it is difficult to know the extent to which this overlaps with the linguistic experience of a particular individual (Taylor 2012: 15). However, it is likely that some phenomena are common to all genres and registers and will be adequately represented in any large corpus (ibid.).

Entrenchment may also cause complex phrases to be stored and retrieved as whole units, or 'chunked' (Bybee 2006: 720; Divjak & Caldwell-Harris 2015: 61; Behrens & Pfänder 2016: 4-5). Chunking comes about by automisation (Diessel 2007: 16), a cognitive process that is not unique to language (Schneider & Chein 2003). Chunks are retrieved whole rather than by combining their parts by a schema (Divjak & Caldwell-Harris 2015: 61). Idioms provide convincing evidence that chunks are stored whole in memory. For example, the word *hale* appears only in the idiom *hale and hearty*, having been lost elsewhere in the language (Bybee 1998: 424). Further, although *disaster* is much less frequent than semantically related nouns such as *accident* and *mishap*, people associate the word *unmitigated* with *disaster*, because *unmitigated disaster* has been chunked but *unmitigated accident* and *unmitigated mishap* have not (Hilpert & Diessel 2016: 6; see also Taylor 2012: 158-161). The chunking of *unmitigated disaster* but not of the other two collocations may be explained in terms of scales; *unmitigated* suggests 'extreme', and *disaster* is nearer to the extreme end of the scale than *accident* or *mishap*, although *catastrophe* is at the far end of the scale and *unmitigated catastrophe* is not chunked either.[31]

Instead of counting raw token frequencies, Stefanowitsch and Flach (2016: 105) propose that the *p*-value of the Fisher-Yates exact test is the best approximation of entrenchment. In the following subsection, I test this hypothesis by exploring the correlation between the token frequency of a verb in a construction and the *p*-value of the Fisher-Yates exact test. Although the raw token frequencies are used in calculating the *p*-value, the calculation is based on a comparison between the observed and expected frequencies. An item can therefore have a very high token frequency but a low *p*-value (i.e. not significant attraction to the construction), if the expected frequency is even higher than the observed frequency. A correlation analysis between the token frequency of a verb in a construction and the *p*-value of the Fisher-Yates exact test is therefore not a circular test.

A strong positive correlation between an item's token frequency and its *p*-value of the Fisher-Yates exact test would support the hypothesis that token frequency is an accurate predictor of entrenchment. I test this hypothesis by examining data from the CLMET3.1 and COHA corpora in English. I do not perform this qualitative analysis on

[31] The string *unmitigated catastrophe* is attested only once in COHA; *unmitigated disaster* is attested 11 times.

the COCA data, because its temporal range overlaps with the latter part of COHA and COCA makes use of some of the same texts as COHA, which means a qualitative analysis on this corpus would yield no additional insight in comparison with COHA. I divide the corpora into periods of 20 years, ignoring the VNC periodisation algorithms in order to obtain as many datapoints as possible. I compare the change in token frequency with the change in the *p*-value of the Fisher-Yates exact test for every verb in the construction, and calculate the correlation between the two measures using Pearson's product moment correlation coefficient *r*. This coefficient has a value between -1 and 1. A value of 1 indicates a perfect positive correlation between the two variables, and a value of -1 indicates a perfect negative correlation; a correlation coefficient of 0 suggests that there is no relationship between the two variables.

Because I do not have information on the word counts per decade for any of the Dutch and German data except the *De Gids* corpus, where the type and token frequency of the construction is too low for a correlation analysis to be informative, I use Kendall's rank correlation coefficient to compare the rank of each verb in terms of token frequency and the *p*-value of the Fisher-Yates exact test in order to determine whether there is a statistically significant difference between the two ranks. Kendall's rank correlation coefficient is a τ value of between -1 and 1. A τ coefficient of 1 indicates that all items have an identical rank between the two variables (in this case, token frequency and the *p*-value of the Fisher-Yates exact test), and a τ coefficient of -1 indicates that the ranks are maximally different, i.e. that one rank is the reverse of the other. If the two variables are independent, a τ coefficient of approximately 0 is expected (Kendall 1938).

3.5.2 Results

3.5.2.1 English

Figure 5 below shows the correlation coefficients[32] between token frequency and the *p*-value of the Fisher-Yates exact test of each verb in the CLMET3.1 corpus. To calculate these, the change in token frequency of each verb over each 20 year period was subjected to a correlation analysis with the change in the *p*-value of the Fisher-

[32] Unless stated otherwise, all correlation tests performed in this study use Pearson's product moment correlation coefficient.

Yates exact test. The average correlation coefficient is 0.672, which suggests that there is a somewhat strong correlation between token frequency and entrenchment. This correlation coefficient was found to be statistically significant at $p < 0.01$.

Figure 5: Correlation coefficients between token frequency and the p-value of the Fisher-Yates exact test of each verb in CLMET3.1

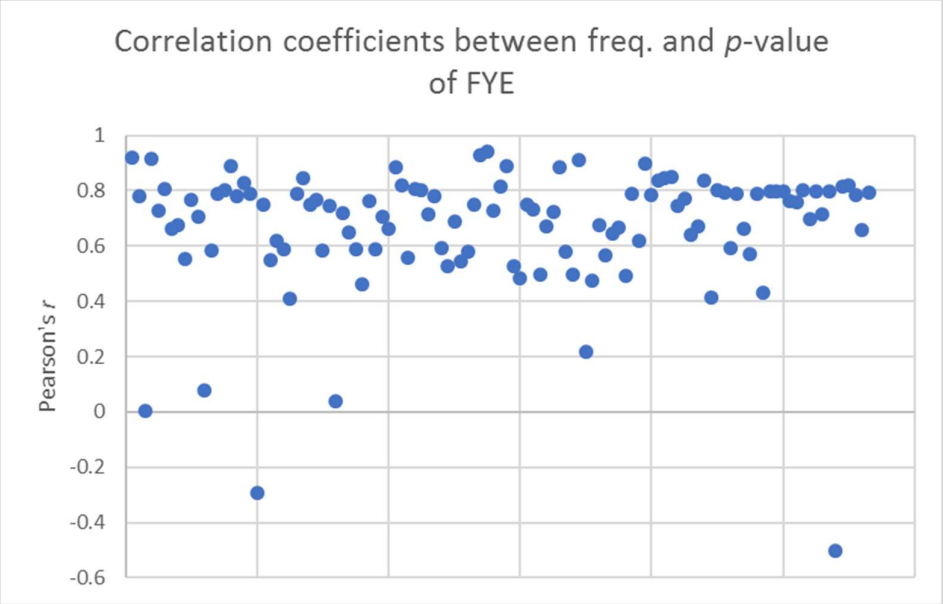
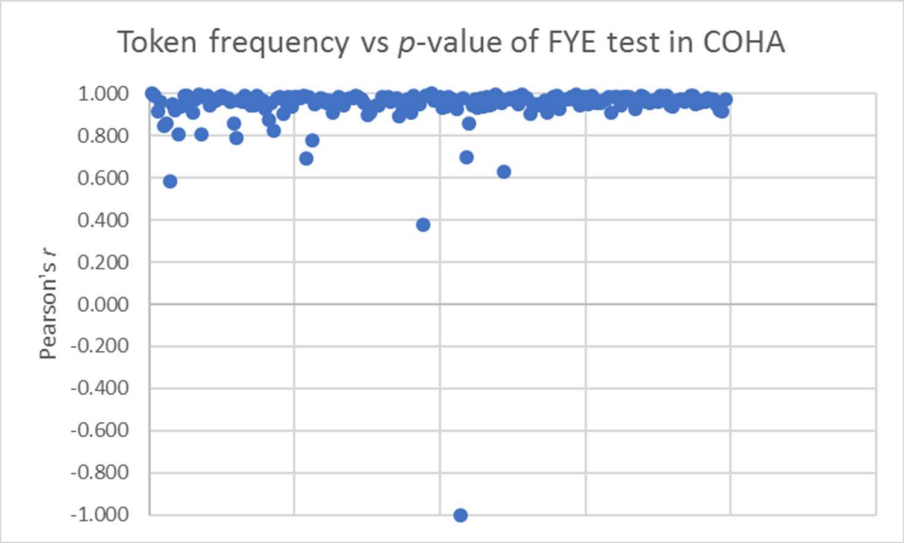


Figure 6 shows the correlation coefficients between token frequency and the *p*-value of the Fisher-Yates exact test for the verbs in COHA. For ease of interpretation, I only plot the 200 most frequently occurring verbs. There is a very strong correlation between these two variables; the average correlation coefficient is 0.937, and is statistically significant at $p < 0.01$. Notably, one of the verbs shows a correlation of -1, i.e. that as its frequency increased, its *p*-value decreased.

Figure 6: Correlation coefficients between token frequency and the p-value of the Fisher-Yates exact test for 200 most frequently attested verbs in COHA



In sum, the data presented here suggest that token frequency and entrenchment are closely linked, and that relying on token frequencies may not be as simplistic as is sometimes assumed (see e.g. Gries 2014b).

3.5.2.2 Dutch

Figure 7 below presents the rank of each verb in the SoNaR in terms of its token frequency and its *p*-value of the Fisher-Yates exact test. Figure 50 shows that there is a strong correlation between these two variables (Kendall's $\tau = 0.800$, statistically significant at $p < 0.01$), which suggests a close link between token frequency and entrenchment.

Figure 7: Token frequency and p-value rank for each verb in the SoNaR corpus

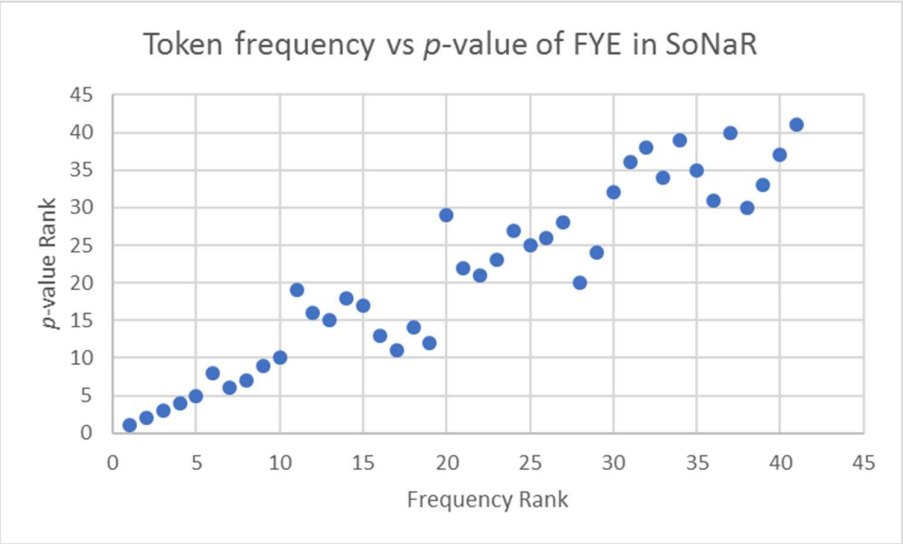
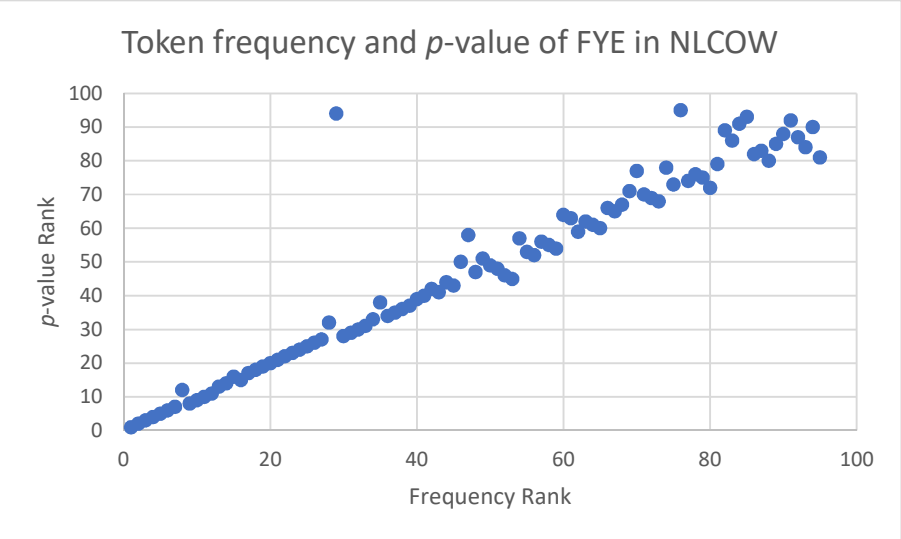


Figure 8 below shows the correlation between the same two variables in the NLCOW corpus. Only the 100 most frequently occurring verbs are shown here.

Figure 8: Token frequency and p-value rank for each verb in the NLCOW corpus



The data from NLCOW also show a very strong positive correlation between the token frequency of an item and the *p*-value of the Fisher-Yates exact test (Kendall's τ = 0.899, statistically significant at $p < 0.01$). Data from the Dutch corpora support the

hypothesis that the link between token frequency and entrenchment may not be as weak as sometimes assumed.

3.5.2.3 German

Figures 9 and 10 show the correlation between token frequency and the *p*-value of the Fisher-Yates exact test in the *Berliner Zeitung* and DECOW corpora respectively. The rank of each verb in terms of frequency is plotted on the *x*-axis, and the rank of each verb's *p*-value on the *y*-axis.

Figure 9: Token frequency and p*-value rank for each verb in the* Berliner Zeitung *corpus*

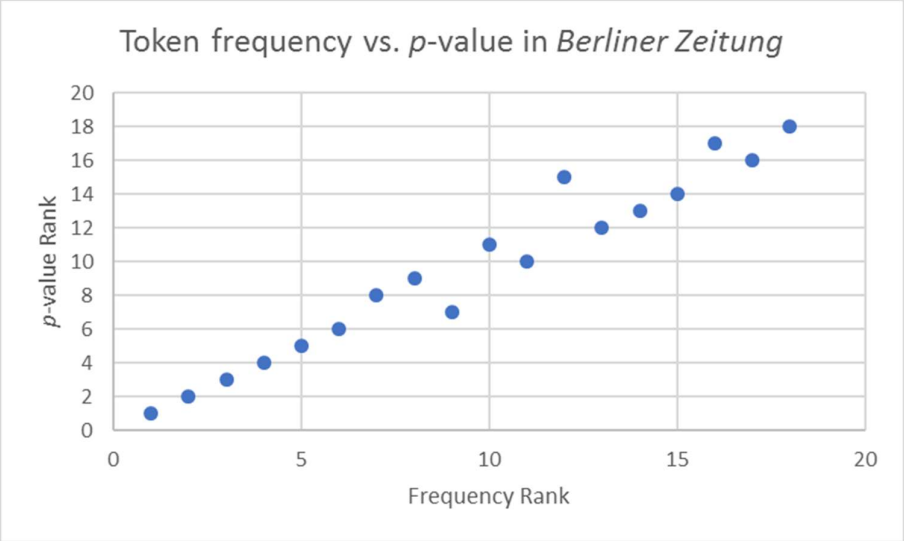
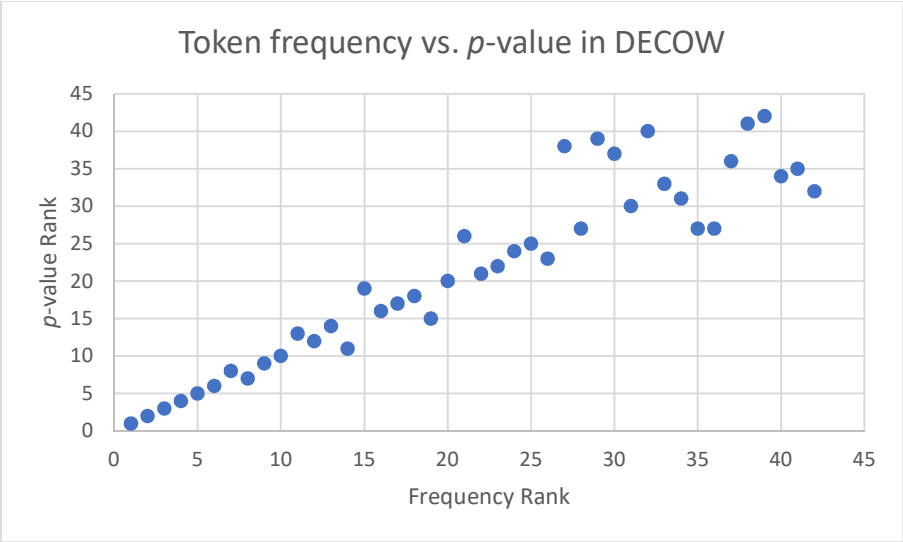


Figure 10: Token frequency and p-value rank for each verb in the DECOW corpus



Data from both corpora show a strong positive correlation between token frequency and entrenchment. The *Berliner Zeitung* data show a Kendall's τ of 0.908, while the DECOW data show a slightly weaker correlation, but nonetheless a very high Kendall's τ coefficient of 0.821. Both coefficients were found to be statistically significant at $p < 0.01$.

In sum, data from all three languages show that raw token frequency may not be as crude a tool as is assumed; the correlation between token frequency and entrenchment was found to be very strong in each language. The following subsection discusses collostructional analysis and its application to my dataset.

3.5.3 Collostructional analysis

Collostructional analysis has its origins in research on collocates (Sinclair 1991, Biber 1993), and focuses on the lexical collocates of grammatical constructions (in e.g. the Goldberg 2006 sense). Pioneered by Stefanowitsch and Gries (2003), collostructional analysis has since been applied in several studies in a diachronic Construction Grammar framework (see e.g. Stefanowitsch 2011 on the ditransitive; Hilpert 2012b on *many a NOUN*). Collostructional analysis is a family of several statistical methods which are used to quantify the strength of the association between an item and one

or more constructions (Gries 2013: 8), and as such, provides a means to track the shift in collocational patterns of a construction over time (Hilpert 2012b: 234).

Collostructional analysis identifies which items are typical in a given construction by comparing the frequency of an item in a construction with its expected frequency. Collostructional analysis therefore builds on work by corpus linguists who conflate prototypicality and frequency; in collostructional analysis, the most frequently attested item in a construction may not be the most typical (Hilpert 2012b: 234). The expected frequency of an item is calculated by comparing the frequency of that item in the construction with the frequency of that item in the rest of the corpus (ibid.). A Fisher-Yates exact test is then used to determine whether the observed frequency is significantly higher or lower than expected; the *p*-value returned by this test is used to calculate how distinctive an item is in a construction (Hilpert 2012b: 238). The more distinct an item is in a construction, the higher its numerical value of collostructional strength. The collostructional strength is a log-transformed probability value based on the *p*-value of the Fisher-Yates exact test (Stefanowitsch & Gries 2003). Collostructional strength values greater than 1.3 indicate that the association between the item and the construction is statistically significant at the level of $p < 0.05$ (Hilpert 2006: 245). A collostructional strength of greater than 2 indicates that this association is statistically significant at $p < 0.01$, and a value of greater than 3 corresponds to a statistical significance at the level of $p < 0.001$ (Gries 2014a). A collostructional strength value of "Inf" means that there is an infinitely small probability that the item is erroneously reported to be attracted to or repelled by the construction (Hilpert 2012a: 142).

Collostructional analysis combines quantitative and qualitative analyses. After the quantitative analysis has been performed, it is up to the researcher to identify semantic categories of the attracted collexemes. For example, in Hilpert's investigation into the *many a NOUN* construction, the distinctive collexemes in each period of his data can be grouped into the following semantic categories: body parts (*eye*, *heart*), human emotion (*thought*, *prayer*), nouns denoting time (*day*, *year*), and human beings (*businessman*, *politician*). While this is, of course, a subjective assessment, his data nonetheless show that quantitative data can be used to inform analysis of qualitative phenomena (Hilpert 2012b: 242). Further, his study undermines

Bybee's (2010: 98) claim that no semantic analysis can arise from collostructional analysis.

In this study, I use two methods from the family of collostructional analysis; a diachronic collostructional analysis and a diachronic distinctive collexeme analysis. Diachronic collostructional analysis is an adaptation of a distinctive collexeme analysis (Hilpert 2012b: 236). A distinctive collexeme analysis compares the collocational patterns of two constructions (see Gries & Stefanowitsch 2004 for a distinctive collexeme analysis of *will* and *BE going to*). A collostructional analysis, on the other hand, analyses the collocational patterns of only one construction. In a diachronic distinctive collexeme analysis, the same construction can be investigated at different points in time (Hilpert 2012a: 142); this is therefore a useful tool to investigate diachronic developments in the *way*-construction. Diachronic distinctive collexeme analysis abstracts away from items that are equally typical in all periods, focusing instead on those that are significantly more frequent than expected in one particular period (Hilpert 2012a: 143). Combined with qualitative analysis, this allows the researcher to investigate whether there are semantic commonalities in these items, and thus identify changes to the semantics of a construction.

There are, however, some problems and challenges concerning collostructional analysis. Collostructional analyses to date have worked on the basis of raw frequencies, rather than normalised frequencies; if certain items are unevenly distributed in a corpus, this will affect the results (Hilpert 2012b: 242). Variability between genres is also likely to affect the results, given that some items are more typical of certain registers (Hilpert 2012b: 243). I build on previous studies by taking dispersion into account (cf. Hilpert & Correia Saavedra 2017), and testing for genre effects in order to mitigate the possibility of variability between genres affecting the results.[33]

A further challenge to collostructional analysis is that tagged corpora are required in order to find the frequency of a given verb in the corpus (Hilpert 2006: 250). Searching a corpus for all instances of *force* would lead to inaccurate results, because this search will also retrieve sentences where *force* is used as a noun. A collexeme analysis also cannot take polysemous verbs into account (Stefanowitsch & Gries

[33] I test for genre effects in the English *way*-construction in Section 6.3.3.

2008: 149). For example, when searching a corpus for the verb *make*, uses of *make* in a light verb construction, as in *make a request*, are also retrieved. Due to the extremely high token frequency of *make*, it is impractical to inspect all tokens of the verb. Despite this challenge, collostructional analysis remains a highly informative tool, and I use it even for polysemous verbs.

Previous applications of collostructional analysis have contained methodological flaws. Baayen (2011b) points out that deriving a measure of collostructional strength from the *p*-value of a statistical test is odd; it is generally accepted in the statistical literature that *p*-values must not be seen as effect measures (Schmid & Küchenhoff 2013: 539; fn. 5). What the *p*-value actually measures is the likelihood of rejecting the null hypothesis, i.e. that there is no attraction between the word and the construction (ibid.). A further problem arises from using *p*-values when they are not done on identical corpus sizes (cf. Gries 2005). If the corpus size increases, the *p*-value decreases (Schmid & Küchenhoff 2013: 540). *P*-values can also be affected by high absolute frequencies of items outside the construction (Schmid & Küchenhoff 2013: 547).

Perhaps the most significant methodological flaw in many studies based on collostructional analysis is the assumption that the two-way association between an item and a construction can be captured by one single measure; the use of *p*-values often conflates different types of associations that should be kept apart (Schmid & Küchenhoff 2013: 545-546). As Ellis and Ferreira-Junior point out (2009: 198), these associations may not necessarily be reciprocal in strength. The difference in directionality of the associations is summed up by Ellis and Ferreira-Junior (2009: 203) as follows:

> "When a construction cues a particular word, that word occurs very often in that construction and it tends to be very generic. When a word cues a particular construction, it may be a lower frequency word, quite specific in its [. . .] semantics and thus very selective of that construction."

(as quoted in Schmid & Küchenhoff 2013: 565).

The difference between these two associations can be seen in the English *way*-construction; the verb *make* occurs very often in the construction and so the *way*-

construction can be said to cue *make*, whereas the verb *wend* cues the *way*-construction, because it is only found in this construction.

In order to avoid conflation of the two types of association, Gries (2014) follows Ellis and Ferreira-Junior's (2009) suggestion and computes in his collostructional analysis script two *ΔP* measures known as 'attraction' and 'faith'; the latter measure is sometimes known in the statistical literature as 'reliance' (see e.g. Schmid 2000). The faith, or reliance, of an item in a construction is a percentage of how many instances of this item occur in the construction. The *ΔP* measure is the contingent probability of a given construction attracting a given lexeme (Schmid & Küchenhoff 2013: 551). *ΔP* measures yield effect sizes rather than *p*-values as measures of attraction or repulsion (Schmid & Küchenhoff 2013: 555), thus responding to Stefanowitsch and Gries's (2009: 943) suggestion to use effect sizes as an alternative measure of association strength. For this reason, I use *ΔP* measures in this study.

Diachronic distinctive collexeme analysis has also been applied improperly in prior studies (Hilpert 2012a: 157). In some of these studies (see e.g. Hilpert & Gries 2012), it has been assumed that all lexical items have an equal chance of occurring in each of the subperiods. This is an erroneous assumption; some items may increase in frequency over time or even fall out of use altogether (Hilpert 2006: 250). The methodology of the present study builds on that of previous studies by taking into account the frequency of a verb in the corpus in each period, where possible, and performing a collostructional analysis based on these frequencies.

3.5.4 Variability-based neighbour clustering

Many researchers assume that there is only one sensible way of dividing a corpus into parts (Gries & Hilpert 2008: 61). To do this, they often divide the corpus into arbitrary periods of time such as 10 or 50 years, or divide the corpus into parts based on visual inspection of the data (Gries & Hilpert 2008: 59; Hilpert & Gries 2009: 386). These approaches are flawed; in the former approach, higher-level generalisations that arise from grouping different temporal stages may be lost. In addition, arbitrary divisions of time may not correspond with the actual stages of development in a construction (Hilpert 2012a: 135). In the latter approach, different researchers may arrive at different conclusions based on the same dataset (ibid.). Hilpert and Gries (2012) propose exploratory bottom-up approaches as a solution to this problem. In

such approaches, data is processed in a way that reveals structures that may not necessarily have been anticipated by the researcher. VNC is an example of such an approach (Hilpert & Gries 2016: 48). In VNC, data from different time periods are grouped together on the basis of their similarity, but only immediately temporally adjacent datapoints are merged (Hilpert & Gries 2009: 389). The VNC algorithm takes as its input the normalised frequency of a construction in each decade of the corpus per million words. Gries and Hilpert (2008) apply VNC to two constructions: verbal complements with *shall*, and the development of perfective aspect markers in English. They show that applying VNC to diachronic data allows the researcher to posit historical stages of a construction that one chooses to study, and allows discrimination of distinct developments that happen in these stages (2008: 75). However, they advise caution when applying VNC. They suggest implementing different similarity measures to get a maximally clear picture of what the data look like (2008: 77), and point out that the results can be influenced either by the phenomenon that is being investigated or by the corpus. Follow-up analyses of genre effects were therefore carried out in order to control for this confounding variable.

3.5.5 Applying collostructional analysis and VNC to the dataset

3.5.5.1 Application to the English data

A collostructional analysis was applied to each of the BNC, CLMET3.1, COCA and COHA datasets. Data from CLMET3.1 and COHA were periodised using the VNC algorithm based on normalised frequencies per million words; the BNC and COCA data were not periodised as these corpora cover a relatively short time span and therefore do not require further division into smaller subperiods. The outputs of the VNC algorithms of the entire CLMET3.1 and COHA datasets are shown in Figures 11 and 12 below.

Figure 11: CLMET3.1 VNC algorithm output

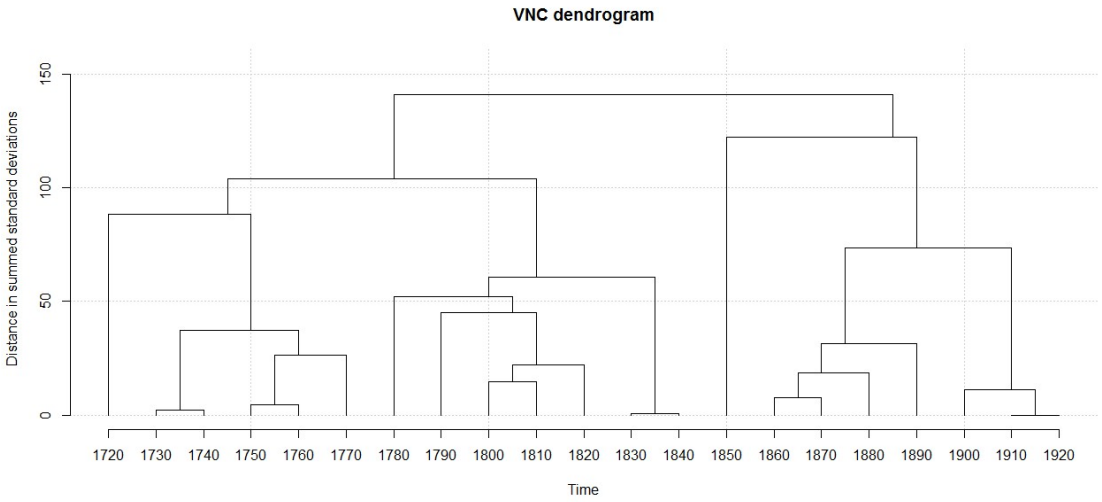
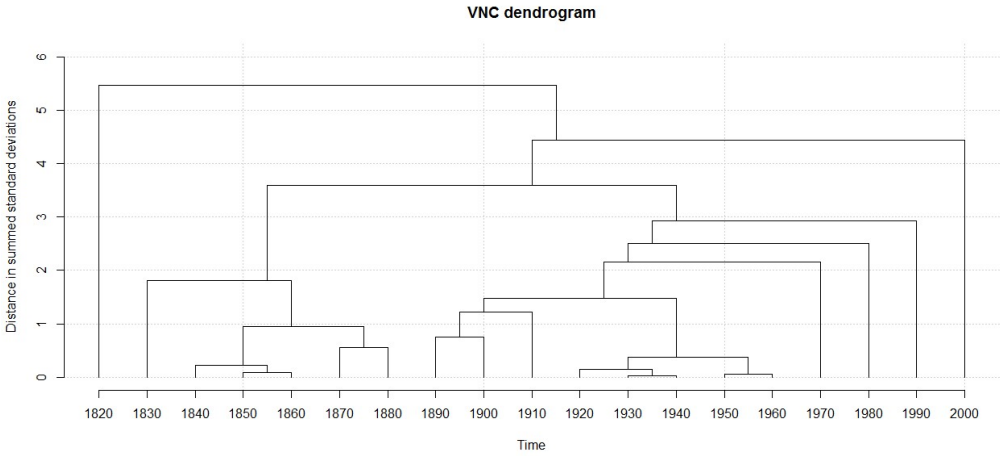


Figure 12: COHA VNC algorithm output



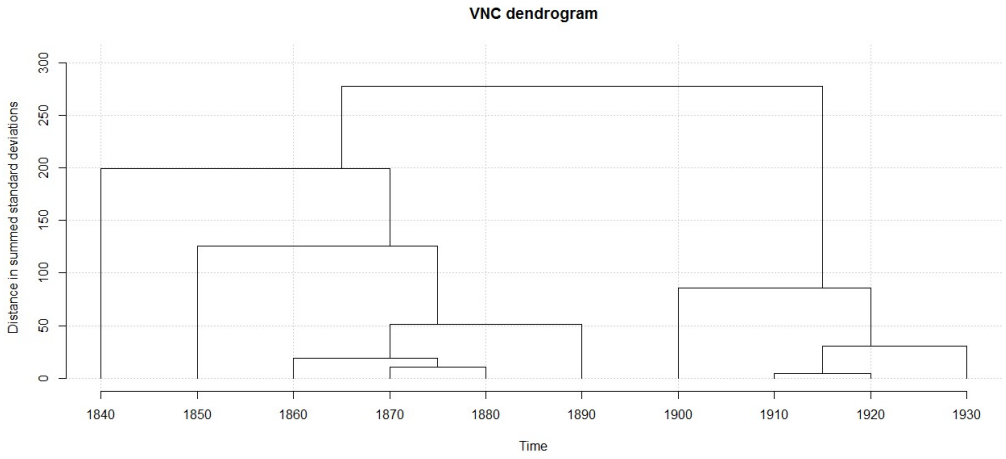
Based on the output of these algorithms, I grouped the CLMET3.1 data into the following subperiods: 1720-1750, 1750-1780, 1780-1890, and 1890-1920. Figure 5 shows that De Smet's (2005) grouping of the CLMET3.1 corpus into 70 year subperiods does not make sense for an analysis of the *way*-construction. Based on Figure 6, the COHA corpus was split into two subperiods: one covering the period from 1820 to 1910, and the second covering the period from 1910 to 2000. Diachronic distinctive collexeme analyses were carried out based on these periodisations, and

the VNC algorithm is run for every dataset, each time producing a different partition. In doing this, I build on Hilpert's (2012a) methodology. Hilpert does not perform the distinctive collexeme analysis based on the frequency of each item in the subperiods of the corpus; as he (2006: 250) notes, words may have a different frequency in each of the subperiods. Following Stefanowitsch and Gries (2003), I took this into account by calculating the corpus frequency of each verb in each subperiod.

3.5.5.2 Application to the Dutch data

Applying collostructional analysis and VNC algorithms to the Dutch dataset posed far more problems than applying these to the English dataset. Although the NLCOW14 corpus is tagged, a collostructional analysis may lead to inaccurate results because this corpus contains several instances of repeated data. While repeated tokens of the *way*-construction were able to be excluded, it would be impractical to inspect every token of every verb occurring in both the corpus and the construction and exclude the repeated ones manually. For this reason, accurate frequencies of each verb in the corpus could not be calculated, and given the extensive nature of the duplication, the frequencies are likely to be inaccurate; for this reason, I apply a combination of qualitative and quantitative analysis to analyse the Dutch data. Information on the year of publication is also missing from almost all texts in the NLCOW and SoNaR corpora; VNC algorithms therefore could not be applied to these datasets. Collostructional and diachronic distinctive collexeme analyses and a VNC analysis were applied to data from *De Gids*. The output of the VNC algorithm used to analyse the *De Gids* corpus is shown in Figure 13.

Figure 13: De Gids *VNC algorithm output*



Based on this output, I made the following periodisation: 1840-1870, 1870-1915, 1915-1930. A diachronic distinctive collexeme analysis was not applied to this dataset, because the type frequency of verbs in the construction in this corpus is very low, and almost all tokens of the construction contain the verb *banen*; a diachronic distinctive collexeme analysis of this dataset would therefore yield very little insight.

3.5.5.3 Application to the German data

As was the case for the Dutch dataset, applying collostructional analysis and VNC algorithms to the German dataset was challenging. The DECOW16A-NANO corpus, as it is hosted by the same interface as its Dutch counterpart, also contains several instances of duplicated data, rendering an accurate collostructional analysis impossible. I therefore combined quantitative and qualitative approaches to analyse the German data. Information regarding the year of publication of the text is also missing, and therefore a VNC algorithm could not be applied. A VNC analysis was also not applied to the *Berliner Zeitung* or *Deutsches Textarchiv* data. In the former case, because the corpus only covers a very short period of time (1994-2005), and in the latter case because the token and type frequency of the *Way-construction* in this corpus is so low that it is unclear what insights could be gained from performing a VNC analysis. Instead, a collostructional analysis was applied to these datasets.

3.6 Conclusion

This chapter of the thesis has described the methods of the investigation and explained how they are novel in both scope and method. Section 3.2 presented some justifications for the application of corpus linguistics in a diachronic Construction Grammar framework, showing that there are many processes of constructional change that can be discovered by looking at corpus data. In Section 3.3, I listed the corpora used in the study, and explained the reasons for using these corpora. It was shown that this investigation uses corpora from a variety of genres and covering a vast time span; the texts used range from very formal texts from the 8th century to highly informal web-based data. Some disadvantages of using these corpora were presented in this chapter, in particular the very small size of some of the corpora. Section 3.4 outlined the process of data collection and coding the tokens of the *way*-construction in the three languages. It was shown that the process of data collection was problematic in some cases because some of the corpora are either untagged, or so large that several smaller queries had to be carried out. In Section 3.5, I showed that there is a strong positive correlation between token frequency and the *p*-value of the Fisher-Yates exact test, which shows that raw token frequency may not be as crude a tool as has sometimes been assumed (cf. Gries 2014b). Section 3.5 also described the statistical methodology employed in this study. I showed that, while statistical analyses such as collostructional analysis and VNC are useful, they have been misapplied in many studies, and that this study corrects some of the flaws of previous studies. The following chapter deals with the role of reanalysis in the development of the English, Dutch, and German *way*-construction. In this chapter I summarise the relevant debates on reanalysis both in a diachronic Construction Grammar framework and more generally. I also discuss the theoretical implications of my data for some of the positions taken in these debates, and use my data to refine some of the principles of reanalysis.

This page intentionally left blank

4. The constructionalization and constructional changes of the *way*-construction in English, Dutch, and German

4.1 Introduction

This chapter of the thesis is concerned with the constructionalization and constructional change of the *way*-construction in the three languages. This chapter answers my third research question: what led to the constructionalization of the *way*-construction in each language, and what post-constructionalization constructional changes took place in each language? The remainder of the chapter is structured as follows. In Section 4.2 I define constructionalization and constructional changes. I follow Traugott and Trousdale (2013) in defining constructionalization as the emergence of a new form-meaning pairing in the constructional network, and constructional changes as changes affecting the dimension of an existing construction. Section 4.3 assesses the role of ambiguity in mechanisms of constructionalization and constructional changes. In this section, I show that ambiguous uses of *way* may have played a role in the constructionalization of the *way*-construction by facilitating the neoanalysis of *way* as a meaningless syntactic marker of the construction. On the other hand, I find this explanation not to be plausible in the case of the Dutch and German constructions, where non-referential paths are more strongly associated with the reflexive construction. Section 4.4 presents evidence for the constructionalization of the *way*-construction in each language. I demonstrate that a new form-meaning pairing has emerged in each language. In all three languages, a post-verbal directional argument is now required, whereas it was not required in the precursor constructions. Further, the English construction contains non-referential uses of *way*, whereas *way* functioned as a referential noun in the precursor constructions. The possessive pronoun before *way* is now fixed; in the precursor constructions, *way* could occur with a range of determiners and even in the plural (Traugott & Trousdale 2013: 87). The item occurring before *weg* in Dutch has also been fixed; while indefinite and definite articles

were attested in the precursor constructions, an indefinite article is now obligatory in the modern Dutch *way*-construction. In German, on the other hand, the item occurring before *Weg* is not fixed; a definite or indefinite article as well as a possessive pronoun may occur. Section 4.5 recasts traditional accounts of language change in more constructional terms. I show that, rather than relying on notions such as rebracketing and relabelling, the neoanalysis of *way*/*weg* is actually a decategorialization; it has lost some nominal properties in that it cannot be pluralised, but it has retained some nominal properties in that it can be selected by a possessive pronoun or article. Section 4.6 examines the post-constructionalization constructional changes affecting the *way*-construction in each language. Host-class expansion is said to be a typical such change (Traugott & Trousdale 2013: 27); I show that host-class expansion has affected the *way*-construction in all three languages, because the number of verbs that can occur in the construction has increased considerably. Further, I show that the schema of the English and Dutch construction has expanded to allow an adjectival modifier of *way*/*weg*, but that this is decreasing in frequency.

4.2 Defining and operationalising constructionalization and constructional changes

As stated in my introduction, some scholars (e.g. Boye & Harder 2012; Traugott & Trousdale 2013) distinguish between constructionalization and constructional changes, although not all scholars make this distinction (e.g. Hilpert 2013). Constructionalization is defined as the creation of a $form_{new}$-$meaning_{new}$ pairing (Traugott & Trousdale 2013: 1). We cannot know constructionalization has taken place until new constructs emerge that cannot have been sanctioned by pre-existing types (Traugott & Trousdale 2013: 22). This suggests that constructionalization of the *way*-construction has taken place in English, Dutch, and German, because in all three languages verbs with no causal relation to motion are used in a construction that expresses motion; *whistle one's way home* was not sanctioned by the either of the precursor constructions to the English *way*-construction, e.g. *He went his way* and *He took his way home*. Constructional changes, on the other hand, only affect features of an existing construction, and do not result in the creation of a new node in the constructional network. As I will show in the following sections, the development of

the *way*-construction in the three languages consisted of a constructionalization which was preceded by and followed by a series of micro-step constructional changes.

In their discussion of constructionalization, Traugott and Trousdale (2013) make a further distinction between lexical constructionalization and grammatical constructionalization. In lexical constructionalization, a contentful $form_{new}$-$meaning_{new}$ pairing emerges in the constructional network. By 'contentful', Traugott and Trousdale (2013: 12) refer to material that can be used referentially in the discourse. Grammatical constructionalization, on the other hand, results in the creation of a new procedural $form_{new}$-$meaning_{new}$ pairing; procedural material cannot be used referentially. The *way*-construction straddles the boundary between contentful and procedural. It contains contentful elements, because there are referential differences between *force* and *giggle one's way through the room*, and it also contains procedural elements (i.e. the encoding of a motion event) (Traugott & Trousdale 2013: 13). Nonetheless, a new grammatical construction has been created, which suggests that grammatical constructionalization has taken place here.

As mentioned above, constructional changes may precede and follow constructionalization. Constructional changes that precede constructionalization usually involve a mismatch between form and meaning (Traugott & Trousdale 2013: 27); as will be shown in Sections 4.3 and 4.4, this occurred in the *way*-construction in each of the three languages, because the neoanalysis of *way*/*weg* from referring to a literal path to a meaningless syntactic marker of the construction led to a situation where the syntax of the verb and the semantics of the construction no longer corresponded. Constructionalization may also lead to further constructional changes; these typically involve host-class expansion (Himmelmann 2004). Since its constructionalization in the three languages, the *way*-construction has undergone considerable host-class expansion, because the collocational range of the verbal slot has increased considerably, from a handful of basic motion verbs to hundreds of verbs, many of which have no causal relation to motion.

Trousdale (2008, 2010, 2012) argues that in grammatical constructionalization, the schematicity and productivity of a construction increases, and its compositionality decreases. However, in a later work, Trousdale and Traugott (2013) find that an increase in productivity and schematicity is common to both lexical and grammatical

constructionalization, rather than being characteristic only of grammatical constructionalization. The *way*-construction in English, Dutch, and German has certainly undergone an increase in schematicity, because the verbal slot of the schema itself has expanded to allow a wider range of verbs (cf. Traugott & Trousdale 2013: 116). The productivity of the construction in each language has also increased, because the construction allows for the use of hundreds of novel verbs. By "decrease in compositionality", Traugott and Trousdale (2013: 120-121) refer to an increase in the mismatch between the syntax and the semantics of the construction; this has arisen in all three languages, because a superficially transitive structure (i.e. a verb plus a determiner and *way*/*weg* in direct object position) is being used to encode an intransitive motion event.

The following section discusses neoanalysis[34] as a mechanism of constructional change preceding constructionalization. Neoanalysis has been the primary mechanism of change discussed in the grammaticalization literature (Traugott & Trousdale 2013: 35), and is regarded by Traugott and Trousdale (2013: 36) as a micro-step in a constructional change. In the following section I give particular focus to the role of ambiguity in the pre-constructionalization constructional change of neoanalysis. I show that, while ambiguity may play a role, other explanations for neoanalysis need to be considered; in the case of the Dutch and German *way*-constructions, ambiguous uses of the noun *weg* 'path' is not a plausible explanation. I hypothesise in this section that metaphorical uses of the noun *way* outside of the English *way*-construction led to its constructionalization, because these metaphorical uses enabled the neoanalysis of *way* as a meaningless syntactic marker of the construction; the noun *way* referencing a literal path cannot be discursively inert.

[34] 'Neoanalysis' is mostly known in the literature as 'reanalysis'. Traugott and Trousdale (2013) prefer the term 'neoanalysis' because a speaker does not reanalyse, only analyse; for this reason, I use the term 'neoanalysis' in the remainder of the chapter.

4.3 The role of ambiguity in processes of constructionalization and constructional changes

4.3.1 Literature Review

Even agreeing upon a definition of ambiguity has proved problematic. One definition that has been proposed is that a sentence is ambiguous if it has two interpretations which are independently supported (Harris & Campbell 1995: 70; Fortson 2003: 651; De Smet 2014: 33). Another definition states that a sentence is ambiguous if it has more than one interpretation and each possible interpretation conforms to an existing structure in the language, but this understanding of structural ambiguity is undermined by instances of innovative reanalysis where a completely new structure has been created, as happened in the development of English auxiliaries (Harris & Campbell 1995: 71).

There is also considerable debate surrounding the extent of the role of ambiguity in neoanalysis.[35] Several linguists (see, among many others, Timberlake 1977: 142; Harris & Campbell 1995: 72; Haspelmath 1998: 326; Andersen 2006: 71; Eckardt 2011: 36) argue that neoanalysis requires ambiguity, though this has been seriously challenged by other linguists. De Smet (2009: 1729-1737), for instance, challenges this view with two case studies, one from English, the other from Dutch. He shows that the role of ambiguity in the neoanalysis of *worthwhile* as an intransitive adjective is difficult to justify, because he finds in the CLMETEV corpus only one sentence where *worthwhile* is ambiguous between the transitive reading (e.g. *It is worthwhile to see a doctor*) and the intransitive reading (e.g. *His hard work was worthwhile*), and only four ambiguous sentences in the CEN corpus, though he warns that a lack of evidence does not constitute counterevidence. His second case study concerns the development of the Dutch intensifying prefix *kei-*, as in *keimooi* 'very beautiful', which developed from the noun *kei* 'pebble'. He points out that *keihard* 'rock hard' is a possible source of this neoanalysis, but it is ambiguous now only because neoanalysis has taken place (De Smet 2009: 1729). The ambiguities that have been said to motivate neoanalysis are instead often the result of neoanalysis (ibid.). He further points out (2009: 1730) that the noun *kei* underwent semantic developments

[35] Ambiguity is sometimes known in the literature as 'opacity' (Hopper & Traugott 2003: 52).

independently of this prefix, which may have triggered its emergence. For instance, *een kei in fysica* (lit. 'a rock in physics') means 'someone who is very good at physics'. Based on this, it appears that there is often much more to neoanalysis, and therefore to constructional changes, than two interpretations and their ambiguity (De Smet 2009: 1730); other potential sources of reanalysis therefore must be considered. The following subsection aims to do this; I test the hypothesis that metaphorical uses of the noun *way* in English and *weg* in Dutch and German outside of the *way*-construction triggered a constructional change in which *way*/*weg* came to be seen as a non-referential object and a meaningless syntactic marker of the construction.

4.3.2 Results

4.3.2.1 English

The first canonical instance of the *way*-construction (i.e. a verb followed by a possessive pronoun, the noun *way*, and a directional argument) is first attested in my dataset 1658; this token is shown in (107).

(107) the then randesvouz of the contrary side [...] fought their way in (ARCHER, 1658).

The first attestation of the noun *way* used metaphorically outside the construction predates this; it dates to the late 15th century, and this token is shown in (107) below.

(108) Thou haste passid the strayte waye and passage of Infortune
You have passed the straight way and passage of Infortune
'You have passed the straight way and passage of infortune' (Caxton tr. R. Le Fèvre Recuyell Hist. Troye (1894), 1473?)

Notably, the metaphorical use of *way* in (108) occurs with a verb of motion; in the mid-16th century, precursor constructions to the *way*-construction with verbs of motion are attested, as in (109).

(109) I toke my waggon and wente my wayes (Helsinki, 1500-1570).

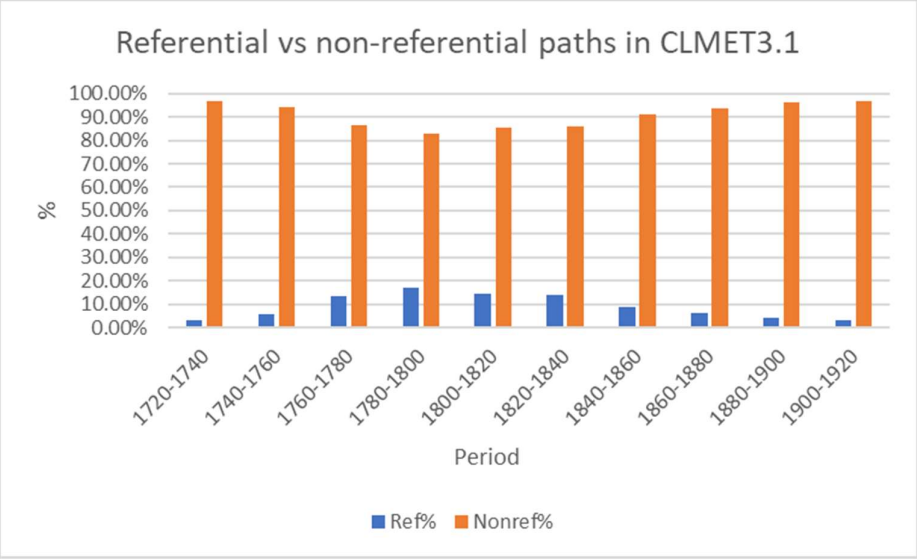
Given that the first attestation of metaphorical *way* predates the constructionalization of the *way*-construction, and that this use occurred with a verb of motion, as did 14 of 16 (87.5%) tokens of precursor constructions to the *way*-construction found in the Helsinki corpus, metaphorical uses of the noun *way* outside the construction may plausibly be said to be a source of the neoanalysis of *way* as a non-referential object. This is because a literal path must always be referential; while metaphorical paths may also be referential in the discourse, I hypothesise that non-literal paths are easier to construe as non-referential. I therefore argue that metaphorical uses of the noun *way* outside the construction triggered its neoanalysis as a meaningless syntactic marker of the construction. This is an instance of a pre-constructionalization constructional change, which are said to involve a mismatch of form and meaning (Traugott & Trousdale 2013: 27); the use of *way* as a meaningless syntactic marker of the construction is a mismatch between the form *way* and the old meaning associated with this form, i.e. a referential object.

These developments illustrate the role of bolstering in the constructional network. The term 'bolstering' refers to the phenomenon whereby a construction is linked to another construction by a formal or functional alignment, or both (McColm & Trousdale 2019); in the case of the *way*-construction, its status has been affected by a construction that is linked to it by the formal pole, i.e. metaphorical uses of *way* outside the *way*-construction. The following paragraphs show the change in frequency of referential and non-referential paths in the English *way*-construction. I hypothesise that an increase in non-referential paths is a result of metaphorical uses of *way* outside the *way*-construction, because only non-literal objects can be non-referential.

All 16 tokens of the *way*-construction and its precursors in the Helsinki corpus contained a referential path. By contrast, only 4 of 67 tokens in ARCHER contained a referential path. This suggests that between Old and Middle English, that *way*-construction had been constructionalized, such that a new form-meaning pairing had emerged in which the phrase headed by POSS *way* no longer had to be referential.

Figure 14 below shows the relative percentage of referential and non-referential paths in the CLMET3.1 corpus. The output of the VNC algorithm is ignored here in order to obtain as many datapoints as possible. The corpus was divided into chunks of 20 years.

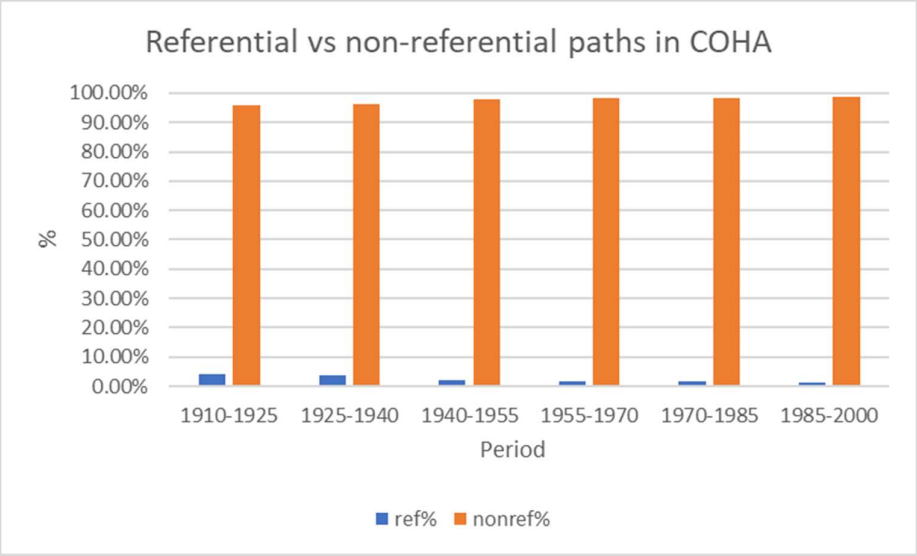
Figure 14: Percentage of referential and non-referential paths in CLMET3.1



The graph above suggests that the preference for a non-referential path decreased between 1720 and 1800. From 1800 onwards, non-referential paths in the construction increased in frequency in each period, such that 96.95% of tokens of the *way*-construction between 1900 and 1920 contained a non-referential path. Performing a correlation analysis on the data from 1800 onwards yields a correlation coefficient of 0.985, which was found to be statistically significant at $p < 0.05$.[36] In Figure 15 I show the relative percentage of referential and non-referential paths in the second half of COHA (i.e. 1910-2000). Once again, the output of the VNC algorithm was ignored in order to obtain as many datapoints as possible. The second half of the corpus was divided into 6 equal periods of 15 years.

[36] The correlation tests and the calculation of the significance of the correlation coefficient were performed in R. All correlation coefficients are Pearson's unless otherwise stated.

Figure 15: Relative percentage of referential and non-referential paths in COHA (1910-2000)



The graph above shows that the preference for a non-referential path was already very strong (95.78%) in the first period analysed. This preference has become even stronger over time, such that 98.59% of tokens of the *way*-construction in the latter period contain a referential path. Performing a correlation analysis on this data yielded a correlation coefficient of 0.911, which was found to be significant at $p < 0.05$. In sum, the findings from CLMET3.1 and COHA appear to confirm my hypothesis that *way* may have been neoanalysed as a non-referential object due to metaphorical uses of *way* outside the construction.

4.3.2.2 Dutch

Metaphorical uses of the noun *weg* in Dutch are attested as far back as the 17th century; one such example is provided in (110).

(110) Volght doch niet den ghemeynen breeden wegh van den grooten Hoop
Follow but not the common wide way of the big hope
'But don't follow the common wide way of the great hope' (Sprankhuisen 7, 50b, 1647).

This token far predates the first instance of a metaphorical path in the Dutch *way*-construction. Metaphorical paths were not found at all in the *De Gids* corpus. One example of a metaphorical path in the *way*-construction is provided below in (111).

(111) Of klikte je je een weg op het Wereldwijde Web
Or clicked you REFL a way onto the worldwide web
'Or you clicked your way onto the worldwide web' (SoNaR, date unknown)

As was the case of the English construction, given that metaphorical uses of *weg* outside the construction predate metaphorical uses of *weg* within the construction, metaphorical uses of *weg* outside of the *way*-construction may plausibly have been a source of its neoanalysis as a non-referential object, because non-literal objects are easier to construe as being non-referential; a literal object must, by definition, be referential. As in the English construction, I argue that the neoanalysis of *weg* as a non-referential object was a pre-constructionalization constructional change. In the paragraph below, I document the change in frequency of referential and non-referential paths in the Dutch *way*-construction. I hypothesise that an increase in frequency of non-referential paths is a result of metaphorical uses of *weg* outside of the Dutch *way*-construction, since only non-literal objects can be non-referential.

Only one of the 133 tokens of the *way*-construction in *De Gids* contained a non-referential path. In the SoNaR corpus, 4.34% of tokens contained a non-referential path, compared to 2.80% in the more modern NLCOW corpus. This shows that, contrary to my initial hypothesis, the preference for a non-referential path is becoming weaker, rather than stronger, in the Dutch *way*-construction. This is perhaps because the most frequent verb in the construction, *banen*, came to be used in the construction after the conventionalisation of the construction *weg banen* ‘smooth a [literal] path’ (Verhagen 2002: 423-424). This suggests that neoanalysis of *weg* as a non-referential object due to metaphorical uses of *weg* outside the construction is not a plausible scenario.

4.3.2.3 German

As is the case in English and Dutch, metaphorical uses of the noun *Weg* in German predate metaphorical uses in the German *way*-construction. One example of a

metaphorical use of *Weg* that exists independently of the *way*-construction is given in (112); this sentence, too, is idiomatic.

(112) Sie seind gangen den weg alles fleisch
They are gone the way of.all flesh
'They have gone the way of all flesh' (Lutz, Wilhelm Friedrich: Ein Christliche Predig. Tübingen, 1585)

Similarly to the English and Dutch constructions, metaphorical uses of *Weg* outside the *way*-construction may be a plausible source of its neoanalysis as a non-referential object. The following paragraph tests this hypothesis in more detail by examining the extent to which non-referential paths are becoming a property of the construction. If metaphorical uses of *weg* were the source of its neoanalysis as a non-referential object, it follows that non-referential paths will have increased in frequency, because only metaphorical uses of *weg* can be non-referential.

As in the Dutch construction, non-referential paths are very rarely attested in the German *way*-construction. In the *Deutsches Textarchiv* corpus, all 46 tokens of the *way*-construction contain a referential path. In the *Berliner Zeitung* corpus, non-referential paths are attested, but only in 2.68% of tokens. In the DECOW corpus, this figure has decreased to 1.39%. This suggests that, while non-referential paths are possible in the German *way*-construction, they are more strongly associated with the reflexive pattern [NP_i V [$REFL_i$ DIR]], as Ludwig (2005) hypothesised.

Summarising my findings on the role of ambiguity in mechanisms of constructional change and constructionalization, the idea that *way* was neoanalysed as a non-referential object of the English *way*-construction due to metaphorical uses of this noun outside the construction is a plausible one. Metaphorical uses of *way* outside the *way*-construction were attested before the neoanalysis of *way* as a non-referential object, and once *way* was neoanalysed as a non-referential object, there was a gradual increase in token frequency of non-referential paths in the construction. These findings do not hold true of the Dutch and German data, however. Although there is a very high type frequency of verbs in the Dutch construction, most of the tokens contain the verb *banen* being used with a referential path. Further, the token frequency of non-referential paths in the Dutch *way*-construction is decreasing, not

increasing, contrary to my hypothesis. In the German construction, the verbs used are mostly of path traversal or path clearing, and referential paths are strongly preferred. The token frequency of non-referential paths has decreased over time, as was found of the Dutch construction. The Dutch and German data suggest that metaphorical uses of *weg* outside of the construction may not have been the source of neoanalysis, and that an alternative explanation may have more merit. The following section is concerned with the constructionalization of the *way*-construction in English, Dutch, and German. I demonstrate in this section that constructionalization has taken place in all three languages, because a new form-meaning pairing has emerged in each of them.

4.4 Evidence for the constructionalization of the *way*-construction in English, Dutch, and German

4.4.1 English

This subsection presents evidence for the constructionalization of the English *way*-construction. I show that the *way*-construction in English is distinct from its precursor constructions, and that a $form_{new}$-$meaning_{new}$ pairing has emerged, and that therefore constructionalization has taken place. I stated in Section 2.2.2 that the English *way*-construction had two precursor constructions; one intransitive, the other transitive. The modern day English *way*-construction differs from these precursor constructions, because the noun *way* in the modern construction does not function as a referential object. Further, the possessive pronoun has become fixed; the precursor constructions allowed a wider range of determiners to occur before *way* (Traugott & Trousdale 2013: 83). I also show in this subsection that an oblique argument after the verb specifying the path is now virtually obligatory; the precursor constructions did not require this oblique argument.

Some of the slots in the constructional schemas of the precursors to the English *way*-construction were less fixed than in the present-day *way*-construction. Instances of *way* in the plural are attested (113), as well as instances of *way* with a definite article (114), an indefinite article (115), with no determiner (116), and with a preposition before the string *POSS way* (117).

(113) I toke my waggon and wente my wayes (Helsinki, 1500-1570).
(114) From Sterling they took the way to Glosco (ARCHER, 1677).
(115) [...] and with our daggers force a way to freedom (ARCHER, 1773).
(116) In the evening the wind grew fresh, and increased till three a'clocke the next morning, so that they made good way in their course (ARCHER, 1654).
(117) And anone as he went on his waye, his servantes met him (Helsinki, 1500-1570).

Pluralisation of *way* is now strongly dispreferred. When the noun *way* occurs in the plural, this is usually in an idiom such as *GO POSS separate/own ways*. However, the following three tokens from the BNC appear to be genuine instances of the *way*-construction with plural marking on *way*.

(118) Both Kenneth and Henry, on the stage or battlefield and off it, talk their different ways to triumph (1989).
(119) Even now, when the boys responsible had become young men and gone off to make their ways in the world (1993).
(120) [T]hen once more we made our ways home stimulated, refreshed and already looking forward to next year (1984).

However, these three tokens represent only 0.05% of the tokens in the BNC. It is therefore not controversial to say that singular *way* has been conventionalised. This further points to the status of *way* as a non-referential object; if *way* functioned as a referential object, pluralisation would be more frequently attested when the subject is plural, because pluralisation of referential *way* is frequently attested.

I turn now to the conventionalisation of the string *POSS way*, by examining data from the CLMET3.1 corpus, which is shown in Table 15 below. 1,862 tokens were examined in total. For the purposes of the analysis, the corpus was divided into equal periods of 20 years.

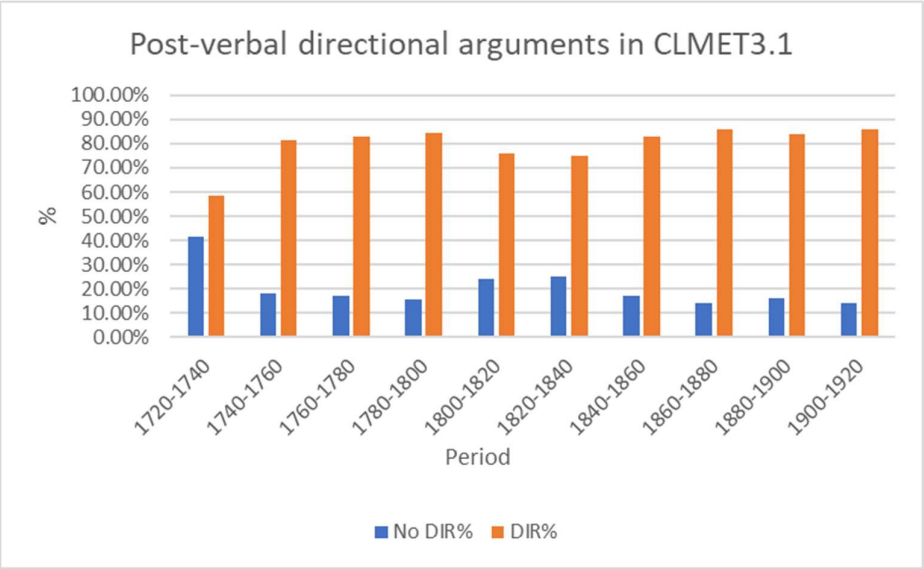
Table 15: Items occupying the slot before the noun way in CLMET3.1

Period	1720-1740	1740-1760	1760-1780	1780-1800	1800-1820	1820-1840	1840-1860	1860-1880	1880-1900	1900-1920
Possessive	96.15%	89.93%	93.06%	90.77%	96.92%	94.51%	95.90%	96.02%	97.25%	95.18%
Article	3.85%	5.76%	2.08%	5.38%	1.54%	3.92%	3.19%	1.77%	1.65%	4.22%
Other determiner	0.00%	2.88%	0.69%	0.77%	0.00%	0.00%	0.23%	0.00%	0.55%	0.60%
No determiner	0.00%	1.44%	4.17%	3.08%	1.54%	1.57%	0.68%	2.21%	0.55%	0.00%

This suggests that the conventionalisation of POSS *way* may have taken place even before 1720, given that 96.15% of tokens between 1720 and 1740 contained POSS *way*. The preference for POSS *way* has fluctuated from 1740 onwards, though it is strongly preferred in each period of the corpus. Subjecting the frequency trend to a correlation analysis yielded a correlation coefficient of 0.509, which was not found to be statistically significant. However, all tokens of the *way*-construction in the second half of COHA (1920-2000) contained the string POSS *way*; this suggests that the possessive pronoun before *way* has become steadily more conventionalised over time, and that the neoanalysis of *way* was a pre-constructionalization constructional change that has led to the constructionalization of a new construction containing POSS *way* in direct object position.

I turn now to the increasing obligatorification of a post-verbal oblique argument in the English *way*-construction. Figure 16 below shows the change in frequency of the oblique argument in CLMET3.1.

Figure 16: Post-verbal directional arguments in CLMET3.1



The graph above shows that between 1720 and 1740, the preference for a directional argument after the verb was not very strong (58.62%). After 1740, roughly 80% of tokens of the construction contain a directional argument, but there is no statistically significant frequency trend in the data at $p < 0.05$ ($r = 0.598$, $p = 0.068$). Data from the BNC, which covers the late 20th century, suggest that the preference for a directional argument has become stronger still; out of 5,962 tokens of the *way*-construction, 5278 contained a directional (88.53%). In the latter part of the COHA corpus (1920-2000), 99.74% of tokens of the *way*-construction contain a directional argument; in combination with the findings from CLMET3.1 and the BNC, this suggests that the post-verbal directional argument has been increasingly conventionalised over time, to the point where it is now virtually obligatory. This is further evidence that a constructionalization has taken place, because the directional argument was not required in the precursor constructions; this points to the existence of a new construction in the constructional network.

4.4.2 Dutch

This subsection presents evidence for the constructionalization of the Dutch *way*-construction. I show that the Dutch *way*-construction differs from its precursor, which contained *banen* 'flatten out' and *weg* 'way', but no oblique argument after the verb (Verhagen 2003b: 231); the modern Dutch *way*-construction must contain a post-

verbal oblique argument. In this precursor construction, the noun *weg* could occur with both definite and indefinite articles (ibid.); as I show in the following paragraphs, the Dutch *way*-construction as it is found today must contain an indefinite article. This suggests that the Dutch *way*-construction has undergone constructionalization, because a $form_{new}$-$meaning_{new}$ pairing has emerged in the constructional network of Dutch speakers.

All instances of the *way*-construction in *De Gids* contained an indefinite article. No instances of a reflexive pronoun followed by a definite article and *weg* were found in the SoNaR corpus, although 41 of the 807 tokens (5.08%) of the *way*-construction contained a definite article with no reflexive pronoun. Of these, 31 contained *vinden* 'find' or *terugvinden* 'find back'. This suggests that, rather than a definite article becoming more preferable, a separate constructional idiom *de weg vinden* 'find the way' has emerged; while this construction is formally similar to the *way*-construction in that both contain the noun *weg*, this construction does not instantiate the *way*-construction. Only 213 instances of a reflexive pronoun followed by a definite article and *weg* were found in the NLCOW corpus, compared to 8,680 instances of the canonical pattern with a reflexive pronoun followed by an indefinite article and *weg* (2.40% vs. 97.60%). Overall, the Dutch data suggest that the pattern with an indefinite article was always very strongly preferred, and that this has not changed significantly over time. The paragraphs below discuss the conventionalisation of the post-verbal directional argument in the Dutch *way*-construction.

Figure 17 below shows the change in frequency in the post-verbal directional argument of the Dutch *way*-construction in the *De Gids* corpus. In order to obtain as many datapoints as possible, I disregarded the output of the VNC algorithm. 131 tokens were examined in total.

Figure 17: Post-verbal directional arguments in De Gids

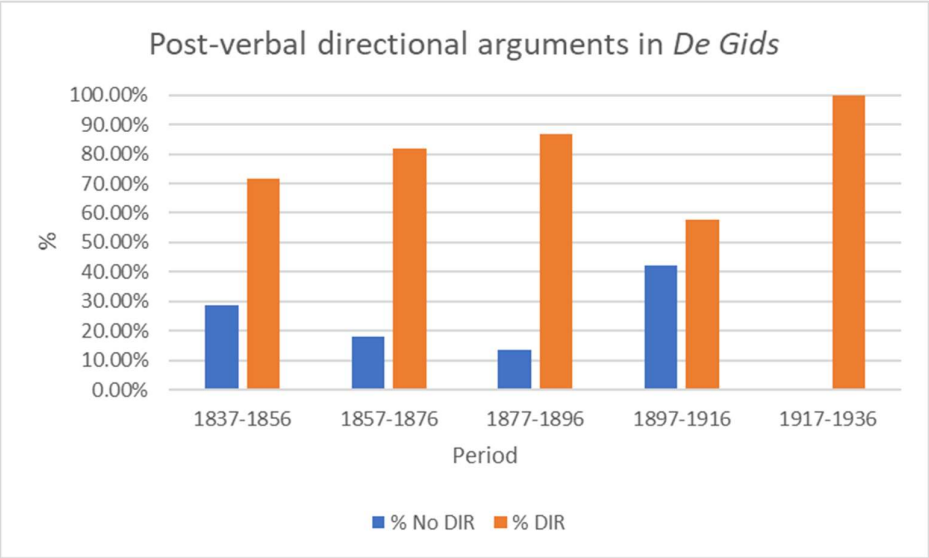


Figure 17 shows that initially, the preference for a directional argument was quite strong (71.43%) and that, with the exception of the period 1897-1916, this preference has increased, such that 100% of tokens in the latter period contain a directional argument. There is no statistically significant frequency trend in the data ($r = 0.330$, $p = 0.587$).

My findings from the SoNaR Corpus, which covers the period from 1954 to 2011, suggest that the preference for a post-verbal directional argument is still very strong. Out of 807 tokens in this corpus, 738 contained a post-verbal directional argument (91.45%). In the NLCOW corpus, which contains data obtained by crawling the web between 2011 and 2014 (Schäfer & Bildhauer 2012; Schäfer 2015), the percentage of tokens of the Dutch *way*-construction containing a post-verbal directional argument is similarly high. Out of 8,680 tokens in that corpus, 7,898 contained a post-directional verbal argument (90.99%).

In sum, my findings from the Dutch corpora suggest that a post-verbal directional argument has become increasingly more obligatory. However, these findings are only tentative given the lack of data between 1930 and 1954, and the near-total absence of information on the year of authorship of texts in the SoNaR corpus. It has been conclusively established, however, that the preference for a post-verbal directional argument is now very strong compared to the earliest data from *De Gids*, and that

therefore a new construction has emerged with a directional. I argue that this constructionalization was facilitated by the neoanalysis of *weg* as a constructional marker.

4.4.3 German

This subsection presents evidence for the constructionalization of the German *way*-construction. As is the case of the English and Dutch constructions, the modern German *way*-construction differs from its precursor constructions. German had two precusors to the *way*-construction; one transitive, the other intransitive.[37] The German *way*-construction as it is found today differs from these precursor constructions in that a post-verbal oblique argument specifying the path is now obligatory; the precursor constructions did not require this argument. The German *way*-construction differs from its English and Dutch counterparts in that the item filling the slot before *weg* is not fixed; a possessive pronoun, a definite article, and an indefinite article can all occur in this slot, though I show below that the preference for a possessive pronoun is increasing over time.

Table 16 below shows the percentage of tokens in each corpus with a possessive, definite, and indefinite article.

Table 16: Percentage of tokens with possessive, definite and indefinite articles by corpus

Corpus	Poss.	def. art.	indef. art.
Deutsches Textarchiv (1473-1927)	26.09%	69.57%	4.35%
Berliner Zeitung (1994-2005)	51.17%	30.43%	18.39%
DECOW	48.21%	33.76%	18.03%

This shows that, over time, a definite article has become markedly less preferred, with the possessive pronoun and indefinite article increasing in frequency at the expense of the definite article.

[37] These precursor constructions are discussed in more detail in Section 5.2.2.3.

I now address the change in preference for a post-verbal directional argument in the German *way*-construction. Given the sparsity of diachronic data found in the German corpora, establishing reliable conclusions on the gradual nature of change is almost impossible. The extremely low frequency of the construction in *Deutsches Textarchiv*, which spans the period from 1473 to 1927, also makes a quantitative analysis difficult; only 46 tokens of the *way*-construction and its precursors were found. I therefore do not provide a quantitative analysis of the *Deutsches Textarchiv* data here.

In the *Berliner Zeitung* corpus, which spans the period from 1994 to 2005, the number of tokens with a post-verbal directional argument has decreased; of the 300 tokens of the *way*-construction in this corpus, 195 contain a post-verbal directional argument, and 105 do not (65% vs 35%). In the DECOW corpus, which contains the most modern data, the number of tokens with a post-verbal directional argument has increased in comparison to the *Deutsches Textarchiv* data. Only 188 of the 865 tokens in this corpus do not contain a post-verbal directional argument, whereas 677 do (78.27% vs 21.73%). This shows that the preference for a post-verbal directional argument has become steadily stronger over time, although the reliability of this finding is challenged by the lack of information on the date of publication of a text in the DECOW corpus. The relatively small number of tokens compared to the English and Dutch corpora, especially in the diachronic *Deutsches Textarchiv* corpus, also poses a significant problem to the possibility of a long-term quantitative analysis of the construction from its first attestation in *Deutsches Textarchiv* to the present day. Despite the low token frequency of the construction in the corpora investigated here, I hypothesise that a constructionalization has taken place, because a directional argument is now strongly preferred, whereas it is not required in the precursor constructions. I argue that this constructionalization was enabled by a pre-constructionalization constructional change that consisted of the neoanalysis of *Weg* as a constructional marker, rather than fully referential noun.

4.4.4 Summary of results on constructionalization

Since the neoanalysis of *way* as a non-referential object, the string *POSS way* has become increasingly conventionalised in the English *way*-construction. Its status as a non-referential object is confirmed by the fact that plural marking on the noun *way* is now only extremely rarely attested. While other determiners were possible in precursors to the *way*-construction, the possessive gradually came to be preferred, a

change that occurred before 1720. It was also shown that the preference for a post-verbal directional argument has become increasingly conventionalised, and is now virtually obligatory. This suggests that a new construction has emerged in the constructional network of English, and that therefore constructionalization of a distinct construction has taken place.

In Dutch, the string *REFL een weg* was very strongly preferred from the beginning of my dataset in 1840, and this has not changed over time. Access to data from before this period would allow precursor constructions to the Dutch *way*-construction to be examined, in order to determine when and how the conventionalisation of *REFL een weg* took place. This in turn would allow the researcher to identify the period when the Dutch *way*-construction was constructionalised. It was also shown that the preference for a post-verbal directional argument in Dutch has become steadily stronger over time and is now much stronger in present-day Dutch than in early data from the *De Gids* corpus. However, the reliability of these findings is challenged by the lack of information regarding the year of publication of the text in the SoNaR and NLCOW corpora.

In the German data, there was a marked decrease in occurrences of a definite article in the *way*-construction between the end of the *Deutsches Textarchiv* corpus in 1927 and the beginning of the *Berliner Zeitung* corpus in 1994. However, given the large gap in the data, it is possible that this change was not as abrupt as my findings suggest; only with access to corpora from this period can the nature of this change be determined more fully. The reliability of this finding is also compromised by the very low token and type frequency of the *way*-construction in the *Deutsches Textarchiv* corpus. The post-verbal directional argument has been steadily increasing in frequency in the German *way*-construction, which points to its conventionalisation; however, the token and type frequency of the *way*-construction in the three corpora investigated is relatively small, which poses a problem to the reliability of these findings. Despite the low token frequency, there is evidence of constructionalization of the German *way*-construction, because a new pattern (NPi V REFL Det *Weg* DIR) has emerged, with the meaning ‘traverse a path DIR by/while V’. This form-meaning pairing is different from the precursor constructions to the German *way*-construction. The following section analyses the constructionalization of the *way*-construction in the three languages in terms of decategorialisation and a movement along the contentful-

procedural cline, rather than in terms of rebracketing and relabelling, as is common in generative approaches. I show that the notions of rebracketing and relabelling lack explanatory power in a constructionist framework, where it is assumed that categories and constituency structure are gradient.

4.5 Rethinking traditional accounts of language change in terms of constructionalization and constructional changes

4.5.1 Literature Review

Reanalysis is considered an important mechanism of language change in both generativist (e.g. Lightfoot 1977; Andersen 2001; Whitman 2012) and functionalist frameworks (e.g. Langacker 1977; Heine et al. 1991; Harris & Campbell 1995; Hopper & Traugott 2003). In constructional terms, Traugott and Trousdale (2013: 36) regard neoanalysis as a micro-step in a constructional change. This section discusses traditional accounts of language change, many of which privilege reanalysis, and casts them in terms of constructionalization and constructional changes. The theoretical issues raised in this subsection are examined in light of my data on the *way*-construction in the three languages in Section 4.5.2.

It is common in generativist approaches to language change to state that reanalysis is a primary mechanism of language change. This reanalysis is said to be motivated by rebracketing, or misparsing, in Hopper and Traugott's (2003) terms (Detges & Waltereit 2002: 153). In rebracketing, the internal structure of a construction is affected, but not its surface form; the hearer simply understands an item to have a different structure from that intended by the speaker (Hopper & Traugott 2003: 50). Thus the reanalysis of the noun *hamburger* 'item of food from Hamburg' involved a hearer assigning to it the internal structure [*ham*][*burger*] instead of the structure intended by the hearer, viz. [*hamburg*][*er*]. It became clear that a reanalysis had taken place when speakers began to coin new items such as *beefburger* and *cheeseburger*, which unambiguously have the structure [*beef*][*burger*] and [*cheese*][*burger*] (ibid.).

The role of rebracketing in reanalysis has been questioned, however, by some linguists. The development of the English preposition *back of* is a much-cited example

of rebracketing (Hopper & Traugott 2003: 41). According to the traditional account, *back of* arose via the following rebracketing: *[back [of the barn]] > [back of] [the barn]*. However, Haspelmath (1998) disputes this account, arguing that there is no evidence for this rebracketing; he argues instead that *back* takes a prepositional complement. He (1998: 344) further argues that if reanalysis is simply the change resulting from the rebracketing of an existing structure, it is less important than is usually assumed. Detges and Waltereit (2002: 154) further argue that rebracketing is epiphenomenal, and is merely a by-product of semantic change. Not all instances of reanalysis arise as a result of rebracketing; adjectival uses of *fun*, for instance, are not the result of rebracketing.

The role of rebracketing is also disputed by Whitman (2012: 73). He argues that the development of *have* perfects, as in *He has written a letter*, did not involve rebracketing. The traditional account of this development states that rebracketing occurred when *letter* was reanalysed as a complement of *written*: *He has [a letter [written]] > He has [written a letter]*, but Whitman (2012: 74) argues instead that the only change is in the label of the projection headed by *have*; according to his account, *have* now selects a clausal argument rather than a direct object. He does not dismiss the notion of rebracketing completely, however, conceding (2012: 81) that the development of the colloquial French question marker *-ti* did involve rebracketing. This change is exemplified in (121) and (122) below.

(121) Votre père part-il?
Your father leaves-he
'Is your father leaving?' (Whitman 2012: 81, ex. 35a)

(122) Votre père par-ti?
Your father leaves-ti
'Is your father leaving?' (Whitman 2012: 81, ex. 35b)

As well as rebracketing, category relabelling is also said to be a typical feature of reanalysis (Detges & Waltereit 2002: 153). For Whitman (2000), relabelling, rather than rebracketing, is the crucial feature of reanalysis, though traditional accounts of the *for...to* infinitive challenge this view. These accounts state that the preposition *for* was reanalysed as part of the *to*-infinitive with which it occurred, i.e. that the structure *V [for NP] [to VP]* was reanalysed as *V [[for [NP to VP]* (De Smet 2009: 1743). If the

traditional account is correct, relabelling cannot have taken place, because there is no plausible relabelling that can have led to the object of the prepositional phrase *for NP* being reanalysed as the subject of a subordinate clause (Whitman 2000: 224). However, De Smet (2009: 1744-1745) points out that the traditional account is problematic, because scholars have focused on *for...to* clauses functioning as extraposed subjects, ignoring them in other contexts. In fact, *for...to* clauses not functioning as extraposed subjects are attested earlier than instances of *for...to* infinitives functioning as extraposed subjects (ibid.). De Smet (2009: 1747-1748) instead proposes that the *for...to* infinitive arose via a series of smaller changes, some of which can be viewed as a relabelling; when *for* was used to reinforce infinitival *to*, a change in category from preposition to infinitival marker occurred.

In the following subsection, I will show that it does not make sense to view the neoanalysis of *way*/*weg* (a pre-constructionalization constructional change) and the subsequent constructionalization of the *way*-construction in each language in terms of rebracketing and relabelling. Instead, I couch my analysis in more constructional terms, and argue that the neoanalysis of *way*/*weg* has led to a decategorialisation, defined as the process whereby an item in one category loses some of the properties typically associated with other members of that category (Hopper & Traugott 2003); this is said to be a typical feature of lexical items undergoing grammaticalization (Traugott & Trousdale 2013: 116). In a framework such as Construction Grammar, where it is assumed that categories are gradient (see e.g. Denison 2001, 2006; Aarts 2004, 2007; Croft 2007), a change may consist of a change within the same category, rather than between categories; such a change is not, strictly speaking, a relabelling. Further, the notion of 'relabelling' in this case lacks explanatory power unless the label applies to the entire construction, i.e. *way*/*weg* becoming a constructional marker.

4.5.2 Results

4.5.2.1 English

As mentioned in Chapter 2, there were two precursor constructions to the *way*-construction. One of these was a transitive construction in which the noun *way* appears with a verb of acquisition, usually *niman* 'take'. In this transitive precursor construction, *way* functioned as a semantic argument of the verb. The other of these

precursor constructions was an intransitive construction, and occurred with basic motion verbs such as *go* and *wend*. Though I found no instances of the transitive precursor construction in my data, Fanego dates it to the 14th century, with the earliest attestation of this construction in her data shown in (123). The earliest attestation of the intransitive construction in my data dates to between the 13th and 14th century, and is shown in (124). In this intransitive construction, *way* cannot be functioning as a direct object of the verb, because verbs of motion do not ordinarily take direct objects.

(123) Corineus [...] harde smot [...] & made is wey bi eiþer side
Corineus hard struck and made his way in every direction
'Corineus struck hard and made his way in every direction' (Fanego 2017: 16, ex. 21a).

(124) He ete and dranc and went his wai
'He ate and drank and went his way' (Helsinki, 1250-1350).

When these two precursor constructions were subsumed into the schema [NP_i [V $POSS_i$ *way* DIR]], the noun *way* changed its status (cf. Israel 1996; Traugott & Trousdale 2013). Rather than being an argument of the verb, it is now an argument of the construction; while in (124) above, the semantics is provided by the verb, in sentences such as (125), the lexical semantics is provided by the construction; *bawling and banging his way* means 'move while bawling and banging', and this meaning is imposed by the construction. If the meaning were not imposed by the construction, we would have to posit that *bawl* and *bang* (as well as every other verb that occurs in the *way*-construction) is polysemous, with an additional sense of 'move while X'. This explanation runs into problems, however, because sentences such as (126) are ungrammatical. To reject the constructional account while accommodating this fact would therefore require that *bawl* and *bang* have an additional meaning of 'move while bawling/banging', but only when they occur with a possessive pronoun and the noun *way*. Given the very high number of verbs that occur in the *way*-construction, it is very unlikely that speakers store this additional meaning for all these verbs. A more likely explanation is to assume that the lexical semantics of the verb can be coerced by the meaning of the construction, rather than the properties of the lexical item in isolation (Goldberg 2006: 9; Traugott 2008: 223). The neoanalysis of *way* as a non-referential object has therefore led to a situation where the lexical

semantics of the verb and the semantics of the construction no longer need to correspond; this is typical of a pre-constructionalization constructional change (Traugott & Trousdale 2013: 27).

(125) I could hear him bawling and banging his way through the Nissen hut across the way (BNC, 1991).
(126) *He bawled and banged through the Nissen hut.

As well as the *way*-construction imposing a meaning of 'move while X' on the verb, a meaning of 'move by X' is also possible. In (127), the hammer moves through the hard materials *by* smashing, rather than while smashing.

(127) The new rotary hammer virtually smashes its way through hard materials (BNC, 1991).

As well as a mismatch between the semantics of the verb and the construction, there is also now a mismatch between the syntax and semantics of the construction. The semantics of the *way*-construction involves motion along a literal or metaphorical path, by means of or while performing the verb. The syntax of the *way*-construction does not correspond with this; verbs of motion are usually intransitive, but in the *way*-construction, the string *POSS* way is in direct object position. The meaning of motion cannot be attributed to any of the individual parts of the construction, either; it is the construction as a whole that imposes the meaning of motion; this suggests that grammatical constructionalization has taken place, because there is now a new form-meaning pairing in the constructional network.

While traditional accounts of reanalysis usually focus on notions such as rebracketing and relabelling, these notions do not account for the development described above. There has been no change of constituency; *He took his way home* and *He whistled his way home* have the same constituent structure, even though *way* is an argument of the verb in the former sentence and a constructional argument in the latter. There has also not been a change of category label; the noun *way* has undergone decategorialisation (cf. Hopper & Traugott 2003), rather than changed label, because although it has lost many of the properties associated with nouns, such as the ability to undergo pluralisation, it retains some nominal properties in that it can be selected

by a possessive pronoun, a property typically associated with nouns. The decategorialisation of *way* suggests that *way* in the construction has undergone grammaticalization.

4.5.2.2 Dutch

The earliest tokens of the *way*-construction found in *De Gids* almost exclusively contain the verb *banen*, where the noun *weg* functions as a referential object of the verb. Apart from this and two tokens of the force-dynamic verb *openen* ‘open’, no other verb types are attested between 1840 and 1889. In the 1890s, a cluster of verbs of fighting and path clearing emerges: this cluster consists of *graven* ‘dig’, *kappen* ‘cut’, and *houwen* ‘cut, hack’. A token of each of these verbs is shown in (128) to (130).

(128) [R]itten, door welke de rivier zich een weg heeft gegraven
Rides through which the river REFL a way has dug
'Rides, through which the river has dug its way' (De Gids, 1890)

(129) Vindt je het niet prachtig dwars door een
Find you it not wonderful across through a
ondoordringbaar bosch je een weg te kappen?
impenetrable forest REFL a way to cut
'Don't you find it wonderful to cut your way through an impenetrable forest?' (*De Gids*, 1890)

(130) een moed om aan te pakken en
a courage to PARTICLE to tackle and
zich een weg uit te houwen
REFL a way out to hack
'A courage to tackle [the problem] and hack one's way out' (De Gids, 1891)

In these sentences, *weg* shows differing degrees of referentiality. In (128), it is easy to imagine a literal path being dug out by the river. However, in (129) and (130), it is not a path that is being hacked or cut, strictly speaking; rather, a path has been cleared as a result of hacking or cutting an obstacle.

In the SoNaR corpus, more verbs of motion begin to be attested. These include *wurmen* ‘worm’, *slalomen* ‘slalom’, and *fietsen* ‘cycle’. When the Dutch *way*-construction occurs with verbs of motion, the semantics of the construction is provided by the verb; *slalomen* means ‘move by slaloming’ regardless of whether it occurs in the *way*-construction. However, evidence begins to emerge in the SoNaR corpus that the lexical semantics of the verb and the semantics of the construction no longer need to correspond, as (131) and (132) show; this suggests that grammatical constructionalization has taken place.

(131) We klaxonneren ons een weg door kuddes scharminkelschapen
We claxon REFL a way through herds scrag.sheep
'We claxon our way through herds of scraggy sheep' (SoNaR, date unknown).

(132) Het trio hoestte zich een weg naar het miljoen
The trio coughed REFL a way to the million
'The trio coughed their way to the million [pounds]' (SoNaR, date unknown)

The two sentences above can be paraphrased as ‘move through the herds of sheep while claxoning’ and ‘get the million [pounds] by coughing’. The sense of movement and attainment is incompatible with the lexical semantics of the verb in isolation; rather, the lexical semantics is provided by the construction. As is the case of the English *way*-construction, the notions of rebracketing and relabelling do not provide an adequate account of the status of *weg* as a non-referential object. *Zich een weg banen* ‘make one’s way’ and *zich een weg claxonneren* ‘claxon one’s way’ have identical constituency structures, and therefore rebracketing has not taken place here. As in the English construction, the noun *weg* has undergone decategorialisation, rather than a relabelling; in tokens of *weg* with an adjectival modifier, the adjective regularly modifies the entire verbal subevent rather than the path. Given this fact, and the fact that intransitive verbs can now appear with *weg* in direct object position, the noun *weg* can be said to have changed status from a strictly lexical, referential noun,

to a highly bleached grammatical marker of the construction, though it retains some nominal properties in that it can be modified by an indefinite article.

4.5.2.3 German

The vast majority of the tokens found in the *Deutsches Textarchiv* corpus are of precursors to the *way*-construction. These typically contain transitive verbs of acquisition, path creation, and the creation or removal of obstacles on a path. These verbs include *bahnen* ‘pave’, *legen* ‘lay’, *öffnen* ‘open’, *bereiten* ‘prepare’, *preparieren* ‘prepare’, *nehmen* ‘take’, *schlichten* ‘smooth’, and *versperren* ‘block’. When these verbs are used in the transitive precursor construction, *Weg* ‘way’ is unambiguously a semantic argument of the verb. In (133), the relative pronoun refers back to the noun *Weg*; this is not possible when *Weg* is used as a non-referential object, as the ungrammaticality of (134) shows. In these sentences, the lexical semantics of the construction is provided by the verb.

(133) Und seh genau [...] den weg den ich in acht [...] nahm
And see exactly the way that I in care took
'And I see exactly the way I carefully took' (*Deutsches Textarchiv*, 1668).

(134) *Der Weg, den ich freigekämpft habe
The way that I free.fought have
'*The way that I fought free'

There are also four tokens in the *Deutsches Textarchiv* corpus of a precursor construction with an intransitive verb of motion. These verbs are *reisen* ‘travel’, *gehen* ‘go’, *laufen* ‘walk/run’, and *wandeln* ‘stroll’. In these cases, the lexical semantics of the construction are still provided by the verb, and therefore there is no mismatch between the semantics of the verb and the semantics of the construction. This suggests that, unlike the English and Dutch constructions during the same period, a pre-constructionalization constructional change had not yet taken place in the German construction at this time.

In the more modern data from the *Berliner Zeitung* and DECOW corpora, tokens are attested where *Weg* functions as a non-referential object. Rather than acting on a literal path, the verb denotes an action performed on obstacles that are blocking this

path. There is therefore a mismatch between the syntax and the semantics of the construction; the meaning of (135) is roughly ‘make oneself free by bombing’, but the syntax has two direct objects: a reflexive pronoun, and the NP *den Weg* ‘the way’. This development is typical of grammatical constructionalization.

(135) [D]ie Terroristen glauben, sich den Weg
the terrorists believe REFL the way
in die Friedengespräche freibomben zu können
into the peace.talks free.bomb to be able
'The terrorists believe they can bomb their way into peace talks' (*Berliner Zeitung*, 1996).

Other verbs attested in the construction which encode the removal of obstacles by force include *freikämpfen* ‘fight free’, *freischießen* ‘shoot free’, and *sprengen* ‘blow up’. In the former three verbs, the directional *free* that would usually be expressed as a post-verbal argument in the English construction is realised as a verbal particle *frei* in German.

In the DECOW corpus, there begins to emerge a mismatch between the lexical semantics of the verb and the semantics of the construction, but far fewer tokens of the German *way*-construction show this mismatch compared to the English and Dutch constructions. In (136), the meaning is roughly ‘get into the Guinness book of records by munching’; this meaning of attainment cannot be attributed to the verb in isolation, which suggests that, as is the case in English and Dutch, the construction must be imposing this meaning and that the German *way*-construction had therefore been constructionalized by this point.

(136) Diese Kuh mampft sich den Weg
This cow munches REFL the way
ins Guinness Buch der Rekorde
into.the Guinness book of.the records

'This cow munches its way into the Guinness book of records' (DECOW, date unknown)

Unlike the English and Dutch constructions, there is no mismatch between the syntax of the verb and the syntax of the construction, because all of the verbs that occur in the construction are transitive. In sum, data from the German corpora suggests that the actualization of *Weg* as a non-referential object is at a different stage than in the English and Dutch constructions. Intransitive verbs are far less frequently attested, and the noun *Weg* is used referentially in most cases. Like the English and Dutch constructions, however, this change cannot be thought of as a rebracketing, because no change of constituency has taken place. Neither is this a relabelling, because *Weg* functions as a fully referential noun in most tokens of the German *way*-construction. Unlike the English and Dutch constructions, intransitive verbs are not used, which means that decategorialisation has not taken place here.

In sum, the noun *way* and its Dutch equivalent have certainly undergone a change of status in the *way*-construction; while early tokens of the *way*-construction in both languages contained referential uses of *way*/*weg*, this noun has lost many nominal properties such as ability to undergo pluralisation. I proposed that this change can be thought of as decategorialisation, which is typical of lexical items undergoing grammaticalization (Traugott & Trousdale 2013: 116). The noun *way*/*weg* retains some nominal properties such as the ability to be selected by a possessive pronoun (in English) and an indefinite article (in Dutch). This change has affected the German *way*-construction to a far lesser extent; tokens of *Weg* functioning as a non-referential object are far less frequent than in the English and Dutch constructions. This suggests that the noun *Weg* in the German construction has undergone decategorialisation to a lesser degree.

The following section is concerned with the constructional changes the *way*-construction underwent after its constructionalization in each language. Traugott and Trousdale (2013: 27) argue that post-constructionalization constructional changes usually involve host-class expansion (cf. Himmelmann 2004); I show that this is the case in all three languages, because the number of verbs that are now permissible in the construction is considerably higher than the number of verbs attested in the construction shortly after its constructionalization. I also show that the schema of the construction itself has expanded in English and Dutch, because an adjectival modifier of *way*/*weg* is now attested (e.g. *he whistled his merry way home*).

4.6 Post-constructionalization constructional changes in the *way*-construction in English, Dutch, and German

4.6.1 English

Table 17 shows the change in type frequency of the *way*-construction in the CLMET3.1 corpus.

Table 17: Type frequency of the way*-construction in CLMET3.1*

Period	Types (cumulative)
1720-1740	5
1740-1760	17
1760-1780	25
1780-1800	29
1800-1820	31
1820-1840	48
1840-1860	74
1860-1880	87
1880-1900	99
1900-1920	112

This shows that the English *way*-construction has undergone considerable host-class expansion since 1720; the number of types has increased from 5 to 112. This frequency trend is statistically significant; a correlation analysis yielded a coefficient of 0.978, which was found to be statistically significant at $p < 0.05$. The following table shows the type frequency of the *way*-construction at the end of each decade in COHA.

Table 18: Type frequency of the way-construction in COHA from 1920-2010

Decade	Type freq.
1920s	176
1930s	253
1940s	315
1950s	386
1960s	457
1970s	523
1980s	590
1990s	664
2000s	755

The findings above suggest that the host-class expansion of the *way*-construction is continuing up to the present day; the number of unique verbs in the construction has increased almost fivefold since 1920. The correlation coefficient of the frequency with respect to time is 0.999 and was found to be statistically significant at $p < 0.01$. Combined with the findings from CLMET3.1, this suggests that the *way*-construction has undergone considerable host-class expansion up to the present day, which is a typical post-constructionalization constructional change.

I turn now to the adjectival modifier of *way*. Since the constructionalization of the *way*-construction, the schema of the construction itself has expanded to allow an intervening adjective between the possessive pronoun and the noun *way*, resulting in sentences such as *He whistled his merry way home*. In the following paragraphs, I examine the change in type and token frequency of the adjectival modifier slot in CLMET3.1 and in COHA from 1920-2010. The change in type and token frequency of the adjective slot in CLMET3.1 is shown in the table below.

Table 19: Token and type frequency of the adjective slot in CLMET3.1

Period	Tokens	Tokens (cumulative)	Types
1720-1740	0	0	0
1740-1760	4	4	4
1760-1780	3	7	7
1780-1800	3	10	9
1800-1820	1	11	10
1820-1840	14	25	23
1840-1860	17	42	34
1860-1880	2	44	36
1880-1900	17	61	48
1900-1920	7	68	50
Total	**68**	**68**	**50**

This suggests that there is no clear token frequency trend; the token frequency of the adjectival modifiers is much higher between 1820 and 1860 than in any other period, with the exception of 1880-1900. This in turn suggests that the adjectival modifier of *way* may be an idiosyncrasy of a particular author or authors who were writing in those periods. The type frequency undergoes a steady increase between 1720 and 1820, with a sharp increase between 1820 and 1860. There is another sharp increase between 1880 and 1900. A correlation analysis of the change in frequency over time yielded a correlation coefficient of 0.562, a moderately strong correlation, but this

coefficient was not found to be significant at the level of $p < 0.05$. The following table shows the change in type and token frequency of the *way*-construction in the second half of COHA.

Table 20: Token and type frequency of the adjective slot in COHA from 1920-2010

Period	Tokens	Tokens (cumulative)	Types
1920-1930	31	31	26
1930-1940	24	55	44
1940-1950	26	81	61
1950-1960	18	99	74
1960-1970	9	108	81
1970-1980	21	129	94
1980-1990	14	143	102
1990-2000	15	158	110
2000-2010	14	172	118

The data from COHA reveals a general trend of decrease, while the type frequency increases steadily over time. Subjecting the change in frequency to a correlation analysis yielded a correlation coefficient of -0.759, a strong negative correlation. This correlation coefficient was found to be statistically significant at $p < 0.05$. In sum, an analysis of the change in type and token frequency of the adjectival modifier slot in the construction shows that two post-constructionalization constructional changes have taken place: firstly, the schema of the *way*-construction expanded to allow an adjective to occur between the possessive pronoun and the noun *way*. Secondly, the token frequency of that adjectival modifier decreased between 1920 and the present day. The following subsection deals with the post-constructionalization constructional changes of the Dutch *way*-construction.

4.6.2 Dutch

Because information on the year of publication of the text is not available in the SoNaR and NLCOW corpora, a detailed analysis of the change over time in token and type frequency of the verbs and adjectival modifiers is not possible. However, given the NLCOW corpus contains data scraped from the web in 2014 and the SoNaR corpus ends in 2011, it is likely that much of the data in NLCOW was published after the texts in SoNaR. Therefore, I examine the change in frequency in the order *De Gids* > SoNaR > NLCOW. To ensure an appropriate comparison of token frequencies, I

normalise them by calculating the token frequency per million words. A further problem is that in the corpus for which dates of publication are available, *De Gids*, the type and token frequency of the *way*-construction is very low; there are only 131 tokens (1.68 pmw), almost all of which contain the verb *banen*; besides 123 tokens of *banen*, there are 7 other verb types. An analysis of the SoNaR corpus shows that there are 807 tokens of the *way*-construction (1.614 pmw) and 108 types. In NLCOW, these figures are 8,680 tokens (1.258 pmw) and 423 types. These findings suggest that the *way*-construction is actually decreasing in frequency over time, which is the opposite to the post-constructionalization constructional change that is expected. However, these findings are extremely tentative given there are only three datapoints and the information on date of publication is mostly missing. The type frequency has increased over time, which suggests that the construction has become more productive; an increase in productivity is also an instance of a post-constructionalization constructional change (Traugott & Trousdale 2013: 27).

I turn now to the change in type and token frequency of the adjectival modifier slot. Rather than expressing the token frequency as a per million words figure, I instead examine what percentage of tokens contained an adjectival modifier of *weg*. In *De Gids*, 11 tokens (8.40%) contained an adjectival modifier, and there were 4 types, yielding a type/token ratio of 0.36. In SoNaR, the relative frequency of adjectival modifiers decreases; only 23 of 807 (2.85%) contained an adjectival modifier. There were, however, 19 types, giving a type/token ratio of 0.826, which suggests that the adjectival modifier slot has increased in productivity. In NLCOW, 282 tokens contained an adjectival modifier of *weg* (3.25%). These 282 tokens contained 62 types, which is a type/token ratio of 0.220. These findings suggest that after the neoanalysis of *weg* as a meaningless syntactic marker of the construction, a post-constructionalization constructional change took place, which consisted of the schema of the construction expanding to allow an adjectival modifier of *weg*. The token frequency of this adjectival modifier has decreased in comparison with the earlier data from *De Gids*, which suggests the opposite trend predicted in cases of post-constructionalization constructional changes; it is posited that the token frequency increases, rather than decreases. However, these findings are tentative given the sparsity of the information concerning the date of publication of each text. The following subsection deals with the post-constructionalization constructional changes of the German *way*-construction.

4.6.3 German

As is the case of the majority of the Dutch data, information on the year of publication of each text is missing from the DECOW corpus. In addition, most of the tokens found in *Deutsches Textarchiv* are of precursor constructions, rather than the *way*-construction itself. Given I am examining post-constructionalization constructional changes, it does not make sense to consider tokens of these precursor constructions. The following table shows the change in type and token frequency of the verbs in the *way*-construction over time in the *Berliner Zeitung* corpus.

Table 21: Token and type frequency of the way*-construction in Berliner Zeitung*

Year	Tokens	Tokens (cumul.)	Types (cumul.)
1994	31	31	13
1995	30	61	22
1996	35	96	26
1997	23	119	31
1998	24	143	35
1999	21	164	39
2000	20	184	41
2001	25	209	42
2002	22	231	44
2003	27	258	44
2004	22	280	46
2005	19	299	48
Total	**299**	**299**	**48**

The table above suggests that the token frequency of the construction is decreasing rather than increasing. The correlation between frequency and time is strongly negative (-0.675), and this coefficient was found to be statistically significant at $p < 0.05$. The increase in type frequency also appears to have slowed down over time. The total token frequency of the construction in *Berliner Zeitung* is 1.28 pmw, and type/token ratio is 0.161. In DECOW, there are 865 tokens (0.874 pmw). Given that DECOW contains data scraped from the web in 2016, and that the *Berliner Zeitung* corpus ends in 2005, it is highly likely the the majority of the data in DECOW was published after the data in *Berliner Zeitung*. This suggests that the token frequency of the construction is decreasing, though the reliability of these findings is seriously challenged by the lack of datapoints. The 865 tokens of the *way*-construction in DECOW contain 104 types, yielding a type/token ratio of 0.120. This suggests that

the construction is becoming less productive, although the reliability of this result is again challenged by insufficient data. Unlike the English and Dutch constructions, adjectival modification of *Weg* is not attested in the German construction.

4.7 Conclusions

This chapter of the thesis has provided evidence for the constructionalization and post-constructionalization constructional changes affecting the *way*-construction in English, Dutch, and German. A justification for distinguishing between constructionalization and constructional changes was presented in Section 4.2. Some scholars (e.g. Hilpert 2013) who do not make this distinction and therefore deny the existence of constructionalization are unwittingly saying that no new constructions are ever attested, which is an incorrect conclusion. Section 4.3 assessed the role of ambiguity in mechanisms of constructionalization and constructional changes. It was found that ambiguity was a plausible explanation for the neoanalysis of *way* in English, but not in Dutch or German. Section 4.4 provided evidence for the constructionalization of the *way*-construction in the three languages. It was shown that a $form_{new}$-$meaning_{new}$ pairing has emerged in each language, because the modern *way*-construction in each language differs from the precursor constructions in many key respects. Firstly, a post-verbal directional argument is now virtually obligatory in each language. Secondly, the item occurring before *way*/*weg* is now fixed in English and Dutch, but not German. This suggests that the constructionalization of the construction is at different stages in each language. In Section 4.5, I showed that rethinking concepts such as rebracketing and relabelling provides a better explanation of the constructionalization and constructional changes of the *way*-construction in each language. I propose that instead of relabelling, the noun *way*/*weg* has undergone decategorialization; the noun has not changed label, but has lost some of the properties typically associated with nouns, such as the ability to undergo pluralization. However, some nominal properties have been retained; *way*/*weg* can be selected by a possessive pronoun or article, and can undergo adjectival modification. Section 4.6 was concerned with the post-constructionalization constructional changes affecting the *way*-construction in the three languages. I showed that host-class expansion has affected all three languages, because the range of verbs that can occur in the construction has increased substantially. The schemas of the English and Dutch construction have themselves expanded, now

allowing an adjective to modify *way*/*weg*. The following chapter is concerned with the role of analogy in the development of the *way*-construction in English, Dutch, and German. In this chapter, I explore the hypothesis that the synchrony of the construction in the three languages can be explained in terms of a long series of local analogical extensions. I also test a number of hypotheses concerning the relationship between analogy and type and token frequency, as well as the hypothesis that there is constructional contamination between the *way*-construction and FRR.

This page intentionally left blank

5. The role of analogy in the development of the *way*-construction in English, Dutch, and German

5.1 Introduction

The present chapter answers the fourth of my research questions: what was the role of analogical thinking and analogization in the development of the *way*-construction in the three languages? This chapter also answers a number of sub-questions: i) Can the development of the English *way*-construction be considered a long series of local analogical extensions, as Israel (1996: 217 *et passim*) claims? ii) Can the development of the Dutch and German *way*-constructions be thought of in a similar way? iii) How can the notions of constructional contamination and bolstering be incorporated into a usage-based theory of analogy? iv) In light of the data on the *way*-construction in English, Dutch, and German, what are the roles played by analogy and reanalysis in language change? Which mechanism, if any, plays a primary role? v) What is the relation between analogy, token frequency, and type frequency in the development of the *way*-construction in the three languages?

Some terminological clarifications are necessary at this point, because the term 'analogy' has been defined in a variety of ways in the literature (Lightfoot 1979: 359), including as rule regularisation (Meillet 1912), "the attraction of new forms to already existing constructions" (Hopper & Traugott 2003: 63-64), "the generalisation of a rule or construction" (Hopper & Traugott 2003: 66), and the process whereby one item becomes more similar to another due to an abstract generalisation made by the speaker (Blevins & Blevins 2009: 3). The term 'analogy' has also been used to describe a wide range of processes, including phonological restructuring of stems and affixes, reanalysis, contamination, folk etymology, and back-formation (McMahon 1994: 74; Lahiri 2000: 10). Metaphorical change has also been understood as analogy (Fischer 2007: 121). To avoid this terminological inconsistency, I follow Traugott and Trousdale (2013) in distinguishing between analogization (the mechanism of

language change itself) and analogical thinking (the motivation for this change), because while analogical thinking may precede much change, not all instances of analogical thinking will lead to analogization (Givón 1991: 258; Fischer 2007; Traugott & Trousdale 2010: 7; Traugott 2011: 25).

The remainder of the chapter is structured as follows. Section 5.2 discusses analogical extension. I show in this section that, while the term has usually been used to refer to a linguistic rule being applied to a new item, it can also be used to describe the process whereby the range of items that can appear in the slot of a construction expand as a result of analogization. In this section, I test Israel's (1996) hypothesis that the development of the English *way*-construction can be considered a long series of analogical extensions, and test this hypothesis for the Dutch and German constructions. I demonstrate that, as Israel (1996) claims of the English construction, the development of the construction in the three languages can be attributed to a long chain of analogical extensions. Section 5.3 explores the relationship between analogy, type frequency and token frequency by testing three hypotheses put forward by Israel (1996), Barðdal (2008), and Zeschel (2010). Israel (1996) proposes that the likelihood of an item undergoing semantic extension is dependent on the number of tokens previously experienced by the user. By "previously experienced by the user", Israel intends "previously attested in a corpus"; as I will demonstrate in a later section, the relationship between what is attested in a corpus and what is part of a user's constructional network is not as simple as Israel assumes. Barðdal (2008) argues that frequently occurring items may serve as a basis for analogical extension and that analogical extension is dependent on low type frequency, whereas Zeschel (2010) claims that the type frequency of a construction needs to be relativized to the number of distinct semantic categories within the construction; it is this variable, in his view, that correlates with the likelihood of an item undergoing analogical extension. I show that the measures proposed by Israel, Barðdal and Zeschel account for the type and token frequency data of the the *way*-construction in the three languages. Section 5.4 deals with the notion of support constructions (Abbot-Smith & Behrens 2006), constructional contamination (Pijpops & van de Velde 2016), and bolstering (McColm & Trousdale 2019), and whether these are compatible with a usage-based theory of analogy. This section also tests the hypotheses that there is constructional contamination between the fake reflexive resultative (FRR) construction and the *way*-construction, and that the *way*-construction is encroaching on the functional territory

of the *way*-construction, as Mondorf (2011) claims; I show that Mondorf's claim is incorrect, because there are important differences between the FRR and the English *way*-construction. Section 5.5 examines the relationship between analogy and reanalysis in light of my data on the *way*-construction in the three languages, and discusses whether either mechanism is primary, or whether reanalysis and analogy work in tandem, as proposed by Fischer (2007: 123). I show that neoanalysis and analogy have worked in tandem in the development of the English, Dutch, and German *way*-constructions. Section 5.6 concludes the chapter.

5.2 Analogical extension

5.2.1 Literature Review

In Dörner's (1977: 75-76) view, analogy consists of three stages. In the first stage, the speaker matches two items that are considered similar in some way (Fischer 2007; Fischer 2011: 36; De Smet 2013: 82; Traugott & Trousdale 2013: 35; see also Wilson 2014: 17). This process affects all meaningful units, from single morphemes and lexical items to more abstract patterns (Fischer 2011: 39). Because a construction is a form-meaning pairing that contains phonological, morphological, syntactic, and pragmatic information, multiple sources of similarity exist within a construction (Blevins & Blevins 2009: 7). In the second stage, the speaker abstracts a more general pattern from these items (cf. Langacker 2000: 60; De Smet 2013: 83). The third stage involves extending this pattern to a new item in a process known as analogical extension (cf. Sommerer 2015: 117). It is this third stage that will be the focus of this subsection; for instance, when speakers first used *push* in the *way*-construction on the basis of existing tokens of *force*, they did so on the basis of analogical extension, because they abstracted the more general pattern that *push* and *force* both describe the exertion of force on an entity.

Analogical extension was first documented by Whitney (1867), who described children's overgeneralisation of plural *-s* and past tense *-ed* as "extension of prevailing analogies beyond their historically correct limits" (as quoted in Lahiri 2000: 5; cf. also Traugott 2011: 25). Meillet (1912) defines this as rule regularisation. Analogical extension leads to a change in the surface manifestation of a construction, but not its underlying structure (Harris & Campbell 1995: 51; see also Meillet 1958: 133). This is

true of the *way*-construction, because the structure of the construction does not change when a novel verb is used in it on the basis of analogical extension.

Analogical extension is sometimes known in the literature as 'four-part analogy' or 'proportional analogy' (see e.g. Blevins & Blevins 2009), i.e. A is to B as C is to D. This type of analogy allows the relation between a pair of items A and B to identify an unknown item D given an item C that is similar to either A or B (Blevins & Blevins 2009: 2). For instance, in the four-part analogy 'wings : fins = lungs : *x*', we can identify item *x* as 'gills', because the relation between wings and fins is similar to the relation between lungs and gills; the pairs of items in question serve similar functions (Itkonen 2005: 1). Applying this type of analogy to language, speakers can apply a morphological generalisation to an item that previously was not subject to this generalisation. In the four-part analogy 'dog : dogs = cow : *x*', speakers identified *x* as *cows*, replacing the earlier plural form *kine* (Hock 2003: 441).

Although most definitions of analogical extension involve applying a rule to a new item, or a new form replacing an older form, analogical extensions can also take place within constructions when the range of items that can fill a slot of that construction expands; this is known as host-class expansion (Himmelmann 2004). De Smet and Fischer (2017: 242) observe that the Dutch reflexive construction is undergoing a long chain of local analogical extensions in psych verbs. Verbs canonically associated with the reflexive construction such as *zich herinneren* 'to remember', *zich realiseren* 'to realise', and *zich ergeren* 'to be annoyed' have been joined by *beseffen* 'realise' and *irriteren* 'to be irritated', which older speakers do not use in the reflexive construction; younger speakers have recruited these verbs to the reflexive schema by analogization with other psych verbs. Delbecque and Verveckken (2014) show that the development of the Spanish binominal quantifier construction, e.g. *una pila de años*, lit. 'a pile of years' i.e. 'many years', can also be thought of as a long chain of local analogical extensions. Israel (1996: 217) claims that the development of the English *way*-construction can also be thought of in this way; the following subsection tests this hypothesis, and the hypothesis that the development of the Dutch and German *way*-constructions can also be considered a long chain of analogical extensions. I show that the English, Dutch, and German *way*-constructions can plausibly be considered a long series of analogical extensions.

The discussion in the following subsections (and in 6.2) relies on grouping the items that occur in the verb slot of the *way*-construction into semantic categories. The divisions employed here mostly follow Perek's (2018). If I felt that a verb could not be placed into one of Perek's categories, new categories were created. Occasionally, a verb could plausibly be placed into more than one category; for instance, *sing* is both a verb of performance, but it is also an action involving the mouth. Such examples were inspected on a case-by-case basis. For example, in *He sang his way into the final of the talent show*, it is clearly the performance that is salient, rather than the fact this action involved the subject's mouth, so this token would be placed in the 'performance' category. This classification system differs from Perek's (2018) approach in that it is strictly a qualitative analysis, which is the approach Hilpert (2012b) employs in his study on the *many a* NOUN construction, although nouns are easier to classify than verbs. Perek (2018) had a random sample of his tokens annotated by a second annotator and calculated an inter-rater agreement score; the present study would have benefited from this statistical rigour.

5.2.2 Analogical extension in the *way*-construction

5.2.2.1 Analogical extension in the English *way*-construction

By 1658, verbs encoding force enter the way-construction, as in (137). These are probably an analogical extension of *take*. This is a plausible extension, because *take* can sometimes connote taking something by force; it is likely, therefore, that speakers would analogically extend this verb to other verbs encoding the exertion of force on an entity.

(137) [They] fought their way in, and beat them all great and small (ARCHER, 1658).

A cluster of basic motion verbs emerges by 1708, as (138) shows. Tokens such as (138) probably have as their analogical source a precursor construction with motion verbs such as *go* but without a directional, found in tokens such as (139); it is highly likely that *go* would be extended to other basic verbs of motion, because they are very closely related in their semantics.

(138) However, he made his way through (ARCHER, 1708).

(139) Ignoring her thanks, he went his way (Traugott & Trousdale 2013: 88, ex. 48; cited from 2006 Stroud, *The Golom's Way* [COCA]).

By 1726, verbs of merely finding a path join the *way*-construction, as (140) shows. This is probably an extension from the category of force dynamic verbs; before a person can exert force on an object, the person must first find this object. It is logical, therefore, that these verbs appear in the construction.

(140) An upright Man, let him be Jew, Turk, or Christian, may find his Way to Heaven (CLMET3.1, 1726).

Verbs of path creation are first attested in 1744, as (141) shows. A possible source of these verbs is the cluster of verbs relating to finding a path. This is a plausible explanation; if a speaker cannot find a path, they must create one by some means. After *weather* is first attested in the construction in 1744, more abstract ways of creating a path emerge.

(141) We had ventured [...] in hopes to weather our way into the Great South Seas (CLMET3.1, 1744).

Verbs describing the manner of motion, as in (142), also emerge by 1744. These verbs may have had as their analogical source the verbs of basic motion. This is a highly likely extension; these clusters of verbs are closely related in that both describe motion, but rather than describing basic motion, speakers used analogical thinking to extend the collocational range of the construction to include other types of motion.

(142) From the vast flocks of birds they had repeatedly seen, which always winged their way to the south (CLMET3.1, 1744).

The manner of motion verbs undergo an analogical extension to verbs encoding the path shape, as in (143). Again, this is a plausible extension; the speaker perceives a semantic similarity between the manner in which the subject moves and the manner in which the path moves, and creates innovative tokens such as (143). While *wend* in (143) may be a result of constructional contamination with *wind* and *went*, verbs which

unambiguously encode the path shape are found as early as 1753; an example is given in (144).

(143) Take a stroll in the Park and [...] wend our way towards this Pool (ARCHER, 1752).
(144) The scout perceived her coming out, and bent her way towards the church in good earnest (CLMET3.1, 1753).

By 1761, verbs encoding actions involving the mouth emerge, possibly by analogical extension of more general force dynamic verbs. The dating of this cluster is considerably earlier than Perek's (2018), who dates the emergence of this cluster to ca. 1880, but he did not use the CLMET3.1 corpus in his study. This extension is made plausible by the strong semantic link between applying force to an entity and applying force to an entity using the mouth. (145) shows the first attestation of a verb encoding an action involving the mouth.

(145) The maggot [...] will eat its way through (CLMET3.1, 1761).

Verbs relating to commerce and finance and performance are attested by 1824, as shown in (146) and (147); unlike the clusters of verbs described so far, these have no obvious analogical source in my dataset. Perek (2018) dates the cluster of commerce and finance verbs to ca. 1880, though given he excluded data from the 1810s and 1820s from his study, this is not surprising.

(146) There would not be as much left for Mr. Birkbeck as would be required to pay his way back again to the Land of Promise (CLMET3.1, 1824).
(147) Oh, these are the hymns and praises o' the auld wives and creeshy louns o' Auchtermuchty, wha are gaun crooning their way to Heaven (CLMET3.1, 1824).

By 1826, verbs encoding laborious motion emerge, as in (148). There is a very strong semantic similarity between these verbs and verbs describing manner of motion; an analogical extension from describing the manner in which the subject moves to describing this subject as moving laboriously is therefore very likely.

(148) Others were already toiling their way up the neighboring hills (COHA, 1826).

Verbs of vehicular motion also surface by 1826, as in (149). The source of analogization for verbs of vehicular motion is probably the intransitive precursor construction; Israel (1996: 222) points out that *ride* could be used in this precursor construction. Another possible analogical source is the verbs of manner of motion, as there is a strong semantic link between travelling by vehicle and other types of motion.

(149) Vivian steered his way through the old narrow winding streets (CLMET3.1, 1826).

By 1834, verbs of path shape undergo an analogical extension to verbs encoding the terrain of motion, as shown in (150). This extension can be explained by a semantic similarity between the manner in which the path moves, to the terrain through which the speaker moves on this path. It is possible that verbs of basic motion are the analogical source of tokens such as (150), but verbs of path shape are more closely semantically related, and so the analogical extension described in this paragraph seems a more likely development.

(150) Through this awful scene did the Athenian wade his way (CLMET3.1, 1834).

Possibly by analogical extension of *croon*, verbs of sound emission enter the construction by 1842, as shown in (151) below. While there is a strong semantic link between *croon* and other verbs of sound emission in that both involve sound, in (151) the sound is not emitted intentionally, but as a result of the motion; this corresponds to Fanego's (2018) sound-emission schema.

(151) It came at last in sight, plashed its way forward (CLMET3.1, 1842).

By 1852, verbs of clearing a path are attested in the construction; the first attestation of such a verb is shown in (152). These are possibly an analogical extension of the force dynamic verbs; the two clusters of verbs are very closely linked, because both involve exerting force on an obstacle to clear a path.

(152) The boughs toss like foam raking its way down a pebbly shore (CLMET3.1, 1852).

Verbs encoding olfaction are attested from 1884. In (153), smelling is the means by which Don Gortez is expected to leave the house; this suggests that *smell* may have arisen by analogy with the other verbs of path creation.

(153) Gor. Don Gortez I wish you'd smell your way out of my house (COHA, 1884).

Verbs relating to cognition are attested from 1918, as (154) shows; these verbs have no obvious analogical source. Verbs encoding incompetence emerge by 1968, as in (155). *Fumble* in (155) is probably an analogical extension of the manner of motion verbs, as they are closely linked semantically; both involve motion along a path, but in (155) the motion is incompetent or clumsy.

(154) A man who had worried his way to the top of his profession (CLMET3.1, 1918).
(155) In the end usually fumbling her way downstairs to get herself a drink (BNC, 1968).

Verbs of obtainment found in tokens such as (156) are first attested in 1972, and can also be said to be an analogical extension of verbs of taking a path, because *take* outside of the *way*-construction can encode obtainment; there is therefore a link between *take* and other verbs of obtainment which are used in the *way*-construction.

(156) He was against the king himself in Westminster, though he seldom won his way there (BNC, 1972).

By 1985, verbs relating to sex begin to be attested in the way-construction, as shown in (157). These have no obvious analogical source, although there is a somewhat weak link between these verbs and verbs encoding force; in (157), force is being applied to an entity, but a force of a very specific type.

(157) I’m fucking my way through college (BNC, 1985).

Possibly by analogy with verbs of performance such as *croon*, verbs of sporting performance are attested from 1973, as (158) shows. These verbs are semantically linked, because both involve some kind of performance, which makes this hypothesised analogical extension a plausible one.

(158) Monzon, Dezotti and others all desperately trying to play their way into my World Cup team (BNC, 1989).

The analogical source of tokens such as (159) is unclear. From these tokens, more general verbs of social interaction emerge by analogical extension, as in (160).

(159) I was obliged to beg my way up town in the habit of a sailor (CLMET3.1, 1751).
(160) The minister of Dunlop joked his way around the world at various functions (BNC, 1989).

An unaccusative verb, *die*, is first attested in the way-construction in 1998. This is possibly by analogization with *melt*, which is attested in 1991, although this is the only token of an unaccusative verb before 1998. A token of each verb is shown in (161) and (162).

(161) The stuff just melts its way down (BNC, 1991).
(162) Sondheim is sort of dying his way into the canon, and it's not pretty (COHA, 1998).

In sum, although not all of the clusters of verbs can be explained by analogical extension, the English data strongly suggest that Israel's hypothesis is correct; the English *way*-construction can certainly be said to have arisen from a long series of local analogical extensions. These extensions involved speakers making categories more specific by semantic narrowing, such as verbs of basic motion being extended to specific types of motion, including vehicular motion, or making a category more general, such as when verbs of acquisition were generalised to other verbs encoding force. The analogical extensions I hypothesise took place in the English *way*-construction are summarised in the diagram below.

Figure 18: Hypothesised analogical extensions in the English way*-construction*

take > force > finding a path > path creation

obtainment actions involving the mouth

clearing a path

sexual verbs

go > basic motion > manner of motion > path shape > terrain of motion

laborious motion

vehicular motion

croon > sound emission

sporting performance

beg > social interaction

No analogical source: commerce/finance, cognition, unaccusatives (*melt*, *die*)

5.2.2.2 Analogical extension in the Dutch *way*-construction

Given that the year of publication of a text is missing from the SoNaR and NLCOW data, constructing a detailed timeline of the local analogical extensions of the Dutch way-construction is very difficult. We only know that the SoNaR data spans the period 1954-2011, and that the NLCOW corpus contains web-based data scraped in 2014, but the texts in this corpus could have been published earlier than this.

For the data from the *De Gids* corpus, where the year of publication is available, the verbs attested in the Dutch *way*-construction can be grouped into two semantic categories: verbs of basic motion (163) (*banen* is the only such verb attested in this corpus), and verbs encoding the exertion of force on an entity (164).

(163) [...] baant het Evangelie zich een weg door Azië
makes the Gospel REFL a way through Asia
'The gospel makes its way through Asia' (De Gids, 1860).

(164) [...] ten einde zich een eigen weg naar 't Oosten te openen
at.the end REFL a own way to the East to open
'To finally open their own way to the east' (De Gids, 1872).

As in the English construction, the verbs of basic motion can be said to undergo an analogical extension to verbs describing other manners of motion, a very likely analogical extension. This is confirmed by the fact that *banen* is the only verb attested in the construction before *kronkelen* ‘wriggle’ joins the construction; there can, therefore, be no other analogical source.

Data from the NLCOW corpus show that the verbs encoding the manner of motion lead to an analogical extension to verbs of vehicular motion (165) and laborious motion (166). This is due to the semantic similarity of these sets of verbs; all of these verbs encode motion. This analogical extension is similar to the one that took place in the English *way*-construction.

(165) Je fietst je een weg tussen de verschillende
You cycle REFL a way between the various
gangen van een uitgebreide maaltijd
courses of a extended meal
'You cycle your way through the various courses of an extended meal' (NLCOW).

(166) Ik ploeter me een weg door
I toil REFL a way through
wat het leven van Pieter was
what the life of Pieter was
'I toil my way through what Pieter's life was' (NLCOW).

It is plausible that the manner of motion verbs also undergo analogical extension to verbs encoding the terrain of motion, as in the English construction, as (167) shows. These two clusters of verbs are highly semantically similar; there is a strong link between verbs describing the manner in which the entity moves and verbs describing the manner in which the terrain of motion is forcing the entity to move. *Zwemmen* ‘swim’ describes the metaphorical motion of the subject, but also encodes the fact that the subject is moving laboriously through an obstacle.

(167) Ondertussen zwem ik me een weg
Meanwhile swim I REFL a way
door de lagen bestrating
through the layers paving
'Meanwhile I swim my way through layers of paving' (NLCOW).

After *openen* ‘open’ enters the construction, other force dynamic verbs arise by analogical extension in the 1890s; *graven* ‘dig’ and *kappen* ‘hack’ emerge in 1890, while *houwen* ‘cut’ is attested in 1891. The last force dynamic verb to emerge in the *De Gids* data is *breken*, which is first attested in the construction in 1895.

Data from the SoNaR corpus reveals a cluster of verbs relating to performance: *dansen* ‘dance’ (attested three times), *kniezwengelen*[38] ‘to perform a dance with the knee’ and *zingen* ‘sing’. It is possible that tokens such as (168) were a source of analogization to verbs of sound emission and verbs of sporting performance, as in (169) and (170). This is because all of these verbs encode some kind of performance, and there is a link between *zingen* ‘sing’ and *toeteren* ‘toot’ in that both involve the emission of sound.

(168) Zoals alleen hij kan, zingt Mick Hucknall zich een
As only he can sings M. H. REFL a
weg door een cocktail van eigen songs en covers
way through a cocktail of own songs and covers
'As only he can, Mick Hucknall sings his way through a cocktail of his own songs and covers' (SoNaR).

(169) Bussen [...] toeteren zich een weg door de massa
Buses toot REFL a way through the crowd
'Buses toot their way through the crowd' (SoNaR).

[38] This verb is difficult to translate because it concerns a dance performed by Kate Ryan in Belgium’s 2006 Eurovision entry ‘Je t’adore’. This dance involves a particular movement with the knee.

(170) In de slotfase counterden we ons
In the end.phase countered we REFL
een weg naar de volgende ronde
a way to the following round
'In the last phase we counter-attacked our way into the next round' (NLCOW).

The verbs of sound emission undergo a further analogical extension to verbs of light emission, as in (171). This is a plausible analogical extension, because sound and light emission are linked by waves travelling through the air causing the perception of a stimulus; the difference is that one stimulus is aural and the other is visual.

(171) Een vlammetje van warmte en licht
a flame.DIM of warmth and light
[f]lakkert zich een weg omhoog
flickers REFL a way upwards
'A little flame of warmth and light flickers its way upwards' (NLCOW).

As in the English *way*-construction, the Dutch *way*-construction contains verbs encoding deceit. Two such verbs are attested in the SoNaR corpus: *bluffen* 'bluff' and *manipuleren* 'manipulate'. An example token is given in (172). A further 10 such verbs are attested in the NLCOW data. It is possible that these verbs were analogically extended to verbs encoding other types of social interaction, including verbs relating to commerce and finance, as in (173), but a lack of information concerning the year of publication of each texts made it impossible to confirm this hypothesis. This is a plausible hypothesis, however, given that deceit is a particular type of social interaction, and commerce and finance involves a kind of social interaction between the buyer and seller.

(172) En zo bluf je je een
and so bluff you REFL a
weg naar de VIP parkeerplats
way to the VIP parking.space
'And that's how you bluff your way to a VIP parking space' (NLCOW).

(173) We dachten altijd dat we ons een
We thought always that we REFL a
weg uit een recessie konden kopen
way out a recession could buy
'We always thought we could buy our way out of a recession' (NLCOW).

Remarkably, there are several clusters of verbs in the SoNaR data that are identical to ones found in the English construction that cannot have arisen by analogy with other Dutch verbs. These include cognitive verbs (174), verbs relating to change of state (175), verbs of incompetence (176), and bodily functions (177). As in the English construction, there is also a cluster of sexual verbs (178), although there is a cluster of verbs that is not found in the English construction containing verbs relating to surfing the internet (179). I hypothesise that these verbs arose by a cross-linguistic analogization with the English *way*-construction; this hypothesis is explored in more detail in Chapter 7.

(174) De miserie van vandaag even vergeten en ons
The misery of today for.a.bit forget and REFL
een weg dromen naar een betere wereld
a way dream to a better world
'To forget the misery of today for a bit and dream our way to a better world' (NLCOW).

(175) Als door de koude grond de krokus
Like through the cold ground the crocus
zich een weg groeit naar het licht
REFL a way grows to the light
'Just as the crocus grows its way through the cold ground to the light' (NLCOW).

(176) [...] blunderen veel heren en dames zich
blunder many men and women REFL
een weg omhoog door het leven
a way upwards through the life
'Many men and women blunder their way upwards through life' (NLCOW).

(177) In de brandende zon zweten we ons een weg omhoog.
In the burning sun sweat we REFL a way upwards
‘We sweat our way upwards in the burning sun’ (NLCOW).

(178) Is er nog de groep vrouwen en
Is there still the group women and
mannen die zich een weg omhoog neuken
men who REFL a way upwards fuck
'There is still the group of women and men who fuck their way to the top' (NLCOW).

(179) Nu google je je een weg naar Wikipedia
Now google you REFL a way to Wikipedia
'Now google your way to Wikipedia' (NLCOW).

Overall, the Dutch data yield the same conclusions as the English data; there has been a long series of local analogical extensions which has expanded the range of verbs that occur in the construction considerably. As in the English construction, these analogical extensions involved making a category more specific, as in the case of manner of motion verbs undergoing an analogical extension to verbs of vehicular motion, or more general, as is the case of a very specific verb *openen* ‘open’ being extended to other verbs encoding force. I summarise the hypothesised timeline of analogical extensions in the Dutch *way*-construction in Figure 19.

Figure 19: Hypothesised analogical extensions in the Dutch way*-construction*

Basic motion > manner of motion > speed of motion

vehicular motion

laborious motion

terrain of motion

openen 'open' > force > sexual verbs

deceit > social interaction

performance > sound/light emission

Borrowed from English: cognitive verbs, change of state, incompetence, sexual verbs

No analogical source: surfing the internet

5.2.2.3 Analogical extension in the German *way*-construction

As in the Dutch data, there is a problem of the year of publication of the text not always being available; this problem, however, only affected the DECOW data. The first verb to be attested in the German data is a verb of path-creation, *legen* 'lay', shown in (180), in a precursor construction; the meaning of this construction is roughly 'to pave the way for somebody'. This precursor construction also existed in Dutch (Kramer 2002). As was the case of the Dutch construction, I hypothesise that the constructionalization of the German *way*-construction can be attributed to this precursor construction merging with a reflexive construction, by which point the notion of path traversal became conventionalised; it is unlikely that somebody will make a path for themselves and not traverse it.

(180) Aquila vnd Priscilla [...] legten yhm den weg Gottis
A. and P. laid him the way God.GEN
'Aquila and Priscilla paved the way of God for him' (Deutsches Textarchiv, 1522).

Verbs encoding clearing a path are first attested in 1584, again in a precursor construction, as (181) shows. An analogization from paving a path for somebody and

removing obstacles from a path for somebody is a very plausible one; the same analogical extension occurred in the Dutch precursor construction (Verhagen 2002: 423).

(181) Keiner andern vrsach denn das
No other cause than that
er jhnen den weg öffnete
he him.DAT the way opened
'No other cause than that he opened the way for him' (Deutsches Textarchiv, 1584).

Verbs of vehicular motion are first attested in the construction before verbs of basic motion. The first verb of vehicular motion, *reisen* 'travel', is first attested in 1615 and is shown in (182), while the first verb of basic motion, *gehen* 'go' is not attested until 1627 and is shown in (183); the order of this development is the opposite from that of the English and Dutch constructions. I hypothesise that verbs of vehicular motion underwent a semantic generalisation, which led to basic motion verbs being used in the construction. However, given the gap in the attestation of these two verbs is only 12 years, it is possible that verbs of basic motion were used in the construction before verbs of vehicular motion.

(182) Dann als ich meinen Weg fort reisen
Then when I my way forward travel
wollte sahe ich zwen Fuhrleut voran fahren
wanted saw I two wagoners ahead go
'Then when I wanted to travel my way onwards I saw two wagoners going ahead (Deutsches Textarchiv, 1615).

(183) Der Hencker [...] gehet er seinen weg fort
The H. goes he POSS way onwards
'Hencker goes his way onwards' (Deutsches Textarchiv, 1627).

Verbs of merely finding a path, as in (184), are attested in the construction from 1647. The analogical source of this cluster of verbs is probably the cluster of verbs relating to clearing a path in the transitive precursor construction which was shown in (180).

Before a path can be cleared, it must be found; this hypothesised analogical extension is therefore a plausible one.

(184) Oder das er den weg zur warheit dadurch funde
Or that he the way to.the truth through.it find.SUBJ
'Or that he might find the way to truth through it' (Deutsches Textarchiv, 1647).

Possibly by analogical extension of the verbs of basic motion, verbs of manner of motion emerge by 1700, as in (185), although this is a precursor construction, rather than the *way*-construction itself, because there is no oblique argument. This analogical extension also took place in the English and Dutch constructions. Given this fact, and that there is a very strong semantic link between verbs of basic motion and verbs describing other types of motion, this analogical extension is highly plausible.

(185) So lauffe ich den weg deiner gebotten
So walk I the way your.GEN commands
'I walk the way of your commands' (Deutsches Textarchiv, 1700).

Verbs encoding taking a path are attested in the construction from 1780, as (186) shows. I hypothesise that these arose by analogical extension of verbs of clearing a path. This development proceeded in a different order from that of the English construction, where *niman* 'take' was attested in the transitive precursor construction as early as Old English.

(186) Daß ihr den weg in dieses drachennest
That you the way into this dragon.nest
Um meinetwillen nicht genommen
for my.sake not taken
'That you didn't take the way into this dragon's nest for my sake' (Deutsches Textarchiv, 1780).

The hypothesised analogical source of verbs relating to commerce and finance in the English and Dutch data, i.e. verbs of social interaction, is not present in the German data. Verbs relating to commerce and finance emerge by 1994, as (187) shows. As

there is no obvious analogical source in the German, I hypothesise that these verbs arose by cross-linguistic analogization with verbs of commerce and finance in the English construction; the same goes for verbs of sound emission, which are attested in 1996 (cf. (195)).

(187) Live Modelle, die sich ihren Weg durchs College verdienen
Live models who REFL POSS way through.the College earn
'Live models who earn their way through college' (Berliner Zeitung, 1994).

(188) Doch der Mercedes 280 hupt sich den
But the M. 280 beeps REFL the
weg frei und beschleunigt auf dem Werftgelände
way free and accelerates on the shipyard
'But the Mercedes 280 beeps its way free and accelerates on the shipyard' (Berliner Zeitung, 1986).

As in the Dutch data, there is a verb related to surfing the internet, which is not present in the English data. There is only one such token, which is shown in (189) below. Cross-linguistic analogization with internet verbs in the Dutch construction is unlikely, because these are only found in the NLCOW corpus, which contains data scraped from the web in 2014, although it is possible that there were tokens in the corpus from texts published before 1998.

(189) So klickt er sich seinen Weg
So clicked he REFL POSS way
durch die verschiedenen Stallungen
through the different stables
'That's how he clicked his way through the different stables' (Berliner Zeitung, 1998).

In the DECOW corpus, which contains data scraped from the web in 2016, several clusters of verbs are attested which cannot plausibly be explained by analogization with other verbs in the German construction. Instead, as for the verbs of sound emission and commerce and finance, I hypothesise that these arose by cross-linguistic analogization with the English construction. These clusters of verbs relate to

verbs of performance (190), bodily functions (191), sexual verbs (192), actions involving the mouth (193), and deceit (194).

(190) Er tanzt sich den Weg durch den Matsch
He dances REFL the way through the mud
'He dances his way through the mud'.

(191) Scheiß Dir deinen Weg in ein besseres Leben!
Shit REFL POSS way into a better life
'Shit your way to a better life![39]'

(192) Er küsste, leckte und nippte sich den
He kissed licked and nipped REFL the
Weg über jeden Zentimeter meines Körpers
way over every centimetre my.GEN body.GEN
'He kissed, licked, and nipped his way over every centimetre of my body'.

(193) Diese Kuh mampft sich den Weg
this cow munches REFL the way
ins Guinness Buch der Rekorde
into.the G. book of.the records
'This cow munches its way into the Guinness Book of Records'.

(194) Claudia würde sich den Weg auf
C. would REFL the way onto
die Anwender-PCs geradezu erschleichen
the user.PCs downright obtain.by.fraud
'Claudia would downright fraudulently trick her way into the user PCs'.

Also present in the DECOW data is an analogical extension of verbs of taking a path to other verbs of obtainment, as in (195); this analogical extension is made plausible by the very strong semantic link between these two clusters of verbs.

[39] Because *scheißen* 'shit' is orthographically very similar to *schießen* 'shoot', I checked the original source of this token to check whether this was a typo. I concluded that it was not a typo, because the source was a blog post containing famous quotations altered to include the verb *scheißen*, e.g. *Ich scheiße also bin ich* 'I shit therefore I am'.

(195) [...] musste sie sich den Weg
must.PAST she REFL the way
zu ihrer Karriere hart erarbeiten
to her career hard obtain.by.working
'She had to work her way into her career'.

The DECOW data also reveals a possible analogical extension from verbs of the manner of motion to verbs encoding the terrain of motion, as in (196). This is a highly likely analogical extension, because there is a strong semantic link between verbs describing the manner in which the subject moves, and verbs describing the manner in which the terrain is causing the subject to move.

(196) Erklimmt euch den Weg in den nördlichen Teil
Climb REFL the way into the northern part
'Climb your way into the northern part'.

The German data lead to the same conclusions as the English and Dutch data; the development of the German *way*-construction can be considered a long series of local analogical extensions. The semantic categories present in the construction in the three languages are very similar. There are many clusters of verbs in the Dutch and German constructions that have no plausible analogical source in those languages; I propose that the presence of this clusters is a result of cross-linguistic analogization with the English *way*-construction. In Figure 20 below, I summarise the sequence of analogical extensions I hypothesise took place in the German *way*-construction.

Figure 20: Hypothesised analogical extensions in the German way*-construction*

legen 'pave' (in precursor construction) > clearing a path > finding a path

taking a path > obtainment

vehicular motion > basic motion > manner of motion > terrain of motion

Borrowed from English: commerce/finance, sound emission, performance, bodily functions, sexual verbs, actions involving the mouth, deceit

The following section explores the relationship between analogy, productivity, and frequency. I show in this section that there is a positive correlation between high token frequency and the likelihood of an item undergoing analogical extension, as claimed by Israel (1996) and Barðdal (2008). I also demonstrate that, contrary to Barðdal's (2008) claim, items with a low token frequency may also be analogically extended. Barðdal also claims that analogical extension is dependent on low type frequency, but, as the following section will show, counterexamples to this claim are found in the English and German data. The following section also tests a hypothesis by Zeschel (2010); he claims that there is a correlation between the number of established (i.e. attested more than once) types in a semantic domain and the number of novel (i.e. attested only once) types in that domain. His hypothesis was found to hold for the data in all three languages.

5.3 The relationship between analogy, productivity, and frequency

5.3.1 Literature Review

This subsection presents two case studies from Barðdal (2008) and Zeschel (2010); the hypotheses put forward in these case studies will be tested against my own data in Section 5.3.2. Barðdal (2008: 89) proposes that the higher the type frequency of a construction, the lower the degree of semantic coherence is needed for it to be productive, and vice versa. She further states that analogical extension is based on the lowest possible type frequency, i.e. one, and the highest degree of semantic coherence, i.e. full synonymy between the source and target item. However, Barðdal

is here referring to analogical extensions taking place in morphology; the analogical extensions described in the subsection above cannot have involved full synonymy, because they involved analogical extension between different semantic domains.

Barðdal (2008: 89-90) finds that 88% of borrowed verbs in Icelandic contain an analogical extension based on only one model verb, and that there was full synonymy between the source item and the target, since they were translational equivalents. For instance, in (197) below, the English verb *dawn* was borrowed into Icelandic and used in a particle verb construction *dona uppi* on the basis of the existent particle verb *daga uppi* 'be forgotten/end up somewhere'; the latter particle verbs consists of *daga* 'dawn' and *uppi* 'up'. *Dona uppi* was therefore formed on the basis of the lowest possible type frequency, one, and the highest degree of semantic coherence since *daga uppi* and *dona uppi* are translational equivalents.

(197) Vörubíllinn donaði uppi í brekkunni
truck.the dawned up in slope.the
'The truck ended up in the slope' (Barðdal 2008: 91, ex. 17).

In the case of lower-level verb-specific analogical extensions, Barðdal (2008: 94) argues that high token frequency is more important than it is for the productivity of high type frequency constructions; this hypothesis will be tested on my data in Section 5.3.2. She cites a conversation between two speakers in which Speaker A borrowed the verb *applisera* 'apply' from Scandinavian into Icelandic and used it in a Nominative-Dative construction. Speaker B corrected Speaker A, claiming that the verb should be used in a Nominative-Accusative construction. Speaker A explained that *applisera* was used in the Nom-Dat construction on the basis of its Icelandic equivalent *beita*, whereas Speaker B associated *applisera* with the more general Icelandic verb *nota* 'use', which is used in the Nom-Acc construction (Barðdal 2008: 93). The reason Speaker B associated *applisera* with *nota* is that *nota* is much more frequent; it is attested 1,210,000 times in the Icelandic corpora used in Barðdal's study, whereas *beita* is attested only 223,000 times.

Bybee and Eddington (2006) state that an expression is acceptable to the extent it is retrievable from memory (because of its high token frequency), or semantically similar

to an entrenched unit in the speaker's network. These facts account for the differences between (198) and the closely related set of expressions in (199) to (201).

(198) To go ADJ
(199) To go mad/bankrupt/quiet
(200) To go crazy/insane/mental ... broke/bust/?skint ... calm/silent/?still
(201) To go ?famous/?popular/??renowned
(examples 1a-1d in Zeschel 2010: 201).

This leads Zeschel (2010: 202) to propose that the schema [*go* ADJ] does not exist; instead, speakers form new expressions by analogy based on the type frequency of each semantically distinct usage class. He further argues that type frequencies should be counted per semantically distinct usage class; thus, instead of counting the type frequency of [*go* $ADJ_{PROPERTY}$], Zeschel counts the type frequencies of [*go* $ADJ_{NORMAL\ MENTAL\ PROCESSING}$] and [*go* $ADJ_{ABLE\ TO\ PAY}$] separately. For this reason, when testing this hypothesis on my own data, I divide the verbs into semantic categories and count the type frequency of verbs within each semantic domain.

The likelihood of an analogical extension taking place is said to correlate with the number of similar expressions previously known to the speaker (Israel 1996). However, measuring what is known to the speaker is problematic; an item being attested in a corpus does not mean that a speaker will have experienced this item. The relationship between frequency counts in corpora and the degree of entrenchment in the mind of a speaker is unclear, because speakers' constructional networks are constantly changing (Zeschel 2010: 217-218, fn. 4).

This is especially true of large corpora; for instance, COHA contains around 400 million words, and it is estimated that we hear about 26 million words a year (Biber et al. 1998: 27). If we assume that COHA is representative of a speaker's knowledge of English and find that an item in that corpus has one or two attestations, we would have to assume that this expression is only encountered about once every 8-16 years. This is unlikely; it may be the case that the speaker simply does not know the expression (Taylor 2012: 15). The opposite phenomenon has also been observed; expressions that would be considered everyday language to a native speaker such

as *speak for yourself* often have no occurrences even in corpora of hundreds of millions of words (Foster 2001: 81).

Zeschel's case study concerns uses of the German adjective *tief* 'deep' in attributive Adjective-Noun contexts. The first level of semantic analysis is the distribution of *tief* in attributive contexts more generally (2010: 205). The second level of analysis is the most frequent of these figurative senses. Finally, the third level concerns the most frequent conceptual mapping within this sense. Zeschel predicts that the distribution of novel Adj-N combinations is not random, but that it will correlate with the number of established Adj-N combinations in different regions of the overall semantic map (2010: 205). The notion of 'established' and 'novel' run into problems, however; a token that is only attested once is not necessarily novel, and a token that is attested more than once is not necessarily well-established in the speech community (Zeschel 2010: 217-218, fn. 4).

Zeschel extracted 10,000 tokens of the adjective *tief* in all relevant inflectional forms, and removed all tokens in which *tief* did not function as an attributive adjective. The uses of *tief* were classified semantically according to a monolingual dictionary of German, which lists seven senses for *tief*. Classifying these was not always straightforward, as the categories sometimes overlapped (2010: 207). For example, is *tiefe Überzeugung* a deep conviction (corresponding to his sense 5), or a profound conviction (sense 6)? The data were analysed by two separate coders and interrater agreements were calculated. The problem of overlapping categories was also present in the *way*-construction in the three languages; for instance, *sing* is both a verb of performance and an action involving the mouth.

Zeschel (2010: 211) performed a correlation analysis between the number of established types and the number of novel types in a given category, in order to test the correlation between type frequency and productivity (i.e. the ability of a given schema to be applied to a new item), and found a strong correlation between the two variables, which suggests that the distribution of novel Adj-N combinations is not random; rather, it correlates with the number of established Adj-N combinations in all domains. I am of course not dealing with Adj-N combinations in this study; when testing this hypothesis against my own data, I consider the type frequency of verbs in the *way*-construction in each semantic domain, and measure the correlation between

the number of novel verbs in each domain with the number of established verbs across all semantic domains. I find that there is a strong correlation between the number of established verbs in the *way*-construction in all semantic domains and the likelihood of analogical extension within a particular semantic domain.

5.3.2 Results

5.3.2.1 English

The following paragraphs are concerned with the relationship between the likelihood of an item undergoing analogical extension, type frequency, and token frequency. I test two hypotheses by Barðdal (2008) against my data on the development of the English *way*-construction; a) that analogical extension is dependent on high token frequency; and b) that analogical extension is dependent on low semantic type frequency. The type and token frequencies were calculated by counting the number of types and tokens across all corpora and calculating a normalised value per million words.

I documented in Section 5.2.2.1 that a cluster of force dynamic verbs emerges in the English *way*-construction from 1658, and that verbs of finding a path emerge by analogization with this cluster from 1726. In 1726, the normalised type and token frequency of the force dynamic verbs is 0.53 per million words; this is a rather low token frequency, which suggests that low frequency items may serve as the basis for analogical extension. This finding is consistent with Barðdal's (2008) hypothesis that analogy is dependent on low type frequency.

By 1744, when other verbs of path creation are first attested in the construction, there is an increase in token frequency in the force dynamic verbs from 0.50 to 0.70 per million words, while their type frequency remains stable at 0.53 per million words. This is further evidence that items of low token frequency may serve as the basis of analogical extension, contrary to Barðdal's (2008) claim.

When verbs describing the manner of motion are first attested in 1744 by analogical extension of verbs of basic motion, basic motion verbs have a normalised token frequency of 7.19 per million words, and a normalised type frequency of 2.11 per

million words. This high token frequency supports Barðdal's (2008) hypothesis that items of high token frequency may serve as the basis for analogical extensions. However, the relatively high type frequency of these verbs suggests that low semantic type frequency is not necessary for analogical extension to take place.

I hypothesised above that by 1752, the verbs describing the manner of motion underwent an analogical extension to verbs describing the path shape. At this point, the manner of motion verbs had a relatively low normalised type and token frequency (both 0.05 per million words). If my hypothesis is correct, this suggests that items of both low and high token frequency may serve as the basis for analogical extensions. This also confirms the hypothesis that analogy is dependent on low type frequency, as Barðdal (2008) claims.

I also hypothesised that by 1834, the verbs encoding the path shape underwent an analogical extension to verbs encoding the terrain of motion. The path shape verbs have a normalised token frequency of 1.89 per million words, and a normalised type frequency of 0.84 per million words by this time. This is further evidence that items of low token frequency may be the source of an analogical extension, and supports the notion that analogy is dependent on low type frequency.

When verbs of clearing a path, which were first attested in my data in 1852, arose by analogical extension of verbs of taking a path, the type frequency of the latter cluster of verbs is relatively low: 0.20 per million words. However, their token frequency is considerably higher, at 0.92 per million words. The token frequency is not as high as that of some of the other clusters, which suggests that Barðdal's (2008) hypothesis concerning token frequency cannot be maintained in this case, though her hypothesis that analogical extension is dependent on low type frequency can be supported.

Also by analogical extension of *croon*, verbs of sporting performance are first attested in 1973; I proposed this analogical extension because both verbs encode a kind of performance. Between 1824 and 1973, the token frequency of performance verbs increases from 0.02 to 0.37 per million words, while the type frequency of these verbs remains relatively low, although it undergoes a slight increase from 0.02 to 0.04 per million words. Once again, Barðdal's (2008) hypothesis on type frequency can be supported, as the type frequency of these verbs is one of the lowest of all of the

semantic clusters. However, this finding suggests that, contrary to her hypothesis, items of low token frequency may also serve as the basis for analogical extensions.

I also proposed in Section 5.2.2.1 that *joke* arose by analogical extension with *beg*, as both verbs encode a type of social interaction. By the first attestation of *joke* in 1989, *beg* was the only other verb of social interaction attested, and it was attested only once. This finding appears to support the notion that items of low token and type frequency may serve as the basis for analogical extensions. While it is unlikely that hapaxes may serve as the basis of analogical extensions, no other plausible source of such an extension is present in my data.

In sum, the English data on analogical extension in the *way*-construction between different semantic categories show that the relationship between analogical extension, token frequency, and type frequency may be rather different than Barðdal (2008) has suggested. It was shown that, contrary to her claim, items of low token frequency were often the locus of an analogical extension. Counterevidence to her claim that analogical extension is dependent on low semantic type frequency was also presented.

I have considered up to this point the possibility of analogical extensions between semantic categories. However, analogical extension also takes place within the same semantic category; for instance, when a speaker uses *push* in the *way*-construction, this is by analogical extension of other verbs encoding force dynamics, such as *force*. The remainder of this subsection documents the analogical extensions within each semantic category in each of the English corpora, and tests the following two hypotheses: that the number of novel items in a semantic category correlates positively with the number of similar tokens previously experienced (Israel 1996), and that the number of novel verbs in each semantic category shows a positive correlation with the number of established (i.e. occurring more than once) verbs in all semantic domains, as Zeschel (2010) claims. I perform this analysis on the CLMET3.1 and COHA corpora only, because these two corpora alone cover a wide time period (1720-2000), and these two corpora are where the vast majority of the analogical extension took place. I demonstrate that these two hypotheses can convincingly be upheld based on my data; this is evidence of frequency effects in the development of the

English *way*-construction, because the analogical extensions that took place are dependent on type and token frequency.

Table 22 below illustrates the number of verbs in each semantic category in the CLMET3.1 corpus, and how many of these are established and novel. The value in the 'verbs' column denotes the type frequency of verbs in the corpus in that semantic domain. The values in the other two columns denote the number of verbs with a token frequency of greater than one ('established') and the number of verbs with a token frequency of one ('novel') respectively.

Table 22: Number of verbs per semantic domain in CLMET3.1

Domain	Verbs	Est.	Novel	Domain	Verbs	Est.	Novel
actions involving the mouth	4	1	3	other activities	8	4	4
attack/coercion/deceit	2	1	1	path creation	30	13	17
basic motion verbs	4	4	0	path shape	6	3	3
body-internal motion	5	3	2	performance	1	0	1
clearing a path	3	0	3	seeing a path	1	0	1
clumsy motion	3	0	3	sound emission	5	0	5
cognitive verbs	1	0	1	speed of motion	8	3	5
commerce/finance	1	1	0	taking a path	5	2	3
force dynamics	17	7	10	terrain of motion	2	0	2
finding a path	1	1	0	vehicular motion	3	1	2
laborious motion	2	1	1				

Zeschel (2010) calculates the correlation between the numbers in the rightmost two columns in his case study. The CLMET3.1 data yielded a correlation coefficient of 0.860, which indicates a very strong positive correlation between the two variables. This correlation was found to be statistically significant at $p < 0.05$. This suggests a very strong link between the number of novel verbs in a semantic domain and the number of established verbs in all domains, and that there are therefore frequency effects in the analogical extensions of the English *way*-construction, because the possibility of analogical extension is dependent on the type frequency of established verbs. This table also shows that almost all of the semantic categories have undergone their own analogical extensions; only the clusters of verbs encoding

cognition, commerce and finance, finding a path, performance, and seeing a path have not undergone analogical extension.

The following table shows the results of testing Israel's (1996) hypothesis that the likelihood of an item undergoing analogical extension depends on the number of similar tokens (i.e. in the same semantic domain) experienced by the speaker. The value returned by this calculation is Pearson's correlation coefficient between the following two variables. The first variable is the number of similar tokens in a semantic category at the date of attestation of a particular token. The second variable is the number of novel items (i.e. with a token frequency of one) at that date. For each semantic category, all tokens were examined in chronological order with the number of similar and novel items being recalculated after each token. Only semantic domains with a type frequency of 5 or greater were analysed, in order to ensure reliability of the results. This constraint is applied to data from both of the corpora analysed here.

Table 23: Correlation between the number of similar and novel items per semantic domain in CLMET3.1

Domain	Correl. Sim/novel
body-internal motion	0.933
force dynamics	0.938
other activities	0.975
path creation	0.971
path shape	0.968
sound emission	0.959
speed of motion	0.961
taking a path	0.920

Data from CLMET3.1 support Israel's hypothesis that the likelihood of analogical extension is dependent on the number of similar tokens previously attested; there is a very strong positive correlation between these two variables for all of the semantic domains tested here. All correlations were found to be significant at $p < 0.05$; this is further evidence of frequency effects in the development of the English *way*-construction, because the analogical extensions that took place were dependent on the token frequency of similar items attested. In the following table, I show the analogical expansion taking place within each semantic category in COHA.

Table 24: Number of verbs per semantic domain in COHA

Domain	No. V	Est.	Novel	Domain	No. V	Est.	Novel
actions involving the mouth	19	11	8	obtainment	2	2	0
attack/deceit/misconduct	27	20	7	path creation	9	9	0
basic motion	5	5	0	path shape	23	14	9
bodily functions	7	3	4	path terrain	5	3	2
body-internal motion	33	26	7	performance	8	4	4
change of state	1	1	0	removal of obstacles	3	3	0
clearing a path	32	29	3	social interaction	17	9	8
cognitive verbs	8	8	0	sound emission	81	44	37
commerce/finance	13	8	5	sporting verbs	1	1	0
force dynamics	53	38	15	taking a path	3	3	0
finding a path	4	3	1	terrain of motion	2	1	1
incompetence	7	6	1	tracing a path	8	8	0
laborious motion	12	11	1	unaccusative verbs	1	0	1
light emission	5	3	2	vehicular motion	9	7	2
manner of motion	52	24	28				

As this table shows, every semantic domain other than verbs encoding a change of state, sporting verbs, and unaccusative verbs have undergone analogical extension when compared with the CLMET3.1 corpus. Subjecting the number of established and novel verbs to a correlation analysis yielded a correlation of 0.916, an extremely strong correlation which was found to be significant at $p < 0.05$. This suggests that the distribution of novel items in each semantic domain is not random; rather, it correlates with the number of established verbs in all domains, which points to frequency effects in the development of the construction. The following table shows the measure of correlation between the number of novel verbs in a domain and the number of similar tokens previously experienced.

Table 25: Correlation between the number of similar and novel items per semantic domain in COHA

Domain	Correl. Sim/novel	Domain	Correl. Sim/novel
actions involving the mouth	0.989	light emission	0.981
attack/deceit/misconduct	0.964	manner of motion	0.987
basic motion	0.894	path creation	0.853
bodily functions	0.990	path shape	0.976
body-internal motion	0.970	path terrain	0.981
clearing a path	0.926	performance	0.986
cognitive verbs	0.957	social interaction	0.968
commerce/finance	0.966	sound emission	0.989
force dynamics	0.984	tracing a path	0.904
incompetence	0.958	vehicular motion	0.985
laborious motion	0.948		

These data once again show that Israel's hypothesis accounts for the data; each semantic domain shows a very strong positive correlation between the number of analogical extensions and the number of tokens previously attested by the date of that extension. All correlations were found to be statistically significant at $p < 0.05$. This provides convincing evidence of frequency effects in the development of the English *way*-construction, because it was shown that the analogical extensions of the construction were dependent on token frequency.

In sum, the data from the English corpora have shown that Barðdal's (2008) hypothesis that frequently occurring items are the basis for analogical extension does not always account for the development of the *way*-construction, though her hypothesis that analogical extensions are dependent on low semantic type frequency was upheld in most cases. The data also reveal that analogical extensions have caused a marked increase in the productivity and schematicity of the construction, as claimed by Israel (1996) and Traugott and Trousdale (2013: 148), among others. Israel's (1996) hypothesis concerning category-internal analogical extension was shown to have provided a convincing account of the data, because a very strong positive correlation between the number of novel items of the *way*-construction in a domain and the number of previously attested similar tokens was found. Further, Zeschel's (2010) hypothesis that the distribution of novel items in a domain is not random was found to hold for the data from both the CLMET3.1 and COHA corpora; the number of novel items in each semantic domain correlates strongly with the

number of established tokens in all domains. It was also shown that the analogical extensions of the *way*-construction were strongly correlated with the token frequency of semantically similar items, and that there are therefore frequency effects in the development of the English *way*-construction.

5.3.2.2 Dutch

The following paragraphs explore the relation between analogical extension, type frequency, and token frequency in the development of the Dutch *way*-construction. I show that, as is the case of the English data, high token frequency is not necessary for analogical extension to take place, and that low semantic type frequency may drive analogical extension.

As mentioned above, the lack of information on the year of publication of a text in the SoNaR and NLCOW corpora makes a quantitative analysis on the change in frequency over time uninformative. Further, the type and token frequency of the *way*-construction in *De Gids*, where information on the year of publication of the text is available, is very low. I therefore do not perform a quantitative analysis on data from *De Gids* here.

I hypothesised above that the cluster of verbs in the SoNaR corpus relating to sound emission and sporting performance arose by analogization with verbs encoding another kind of performance, such as *zingen* 'sing'. The type and token frequency of verbs encoding performance is low; only 3 of the 107 verbs (2.80%) attested in the construction encoded performance, and these verbs represented only 5 of 807 tokens (0.62%). This suggests that high token frequency is not required for analogical extension to take place, contrary to Barðdal's (2008) claim, and that analogical extensions can take place even if the item has a low semantic type frequency; this latter finding is consistent with what Barðdal (2008) suggests.

I proposed that in the SoNaR and NLCOW corpora, the cluster of verbs of sound emission undergoes an analogical extension to verbs of light emission. The type frequency of the sound emission verbs is relatively high; of the 442 verbs attested in these corpora, 34 were a verb of sound emission (7.69%). The token frequency of these verbs is relatively low; of the 9,487 tokens of the construction in the two corpora,

only 62 contained a verb of sound emission (0.65%). Taken together, these findings suggest that both low token frequency and low type frequency may lead to analogical extension.

I turn now to category-internal analogical extension in the Dutch *way*-construction. Given the lack of availability of the year of publication of the texts in the SoNaR and NLCOW corpora, and that testing Israel's hypothesis (i.e. that the likelihood of analogical extension correlates with the token frequency of semantically similar items) requires a calculation of the number of established and novel types up to the date of an analogical extension, only Zeschel's hypothesis will be tested in this subsection. Table 26 below shows the category-internal analogical extensions in the Dutch *way*-construction found in the SoNaR corpus.

Table 26: Number of verbs per semantic domain in SoNaR

Domain	No. V	Est.	Novel	Domain	No. V	Est.	Novel
sports/games	4	3	1	sound emission	9	1	8
basic motion	2	1	1	commerce/finance	1	0	1
manner of motion	7	4	3	laborious activity	4	2	2
force dynamics	31	17	14	social interaction	1	0	1
taking/clearing a path	5	2	3	performance	3	1	2
finding a path	4	1	3	bodily functions	1	0	1
actions involving the mouth	8	3	5	light emission	1	0	1
internet verbs	2	1	1	vehicular	1	0	1
deceit	2	1	1	incompetence	1	0	1

Data from the SoNaR corpus reveal that many of the semantic categories of verbs have undergone analogical extension since their emergence. The correlation between the number of novel verbs in a domain and the number of established verbs in all domains was 0.856 (statistically significant at $p < 0.05$); a very strong correlation which suggests that Zeschel's hypothesis may be upheld in this case, and therefore confirms that there are frequency effects in the development of the Dutch *way*-construction, because the possibility of analogical extension is dependent on the type frequency of other verbs in the construction.

Table 27 shows the semantic categories of verbs found in the construction in the NLCOW corpus, and how many of these verbs are established and novel.

Table 27: Number of verbs per semantic domain in NLCOW

Domain	No. V	Est.	Novel	Domain	No. V	Est.	Novel
force dynamics	53	37	16	vehicular motion	7	3	4
basic motion	4	3	1	commerce/finance	2	1	1
cognitive verbs	12	3	9	attack/coercion/misconduct	12	4	8
path-clearing	12	6	6	terrain of motion	4	3	1
path-creation	79	37	42	social interaction	9	0	9
manner of motion	32	15	17	sex	7	3	4
actions involving the mouth	13	8	5	speech	10	4	6
performance	16	3	13	bodily functions	5	1	4
path shape	15	7	8	sound/light emission	31	12	19
Playing	11	4	7	ingestion	2	1	1
body-internal motion	14	11	3	verbs of appearance	2	0	2
taking a path	1	1	0	incompetence	8	2	6
internet verbs	8	3	5	achievement verbs	1	0	1
laborious motion	7	5	2	change of state	1	1	0
sporting verbs	11	4	7	olfaction	1	0	1

The NLCOW data show that the collocational range of the construction has expanded even further in comparison with the SoNaR corpus. The correlation between the number of novel verbs in a domain and the number of established verbs in all domains was found to be 0.804 (statistically significant at $p < 0.05$); this suggests that Zeschel's hypothesis provides a convincing explanation of the Dutch data, because the distribution of novel items in each semantic domain correlates positively with the number of established verbs in all domains. This in turn points to the existence of frequency effects in the development of the Dutch *way*-construction.

The Dutch data presented in this chapter yield much the same conclusions as the English data. The Dutch *way*-construction has undergone a long series of local analogical extensions which has led to the construction becoming much more productive. The construction has extended its collocational range from only *banen* to several kinds of manner of motion verbs, and verbs which encode an incidental activity accompanying motion. Barðdal's (2008) hypothesis that analogical extension is dependent on low semantic type frequency was found to account for the Dutch data. However, contrary to her hypothesis on token frequency, it was shown in this subsection that items of low token frequency may serve as the basis of an analogical extension. It was also found that the likelihood of an item undergoing analogical extension was found to be strongly correlated with the number of established verbs

in the construction in all semantic domains, as Zeschel (2010) claims. This suggests that part of the development of the Dutch *way*-construction can be explained by frequency effects.

5.3.2.3 German

The following paragraphs explore the relationship between analogical extension, type frequency, and token frequency in the development of the German *way*-construction. Because the type and token frequency of the construction in the *Deutsches Textarchiv* corpus is very low in comparison with the other German corpora (only 46 tokens and 25 types were found), I perform a qualitative analysis only on the *Berliner Zeitung* and DECOW data. In particular, this subsection seeks to answer whether there is a correlation between the possibility of analogical extension of an item and its token frequency, as Barðdal (2008) claims. I also test Barðdal's (2008) hypothesis that analogical extension is dependent on low semantic type frequency, and find that both of her hypotheses provide a convincing explanation of the data.

I demonstrated in Section 5.2.2.3 that the DECOW data reveal an analogical extension of verbs of taking a path to other verbs of obtainment. In the DECOW corpus, the normalised token frequency of verbs of taking a path increased from 0.005 to 0.23 per million words, with the type frequency increasing from 0.005 to 0.09 per million words. The type frequency is relatively high compared to the other semantic clusters of verbs, which contradicts the hypothesis that low semantic type frequency is necessary for analogical extension to take place. The normalised token frequency is also much higher than that of the other semantic clusters of verbs; Barðdal's (2008) hypothesis that high token frequency may lead to analogical extension can therefore be supported in this case.

The DECOW data also reveal an analogical extension from verbs of the manner of motion to verbs encoding the terrain of motion. By the time this analogical extension took place, there was an increase in the normalised token frequency of basic motion verbs from 0.005 to 1.15 per million words, while the type frequency increased from 0.005 to 0.014 per million words. Although there was a sharp increase in token frequency, the type frequency remained relatively low; this supports both of Barðdal's (2008) hypotheses concerning the relation between analogical extension, type

frequency, and token frequency. The following paragraphs discuss analogical extension within the semantic categories of verbs found in the German *way*-construction in my data.

Table 28 shows the number of verbs in each semantic domain in the German *way*-construction in the *Berliner Zeitung* corpus.

Table 28: Number of verbs per semantic domain in Berliner Zeitung

Domain	No. V	Est.	Novel
basic motion	3	3	0
clearing a path	9	2	7
commerce/finance	2	0	2
finding a path	2	1	1
force dynamics	7	3	4
internet verbs	1	0	1
manner of motion	4	0	4
path-creation	7	2	5
sound emission	1	1	0
taking a path	6	2	4
tracing a path	3	1	2
vehicular motion	1	1	0

A very weak positive correlation was found between the number of novel verbs in a domain and the number of established verbs in all domains; the correlation coefficient between these two variables is 0.166, and was not found to be statistically significant. These findings contradict the hypothesis that there is a correlation between the number of novel items in a domain and the number of established verbs in all semantic domains. The following table shows the results of testing Israel's hypothesis that the likelihood of an item undergoing analogical extension is dependent on the number of previously attested similar tokens.

Table 29: Correlation between the number of similar and novel items per semantic domain in Berliner Zeitung

Domain	Correl. Sim/nov
clearing a path	0.950
force dynamics	0.953
manner of motion	0.982
path-creation	0.920
taking a path	0.992

These findings suggests that Israel's hypothesis accounts for the data; there is a strong positive correlation between the possibility of analogical extension and the number of tokens previously attested. The following table shows the number of verbs in each semantic domain in the DECOW corpus. Given that the year of publication of each text is not available, I do not test Israel's hypothesis against the data from this corpus.

Table 30: Number of verbs per semantic domain in DECOW

Domain	No. V	Est.	Novel	Domain	No. V	Est.	Novel
basic motion verbs	2	2	0	sexual verbs	3	1	2
force dynamics	15	7	8	attack/coercion/deceit	2	0	2
clearing a path	35	15	20	actions involving the mouth	2	1	1
bodily functions	2	0	2	commerce/finance	1	0	1
tracing a path	6	4	2	sound emission	2	0	2
body-internal motion	1	1	0	taking a path	1	1	0
obtainment	12	5	7	path shape	2	0	2
terrain of motion	1	0	1	sporting verbs	1	0	1
path creation	5	2	3	speech	1	0	1
performance	1	0	1				

The DECOW corpus reveals a number of additional semantic clusters in comparison with the data from the *Berliner Zeitung* and *Deutsches Textarchiv* corpora. A strong positive correlation was found between the number of novel items in a semantic domain and the number of established verbs in all domains; the correlation coefficient is 0.871, and is statistically significant at $p < 0.05$. This suggests that there is a strong link between analogical extension and type frequency.

The data presented in this section have shown that analogical extension has caused the construction to become far more productive since its emergence in Middle German. Barðdal's (2008) hypotheses were often contradicted by the data, with items of high semantic type or low token frequency serving as the basis for analogical extensions. Israel's hypothesis provided a convincing account of the *Berliner Zeitung* data; I found a strong positive correlation between the number of novel items in a domain and the number of similar tokens attested by the date of that analogical extension. I also found a strong positive correlation and between the number of analogical extensions within a semantic domain and the number of established verbs in all domains in the DECOW data, confirming Zeschel's (2010) hypothesis. The *Deutsches Textarchiv* and *Berliner Zeitung* data did not support this hypothesis, with weak positive or even negative correlation found between these two variables. However, these correlations were not found to be statistically significant, and therefore do not pose a serious problem to Zeschel's hypothesis, especially given the relatively low type and token frequency of the construction in these corpora.

When the English, Dutch, and German data are viewed together, it has been shown conclusively that items of low token frequency may be analogically extended; this contradicts Barðdal's (2008) claim. Counterexamples to her hypothesis that analogical extension depends on low type frequency were found in the English and German data. Israel's hypothesis concerning analogical extension, i.e. that the likelihood of an item undergoing analogical extension is dependent on the number of similar expressions previously known to the speaker, was convincingly upheld in all cases in which it was able to be tested. Zeschel's proposed correlation between the number of established and novel types was also found in the three languages. As a result of the long series of analogical extensions, the *way*-construction in all three languages has become more productive, with the number of verbs that can appear in the construction expanding considerably since its constructionalization.

5.4 Bolstering, support constructions, and constructional contamination in a usage-based theory of analogical thinking and analogization

5.4.1 Literature Review

Although analogy is very commonly applied to morphology (see e.g. McMahon 1994, Lahiri 2000, Deutscher 2001) and sound change (Hock 2003: ch. 4 and references there), it is much less often applied to syntactic change or constructions (though see Itkonen & Haukioja 1997 for a computer program for syntactic analogy). The following paragraphs explain how processes in constructional change may be considered instances of analogical extension and analogization.

Abbot-Smith and Behrens (2006) invoke the notion of supporting constructions. The term 'supporting construction' refers to a construction that has lexical or morphological similarities with another construction, but this term is problematic because it ignores functional similarities that may exist between constructions. McColm and Trousdale (2019) instead couch their discussion in terms of bolstering. This term captures the fact that there may be a formal or functional alignment (or both) between two constructions, with this alignment strengthening the representation of another construction. De Smet and Fischer (2017: 243) argue that the notion of supporting constructions (or bolstering) can explain how constructions in diachrony spread analogically; the more similar the outcome of analogy resembles an existing construction, the more likely it is to be analogically extended (cf. the references in De Smet & Fischer 2017: 243). This is described by Rosenbach (2010: 171) as structural analogy, a process whereby the properties of one construction are transferred to another one. This hypothesis will be tested of the *way*-construction and the fake reflexive resultative construction in the following subsection; the two constructions share formal and functional similarities, which makes the possibility of constructional contamination more likely.

A notion similar to supporting constructions and bolstering is that of constructional contamination (Pijpops & van de Velde 2016). In constructional contamination, a token of a construction may be produced in a deviant manner due to its superficial

resemblance with another construction; the latter construction is known as the contaminating construction, and usually does not merge with the target construction (Pijpops & van de Velde 2016: 543-544). Mondorf (2011) suggests that the fake reflexive resultative construction may be contaminating the *way*-construction; this hypothesis will be tested in the following subsection. The notion of constructional contamination has implications for the nature of the constructional network; constructional contamination suggests that the constructional network is more than a series of constructions with vertical inheritance links and horizontal opposition links (Pijpops & van de Velde 2016: 549).

Diachronically, constructional contamination may arise from a similarity on the formal or semantic level; formal similarities between two constructions may lead to semantic convergence, and semantic similarities between two constructions may lead to a formal convergence (Pijpops & van de Velde 2016: 547). An example of the latter kind of contamination can be found in German utterances such as the ungrammatical **das gehört mein*, a combination of the target construction *das gehört mir* 'that belongs to me' and *das ist mein* 'that is mine' (Harris & Campbell 1995: 118, exx. 43a-b, 44). This contamination is an instance of analogical thinking; the speaker perceives a functional similarity between the two constructions and this leads to the contamination. Synchronically, contamination effects may arise because a given construction has multiple sources (Pijpops & van de Velde 2016: 547; see also the papers in van de Velde et al. 2013).

Although the instance of constructional contamination discussed in the paragraph above has not led to change, examples that do lead to change are well attested. For such an example, consider the two Dutch sentences in (202) and (203).

(202) in begin van de week iets verkeerd gegeten
in beginning of the week something wrong eaten
'I ate something wrong at the start of the week' (Pijpops & van de Velde 2016: 544, ex.1)

(203) dat iets verkeerd geïnterpreteerd wordt?
that something wrong interpreted gets
'That something gets interpreted wrongly?' (Pijpops & van de Velde 2016: 545, ex. 2).

(202) is an example of a partitive genitive construction in which the adjective usually carries an *-s* suffix. In (203), an adverbial construction, an *-s* suffix on the adverb is ungrammatical, but examples of *-s* suffixes on adverbs are now attested as a result of contamination from the partitive genitive construction. This process, again, is driven by analogical thinking; when speakers make such analogies, they attempt to align whole constructions and all of the elements in these constructions on a one-to-one basis, guided by the functional roles of these elements (Tomasello 2003: 165). In this case, there is an alignment between the adjective in the partitive genitive construction and the adverb in the adverbial construction.

To measure constructional contamination, instances of the target construction need to be identified that resemble instances of the contaminating construction; it also needs to be investigated whether strictly unambiguous instances of the target construction are affected (Pijpops & van de Velde 2016: 559). Pijpops and van de Velde (2016) propose four measures of constructional contamination. The first of these is a 'type-adjective' measure, which looks at the semantic categorisation of adjectives that appear in both constructions. The other three measures are numerical variables which calculate how likely a token is to be affected by a contaminating construction (Pijpops & van de Velde 2016: 563). One of these variables is known as 'partial string resemblance', which requires a partial overlap between the target construction and contaminating construction. The other two variables are 'string resemblance', which requires a full overlap between the two constructions, and 'semantic string resemblance', which requires a form of semantic resemblance between the two constructions, and measures how often a phrase occurs in a string that is syntactically ambiguous (ibid.). In the following subsection, I test whether there is constructional contamination between the *way*-construction and fake reflexive resultative construction (FRR). Given I only examined verbs which occurred in both constructions, using a variable such as partial string resemblance would lead me to conclude that constructional contamination is 100% likely to occur between the two

constructions, which is incorrect; for this reason, I adopt a qualitative approach to the analysis in this section.

5.4.2 Constructional contamination between the *way*-construction and the FRR

5.4.2.1 English

Mondorf (2011: 399) points out that there are several constructions that are functionally similar to the English *way*-construction. One of these is the fake reflexive resultative construction (cf. Simpson 1983). The FRR is attested earlier than the *way*-construction, and is sometimes used in the same context as the *way*-construction, e.g. *He worked his way/himself* to the top (Mondorf 2011: 398). As well as the FRR, Mondorf mentions three other functionally similar constructions: a construction with a determiner and a noun such as *path* (204), a construction with dummy *it* in object position (205), and a zero variant (206).

(204) Max forced an entrance to the house.
(205) Max bicycled it through the country.
(206) Max hiked across the US.
(Mondorf 2011: 399, exx. 4c-4e).

Due to the vast number of tokens, a detailed analysis of all of these functionally similar constructions is beyond the scope of this thesis. The analysis in this section will therefore be restricted to verbs which occur in both the *way*-construction and the FRR. I searched the English corpora[40] for the verbs attested in the *way*-construction in my data followed by a reflexive pronoun and a preposition. For non-tagged corpora, I searched for all spelling variants of a reflexive pronoun[41] and filtered out irrelevant results manually. This part of the thesis builds on Mondorf's (2011) study; Mondorf's investigation is limited to 10 verbs that occur in the *way*-construction and FRR, whereas I found 58 such verbs. In this section, I test Mondorf's (2011) hypothesis that the *way*-construction is supplanting the FRR, and investigate whether there is constructional contamination or bolstering between the two constructions. Because

[40] A list of these corpora can be found in Section 3.3.1.
[41] See Mondorf (2011: 401) for a complete list of these spelling variants.

of the high type frequency of the verbs, I do not present tokens of every verb here; instead, I use a few examples to illustrate that, while the FRR and *way*-construction are similar, there are some important differences in aspect, telicity, and function.

Of the 58 verbs that occur in my data in both the FRR and *way*-construction, 29 were attested in the FRR before the *way*-construction. These verbs, along with the date of their first attestation and frequency in both constructions are shown in Table 31 below.

Table 31: Verbs attested in the FRR before the way*-construction*

Verb	First att. in FRR	Freq. in FRR	First att. in way	Freq. in way
toast	1670	1	2009	1
sleep	1670	1	1822	24
extend	1721	1	1906	1
toil	1759	1	1826	15
intrude	1790	1	1999	1
weep	1822	5	1922	1
talk	1823	20	1894	397
press	1826	40	1843	40
drink	1832	128	1868	58
argue	1833	21	1892	18
read	1834	2	1921	41
stamp	1835	3	1922	5
dream	1849	4	1962	17
play	1852	4	1946	72
cry	1853	5	1991	8
drill	1855	4	1951	28
laugh	1861	11	1913	34
sob	1869	2	1916	4
worry	1874	2	1894	8
breathe	1874	1	1988	6
whisper	1889	1	1993	11
screech	1904	1	1990	4
scream	1913	3	1932	10
howl	1918	1	1961	6
chant	1920	2	1952	3
pry	1958	2	1976	7

At first glance, it may seem implausible that the FRR is contaminating the *way*-construction, given its much lower token frequency; however, Lieven (2017) and Ellis (2017) argue that speakers may analogize infrequent constructions to extant constructions provided there exists a similarity in form and meaning between the two constructions. A formal similarity exists between the two constructions; the FRR has

the formal schema [NP_i [V $REFL_i$ DIR]], and the schema of the *way*-construction differs only in that there is a possessive pronoun followed by the noun *way* instead of a reflexive pronoun. There also exists a semantic similarity; the sentences *He worked his way to the top* and *He worked himself to the top*[42] both code the event of a person reaching the top (of his field) by hard work.

As Mondorf (2011) claims, the FRR can be used for the same function as the *way*-construction, as the examples below show.

(207) Even now I could lie down, and weep myself to death (COHA, 1822).
(208) One young woman claimed to have wept her way into a degree (BNC, 1991).

These constructions are very similar, because both encode the speaker reaching a goal, and this goal is expressed by a PP in both constructions. Further, both (209) and (210) are telic. This is confirmed by Garey's (1957: 105) test for differentiating telic and atelic activities. To differentiate between the two, Garey asks the following question: "if one was *verb*ing but was interrupted while *verb*ing, has one *verb*ed?' (italics original). If the situation is atelic, the answer may be 'yes', but in telic situations the answer must be 'no'" (as quoted in Brinton 1988: 25-26). If you are interrupted while weeping yourself to death or weeping your way into a degree, you have not wept yourself to death or wept your way into a degree; therefore both (207) and (208) are telic.

The similarity between the two constructions also holds of other verbs as (209) and (210) show.

(209) I'd rather talk myself into a consumption [...] than scribble myself cold (COHA, 1823).
(210) It was like talking my way into protected or restricted places so I could gain intel (COCA, 2010).

[42] Trousdale (pers. comm.) points out that this latter sentence is not grammatical for all speakers.

Both (209) and (210) are telic, and contain a goal that is reached. This goal is expressed by a PP headed by *into* in both cases. However, there is a small difference in that the goal in (209) is unintended, while the goal in (210) is intended.

Although the FRR and *way*-construction can often be used to express the same event, there are nonetheless important differences between the constructions in their aspect and telicity, as (211) and (212) show.

(211) Drink yourself to oblivion. See how far you can go (COCA, 1999).
(212) Jabelman had been an urbane , well-educated man who was drinking his way to oblivion (BNC, 1992).

(211) and (212) differ in that (211) is an imperative perfective and presumably the subject is being told to drink until he reaches the endpoint (i.e. oblivion). (212), on the other hand, is a progressive imperfective and does not entail that the endpoint has been reached; the subject could be drinking his way to oblivion but later become teetotal and not reach the endpoint.

The difference in aspect and telicity between the two constructions is also present in (213) and (214); (213) is perfective and telic, whereas (214) is progressive, imperfective and atelic. Further, the FRR cannot be used to express the meaning of (214) (**She breathed herself along with the music*).

(213) Let your anger breathe itself on me (COHA, 1874).
(214) [W]ith a round Peter Pan collar she unbuttoned slowly, breathing her way along with the music (BNC, 1988).

The difference between the two constructions is also illustrated by (215) and (216).

(215) At one time he would be devoured by suspicions, at another he would try to laugh himself out of them (COHA, 1861).
(216) He laughed his way through life, apparently without care (COHA, 1913).

In (215) the goal is not necessarily reached, because of *try to*; (216) is iterative and refers to many instances of laughing over the course of the subject's life. The meaning of (216) cannot be conveyed by the *way*-construction; **He laughed himself through life* is ungrammatical.

When considering the possibility of constructional contamination, the gap between the year of first attestation in both constructions must be taken into account. A token that is very infrequent and attested a long time ago is unlikely to be a source of constructional contamination, because it is unlikely to be very entrenched in speakers' constructional networks. Table 32 below shows the gap between the first attestation of each verb in the FRR and the *way*-construction.

Table 32: Gap between first attestation of each verb in the FRR and the way-*construction*

Verb	First att. FRR	First att. Way	Gap (yrs)	Verb	First att. FRR	First att. way	Gap (yrs)
toast	1670	2009	339	*play*	1852	1946	94
sleep	1670	1822	152	*cry*	1853	1991	138
extend	1721	1906	185	*drill*	1855	1951	96
toil	1759	1826	67	*laugh*	1861	1913	52
intrude	1790	1999	209	*grind*	1864	1890	26
weep	1822	1922	100	*sob*	1869	1916	47
talk	1823	1894	71	*worry*	1874	1894	20
press	1826	1843	17	*breathe*	1874	1988	114
drink	1832	1868	36	*whisper*	1889	1993	104
argue	1833	1892	59	*screech*	1904	1990	86
read	1834	1921	87	*scream*	1913	1932	19
stamp	1835	1922	87	*howl*	1918	1961	43
fret	1835	1902	67	*chant*	1920	1952	32
dream	1849	1962	113	*pry*	1958	1976	18

With an average gap of 86 years between attestation in the two constructions, and given the difference in aspect and telicity between the constructions, it seems implausible for all but a handful of verbs to have arisen in the *way*-construction by constructional contamination with the FRR. This contradicts Mondorf's (2011) claim that the *way*-construction is increasingly encroaching on the functional territory of the FRR.

5.4.2.2 Constructional contamination in the Dutch and German *way*-constructions

As mentioned in Section 2.3.3, Dutch has a construction functionally similar to the *way*-construction known as the Transition to Location Construction (TLC), which has

the formal schema [NP_i [V $REFL_i$ DIR]]. An example of this construction is repeated here for convenience in (217).

(217) Marien zwemt zich in de finale
Marien swims REFL into the final
'Marien swims his way into the final' (van Egmond 2006: 90, ex. 1a).

German has the same construction, an example of which is shown in (218).

(218) Die Titelrolle des Gangsters, der sich
the title.role the.GEN gangster.GEN REL.PRON REFL
nach oben mordet, spielt Martin Wuttke
to up murders plays M. W.
'Martin Wuttke plays the lead role of the gangster who murders his way to the top' (Berliner Zeitung, 1995).

Measuring constructional contamination between these constructions and the Dutch and German *way*-constructions proved impractical. Even limiting my investigation to verbs which occurred in both constructions, there were far too many irrelevant results to filter out manually; this is because the preposition *in* in both languages can mean both ‘in’ and ‘into’. While in German these two meanings are distinguished by putting the following noun into the dative and accusative cases respectively, no such distinction exists in Dutch. My corpus searches therefore returned many instances of reflexive verbs taking place within the bounds of a particular space, as well as genuine TLCs. As such, an investigation into the constructional contamination of the TLC affecting the *way*-construction is beyond the scope of this thesis, but represents fertile ground for future research on constructional contamination in Dutch and German resultative constructions.

5.5 Analogical thinking, analogization, and reanalysis

In cognitive linguistics, there are two models of grammatical change; one reanalysis-based, and the other analogy-based. The analogy-based model is more radically usage-based than the reanalysis model, because it does not involve a constructicon that is external to the speaker, whereas, according to Noël (2017: 73), the reanalysis

model hinges on a mismatch between the speaker's internal constructicon and an external one. However, there is no mismatch, because it is not necessary to posit an external constructicon; neoanalysis involves a speaker assigning a new analysis to an item, which is a process internal to that speaker's constructicon. In the analogy-based model, the speaker replaces one pattern with another by analogy rather than by reanalysis (Fischer 2009: 7). By 'replace', Fischer refers to innovative ways of expressing certain meanings and applying new meanings to old surface forms. She (2007, 2008, 2013) illustrates the former kind of replacement with the development of modal auxiliary *may* as a replacement for *it may be that*, stating that modal auxiliary *may* arose due to analogy with other modals and subject-raising structures with verbs such as *seem*. Likewise, De Smet (2016: 100) argues that expressions increase their collocational range by a series of local analogies determined by similarity relations; this was convincingly shown to be the case for the *way*-construction in English, Dutch, and German. The analogy-based replacement account of constructionalization runs into problems, however, in cases where a brand new structure has been created, such as auxiliaries in English, as there was no existing structure in the language with which speakers could have perceived a similarity. De Smet (2009: 1751) responds to this point by arguing that the first auxiliary could not be analysed as an auxiliary until there was a second auxiliary.

Instead of assigning primacy to reanalysis or analogization (which in turn drives actualization), Barðal and Gildea (2015: 17-18) claim that both reanalysis and analogization are necessary for constructionalization. This point of view is supported by Traugott (2015: 54), who argues that constructionalization arises through a sequence of small-step reanalyses of both form and meaning. The contexts that enable this reanalysis to take place are termed 'critical contexts' (Smirnova 2015: 89). Constructionalization can only occur when there have been morphosyntactic and semantic reanalyses that are shared across a network of speakers (Traugott 2015: 56); this manifests itself when analogical thinking and analogization causes constructs to be attested which could not have been fully sanctioned by pre-existing constructional types (Traugott & Trousdale 2013: 22). This results in the addition of a micro-construction to the network, because a new conventional symbolic unit has been created (Traugott 2015: 56).

The data presented in this chapter and in the previous chapter support the hypotheses presented above; the long series of local analogical extensions in each of the three languages could not have taken place without the neoanalysis of *way*/*weg*. Further, tokens such as *He whistled his way home* (and its Dutch and German equivalents) could not have been sanctioned by pre-existing constructional types, in which *way*/*weg* functioned as a referential object, because a literal path is not something that can be whistled.

Trousdale (2013: 9) argues that after constructionalization has taken place, each subsequent constructional change is said to be the result of neoanalysis. Smirnova (2015: 93), on the other hand, hypothesises that each subsequent constructional change is the result of step by step analogical extensions of the material to more and more contexts. The data presented in this chapter favour both Smirnova's and Trousdale's accounts; after *way*/*weg* was neoanalysed as a non-referential object in the three languages, the *way*-construction underwent a long series of local analogical extensions, and when new verbs were used in the construction, the speaker assigned a new analysis to these verbs, i.e. that they could appear with a possessive pronoun and the noun *way*, in the case of English, or with a reflexive pronoun and the noun *weg* in the case of German and Dutch.

In sum, the data presented in this chapter have led me to conclude that neither neoanalysis or analogy was primary in the development of the English, Dutch, and German *way*-constructions. Instead, I propose that the reanalysis of *way*/*weg* as a non-referential object facilitated the long series of local analogical extensions within the construction in each language. The data in each of the three languages suggest that analogization and reanalysis therefore both play key roles in language change, which supports Fischer's (2007: 123) claim that analogy and reanalysis work in tandem.

5.6 Conclusions

This chapter of the thesis has followed Traugott and Trousdale (2013) in distinguishing between analogical thinking and analogization, because it is useful to distinguish between the motivation for change and the mechanism of change. Further, not all instances of analogical thinking will lead to change. It was shown in this chapter

that the development of the *way*-construction in all three languages can be considered a long series of local analogical extensions, as Israel (1996) claims of the English construction, and that this series of analogical extensions was triggered by the neoanalysis of *way*/*weg* as a non-referential object. Some hypotheses concerning the relationship between analogy, frequency, and productivity were tested in Section 5.3. It was found that low semantic type frequency is not necessary for analogization to take place, and neither is high token frequency (contra Barðdal 2008). Corpus data from the three languages confirmed Israel's (1996) hypothesis that the likelihood of an item undergoing analogical extension is dependent on the number of similar tokens previously attested in the corpus. The corpus data also convincingly confirmed Zeschel's (2010) hypothesis that the distribution of novel verbs in each semantic domain is not random, but that it correlates positively with the number of established verbs in all semantic domains; this demonstrates that there are frequency effects in the development of the *way*-construction in the three languages, because the analogical extensions that took place correlated strongly with the type frequency of verbs in the construction. In Section 5.4, I showed that the notions of constructional contamination and bolstering can be integrated into a construction-based theory of analogy, because these processes involve analogical thinking and analogization between constructions. I also tested the hypotheses that the FRR is contaminating the *way*-construction, and that the *way*-construction is encroaching on the functional territory of the FRR, as Mondorf (2011) claims; neither hypothesis was borne out by my data. It can therefore be concluded that the parsing of *way* as a non-referential object in the construction was an instance of neoanalysis, rather than analogization with another resultative construction. The following chapter deals with frequency effects in language, and examines whether some of the changes to the *way*-construction in English, Dutch, and German can be explained in terms of frequency effects. I also discuss exemplar and prototype models of language and how they can be applied to the *way*-construction in each language.

6. Frequency effects, entrenchment, and exemplar representations of language in the development of the *way*-construction in English, Dutch, and German

6.1 Introduction

The present chapter of the thesis answers the fifth of my research questions: is there evidence for frequency effects and exemplar representation of language in the *way*-construction in English, Dutch, and German? I show that the items that can occur in the slot of a construction can be grouped into exemplar clouds of semantically similar items, as Bybee (2013) claims. In this chapter I also explore whether there may be genre effects in the *way*-construction in the three languages. The remainder of the chapter is structured as follows. Section 6.2 deals with exemplar and prototype models of language and discusses their application to the *way*-construction in the three languages. I show that an exemplar model is very useful in describing the construction, because the verbs that can occur can be divided into exemplar clouds of semantically related items. In Section 6.3, I evaluate the hypothesis that factors other than frequency may affect the development of a construction. In order to test this hypothesis, I perform statistical analyses on my dataset that depend on variables other than frequency, such as genre, and an item's dispersion in the corpus (cf. Hilpert & Correia Saavedra 2017). I show in this section that factors other than frequency have played a role in the development of the English *way*-construction, because genre effects are present.

6.2 Exemplars, prototypes, and frequency

6.2.1 Literature Review

Cognitive linguists to date have employed two main approaches to the representation of categories: one approach is exemplar-based, and the other is prototype-based. The prototype model states that categories are represented by a weighted set of features abstracted from specific instances of a concept; for instance, the concept *fruit* contains information about what the average fruit is like. Prototype effects arise from graded category membership in which some members of the category are considered more central than others (Bybee 2010b: 18); prototype effects have been shown to be pervasive in language (see e.g. Lakoff 1987, Taylor 1995). Indeed, Medin and Schaffer (1978) show that similarity even to a marginal member of a category facilitates category assignment. For example, if a person is familiar with an ostrich and has assigned to it the category 'bird', the assignment of an emu to the same category is facilitated, even though an emu is very far from the prototypical bird (Bybee 2010b: 19). A prototype model does not store the exemplars upon which the categorisation is based (Taylor 1995: 59ff.). In an exemplar model, on the other hand, an item a speaker hears is mapped onto an existing stored exemplar, or a new exemplar cloud is created if there is no similarity between the heard item and any of the stored exemplars (Bybee & Eddington 2006: 326). An exemplar model provides a more accurate representation of the *way*-construction in the three languages, because, as the previous chapter demonstrated, speakers analogically extended the collocational range of the construction on the basis of very fine-grained categories; a prototype model cannot account for this.

The exemplar view states that a category is represented by instances that have been previously encountered (Voorspoels et al. 2008: 630). The concept of *fruit* in this view would contain information about the instances of fruit that a person has experienced. In both the exemplar and prototype model, the observer classifies items based on their similarity to the central member of the category, i.e. the prototype (Nosofsky 1992: 149; Taylor 2015: 565). However, the concept of 'prototype' has been criticised for not being a unified concept, and because it can be understood in different ways with respect to different categories (Taylor 2015: 262); 'prototype' is itself a prototype concept (Geeraerts 1987: 592). Results from the majority of psycholinguistic studies

favour the exemplar view (cf. the references in Voorspoels et al. 2008: 630; but see especially Nosofsky 1992). For instance, people know that if a bird sings, it is more likely to be a small bird than a large bird (Malt & Smith 1984). This knowledge would not be possible if people represented their knowledge of birds through an abstract "bird" prototype, discarding individual exemplars (Bybee & Beckner 2010: 832), and the analogical extensions that took place in the *way*-construction in the three languages also would not have been possible.

Exemplar theory was first applied to phonological categories (Bybee 2002; Pierrehumbert 2003). Bybee (2006) extends this, and proposes an account of all aspects of language in terms of exemplar representation. In this theory, every token of linguistic experience is classified and placed in a network (Bybee 2006: 716). Each token is stored in a highly detailed fashion, down to minute details of sound (Bybee 2010b: 14; Hilpert 2015: 348). As well as phonetic detail, exemplar representations contain the lexical items and constructions used in an utterance, the semantics and pragmatics of the utterance, and its wider context (Bybee 2010b: 14). This model of language is markedly different from the generativist approach, which states that redundant information is discarded, not stored in memory (Bybee 2010b: 15). This also differs from the prototype view, which allows the prototype to be defined in terms of features, rather than a particular instance of that category (Rosch & Mervis 1975). Exemplars are stored in clouds, which are categories that exhibit prototype effects, and are organised in terms of members that are more or less central to the category (Bybee 2006: 717). Exemplar clouds may also be pluricentric, i.e. have more than one central member (Hilpert 2015: 349); as the following subsection will demonstrate, this is true of the *way*-construction in the three languages.

Although considerable detail is represented according to the exemplar model, language users do not encode every single token of linguistic experience (Pierrehumbert 2006: 525). Rather, a token of linguistic experience that is identical to an existing exemplar is mapped onto that exemplar, strengthening it (ibid.). Exemplar clouds can be formed on phonetic, semantic, pragmatic, or contextual grounds (Bybee 2013: 54). A number of marginal items in the original exemplar cloud may become frequent enough to constitute their own exemplar cloud, and these two exemplar clouds may overlap at first (Hilpert 2015: 349); as I will demonstrate later, this is true of the *way*-construction in the three languages. A new exemplar cloud can

also be formed by analogy with previously experienced exemplars of a construction (Bybee 2010b: 26); as I showed in Chapter 5, this is true of the *way*-construction in English, Dutch, and German.

Modelling language in terms of exemplars allows both type and token frequency to be represented (Bybee 2013: 52). These frequency patterns are important for understanding the categories that are formed for the slots in constructional schemas (Bybee 2013: 53). For instance, it has often been observed that the set of lexical items that can occupy a slot of a construction consists of two or more clusters of closely related items (Bybee 2013: 58; see also Goldberg 1995, Israel 1996). This is certainly the case of the *way*-construction in the three languages, where the number of clusters of semantically related items is very large. These exemplar clusters are the basis of new extension of the construction (Bybee 2013: 58); I show in Section 6.2.2 that this hypothesis holds of the *way*-construction in each language.

Bybee and Eddington's (2006) study supports the notion that the set of items in a construction can be grouped into exemplar clouds. Their study concerns four Spanish reflexive verbs that express a change of state when used with adjectives and an animate subject: *ponerse* 'to put', *volverse* 'to turn', *quedarse* 'to remain', and *hacerse* 'to make' (2006: 323). They found in their study that the adjectives occurring in the slot after the verb could be grouped into clusters of semantically similar exemplars. For instance, they found that the adjectives occurring in the *quedarse* + ADJ construction could be divided into clusters of adjectives relating to being alone, to being surprised, to feelings of sadness or nostalgia, and to adjectives encoding a physical state (Bybee & Eddington 2006: 337-339).

Bybee and Eddington hypothesise (2006) that the central member of the exemplar cloud is the most frequently occurring one. To test the relation between frequency and centrality, Bybee and Eddington (2006: 349-350) devised an experiment where native speakers were asked to rate how natural-sounding an expression was. For the *quedarse* + ADJ and *ponerse* + ADJ constructions, they tested participants on the high-frequency members which were hypothesised to be central to the exemplar cloud, low-frequency items semantically related to one of the central high-frequency members, and low-frequency items that were not semantically related to the high-frequency ones (Bybee & Eddington 2006: 350). Bybee and Eddington (2006: 351)

found that the participants were more likely to judge the high-frequency items and the low-frequency items semantically related to these as acceptable. The low-frequency items not semantically related to the high-frequency items were judged to be much less acceptable than the ones that were (ibid.).

However, the participants in Bybee and Eddington's (2006) study were asked to rate items on a Likert scale from 1 to 5, which is a crude method, because these scales are too restrictive in their values (Bard et al. 1996: 41). Magnitude estimation studies are preferred in studies on acceptability judgements, because they allow informants to make very fine-grained judgements, differentiating as many levels of acceptability as they deem necessary (Hoffmann 2013: 105). Further, Taylor (2012: 187) states that the **statistical** centre of the exemplar cloud is the central member; raw token frequency is a very crude statistical measure. It may be the case that the most frequent member of the exemplar cloud is not the one that is most strongly attracted to the construction; I test whether this is true of the exemplar clouds in the *way*-construction in the following subsection. Because most of the exemplar clouds found in the German data have very low type and token frequency, I do not test Bybee and Eddington's (2006) hypothesis relating to frequency and centrality on the German data.

Exemplar clouds are maintained by both frequency and recency (Bybee 2010b: 28). Similarly to memory of non-linguistic objects, marginal exemplars that are not reinforced may be lost (Bybee & Beckner 2010: 833). Experimental data from Caldwell-Harris et al. (2012) have shown that there is no minimum threshold for the storage of exemplars; this is supported by data on the *way*-construction in the three languages, because clusters of semantically related verbs emerged even when the type and token frequency of the construction was very low.

Bybee (2013) argues that an exemplar model is essential to explaining how constructions acquire idiosyncrasies. In a model of language where semantic representations only contain an abstract meaning, it is impossible for an implication or inference to become associated with a construction. However, an exemplar model would record the inferences made in each instance of use, and if the same inference is made on multiple occasions, the strength of that inference increases (Bybee 2013: 56). This is true of the *What's X doing Y* construction (Kay & Fillmore 1999), which

had its origin in conversational implicatures. The hypothesis that an exemplar model is required to explain the idiosyncrasies of the *way*-construction in English, Dutch, and German will be tested in the following subsection.

The following subsection also tests the hypothesis that the diachrony of the *way*-construction in English, Dutch, and German can be explained in terms of clusters of semantically related exemplars, with marginal members of the exemplar cloud serving as a locus of analogical change for the formation of a new exemplar cloud. In this subsection, I also explore the relation between the exemplar cloud and prototypes of the construction in each language, and how the collocational preferences of the construction have changed over time. Section 6.3.2.1 below answers these questions by examining data from two corpora; the CLMET3.1 corpus, covering the period 1720-1920, and COHA, from which I use data from the period 1920-2000. I use this combination of corpora in order to cover the period where the increase in type and token frequency of the construction was most prevalent.

The exemplar clouds in the following sections have been classified on the basis of the interaction between the verb and the construction; e.g. *make* was classified as a basic motion verb, because when it is used in the *way*-construction the meaning is of basic motion. The classification was performed on a token-by-token basis.

6.2.2 Results

6.2.2.1 English

Figure 21 below shows the exemplar clouds found in the CLMET3.1 corpus. Where an exemplar cloud has an extremely high type frequency, only the ten most frequent types are shown. Some verbs could be assigned to more than one category; however, for ease of exposition, I only assign them to one category in the exemplar clouds shown in this section.

Figure 21: Exemplar clouds in CLMET3.1

Cloud	Items
basic motion	*MAKE* (510), *go* (82), *walk* (2)
vehicular motion	*steer* (4), *drive* (2), *pilot* (1), *ride* (1)
manner of motion (6 not shown)	*WING* (16), *worm* (5), *plod* (3), *edge* (3), *writhe* (2) *flap* (2), *toddle* (1), *squirm* (1), *foot* (1)
finding a path	*FIND* (390), *explore* (4)
path shape	*WIND* (35), *wend* (33), *thread* (23), bend (17), *twine* (1) *screw* (1), *bounce* (1)
force dynamics (28 not shown)	*FORCE* (142), *pick* (45), *push* (42), cut (34), *elbow* (12) *break* (7), *cleave* (6), *burst* (5), *battle* (3), *poke* (3)
taking a path (5 not shown)	*PURSUE* (77), *take* (66), *grope* (36), *trace* (7), *measure* (2) *cross* (1), *mine* (1), *follow* (1), *rake* (1), *delve* (1)
laborious motion	*WORK* (72), *plough* (9), *plod* (3), *drag* (2), *toil* (1) *wear* (1), *weather* (1)
actions w/ mouth	*EAT* (12), *nibble* (1), *bite* (1), *devour* (1)
commerce/finance	*pay* (3)
deceit	*steal* (2)
terrain of motion	*wade* (1), *bob* (1)
sound emission	*bark* (1), *splash* (1), *gasp* (1), *plash* (1), *purr* (1)
performance	*croon* (1)
cognition	*worry* (1)

These findings support Bybee and Eddington's (2006) hypothesis that items with high token frequency are central members of the construction; *make*, a highly general verb, is the most frequently attested verb and is used to connote basic motion along a path, a central sense of the construction. *Force* is also highly frequent, as are other verbs encoding force; the removal of obstacles on a path (usually by force) is another central sense of the construction. The findings shown above also support Bybee's (2013: 56) hypothesis that an exemplar model is necessary to explain the idiosyncrasies of the construction. A prototype analysis cannot explain why verbs of sound emission, e.g. *bark*, are now attested, because the verbs of sound emission in this cluster are very far from being prototypical of the construction. Further, in all exemplar clouds where the most frequently occurring member had a token frequency greater than 10, the most frequent member of the cloud was found to be the most strongly attracted to the construction, which supports Bybee and Eddington's (2006) claim that these are central members.

The following table zooms in on the CLMET3.1 data and examines how the collocational preferences of the construction changed between 1720 and 1920. To achieve this, a diachronic distinctive collexeme analysis was used, the results of which are shown in Table 33 below. A diachronic distinctive collexeme analysis ignores items that are equally frequent in all periods, instead focusing on items that are unusually frequent in a particular period and therefore characteristic of that period.

Table 33: Diachronic distinctive collexeme analysis of verbs occurring in the way-*construction in CLMET3.1*

Verb	Typical of?	Pbin	Sig. at *p* <
make	1720-1750	9.1545	0.001
take	1780-1885	2.5873	0.01
go	1780-1885	1.4166	0.05
feel	1900-1915	2.9111	0.01
storm	1900-1915	1.9833	0.05
work	1900-1915	1.8423	0.05
thread	1900-1915	1.4221	0.05

This shows that *make* is both very frequent in and typical of 1720-1750; another verb of basic motion, *go*, was found to be typical of 1780-1885. The findings above suggest that a cluster of verbs encoding laborious motion (*work*, *thread*) began to be preferred in 1900-1915. These findings also show that the most frequent item may not be the most typical one in a period; *take* is less frequent than *pursue* in the construction between 1780 and 1885, but the former was found to be typical of that period whereas the latter was not.

The following paragraphs discuss the data from the 1920-2000 portion of COHA. Figure 22 shows the exemplar clouds in COHA between 1920 and 2000. In comparison to the CLMET3.1 data, an additional two exemplar clouds have emerged; one containing verbs encoding social interaction, and one containing verbs denoting bodily functions. I hypothesised above that verbs of social interaction entered the construction by analogical extension of *beg*; this supports Bybee's (2013: 56) hypothesis that marginal members of an exemplar cloud extend into new exemplar clouds.

Figure 22: Exemplar clouds in COHA from 1920-2000

Cloud	Items
basic motion	*MAKE* (2656), *go* (61), *walk* (7), *run* (5)
vehicular motion	*DRIVE* (10), *steer* (3), *pilot* (3), *sail* (2), *row* (1) *fly* (1), *wheel* (1), *pole* (1)
manner of motion (60 not shown)	*WORM* (112), *inch* (78), *edge* (53), *nose* (24), *ease* (19) *thrust* (17), *bull* (16), *twist* (11), *crash* (4), *writhe* (4)
finding a path	*FIND* (1165), *seek* (3), *search* (3), *explore* (1)
path shape (12 not shown)	*THREAD* (204), *wind* (114), *wend* (77), *weave* (59), *wing* (46) *snake* (17), *sweep* (6), *zigzag* (5), *climb* (4), *bend* (1)
force (48 not shown)	*FIGHT* (427), *force* (306), *cut* (66), *elbow* (63), *claw* (42) *beat* (38), *shoot* (34), *battle* (33), *hack* (32), *hew* (7)
taking/clearing a path (44 not shown)	*PUSH* (361), *feel* (254), *grope* (165), shoulder (60), *take* (45), *plough* (39), *tear* (21), *break* (21), *pursue* (19), *retrace* (13)
laborious motion (2 not shown)	*WORK* (798), *pick* (390), *crowd* (11), *slog* (8) *toil* (5), *churn* (5), *struggle* (3), *wear* (3), *slug* (3)
actions w/ mouth (4 not shown)	*EAT* (71), *gnaw* (7), *drink* (10), *bite* (9), *chew* (9) *kiss* (5), *munch* (5), *chug* (5), *nibble* (3)
commerce/finance (2 not shown)	*PAY* (85), *buy* (59), *bribe* (12), *negotiate* (11), *spend* (8) *borrow* (3), *purchase* (1), *advertise* (1), *finance* (1)
deceit/misconduct (16 not shown)	*LIE* (20), *bluff* (19), *bully* (8), *steal* (7), *sleep* (5) *bargain* (5), *con* (3), *cheat* (3), *wangle* (3), *charm* (3)
terrain of motion	*TREAD* (3), *wade* (2), *ripple* (2), *swim* (1), *flow* (1) *paddle* (1)
sound emission (67 not shown)	*TALK* (57), *puff* (10), *laugh* (7), *roar* (7), *bang* (6) *wheeze* (6), *whistle* (5), *crunch* (5), *yawn* (5), *tap* (5)
achievement	*WIN* (63), *earn* (29)
incompetence	*FUMBLE* (16), *bum* (5), *blunder* (4), *sleepwalk* (2) *bungle* (1), *flounder* (1)
olfaction	*smell* (4), *sniff* (3)
lighting a path	*LIGHT* (14)
light emission	*flash* (2), *flicker* (2), *shine* (1)
change of state	*melt* (4), *die* (1)
performance	*DANCE* (15), *sing* (10), *play* (6), *act* (2), *boogie* (1) *singsong* (1)
cognition	*THINK* (24), *learn* (11), *reason* (9), *dream* (4), *worry* (2) *puzzle* (2), *imagine* (1)
persuasion	*argue* (4), *grovel* (1), *strongarm* (1), *reargue* (1), *flatter* (1), *squabble* (1)
social interaction	*joke* (6), *bluster* (4), *chatter* (3), *hug* (2), *chat* (2), *complain* (1)
bodily functions	*sweat* (5), *sniffle* (1), *vomit* (1)

A comparison between the exemplar clouds in the CLMET3.1 data and the ones shown in Figure 22 above shows that *make*, *force*, and *find* continue to be highly frequent and central members of the construction, as Bybee and Eddington's (2006) hypothesis suggests. This hypothesis is supported by the results of the collostructional analysis of each exemplar cloud; the most frequent member of an

exemplar cloud was found to be the one most strongly attracted to the construction in all cases. Most of the exemplar clouds have also undergone a considerable increase in type frequency. The collocational preferences of the construction have changed slightly; *push* and *fight* are extremely highly frequent in COHA but were only attested twice and once respectively in CLMET3.1.

The most frequently occurring type of verb of vehicular motion in the CLMET3.1 data, *steer*, has licensed an extension to verbs relating to controlling other types of vehicle, with *fly*, *sail*, and *row* later being attested in the construction. This supports Bybee and Eddington's (2006) hypothesis that analogical extension within exemplar clouds is based on the most frequent member of these clouds. This hypothesis is also upheld in the case of verbs encoding the manner of motion; the most frequent such verb in the CLMET3.1 data, *wind*, has served as a locus for analogical extension to *snake*. Further, the most frequently occurring verb encoding speed of motion in that corpus, *edge*, is the basis of an extension to *inch*.

The most frequently occurring verbs encoding path shape in the CLMET3.1 corpus usually encode winding motion (*wend*, *wind*, *bend*), and it is these verbs that have licensed the extension of the *way*-construction to *zig-zag*. The COHA data also shows that, in the exemplar cloud of verbs encoding force, the most frequent such verb in CLMET3.1 (*force*) has licensed extension to other such verb, such that *force* is no longer the most frequently attested verb encoding force. These extensions support Bybee and Eddington's (2006) hypothesis that category-internal analogical extensions are based on the most frequent members of the category. Further evidence supporting this hypothesis can be found in the exemplar cloud of laborious activity verbs, with more specific types of *work* (the most frequent such verb in CLMET3.1) being attested in the construction in COHA up to 2000.

The exemplar cloud of actions involving the mouth now also exhibits pluricentricity. The verb *eat* was the most frequent such verb in CLMET3.1, and this verb has licensed the extension of the *way*-construction to *munch*. The most frequently occurring verb of commerce and finance in CLMET3.1, *pay*, has licensed the extension of the construction to more specific verbs, *finance* and *purchase*. This does not hold true, however, of the exemplar cloud containing verbs of deceit. *Lie*, which was not the most frequently attested verb in CLMET3.1, has led to *cheat*, *con, wangle*,

and *bullshit* entering the construction; the most frequently attested verb, *steal*, is a less plausible source for this extension. Concerning the exemplar cloud of verbs relating to the terrain of motion, the verbs that enter the construction after 1920 are extensions of *wade* and *bob*, and encode motion through water (*ripple*, *swim*, *flow*). The pluricentric exemplar cloud of sound emission verbs undergoes a substantial increase in type frequency after 1920; verbs entering the construction after this period are the result of analogy with the existing verbs of human speech or sound (*talk*, *cheer*, *laugh*, *whistle*, *yawn*), and other miscellaneous sounds. Verbs of light emission also undergo analogical extension with *flash*, with *flicker* and *shine* entering the construction before 2000. Finally, the verbs that enter the exemplar cloud of verbs encoding incompetence (*blunder*, *bungle*) are more specific types of *mistake*, but this was not the most frequently attested verb in CLMET3.1. In sum, however, it was found that the most frequently attested verb was usually the source of an extension of an exemplar cloud in the *way*-construction.

Table 34 below shows the results of the diachronic distinctive collexeme analysis of the COHA dataset. For the purposes of the analysis, I ignored the VNC dendrogram and divided the corpus into three equal periods in order to obtain more datapoints: 1820-1880, 1880-1940, and 1940-2000. Because the first two periods largely overlap with the period of the CLMET3.1 corpus, I only show the most typical collexemes of 1940-2000 here. 55 verbs were found to be typical of 1940-2000; Table 34 shows only the 20 most typical of these.

Table 34: Verbs in the way*-construction in COHA typical of 1940-2000*

Verb	Typical of?	Pbin	Sig. at $p <$	Verb	Typical of?	Pbin	Sig. at $p <$
work	1940-2000	29.4073	0.001	*hack*	1940-2000	4.7471	0.001
inch	1940-2000	22.4341	0.001	*navigate*	1940-2000	4.5201	0.001
weave	1940-2000	12.1959	0.001	*pound*	1940-2000	4.1092	0.001
talk	1940-2000	11.4132	0.001	*punch*	1940-2000	3.7084	0.001
claw	1940-2000	10.0030	0.001	*chew*	1940-2000	3.6983	0.001
buy	1940-2000	7.6175	0.001	*bluff*	1940-2000	3.6763	0.001
bull	1940-2000	6.5748	0.001	*manoeuvre*	1940-2000	3.6320	0.001
ease	1940-2000	6.3170	0.001	*muscle*	1940-2000	3.6320	0.001
shove	1940-2000	4.8544	0.001	*snake*	1940-2000	3.5047	0.001
battle	1940-2000	4.7471	0.001	*slog*	1940-2000	3.2874	0.001

These findings reveal a change in the collocational preferences of the construction compared with the CLMET3.1 data, because verbs of deceit and manipulation (*bluff*, *manoeuvre*) are now typical of the construction. Many verbs encoding force were also found to be typical of this period (*claw*, *bull*, *shove*, *battle*, *hack*, *pound*, *punch*, *muscle*); this matches the increase in token frequency of the force dynamic verbs in the latter part of COHA. Verbs of laborious motion were also found to be typical of this period (*work*, *slog*), as well as verbs encoding the path shape (*weave*, *snake*). This suggests a mismatch between typicality and frequency; although these verbs were highly frequent before this period, they did not become typical of the construction until 1940.

In sum, corpus data on the English *way*-construction convincingly support the hypothesis that the items that can fill the slot of a construction can be grouped into exemplar clouds of semantically similar verbs; several such clusters were identified in the English dataset. The complexity of the *way*-construction would not be captured by a prototype model, unless it was posited that each semantic cluster had its own prototype; however, even this would not fully account for the developments described above. Even if 'prototype' is erroneously understood to be the most frequently occurring item, as is common in many corpus approaches to language (Gilquin 2006), this would lead to the conclusion that *make*, *find*, and *go* were central to the development of the *way*-construction, which is not the case, because this does not account for the development of the incidental activity reading, which is now very productive. If 'prototype' is understood to mean the most entrenched item, using the *p*-value of the Fisher Yates exact test as an approximation to this[43], this would imply that *make*, *find*, *force*, and *wend* were key to the development of the *way*-construction. It is certainly true that a very high number of force dynamic verbs arose by analogization with *force*, and other motion verbs by analogical extension of *go*, this does not fully account for the complexity of the construction today. Bybee's (2013) hypothesis that an exemplar model is required to explain the idiosyncrasies of a construction is therefore a better explanation of the data than a prototype analysis; a prototype analysis does not explain how the construction now sanctions several verbs that are not related to motion at all, because the prototypical verbs in the construction necessarily connote motion, as the meaning of the construction is '$SUBJ_i$ move along a path OBL by/while V'. It was also found that exemplar clouds can be pluricentric, as

[43] This is dealt with in more detail in Section 6.5.

Hilpert (2015: 349) claims; several exemplar clouds presented in this subsection exhibited pluricentricity, i.e. had more than one central member.

6.2.2.2 Dutch

Only one verb is attested in the Dutch *way*-construction in *De Gids* between 1840 and 1915; *banen* 'make'. By 1915, an exemplar cloud of verbs encoding the exertion of force on an entity emerges: *openen* 'open', *graven* 'dig', *kappen* 'cut', *houwen* 'hack', and *breken* 'break'. *Maken* 'make' also joins the construction at around this time. No new verbs are attested in *De Gids* after this period. In the SoNaR corpus, exemplar clouds similar to those of the English *way*-construction emerge. These are shown in Figure 23.

Figure 23: Exemplar clouds in the SoNaR corpus

Cloud	Items
basic motion	*BANEN* 'make' (534)
vehicular motion	*fietsen* 'cycle' (1)
manner of motion	*wurmen* 'worm' (7), *kronkelen* 'writhe' (4), *slalomen* 'slalom' (3) *slingeren* 'swing' (3), *spinnen* 'spin' (1), *bewandelen* 'wander' (1) *wriemelen 'wriggle' (1)*
finding a path	*vinden* 'find' (36), *uitstippelen* 'map out' (2), *terugvinden* 'find back' (1) *navigeren* 'navigate' (1), *fouilleren* 'search' (1)
force dynamics (21 not shown)	*VECHTEN* 'fight' (26), *schieten* 'shoot' (7), *boren* 'bore' (7) *knokken 'knock' (7), kappen* 'cut' (6), *hakken* 'hack' (5), *slaan* 'slay' (4), *openen 'open' (3), wroeten 'root' (2), moorden 'murder' (2)*
taking/clearing a path	*nemen* 'take' (4), *effenen* 'smooth' (1) *tekenen 'design' (1), verschaffen 'provide' (1)*
laborious motion	*ploegen* 'plough' (3), *ploeteren* 'plod' (2), *werken* 'work' (1), *zwoegen* 'toil' (1)
actions w/ mouth	*VRETEN* 'eat' (16), *eten* 'eat' (4), *bijten* 'bite' (3), *likken* 'lick' (1) *zoenen* 'kiss' (1), *spitten* 'spit' (1), *drinken* 'drink' (1), *snoepen* 'snack' (1)
commerce/finance	*kopen* 'buy' (1)
deceit	*bluffen* 'bluff' (3), *manipuleren* 'manipulate' (1)
sound emission	*toeteren* 'toot' (2), *claxonneren* 'claxon' (1), *hoesten* 'cough' (1) *zuchten 'sigh' (1), jodelen 'yodel' (1), brullen* 'roar' (1), *janken* 'whine' (1) *schreeuwen 'scream' (1), huilen 'howl' (1), trommelen 'drum' (1)*
performance	*dansen* 'dance' (3), *kniezwengelen* 'knee dance' (1), *zingen* 'sing' (1)
social interaction	*flirten* 'flirt' (1)
incompetence	*blunderen* 'blunder' (1)
sports/games	*dribbelen* 'dribble' (4), *puzzelen* 'puzzle' (2), *boksen* 'box' (2), *racen* 'race' (1)
bodily functions	*plassen* 'pee' (1)

This represents a considerable increase in type and token frequency compared with the *De Gids* data, where only verbs of motion and verbs encoding force were attested. The findings above support Bybee and Eddington's (2006) hypothesis that the most frequent member of the exemplar cloud is the most central; *banen* is a highly general motion verb and this sense of motion is central to the construction, as is a sense of fighting one's way through obstacles, which explains the high token frequency of *vechten* 'fight'. Both *banen* and *vechten* were found to be the most strongly attracted member of their exemplar clouds to the construction. A cluster of verbs encoding actions involving the mouth has emerged; this cluster is relatively high in both token and type frequency. This cluster has *vreten* 'eat' as its central member; its centrality is supported both by the fact it has the highest token frequency in the cluster, and that it is the most strongly attracted to the construction, as the collostructional analysis of this cluster showed. Figure 24 shows the exemplar clouds in the NLCOW data.

Figure 24: Exemplar clouds in NLCOW (continued overleaf)

Cloud	Items
basic motion	*BANEN* 'make' (4872), *gaan* 'go' (92), *maken* 'make' (20)
vehicular motion	*rijden* 'ride' (6), *peddelen* 'pedal' (3), *fietsen* 'cycle' (2)
	berijden 'ride' (1), *inrijden* 'ride in' (1)
	oprijden 'ride up' (1), *heenvliegen* 'fly to' (1)
manner of motion (38 not shown)	*SLINGEREN* 'swing' (82), *kronkelen* 'writhe' (60), *wurmen* 'worm' (36)
	bewandelen 'wander' (27), *manoeuvreren* 'manouevre' (12)
	stuntelen 'bumble' (8), *benen* 'leg' (8), *meanderen* 'meander' (7)
	strompelen 'stumble' (6), *draaien* 'turn' (4)
finding a path	*ZOEKEN* 'seek' (602), *vinden* 'find' (308), *uitstippelen* 'trace out' (10)
	navigeren 'navigate' (4), *voelen* 'feel' (2)
path shape	*KRUISEN* 'cross' (18), *slalomen* 'slalom' (17), *oversteken* 'cross over' (17)
	klimmen 'climb' (7), *springen* 'jump' (6), *opgaan* 'go up' (3)
force dynamics (56 not shown)	*VECHTEN* 'fight' (401), *slaan* 'slay' (117), *boren* 'bore' (87)
	worstelen 'wrestle' (67), *hakken* 'hack' (57), *ploegen* 'plough' (49)
	schieten 'shoot' (47), *knokken* 'knock' (38), *wriegen* 'wring' (27)
	dringen 'push' (20)
taking/clearing a path	*NEMEN* 'take' (22), *tekenen* 'trace' (8), *aftekenen* 'trace off' (2)
	plaveien 'pave' (1), *verleggen* 'shift' (1), *snoeien* 'shear' (1)
	vegen 'sweep' (1)
laborious motion	*WERKEN* 'work' (111), *ploeteren* 'plough' (30)
	zwoegen 'toil' (8), *slepen* 'drag' (2), *banjeren* 'swagger' (1)
actions w/ mouth (9 not shown)	*VRETEN* 'eat' (127), *eten* 'eat' (59), *likken* 'lick' (15), *bijten* 'bite' (11)
	knagen 'gnaw' (11), *kussen* 'kiss' (11), *zoenen* 'kiss' (5), *kauwen* 'chew' (3)
	zuigen 'suck' (3), *happen* 'bite' (2)
commerce/finance	*kopen* 'buy' (10), *onderhandelen* 'negotiate' (1)
deceit	*BLUFFEN* 'bluff' (11), *stelen* 'steal' (2), *hacken* 'hack (a computer)' (2)

	handbalverdedigen 'defend by handball' (1), *bedriegen* 'deceive' (1)
terrain of motion	*waden* 'wade' (9), *zwemmen* 'swim' (3), *glijden* 'glide' (2)
	walljumpen 'walljump' (1), *stromen* 'flow' (1)
sound emission (31 not shown)	*kreunen* 'groan' (7), *toeteren* 'toot' (4), *drummen* 'drum' (4)
	fluisteren 'whisper' (3), *janken* 'whine' (3), *krijsen* 'shriek' (3)
	schreeuwen 'scream' (3), *ritselen* 'rustle' (2), *brullen* 'roar' (2)
	grommen 'growl' (1)
performance	*ZINGEN* 'sing' (26), *dansen* 'dance' (13), *improviseren* (3)
	vuvuzelaen 'play vuvuzela' (1)
	headbangen 'headbang' (1), *croonen* 'croon' (1), *walsen* 'waltz' (1)
	honky-tonken 'play the honky-tonk' (1), *harmoniseren* 'harmonise' (1)
social interaction	*flirten* 'flirt' (1), *flaneren* 'hang out' (1), *keuvelen* 'chat' (1)
cognition	*WANEN* 'believe falsely' (28), *dromen* 'dream' (3)
	fantaseren 'fantasise' (1), *leren* 'learn' (1)
	redeneren 'reason' (1), *inbeelden* 'imagine' (1), *herinneren* 'remember' (1)
	denken 'think' (1), *verbeelden* 'imagine' (1), *filosoferen* 'philosophise' (1)
incompetence	*spartelen* 'struggle' (3), *botsen* 'crash' (2), *frommelen* 'fumble' (2)
	blunderen 'blunder' (1), *baggeren* 'flounder' (1)
olfaction	*ruiken* 'smell' (1)
sports/games (9 not shown)	*PUZZELEN* 'puzzle' (15), *spelen* 'play' (11), *dribbelen* 'dribble' (8)
	golfen 'golf' (2), *quizzen* 'quiz' (2), *voetballen* 'play football' (2)
	dobbelen 'gamble' (1), *pokeren* 'play poker' (1), *coachen* 'coach' (1)
bodily functions	*zweten* 'sweat' (4), *braken* 'vomit' (1)
light emission	*bikkelen* 'flicker' (2), *flakkeren* 'flare' (1), *glunderen* 'shine' (1)
change of state	*smelten* 'melt' (2), *groeien* 'grow' (1)
surfing the internet	*KLIKKEN* 'click' (14), *googelen* 'google' (2), *surfen* 'surf' (2)
	dubbelklikken 'double click' (1), *printen* (1), *bloggen* 'blog' (1)
	skypen 'skype' (1)
building a path	*AANLEGGEN* 'construct' (15), *bouwen* 'build' (5), *vormen* 'form' (5)
	creeren 'create' (4), *scheppen* 'create' (1)
persuasion	*lobbyen* 'lobby', *flikflooien* 'flatter' (1), *charmeren* 'charm' (1)
sex	*neuken* 'fuck' (3), *swaffelen* 'hit with penis' (1), *vingeren* 'finger' (1)
	fikken 'fuck' (1), *strelen* 'caress' (1)

Data from NLCOW and SoNaR show convincingly that the verbs that can occur in the Dutch *way*-construction can be grouped into clusters of semantically similar verbs, as Bybee (2013) claims. The exemplar clouds of the Dutch construction are very similar to those of the English construction, with verbs of olfaction, bodily functions, persuasion, and social interaction now attested. There is a cluster of verbs in the Dutch construction of verbs relating to surfing the internet that is not attested in the English construction; the analogical source of these verbs is not clear. A prototype analysis would not account for the complexity of the Dutch *way*-construction. As in the English construction, this is because the prototypical verbs in the construction necessarily encode motion, because the meaning of the construction is one of motion along a path. The prototypical verb in the construction is certainly *banen*, because it

is the most frequently occurring verb and *banen* is a highly specific verb occurring only in this construction, but many of the verbs attested in the construction have only a very weak semantic link with *banen*.

There has been an increase in type and token frequency in the NLCOW data compared with the SoNaR data, although the NLCOW corpus is much bigger. Some of the Dutch exemplar clouds exhibit pluricentricity; the exemplar cloud of verbs of sound emission contains verbs of human (*kreunen* 'groan', *fluisteren* 'whisper') and animal sounds (*brullen* 'roar', *grommen* 'growl'). The exemplar cloud of actions involving the mouth support Hilpert's (2015: 349) that marginal members may constitute their own exemplar cloud; an exemplar cloud of verbs relating to sex emerge in NLCOW, and *kussen*/*zoenen* 'kiss' may be the source. This also shows that exemplar clouds may overlap, because *kussen* could plausibly be assigned to either exemplar cloud; this ambiguity is what facilitates analogization between semantic domains. As was the case of the SoNaR data, the most frequently occurring member of an exemplar cloud is the most central one, both in terms of token frequency and attraction to the construction.

6.2.2.3 German

The verbs occurring in the German *way*-construction in *Deutsches Textarchiv* can be grouped into the following semantic categories: a) basic motion verbs; b) vehicular motion verbs; c) verbs encoding the manner of motion; d) verbs relating to finding or seeking a path; e) verbs encoding force; f) verbs of taking or clearing a path; g) verbs of achievement; and h) verbs of path creation. However, exemplar clouds have not yet formed, because almost all of these verbs are attested only once. The exceptions are *bahnen* 'make' (attested 7 times), *gehen* 'go' (2x), *nehmen* 'take' (9x), *bereiten* 'prepare' (4x), and *öffnen* 'open' (2x). Similarly to the early English *way*-construction, intransitive basic motion verbs and transitive verbs of acquisition (*nehmen* 'take') are the most frequently occurring verbs at this stage. Figure 25 shows more modern German data from the *Berliner Zeitung* corpus, which spans the period 1994 to 2005.

Figure 25: Exemplar clouds in the Berliner Zeitung *corpus*

Cloud	Items
basic motion	*BAHNEN* 'make' (107), *gehen* 'go' (46), *machen* 'make' (3)
vehicular motion	*fahren* 'go by transport' (5)
manner of motion	*erkleckern* 'spill messily' (1), *hinaufschleppen* 'slog one's way up' (1)
	ablaufen 'run off' (1)
finding a path	*FINDEN* 'find' (30), *suchen* 'seek' (1), *spuren* 'feel' (1)
	aussuchen 'seek out' (1), *ertasten* 'feel out' (1)
	erwandern 'discover on foot' (1), *nachzeichnen* 'trace' (1)
force dynamics	*freischießen* 'shoot one's way free' (5)
	freikämpfen 'fight o.w. free' (4), *schlagen* 'strike' (3)
	freibomben 'bomb o.w. free' (2), *sprengen* 'blow up' (1)
	freiballern 'shoot o.w. free' (1), *buddeln* 'dig' (1)
	graben 'dig' (1), *rempeln* 'shove' (1), *freipicken* 'pick free' (1)
	freisprengen 'bomb o.w. free' (1), *schaufeln* 'shovel' (1)
taking/clearing a path	*NEHMEN* 'take' (8), *fortsetzen* 'pursue' (5), *freimachen* 'make free' (5)
	freigeben 'make free' (3), *freischalten* 'unlock' (1)
	freischreiben 'write o.w. free' (1), *offenhalten* 'keep open' (1)
	beschreiten 'pursue' (1), *antreten* 'set on' (1), *verfolgen* 'pursue' (1)
path creation	*ebnen* 'pave' (11), *glätten* 'smoothe' (1), *bereiten* 'prepare' (1)
lighting a path	*leuchten* 'light' (1)
achievement	*verdienen* 'earn' (1), *hocharbeiten* 'work o.w. up' (1)
sound emission	*hupen* 'beep' (2)
surfing the internet	*klicken* 'click' (1)

Compared with the *Deutsches Textarchiv* data, the exemplar cloud of basic motion verbs has increased in type frequency, with *machen* 'make' now entering the construction, possibly as a result of analogization with English *make one's way*.[44] The exemplar cloud of manner of motion verbs increases in type frequency, but the semantics is now different; verbs of leisurely motion (*laufen* 'walk', *wandeln* 'stroll') have been replaced by verbs of laborious or hurried motion (*hinaufschleppen* 'slog one's way up', *ablaufen* 'run off'). The exemplar cloud of verbs relating to finding or seeking a path now exhibits pluricentricity, with a cluster of verbs denoting the entity tracing out a path emerging (*ertasten* 'feel out', *spuren* 'feel', *nachzeichnen* 'trace'); these verbs denote the method of finding the path, and have emerged by extension of *finden* 'find', which is the most frequently attested verb in this cluster in the *Deutsches Textarchiv* corpus. In the exemplar cloud of verbs relating to force, verbs encoding more specific types of force are now attested, including a cluster of verbs with a *frei-* 'free' prefix (*freischießen* 'shoot one's way free', *freikämpfen* 'fight one's

[44] This hypothesis is explored in more detail the following chapter.

way free', *freibomben* 'bomb one's way free', *freisprengen* 'bomb one's way free'). There is also a cluster of verbs relating to digging (*graben* 'dig', *buddeln* 'dig', *schaufeln* 'shovel'). The exemplar cloud of verbs relating to taking or clearing a path increases in type frequency. More verbs of pursuing enter the construction (*fortsetzen*, *verfolgen*), as well as more verbs relating to keeping or making a path free of obstacles (*freigeben* 'make free', *freimachen* 'make free', *freischalten* 'unlock'). The exemplar cloud of verbs of path creation decreases in type frequency; *ebnen* 'pave' enters the construction, and is by far the most frequently attested verb in this cluster. As well as these exemplar clouds, several new exemplar clouds have emerged; these are also found in the Dutch and English constructions. These exemplar clouds contain verbs of lighting a path, verbs relating to commerce and finance, verbs of sound emission, and verbs pertaining to the Internet and computing.

Figure 26 shows how the exemplar clouds in the *Berliner Zeitung* corpus have changed by comparing them to the most modern data from the DECOW corpus.

Figure 26: Exemplar clouds in DECOW

Cloud	Items
basic motion	*BAHNEN* 'make' (432), *machen* 'make' (7)
manner of motion	*schlängeln* 'wriggle' (4)
finding a path	*SUCHEN* 'seek' (188), er*tasten* 'feel out' (3)
	tasten 'feel' (2), *zeichnen* 'sketch' (1)
	erwandern 'discover on foot' (1), *nachzeichnen* 'trace' (1)
force dynamics (31 not shown)	*KÄMPFEN* 'fight' (19), *erkämpfen* 'fight o.w. out' (18)
	freischießen 'shoot o.w. free' (15), *graben* 'dig' (8)
	schlagen 'strike' (5), *brechen* 'break' (4), *freikämpfen* 'fight o.w. free' (4)
	schneiden 'cut' (3), *freibomben* 'bomb o.w. free' (2)
	schießen 'shoot' (2)
taking/clearing a path	*erleichtern* 'make easier' (5), *freimachen* 'make free' (5)
	eröffnen 'open' (3), *freiräumen* 'clear' (3), *offenhalten* 'keep open' (2)
	nehmen 'take' (2), *öffnen* 'open' (1), *räumen* 'clear' (1)
	freigeben 'make free' (1)
path creation	*EBNEN* 'pave' (9), *bauen* 'build' (3), *bereiten* 'prepare' (1)
	verschaffen 'provide' (1)
lighting a path	*leuchten* 'light' (2)
achievement	*ERARBEITEN* 'earn' (6), *arbeiten* 'work' (2)
	erspielen 'gain by playing' (2), *erstreiten* 'contend successfully' (1)
	ersingen 'gain by singing' (1)
commerce/finance	*kaufen* 'buy' (1), *erkaufen* 'buy' (1)
sound emission	*schreien* 'scream' (1), *schrammeln* 'play schrammel music' (1)
	beleidigen 'insult' (1)
performance	*tanzen* 'dance' (1)
path shape	*hinaufbewegen* 'move up' (1)

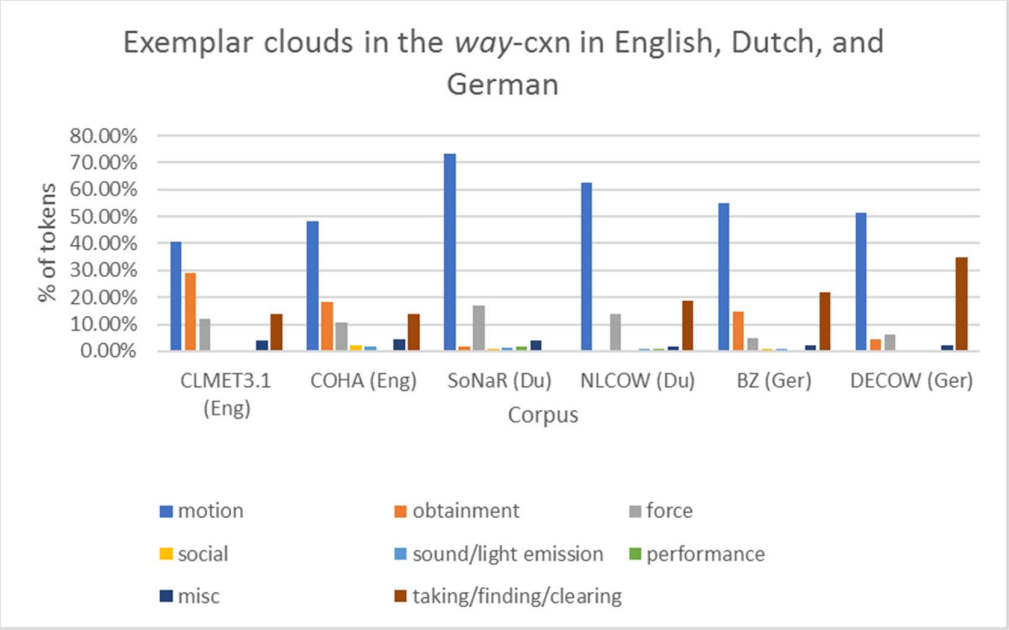
terrain of motion	*erklimmen* 'climb' (1)
deceit	*erschleichen* 'trick' (1)
persuasion	*schmeicheln* 'flatter' (1)
sport	*bolzen* 'kick a ball around' (1)
actions w/ mouth	*fressen* 'eat' (3), küssen 'kiss' (3), *mampfen* 'munch' (1) *lecken* 'lick' (1), *nippen* 'nip' (1), *spucken* 'spit' (1)
bodily functions	*scheißen* 'shit' (1)
cognition	*erdenken* 'think up' (1), *ausdenken* 'think out' (1)

As in the *Berliner Zeitung* data, *bahnen* remains by far the most frequent verb in the German *way*-construction. The cluster of verbs encoding vehicular motion has fallen out of use, as has the cluster of verbs relating to surfing the internet, although these were already very low in type and token frequency in the *Berliner Zeitung* corpus. A change has taken place in the exemplar clouds of verbs of finding a path; *finden* 'find' was very frequent in the *Berliner Zeitung* corpus, but it is not attested at all in the DECOW corpus. The most frequent verb of finding a path in that corpus is now *suchen* 'seek', which was very rare in the *Berliner Zeitung* corpus, occurring only once. The cluster of verbs encoding the manner of motion has undergone a decrease in type frequency, while the clusters of verbs relating to force and sound emission have greatly increased in type frequency, though each type in the latter cluster is of low token frequency.

A comparison with the *Berliner Zeitung* data also reveals that several new clusters of verbs have emerged. These relate to sport, deceit, persuasion, path shape, path terrain, actions involving the mouth, bodily functions, and cognition. Most of these are found in the English construction, and no plausible source of analogization exists in the German construction; I hypothesise therefore that these clusters of verbs arose by cross-linguistic analogization with the English construction, a hypothesis I explore in more detail in the following chapter. Data from both the *Berliner Zeitung* and DECOW corpora show that the German *way*-construction now permits several verbs with the prefixes *frei-* 'free' and *er-* 'obtain by V'; this shows that movement away from an obstacle and obtainment are central to the construction. The German data also support Bybee's (2013) hypothesis that an exemplar model of language is necessary to fully account for the idiosyncrasies of a construction; a prototype model is inadequate in explaining why verbs of sound emission, for instance, are now attested in the construction, because these have very little relation to the prototypical verbs *bahnen* 'make one's way', *suchen* 'seek' and *kämpfen* 'fight'.

In sum, data from the *way*-construction in all three languages convincingly support the hypothesis that the items that can fill a particular slot of a construction can be grouped into clusters of semantically similar verbs. It was also shown in this subsection that an exemplar model of language is essential to explaining how the *way*-construction has acquired its idiosyncrasies; an analysis of the *way*-construction in terms of prototypes lacks explanatory power. The data showed that each language has almost identical clusters of verbs; the only exception is the cluster of verbs relating to surfing the internet, which are attested only in the Dutch and German corpora. The more modern Dutch and German data reveal the emergence of semantic clusters identical to those of the English construction, which have no plausible analogical source in the Dutch and German constructions; I propose therefore that these arose as a result of language contact and speakers' exposure to English. The differences between the exemplar clouds in the three languages are summarised in the diagram below. For ease of interpretation, I group the exemplar clouds into much broader categories, and classify a verb as miscellaneous if it does not easily fit into one of the broad categories. This diagram reveals that the exemplar clouds found in each language are very similar, and that they constitute roughly the same proportion of the construction in each language.

Figure 27: Similarities and differences between exemplar clouds in the way*-construction in the three languages*



In the following section, I present a review of prior literature on effects other than frequency effects, and statistical measures more sophisticated than raw token frequency. I show that genre effects are present in the English *way*-construction, and that statistical testing can be used to investigate these both in my dataset and more generally. I also investigate a statistical measure known as dispersion, which has been hypothesised to be strongly correlated with token frequency (Hilpert & Correia Saavedra 2017) and show that this correlation is not as strong as assumed in my data.

6.3 Factors and statistical measures other than frequency in language change

6.3.1 Factors other than frequency

Although frequency undoubtedly plays a role in language change, referring to "frequency effects" obscures the fact that change is driven by speakers themselves, rather than any inherent property of an item such as high frequency. Experimental data suggests that at least part of the frequency effect can be attributed to speakers' experience and knowledge of the words and phrases concerned. Gardner et al. (1987) showed that nurses, law students, and engineers responded more quickly in a lexical decision task to items relevant to their field, both at the word and phrase level. Caldwell-Harris et al.'s (2012) study provides further evidence for this; they showed that that religious Jews process religious phrases in Hebrew more quickly than secular Jews.

Frequency effects can also be explained in terms of a number of other factors. Highly frequent words are typically short, and denote concrete rather than abstract items (Divjak & Caldwell-Harris 2015: 57). Highly frequent words also tend to be more easily imaginable and are acquired earlier (ibid.). There is also a positive correlation between frequency and other lexical attributes, such as orthographic neighbourhood density, syntactic family size, noun-verb ratio and number of meanings; it has been proposed that these factors, rather than frequency, are what contribute to increased processing speed. It has also been suggested that factors which are correlated with frequency are more strongly correlated with behavioural outcomes than frequency.

One of these factors is the typical context of co-occurrence of words. Some studies have concluded that contextual diversity is a psychologically more relevant variable than frequency (McDonald & Shillcock 2001; Adelman et al. 2006). This conclusion is supported by Raymond and Brown (2012), who find that word frequency plays no role in reductive processes once contextual factors are taken into account. Baayen (2011a) also provides evidence for contextual and entropy-based measures.

It is often claimed that highly frequent words are more likely to be phonologically reduced (see e.g. Bybee & Thompson 2000). However, Bell et al. (2009) showed that there is in fact a strong negative correlation between frequency of occurrence and the degree of phonetic reduction if all other factors are controlled for, though this correlation is not uniform across expressions; they observed a correlation between frequency and phonetic reduction of content words, but not of function words. More specifically, they observed that frequent content words are more likely to be reduced than infrequent ones, regardless of linguistic context, whereas function words are only phonetically reduced if their occurrence is predictable from the linguistic context, regardless of their frequency. Jurafsky et al. (2001) also found that some high-frequency words had low rates of phonological reduction, even when controlling for phonological context. Raymond and Brown (2012: 36), on the other hand, found that high-frequency items in Spanish show word-initial fricative reduction, but that other factors are responsible for this reduction, including lexical structure and class, extra-lexical phonological context, prosodic environment, speech rate, and sociolinguistic factors.

However, the criticisms of accounts of linguistic change in terms of frequency effects are attacking a straw man. Very few usage-based linguists claim that frequency alone is a determinant of linguistic change. Bybee (2007: 17-18) acknowledges that frequency interacts with other factors such as phonological and semantic similarity, categorization, and semantic/pragmatic change, as well as finding a relationship between frequency and generality of meaning. Croft (2006: 70) also finds that high frequency is motivated by joint discourse salience. This is not to say that sometimes type and token frequency does not have a direct effect, but in other domains factors such as recency and salience are more important (Ellis 2012: 7; Behrens & Pfänder 2016: 1). Register and genre also needs to be taken into account, as different registers exhibit different frequency profiles (Taylor 2012: 148); genre effects are

tested on my dataset in Section 6.3.3. The following subsection discusses some statistical measures that do not rely on raw token frequencies alone, and how they were employed in the analysis of my data.

6.3.2 Statistical measures other than frequency

A number of alternative statistical measures have been suggested in the literature. Hilpert and Correia Saavedra (2017), for instance, suggest using dispersion as a variable rather than token frequency alone, because evenness of dispersion of grammatical elements is an effect of high frequency (Hilpert & Correia Saavedra 2017). Using dispersion is also relevant because high frequency or high attractions are important when they are attested in many different registers or situations (Gries 2014b: 40). For instance, Stefanowitsch and Gries (2003) found that *fold* is attracted to the imperative construction, but all of the occurrences of *fold* in this construction occurred in a single text, about origami (Taylor 2012: 15). Further, Gries (2010b) finds that dispersion is sometimes a better predictor of reaction times than frequency, and observes (2008: 428) that dispersion measurements can be used to adjust observed text frequencies, so that essentially dispersion is added as a covariate to analyses that investigate a relation between frequency and some cognitive response.

Gries (2008: 415) measures dispersion by the deviation of proportions in the following way. First, a corpus is divided into parts. For each part, it is determined what percentage of the corpus it contains. The frequency of the linguistic item under investigation in the corpus and in all the corpus parts is then calculated. For all corpus parts, the differences between the observed and expected percentages are summed up, and the sum of these differences is divided by 2. This calculation yields a value between 0 and 1, where 0 indicates a perfectly even dispersion and 1 indicates a maximally uneven dispersion.

Another proposed measure is conditional and transitional probabilities. The entrenchment of complex units can be measured by relating the frequency of the expression as a whole to the frequency of the individual parts (Stefanowitsch & Flach 2016: 111). This is done by calculating the conditional probability $p(w_{n+1}|w_n)$, i.e. the likelihood that we encounter the word w_{n+1} given that we have just encountered the word w_n. This conditional probability is calculated by dividing the frequency of the

bigram w_n w_{n+1} by the frequency of w_{n+1}; this is known as the transitional probability (Stefanowitsch & Flach 2016: 111). The entrenchment can also be measured in the opposite direction, i.e. $p(w_n|w_{n+1})$. The product of these two probabilities is known as the cue validity (Stefanowitsch & Flach 2016: 112-113). Experimental evidence suggests that conditional probabilities are an accurate predictor of a range of linguistic behaviours (Divjak & Caldwell-Harris 2015: 67 and references therein). However, this approach has as an unintended consequence that the importance of rare collocations is often overestimated (Stefanowitsch & Flach 2016: 114). Probability-based measures react more sensitively to small differences for low frequencies than for high frequencies (ibid.). Measuring the type/token ratio has also been suggested in the literature, because the entrenchment of a schema depends on its type frequency, and token frequency in turn depends on type frequency (Stefanowitsch & Flach 2016: 118).

A third way of measuring the entrenchment of complex units involves statistical testing. This approach takes the complexity of the units into account but measures frequency in terms of measures derived from contingency tests (Stefanowitsch & Flach 2016: 115). These measures include statistics such as G^2 from the log-likelihood test and χ^2 from the chi-squared test, and the *p* values of exact tests such as the Fisher-Yates test (ibid.); collostructional analysis uses the latter measure, and this is the family of methods that I employ in this study. This approach is advantageous in that it also takes into account the frequency of co-occurrence of the elements in question relative to the overall size of the corpus, as well as the co-occurrence frequency of elements relative to their individual frequencies (Stefanowitsch & Flach 2016: 115). Another advantage is that the *p* value of the Fisher-Yates test downgrades the influence of words that are frequent everywhere and weighs more highly observed relative frequencies of co-occurrence that are based on high absolute frequencies of co-occurrence (Gries 2014b: 22). The *p*-value of the Fisher-Yates exact test is also highly correlated with statistics that are known to be relevant in cognitive contexts (Gries 2014b: 29).

The statistically-minded approach has been criticised by Bybee (2010b: 97), who argues that words do not appear in a corpus by pure chance, and that the factors that make a word high frequency in a corpus are the factors that make it a central member of the category. However, this logic would lead linguists to conclude that the *as-*

predicative (found in sentences such as *He is regarded as a brilliant linguist*) is defined by *see* and *describe*, not by *regard*, which has the highest collostruction strength (Gries 2014b: 25). Given that *see* occurs in a wide range of constructions and *regard* almost always occurs in the *as*-predicative construction, saying that the *as*-predicative is characterised by *see* is an unintuitive result (ibid.).

Collostructional analysis involves counting the number of constructions in a corpus, which Bybee (2010b: 98) also takes issue with, because a given sentence may instantiate multiple constructions. The sentence *He is going to bake me a cake*, for example, contains the *be going to* future construction, the ditransitive construction, and the individual words that make up that sentence are also constructions in their own right. However, in a study conducted with Eddington (2006), she concedes that different corpus sizes yielded similar results, which suggests that the number of constructions in the corpus is irrelevant and it is therefore not a problem that the exact number of constructions in the corpus cannot be counted. Bybee (2010b: 101) also asserts that collostructional analysis ignores low-frequency collexemes, which she finds to be problematic because low-frequency lexemes show the productive expansion of a category. However, many of the low-frequency collexemes in the *way*-construction have a high collostructional strength, and so Bybee's claim is inaccurate.

Using statistical measures other than frequency is also advantageous because using raw frequencies alone assumes that frequency effects are linear; according to this approach, an item that is twice as frequent in a corpus is twice as entrenched. This is in error, as studies have shown that many effects in learning, memory, and cognition are not linear, such as the power law of learning (Anderson 1982). Tryk (1986) further shows that word frequency effects are logarithmic, not linear. It has also been demonstrated in many studies that forgetting curves are logarithmic (see e.g. Gries 2005a, Szmrecsanyi 2006). The *p*-value of the Fisher-Yates test is not linear, and therefore this approach yields more meaningful results than studies which use raw frequencies alone. The following subsection explores one of the alternative statistical measures described above, dispersion. In addition, I test for any genre effects in the English data, and find that genre effects are present. I analyse data from COHA and COCA; although the temporal range of COCA partially overlaps with COHA, COCA contains more informal genres and it is possible that there are genre effects in these informal genres.

6.3.3 Results

6.3.3.1 English

In order to calculate the degree of dispersion in the CLMET3.1 corpus, the corpus was split into 21 roughly equal chunks of around 2 million words each, and the dispersion coefficient was calculated as described in Section 6.4.2 above. Calculating the dispersion coefficient proved more problematic for the COHA corpus, because the corpus is divided into chunks of one decade and the word count is not equal across each decade. The word count in each decade of COHA is presented here again in Table 35 for convenience.

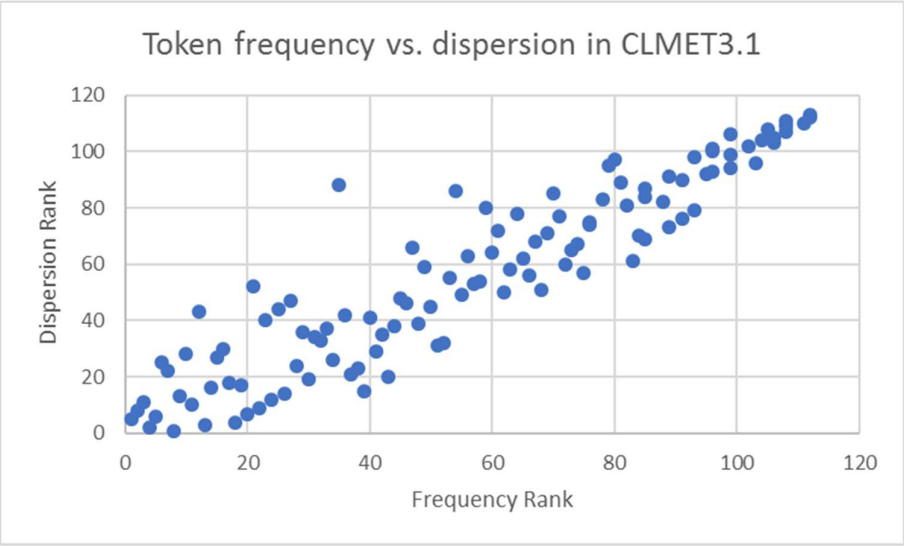
Table 35: Word count in COHA per decade

Decade	# Words	Decade	# Words
1810s	1,181,022	1910s	22,655,252
1820s	6,927,005	1920s	25,632,411
1830s	13,773,987	1930s	24,413,247
1840s	16,046,854	1940s	24,144,478
1850s	16,493,826	1950s	24,398,180
1860s	17,125,102	1960s	23,927,982
1870s	18,610,160	1970s	23,769,305
1880s	20,872,855	1980s	25,178,952
1890s	21,183,383	1990s	27,877,340
1900s	22,541,232	2000s	29,479,451

As this table shows, the 1810s and 1820s contain very few words compared to the other decades; data from these decades were therefore excluded from the calculations of the dispersion coefficient. The distribution of words across the other decades is unequal, which Gries (2008) warns may be problematic when calculating dispersion. Because I do not have an offline copy of COHA, I could not work out the word count in each year; the COHA interface I used only lists the word count per decade. I therefore could not divide the corpus into equal chunks.

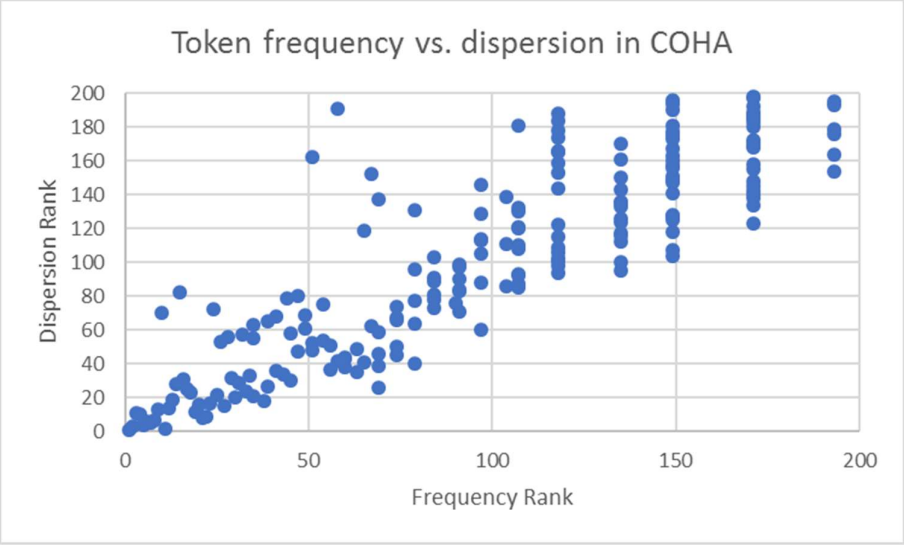
A Kendall's rank correlation test was applied to the data. The rank of each item in terms of token frequency and dispersion was calculated and the correlation between the two ranks was calculated. Figure 28 shows the rank of each verb in the CLMET3.1 corpus in terms of these two variables.

Figure 28: Rank of each verb in CLMET3.1 in terms of frequency and dispersion



The graph above suggests a strong positive correlation between an item's token frequency and its dispersion. This hypothesis is confirmed by Kendall's τ coefficient ($\tau = 0.777$, significant at $p < 0.001$). The correlation between token frequency and dispersion in COHA is shown in Figure 29. Only the 200 most frequently occurring verbs are considered.

Figure 29: Rank of each verb in COHA in terms of frequency and dispersion



This suggests that the correlation between token frequency and dispersion is weaker in COHA than in CLMET3.1 (Kendall's τ = 0.716, significant at $p < 0.01$), although this is still a strong positive correlation. The findings from the CLMET3.1 and COHA corpora support Hilpert and Correia Saavedra's (2017) hypothesis that there is a correlation between evenness of dispersion and high token frequency.

In order to test for genre effects, I calculated each verb's genre dispersion in COHA and COCA in the following way. I did not perform this analysis on the CLMET3.1 corpus because it is much more homogeneous than COHA and COCA in terms of genre; the texts are mostly from very formal registers (De Smet 2005). Firstly, I counted the number of observed tokens of each verb in the construction and in the corpus by genre. The expected number of tokens by genre in the construction was then calculated by multiplying the total frequency of each verb in the construction by the percentage that each genre represents in the corpus. The differences between the observed and expected percentages were added together and divided by 2 into order to yield a measure of genre dispersion between 0 and 1. A value of 0 would suggest that the verb in the construction is distributed by genre in exactly the same way as the corpus, and a value of 1 would indicate a maximally different genre distribution. The rank of each verb in terms of token frequency and genre distribution was calculated, and Kendall's τ was computed in order to assess whether these genre dispersion effects are related to token frequency.

Table 36 below shows the percentage of the COHA corpus that each genre contains, and the number of tokens of the *way*-construction in each genre.

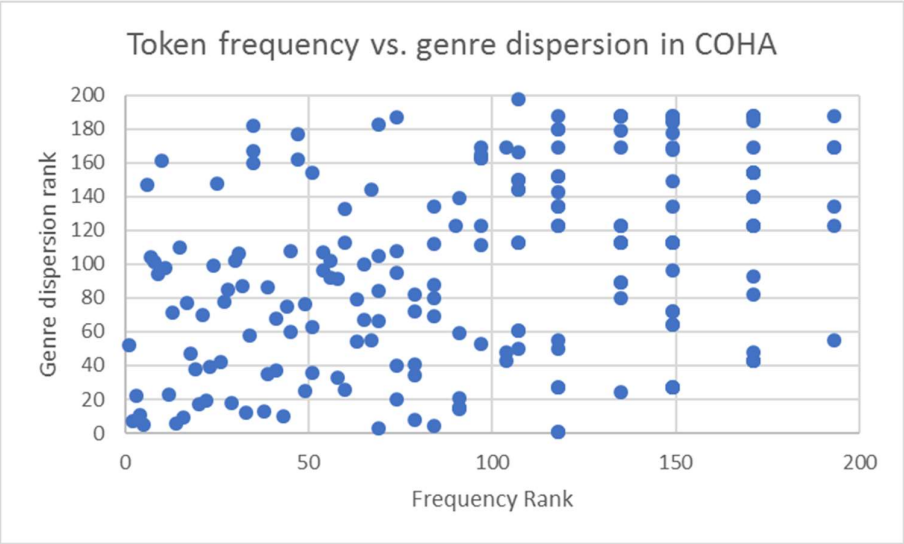
Table 36: Distribution of texts and tokens of the way*-construction in COHA by genre*

Genre	# Words	%	# Tokens	%
Fiction	207,633,395	51.11%	12,000	62.81%
Magazines	97,207,399	23.93%	4,133	21.63%
Newspapers	40,124,656	9.88%	1,076	5.63%
Non-fiction	61,266,574	15.08%	1,896	9.92%
Total	**406,232,024**	**100.00%**	**19,105**	**100.00%**

Table 36 shows that the distribution of tokens of the *way*-construction by genre is roughly what is expected based on the percentage of the corpus each genre contains.

The genre dispersion and token frequency rank of the 200 most frequently occurring verbs in COHA is shown in Figure 30.

Figure 30: Genre dispersion and token frequency ranks of 200 most frequent verbs in COHA



The average genre dispersion coefficient of each verb in the corpus was 0.224, which indicates a very even dispersion. A Kendall's τ coefficient of 0.276, although statistically significant at $p < 0.001$, indicates a very weak relationship between genre dispersion and token frequency. This shows that there are effects other than frequency effects at work here, since the genre dispersion effects cannot reliably be explained in terms of token frequency. Table 37 below shows the distribution of the corpus and *way*-construction in COCA by genre.

Table 37: Distribution of texts and tokens of the way*-construction in COCA by genre*

Genre	# Words	%	# Tokens	%
Spoken	116,748,578	20.47%	2,774	11.72%
Fiction	111,845,122	19.61%	9,904	41.84%
Magazines	117,354,113	20.58%	5,331	22.52%
Newspapers	112,995,407	19.81%	3,983	16.83%
Academic	111,410,528	19.53%	1,680	7.10%
Total	**570,353,748**	**100.00%**	**23,672**	**100.00%**

Table 37 reveals a greater disparity in genre between the construction and corpus than the COHA data. In comparison to the distribution of the corpus by genre, the *way*-construction is dispreferred in the spoken and academic genres, and strongly preferred in the fiction genre. The average genre dispersion coefficient of each verb in the corpus was 0.412, which indicates a relatively uneven distribution of the construction by genre. In Figure 31 below, I plot the rank of the 200 most frequently occurring verbs in COCA in terms of their token frequency and genre dispersion.

Figure 31: Genre dispersion and token frequency ranks of 200 most frequent verbs in COCA

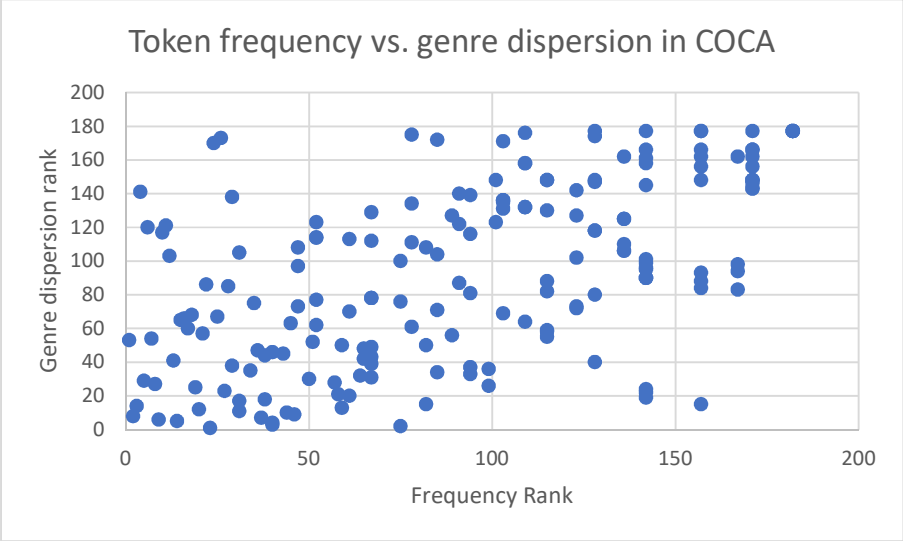


Figure 40 suggests only a moderate correlation between an item's token frequency and its genre dispersion. This hypothesis is confirmed by the value of Kendall's τ coefficient: $\tau = 0.410$, statistically significant at $p < 0.01$; this shows that there is not a strong relationship between token frequency and genre dispersion. When the English data on dispersion and genre effects are viewed together, it is clear that both frequency effects and genre effects have played a role in the development of the English *way*-construction, because not all of the genre dispersion can be explained in terms of frequency. COCA revealed stronger genre effects than COHA, especially in the fiction genre.

6.4 Conclusion

This chapter of the thesis has provided evidence of frequency effects in language and in the *way*-construction in English, Dutch, and German. It was demonstrated in Section 6.2 that the lexical items that can fill the slot of a construction can be grouped into exemplar clouds of semantically related verbs, and that the semantic categories are very similar in the English, Dutch, and German *way*-construction. It was further shown that the extension of each exemplar cloud usually, but not always, proceeds on the basis of the most frequently attested item in that exemplar cloud, confirming Bybee and Eddington's (2006) hypothesis. Section 6.3 showed that there are genre effects in the English *way*-construction which cannot plausibly be explained in terms of frequency effects alone, and showed that statistical testing can be used to measure genre effects. There are promising avenues for future research if Dutch and German corpora are used that allow the calculation of dispersion and genre dispersion coefficients; the corpora chosen for the investigation in those languages unfortunately did not allow for this. It was also shown in this section that there is a strong positive correlation between dispersion and token frequency in the CLMET3.1 and COHA data, but not in COCA, which suggests that a high degree of dispersion may not always be a good predictor of high token frequency, as Hilpert and Correia Saavedra (2017) claim. The following chapter is concerned with the phenomena of language contact and borrowing and their role in the development of the Dutch and German *way*-constructions. In this chapter, I explore the hypothesis that the Dutch and German *way*-constructions have undergone considerable analogical extension as a result of cross-linguistic analogization with the English *way*-construction, which in turn was caused by speakers' exposure to English language in the Internet era.

This page intentionally left blank

7. The role of contact and borrowing in the development of the Dutch and German *way*-constructions

7.1 Introduction

This chapter of the thesis answers the final of my research questions: what is the role of contact and borrowing in the development of the Dutch and German *way*-construction, and how can a constructional theory of language accommodate this borrowing? Though contact and borrowing are closely related, they are nevertheless phenomena that should be kept apart. The term 'language contact' refers to contact between speakers of different languages (Milroy 1997: 311), and 'borrowing' is the transfer of features from one language to another as a result of language contact (Aikhenvald 2006). Much of the literature on borrowing has focused on lexical borrowing; comparatively little attention has been paid to the borrowing of complex constructions in the Construction Grammar sense. The analysis in this chapter is couched in a Diasystematic Construction Grammar framework (see especially Höder 2012; but also Höder 2011, 2014a, 2014b, 2016), a strain of Construction Grammar which states that multilingual speakers organise their knowledge of multiple languages in a diasystem (see also Mufwene 2001: 4 on the feature pool). The remainder of the chapter is organised as follows. Section 7.2 describes how contact and borrowing phenomena can be integrated into a Construction Grammar framework. This section begins by outlining the motivations for borrowing. I go on to show that the borrowing of complex constructions is not impossible, as claimed by some linguists (e.g. Sapir 1921; Gerritsen 1984; Sankoff 2002), but widely attested. I also show that the borrowing of only the formal or functional pole of a construction is possible. Section 7.3 discusses Höder's Diasystematic Construction Grammar and describes the factors that may lead to the establishment of a diasystem. I demonstrate in this section that Dutch and German speakers' increasing exposure to English has

led to the establishment of a diasystem between these two languages and English. Section 7.4 argues that the incidental activity reading of the Dutch and German *way*-constructions is a constructional borrowing from English. This hypothesis is supported by three pieces of evidence. Firstly, the incidental activity reading was not attested in Dutch and German until the rapid increase in exposure to English took place in the Internet era. Secondly, the emergence of the incidental activity reading in the three languages is not equally likely, because the input verbs and pathway of analogical extensions are not the same in each language. Thirdly, many of the verbs used in the Dutch and German incidental activity reading are of English origin; verbs of English origin are not attested at all in my dataset until the Internet era. Section 7.5 concludes the chapter.

7.2 Contact and borrowing in constructionist frameworks

7.2.1 What motivates borrowing?

Winford (2010: 177) points out that borrowing may be motivated by prestige (see also Matras 2009: 149). This explains the high degree of borrowing from English found in Dutch; English has held prestige in the Netherlands since the Second World War (Edwards 2016). The adoption of English words and phrases became much more frequent after this period, as English was seen as the language of the liberators (Ridder 1995: 44); this view, of course, was not shared by Germans. English has been an important influence on Dutch since before this period, however; Ridder (1995: 44) dates the beginning of the influence of English to the mid-19th century. German speakers show greater reluctance to borrow from English; this may be understood against the background of a German national identity, of which the German language is a core value (Barbour 2005: 154; see also Barbour 2000).

Despite the prestige that English enjoys in the Netherlands, numerous prescriptivist organisations against anglicisms in Dutch have been founded, such as the *Stichting Nederlands* ('Dutch language foundation') and the *Stichting Taalverdediging* ('Language Defence Foundation') (Vriesendorp & Rutten 2017: 47). English also appears to be losing prestige in the realm of advertising. Korzilius et al. (2004: 102) found that the use of English in Dutch-speaking advertisements did not affect applicants' attitude to the advertisement, the job or to the organisation, even when the

advertisement was completely in English, but a more recent study by Van Meurs et al. (2017) found that in job advertisements, applicants had a more positive attitude to the Dutch version of a job title than its English equivalent, preferring titles such as *hoofdredacteur* 'editor-in-chief' to their English equivalents.

As well as prestige, Matras (2009: 237) points out that borrowing among languages is a common feature of second language acquisition; this explains the greater degree of borrowing from English found in Dutch than German, because English is compulsory in Dutch secondary schools but not in German ones (Hilgendorf 2005; Edwards 2016). Meney (1994: 930) argues that sociolinguistic factors such as sex, social status, communicative situation, and register may also be secondary factors motivating borrowing. Borrowing may also be motivated by multilingual diffusion in an area (Aikhenvald 2002: 13). However, the interaction between these sociolinguistic factors and linguistic change is not well known, because the short-term and long-term effects of contact are rarely brought together (Backus et al. 2011: 738). Haase (1991) points out that borrowing may have functional motivations. He argues that bilingual speakers are motivated to use the expressive means of both their languages, and wish to have an equal number of constructions at their disposal in each language. Speakers' desire to have an equal number of constructions at their disposal in each language would explain why the incidental activity reading of the English *way*-construction has been borrowed from English into Dutch and German, as I demonstrate in Section 7.4.

7.2.2 The borrowing of (parts of) constructions and the resultant changes in the constructional network

Several scholars have attempted to rank items along a borrowability hierarchy (see, among many others, Haugen 1950; Muysken 1981; Muysken and van Hout 1994). Matras (2011: 208), for instance, presents the following borrowability hierarchy, based on that of Moravcsik (1978) and Field and Comrie (2002).

a) nouns > non-nouns, function words

b) free morphemes > bound morphemes

c) derivational morphology > inflectional morphology

d) agglutinating suffix > fusional affix

Based on these borrowability hierarchies, it has widely been claimed that grammatical categories are less likely to be borrowed than lexical ones (Harris & Campbell 1995: 132-133 and references there). For instance, it has been argued that morphologically complex lexical categories are less amenable to borrowing; this constraint explains why verbs are less likely to be borrowed than nouns (Winford 2010: 178). Closed-class items such as prepositions are also said to be less likely to be borrowed than open-class items such as nouns and verbs due to their greater structural cohesion (Winford 2010: 178). Moravcsik (1978: 111) even claims that the borrowing of verbs is impossible: "[a] lexical item whose meaning is verbal can never be included in the set of borrowed properties". This has been refuted by many studies, for instance by Dux (2017), who shows that Texas German extensively borrows English verbs. Further, I will show in Section 7.4 that Dutch and German have borrowed several English verbs and used these in the *way*-construction.

Contrary to the claims of many studies, borrowing is not limited to items at the lexical end of the lexical-grammatical continuum (Heine & Kuteva 2005: 1). Hasselmo (1970), for instance, demonstrates that discourse markers such as *of course* are transferred into American-Swedish discourse, and Clyne (1978) showed that the co-ordinators *and* and *but* are transferred into German discourse among the English-German bilinguals in his study. Borrowing of these closed-class items may also take place between typologically unrelated languages. For example, Alvanoudi (2017) demonstrates that the variety of Greek spoken by immigrants in Far North Queensland, Australia has borrowed discourse patterns from English. Further, Mithun (2008) shows that demonstrative constructions were borrowed between two unrelated languages on the northwest coast of the US. A study by Doğruöz and Backus (2009) adds to this growing body of evidence; they find that the variety of Turkish spoken by second generation immigrants to the Netherlands contains some function words and

other morphosyntactic constructions borrowed from Dutch. The borrowing of the incidental activity reading of the *way*-construction is yet another example of an item at the grammatical end of the lexical-grammatical continuum; in Section 7.4, I provide evidence that this borrowing took place.

While prior accounts of contact and borrowing phenomena distinguish between many types of borrowing, this is unhelpful in a constructionist framework, where it is assumed that the entirety of a speaker's knowledge of language consists of constructions. It is therefore constructions or parts of constructions (i.e. the form or meaning pole, or both) that are borrowed. Ranking items in a borrowability hierarchy according to their part of speech is also unhelpful in a constructionist framework, because in many constructionist theories it is assumed that grammatical categories are not cross-linguistic, but language-specific (Haspelmath 2012: 109) or even construction-specific (Croft 2001: 13-14). Instead, explaining borrowability in terms of an item's position on the lexical-grammatical continuum or in terms of its complexity has greater explanatory power; items on the lexical end of the continuum are more likely to be borrowed, and minimally complex items (i.e. single words) are more likely to be borrowed than more complex constructions.

Borrowings of the formal pole of a construction alone are widely attested; *ersatz* in English connotes a replacement of poor quality, whereas in the donor language, German, it refers to any type of replacement. The meaning of the German construction has therefore not been borrowed into English, only the form. Examples of borrowing the functional pole of a construction are also widely attested. Pietsch (2010: 121) provides numerous convincing examples from Hiberno-English, which has gained a marker of habituality with periphrastic *do* and *be* on the basis of a functionally similar construction in Irish. Other examples of constructional borrowing from Irish in Hiberno-English include the *after*-perfect, found in sentences such as *I'm after selling the boat*, and the preservation of verb-subject order in embedded questions, e.g. *He asked me what on earth was I doing* (Pietsch 2010: 121). As a result of the syntactic influence of Irish, speakers of Hiberno-English use cleft constructions much more frequently than speakers of other dialects of English (Pietsch 2010: 122). This shows that function alone can be borrowed; speakers of Hiberno-English have borrowed several grammatical functions from Irish, but not the phonological form associated with these functions. As I demonstrate in Section 7.4,

Dutch and German speakers have borrowed the function of the incidental activity reading of the English *way*-construction, but not its phonological form. A further example of functional borrowing can be found in French; the French word *gratte-ciel* 'scrape-sky' (i.e. skyscraper) contains a borrowing of the meaning from English, but not of the phonological form.

Dux (2017: 416) identifies verbal constructions in the codeswitching of Texas German speakers where either the form or function alone have been transferred from English. In (219), the speaker borrows the English verb *look up* (information) but translates the English particle *up* with the functionally identical German prefix *auf*, where Standard German has *nach*. This prefix is combi0ned with the German verb *gucken* 'look'.

(219)	Ich	hab	Papiere	irgendwo	dann	kann	ichs	aufgucken
	I	have	papers	somewhere	then	can	I.it	up.look

'I have papers somewhere, then I can look it up' (Dux 2017: 416, ex. 24).

Further, he points to the transfer of the functional pole of the English [NP *better* V] construction found in *you better have a big coat*, which is loan-translated into German as *du besser hast 'n grossen Jacke*, which is unidiomatic in Standard German. He also gives examples of the formal pole alone being transferred, such as in *und hat helfed die Kirche* 'and has helped the church', where the fixed part of the English past tense construction *-ed* is combined with the root of the German verb *helfen* 'help'.

The speech of bilinguals provides ample evidence that constructions can be borrowed. Every item that a speaker has experienced in a given language has an impact on speakers' mental representation of language; if speakers speak more than one language, it follows that their constructional networks will contain items from multiple languages. Constructional borrowing occurs when speakers take an element that previously only occurred in one language and use it to form an utterance in a different language (Backus et al. 2011: 740); this is a type of cross-linguistic constructional contamination. Backus et al. (2011: 740) state that if speakers do this often, and these novel utterances become conventionalised, contact-induced change will occur, although this raises the question of how 'often' is considered often enough to take place. Another source of constructional transfer is when previously monolingual speakers acquire a second language; acquisition of a second language

can lead to competition between constructions in L1 and L2, which in turn can lead to constructional transfer (cf. Ellis 2008; Ellis & Cadierno 2009; Hendrikx et al. 2015: 8).

An example of competition between constructions in L1 and L2 can be found in Texas German (Dux 2017). Dux's study shows that speakers of Texas German, who are all bilingual with English as L1, readily insert English nouns into German structures (Dux 2017: 380). Inserting English verbs into German structures presents more of a challenge, as English differs from German in its verbal inflectional morphology, and collocations with objects and prepositions (ibid.); verbs are an integral part of the argument structure constructions of a language and are therefore very difficult to insert into foreign structures (Dux 2017: 397-398). Nevertheless, Texas German does borrow English verbs, and integrates them into existing structures in the language by adding verbal inflectional morphology from German, e.g. *cranken*, *crankte*, *habe gecrankt* 'crank, cranked, have cranked' (Dux 2017: 386). This phenomenon is known in Goldberg's (1990) terminology as "factoring out"; if part of a construction from a donor language is considered non-essential, it will not be borrowed into the recipient language. In this example, the English past and perfect tense inflectional morphology has not been borrowed into Texas German. In the borrowing of the incidental activity reading of the *way*-construction into German and Dutch, the entire formal pole of the English construction has been factored out; all of the formal elements in the borrowed construction come from Dutch and German.

Contact may lead to constructional changes in the recipient language. Backus et al. (2011) present a case study of Netherlands-Turkish (NL-Turkish), the variety of Turkish spoken by second-generation Turkish immigrants to the Netherlands. All speakers of NL-Turkish investigated in their study are completely fluent in Dutch. NL-Turkish has undergone constructional changes due to Dutch influence; one of these is the context extension of lexical items, whereby the use of a native lexical item and the construction associated with it is extended to novel contexts on the basis of its equivalent in the other language (Backus et al. 2011: 742). Backus et al. found that speakers of NL-Turkish extend the context of the verb *almak* 'take' on the basis of the equivalent Dutch verb *nemen*, using it in sentences equivalent to *I took the train*, which is ungrammatical in Standard Turkish, which prefers the equivalent of *I got on the train*. This kind of host-class expansion is very common in language contact situations (ibid.), and has also affected the Dutch and German *way*-constructions, as I

demonstrate in Section 7.4. The verb *yapmak* 'do' has also extended its context in the [N *yapmak*] construction in NL-Turkish, and several constructs are attested where the noun slot in the construction is filled by a deverbal noun of Dutch origin, such as *winkelen yapmak* 'do shopping' (Backus et al. 2011: 743).

Although constructional changes are a common outcome of language contact, the creation of a totally new construction, i.e. constructionalization, is quite rare (Backus et al. 2011: 743). Traugott and Trousdale (2013), for instance, do not mention contact-induced change in their analysis of constructionalization. They briefly remark on instantaneous constructionalization (2013: 29-30), though their examples are limited to lexical borrowings, e.g. *sushi*. Colleman (2016) shows that instantaneous constructionalization is not limited to the lexicon. He shows that Dutch has borrowed the English time-*away* construction (cf. Jackendoff 1997), found in sentences such as *He danced the night away*. He hypothesises that a form-meaning unit was copied into Dutch. While the verbal prefix *ver-* has usually been used in the function of 'spend [unit of time] doing V', the particle *weg* 'way' is now used. Sentences such as *Tim danst de nacht weg* 'Tim dances the night away' are now regularly attested. A further example is given in Colleman and Noël (2014), who show that Dutch NCI (*nominativus cum infinitivo* 'nominative with infinitive') patterns such as *geacht worden te* 'be deemed to' and *verondersteld worden te* 'be supposed to' are a case of polysemy copying and are modelled on English *be supposed to*. This English construction in turn was calqued from an equivalent Latin pattern (Colleman and Noël 2014: 221). In polysemy copying, an item of one language and an item of another language have overlapping functions, and as a consequence of language contact, the functional range of the two items becomes more aligned so that the item of the target language comes to express the same functions as the item of the model language (Colleman & Noël 2014: 225). In Section 7.4, I test the hypothesis that the incidental activity reading of the Dutch and German *way*-constructions are an instance of contact-induced instantaneous constructionalization. The following section describes Höder's Diasystematic Construction Grammar; the application of this theory to my data is presented in Section 7.4.

7.3 Höder's Diasystematic Construction Grammar

7.3.1 Factors leading to the establishment of a diasystem

Under the Construction Grammar view that all linguistic patterns are constructions (see e.g. Goldberg 2006: 5), speakers do not transfer individual words, but entire constructions, with both formal and semantic properties, or parts of constructions (Dux 2017: 392; cf. also Bisang 2001: 188). Construction Grammar approaches to language all agree that speakers learn schematic constructions by categorising them on the basis of the available input. If speakers have a multilingual input, this categorisation process will affect all languages (Höder 2012: 251). Cross-linguistic categorisation occurs when multilingual speakers establish an equivalence between elements of different languages (Höder 2014a: 143). This process leads to the establishment of a 'diasystem' (ibid.). The term 'diasystem' was coined by Weinreich (1953: 390) in order to account for regular correspondences between different phonological structures in closely related dialects; Höder extends the term to apply to other cross-linguistic similarities, and has integrated the notion of diasystems into a Construction Grammar framework, coining the term Diasystematic Construction Grammar. The diasystem is similar to Mufwene's (2001: 4) notion of the 'feature pool'; a range of input varieties from which speakers can select combinations of features.

Although Weinreich stipulates that the two systems in a diasystem must have partial similarity, typological similarity between languages is not a prerequisite for the establishment of a diasystem; Höder (2012) applies a diasystematic Construction Grammar approach to medieval contact between Latin and Old Swedish and contact-induced changes in written Old Swedish (see also Höder 2014a: 152). However, greater typological similarity between the two languages does facilitate their participation in a diasystem; Danish and Swedish are more likely to develop diasystematicity than French and Japanese, for instance (Höder 2012: 250; see also Höder 2011, 2014a).

Interlingual identification (cf. Weinreich 1953) is key to the establishment of a diasystem. Interlingual identification is based on cross-linguistic analogization; elements in different languages that are perceived as similar form a system of overarching structures in the diasystem (Höder 2012: 242). These similarities may be

perceived on structural, phonetic, semantic, functional, pragmatic, or even frequency-related grounds (Höder 2012: 249).

The establishment of a diasystem in turn leads to the establishment of diaconstructions. In a diasystematic Construction Grammar framework, it is assumed that some constructions, such as abstract syntactic patterns, are unspecified for language, while others, such as lexical items, are language-specific (Höder 2012: 251). The establishment of these diaconstructions at least partly depends on the degree to which they are conventionalised within the speech community (Höder 2014a: 143). Wasserscheidt (2014: 307) asks the question that, if multilingual speakers develop diaconstructions, how do they know that the language-specific constructs they are generalising are equivalent? Bilingualism research to date has mostly favoured an explanation in terms of formal equivalence, but an explanation in terms of formal and functional equivalence may have more explanatory power in a Construction Grammar framework (Wasserscheidt 2014: 307-308). If cross-linguistic generalisations were operationalised solely in terms of functional equivalence, this would incorrectly predict that speakers could randomly choose between the realisation of objects in different languages within a single sentence (Wasserscheidt 2014: 312-313). This is not supported by the empirical evidence; when speakers codeswitch between typologically distinct languages, grammatical markers are typically produced in a single language, while the lexical elements they combine with can come from any language (Wasserscheidt 2014: 313). Wasserscheidt cites the following example from Budzhak-Jones (1998), where a lexical item has been borrowed from English (*soda*), but the grammatical markers come from Ukrainian.

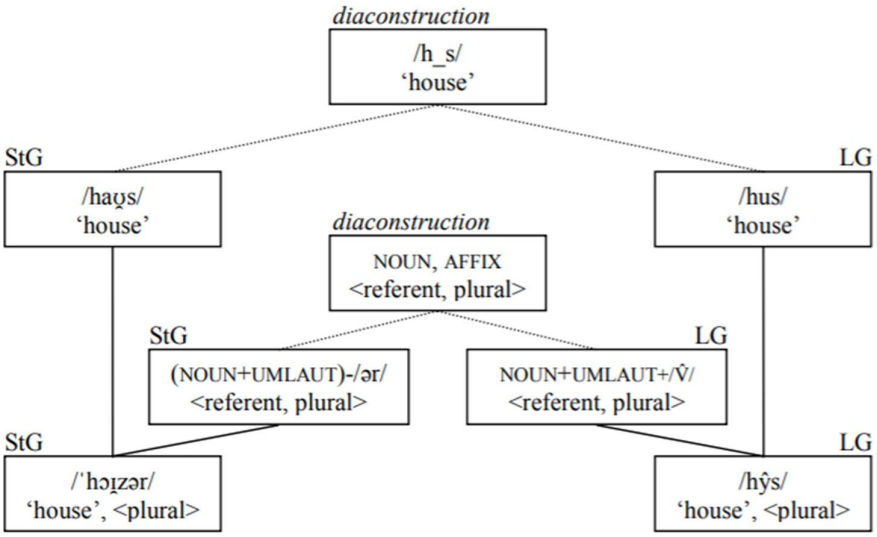
(220) Vin tam prodavav sodu
He there was.selling soda-ACC.F.SG
vsjaki taki rody sody
various such kinds soda.GEN.F.SG.
'He was there selling soda, various such kinds of soda' (example and gloss from Wasserscheidt 2014: 313, ex. 6).

In a diasystematic Construction Grammar framework, two language-specific elements are linked by a diasystematic link. Diasystematic links typically facilitate interlingual convergence in situations of stable language contact, which leads to a higher number

of common structures between the two languages (Höder 2012: 252). However, in a later work, Höder (2014b) points out that the distinction between convergence and divergence is not as clear as it may seem, and proposes instead to think in terms of pro- and counter-diasystematic change. In pro-diasystematic change, a construction can lose its language-specific contextual restriction and become more productive in another language (Höder 2012: 252). In counter-diasystematic change, on the other hand, a construction becomes more language-specific and becomes less productive in another language. The two language-specific elements and the diasystematic link between them constitute a more abstract item within an overarching system shared by the two languages; this more abstract item is known as a dia-element (Höder 2012: 249). These diasystematic links and dia-elements constitute a network through which two or more language systems used within a multilingual speaker group are inter-connected, and contains constructions of multiple languages with different degrees of schematicity (Höder 2012: 246).

The word for *house* is pronounced differently in Standard German and Low German: the Standard German pronunciation contains a diphthong, but in Low German there is a high back monophthong. The diaconstruction specifies what is common to both: in this case, the glottal fricative /h/ and the voiceless alveolar fricative /s/, and the concept of *house*. Figure 32 below illustrates the diasystem for this case. This figure also shows that Standard German (StG) and Low German (LG) form the plural for this noun differently; with an umlaut and *-er* suffix in Standard German, and with a vowel change in Low German. The diaconstruction for the plural formation captures the fact that a noun and an affix are common to both varieties.

Figure 32: Diasystem of the word for house *in Germanic varieties (reproduced from Figure 1 in Höder 2014a)*



The following section discusses the extent to which a diasystem has emerged between Dutch, German, and English. I show that Dutch and German speakers' increased proficiency and exposure to English has facilitated the participation of these languages in a diasystem with English. I also show that this diasystem serves various cultural and social functions.

7.3.2 The Dutch, German, and English diasystem

Previous studies on language contact in multilingual societies have typically focused on multilingual societies within the same nation-state (see e.g. Matras 2009), but Europe is increasingly becoming a single society with a high degree of interaction between people of different nation-states (Bruter 2005). In an increasingly globalised society, most people cannot live their everyday lives without making use of several linguistic varieties; multilingualism by far outweighs monolingualism, and Europe is no exception (see e.g. Lüdi 1996). Due to its role as global contact language (Görlach 2002), many people learn English, and this is certainly true of the Netherlands and Germany. Data from the English Proficiency Index (EPI) shows that the population of the Netherlands has the highest English proficiency in the world of a non-English speaking country, with 90% of the population speaking English as a second language

("EF English Proficiency Index", 2018). The German population also has a very high degree of proficiency in English, ranking 9th. The proficiency in English found in both countries has facilitated the emergence of a diasystem between these languages and English.

This diasystem has functional motivations. English serves various social functions in the Netherlands; Vriesendorp and Rutten (2017) observe that Dutch young gay men sometimes codeswitch to English in their online messaging. This is an identity practise for young gay men in the Netherlands because their gay role models are found in English-speaking entertainment. Further, Lettinga et al. (2017) observe that native speakers of Dutch use English in their text messages, typically to convey familiarity and as a function of lower register. When addressing more senior recipients, English is used much less frequently (ibid.). In Germany, lexical borrowing from English is associated with domains such as advertising and computing (see e.g. Ruhnkehl et al. 1998; Eisenberg 1999, 2001) and the media; English-language newspapers are now commonplace in Germany, and American films constitute a large part of what is shown in German cinemas (Hilgendorf 1996: 4).

As well as serving social functions, Berns (1995: 8-9) observes that in the Netherlands and Germany, English also serves various cultural, commercial and educational functions. English is used in an increasing number of domains in Dutch society, including in the media (see, among others, Claus & Taeldeman 1989; Ridder 1995; van der Sijs 1996; Gerritsen & Nickerson 2004) and advertising (Gerritsen 1995; Gerritsen et al. 2000). De Bot (1994), for instance, remarks that 40-60% of the television programmes shown in the Netherlands are in English. Dutch speakers are also becoming increasingly exposed to English on the Internet, especially younger speakers; more people aged 16-34 use the Internet in the Netherlands than any other group.[45] Gerritsen (1996) found that almost 20% of the pages in Dutch newspapers and magazines contain advertisements in English; he adds that advertisements entirely in English are now a well-established phenomenon in the Netherlands. Korzilius et al. (2006) found that 39% of job advertisements in the Dutch newspaper *de Volkskrant* contained at least one English word, and that 2.4% of the advertisements were completely in English. Job advertisements in English are also commonplace in Germany; Hilgendorf (1996: 10-11) gives various examples of

[45] https://www.statista.com/statistics/348036/daily-internet-usage-age-group-netherlands/.

English being used in German job advertisements in the *Frankfurter Allgemeine Zeitung*. In the area of education, Edwards (2016: 29) points out that English is compulsory in all Dutch high schools. In a small number of Dutch secondary schools, English is even used as the language of instruction, a practice which dates to 1989 (Admiraal et al. 2006: 77). In Germany, though not compulsory, English is the most widely taught second language by a considerable margin, with 94% of secondary school pupils in Germany learning English in 1996/7 (Hilgendorf 2005: 56).

The establishment of a diasystem has led to contact-induced changes in Dutch (van de Velde & Zenner 2010: 52). Ridder (1995) suggests that a 'Dutchlish' hybrid is emerging, though she concedes that the influence of English on Dutch is largely restricted to the lexicon, both at the single-word and phrasal level (see also De Decker & Vandekerckhove 2012; Zenner et al. 2013, 2015). German is also undergoing contact-induced change due to English (cf. Barbour 2005), though this influence is often overstated by German prescriptivists (see e.g. Zimmer 1997, who claims that English is transforming German into a 'different entity'). However, linguists generally agree that the influence of English on German is limited to lexical borrowings.

The following section explores the notion of the diasystem and diaconstructions in the Dutch and German *way*-constructions. I show that the Dutch *way*-construction has undergone pro-diasystematic change, because a diaconstruction has been established in Dutch, namely the incidental activity reading. An incidental activity reading is also emerging in German, but it is far less productive. The Dutch *way*-construction where *weg* is modified by an adjective has also undergone contact-induced constructional changes under the influence of English, with an increasing range of modifiers being used. This is in contrast to the German *way*-construction, where adjectival modification of *weg* is not attested.

7.4 Pro-diasystematic change in the Dutch and German *way*-constructions

7.4.1 The incidental activity reading

This subsection begins by examining the collostructional strength and frequency of verbs in the incidental activity reading in the English *way*-construction. I analyse data from the CLMET3.1, COHA and COCA corpora here; although the temporal range of COCA partially overlaps with that of COHA, COCA contains informal spoken data and it is possible that the incidental activity reading is associated with an informal register. I then analyse the verbs in the incidental activity readings of the Dutch and German *way*-constructions, elucidating the similarities and differences between the three languages and examining the nature of the diaconstructions and the diasystem.

7.4.1.1 English

In the CLMET3.1 corpus, 9 tokens out of 1,862 contained the incidental activity reading, and these 9 tokens contained 9 verb types: *lounge*, *bark*, *splash*, *fumble*, *worry*, *gasp*, *rub*, *purr*, *croon* were each attested once. The degree of collostructional strength of each of these verbs is presented in Table 38 below. Note: a coll.strength value of greater than 2 indicates that the association between the word and construction is statistically significant at $p < 0.01$ (Gries 2014a).

Table 38: Collostructional strength of verbs in the incidental activity reading in CLMET3.1

Verb	Freq. in cxn	relation	coll.strength
croon	1	attraction	2.77525
purr	1	attraction	2.55836
fumble	1	attraction	2.22711
lounge	1	attraction	1.96886
splash	1	attraction	1.95283
bark	1	attraction	1.83974
worry	1	attraction	1.60892
gasp	1	attraction	1.56166
rub	1	attraction	1.34201

A cluster of sound emission verbs has emerged here (cf. Fanego 2017, Perek 2018): *croon*, *purr*, *bark*, *gasp*, which are significantly attracted to the construction at $p < 0.05$.

By the beginning of the 19th century, when the COHA corpus begins, the type frequency of verbs in the incidental activity reading increases considerably, though the token frequency is relatively small. Of the 19,106 tokens of the *way*-construction in COHA, only 104 contained the incidental activity reading (0.54%), though these 104 tokens contained 48 verb types. The 20 most frequent of these and their frequency are presented in Table 39 and the 20 most strongly associated verbs and their degree of association are given in Table 40. A coll.strength value of greater than 3 indicates that the association between the word and construction is statistically significant at $p < 0.001$.

Table 39: 20 most frequently attested verbs in the incidental activity reading in COHA

Verb	Freq	Verb	Freq
eat	8	*puff*	2
fumble	4	*nibble*	2
dance	4	*hiss*	2
drink	4	*joke*	2
smell	3	*whisper*	2
dream	3	*slip*	2
smile	3	*argue*	2
clown	2	*kiss*	2
mumble	2	*laugh*	2
act	2		

Table 40: 20 most strongly associated verbs in the incidental activity reading in COHA

Verb	relation	coll.strength	Verb	relation	coll.strength
eat	attraction	23.57192	*bullshit*	attraction	3.00559
fumble	attraction	8.85487	*smell*	attraction	2.64465
clown	attraction	5.40805	*slosh*	attraction	2.55734
mumble	attraction	3.89664	*dream*	attraction	2.46029
puff	attraction	3.86672	*fart*	attraction	2.40936
nibble	attraction	3.82604	*woof*	attraction	2.25692
hiss	attraction	3.73942	*pout*	attraction	2.19544
joke	attraction	3.15081	*whisper*	attraction	2.08674
burble	attraction	3.09732	*smile*	attraction	2.08515
dance	attraction	3.03369			

Tables 39 and 40 show that the cluster of sound emission verbs has attracted new types: *mumble*, *puff*, *hiss, fart*, *woof*, *whisper* are now attested. A cluster of verbs relating to social interaction has also emerged: *clown*, *joke*, and *smile*.

In the most modern data from COCA, the trend of increasing type frequency of verbs in the incidental activity reading continues, though the token frequency remains relatively stable. Of the 23,672 tokens of the *way*-construction in COCA, 957 contained the incidental activity reading (4.04%), and these 957 tokens contained 166

verb types. The 30 most frequent and most strongly associated of these are presented in Tables 41 and 42 below.

Table 41: 30 most frequent verbs in the incidental activity reading in COCA

Verb	Freq.	Verb	Freq.
eat	73	*nibble*	11
dance	60	*sweat*	11
sing	38	*argue*	10
fumble	36	*smile*	9
drink	35	*chug*	9
laugh	26	*flirt*	9
click	25	*joke*	9
munch	22	*cry*	8
crunch	15	*whisper*	8
puff	15	*scream*	8
act	14	*cough*	7
sniff	12	*whistle*	7
slosh	11	*huff*	6
splash	11	*wheeze*	6
dream	11		

Table 42: 30 most strongly associated verbs in the incidental activity reading in COCA

Verb	relation	coll.strength	Verb	relation	coll.strength
eat	attraction	230.35693	*flirt*	attraction	18.91204
fumble	attraction	106.61552	*sweat*	attraction	17.73045
dance	attraction	101.86698	*wheeze*	attraction	16.49781
munch	attraction	59.87953	*joke*	attraction	16.47345
sing	attraction	49.75263	*huff*	attraction	15.87057
drink	attraction	44.4102	*bullshit*	attraction	15.58432
puff	attraction	39.76714	*dream*	attraction	14.09356
click	attraction	36.65028	*creak*	attraction	14.047
crunch	attraction	34.75985	*grumble*	attraction	14.03912
slosh	attraction	32.48031	*squeak*	attraction	14.01562
laugh	attraction	32.16648	*whistle*	attraction	13.89787
nibble	attraction	26.29761	*clown*	attraction	13.66379
chug	attraction	24.59636	*mumble*	attraction	13.59206
sniff	attraction	24.49453	*blunder*	attraction	13.26044
splash	attraction	23.11928			

Tables 41 and 42 show that the cluster of verbs encoding sound emission remain strongly associated to the *way*-construction, as do the cluster of verbs encoding actions involving the mouth, social interaction, and performance.

7.4.1.2 Dutch

Whereas the *De Gids* corpus contained no tokens of the incidental activity reading, the data from SoNaR and NLCOW14A contain several such tokens; therefore, only data from these corpora is analysed here. In SoNaR, 31 of the 807 tokens (3.84%) of the *way*-construction instantiated the incidental activity reading, and these 31 tokens contained 26 verb types. The token frequencies of these verb types and their translation is illustrated in Table 43, and the collostructional strength of these verbs is shown in Table 44. All of the verbs shown in Table 44 were found to be attracted to the construction.

Table 43: Token frequency of verbs in the incidental activity reading in SoNaR

Verb	Translation	Freq	Verb	Translation	Freq
bijten	'bite'	3	*gokken*	'gamble'	1
dansen	'dance'	3	*huilen*	'howl'	1
toeteren	'toot'	2	*jodelen*	'yodel'	1
surfen	'surf (the internet)'	1	*brullen*	'roar'	1
lezen	'read'	1	*kniezwengelen*	'a dance with the knee'	1
claxonneren	'claxon'	1	*zingen*	'sing'	1
blauwbekken	'shiver'	1	*schreeuwen*	'scream'	1
zoenen	'kiss'	1	*drinken*	'drink'	1
smachten	'sigh'	1	*bikkelen*	'flicker'	1
flirten	'flirt'	1	*trommelen*	'drum'	1
hoesten	'cough'	1	*stompen*	'stomp'	1
plassen	'piss'	1	*blunderen*	'blunder'	1
zuchten	'sigh'	1	*strelen*	'caress'	1

Table 44: Collostructional strength of verbs in the incidental activity reading in SoNaR

Verb	Translation	coll.str.	Verb	Translation	coll.str.
bijten	'bite'	5.9408	*flirten*	'flirt'	2.4031
toeteren	'toot'	5.8928	*hoesten*	'cough'	2.3959
dansen	'dance'	5.2774	*strelen*	'caress'	2.3841
kniezwengelen	'a dance with the knee'	4.7507	*brullen*	'roar'	2.3364
blauwbekken	'shiver'	4.2358	*surfen*	'surf (the internet)'	2.0711
claxonneren	'claxon'	3.5566	*gokken*	'gamble'	2.0467
jodelen	'yodel'	3.4955	*zingen*	'sing'	2.0012
stompen	'stomp'	3.4955	*zuchten*	'sigh'	1.9419
bikkelen	'flicker'	3.1498	*schreeuwen*	'scream'	1.7268
smachten	'sigh'	3.0994	*huilen*	'howl'	1.6160
trommelen	'drum'	2.9573	*drinken*	'drink'	1.0692
blunderen	'blunder'	2.9185	*lezen*	'read'	0.6898
zoenen	'kiss'	2.4516	*zweten*	‘sweat’	0.6653

The data in Tables 43 and 44 show that the emergence of the incidental reading is remarkably similar to that in the English *way*-construction. All of the verbs are attracted to the construction, and their token frequency is very low. The very low token

frequency of these structural innovations supports Backus et al.'s hypothesis (2011: 744) that contact-induced structural innovations typically, of course, have very low token frequency at the initial stages of change. Further, there is a cluster of sound emission verbs: *toeteren* 'toot', *claxonneren* 'claxon', *jodelen* 'yodel', *smachten* 'sigh', *hoesten* 'cough', *brullen* 'roar', *zuchten* 'sigh', *schreeuwen* 'scream', and *huilen* 'howl'. A cluster of verbs relating to performance has emerged, as in the English construction, with *dansen* 'dance', *kniezwengelen* 'dance with the knee' and *zingen* 'sing' being attracted to the construction. There are also two verbs encoding actions involving the mouth (*bijten* 'bite', *zoenen* 'kiss'), which suggests that a cluster of such verbs may be emerging, as in the English construction.

In the NLCOW corpus, the token frequency of the incidental activity has decreased, relatively speaking: only 229 of the 8,680 tokens instantiated the incidental activity reading (2.64%). However, the type frequency of these verbs has increased considerably, with 145 verb types now attested. The 30 most frequent and most strongly attracted of these are shown in Tables 45 and 46 below. All of the verbs shown in Table 41 were found to be attracted to the construction.

Table 45: 30 most frequent verbs in the incidental activity reading in NLCOW14A

Verb	Translation	Freq	Verb	Translation	Freq
branden	'burn'	20	*graaien*	'rummage'	3
puzzelen	'puzzle'	15	*dromen*	'dream'	3
dansen	'dance'	13	*fluisteren*	'whistle'	3
spelen	'play'	11	*janken*	'whine'	3
kussen	'kiss'	11	*krijsen*	'scream'	3
kreunen	'moan'	7	*koken*	'cook'	3
zoenen	'kiss'	5	*spartelen*	'flounder'	3
schrijven	'write'	5	*schreeuwen*	'scream'	3
toeteren	'toot'	4	*vissen*	'fish'	2
zweten	'sweat'	4	*bulldozeren*	'bulldozer'	2
drummen	'drum'	4	*proeven*	'try'	2
blazen	'blow'	4	*ritselen*	'riddle'	2
grazen	'graze'	3	*lezen*	'read'	2
tikken	'tap'	3	*botsen*	'crash'	2
improviseren	'improvise'	3	*brullen*	'roar'	2

Table 46: 30 most strongly attracted verbs in the incidental activity reading in NLCOW14A

Verb	Trans.	coll.str.	Verb	Trans.	coll.str
puzzelen	'puzzle'	33.2589	*swipen*[46]	'swipe'	5.8019
branden	'burn'	29.3661	*keuvelen*	'chat'	5.8019
dansen	'dance'	16.3805	*bloggen*	'blog'	5.8019
kussen	'kiss'	16.2997	*swaffelen*	'hit with penis'	5.8019
kreunen	'moan'	12.0462	*headbangen*	'headbang'	5.8019
bulldozeren	'bulldozer'	11.6039	*grimasen*	'grimace'	5.8019
bikkelen	'flicker'	11.6039	*modereren*	'moderate'	5.8019
quizzen	'quiz'	11.6039	*skypen*	'skype'	5.8019
drummen	'drum'	8.6576	*stumperen*	'stump'	5.8019
toeteren	'toot'	8.6410	*polderen*	'polder'	5.8019
zoenen	'kiss'	8.2065	*frissen*	'freshen'	5.8019
spartelen	'flounder'	6.9470	*schokschouderen*	'shrug shoulders'	5.8019
zweten	'sweat'	6.1687	*smoken*	'smoke'	5.8019
krijsen	'scream'	5.9403	*slebberen*	'sob'	5.8019
vuvuzelaen	'play vuvuzela'	5.8019	*honky-tonken*	'play honky-tonk'	5.8019

The findings illustrated in Tables 45 and 46 support the existence of a cluster of sound emission verbs. As well as the sound emission verbs in the SoNaR corpus remaining attracted to the construction, the cluster has attracted new verbs, such as *krijsen* 'scream', *slebberen* 'sob' and *kreunen* 'moan'. The cluster of verbs encoding performance has also attracted new members, with *honky-tonken* 'play the honky-tonk' and *vuvuzelaen* 'play the vuvuzela' now joining the construction. The cluster of verbs encoding actions involving the mouth has not attracted any new members, but as in the English construction, a cluster of verbs encoding social interaction has emerged: *swipen* 'swipe', *keuvelen* 'chat' and *schokschouderen* 'shrug one's shoulders'.

[46] On a mobile dating app, e.g. Tinder.

7.4.1.3 German

No instances of the incidental activity reading were found in the *Deutsches Textarchiv* (1473-1927) corpus; this suggests that the emergence of the incidental activity reading in German is a relatively recent phenomenon. In the *Berliner Zeitung* corpus (1994-2005), 3 tokens out of 299 instantiated the incidental activity reading (1.00%): *hupen* 'beep' was attested in the construction twice, and *klicken* 'click' was attested once. These two verbs have a collostructional strength of 7.61 and 3.15 respectively, which suggests that they are attracted to the construction, and that this attraction is statistically significant at $p < 0.001$ (Gries 2014a).

The type and token frequency of verbs in the German incidental activity reading undergoes a slight increase in the DECOW16A-NANO corpus; 14 of the 865 tokens of the construction contained the incidental activity reading (1.62%), and these 14 tokens contained 11 verb types. The token frequency of these verb types and their collostructional strength is shown in Table 47 below. All verbs shown were found to be attracted to the construction.

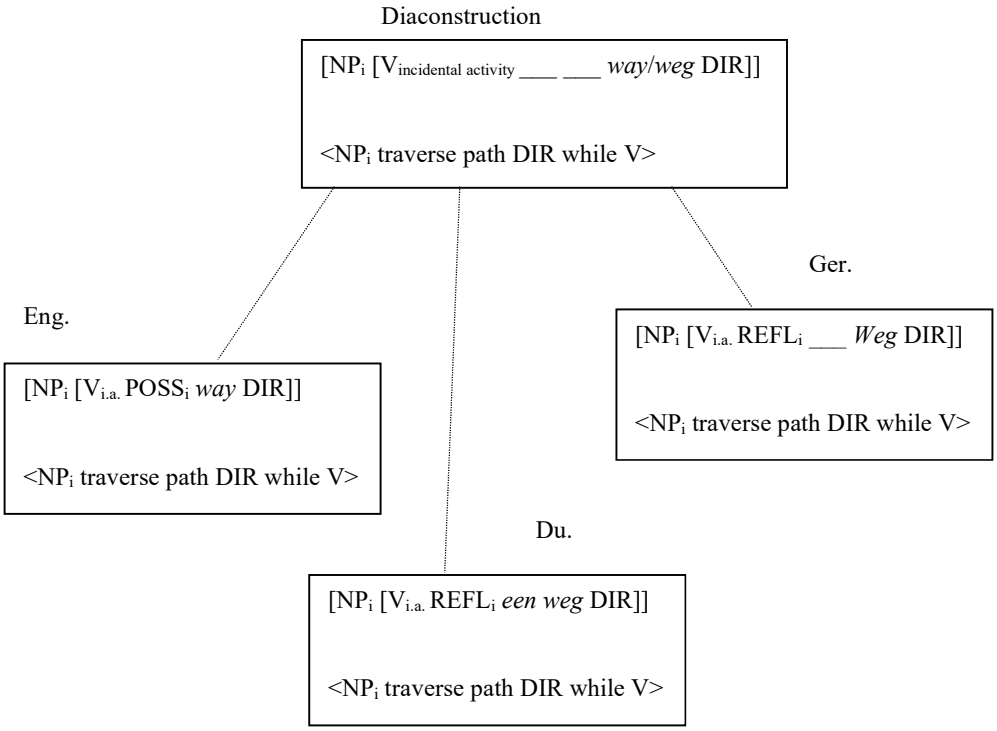
Table 47: Token frequency of verbs in the incidental activity reading in DECOW16A-NANO and their collostructional strength

Verb	Translation	Freq	coll.str.
küssen	'kiss'	3	6.9295
schrammeln	'play Schrammel music'	1	5.9569
bolzen	'kick (a ball) around'	1	5.9569
nippen	'nip'	1	3.3539
lecken	'lick'	1	2.5176
spucken	'spit'	1	2.5029
beleidigen	'insult'	1	2.1836
scheißen	'shit'	1	1.9240
loben	'praise'	1	1.7725
tanzen	'dance'	1	1.7527
schreien	'scream'	1	1.7230

The data in Table 47 in combination with data from the *Berliner Zeitung* corpus suggest a similar development to the incidental activity reading in English and Dutch. There is a cluster of verbs encoding actions involving the mouth (*küssen* 'kiss', *nippen* 'nip', *lecken* 'lick', *spucken* 'spit'), a cluster of verbs encoding performance (*schrammeln* 'play Schrammel music', *tanzen* 'dance') and a cluster of verbs of sound emission and speech (*hupen* 'beep', *schreien* 'scream', *beleidigen* 'insult', *loben* 'praise'). Given the similarity in the range of verbs that can occur in the German and English incidental activity readings, I propose that a contact-induced constructionalization has taken place. This is an instance of pro-diasystematic change; a new diaconstruction has emerged which is underspecified for language.

The emergence of an incidental activity reading in Dutch can also be thought of as an instance of pro-diasystematic change. Bilingual speakers of Dutch perceived a similarity between the incidental activity reading in English, and the means/manner reading in Dutch on the grounds of their semantic similarity; both constructions involve motion along a path, but in the incidental activity reading, the activity denoted by the verb merely accompanies the motion, rather than being the manner in which the entity moves. A diaconstruction has therefore emerged, and this can be thought of as a contact-induced instantaneous constructionalization, supporting Colleman's (2016: 105-106) hypothesis that contact-induced instantaneous constructionalization is not limited to the lexicon. Figure 33 below illustrates the diasystem of the incidental activity reading in English, Dutch, and German. I adopt the following notational conventions from Höder (2012). The meaning and form of a construction are arranged on separate lines. Grammatical meanings are enclosed in angle brackets. Square brackets indicate the form of a construction, with schematic components in small capitals. An underscore indicates an unspecified component; in this case, only the noun phrase, verb and directional argument are common to both the English, Dutch, and German constructions. The Dutch construction uses a reflexive pronoun, indefinite article, and the noun *weg*, while the English construction uses a possessive pronoun and the noun *way*. The German construction uses a reflexive pronoun, article or possessive pronoun, and the noun *Weg*.

Figure 33: Diasystem of the incidental activity reading in English, Dutch, and German



7.4.2 Adjectival modification of *way*/*weg*

Another way in which the English *way*-construction has influenced its Dutch counterpart is in the adjectival modification of *way* or *weg*. I first describe the development of the adjectival modifier slot in the English *way*-construction, and then compare it with the development of the adjectival modifier slot in the Dutch construction.

7.4.2.1 English

In CLMET3.1, 69 of the 1,862 tokens contained an adjectival modifier (3.70%), and 54 adjective types were attested. These either describe the path along which the subject moves (*shortest*, *upward*, *shadowy*), or modify the entire verbal subevent and describe the manner in which the subject moves (*lonely*, *solitary*, *weary*, *stealthy*). The trend of increasing type frequency continues in COHA, with 210 adjective types attested, though the token frequency decreases relatively speaking, with 394 of

19,106 tokens containing an adjectival modifier (2.06%). The 30 most frequently attested of these are presented in Table 48 below.

Table 48: 30 most frequent adjectival modifiers in COHA

Adjective	Freq	Adjective	Freq
slow	29	*quiet*	4
solitary	17	*sluggish*	4
weary	16	*angry*	3
cautious	11	*sinuous*	3
painful	11	*proud*	3
lonely	9	*desperate*	3
leisurely	9	*serpentine*	3
devious	6	*early*	3
swift	6	*careful*	3
homeward	6	*tortured*	3
toilsome	5	*separate*	3
easy	5	*dainty*	3
difficult	5	*peaceful*	3
silent	4	*merry*	3
uncertain	4	*crooked*	3

Most of these adjectives describe the manner in which the subject moves, but adjectives describing the path shape are also attested, e.g. *serpentine*, *crooked*. The most modern data from COCA shows a continuing decrease in token frequency, with only 138 of the 23,672 tokens containing an adjectival modifier. Only two of these were in the spoken genre, which suggests that adjectival modification of *way* is now seen as very formal. These 138 tokens contained 100 adjective types, and all of these adjectives functioned as adverbial modifiers of the entire verbal subevent.

7.4.2.2 Dutch

Data from *De Gids* suggest that the possibility of adjectival modification of *weg* was initially very limited; of the 131 tokens in that corpus, only 11 contained an adjectival modifier (0.83%). These adjectives were: *eigen* ‘own’, *nieuw* ‘new’, *veilig* ‘safe’ and *ander* ‘other’. Data from the SoNaR0 corpus reveals an increase in type and token

frequency thereafter; 23 of the 807 tokens in SoNaR contained an adjectival modifier of *weg* (2.85%), and these tokens contained 19 adjective types. The frequency of these is shown in Table 49.

Table 49: Frequency of adjectival modifiers in SoNaR

Adjective	Translation	Freq	Adjective	Translation	Freq
eigen	'own'	3	*cynisch*	'cynical'	1
juist	'correct'	3	*heel*	'whole'	1
bloederig	'bloody'	1	*lang*	'long'	1
traag	'slow'	1	*oostelijk*	'eastward'	1
muzikaal	'musical'	1	*verkeerd*	'wrong'	1
eng	'scary'	1	*diep*	'deep'	1
vrij	'free'	1	*recht*	'right'	1
sensueel	'sensual'	1	*veilig*	'safe'	1
nieuw	'new'	1	*kalmer*	'calmer'	1
betaalbaar	'affordable'	1			

This shows that adjectives that describe the path itself are acceptable (*juist* 'correct', *eng* 'scary', *lang* 'long', *oostelijk* 'eastward', *verkeerd* 'wrong', *diep* 'deep'). The remainder of the adjectives, as in the English construction, describes the manner in which the subject moves.

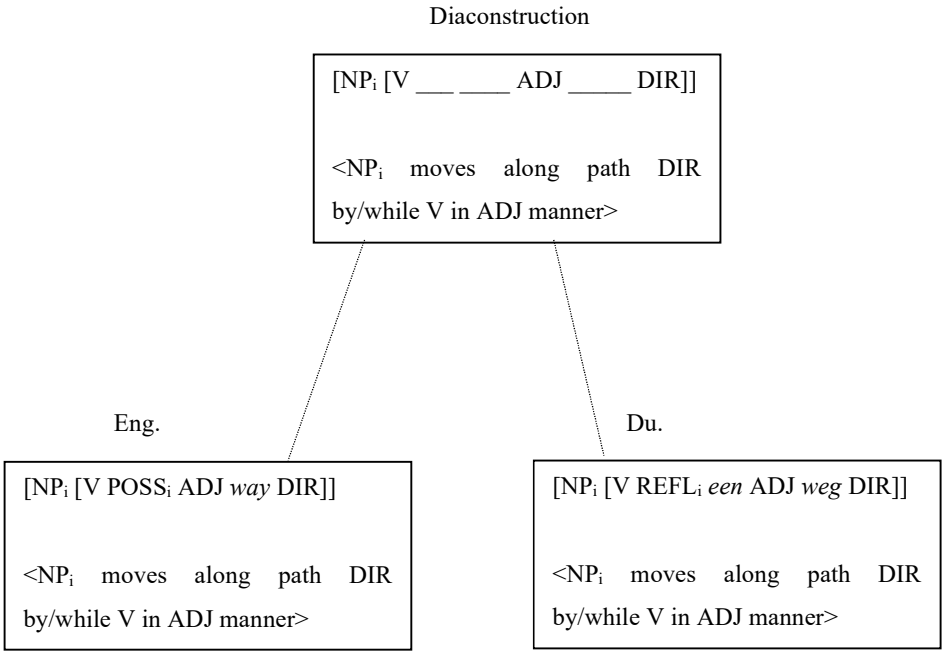
The most modern data from the NLCOW corpus shows a slight increase in the token frequency of adjectival modifiers. Of the 8,680 tokens, 282 contained an adjectival modifier (3.25%). The type frequency also increased to 62, though it is worth noting that NLCOW is ten times bigger than SoNaR, and that this difference in size may account for the higher type frequency. The 20 most frequently attested adjectival modifiers of *weg* are presented in Table 50.

Table 50: 20 most frequent adjectival modifiers in NLCOW14A

Adjective	Translation	Freq	Adjective	Translation	Freq
ander	'other'	81	*korter*	'shorter'	3
nieuw	'new'	42	*verharde*	'hardened'	3
lang	'long'	41	*recht*	'right'	3
eigen	'own'	22	*begaanbaar*	'passable'	3
bepaald	'determined'	8	*smal*	'narrow'	2
druk	'busy'	5	*breed*	'wide'	2
verkeerd	'wrong'	4	*gevaarlijk*	'dangerous'	2
moeilijk	'difficult'	4	*moeizaam*	'laborious'	2
veilig	'safe'	4	*heel*	'whole'	2
goed	'good'	4	*vertikaal*	'vertical'	2

The NLCOW data supports the hypothesis that the adjective may describe the path, or modify the entire verbal subevent. Overall, the data on adjectival modifiers in English and Dutch shows that Dutch is at a different stage of the development than English. Between the 18th and 20th centuries, the type and token frequency of adjectival modifiers increased considerably, and then began to decrease; the Dutch construction has not yet undergone this decrease. Further, instances in English where the adjective describes the path are no longer grammatical, but Dutch does not show this restriction. The *way*-construction with an adjectival modifier where the adjective modifies the entire verbal subevent may be thought of as a diasystem with the form shown in Figure 34. In this case, the unspecified elements are the possessive/reflexive pronoun, the indefinite article in Dutch, and the noun *way* or *weg*.

Figure 34: Diasystem of the way-construction with adjectival modifier in English and Dutch



7.4.2.3 German

Adjectival modification of *Weg* was not attested in any of the German corpora investigated.

7.4.3 Other evidence of the influence of English on the Dutch and German *way*-constructions

The Dutch *way*-construction contains a number of verbs of English origin that were absent from the construction prior to when the contact situation with English began to take hold. In the SoNaR corpus, the attested verbs of English origin are *klikken* (4 times), *bluffen* (3 times), *racen*, *surfen*, and *flirten* (all once each). All of these verbs are significantly attracted to the construction at $p < 0.01$, as Table 51 shows.

Table 51: Collostructional strength of verbs of English origin in SoNaR

Verb	Freq	relation	coll.strength
bluffen	3	attraction	10.30527
klikken	4	attraction	8.83981
racen	1	attraction	2.5648
flirten	1	attraction	2.40308
surfen	1	attraction	2.07106

The NLCOW corpus contained 32 verbs of English origin, which accounted for 71 of the 8,680 tokens (0.82%). This higher type and token frequency may be explained in part by the fact this corpus contains only data from the Internet, where exposure to English is very high. The token frequency and collostructional strength of these verbs are shown in Table 52.

Table 52: Token frequency of verbs of English origin in NLCOW14A and their collostructional strength

Verb	Freq	relation	coll.str.	Verb	Freq	relation	coll.str.
puzzelen	15	attraction	33.2589	*skypen*	1	attraction	2.2682
bluffen	11	attraction	31.5845	*slurpen*	1	attraction	2.0482
meanderen	7	attraction	23.6075	*surfen*	2	attraction	1.7692
zigzaggen	4	attraction	10.0213	*lobbyen*	1	attraction	1.7022
bulldozeren	2	attraction	8.1042	*flirten*	1	attraction	1.6697
smashen	2	attraction	6.6870	*kicken*	1	attraction	1.5750
bodyslammen	1	attraction	5.8019	*sprinten*	1	attraction	1.5266
honky-tonken	1	attraction	5.8019	*pokeren*	1	attraction	1.5018
walljumpen	1	attraction	5.0238	*racen*	1	attraction	1.1236
quizzen	2	attraction	4.2259	*racen*	1	attraction	1.1236
googelen	2	attraction	4.1573	*counteren*	1	attraction	1.1031
smoken	1	attraction	3.1636	*bloggen*	1	attraction	1.0407
headbangen	1	attraction	2.9442	*coachen*	1	attraction	0.9008
swipen	1	attraction	2.9095	*interviewen*	1	attraction	0.8725
framen	1	attraction	2.8356	*printen*	1	attraction	0.5284
swingen	2	attraction	2.6920	*trainen*	1	attraction	0.4193

In all of the German corpora, only one verb of English origin was found, *klicken*. This verb is significantly attracted to the construction, with a collostructional strength of 3.15. This shows that speakers of German have a greater reluctance to borrow from English than speakers of Dutch, and one possible reason for this may be that they wish to keep the language 'pure', because the German language is a core part of their German identity (cf. Barbour 2005).

A further piece of evidence of English influence is that speakers of Dutch have extended the context of *maken* to the *way*-construction, on the basis of English *make one's way*, as in the following example, where *banen* would usually be preferred.

(221) Hij maakt zich een weg door het menselijk lichaam
He makes REFL a way through the human body
'He makes his way through the human body' (NLCOW14A)

German has also undergone this context extension, using *machen* where *bahnen* would be preferred, as (222) shows.

(222) Mache dir den Weg zu deinem Platz
Make REFL the way to your seat
'Make your way to your seat' (DECOW).[47]

Thomason (2001) defines a contact-induced change as a change that would have been less likely to occur if a contact situation had not taken place; the emergence of the incidental activity reading in the Dutch and German *way*-construction fits this definition. If we say that the incidental activity reading was not borrowed into Dutch and German, we would have to say that the same sub-construction emerged in the three languages by very different pathways, which is a less plausible explanation than a contact-based account of this change. The Dutch *way*-construction was not very productive before contact with English, with only a handful of verbs describing the manner of motion or force attested, as well as *banen*. The English construction, on the other hand, was much more productive before the emergence of the incidental activity reading, with hundreds of verbs in several semantic domains. Further, there

[47] Stefan Hartmann (pers. comm.) points out that this could be a typo of the idiom *sich auf den Weg machen* 'make one's way'.

is no plausible analogical source of the incidental activity verbs in Dutch. The German construction was also very unproductive before contact with English and the emergence of the incidental activity reading, with only a few verbs describing the manner of motion and verbs of paving and clearing a path in a precursor construction; these cannot plausibly be said to be the analogical source of the incidental activity reading in German.

In sum, it was shown in this section that the English *way*-construction has influenced the Dutch and German constructions considerably. The incidental activity reading has emerged in Dutch and German, with semantic clusters of verbs identical to those found in the incidental activity reading of the English construction: performance, sound emission, bodily functions, and human speech. The English construction has influenced the Dutch construction because it has caused the range of adjectival modifiers of the noun to increase considerably. However, unlike the English construction, these may describe the path itself, as well as the manner in which the subject moves. Several verbs of English origin were found to be strongly attracted to the Dutch construction, which supports the hypothesis that the Dutch incidental activity reading is an instance of contact-induced constructionalization. Only one verb of English origin was attested in the German data, but this was found to be strongly attracted to the construction. The Dutch and German verbs *maken*/*machen* 'make' were shown to have undergone context extension, now occurring in the *way*-construction where *banen*/*bahnen* would be preferred. I therefore argue that the case outlined here for contact-induced constructionalization here is a convincing one.

7.5 Conclusions

This chapter of the thesis has shown that English has influenced the Dutch and German languages considerably, to the point where English goes beyond being an L2 variety for speakers in those countries; in both countries, English serves various social, commercial, and educational functions. While the influence of English on both languages has been considerable, Dutch shows greater influence. It was hypothesised that this was due to various factors. Firstly, English is compulsory in Dutch secondary schools, which explains the higher proficiency in English in the Netherlands compared to Germany. Secondly, borrowing from English is discouraged in German, and German utterances with English lexical borrowings are pejoratively

referred to as “Denglisch” (from *Deutsch* ‘German’ + *Englisch*). Thirdly, German speakers’ reluctance to borrow from English may be due to wanting to keep the German language ‘pure’ as it is a core part of their German national identity, as Barbour (2005) claims.

It was demonstrated that rather than separating out syntactic and lexical borrowing, both should be united under the umbrella of constructional borrowing, because the entirety of a speakers’ knowledge is made up of constructions (Goldberg 2006: 5). A diasystematic Construction Grammar approach was proposed to account for the constructicon of multilingual speakers; the language-learning process of categorisation and generalisation is not sensitive to language barriers, which suggests that a multilingual speaker’s constructicon contains constructions of multiple languages. The process of cross-linguistic categorisation was shown to lead to the establishment of a diasystem which contains diaconstructions which are unspecified for language.

It was further shown that the developments in the Dutch and German *way*-constructions can be thought of as an instance of pro-diasystematic change. In the case of Dutch, this is because an incidental activity reading has emerged, whose trajectory is remarkably similar to that of the English incidental activity reading. This suggests the possibility of an instance of contact-induced instantaneous constructionalization, which is not limited to the lexicon (cf. Colleman 2016). The development of the adjectival modifier slot in the Dutch *way*-construction also mirrors the development of its English counterpart between the 18th and 20th centuries, and the Dutch *way*-construction now contains several verbs of English origin. German has also gained an incidental activity reading of the *way*-construction, but it is far less productive than in Dutch and English. However, like Dutch, it has also undergone context extension of *machen* ‘make’ in the construction. The following chapter concludes the thesis by answering the research questions posed in Section 1.1.

This page intentionally left blank

8. Conclusions

8.1 Summary of results

The following subsections summarise the main findings for each language, and answer the research questions posed in Section 1.1, which are repeated here for convenience. These research questions are answered of the English, Dutch, and German *way*-constructions in Sections 8.1.1, 8.1.2, and 8.1.3 respectively.

a) What does the *way*-construction look like in the three languages today, and what was the diachronic development of the construction in each language?
b) What can corpora and statistical analysis of corpus data tell us about the development of the English, Dutch, and German *way*-construction?
c) What led to the constructionalizaiton of the *way*-construction in each language, and what post-constructionalization constructional changes took place in each language?
d) What was the role of analogization and analogical extension in the development of the *way*-construction in the three languages?
e) Is there evidence for frequency effects and exemplar representation of language in the *way*-construction in any of the three languages?
f) What is the role of borrowing in the Dutch and German *way*-construction, and how can a constructional theory of language accommodate this borrowing?

8.1.1 English

Traugott and Trousdale (2013: 79-80) identify two precursors to the *way*-construction: one transitive, the other intransitive. This intransitive construction could occur with most verbs encoding forward progress, such as *go*, *wend* or *ride*. Based on data from the Helsinki corpus, I found that this intransitive construction emerged between 1250 and 1350 (*He ete and dranc and went his wai*). In this construction, plural marking on the noun *way* is also attested (*I toke my waggon and wente my wayes*). This construction is also attested in my data with a preposition (in phrases such as *on*

way), and with an article instead of a possessive pronoun. In the transitive construction, which is found as early as Old English, the verb used was typically a verb of acquisition such as *niman* 'take'; I did not find any such examples in my data, although the Old English corpus I used (Helsinki) was very small. I showed that the constructionalization of the English *way*-construction was facilitated by the neoanalysis of *way* from a referential to non-referential object which functioned as a bleached syntactic marker of the construction.

Since its constructionalization, the *way*-construction has undergone several post-constructionalization constructional changes. For one, the *way*-construction has undergone a steep increase in both type and token frequency between 1550 and 1700. Croft and Cruse (2004: 292) hypothesise that an increase in token frequency of a construction leads to its storage as a conventionalised unit. This hypothesis is borne out by the data; by 1650, the noun *way* in the transitive subschema had been neoanalysed as a non-referential object; examples from ARCHER show verbs of fighting such as *fight*, *force*, and *batter* being used in the construction, where *way* cannot plausibly be the object of the verb. This suggests that the noun *way* at that point functioned as a meaningless syntactic marker of the construction. Croft and Cruse's hypothesis is further supported by the fact that a possessive pronoun became obligatory during this period, and plural marking on the noun *way* was lost. This suggests that this increase in token frequency led to the conventionalisation of the construction [NP_i [V $POSS_i$ *way* DIR]].

One side effect of increased type frequency of a construction is its ability to be readily applied to new items (Bybee & Beckner 2010: 841). This is observed in my data between 1550 and 1700. As well as the verb of motion *go*, which also underwent a sharp increase in token frequency in this period, new verbs of motion emerged by analogy; these include *walk*, *wend* and *travel*. The construction was again extended by analogization to include verbs encoding laborious motion, such as *plough*, *trudge*, and *plod*.

The increase in type and token frequency continues in the CLMET3.1 corpus (1710-1920). The data shows that a number of verbs of path creation emerged in this period, such as *trace* and *cut*. The cluster of verbs encoding fighting that emerged between 1550 and 1700 expanded by analogization; *shoot*, *wrestle* and *battle* were attested

after 1700. Perek (2018: 76) proposes that new subtypes of verbs in the *way*-construction were created by a long series of analogical extensions. This hypothesis is confirmed by the data. After 1880, the following pathway of new verb clusters emerged: verbs related to ingestion > olfaction > actions involving the mouth. This series of changes is not surprising; all of these actions involve the body in some way, and the construction extends by analogization to include increasingly more specific verbs (e.g. *eat* > *gnaw*). Two unrelated clusters of verbs emerged at around the same time, one containing verbs related to commerce and finance (*buy*, *borrow*), the other containing verbs encoding deceit or misconduct (*bully*, *cheat*, *bribe*). This again supports Bybee and Beckner's (2010: 841) hypothesis that an increase in type frequency of a construction leads to its application to new items.

Between 1710 and 1920, I found a considerable increase in type and token frequency of verbs relating to motion. Increasingly more precise verbs of motion were used; these include verbs of clumsy motion (*stumble*), verbs of rapid motion (*speed*), and verbs of walking (*shuffle*, *walk*). In addition, a new subschema of the way construction emerged, in which the verb encodes an activity that is not causally related to the motion. These are mostly verbs of sound emission (*bark*, *gasp*), though I find one token of a verb of thought (*He had worried his way to the top of his profession*). After 1920, this subschema of incidental activity verbs increases in type and token frequency. These include verbs of performance (*sing*, *dance*) and verbs of social interaction (*chat*, *smile*).

In sum, the English *way*-construction has undergone a long series of analogical extension from 1250 to the present day, as Israel (1996: 217) claims, such that there are now three distinct subschemas of the construction: one containing verbs encoding the manner of motion, one containing transitive verbs, and one containing verbs encoding actions accompanying motion. This is possible because the noun *way* has been neoanalysed as a non-referential object, and now functions as a highly bleached syntactic marker of the construction. This neoanalysis did not consist of a rebracketing or a strict relabelling, as many traditional accounts of the concept claim are characteristic of this process; the constituency structure of the construction has not changed, and the noun *way* has undergone decategorialisation rather than relabelling; although it is highly bleached, it retains some nominal properties in that it can appear after a possessive pronoun. It was also shown that, while many accounts

of analogical extension focus on the application of a rule to new items, analogical extension can also lead to the expansion of a construction's collocational preferences, as has happened here. Frequency effects were found in the English *way*-construction; the analogical extension of an exemplar cloud usually proceeded on the basis of the most frequently attested item. Genre effects were also found in the English construction.

8.1.2 Dutch

The Dutch construction has undergone similar changes to the English construction. Between 1830 and 1890, the verb used in the construction was exclusively *banen* (when *banen* appears in the construction, it is roughly equivalent to English *make's one way*). By 1890, a cluster of verbs encoding fighting and path clearing emerges, such as *vechten* 'fight', *kappen* 'cut' and *graven* 'dig'. This development mirrors the development of the English *way*-construction in that the noun *weg* 'way' has been neoanalysed as a non-referential object; this led to the constructionalization of the Dutch *way*-construction.

Several post-constructionalization constructional changes have taken place in the Dutch construction, such as host-class expansion; a cluster of body-internal motion verbs also emerges by around 1890, including *kronkelen* 'wriggle' and *wurmen* 'worm'. After 1950, the cluster of fighting verbs expands by analogy; *hakken* 'hack', *knokken* 'knock' and *slaan* 'slay' were attested in the construction at around this time. Contrary to Verhagen (2007), verbs encoding an incidental activity are now widely attested in the construction; given Verhagen's data was collected only 20 years ago, this suggests that the emergence of the incidental activity reading in Dutch is a very recent change. I hypothesise that this is as a result of borrowing and cross-linguistic analogization from the English *way*-construction. This hypothesis is supported by three pieces of evidence. Firstly, the overwhelming majority of these verbs are attested in the Internet era, when Dutch speakers' exposure to English language media increased. Secondly, the verb *maken* 'make' has sharply increased in frequency; before 1950, the verb *maken* was not attested at all, and it was attested only once in the SoNaR corpus (1954-2011). Finally, many of these incidental activity verbs are of English origin; these include *quizzen* 'quiz', *pokeren* 'play poker', *headbangen* 'headbang'. Aside from these verbs of English origin, there is also a

considerable cluster of verbs of sound emission, as in the English construction, such as *huilen* 'howl', *toeteren* 'toot', and *claxonneren* 'claxon'. The development of the Dutch *way*-construction illustrate the role of neoanalysis and analogy in language change; the noun *weg* in the construction has been neoanalysed as a non-referential object, which has facilitated a long series of local analogical extensions in the construction. As in the English construction, this neoanalysis did not consist of a rebracketing or relabelling, because there has been no change in constituency structure. I propose that decategorialisation took place, rather than relabelling, because some nominal properties of *weg* are retained in that it can be selected by a determiner or modified by an adjective. Frequency effects were also found in the Dutch *way*-construction; the analogical extension of an exemplar cloud usually proceeded on the basis of the most frequently attested item. However, the development of the English *way*-construction cannot be attributed to frequency effects alone; genre effects were also found.

8.1.3 German

The development of the German *way*-construction is similar in many ways to the development of its English and Dutch equivalents. In Middle German, there are intransitive and transitive precursors to the construction; the former with *gehen* 'go', the latter with *bahnen* (equivalent to English *make*) or *nehmen* 'take' (cf. the Old English construction with *niman*). As in the Dutch construction, I hypothesise that the constructionalization of the modern-day German *way*-construction arose as a blend of the transitive construction and a reflexive construction, which led to the conventionalisation of the notion of path traversal (see Kramer 2002 for an account of this development in the Dutch construction). This was facilitated by the neoanalysis of *Weg* as a meaningless syntactic marker of the construction.

The German *way*-construction was very unproductive until the mid-1990s. Data from the Berliner Zeitung corpus (1994-2005) shows the emergence of a cluster of verbs encode path clearing (*ebnen* 'pave', *glätten* 'smooth'), verbs of fighting (*freikämpfen* 'fight free', *freischießen* 'shoot free'), and incidental activity verbs (*klicken* 'click', *hupen* 'beep'). This development is consistent with the neoanalysis of the noun *Weg* as a non-referential object, mirroring the developments in the Dutch and English constructions. The development differs, however, in that metaphorical paths are far

less frequent in the German constructions than in its English and Dutch equivalents. This supports Ludwig's (2005) claim that metaphorical paths are more strongly associated with the reflexive pattern [NP_i V [$REFL_i$ DIR]]. As is the case of the English and Dutch constructions, there has been a long series of local analogical extensions in the construction. In addition, I hypothesise that there has been cross-linguistic analogization and borrowing from the English construction; several semantic clusters of verbs are now attested which cannot plausibly have arisen by analogization with other German verbs. Given that the rapid increase in productivity took place in the Internet era, when German speakers' exposure to English was increased, and that the pathways of change are rather different in the English construction, borrowing is a plausible explanation for the synchronic state of the German *way*-construction. The development of the German *way*-construction can also be attributed to neoanalysis and analogy. After the noun *Weg* was neoanalysed, a long series of analogical extensions took place, which considerably expanded the collocational range of the construction. As in the English and Dutch constructions, this neoanalysis was neither a rebracketing nor relabelling, because the constituency structure of the construction has not changed, and the noun *Weg* retains nominal properties in that it can be selected by a determiner. Frequency effects were also found in the development of the German *way*-construction, as the analogical extension within semantic categories often had the most frequently attested item as its locus.

8.1.4 Theoretical implications of the thesis

The novel combination of corpora from a wide range of genres used in this study have allowed me to answer how and when the English, Dutch, and German *way*-constructions have changed over time, from Old English and Middle German to the present day; this question could not have been answered without using corpora. The thesis has also shown that neoanalysis and analogy have worked in tandem in the development of the *way*-construction in each language; in all three languages, neoanalysis of the noun *way*/*weg* as a meaningless syntactic tag of the construction facilitated analogical change in the construction in all three languages. I also found that frequency effects are present in the development of the *way*-construction in all three languages, although there are genre effects in the English *way*-construction which cannot plausibly be explained in terms of frequency effects alone. An important theoretical finding was made, namely that raw frequency is not a crude a tool as is

sometimes assumed. Convincing evidence of borrowing in the Dutch and German *way*-constructions was presented in Chapter 7. It was also shown in this chapter that the concept of borrowing can be integrated into a Construction Grammar framework, as Höder (see especially Höder 2012) has done.

8.2 Possibilities for future research

A significant problem that arose during the investigation was the lack of information relating to the year of publication of the text. This problem particularly affected the Dutch data; this information was missing from the SoNaR and NLCOW corpora, and these corpora yielded the vast majority of the tokens of the Dutch *way*-construction. This meant that a precise chronology of the development of the Dutch *way*-construction could not be established. The same problem affected the German data to a lesser extent, with information about the year of publication missing from the DECOW corpus, although it can be inferred by the nature of this corpus that it contains very modern data.

A further problem was that the diachronic corpus of German that was chosen (*Deutsches Textarchiv*, ca. 213 million words) yielded very few tokens of the German *way*-construction. Future work on the diachrony of the German *way*-construction would benefit from using a larger corpus or combination of corpora.

This thesis has revealed that there is much promising future work to be done on constructional contamination in the *way*-construction in all three languages. I only investigated one contaminating construction in English, the FRR; an investigation into other potential contaminating constructions in other languages could reveal that there is constructional contamination affecting the *way*-construction in those languages. Contaminating constructions in Dutch and German were not investigated at all due to the vast number of tokens of the reflexive construction in these languages; were this investigation to be extended to the reflexive construction, the possibility of constructional contamination could be tested, and the similarities and differences between the two constructions explored in more detail.

Bibliography

Corpora used

ARCHER = A Representative Corpus of Historical English Registers

Berliner Zeitung = Corpus of articles from the newspaper of the same name, 1994-2005

BNC = British National Corpus

CLMET3.1 = Corpus of Late Modern English Texts version 3.1

COHA = Corpus of Historical American English

COCA = Corpus of Contemporary American English

DECOW16A-NANO = German corpus from the web, nano edition

De Gids = Corpus taken from the literary periodical of the same name, 1837-1936

Helsinki = The Helsinki Corpus

NLCOW14A = Dutch corpus from the web

SoNaR = *Stevin Nederlandstalig Referentie corpus* 'Stevin Dutch language reference corpus'

Other works cited

Aarts, B. (2004). Modelling linguistic gradience. *Studies In Language*, *28*(1), 1-49. http://dx.doi.org/10.1075/sl.28.1.02aar

Aarts, B. (2007). *Syntactic Gradience: The Nature of Grammatical Indeterminacy*. Oxford: Oxford University Press.

Abbot-Smith, K., & Behrens, H. (2006). How Known Constructions Influence the Acquisition of Other Constructions: The German Passive and Future Constructions. *Cognitive Science*, *30*(6), 995-1026. doi: 10.1207/s15516709cog0000_61

Admiraal, W., Westhoff, G., & de Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English 1. *Educational Research And Evaluation*, *12*(1), 75-93. doi: 10.1080/13803610500392160

Aikhenvald, A. (2002). *Language contact in Amazonia*. Oxford: Oxford University Press.

Aikhenvald, A. (2006). Grammars in contact: A cross-linguistic perspective. In A. Aikhenvald & R. Dixon, *Grammars in Contact: A Cross-Linguistic Typology* (pp. 1-66). Oxford: Oxford University Press.

Aitchison, J. (1987). *Words in the Mind. An Introduction to the Mental Lexicon*. New York: Basil Blackwell.

Alegre, M., & Gordon, P. (1999). Frequency Effects and the Representational Status of Regular Inflections. *Journal Of Memory And Language*, *40*(1), 41-61. doi: 10.1006/jmla.1998.2607

Allen, C. (1995). *Case marking and reanalysis*. Oxford: Clarendon Press.

Alvanoudi, A. (2017). Language contact, borrowing and code switching. *Journal Of Greek Linguistics*. doi: 10.1163/15699846-01701001

Andersen, H. (2001). Actualization and the (uni)directionality. In H. Andersen, *Actualization: Linguistic Change in Progress [Current Issues in Linguistic Theory 219]* (pp. 225-248). Amsterdam/Philadelphia: Benjamins.

Andersen, H. (2006). Synchrony, diachrony, and evolution. In O. Thomsen, *Different Models of Linguistic Change* (pp. 59-91). Amsterdam/Philadelphia: Benjamins.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89(4),* 369–406.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal Of Memory And Language*, *62*(1), 67-82. doi: 10.1016/j.jml.2009.09.005

Arppe, A., Gilquin, G., Glynn, D., Hilpert, M., & Zeschel, A. (2010). Cognitive Corpus Linguistics: five points of debate on current theory and methodology. *Corpora*, *5*(1), 1-27. http://dx.doi.org/10.3366/cor.2010.0001

Aslin, R., & Newport, E. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions In Psychological Science*, *21*(3), 170-177. doi: 10.1177/0963721412436806

Aston, G., & Burnard, L. (1998). *The BNC handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Asudeh, A., Dalrymple, M. and Toivonen, I. (2008). Constructions with Lexical Integrity: Templates as the lexicon-syntax interface. In Miriam Butt and Tracy Holloway King (eds.) In *Proceedings of the LFG08 Conference* (pp. 68–88). CSLI Publications.

Baayen, H. (2011a). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, *5*(3), 436-461. doi: 10.1075/ml.5.3.10baa

Baayen, R. (2011b). Corpus linguistics and naive discriminative learning. *Revista Brasileira De Linguística Aplicada*, *11*(2), 295-328. doi: 10.1590/s1984-63982011000200003

Baayen, R., Dijkstra, T., & Schreuder, R. (1997). Singulars and Plurals in Dutch: Evidence for a Parallel Dual-Route Model. *Journal Of Memory And Language*, *37*(1), 94-117. doi: 10.1006/jmla.1997.2509

Baayen, R., Wurm, L., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon*, *2*(3), 419-463. doi: 10.1075/ml.2.3.06baa

Backus, A., Seza Doğruöz, A., & Heine, B. (2011). Salient stages in contact-induced grammatical change: Evidence from synchronic vs. diachronic contact situations. *Language Sciences*, *33*(5), 738-752. doi: 10.1016/j.langsci.2011.02.004

Barbour, S. (2000). Germany, Austria, Switzerland, Luxembourg: The Total Coincidence of Nations and Speech Communities?. In S. Barbour & C. Carmichael, *Language and Nationalism in Europe* (pp. 151-168). Oxford: Oxford University Press.

Barbour, S. (2005). Contemporary English influence on German—a perspective from linguistics. In G. Anderman & M. Rogers, *In and Out of English: For Better, For Worse?* (pp. 153-160). Multilingual Matters Ltd.

Bard, E., Robertson, D., & Sorace, A. (1996). Magnitude Estimation of Linguistic Acceptability. *Language*, *72(1)*, 32-68.

Barðdal, J. (2008). *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: Benjamins.

Barðdal, J., & Gildea, S. (2015). Diachronic Construction Grammar: Epistemological context, basic assumptions and historical implications. In J. Barðdal, S. Gildea & L. Sommerer, *Diachronic Construction Grammar* (pp. 1-50). Amsterdam: Benjamins.

Beckner, C., & Bybee, J. (2009). A Usage-Based Account of Constituency and Reanalysis. *Language Learning*, *59*(s1), 27-46. http://dx.doi.org/10.1111/j.1467-9922.2009.00534.x

Behrens, H., & Pfänder, S. (2016). *Experience Counts: Frequency Effects in Language*. Berlin: De Gruyter.

Bell, A., Brenier, J., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal Of Memory And Language*, *60*(1), 92-111. doi: 10.1016/j.jml.2008.06.003

Bergen, B., & Chang, N. (2005). Embodied Construction Grammar in simulation-based language understanding. In J. Östman & M. Fried, *Construction Grammar(s):*

Cognitive and Cross-Language Dimensions (pp. 147-190). Amsterdam: John Benjamins.

Berns, M. (1995). English in the European Union. *English Today*, *11*(3), 3-11. doi: 10.1017/s0266078400008348

Biber, D. (1993). Representativeness in Corpus Design. *Literary And Linguistic Computing*, *8*(4), 243-257. http://dx.doi.org/10.1093/llc/8.4.243

Biber, D., Finegan, E., & Atkinson, D. (1994). ARCHER and its challenges. Compiling and exploring a representative corpus of historical English registers. In U. Fries, G. Tottie & P. Schneider, *Creating and using English language corpora* (pp. 1-14). Amsterdam: Rodopi.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Bisang, W. (2001). Areality, grammaticalization and language typology. On the explanatory power of functional criteria and the status of Universal Grammar. In W. Bisang, *Language Typology and Universals* (pp. 175-223). Berlin: Akademie-Verlag.

Blevins, J., & Blevins, J. (2009). Introduction. In J. Blevins & J. Blevins, *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press.

Blumenthal-Drame, A. (2012). *Entrenchment in Usage-Based Theories. What Corpus Data Do and Do Not Reveal About The Mind*. Berlin: Mouton de Gruyter.

Boas, H., & Sag, I. (2012). *Sign-based construction grammar*. Stanford (California): CSLI.

Boye, K., & Harder, P. (2012). A usage-based theory of grammatical status and grammaticalization. *Language*, *88*, 1-44.

Brinton, L. (1988). *The development of English aspectual systems: Aspectualizers and post-verbal particles*. Cambridge: Cambridge University Press.

Bruter, M. (2005). *Citizens of Europe? The emergence of a mass European identity*. Bastingstoke: Palgrave Macmillan.

Burnard, L. (2007). *Reference Guide for the British National Corpus (XML Edition)*. Oxford: Research Technologies Service at Oxford University Computing Services.

Bybee, J. (1998). The emergent lexicon. *Chicago Linguistic Society*, *34*, 421-435.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation And Change*, *14*(3), 261-290. doi: 10.1017/s0954394502143018

Bybee, J. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language*, *82*(4), 711-733. doi: 10.1353/lan.2006.0186

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Bybee, J. (2010a). Diachronic linguistics. In D. Geeraerts & H. Cuyckens, *The Oxford Handbook of Cognitive Linguistics* (pp. 945-987). Oxford: Oxford University Press.

Bybee, J. (2010b). *Language, usage and cognition*. Cambridge: Cambridge University Press.

Bybee, J. (2013). Usage-based theory and exemplar representations of constructions. In T. Hoffmann & G. Trousdale, *The Oxford handbook of construction grammar* (pp. 49-69). Oxford: Oxford University Press.

Bybee, J., & Beckner, C. (2010). Usage-based Theory. In B. Heine & H. Narrog, *The Oxford Handbook of Linguistic Analysis* (pp. 827-856). Oxford: Oxford University Press.

Bybee, J., & Beckner, C. (2015). Language use, cognitive processes and linguistic change. In C. Bowern & B. Evans, *The Routledge Handbook of Historical Linguistics* (pp. 503-518). London: Routledge.

Bybee, J., & Eddington, D. (2006). A Usage-based Approach to Spanish Verbs of 'Becoming'. *Language*, *82*(2), 323-355. doi: 10.1353/lan.2006.0081

Bybee, J., & Thompson, S. (2000). Three Frequency Effects in Syntax. *Annual Meeting Of The Berkeley Linguistics Society*, *23*(1), 378. doi: 10.3765/bls.v23i1.1293

Caldwell-Harris, C., Berant, J., & Shimon, E. (2012). Entrenchment of phrases with perceptual identification, familiarity ratings and corpus frequency statistics. In D. Divjak & S. Gries, *Frequency Effects in Language Representation: Volume 2* (pp. 165-194). Berlin: Mouton de Gruyter.

Callies, M. and Szczesniak, K. (2008). 'Europa meisterschaft zu erdribbeln. Manner of obtainment constructions in sports reporting'. In E. Lavric, G. Pisek, A. Skinner, and W. Stadler (eds), *The Linguistics of Football*. (Language in Performance 38). Tübingen: Narr (pp. 23–34).

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton and Co.

Christie, E. (2011). Investigating the differences between the English way-construction and the fake reflexive resultative construction. In Armstrong, L. (ed.), *2011 CLA Conference Proceedings*, ISSN 0027-9633

Claus, P., & Taeldeman, J. (1988). De infiltratie van Engelse leenwoorden in het Nederlands en in Nederlandse woordenboeken. In S. Theissen & J. Vromans, *Een Bundel Opstellen Aangeboden aan Joseph Moors ter Gelegenheid van Zijn 75ste Verjaardag* (pp. 11-30). Liège: CIPL.

Clear, J. (1992). Corpus sampling. In G. Leitner, *New Directions in English Language Corpora* (pp. 21-31). Berlin: Mouton de Gruyter.

Closs Traugott, E. (2008). Grammaticalization, constructions and the incremental development of language: Suggestions from the development of degree modifiers in English. In R. Eckardt, G. Jäger & T. Veenstra, *ariation, Selection, Development--Probing the Evolutionary Model of Language Change* (pp. 219-250). Berlin/New York: Mouton de Gruyter.

Closs Traugott, E. (2015). Toward a coherent account of grammatical constructionalization. In J. Barðdal, S. Gildea & L. Sommerer, *Diachronic Construction Grammar* (pp. 51-80). Amsterdam: Benjamins.

Clyne, M. (1978). Some (German-English) language contact phenomena at the discourse level. In J. Fishman, *Advances in the Study of Societal Multilingualism* (pp. 113-128). The Hague: Mouton de Gruyter.

Colleman, T. (2016). A reflection on constructionalization and constructional borrowing, inspired by an emerging Dutch replica of the 'time'-away construction. In K. Beuls & R. van Trijp, *Computational Construction Grammar and Constructional Change* (pp. 91-113). Amsterdam: John Benjamins.

Colleman, T., & Noël, D. (2014). Tracing the History of Deontic NCI Patterns in Dutch: A Case of Polysemy Copying. In I. Taavitsainen, A. Jucker & J. Tuominen, *Diachronic Corpus Pragmatics* (pp. 213-236). Amsterdam: John Benjamins.

Croft, W. (2000). *Explaining language change: an evolutionary approach*. Harlow: Pearson Education.

Croft, W. (2001). *Radical construction grammar*. Oxford: Oxford University Press.

Croft, W. (2006). Evolutionary models and functional-typological theories of language change. In A. van Kemenade & B. Los, *Handbook of the History of English*. Oxford: Blackwell

Croft, W. (2007). Beyond Aristotle and gradience: A reply to Aarts. *Studies In Language*, *31*(2), 409-430. http://dx.doi.org/10.1075/sl.31.2.05cro

Croft, W., & Cruse, D. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.

Davies, M. (2012). Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. *Corpora*, *7*(2), 121-157. http://dx.doi.org/10.3366/cor.2012.0024

De Bot, K. (1994). *Waarom deze rede niet in het Engels is*. Nijmegen: Hertogenbosch.

De Smet, H. (2005). A corpus of Late Modern English texts. *ICAME Journal*, *29*, 69-82.

De Smet, H. (2009). Analysing reanalysis. *Lingua*, *119*(11), 1728-1755. doi: 10.1016/j.lingua.2009.03.001

De Smet, H. (2012). The course of actualization. *Language*, *88*(3), 601-633. http://dx.doi.org/10.1353/lan.2012.0056

De Smet, H. (2013). Change through recombination: blending and analogy. *Language Sciences*, *40*, 80-94. doi: 10.1016/j.langsci.2013.03.003

De Smet, H. (2014). Does innovation need reanalysis?. In E. Coussé & F. von Mengden, *Usage-Based Approaches to Language Change* (pp. 23-48). Amsterdam/Philadelphia: Benjamins.

De Smet, H. (2016). How gradual change progresses: The interaction between convention and innovation. *Language Variation And Change*, *28*(1), 83-102. http://dx.doi.org/10.1017/s0954394515000186

De Smet, H., & Fischer, O. (2017). The role of analogy in language change: Supporting constructions. In M. Hundt, S. Pfenninger & S. Mollin, *The changing English language: Psycholinguistic perspectives* (pp. 240-268). Cambridge: Cambridge University Press.

De Decker, B., & Vandekerckhove, R. (2012). English in Flemish adolescents' computer-mediated discourse: A corpus-based study. *English World-Wide*, *33*(3), 321-352. doi: 10.1075/eww.33.3.04ded

Delbecque, N., & Verveckken, K. (2014). Conceptually-driven analogy in the grammaticalization of Spanish binominal quantifiers. *Linguistics*, *52*(3), 637-684. doi: 10.1515/ling-2014-0002

Denison, D. (1993). *English Historical Syntax*. London: Longman.

Denison, D. (2001). Gradience and linguistic change. In L. Brinton, *Historical Linguistics 1999: Selected Papers from the 14th International Conference on*

Historical Linguistics, Vancouver, 9-13 August 1999 [Current Issues in Linguistic Theory 215] (pp. 119-144). Amsterdam/Philadelphia: Benjamins.

Denison, D. (2006). Category change and gradience in the determiner system. In A. van Kemenade & B. Los, *The Handbook of the History of English* (pp. 279-304). Malden, MA/Oxford: Blackwell.

Detges, U., & Waltereit, R. (2002). Grammaticalization vs. Reanalysis: a Semantic-Pragmatic Account of Functional Change in Grammar. *Zeitschrift Für Sprachwissenschaft*, *21*(2), 151-195. http://dx.doi.org/10.1515/zfsw.2002.21.2.151

Deutscher, G. (2001). On the mechanisms of morphological change. *Folia Linguistica Historica*, *22*, 41-48. doi: 10.1515/flih.2001.22.1-2.41

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas In Psychology*, *25*(2), 108-127. doi: 10.1016/j.newideapsych.2007.02.002

Divjak, D., & Caldwell-Harris, C. (2015). Frequency and entrenchment. In E. Dąbrowska & D. Divjak, *Handbook of Cognitive Linguistics* (pp. 53-75). Berlin: Mouton de Gruyter.

Doğruöz, S., & Backus, A. (2009). Innovative constructions in Dutch Turkish: An assessment of ongoing contact-induced change. *Bilingualism: Language And Cognition*, *12*(01), 41. doi: 10.1017/s1366728908003441

Dörner, D. (1977). Superzeichen und kognitive Prozesse. In R. Posner & H. Reineke, *Zeichenprozesse: Semiotische Forschung in den Einzelwissenschaften* (pp. 73-82). Wiesbaden: Athenaion.

Dutch SoNaR corpus search | Sketch Engine. (2011). *Sketchengine.co.uk*. Retrieved 5 February 2018, from https://www.sketchengine.co.uk/sonar-corpus/

Dux, R. (2017). Classifying Language Contact Phenomena: English Verbs in Texas German. *Journal Of Germanic Linguistics*, *29*(04), 379-430. doi: 10.1017/s1470542717000034

Eckardt, R. (2011). Semantic Reanalysis and Language Change. *Language And Linguistics Compass*, *5*(1), 33-46. http://dx.doi.org/10.1111/j.1749-818x.2010.00260.x

Edwards, A. (2016). *English in the Netherlands*. Amsterdam: John Benjamins.

EF EPI 2017 – Netherlands. (2017). Retrieved from https://www.ef.co.uk/epi/regions/europe/netherlands/

Eisenberg, P. (1999). Stirbt das Deutsche an den Internationalismen?. *Der Deutschunterricht*, *3*, 17-24.

Eisenberg, P. (2001). Die grammatische Integration von Fremdwörtern – was fängt das Deutsche mit seinen Latinismen an?. In G. Stickel, *Neues und Fremdes im Deutschen Wortschatz. Jahrbuch 2000 des Instituts für Deutsche Sprache* (pp. 183-209). Berlin: Mouton de Gruyter.

Ellis, N. (2002). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies In Second Language Acquisition*, *24*, 143-188.

Ellis, N. (2008). Phraseology: The periphery and the heart of language. In F. Meunier & S. Granger, *Phraseology in Foreign Language Learning and Teaching* (pp. 1-14). Amsterdam & Philadelphia: John Benjamins.

Ellis, N. (2012). What can we count in language, and what counts in language acquisition, cognition, and use?. In S. Gries & D. Divjak, *Frequency effects in language learning and processing (Vol. 1)* (pp. 7-34). Berlin: Mouton de Gruyter.

Ellis, N. (2017). Salience in language usage, learning and change. In M. Hundt, S. Mollin & S. Pfenninger, *The Changing English Language: Psycholinguistic perspectives* (pp. 71-92). Cambridge: Cambridge University Press.

Ellis, N., & Cadierno, T. (2009). Constructing a Second Language: Introduction to the Special Section. *Annual Review Of Cognitive Linguistics Published Under The*

Auspices Of The Spanish Cognitive Linguistics Association, *7*, 111-139. doi: 10.1075/arcl.7.05ell

Ellis, N., & Ferreira-Junior, F. (2009). Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review Of Cognitive Linguistics Published Under The Auspices Of The Spanish Cognitive Linguistics Association*, *7*, 188-221. http://dx.doi.org/10.1075/arcl.7.08ell

Evans, N., & Levinson, S. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral And Brain Sciences*, *32*(05), 429. doi: 10.1017/s0140525x0999094x

Fanego, T. (2012). Motion events in English: the emergence and diachrony of manner salience from Old English to Late Modern English. *Folia Linguistica Historica*, *46*(Historica vol. 33), 29-85. http://dx.doi.org/10.1515/flih.2012.003

Fanego, T. (2017). The trolley rumbled through the tunnel: On the history of the English Intransitive Motion Construction. *Folia Linguistica*, *51*(s38), 1-45. doi: 10.1515/flih-2017-0002

Fanego, T. (2018). A construction of independent means: the history of the Way construction revisited. *English Language And Linguistics*, 1-29. doi: 10.1017/s1360674318000059

Farina, C. J. (2012). *Reanalyzing Reanalysis: A Semantic and Pragmatic Investigation of the* Habere- *and* Esse-*Perfects in Romance*. Accessed at https://www.researchgate.net/publication/319130308_Reanalyzing_Reanalysis_A_Semantic_and_Pragmatic_Investigation_of_the_Habere-and_Esse-Perfects_in_Romance

Field, F., & Comrie, B. (2002). *Linguistic borrowing in bilingual contexts*. Amsterdam: J. Benjamins.

Fillmore, C. (1988). The Mechanisms of “Construction Grammar”. *Annual Meeting Of The Berkeley Linguistics Society*, *14*, 35-55. doi: 10.3765/bls.v14i0.1794

Fischer, O. (2007). *Morphosyntactic change*. Oxford: Oxford Univerity Press.

Fischer, O. (2008). On analogy as the motivation for grammaticalization. *Studies In Language*, *32*(2), 336-382. http://dx.doi.org/10.1075/sl.32.2.04fis

Fischer, O. (2009). Grammaticalization as analogically driven change?. *Vienna English Working Papers*, *18*, 3-23.

Fischer, O. (2010). An Analogical Approach to Grammaticalization. In E. Stathi, E. Gehweiler & E. König, *Grammaticalization: Current views and issues* (pp. 181-218). Amsterdam: Benjamins.

Fischer, O. (2011). Grammaticalization as analogically driven change?. In H. Narrog & B. Heine, *The Oxford Handbook of Grammaticalization* (pp. 31-41). Oxford: Oxford University Press.

Fischer, O. (2013). An inquiry into unidirectionality as a foundational element of grammaticalization: On the role played by analogy and the synchronic grammar system in processes of language change. *On Multiple Source Constructions In Language Change*, *37*(3), 515-533. http://dx.doi.org/10.1075/sl.37.3.03fis

Fortson, B. (2003). An Approach to Semantic Change. In B. Joseph & R. Janda, *The Handbook of Historical Linguistics* (pp. 648-666). Oxford: Blackwell.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain, *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing* (pp. 75-94). London: Longman.

Gardner, M., Rothkopf, E., Lapan, R., & Lafferty, T. (1987). The word frequency effect in lexical decision: Finding a frequency-based component. *Memory & Cognition*, *15*(1), 24-28. doi: 10.3758/bf03197709

Garey, H. (1957). Verbal aspect in French. *Language 49*, 890-925.

Geeraerts, D. (1987). Introduction: Prospects and problems of prototype theory. *Linguistics*, *27*(4), 587-612. doi: 10.1515/ling.1989.27.4.587

Gerrig, R., & Gibbs, R. (1988). Beyond the Lexicon: Creativity in Language Production. *Metaphor and Symbolic Activity 3(1)*, 1-19.

Gerritsen, M. (1984). Divergent word order developments in Germanic languages: a description and a tentative explanation. In J. Fisiak, *Historical Syntax* (pp. 107-136). Berlin: Mouton de Gruyter.

Gerritsen, M. (1995). 'English' advertisements in the Netherlands, Germany, France, Italy and Spain. In B. Machová & S. Kubátová, *Uniqueness in Unity. The Significance of Cultural Identity in European Cooperation* (pp. 324-341). Prague: SIETAR Europa.

Gerritsen, M. (1996). Engelstalige productadvertenties in Nederland: Onbemind en onbegrepen. In R. van Hout & J. Kruijsen, *Taalvariaties: Toonzettingen en modulaties op een thema* (pp. 67-83). Dordrecht: Foris Publications.

Gerritsen, M., & Nickerson, C. (2004). Fact or fallacy? English as an L2 in the Dutch business context. In C. Candlin & M. Gotti, *Intercultural Aspects of Specialized Communication. Linguistic Insights. Studies in Language and Communication. Vol. 14* (pp. 105-125). Bern: Peter Lang.

Gerritsen, M., Korzilius, H., van Meurs, F., & Gijsbers, I. (2000). English in Dutch Commercials: Not Understood and Not Appreciated. *Journal Of Advertising Research*, *40*(4), 17-31. doi: 10.2501/jar-40-4-17-31

Geurts, B. (2004). Weak and strong reflexives in Dutch. In *Proceedings of the ESSLLI workshop on semantic approaches to binding theory*. Nancy, France.

Gilquin, G. (2006). The place of prototypicality in corpus linguistics. Causation in the hot seat. In S. Th. Gries & A. Stefanowitsch, *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis* (pp. 159-191). Berlin/New York: Mouton de Gruyter.

Gisborne, N., & Patten, A. (2011). Construction grammar and grammaticalization. In Narrog, H. & Heine, B., *The Oxford Handbook of Grammaticalization* (pp. 92-104). Oxford: Oxford University Press.

Givón, T. (1991). The evolution of dependent clause morpho-syntax in biblical Hebrew. In E. Traugott & B. Heine, *Approaches to Grammaticalization, Vol. II [Typological studies in Language 19]* (pp. 257-310). Amsterdam: John Benjamins.

Goldberg, A. (1990). *Constructional Borrowing and the Process of Factorization*. Unpublished manuscript.

Goldberg, A. (1995). *Constructions*. Chicago: The University of Chicago Press.

Goldberg, A. (1996). Making one's way through the data. In M. Shibatani & S. A. Thompson (Eds.), *Grammatical Constructions: Their Form and Meaning* (pp. 29–53). Oxford: OUP.

Goldberg, A. (2006). *Constructions at work*. Oxford: Oxford University Press.

Goldberg, A. (2013). Constructionist Approaches. In T. Hoffmann & G. Trousdale, *The Oxford Handbook of Construction Grammar* (pp. 15-31). Oxford: Oxford University Press.

Goldberg, A., and Jackendoff, R. (2004). The English Resultative as a Family of Constructions. *Language*, *80(3)*, (pp. 532-568). doi:10.1353/lan.2004.0129

Görlach, M. (2002). *English in Europe*. Oxford: Oxford University Press.

Gries, S. (2005a). Null-hypothesis significance testing of word frequencies: a follow-up on Kilgarriff. *Corpus Linguistics And Linguistic Theory*, *1*(2). http://dx.doi.org/10.1515/cllt.2005.1.2.277

Gries, S. (2005b). Syntactic Priming: A Corpus-based Approach. *Journal Of Psycholinguistic Research*, *34*(4), 365-399. doi: 10.1007/s10936-005-6139-3

Gries, S. (2006a). Exploring variability within and between corpora: some methodological considerations. *Corpora*, *1*(2), 109-151. http://dx.doi.org/10.3366/cor.2006.1.2.109

Gries, S. (2006b). Introduction. In S. Th. Gries & A. Stefanowitsch, *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis* (pp. 1-17). Berlin/New York: Mouton de Gruyter.

Gries, S. (2008). Dispersions and adjusted frequencies in corpora. *International Journal Of Corpus Linguistics*, *13*(4), 403-437. doi: 10.1075/ijcl.13.4.02gri

Gries, S. (2010). Dispersions and adjusted frequencies in corpora: Further explorations. In S. Th. Gries, S. Wulff & M. Davies, *Corpus linguistic applications: Current studies, new directions* (pp. 197-212). Amsterdam: Rodopi.

Gries, S. (2013). Data in Construction Grammar. In G. Trousdale & T. Hoffmann, The Oxford Handbook of Construction Grammar (pp. 93-108). Oxford: Oxford University Press.

Gries, S. (2014a). Coll.analysis 3.5. A script for R to compute perform collostructional analyses.

Gries, S. (2014b). Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics: Some necessary clarifications. In N. Gisborne & W. Hollmann, *Theory and Data in Cognitive Linguistics* (pp. 15-48). Amsterdam/Philadelphia: John Benjamins.

Gries, S., & Hilpert, M. (2008). The identification of stages in diachronic data: variability-based neighbour clustering. Corpora, 3(1), 59-81. http://dx.doi.org/10.3366/e1749503208000075

Gries, S., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal Of Corpus Linguistics*, *9*(1), 97-129. http://dx.doi.org/10.1075/ijcl.9.1.06gri

Gries, S., & Stefanowitsch, A. (2008). Channel and constructional meaning: A collostructional case study. In G. Kristiansen & R. Dirven, Cognitive Sociolinguistics (pp. 129-152). Berlin/New York: Mouton de Gruyter.

Gurevich, O., Johnson, M., & Goldberg, A. (2010). Incidental verbatim memory for language. *Language And Cognition*, *2*(01), 45-78. doi: 10.1515/langcog.2010.003

Haase, M. (1991). *Sprachkontakt und Sprachwandel im Baskenland*. Hamburg: H. Buske.

Harris, A. (2003). Cross-linguistic perspectives on syntactic change. In B. Joseph & R. Janda, *The Handbook of Historical Linguistics* (pp. 529-551). Oxford: Blackwell.

Harris, A., & Campbell, L. (1995). *Historical Syntax in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.

Harris, A., & Campbell, L. (2005). *Historical Syntax in Cross-Linguistic Perspective* (2nd ed.). Cambridge: Cambridge University Press.

Haspelmath, M. (1998). Does Grammaticalization Need Reanalysis?. *Studies In Language*, *22*(2), 315-351. doi: 10.1075/sl.22.2.03has

Haspelmath, M. (1999). Why is grammaticalization irreversible? *Linguistics 37(6)*, 1043-1068.

Haspelmath, M. (2012). Loanwords in the World's Languages: A Comparative Handbook. In T. Graf, D. Paperno, A. Szabolcsi & J. Tellings, *Theories of everything: in honor of Edward Keenan* (pp. 109-130). Los Angeles: UCLA.

Hasselmo, N. (1970). Code-switching and modes of speaking. In N. Hasselmo & G. Gilbert, *Texas Studies in Bilingualism* (pp. 179-210). Berlin: Mouton de Gruyter.

Haugen, E. (1950). The Analysis of Linguistic Borrowing. *Language*, *26*(2), 210-231. doi: 10.2307/410058

Heine, B., & Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Heine, B., & Reh, M. (1984). *Grammaticalization and reanalysis in African languages*. Hamburg: H. Buske.

Heine, B., Clauid, U., & Hünnemeyer, F. (1991). *Grammticalization: A Conceptual Framework*. Chicago: University of Chicago Press.

Hendrikx, I., van Goethem, K., & Meunier, F. (2015). Modelling "Constructional transfer": A comparative corpus study of morphological and syntactic intensifiers in L1 French, L1 Dutch and L2 Dutch. In *Morphology Days*.

Hennemann, A. (2015). A constructionist approach to the development of the Spanish topic marker *en cuanto a* 'in terms of'. *Constructions*, *1*, 1-13.

Hilgendorf, S. (1996). The impact of English in Germany. *English Today*, *12*(3), 3. doi: 10.1017/s0266078400009093

Hilgendorf, S. (2005). "Brain Gain statt[instead of]Brain Drain": the role of English in German education. *World Englishes*, *24*(1), 53-67. doi: 10.1111/j.0883-2919.2005.00387.x

Hilpert, M. (2006). Distinctive collexeme analysis and diachrony. *Corpus Linguistics And Linguistic Theory*, *2*(2), 243-257. http://dx.doi.org/10.1515/cllt.2006.012

Hilpert, M. (2010). What can synchronic gradience tell us about reanalysis? Verb-first conditionals in written German and Swedish. In E. Closs Traugott & G. Trousdale, *Gradience, gradualness and grammaticalization* (pp. 181-203). Amsterdam: Benjamins.

Hilpert, M. (2012a). Diachronic collostructional analysis. How to use it, and how to deal with confounding factors. In J. Robynson & K. Allan, *Current Methods in Historical Semantics* (pp. 133-160). Berlin: Mouton de Gruyter.

Hilpert, M. (2012b). Diachronic collostructional analysis meets the noun phrase. Studying many a noun in COHA. In T. Nevalainen & E. Traugott, *The Oxford handbook of the history of English* (pp. 233-244). Oxford: Oxford University Press.

Hilpert, M. (2013). Corpus-based approaches to constructional change. In G. Trousdale & T. Hoffmann, *The Oxford Handbook of Construction Grammar* (pp. 458-477). Oxford: Oxford University Press.

Hilpert, M. (2014). *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.

Hilpert, M. (2015). Historical Linguistics. In E. Dąbrowska & D. Divjak, *Handbook of Cognitive Linguistics* (pp. 346-366). Berlin: Mouton de Gruyter.

Hilpert, M., & Correia Saavedra, D. (2017). Why are grammatical elements more evenly dispersed than lexical elements? Assessing the roles of text frequency and semantic generality. *Corpora*, *12*(3), 369-392. doi: 10.3366/cor.2017.0125

Hilpert, M., & Diessel, H. (2016). Frequency effects in grammar. In M. Aronoff, *Oxford Research Encyclopedia of Linguistics*. New York: Oxford University Press.

Hilpert, M., & Gries, S. (2009). Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition. *Literary And Linguistic Computing*, *24*(4), 385-401. http://dx.doi.org/10.1093/llc/fqn012

Hilpert, M., & Gries, S. (2012). Variability-based neighbor clustering: a bottom-up approach to periodization in historical linguistics. In T. Nevalainen & E. Traugott, *The Oxford handbook of the history of English* (pp. 134-144). Oxford: Oxford University Press.

Hilpert, M., & Gries, S. (2016). Quantitative approaches to diachronic corpus linguistics. In M. Kytö & P. Pahta, *The Cambridge Handbook of English Historical Linguistics* (pp. 36-53). Cambridge: Cambridge University Press.

Hilpert, M., & Koops, C. (2008). A quantitative approach to the development of complex predicates: The case of Swedish Pseudo-Coordination withsitta“sit”. *Diachronica*, *25*(2), 242-261. http://dx.doi.org/10.1075/dia.25.2.06hil

Himmelmann, N. (2004). Lexicalization and grammaticization: Opposite or orthogonal?. In W. Bisang, N. Himmelmann & B. Wiemer, *What makes Grammaticalization? A Look from its Fringes and its Components* (pp. 21-42). Berlin: Mouton de Gruyter.

Hock, H. (2003). Analogical Change. In B. Joseph & R. Janda, *The Handbook of Historical Linguistics* (pp. 441-460). Oxford: Blackwell.

Höder, S. (2011). Niederdeutsch und Norddeutsch: ein Fall von Diasystematisierung. *Niederdeutsches Jahrbuch*, *134*, 113-136.

Höder, S. (2012). Multilingual Constructions: a Diasystematic Approach to Common Structures. In K. Braunmüller & C. Gabriel, *Multilingual Individuals and Multilingual Societies* (pp. 241-257). Amsterdam & Philadelphia: John Benjamins.

Höder, S. (2014a). Constructing Diasystems: Grammatical Organisation in Bilingual Groups. In T. Åfarli & B. Mæhlum, *The Sociolinguistics of Grammar* (pp. 137-152). Amsterdam & Philadelphia: John Benjamins.

Höder, S. (2014b). Convergence vs. divergence from a diasystematic perspective. In S. Höder, K. Kühl & K. Braunmüller, *Stability and divergence in language contact. Factors and mechanisms* (pp. 39-60). Amsterdam: John Benjamins.

Höder, S. (2016). Though this be contact, yet there is system in't: Was man noch heute von Uriel Weinreich über Sprachkontakt lernen kann. In A. Walker, *Classics Revisited. Wegbereiter der Linguistik neu gelesen* (pp. 157-178). Frankfurt am Main etc.: Lang.

Hoffmann, T. (2013). Obtaining introspective acceptability judgements. In M. Krug & J. Schlüter, *Research Methods in Language Variation and Change* (pp. 99-118). Cambridge: Cambridge University Press.

Hopper, P., & Traugott, E. (2003). *Grammaticalization*. Cambridge: Cambridge University Press.

Hundt, M., Nesselhauf, N., & Biewer, C. (2007). *Corpus Linguistics and the Web*. Amsterdam: Rodopi.

Israel, M. (1996). The Way Constructions Grow. In A. Goldberg (ed.) *Conceptual Structure, Discourse And Language*, (pp. 217-230). Stanford: CSLI.

Itkonen, E. (2005). *Analogy as structure and process*. Philadelphia: J. Benjamins.

Itkonen, E., & Haukioja, J. (1997). A rehabilitation of analogy in syntax (and elsewhere). In A. Kertész, *Metalinguistik im Wandel: Die "kognitive Wende" in Wissenschaftstheorie und Linguistik* (pp. 131-177). Frankfurt: Lang.

J. McCarthy, M., & A. Carter, R. (2003). What Constitutes a Basic Spoken Vocabulary?. *Cambridge ESOL Research Notes*, *13*, 5-7.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, Mass.: MIT Press.

Jackendoff, R. (1992). Babe Ruth Homered His Way into the Hearts of America. In R. Oehrle & T. Stowell (eds.) *Syntax and the Lexicon. Syntax and Semantics, Volume 26* (pp. 155-174). Academic Press Inc.

Jackendoff, R. (1997). Twistin' the Night Away. *Language*, *73*(3), 534. doi: 10.2307/415883

Jackendoff, R. (2002). *Foundations of Language*. New York: Oxford University Press.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2001). Probabilistic relations between words: evidence from reduction in lexical production. In J. Bybee & P. Hopper, *Frequency and the emergence of linguistic structure* (pp. 229-254). Amsterdam: John Benjamins.

Kay, P., & Fillmore, C. (1999). Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. *Language*, *75*(1), 1-33. doi: 10.1353/lan.1999.0033

Keller, R. (1994). *Language Change: The Invisible Hand in Language*. London: Routledge.

Kendall, M. (1938). A new measure of rank correlation. *Biometrika*, *30 (1-2)*, 81-89.

Kilgarriff, A. (2001). Comparing Corpora. *International Journal Of Corpus Linguistics*, *6*(1), 97-133. http://dx.doi.org/10.1075/ijcl.6.1.05kil

Koester, A. (2006). *Investigating workplace discourse*. London: Routledge.

König, E. (1999). On a tué le président! The nature of passives and ultra-indefinites. In B.A. Fox, D. Jurafsky & L.A. Michaelis, *Cognition and Function in Language* (pp. 235-251). Stanford: CSLI.

Kortmann, B., & König, E. (1992). Categorial reanalysis: the case of deverbal prepositions. *Linguistics*, *30*(4). http://dx.doi.org/10.1515/ling.1992.30.4.671

Korzilius, H., van Meurs, F., & Hermans, J. (2006). The use of English in job advertisements in a Dutch national newspaper – On what factors does it depend?. Retrieved from https://www.researchgate.net/publication/254868520_The_use_of_English_in_job_advertisements_in_a_Dutch_national_newspaper_-_On_what_factors_does_it_depend

Korzilius, H., van Meurs, W., & Hermans, J. (2004). The influence of the use of English in Dutch job advertisements: An experimental study into the effects on text evaluation, on attitudes towards the organisation and the job, and on comprehension. *Esp Across Cultures*, *1*(1), 93-110.

Kramer, A. (2002). *De geschiedenis van de Nederlandse weg-constructie*. Unpublished MA-thesis, Dept. of Dutch, University of Leiden.

Kytö, M. (2012). New perspectives, theories and methods: corpus linguistics. In L. Brinton, *English Historical Linguistics: An International Handbook, Vol. 2* (pp. 1509-1531). Berlin/New York: Mouton de Gruyter.

Lahiri, A. (2000). *Analogy, Levelling, Markedness*. Berlin: Mouton de Gruyter.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. Chicago: Chicago University Press.

Langacker, R. (1977). Syntactic Reanalysis. In C. Li, *Mechanisms of Syntactic Change* (pp. 57-193). Austin: University of Texas Press.

Langacker, R. (1986). An Introduction to Cognitive Grammar. *Cognitive Science*, *10*(1), 1-40. doi: 10.1207/s15516709cog1001_1

Langacker, R. (1987). *Foundations of Cognitive Grammar, Volume I, Theoretical Prerequisites.*. Stanford: Stanford University Press.

Langacker, R. (2000). A Dynamic Usage-based Model. In M. Barlow & S. Kemmer, *Usage-based Models of Language* (pp. 1-63). Stanford, CA: CSLI Publications.

Langacker, R. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Lapata, M., McDonald, S., & Keller, F. (1999). Determinants of adjective–noun plausibility. In *Proceedings of EACL '99*.

Leech, G. (1993). 100 million words of English: a description of the background, nature, and prospects of the British National Corpus project. *English Today 33*, *9*(1). http://dx.doi.org/10.1017/s0266078400006854

Lettinga, A., van Wijk, C., & Broeder, P. (2017). The use of English in Dutch text messages as a function of communicative constraints. *Taal En Tongval*, *69*(1), 71-87. doi: 10.5117/tet2017.1.lett

Lieven, E. (2017). Developing Language from Usage: Explaining Errors. In M. Hundt, S. Mollin & S. Pfenninger, *The Changing English Language: Psycholinguistic perspectives* (pp. 321-331). Cambridge: Cambridge University Press.

Lightfoot, D. (1979). *Principles of diachronic syntax*. Cambridge: Cambridge University Press.

López-Couso, M. (2016). Corpora and online resources in English historical linguistics. In M. Kytö & P. Pahta, *The Cambridge Handbook of English Historical Linguistics* (pp. 127-145). Cambridge: Cambridge University Press.

Lüdi, G. (1996). Multilingualism through migration: A comparison of internal and external migrant communities in Switzerland. In M. Hellinger & U. Ammon, *Contrastive Sociolinguistics* (pp. 103-134). Berlin: Mouton de Gruyter.

Ludwig, I. (2005). *The* way-*construction in German*. Honours paper, Department of Linguistics, University of Canterbury.

Mair, C. (2007). Change and variation in Present-day English: integrating the analysis of closed corpora and web-based monitoring. In M. Hundt, N. Nesselhauf & C. Biewer, *Corpus Linguistics and the Web* (pp. 233-247). Amsterdam: Rodopi.

Mair, C. (2012). From opportunistic to systematic use of the web as corpus: do-support with *got (to)* in contemporary American English. In T. Nevalainen & E. Traugott, *The Oxford Handbook of the History of English* (pp. 245-255). Oxford: Oxford University Press.

Malt, B., & Smith, E. (1984). Correlated properties in natural categories. *Journal Of Verbal Learning And Verbal Behavior*, *23*(2), 250-269. doi: 10.1016/s0022-5371(84)90170-1

Marantz, A. (1992). The *way*-construction and the semantics of direct arguments in English: A reply to Jackendoff. In R. Oehrle & T. Stowell (eds.) *Syntax and the Lexicon. Syntax and Semantics, Volume 26* (pp. 179-188). Academic Press Inc.

Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.

Matras, Y. (2010). Contact, Convergence, and Typology. In R. Hickey, *The Handbook of Language Contact* (pp. 66-86). Malden, MA: Blackwell.

Matras, Y. (2011). Universals of structural borrowing. In P. Siemund, *Linguistic universals and language variation* (pp. 200-229). Berlin: Mouton de Gruyter.

McColm, D., & Trousdale, G. (2019). Whatever happened to *whatever*?. In N. Yáñez-Bouza, E. Moore, L. van Bergen, & W. Hollmann (Eds.), *Categories, Constructions and Change in English Syntax* (pp. 81-104). (Studies in English Language). Cambridge: Cambrige University Press.

McDonald, S., & Shillcock, R. (2001). Rethinking the Word Frequency Effect: The Neglected Role of Distributional Information in Lexical Processing. *Language And Speech*, *44*(3), 295-322. doi: 10.1177/00238309010440030101

McMahon, A. (1994). *Understanding language change*. Cambridge: Cambridge University Press.

Medin, D., & Schaffer, M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207-238. doi: 10.1037//0033-295x.85.3.207

Meillet, A. (1912). L'évolution des formes grammaticales. *Scientia (Rivista Di Scienza)*, *12*(6), 130-48.

Meillet, A. (1958). *Linguistique historique et linguistique générale*. Paris: Champion.

Mezhevich, I. (2003). English resultatives: State versus location. In S. Burelle and S. Somesfalean (eds.), *Proceedings of the 2003 Annual Conference of the Canadian Linguistic Association* (pp. 166-177). Québec: Université du Québec à Montréal.

Michaelis, L. (2003). Headless constructions and coercion by a construction. In Francis, L., & Michaelis, L (eds.), *Form-function incongruity and the architecture of grammar* (pp. 259-310). Stanford, California: CSLI Publications.

Michaelis, L. (2013). Sign-Based Construction Grammar. In Hoffman, T., & Trousdale, G. (eds.), *The Oxford Handbook of Construction Grammar* (pp. 133-152). Oxford: Oxford University Press.

Milroy, J. (1997). Internal vs external motivations for linguistic change. *Multilingua - Journal Of Cross-Cultural And Interlanguage Communication*, *16*(4), 311-324. doi: 10.1515/mult.1997.16.4.311

Mithun, M. (2008). Borrowed Rhetorical Constructions as Starting Points for Grammaticalization. In A. Bergs & G. Diewald, *Constructions and Language Change* (pp. 196-230). Berlin: Mouton de Gruyter.

Mondorf, B. (2006). Competing Constructions. Paper presented at the Third Meeting of the German Construction Grammar Network, Language Variation and Change in Construction Grammar. Düsseldorf, Germany. March 31-April 2 2006.

Mondorf, B. (2011). Variation and change in English resultative constructions. *Language Variation and Change*, *22(3)*, (pp. 397–421).

Moravcsik, E. (1978). Universals of Language Contact. In J. Greenberg, *Universals of Human Language* (pp. 94-122). Stanford: Stanford University Press.

Mufwene, S. (2001). *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.

Muysken, P. (1981). Halfway between Quechua and Spanish: the case for relexification. In A. Highfield & A. Valdman, *Historicity and Variation in Creole Studies* (pp. 52-78). Ann Arbor: Karoma.

Muysken, P., & van Hout, R. (1994). Modeling lexical borrowability. *Language Variation And Change*, *6*(1), 39-62. doi: 10.1017/s0954394500001575

Newmeyer, F. (2003). Grammar is Grammar and Usage is Usage. *Language*, *79*(4), 682-707. doi: 10.1353/lan.2003.0260

Noël, D. (2007). Diachronic construction grammar and grammaticalization theory. *Functions Of Language*, *14*(2), 177-202. doi: 10.1075/fol.14.2.04noe

Noël, D. (2016). For a radically usage-based diachronic construction grammar. *Belgian Journal Of Linguistics*, *30*(1), 39-53. http://dx.doi.org/10.1075/bjl.30.03noe

Noël, D. (2017). The development of non-deontic be bound to in a radically usage-based diachronic construction grammar perspective. *Lingua*, *199*, 72-93. http://dx.doi.org/10.1016/j.lingua.2017.07.012

Nosofsky, R. (1992). Exemplars, prototypes, and similarity rules. In A. Healy, S. Kosslyn & R. Shiffrin, *Essays in honor of William K. Estes: Vol. 1. From learning theory to connectionist theory* (pp. 149-167). Hillsdale, NJ: Erlbaum.

O'Keeffe, A. (2003). *Strangers on the Line: A Corpus-based Lexico-grammatical Analysis of Radio Phone-in* (Unpublished PhD thesis). University of Limerick.

Pedersen, J. (2013). The *Way*-construction and Cross-linguistic Variation in Syntax: Implications for Typological Theory. In C. Paradis, J. Hudson & U. Magnusson, *The Construal of Spatial Meaning: Windows into Conceptual Space* (pp. 236-262). Oxford: OUP.

Pateman, T. (1997). *Language in mind and language in society: studies in linguistic reproduction*. Oxford: Clarendon Press.

Perek, F. (2018). Recent change in the productivity and schematicity of the way-construction: A distributional semantic analysis. *Corpus Linguistics And Linguistic Theory, 14(1)*, 65-97. doi: 10.1515/cllt-2016-0014

Pierrehumbert, J. (2003). Phonetic Diversity, Statistical Learning, and Acquisition of Phonology. *Language And Speech*, *46*(2-3), 115-154. doi: 10.1177/00238309030460020501

Pierrehumbert, J. (2006). The next toolkit. *Journal Of Phonetics*, *34*(4), 516-530. doi: 10.1016/j.wocn.2006.06.003

Pietsch, L. (2010). What has changed in Hiberno-English: constructions and their role in contact-induced change. *STUF - Language Typology And Universals*, *63*(2). doi: 10.1524/stuf.2010.0010

Pijpops, D., & Van de Velde, F. (2016). Constructional contamination: How does it work and how do we measure it?. *Folia Linguistica*, *50*(2), 543-581. doi: 10.1515/flin-2016-0020

RStudio Team (2015). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA. URL http://www.rstudio.com/.

Rappaport Hovav, M., and Levin, B. (1995). *Unaccusativity*. Cambridge, Mass.: MIT Press.

Raymond, W., & Brown, E. (2012). Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish. In S. Gries & D. Divjak, *Frequency effects in language learning and processing* (pp. 35-52). Berlin/New York: Mouton de Gruyter.

Rayson, P., & Garside, R. (2000). Comparing Corpora using Frequency Profiling. In *Proceedings of the Workshop on Comparing Corpora* (pp. 1-6).

Ridder, S. (1995). English in Dutch. *English Today*, *11*(4), 44. doi: 10.1017/s0266078400008622

Robinson, P., & Ellis, N. (2008). Conclusion: Cognitive Linguistics, Second Language Acquisition and L2 Instruction—Issues for Research. In P. Robinson & N. Ellis, *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 489-528). London: Routledge.

Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology 7*, 575-605.

Rosenbach, A. (2010). How synchronic gradience makes sense in the light of language change (and vice versa). In E. Traugott & G. Trousdale, *Gradience, gradualness and grammaticalization* (pp. 149-180). Amsterdam: John Benjamins.

Ruhnkehl, J., Schlobinski, P., & Siever, T. (1998). *Sprache und Kommunikation im Internet*. Opladen: Westdeutscher Verlag.

Salkoff, M. (1988). Analysis by Fusion. *Linguisticae Investigationes*, *12(1)* (pp. 49-84).

Sankoff, G. (2002). Linguistic Outcomes of Language Contact. In J. Chambers, P. Trudgill & N. Schilling-Estes, *The Handbook of Language Variation and Change* (pp. 638-668). Malden, MA: Blackwell Publishers.

Sapir, E. (1921). *Language*. New York: Harcourt, Brace.

Schäfer, R. (2015). Processing and Querying Large Web Corpora with the COW14 Architecture. In *Challenges in the Management of Large Corpora (CMLC-3)* (pp. 28-34). Lancaster.

Schäfer, R., & Bildhauer, F. (2012). Building Large Corpora from the Web Using a New Efficient Tool Chain. In *The Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 486-493). Istanbul.

Schmid, H. (2000). *English Abstract Nouns as Conceptual Shells*. Berlin/New York: Mouton de Gruyter.

Schmid, H. (2010). Does frequency in text really instantiate entrenchment in the cognitive system?. In D. Glynn & K. Fischer, *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches* (pp. 101-133). Berlin: Walter de Gruyter.

Schmid, H., & Küchenhoff, H. (2013). Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics*, *24*(3), 531-577. http://dx.doi.org/10.1515/cog-2013-0018

Schneider, W., & Chein, J. (2003). Controlled & automatic processing: behavior, theory, and biological mechanisms. *Cognitive Science*, *27*(3), 525-559. doi: 10.1016/s0364-0213(03)00011-9

Scott, M. (1999). *WordSmith Tools*. Oxford: Oxford University Press.

Seland, G. (2001). *The Norwegian reflexive caused motion construction. A Construction Grammar approach*. Oslo: Department of Linguistics, University of Oslo.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (1999). A Way with Common Words. In H. Hasselgard & S. Okesfjell, *Out of Corpora: Studies in Honour of Stig Johansson* (pp. 157-179). Amsterdam: Rodopi.

Slobin, D. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In Strömqvist, S., & Verhoeven, L. (eds.), *Relating events in narrative, Vol. II: Typological and contextual perspectives* (pp. 219–257). Mahwah, NJ: Lawrence Erlbaum Associates.

Slobin, D. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In Hickmann, M., & Robert, S. (eds.), *Space in languages: Linguistic systems and cognitive categories* (pp. 59–81). Amsterdam & Philadelphia: John Benjamins.

Smirnova, E. (2015). Constructionalization and constructional change. In J. Barðdal, S. Gildea & L. Sommerer, *Diachronic Construction Grammar* (pp. 81-108). Amsterdam: Benjamins.

Sommerer, L. (2015). The influence of constructions in grammaticalization. In J. Barðdal, E. Smirnova, L. Sommerer & S. Gildea, *Diachronic Construction Grammar* (pp. 107-138). Amsterdam/Philadelphia: John Benjamins.

Steels, L. (2011). *Design patterns in fluid construction grammar*. Amsterdam: John Benjamins Pub. Co.

Stefanowitsch, A. (2011). Cognitive linguistics meets the corpus. In M. Brdar, M. Fuchs & S. Th. Gries, *Expanding Cognitive Linguistic Horizons* (pp. 257-290). Amsterdam: John Benjamins.

Stefanowitsch, A., & Flach, S. (2016). A corpus-based perspective on entrenchment. In H. Schmid, *Entrenchment and the psychology of language: How we reorganize and adapt linguistic knowledge*. Berlin: Mouton de Gruyter

Stefanowitsch, A., & Gries, S. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal Of Corpus Linguistics*, *8*(2), 209-243. http://dx.doi.org/10.1075/ijcl.8.2.03ste

Stefanowitsch, A., & Th. Gries, S. (2009). Corpora and grammar. In A. Lüdeling & M. Kytö, *Corpus Linguistics: an International Handbook, Vol. 2* (pp. 933-951). Berlin/New York: Mouton de Gruyter.

Sweetser, E. (1988). Grammaticalization and Semantic Bleaching. In *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society* (pp. 389-405).

Szczesniak, K. (2013). You can't cry your way to candy: Motion events and paths in the x's way construction. *Cognitive Linguistics*, *24*(1) (pp. 159-194).

Szmrecsanyi, B. (2006). *Morphosyntactic persistence in spoken English. A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin: Mouton de Gruyter.

Taylor, J. (1995). *Linguistic categorization*. Oxford: Oxford University Press.

Taylor, J. (2012). *The Mental Corpus: How Language is Represented in the Mind*. Oxford: Oxford University Press.

Taylor, J. (2015). Prototype effects in grammar. In E. Dąbrowska & D. Divjak, *Handbook of Cognitive Linguistics* (pp. 562-579). Berlin: Mouton de Gruyter.

Thomason, S. (2001). *Language contact*. [Edinburgh]: Edinburgh University Press.

Timberlake, A. (1977). Reanalysis and actualization in syntactic change. In C. Li, *Mechanisms of Diachronic Syntax* (pp. 141-177). Austin: University of Texas Press.

Toivonen, I. (2002). The directed motion construction in Swedish. *Journal of Linguistics*, *38(2)*, (pp. 313–45).

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Traugott, E. (2011). Grammaticalization and mechanisms of change. In H. Narrog & B. Heine, *The Oxford Handbook of Grammaticalization* (pp. 19-30). Oxford: Oxford University Press.

Traugott, E. (2015). Toward a coherent account of grammatical constructionalization. In J. Barðdal, E. Smirnova, L. Sommerer & S. Gildea, *Diachronic construction grammar* (pp. 51-80). Amsterdam: John Benjamins.

Traugott, E., & Trousdale, G. (2010). *Gradience, Gradualness and Grammaticalization*. Amsterdam: John Benjamins Publishing Company.

Traugott, E., & Trousdale, G. (2013). *Constructionalization and constructional changes*. Oxford: Oxford University Press.

Tribble, C. (1997). Improvising Corpora for ELT: Quick and Dirty Ways Developing Corpora for Language Teaching (pp. 106-117). Łódź: Łódź.

Trousdale, G. (2008). Constructions in grammaticalization and lexicalization: Evidence from the history of a composite predicate construction in English. In G. Trousdale & N. Gisborne (Eds.), *Constructional Approaches to English Grammar* (pp. 33-67). Berlin: Mouton de Gruyter.

Trousdale, G. (2010). Issues in constructional approaches to grammaticalization in English. In K. Stathi, E. Gehweiler, & E. Konig (Eds.), *Grammaticalization: Current Views and Issues* (pp. 51-72). (Studies in Language Companion). John Benjamins.

Trousdale, G. (2012). Grammaticalization, constructions, and the grammaticalization of constructions. In K. Davidse, T. Breban, L. Brems, & T. Mortelmans (Eds.), *Grammaticalization and Language Change: New Reflections* (pp. 167-198). Amsterdam: John Benjamins.

Trousdale, G. (2013). Gradualness in language change. *Synchrony And Diachrony: A Dynamic Interface*, *133*(27), 27-42.

Tryk, H. (1986). Subjective Scaling of Word Frequency. *The American Journal Of Psychology*, *81*(2), 170-177. doi: 10.2307/1421261

Van de Velde, F., & Zenner, E. (2010). Pimp my Lexis: het nut van corpusonderzoek in normatief taaladvies [Pimp my lexis: the use of corpus research in normative language advice]. In E. Hendrickx, K. Hendrickx, W. Martin, H. Smessaert, W. Van Belle & J. van der Horst, *Liever meer of juist minder? Over normen en variatie in taal [Preferably more or rather less? On norms and variation in language]* (pp. 51-68). Gent: Academia press.

van de Velde, F., Ghesquière, L., & de Smet, H. (2013). *On multiple source constructions in language change*. Amsterdam: John Benjamins.

van der Sijs, N. (1996). *Leenwoordenboek. De invloed van andere talen op het Nederlands*. SDU.

Van Egmond, M. E. (2006). *Two 'way'-constructions in Dutch: motion along a path and transition to a location*.

van Meurs, F., Hendriks, B., & Sanders, D. (2017). Attitudes to English job titles in the Netherlands and Flanders. *Taal En Tongval*, *69*(1), 89-111. doi: 10.5117/tet2017.1.meur

Verhagen, A. (2002). From parts to wholes and back again. *Cognitive Linguistics, 13(4)* (pp. 403-439).

Verhagen, A. (2003a). The Dutch way. In A. Verhagen & J. van de Weijer (Eds.), *Usage-based Approaches to Dutch* (pp. 27–57). Utrecht: LOT.

Verhagen, A. (2003b). *Wie sich Sprachen ihren Weg bahnen. Spezifische vs. allgemeine grammatikalische Konstruktionen im Vergleich*.

Verhagen, A. (2004). Hoe het Nederlands zich een eigen weg baant - Vergelijkende en historische observaties vanuit een constructie-perspectief. *Nederlandse Taalkunde 8* (pp. 327-346).

Verhagen, A. (2007). English constructions from a Dutch perspective: Where are the differences? (pp. 257-274). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Voorspoels, W., Vanpaemel, W., & Storms, G. (2008). Exemplars and prototypes in natural language concepts: A typicality-based evaluation. *Psychonomic Bulletin & Review*, *15*(3), 630-637. doi: 10.3758/pbr.15.3.630

Vriesendorp, H., & Rutten, G. (2017). 'Omg zo fashionably english'. *Taal En Tongval*, *69*(1), 47-70. doi: 10.5117/tet2017.1.vrie

Wasserscheidt, P. (2014). Constructions do not cross Languages: On cross-linguistic generalizations of constructions. *Reflections On Constructions Across Grammars*, *6*(2), 305-337. doi: 10.1075/cf.6.2.07was

Weinreich, U. (1953). *Languages in contact, findings and problems*. New York: Linguistic Circle of New York.

Wendt, G. (1911). *Syntax des heutigen Englisch*. Heidelberg: Winter.

Whitman, J. (2000). Relabelling. In S. Pintzuk, G. Tsoulas & A. Warner, *Diachronic Syntax: Models and Mechanisms* (pp. 220-240). Oxford/New York: Oxford University Press.

Whitman, J. (2012). Misparsing and syntactic reanalysis. In A. van Kemenade & N. de Haas, *Historical Linguistics 2009: Selected papers from the 19th International Conference on Historical Linguistics, Nijmegen, 10-14 August 2009* (pp. 69-88). Amsterdam: Benjamins.

Wilson, D. (2014). *Categorization and constructional change in Spanish expressions of 'becoming'*. Leiden: Brill.

Winford, D. (2010). Contact and Borrowing. In R. Hickey, *The Handbook of Language Contact* (pp. 170-188). Malden, MA: Blackwell.

Zawada, B. (2009). Linguistic creativity and mental representation with reference to intercategorial polysemy. PhD Thesis, University of South Africa.

Zenner, E., Speelman, D., & Geeraerts, D. (2013). What Makes a Catchphrase Catchy? Possible Determinants in the Borrowability of English Catchphrases in Dutch. In E. Zenner & G. Kristiansen, *New Perspectives on Lexical Borrowing: Onomasiological, Methodolocial and Phraseological Innovations* (pp. 41-64). Berlin et al.: Mouton de Gruyter.

Zenner, E., Speelman, D., & Geeraerts, D. (2015). A sociolinguistic analysis of borrowing in weak contact situations: English loanwords and phrases in expressive utterances in a Dutch reality TV show. *International Journal Of Bilingualism*, *19*(3), 333-346. doi: 10.1177/1367006914521699

Zeschel, A. (2010). Exemplars and analogy: Semantic extension in constructional networks. In K. Fischer, *Quantitative methods in Cognitive Semantics: Corpus-driven approaches* (pp. 201-219). Berlin: Mouton de Gruyter.

Zimmer, D. (1997). *Deutsch und anders: die Sprache im Modernisierungsfieber*. Reinbek bei Hamburg: Rowohlt.