UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Magistrale in Linguistica
Classe LM-39

Tesi di Laurea

The Battle between Age and Frequency

Relatrice
Prof.ssa Cecilia POLETTO

Correlatore esterno
Prof. Martin HASPELMATH

Laureanda
Alexandra ZALESKY
n° matr.1190064 / LMLIN

Anno Accademico 2019 / 2020

Table of contents

Abstract: This paper investigates the driving forces behind the development of grammatical structures in the field of tense and aspect, and more precisely regarding the expression of present perfects compared to perfective pasts or general pasts. For this reason, Joan L. Bybee and Östen Dahl's form-meaning correlation hypothesis (1989) is compared to Martin Haspelmath's form-frequency correlation hypothesis (2020). A corpus research of four languages was further carried out to test the supposed correlation between the frequency and the form of a grammatical construction, as argued by Haspelmath. The four examined languages were English, Swedish, Peninsular Spanish and European Portuguese. Six separate corpus researches were carried out to determine the relative token frequency of periphrastic present perfect constructions compared to that of inflectional perfective pasts or general pasts. 30 verbs were searched in first person singular, with or without a pronoun, depending on the language. Constructions with adverbs were also included. The English corpus research was divided into American English and British English, for the language-specific differences regarding the use of the Present Perfect and the Simple Past. A smaller manual research was executed to include the Swedish Present Perfect construction in which the auxiliary can be omitted in subordinate clauses. The results showed a preference for the bound form in Portuguese, English and Swedish. However, this picture is complicated by several language-particular tendencies, such as the increasing use of English Present Perfect constructions together with a past time adverbial. Furthermore, the data from the Spanish research were not as clear-cut. The corpus results were compared to, and complemented by, a study regarding the role of language contact in grammaticalization (Giacalone Ramat 2008) and Susana Azpiazu's concept of "simultaneity" (2018). Although it is clear that the role of frequency is of importance regarding grammatical coding, further research and a broader and more varied language sample is necessary for the possibility to draw more definite conclusions to whether frequency can be considered the ultimate cause.

Introduction

The past can be described from different points of view and expressed through different constructions. For example, a past event which is described as occurring at a specific moment is typically expressed through inflection in Swedish, whereas a past event with current relevance is expressed with a periphrastic construction. This is not specific to Swedish, but reflects a cross-linguistic tendency according to which general pasts or perfective pasts are bound and present perfects have a periphrastic construction (Bybee and Dahl 1989). The aim of the current paper is to understand better why certain constructions tend to be longer, here present perfects, whilst others are shorter. To do this two theories will be presented and compared. The first one is the hypothesis based on form-meaning correlations, presented in *The Creation of Tense and Aspect Systems in the*

Languages of the World (1989) by Joan L. Bybee and Östen Dahl. According to this theory, diachronic paths of grammaticalization explain cross-linguistic tendencies of the tense-aspect systems. Bybee and Dahl argue that the role of time is crucial, since the semantic development of a grammatical construction is paralleled by phonological changes. Their individual developments of the common theory will also be presented, based on *The Growth and Maintenance of Linguistic Complexity* (2004) by Dahl and on *Language change and universals* (2006) by Bybee, for a more recent perspective. The second theory examined is a functional-adaptive approach presented in *Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability* (2020) by Martin Haspelmath. Cross-linguistic asymmetrical patterns are explained as a result of efficient coding: more frequently used constructions do not necessitate extra coding and are therefore left zero-coded or less coded.

If Haspelmath is correct, then present perfect constructions should generally be less frequent. To test this, a corpus-based research will be carried out, in which the relative token frequency of 30 verbs in present perfect constructions is compared to perfective past or general past constructions in four languages, namely English, Swedish, Peninsular Spanish and European Portuguese. The collected corpus data will be compared to the predictions of each hypothesis. The results of the corpus researches will further be considered in the light of two other theories. On the one hand, to the role of language contact, as proposed by Anna Giacalone Ramat (2008), comparing the results to three other Indo-European languages, namely French, Northern Italian and Southern German. On the other hand, to Susana Azpiazu's concept of "simultaneity" (2018), which aims at explaining the reason behind temporal and aspectual differences of the present perfect in Spanish, Portuguese and English. The complete results from the corpus researches are presented in the appendix.

The paper is structured as follows. Chapter 1 discusses different functions and expressions of the present perfect and the perfective, and gives a general overview of their use in twelve different languages, based on reference grammars and linguistic studies. I also consulted some native-speakers, when necessary. In chapter 2 the form-meaning correlation theory by Bybee and Dahl is presented, together with their more recent developments of the theory. In the following chapter the form-frequency correspondence theory by Martin Haspelmath is laid out. Chapter 4 is dedicated to the methodology and the results of the corpus research. In chapter 5 the results are discussed and compared to the different approaches, followed by a conclusion.

CHAPTER 1

Perfective and Perfect

1.1 Terminological confusion

The notions “perfective” and “perfect” have often been confused in language and linguistic studies (Comrie 1976). Perfective is sometimes called “aoristic”, and perfect is sometimes referred to as “anterior”. I use the terms “perfective” and “perfect”, unless otherwise noted. The perfective and the perfect both belong to the tense-aspect system. Aspects can semantically be described as “different ways of viewing the internal temporal constituency of a situation” (Comrie 1976: 3), as Bernard Comrie wrote in his monumental work *Aspects.* A tense can briefly be described as a deictic category, which “relates the time of the situation referred to some other time, usually to the moment of speaking” (Comrie 1976: 1-2). One can, therefore, understand that part of the confusion originates from the fact the both tenses and aspects are, of course, related to time, but in different ways. Once again in Comrie’s words: aspect regards the “situation-internal time”, whilst tense regards “situation-external time” (1976: 5).

Language-particular categories often combine tense and aspect in different ways, which, naturally, complicates cross-linguistic comparison. Part of the terminological confusion concerns the fact whether perfect is to be considered as a tense or as an aspect. Joan L. Bybee and Östen Dahl (1989) propose a different approach to this problem: to focus on comparing expressions of grammatical morphemes of different languages “from similar sources at similar stages of development” (1989: 97), instead of focusing on which “super-category” each grammatical expression is a member of.

I will now describe in more detail perfective and perfect, following Comrie (1976), unless otherwise noted. These two notions are thereafter exemplified in twelve languages from different languages families.

1.2 Perfective

The perfective can be opposed to the imperfective. The former views the situation as bounded, as a single whole, and the latter as non bounded, focusing on the internal structure of the situation. The same situation could, therefore, be described either from a perfective point of view or from an imperfective point of view.

As is typical of linguistic notions, or any notion in general, the notion of "perfectivity" does not have fixed boundaries and can be divided into different subtypes. This fact has sometimes led to inexact definitions of the notion, and Comrie (1976) lists the frequent ones, as here follows. A widespread conception of perfectivity is that it describes a situation of short duration, or even limited duration. The perfective is, furthermore, often said to indicate a completed action, but Comrie argues that it is better described as indicating a complete situation, that is, one with a beginning, middle and end, since the situation is viewed as a whole. Another misconception, according to Comrie, is to regard the perfective as a notion which simply denotes resultativity. Even though this can be true for some verbs, it does not describe perfectivity in general.

Not describing the situation as having "internal temporal constituency" (1976: 21), does not imply that the situation in itself lacks it. It is simply one way of representing the situation. Dahl describes the perfective as something which "will typically denote a single event, seen as an unanalyzed whole, with a well-defined result or end-state, located in the past" (1985: 78). Furthermore, Bybee and Dahl (1989) distinguish two main types of perfective aspect: one in which there is a "total view of the situation", and another, where the focus lies on the "presence of a limit or end-state for the process" (1989: 87-88). For some languages the former description is more adequate, and for others the latter.

1.3 Perfect

The perfect has traditionally been considered an aspectual notion, but, as mentioned above, it is a matter of debate whether perfect is better considered as part of the temporal or the aspectual realm. The perfect indicates that a situation is relevant at a certain point of reference, such as the moment of speech. It, therefore, does not fit in the general description of what an aspect is, not describing the situation itself, but simply its relevance to a point of reference. It connects two time-points, and by doing so it is closer to the above-mentioned definition of a tense.

Comrie (1976) defines different types of perfect, which can be grammaticalized, or not, in a given language. Each type can, furthermore, be combined with different tenses, depending on which the point of reference is. We can hence speak of “present perfect”, “past perfect” or “future perfect”. The types described by Comrie are listed below and are considered in the present tense of the perfect aspect.

1.3.1 Perfect of result

With resultative a “present state is referred to as being the result of some past situation” (1976: 56), e.g. “I have lost my keys” (and I still do not know where they are).

1.3.2 Experiential perfect

“The experiential perfect indicates that a given situation has held at least once during some time in the past leading up to the present” (1976: 58), e.g. “Have you ever eaten sushi?”.

1.3.3 Perfect of persistent situation

This type of perfect is used “to describe a situation that started in the past but continues (persists) into the present” (1976: 60), e.g. “She has lived in Portugal for ten years” (and still lives there). It is referred to as “universal perfect” or “continuative perfect” as well (Walker 2011).

1.3.4 Perfect of recent past

“[T]he present relevance of the past situation referred to is simply one of temporal closeness” (1976: 60), e.g. “I have just had lunch”. It is often described as “hot news”.

1.4 Perfective and perfect in different languages

There is, of course, variety among the languages of the world when it comes to verbal categories, as with the other aspects of the language. Besides the fact that a notion of aspect or tense seldom coincides perfectly in two languages, languages do not always grammaticalize the same notions of tense and aspect.

Dahl (1985), and Bybee and Dahl (1989) argue that a typical tense-aspect system including a perfective aspect is a tripartite system, in which the perfective is only grammaticalized in the past, whilst the imperfective is divided into present and past. This

tripartite system is to be seen as the core, with the possibility of the language having other tenses and aspects. From this follows that “the main difference between a language that has a simple past and one that has a perfective is the presence or absence of a past imperfective” (Bybee and Dahl 1989: 83). This is, however, as Bybee (1995) notes, not the only system in the world, but it is argued to be a common one among languages with inflectional aspect. An alternative tense-aspect system can be found in Russian, see 1.4.7.

The expressions and uses of perfective and perfect are illustrated through twelve languages below. The languages are mainly from the Indo-European language family. I have included one isolating language, Mandarin Chinese, although this study is mainly focused on languages with inflection. I have based the following descriptions both on reference grammars and on more strictly linguistic papers and essays. Different authors may, of course, use different terminology when referring to the language-specific categories. I use as little variation among terms as possible for clarity, unless otherwise noted. Language-specific categories are marked with an initial capital letter, both its original denomination and the English translation, to differentiate them from the cross-linguistic notions.

1.4.1 English

English does not grammatically express the opposition between the perfective and the imperfective. The perfect, on the other hand, is well grammaticalized and is expressed periphrastically: “have” + past participial form of the verb. The English Present Perfect and its semantics can be considered close to prototypical (Dahl 1985) and its uses can accordingly be divided into perfect of result, experiential perfect, perfect of persistent situation and perfect of recent past. See 1.3.1-1.3.4 for examples.

The English Present Perfect does not allow time adverbials indicating the past moment, and a sentence such as “I haven’t read the newspaper this morning” is not considered “correct” if pronounced, for example, during the afternoon of the same day. Time expressions including the current moment are, on the other hand, allowed, e.g. “I haven’t read the newspaper today”. This phenomenon, which is not limited to English, is known as the “Present Perfect Puzzle” (Klein 1992). To be noted is the fact that it is not the past tense expression in itself which causes the problem, but the reference to that specific moment, as can be seen with experiential perfect. Comrie exemplifies this with “I

have (on some occasion in the past) got up at five o'clock". See 1.4.3 for a comparison with the Spanish Present Perfect on this note.

The Present Perfect is more frequently used in British English compared to American English. The distribution of Present Perfect in the English speaking world can be described as a continuum, with American English at one end and British English at the other (Yao and Collins 2012). *The perfect and the preterite in contemporary and earlier English* (1997) by Johan Elsness shows a general increasing frequency of the Present Perfect till the late 18th century, and from there on the opposite development. This decline of the Present Perfect is more significant in American English, but not exclusive to American English. What at first glance could look like a decline only regarding the American English Present Perfect is, therefore, in reality part of a bigger picture which involves the British English Present Perfect as well. According to Xinyue Yao and Peter Collins (2012), this general decline of the English Present Perfect goes against what is happening in several other Indo-European languages, such as German and French, where "the preterit has been giving way to the present perfect" (2012: 387). I discuss this further in chapter 5.

1.4.2 Swedish

The opposition between the perfective and the imperfective is not grammatically expressed in Swedish, in conformity with the rest of the modern Germanic languages, which in general do not express the perfective (Dahl 1985). However, again similarly to other Germanic languages, the perfective can be expressed with some verbs through the addition of a particle following the verb, such as *upp* ("up") and *ut* ("out"): *Hon har gått ut skolan* ("She has finished high school"), compared to *gå i skolan* ("go to (frequent) school") (Strzelecka 2003).

The Swedish Present Perfect is similar to the English Present Perfect. It is expressed through periphrasis: *ha* ("have") + the supine form of the verb. Being close to the English meanings, it is also close to the prototypical perfect and can express perfect of result, experiential perfect, perfect of persistent situation and perfect of recent past. Furthermore, it rarely occurs together with definite time adverbials, but it can however be acceptable in certain constructions, unlike in English. According to Dahl (1985) a sentence like *Jag har besökt England i januari nittonhundrafyrtiotvå* (lit. "I have visited England in January, nineteen forty-two") is acceptable if the time adverbial is stressed,

denoting that this is “new information”. I will come back to this in 5.2.2, for a possible parallel in English.

A particularity regarding the form of the Perfect, is that the auxiliary *ha* can be omitted in a subordinate clause, as in *Han klagar över att han inte* (*har*/*hade*) *fått instruktioner* (“He complains about not having received instructions”), a phenomenon which has been hypothesized as a result of High German influence.[1] The theory of the German influence can, however, be questioned since there is evidence for Perfect constructions lacking the auxiliary even from Late Old Swedish. That is, these constructions have been found in texts preceding the period of the German influence (Sangfelt 2014).

1.4.3 Spanish

Both the opposition perfective/imperfective and the perfect are grammaticalized in Spanish. The perfective is expressed only in the past, as is typical for many Indo-European languages. The language-specific terms for the perfective past and the present perfect are respectively *Pretérito perfecto simple* and *Pretérito perfecto compuesto*, but I refer to them as “Preterite” and “Present Perfect”, respectively. Spanish is one of the most spoken languages in the world and has many varieties. I will focus on Peninsular Spanish, spoken in central Spain, henceforth simply referred to as “Spanish”.

The Preterite is expressed through bound morphology. In *A New Reference Grammar of Modern Spanish* (2019) by John Butt et al. its basic use is described as depicting punctual situations in the past, “events that were completed in the past” (2019: 208), such as *Martín la llamó cuatro veces* (“Martin called her four times”). As can be seen from the example, also from Butt et al., this is a typical example of perfectivity, in which the situation is analyzed as a single whole, and viewed as “completed” in that sense. Butt et al. go on stating that “the preterite tense must be used for events that continued throughout a finite period of time” (2019: 209), and argue that what is considered to be significant is whether the period is finite, not the action. This is, once again, a typical use of perfective, since the situation is seen as temporarily bounded.

Furthermore, the Preterite together with stative verbs, such as *saber* (“know”) and *conocer* (“know”; “be acquainted”) may have an inchoative meaning, indicating the process of beginning, or signal “entering into a state”. Bybee (1995) argues that this is a natural inference, since perfective in general is used to express “something that

[1] Even though this is no longer possible in Modern German.

happened". The inference of this perfective meaning together with a stative verb is "that the subject entered into the state" (Bybee 1995: 449), for example, *Supe lo sucedido* ("I learned what happened").

The Present Perfect is expressed periphrastically, with the auxiliary *haber* ("have") + past participle. It is mainly used to describe situations which have relevance for a subsequent point in time, having, therefore, a typical perfect meaning. It can express resultative perfect, experiential perfect, perfect of persistent situation and "hot news", and is, therefore, once again a typical example of present perfect. A Present Perfect construction can also occur with time adverbials indicating the time of the past situation, unlike in English, as long as the situation has bearing on the current point of time. Charles Hugh Stevenson exemplifies this in *The Spanish Language Today* (1970: 62):

(1) - *Pareces cansado* ("You look tired")
- *Me he levantado muy temprano* (lit. "I have got up very early")

Chad Howe and Scott A. Schwenter (2003) give a similar example: *Me he levantado esta mañana a las siete* (lit. "I have got up this morning at seven"), as an alternative to the Preterite form *Me levanté esta mañana a las siete*. However, this sentence can be said in the afternoon, describing, therefore, a punctual situation in the past, without relevance for present time, such as "hot news" or resultativity. According to the two authors, this is a "diachronic innovation of the language", in which the Present Perfect is taking over functions of the Preterite. This function is temporally limited, and used mainly for situations of the same today, used hence as a hodiernal past. The authors continue, the temporal limits are, however, expanding, and for younger age groups also the following sentence is acceptable in some varieties of Peninsular Spanish[2]: *Lo he visto ayer en el supermercato* (lit. "I have seen him yesterday at the supermarket"). It should be noted that in many varieties of Latin-American Spanish there is, on the other hand, a strong preference for the Preterite, which is used also where the Present Perfect is used in Peninsular Spanish (Butt et al. 2019).

Futhermore, Miranda Stewart in *The Spanish Language Today* (1999) notes that there is a tendency for the Preterite to be used instead of the Present Perfect in news report language when speaking of "recent continuing past", e.g. *Dijo hoy* (lit. "Said today") instead of *Ha dicho hoy* (lit. "Has said today").

[2] Such as the Spanish spoken in Alicante and Madrid (Howe and Schwenter 2003).

1.4.4 Portuguese

Portuguese is the official language in nine countries around the world. I will discuss Portuguese spoken in Portugal. Both the opposition perfective/imperfective and the perfect are grammatically expressed in Portuguese.

The Preterite (*Pretérito Perfeito*) and the Imperfect (*Pretérito Imperfeito*) both denote past tense and are expressed through inflection, but they differ in their aspectual reading. The Preterite typically expresses an episodic reading, so that the events "are delimited within an assumed timeframe by the implicit completion of the action. […] viewed as being bounded in time." (Iverson and Rothman 2008: 71). It can, therefore, be said to express perfective past. The Imperfect, on the other hand, denotes the imperfective aspect, since it generally expresses a habitual or ongoing situation in the past. Micheal Iverson and Jason Rothman exemplify this opposition with the following sentences containing the verb *comer* ("to eat") in the Preterite and in the Imperfect, respectively (2008: 71):

(2) a. *A Maria comiu o sorvete ontem (*"Maria ate the ice cream yesterday")

b. *Durante a sua meninice, a Maria comia muito sorvete* ("During her childhood, Maria ate a lot of ice cream")

The Portuguese Present Perfect (*Pretérito Perfeito Composto*) is a Romance invention, not existing in Latin. It is formed by the auxiliary *ter* ("to have") and the past participle of the verb. The Portuguese Present Perfect denotes meanings which altogether do not correspond to a stereotypical perfect, or to most Romance compound past forms in general. It does relate a past event to the present, but the situation itself can only be either durative or iterative. Eventive verbs can only have a iterative reading. It can, therefore, not refer to a single, terminated event. It can hence be described as a limited Perfect, being closer to the English Present Perfect Progressive than the English Present Perfect when used with dynamic verbs. For example (Santos Lopes et al. 2016: 476):

(3) a. *Ele tem sido um bom rapaz* ("He has been (and still is) a good boy")

b. *Ele tem feito os trabalhos de casa* ("He has done his homework", or, better, "He has been doing his homework")

Furthermore, the verb in the second clause in “Go and open the door quickly, the mailman has already rung the doorbell three times” is in Portuguese expressed with a bound construction, in the Preterite, and not in the Perfect, even though it is highly relevant for the moment of speech: *Abra, rápido, o cartiero já tocuo a campainha três vezes.* Both the Portuguese Present Perfect and Preterite, therefore, can refer to a past event which is relevant for the moment of speech.

The Preterite can hence bear a type of present perfect meaning as well, as in *Quero ir ao brasil porque nunca estive lá* (“I want to go to Brazil because I have never been there”) (Alego 1976). In this experiential reading the Preterite is often preceded by the adverb *já* (“already”, “yet”, “ever”) and which, according to Alego, bears a “quasi-verbal role” (Alego 1976: 206). *Escrevi três cartas ontem* (“I wrote three cards yesterday”), expressed in the Preterite, is grammatically correct and a typical example of the perfective. However, when the Preterite has a present perfect meaning, that is, when the adverb *já* is used with the same sentence, *Já screvi três cartas ontem*, then such a sentence is ungrammatical. For the same reason “I have written three cards yesterday” would be ungrammatical in English.

1.4.5 French

French is the official language in 29 countries, I will focus on the French spoken in France. French has grammatical expressions both for the perfective/imperfective opposition and for the perfect, but only in written language. The language-specific terms are *Passé Simple, Imparfait* and *Passé Compose*, but I refer to them as “Past Definite”, “Imperfect” and “Perfect”, respectively.

Both the perfective, Past Definite, and the imperfective, Imperfect, are expressed through bound morphology: *Il lut* (“He read”) and *Il lisait* (“He was reading”), respectively. The Past Definite analyses the period of time as a single whole and, therefore, denotes perfectivity. It is, however, only used in written language (Comrie 1976).

The Perfect is formed by an auxiliary *avoir/être* + the past participial form of the verb. As mentioned above, the compound pasts of the Romance languages are a Romance invention and did not exist in Latin, where the Perfect denoted both simple past and present perfect. Like in other Romance languages, such as Northern Italian and Romanian, the Perfect does not only denote perfect meaning, but also non-perfect meaning. This is a result of the fact that it has displaced the Past Definite in spoken language. It is, therefore, more accurate to say that both the Perfect and the Past Definite

can denote perfective past. For this reason the term “Compound Past” might be more suitable than “Perfect”.

Perfect of persistent situation is expressed through present tense, *Nous vivons ici depuis des années* (“We’ve lived here for ten years”).

1.4.6 Italian

I will focus on Northern Italian, henceforth simply referred to as “Italian”. I must, however, add that many varieties of Southern Italian tend to use the Simple Past (*Passato Remoto*) with a much higher frequency than the Perfect/Compound Past (*Passato Prossimo*), unlike the varieties of Northern Italian.

Italian has inherited the Imperfect and the Simple Past from Latin, both expressed through bound morphology. The Perfect is a Romance invention and the Italian Perfect, like other Romance perfects, does not express exclusively perfect meaning. Furthermore, perfect of persistent situation is expressed in present tense, as in many other languages (Comrie 1976): *Aspetto da tre giorni* (lit. “(I) wait since three days”) translates into “I have been waiting for three days”.

The Imperfect (*Imperfetto*) expresses imperfective past, denoting, for example, that something was a habit or ongoing. Elisabetta Mauroni (2013) distinguishes Simple Past from Perfect by describing the former as a perfective aspect, in which the action is viewed as “completed”, and the latter as focusing more on the effect of the event on present time, hence as a perfect of result. Both the Perfect and the Simple Past can, however, express perfective past. The Present Perfect is displacing the Simple Past, and is used to indicate both recent and remote past actions. The Simple Past is, therefore, much less frequent than the Present Perfect in spoken language, and belongs to a higher register. According to Comrie “the Perfect has completely supplanted the Simple Past” (1976: 61) in Spoken Italian, “conquering the PFV territory” (Dahl 1985: 171).

1.4.7 Russian

The aspectual studies have their origins in Slavic studies. Russian, as is well known, has a rich aspectual system and grammatically marks the perfective/imperfective opposition. The language, on the other hand, lacks a grammatical marking for the perfect.

As argued by Bybee and Dahl, the Slavic aspectual systems differ from the above-mentioned tripartite system “in their origins, their semantics, their means of expression

and their relation to other parts of the system of verbal grams[3] such as tense" (1989: 85). The perfective is marked by "bounders"[4], and does, consequently, not originate from periphrastic constructions with auxiliaries, as do, for example, the Spanish or Portuguese perfectives. It can, therefore, be considered more derivational than inflectional. Bybee and Dahl (1989) argue that there is a cross-linguistic tendency for these bounders to become grammaticalized as aspectual markers, denoting perfectivity, see 1.4.8 for an example in Hungarian. The particularity of Slavic languages, such as Russian, lies in the systematic way in which these bounders have been prefixed to denote perfectivity. A further sign of the systematicity can be observed regarding the imperfective. Some verbs, already bearing a perfective bounder, can further be marked with an imperfective suffix. Bybee and Dahl (1989) give the example of the verb *pisat* ("write"), which can take the prefix *pere-* ("re-"), acquiring a perfective sense, "rewrite". This verb can further take on a secondary imperfective sense adding the suffix *-va*, *perepisyvat.*

Unlike the tripartite system, where the aspects and tenses are closely related, "the opposition between perfective and imperfective aspect is almost wholly independent of the category tense in Russian" (Bybee and Dahl 1989: 87). In fact, in Russian there is also a non-past (present) perfective, which, however, expresses either non-specific time reference or future time. Furthermore, in 1.2 two different types of perfective were discussed, and Dahl (1985: 74) gives an example to demonstrate how the "totality" view can result inadequate for Slavic languages such as Russian:

(4) - What did you brother do after dinner yesterday?
- He wrote letters.

The focus here is the second sentence, which might look like a "clear case of perfectivity", but which actually would be translated with an imperfective verb in Russian, *pisal*, since the activity of writing is an "unbounded activity as long as we have not delimited the object in any way" (1985: 75), that is, for example, by referring to a specific number of letters. The "boundedness" criterion, therefore, appears as the predominant one in Russian.

[3] Bybee and Dahl use the term "gram" as a shortening for "grammatical morpheme".

[4] Bybee and Dahl use the term "bounder" to indicate verbal particles such as "out", "up", "apart".

1.4.8 Hungarian

This section is based on *Hungarian: An Essential Grammar* (2001) by Carol Rounds, if nothing else is stated. Hungarian is part of the Finno-Ugric languages and further of the Uralic language family, and is an agglutinative language. It has only one past tense, expressed through bound morphology. It does not have a grammatical expression for the perfect. The perfectivity/imperfectivity opposition is grammatically marked, even though only partly.

As mentioned in 1.4.7, Hungarian can mark perfectivity through bounders, or "coverbs" as they are also called. These coverbs are usually derived from adverbs or pospositions. This is, however, not as systematic as it is in many Slavic languages (Comrie 1976). The Hungarian perfectivizing morphemes can either be free (particles) or bound (prefixes), and the two most common ones are *meg* and *el* (lit. "away from"). *Meg* is the only Hungarian coverb without a lexical meaning. If no coverb is added to the sentence, the past verb indicates an action in progress. The addition of a coverb indicates that the action is complete, hence perfective past. Compare *Írtam egy levelet* ("I was writing a letter") and *Megírtam a levelet* (I wrote a letter). Unlike Russian, there is no way to derive verbs with an imperfective meaning from perfective verbs.

The position of the coverb has importance for the aspectual meaning. For a perfective meaning, the coverb usually occupies the immediate preverbal position. If the proverb, on the other hand, follows the verb, the meaning of the verb has an imperfective reading. Éva Kardos illustrates this with the following examples (2016: 7):

(5) *A* *macska* *fel-mászott* *a* *fá-ra*
the cat.NOM PRT-climbed the tree-to
"The cat climbed up the tree"

(6) *A* *macska* *'mászott*[5] *'fel* *a* *'fá-ra,*
the cat.NOM climbed PRT the tree-to

amikor *a* *kutya* *kinézett* *az* *óljá-ból*
when the dog.NOM looked.out the his.dog.house-from
"The cat was climbing up the tree when the dog looked out of his house"

A present tense verb with a coverb often indicates future tense. By adding the coverb *el*, the imperfective reading of *Olvassa a könyvet* ("She is reading the book") takes on a future tense, *Elolvassa a könyvet* ("She will read the book").

[5] There is also a difference in the stress pattern, here marked by apostrophes following Kardos.

Hungarian is a language with a rich case system and aspect can, furthermore, be expressed through locative constructions. The sublative, which is a type of locative, “may be used to mark the result or goal of an action” (Rounds 2001: 107), e.g. *Magyarra*[6] *fordította a könyvet* (“He translated the book into Hungarian”). Dahl and Bybee (1989), in fact, mention a third kind of tense-aspect system, the case-marking systems of Finno-Ugric languages.

1.4.9 Turkish

Turkish is spoken by approximately 50 million people and is part of the Turkic languages and the Altaic language family. It is an agglutinative language.

Turkish marks the perfective/imperfective grammatically. The perfect, on the other hand, is not grammatically expressed, even though it did exist in the past and the mark today denotes quotative (Dahl 1985). Quotative is a type of evidential function and indicates “that the speaker has not experienced himself what he is reporting in the sentence but rather has it from a secondary source” (Dahl 1985: 150).

The Past Definite expresses perfective past, presenting the situation as complete with no temporal subdivisions. The Past Definite is expressed by adding the suffix *-dı* to the verb stem, and this goes for all non-stative verbs. Jaklin Kornfilt exemplifies this in *Turkish* (1997: 355):

(7)	*dün*	*oda*	*-m*	*-ı*	*topla*	*-dı*	*-m*
	yesterday	room	-1.SG	-ACC	tidy up	-PAST	-1.SG
	“Yesterday I tidied up my room”						

The imperfective is expressed through different markers, depending on the type of imperfectivity, and can refer to both present and past tense, unlike the perfective, which only refers to past situations. The Aorist suffix *-(a)r* denotes habitual aspect and the old Progressive suffix *-(i)yor* denotes continuous aspect when added to non-stative verbs. The suffix -(i)*yor* originates from the verb “to go”, “to walk”.

1.4.10 Basque

This section is based on *Tense, Aspect and Mood* (2019) by José Ignacio Hualde and Céline Mounole. Basque is an isolate spoken by approximately 750 000 people in the French and Spanish Pyrenees. The Basque intransitive verbs are divided into unergatives

[6] The sublative mark here is *-ra*, but could also be *-re*, depending on the vowel harmony.

and unaccusatives, and this division determines which auxiliary the verb takes in compound forms. The transitive and unergative verbs take *edun* (“to have”). *Izan* (“to be”) is used with unaccusative verbs. Furthermore, most verbs in Modern Basque only possess compound forms, so both the opposition perfective/imperfective and the perfect are periphrastically expressed.

The Present Perfect is formed by a perfective participle of the verb and a present-tense auxiliary, like in the Romance and Germanic languages, and denotes perfect meaning in the sense of present relevance and, furthermore, functions as a hodiernal or recent past, similarly to Peninsular Spanish. For example, *Gaur goizean Mikel ikusi dut (“*I saw Mikel this morning”). The Basque Perfect is a relatively recent construction and possibly a consequence of language contact (Hualde and Mounole 2019).

The perfective/imperfective opposition is restricted to the past, since the Perfective only exists in this tense. It is, in this sense, similar to many other European languages, even though it is an isolate. A difference can be found in what looks like a pluperfect construction: the construction including the perfective participle of the verb together with the past of “be”/“have” does not denote pluperfect, but a perfective non-hodiernal or remote past. Hualde and Mounole (2019) give the example of the verb *erosi* (participle of “buy”) which takes the auxiliary “have”. “To have” in first person singular translates to *dut* in present tense and to *nuen* in past tense. So, *Erosi dut* corresponds and translates to “I have bought”. *Erosi nuen,* on the other hand, literally corresponds to “I had bought”, but translates to “I bought”.

1.4.11 Mandarin Chinese

Mandarin Chinese is spoken by approximately 920 million people and is part of the Sino-Tibetan language family. It is an isolating language and is, therefore, characterized by almost complete lack of inflectional morphology.

Perfective is mainly expressed with the verbal particle *le*, which is added after the usually not stative verb and indicates past time alongside perfective, or with the experiential marker *guo*, added after the verb as well. If the verb is stative, *le* often denotes resultative (Comrie 1976). According to Dahl (1985), referring to the study *The discourse motivation for the perfect aspect: the Mandarin particle le* (1982) by Li et al., the meaning of the particle *le* depends on its position: it is a perfective marker if it follows the verb directly, but if it is sentence-final, it appears to bear a perfect meaning, or possibly

resultative. Both the perfective and the perfect particle *le* originate from the verb *liao* ("finish") (Bybee and Dahl 1989: 58).

Meichun Liu notes that although *le* is a perfective marker, "an event marked by the Chinese perfective *le* is not necessarily completed or has reached its natural endpoint" (2015: 277), and gives the example *wo zuotian xie-le xin, keshi mei xie-wan* ("I wrote a letter yesterday but didn't finish it"). Liu continues, "it conveys an arbitrary final point, not necessarily the natural final point of the event" (2015: 278). This is, however, in line with Comrie's description of the perfective presented above, in which the perfective does not necessarily denote the action as completed, but rather the completion of the situation.

The aspectual marker *gou* denotes a past experience which is discontinued in the current time. Liu describes it as "the experience of having engaged in an activity or participated in a situation" and illustrates this with the example *guo, ta kai-guo feiiji* ("He had the experience of flying an airplane" or "He (at some point) experienced flying an airplane") (2015: 280).

Richard Xiao and Tony McEnery (2004) argue that the perfective can also be expressed by reduplication to denote the delimitative aspect, or through resultative verb complements to denote the completive aspect.

1.4.12 Swahili

This section is based on *The Acquisition of Swahili* (2005) by Kamil Ud Deen, if nothing else is noted. Swahili is the most prominent Bantu language and is the native language of approximately 2 million people in Kenya and Tanzania, but it also used as a *lingua franca* and is as such spoken by an additional 70 million people in Eastern Africa. Swahili is an agglutinative language. A typical trait of the Bantu languages is the extensive use of prefixes, comparable to the role of suffixes in many European languages. Swahili, however, uses both prefixing and suffixing.

Every indicative utterance is marked for tense/aspect, but there is no grammatical distinction regarding the perfective/imperfective opposition. The past is marked by the prefix *li-* and this past tense marker was originally a copula.

The perfect, on the other hand, can be marked by either the prefix *me-* or the prefix *sha-,* which are positioned after the subject prefix. The two perfect markers are similar in meaning, both denoting present perfects of result, with the difference that *sha-* also

denotes a sense of completion. Deen gives following examples to distinguish them *Ni-me-kul-a* translates to “I have eaten”, and *ni-sha-kul-a* to “I have already eaten”.

Comrie points out that even though the present perfect marker *me-* can be translated into English Present Perfect, it will “often be most naturally translated into English as a stative Present, e.g. *a-me-choka* ‘he is tired’ (lit. ‘he has got tired’)” (1976: 57).

CHAPTER 2

The Bybee and Dahl Theory

2.1 Two studies, one result

In two independent studies based on different methodologies, both from 1985, Joan L. Bybee and Östen Dahl reached similar conclusions about verbal morphology on a cross-linguistic level. Bybee and Dahl both discovered cross-linguistic similarities regarding formal and semantic properties of tense and aspect notions, and, furthermore, correlations between formal and semantic properties. The similarities of their results led to their common study *The creation of tense and aspect systems in the languages of the world* (1989). Through an integration of their separate results, they aim at a better understanding of grammatical meaning of the languages of the world, and more precisely to understand how the meaning and expression of aspectual and temporal categories are related, both on an interlinguistic and an intralinguistic level.

Bybee and Dahl use the term “gram”, short for “grammatical morpheme”, which functions as an umbrella term including affixes, adpositions, clitics, auxiliaries, reduplication, stem change, ablaut and so on. The notion of “gram” facilitates their study, since it makes it possible to analyse both bound morphology and periphrastic constructions in one single domain, instead of the traditional split between morphology, which studies bound expressions, and syntax, which studies periphrastic expressions. This is highly relevant for their study, since part of their main thesis is that there are correlations between the meaning of a gram and its expression, that is, whether it is expressed through bound morphology or through a periphrastic construction. Another advantage with the notion of “gram”, specified by the authors, regards the difference between notional and grammatical morphology: by “gram” Bybee and Dahl refer to grammatical categories, hence categories with a certain meaning *and* a certain expression. Their goal is to postulate the existence of “a small set of cross-linguistic gram-types identifiable by their semantic foci and associated with typical means of expression” (Bybee and Dahl 1989: 52).

As stated by the authors themselves, the hypothesis that there could be cross-linguistically valid grammatical meanings goes against the general idea of American linguistics of the 20th century, in which the idea of semantic universals is not recognized, against structuralism, according to which grammatical meaning is language-specific, and against the Chomskyan view, which treats grammatical and semantic description as autonomous to one another. Bybee and Dahl, however, note that, since about a decade before the publication of their common paper in 1989, contrasting theories had started to emerge, such as Comrie's *Aspect* (1976) and *Tense* (1985). This opened up for the possibility of a "cross-linguistic understanding of grammatical meaning" (Bybee and Dahl 1989: 53), possibly even at a universal level.

In the following section I will give a short summary of the two studies which lie as a basis for the paper from 1989. I will thereafter present the common theory, and finally refer to two more recent studies on the matter, by Dahl and by Bybee, respectively.

2.2 The two independent studies

Morphology: A study of the relation between meaning and form (1985) by Bybee is an attempt to test hypotheses according to which there is a relation "between the meaning of inflectional grams and the degree of fusion they exhibit with a lexical stem" (Bybee and Dahl 1989: 53). Bybee used a world-wide sample with 50 randomly selected languages, to avoid possible genetic and areal biases. The goal was to match verbal inflection with one of the so-called "super-categories", namely valence, voice, aspect, tense, mood and agreement, and, furthermore, to identify to which category the meanings expressed by these inflections belong, such as perfective, imperfective, past and future, as traditionally identified. The study relied mainly on reference grammars, which often already contained labels for the different categories. It was, however, necessary to reanalyse them, to obtain a cross-linguistically valid comparison.

Dahl's study *Tense and aspect systems* (1985) differs both for the scope and for the source of information. Firstly, regarding the scope, since Dahl focused on the way tense and aspect grams are used, he paid less attention to how they are expressed. Secondly, the data were collected with the help of a questionnaire made up of 150 sentences focused on tense, mood and aspect. These sentences were then translated by native speakers of 64 different languages, with the aim to investigate the use of tense and aspect grams. Measures were taken to avoid interference as much as possible from

English, which was the language of the original questionnaire. There was, however, a certain bias towards European languages among the 64 selected languages, partly compensated through a smaller sample of 18 well-selected languages, used as a control sample. The next step was to label the predicates expressing tense or aspect, and to transfer these data to a database for inter- and intralinguistic comparison. Dahl found, even though this was not the main goal of his study, a certain pattern regarding the expression of the grams, namely that perfect and progressive grams tend to be expressed periphrastically, in 88% and 95% of the cases, respectively, and that past, perfective and imperfective grams tend to be bound, in 73%, 85% and 100% of the cases, respectively. Future grams, on the other hand, turned out to be almost equally split between the two means of expression.

As mentioned, the two studies reached similar results regarding the verbal morphology, namely that "certain correlations between meaning and mode of expression exist for grams viewed cross-linguistically" (Bybee and Dahl 1989: 53). Both found that between 70% and 80% of the grams marking tense or aspect, found in the sample and in the translated questionnaires, respectively, can be divided into the following six gram-types: perfective, imperfective, progressive[7], future, past or perfect[8]. Bybee considered the gram-type "present" as unmarked, and Dahl as a default member. Minor gram-types could include combinations of the major ones, such as perfect or progressive combined with past tense. Unique gram-types, existing in single languages, were also found. The six major gram-types resulted, however, by far to be the most common ones.

Following these results, Bybee and Dahl acknowledged the possibility to make further generalizations regarding the semantic content of tense and aspect grams of the languages of the world, and, furthermore, regarding correlations between the meaning of a gram and its mean of expression.

2.3 The thesis: cross-linguistically similar grams

Bybee and Dahl's goal in *The creation of tense and aspect systems in the languages of the world* (1989) is to reach "a general theory of grams" (1989: 52), in which they present a small number of cross-linguistic gram-types, each with a certain meaning and

[7] Referred to as "continuous" in Bybee's study.

[8] Referred to as "anterior" in Bybee's study.

expression. I shall now present the theory more closely, and will refer to their common paper from 1989 if nothing else is noted.

In short, the authors exemplify this proposed form-meaning correlation by comparing perfects and progressives, on the one hand, and pasts, perfectives and imperfectives, on the other. Perfects and progressives, Bybee and Dahl argue, tend to have less general meanings, and show less grammaticalization[9] form-wise, as can be seen from their typically periphrastic constructions. Pasts, perfectives and imperfectives, on the other hand, have more general meanings and show more grammaticalization of form, as confirmed by their tendency to be expressed through bound morphology. In 2.4 I describe in what way the authors argue that, for example, perfects have a less general meaning, whilst perfectives have a more general meaning.

The idea that a periphrastic construction is less grammaticalized than a bound morpheme rests on the authors' thesis that inflectional grams are the most grammaticalized elements, and, therefore, possess the largest quantity of grammatical properties, compared to other constructions. Derivational expressions are, in this sense, considered to be in the middle of bound morphology and periphrastic expressions. The distinction between grammatical and lexical morphemes is, therefore, crucial. But let us first take a step back.

The concept of grammaticalization is central in this study of grammatical meaning. Grammaticalization can briefly be described as the process according to which elements of grammatical meaning develop from lexical material. Bybee and Dahl's specific take on grammaticalization is that the semantic and the phonological development of the lexical item run parallel: the gram evolves both through semantic generalization and through phonological reduction. Diachrony is, as can be understood, a key notion in Bybee and Dahl's theory, and for more than one reason. Firstly, for the simple fact that grammaticalization is a gradual process which takes place over time. Secondly, since the authors have found diachronic relations between gram-types, "a perfect tends to develop into a past or perfective, […] and a progressive tends to develop into an imperfective" (1989: 56). This fact is, in turn, related to the fact that the authors argue for cross-linguistically common sources of the different grams.

Since Bybee and Dahl's theory is based on the idea that lexical material can develop into grammatical material, through the process of grammaticalization, a closer

[9] I use the term "grammaticalization" where Bybee and Dahl (1989), and Bybee in general, use "grammaticization".

distinction regarding grammatical and lexical morphemes is necessary. This is also important to understand better why the authors consider inflectional grams to be more grammaticalized than derivational and periphrastic ones, sharing less properties with lexical morphemes. Just as grammatical meaning can develop out of lexical meaning through generalization, the idea is that a grammatical morpheme can continue to generalize. Bybee and Dahl claim that grammatical morphemes are structurally and semantically different from lexical morphemes, as they tend to be “highly general and relational, serving to relate parts of clauses or parts of discourse to one another” (1989: 63), compared to having specific and referential meaning, as lexical elements do.

As far as the semantic change of the gram is concerned, the authors argue that the components of the lexical meaning of the “source” is gradually lost, even though shades of it can be retained for a long time. They further state that there is a correspondence between the formal properties of the gram and its semantic change. The loss of lexical meaning is reflected in the rigidity regarding the position of the developing gram, which, in turn, has an effect on the possible semantic relations of the gram. The loss of lexical meaning has two main consequences on the gram: loss of semantic autonomy and generalization of meaning. Semantic autonomy regards the types of semantic relations a gram may take part in. Generality of the grammatical meaning regards the “absence of lexical and contextual restrictions” (1989: 63), since a more general meaning leads to a less restricted applicability, which, in turn, results in a “rapid increase in token frequency” (1989: 64). There is hence the possibility that the gram eventually will be able to combine with any kind of subject and verb. The authors illustrate this with the English verbs “want” and “will”. The former one tends to require an animate subject. The latter, on the other hand, which originally meant “to want”, has today generally lost this meaning of “an agent desires” and can now be used with all kinds of subjects. Furthermore, inflectional grams such as perfective, past and imperfective grams can usually be combined with all kinds of verbs, unlike expressions that signal, for example, iteration, which are limited to telic or punctual verbs and usually expressed through derivation. This is, therefore, an example of how inflectional grams share less similarities with lexical morphemes, compared to other temporal and aspectual means of expression.

The correspondence of the formal and the semantic properties of the gram have several expressions. The loss of autonomy can be observed as the distributional properties of the gram become more fixed, since grams, unlike lexical morphemes, are “not manipulable for semantic or pragmatic purposes” (1989: 62), and, furthermore, since

grams tend not to be independently modifiable. A more general meaning of the gram also makes it more dependent on its context and can structurally be defined as belonging to a closed class and as having "a fixed, or grammatically determined, position" (1989: 61)[10]. The notion of "closed class" is, as the authors point out, in itself a complicated notion. Classhood can, briefly put, be determined by formal properties: if a certain element is member of class or not depends, for example, on which positional constraints it obeys to.

Furthermore, Bybee (1985) noted that gram-types which are semantically more specific, such as habitual and continuous grams, tend to be further from the lexical stem and have less influence on stem changes, compared to less semantically specific gram-types. The reason for this would be that they have not developed as much yet, compared to, for example, perfective and imperfective inflections, which have "undergone a longer course of development" (1989: 57). It is more probable that an affix which has been attached to the stem for a longer period of time will have more effect on the stem regarding fusion and reduction. For a parallel, regarding the expression of modals, Bybee and Dahl refer to a paper by Bybee and William Pagliuca (1985) on universals of semantic change studied through areally and genetically unrelated languages. Bybee and Pagliuca show that modals with more specific meaning, e.g. agent-oriented modals, also tend to have periphrastic expression. More general modals expressing epistemic modality, on the other hand, may have bound expression. The reason for this would, again, be that the latter has undergone more semantic and phonological reduction, being older. Bybee and Pagliuca, like Bybee and Dahl, parallel the semantic reduction with phonological reduction. This correlation, between semantic and phonological reduction in grammaticalization, is further demonstrated in *The Evolution of Grammar* by Bybee et al. (1994).

Bybee and Dahl also list typical characteristics of inflectional grams, such as being obligatory in a certain grammatical context, which in the long run, following from increase of generality and frequency, may even lead to the gram becoming redundant in certain contexts. Their definition of an affix is that it has a fixed position, that open class items may not intervene between the gram and its head and, finally, that it is phonologically fused with the stem. Since the authors argue that the "phonological reduction and fusion mirror the loss of grammatical and semantic autonomy in grammaticization" (1989: 66), they see a direct link between affixation and semantic generalization, proposing that "for

[10] A gram can also have more than one position, what matters is that the positions are grammatically determined. Bybee and Dahl add, however, that a gram with more than one position could signify that the gram is not as grammaticalized as one with only one position.

each path along which tense and aspect grams develop, a semantic stage is reached which necessitates the expression [of] properties that lead to affixation" (1989: 67).

2.4 The universal paths

Bybee and Dahl hypothesize three common major paths along which gram-types of tense and aspect evolve (1989: 57):

1. expressions with a copula or possession verb plus a past participle, or verbs meaning "finish", "come from" or "throw away", develop into grams marking anterior or perfect, which in turn develop into perfectives or pasts;
2. expressions with a copula, locative or movement verb develop into progressives which in turn develop imperfectives;
3. expressions with a verb meaning "desire", "movement towards a goal" or "obligation" develop into grams expressing intention and future

There is, hence, a limited amount of sources which later can develop into one of the six gram-types. Bybee and Dahl note that each lexical source and successive grammatical expression can be found in at least three unrelated languages, which strengthens their hypothesis that there is a universal semantic basis for grammaticalization. According to this theory the original meaning of the lexical item has an effect on the path of development for the gram: each gram has "inherent semantic substance reflecting the history of its development as much as the place it occupies in the synchronic system" (1989: 97). In the languages treated in the first chapter of the current paper, some clear examples of these paths can be found. For example, the progressive marker *-(i)yor* in Turkish is an example of the "movement towards a goal" path; the perfective marker *le* in Mandarin Chinese gives an example of how the verb "finish" can evolve into a perfective; the tense marker *li-* in Swahili illustrates the path of pasts which originates from a copula.[11] More examples can be made regarding the well-studied Indo-European languages as well, of course, such as the development of the perfects and perfectives from an auxiliary plus past participle in the Romance and Germanic languages. Further strong evidence for the major paths is presented by Bybee et al. (1994), where each path is illustrated through a large number of languages.

The first out of the three major paths of development of grams regards the perfect, which is also the path which is the most interesting to this paper. This gram-type was

[11] See 1.4.9, 1.4.11 and 1.4.12, respectively.

found in approximately 25% of the languages of Bybee's sample and in 35% of Dahl's sample. Its most prominent characteristic, in the words of the authors, is that "the situation described in the sentence is viewed form the perspective of — or described as being relevant at — a later point in time, most typically the point of speech" (1989: 67). As stated, perfect grams are expressed periphrastically in more than 80% of the languages with a perfect in Dahl's study. Bybee and Dahl list four common types of periphrastic constructions and three languages in which they are found. The constructions also show the sources for perfects. They are as follows (1989: 68):

1. copula + past participle (or similar form) of the main verb (ex.: Hindi, Bulgarian, Tamil);
2. constructions based on original possessive constructions, e.g. auxiliary "have" + past participle of the main verb (ex.: most Germanic and Romance languages, North Russian dialects);
3. main verb + particle with an original meaning "already" (ex.: the Kwa languages Yoruba and Isekiri[12]);
4. constructions involving auxiliaries historically derived from verbs meaning "finish", or less frequently other lexical verbs, such as "throw away" or "come from" (ex.: Sango, Ewe, Fore)[13]

The authors focus on the first two constructions, which they describe as "resultative constructions to perfects'" (1989: 68). In fact, "resultative" is a typical reading of perfects, an indication of their common semantics and common past. Since resultatives tend to focus on the state of the result of the previous event, compared to perfects, which focus on the event itself, the development from resultative to perfect is interpreted as a de-emphasis of the present moment. Together with this semantic change there is a grammatical change: the particle of a perfect does less typically agree with the subject or object, unlike the particle of the resultative construction. This is coherent with the authors' hypothesis regarding the grammaticalization process, according to which the semantic and grammatical changes are parallel.

A following, and the most well-known, step along the path is the development of perfect grams to past or perfective grams. This, once again, happens through lexical generalization, which in this case implies the total loss of relevance when it comes to the present time. A past or perfective gram is, in fact, generally disconnected from the moment of speech. Bybee and Dahl state that this lexical generalization is, as expected, accompanied by generalization of form, which happens through affixation of the

[12] Bybee and Dahl refer to the Itsekiri language.

[13] The authors' references have been excluded.

periphrastic marker or deletion of the auxiliary: the periphrastic form becomes bound, in accordance with the hypothesized form-meaning correlations.

The reason for the form-meaning correlation regarding tense and aspect grams according to Bybee and Dahl lies, hence, in the "age" of the specific gram. A younger gram tends to have a periphrastic mean of expression, since it has not evolved as much as an older gram has, and is according to their theory, therefore, less reduced in its form. The meaning of a younger gram is, in the same way, supposed to be less general.

Progressive grams contribute to confirm their hypothesis. 95% of the progressives in Dahl's sample are expressed periphrastically, and their lexical sources tend to be transparent. If the source is transparent, it means that the lexical source has not undergone much semantic change, which should be a consequence of the gram having existed for a shorter period of time. An example of a transparent progressive gram is the Irish locative construction: copula + *ag* ("at") + verbal noun (1989: 78):

(1) *Tá sé ag dúnandh an dorais*
He is at shutting the door
"He is shutting the door (GEN)"

Further evidence for their hypothesis regards future grams, which, unlike the other major gram-types, do not seem to have a preference regarding a bound compared to a periphrastic expression. However, Bybee and Dahl argue for a difference in meaning of the two types of expression. The bound expressions are also used where the expression of the future tense could be redundant, which is also, according to the authors, where an older gram would be expected. An example of this is subordinate clauses introduced by "whatever", "if", "even if" or "when", where periphrastically expressed futures appears to be rare.

As mentioned in the previous chapter, Bybee and Dahl argue that among languages with an inflectional tense-aspect system, the most common type is the tripartite one, where there is a perfective past, on the one hand, and an imperfective divided into past and non-past, on the other. From Dahl's study of 1985 follows the universal that "all languages with inflectional tense or aspect have grammatical expression for past or perfective or both" (1989: 95), since the respective grams should have come far along the grammaticalization paths. Whether a perfect gram develops into a past or a perfective gram, depends on the existence of an imperfective gram in that language. If there is an imperfective gram, then the perfect can develop into a perfective

gram. On the contrary, if the language lacks a grammaticalized imperfective, the perfect will become a past, since it, unlike the perfective, is compatible with imperfective meaning. The former development can be seen in Modern French, where the Perfect is taking over the perfective past meaning, since there already existed an imperfective past gram. Bulgarian, on the other hand, is an example of a language in which the perfect still indicates perfect meaning, and has, as expected, also kept the auxiliary. Bybee and Dahl do, however, mention a couple of exceptions to the supposed universal: the isolating language Mandarin Chinese has a grammaticalized perfective even though it lacks inflection, and French, which has inflection, has a periphrastically expressed perfective.

Another note regarding types of tense-aspect systems concerns the Slavic languages. As discussed in the previous chapter, many Slavic languages differ from the typical tripartite tense-aspect system. This is partly for the origin of the perfective, which does not come from periphrastic constructions with an auxiliary, but from bounders, giving it a derivational expression. Their different origin led them on a different path, giving them a different semantic character: the Russian Perfective Past, for example, is better described in terms of boundedness than totality. However, according to the authors, at the same time, this strengthens their hypothesis of universal gram-types for perfective aspect, because of the overall similarity in the perfective meanings even though they have different origins.

Furthermore, grams in different languages in a particular moment can be at different points along one of these paths, which would also partly explain differences regarding the meanings of the grams of the world's languages. This is a fact which, according to the authors, should be taken into consideration when grams are compared. Bybee and Dahl further list other language-specific reasons to explain the diversity regarding the tense-aspect systems of the languages of the world: each gram can develop independently from the other; at each stage a language may have more than one gram with similar meaning; the possible combinations of grams may vary, and languages may also vary the possibility to combine less common gram-types with the major ones, etc. However, in order to explain how it is possible that such a limited set of gram-types occur in such a large number of languages, and hence at the basis of a possible universal theory of grammatical meaning regarding tense and aspect, Bybee and Dahl theorize, based on the idea that language change happens through language use, the existence of a "small set of highly generalized discourse or pragmatic functions served by tense and

aspect grams" (1989: 96). These are supposedly general to human communication and expressed on a language-specific level through tense and aspect grams.

I shall now refer to the each author's further development of the common thesis, based on *The Growth and Maintenance of Linguistic Complexity* (2004) by Dahl, and on the article *Language change and universals* (2006) by Bybee.

2.5 Dahl 2004: Predictability and redundancy

In chapter 2 of *The Growth and Maintenance of Linguistic Complexity* (2004) Dahl proposes a reading of the information theory, first laid out by the mathematician Claude Shannon (1949), from a linguistic point of view. Shannon's information theory regards "how signals (e.g. radio signals) are transmitted in a communication channel and the focus is how easy or difficult it is to predict what is going to be transmitted" (Dahl 2004: 7). A key notion here is "predictability". Its relation to information is that the level of predictability is inversely related to the amount of information which needs to be transmitted. That is, for something less predictable to be transmitted, the signal must contain more information. In language, this regards both syntactic and semantic information, and both are, Dahl argues, relevant when a linguistic item becomes more frequent. For a linguistic item to become more frequent in the first place, it is necessary that it is used independently of its amount of semantic information, that is, independently of whether it bears new information or not. As it becomes more frequent, it will also become more predictable, and a highly predictable item has little syntactic information.

Predictability and redundancy are two crucial aspects regarding the complexity of linguistic constructions. A redundant message is one that could have been formulated in a less "expensive" way, that is, using less communicative resources. However, it is more probable that a redundant message will be understood correctly, and Dahl argues that "spreading out information is beneficial to safe transmission" (2004: 10). This is illustrated through co-articulation. Co-articulation, when one phoneme influences another, is often thought of as a matter of ease of pronunciation. However, it can also be interpreted as "spreading out the information", in this case phonemic information, with the effect that the listener can more easily understand the message in case of disturbing elements. But there is another side to this: when the information is spread-out, there is also a bigger difference regarding the mapping from input to output. The mapping has become more complex.

Redundancy can be created in different ways, and some ways are more efficient than others. The example of co-articulation shows how the information of a message becomes more redundant, without making the form of the message any longer, it is an example of "smart redundancy" in Dahl's words. Dahl refers to the balance between costs and benefits as "redundancy management", and differentiates user-level redundancy management from system-level redundancy management. The latter regards the redundancy that the grammar of each language bears and, therefore, does not increase the processing costs as much, since the listener finds it more predictable. For example, in English the plural morpheme "-s" in "five books" is redundant, since "five" already conveys plurality. In Hungarian a noun followed by a numeral remains in singular even if the numeral is plural. For example:

(2) *egy könyv* ("one book")
öt könyv ("five books")

Furthermore, the increase in frequency of an item may lead to it being over-used, which in turn can lead to the item losing informational value. This in turn can have effects on the redundancy management: "the speaker will be less motivated to spend as much energy on producing it as before, which may lead to phonetic reduction" (Dahl 2004: 17). Dahl argues that grammatical maturation[14], and linguistic maturation in general, involves a kind of "information spread-out". In this case it is a grammatical construction or marker which is spread to new domains, and this spread would further lead "to a decrease in the rhetorical and/or informational value" (2004: 121). Dahl calls this "pattern spread". An example of this is a perfect which becomes a perfective or a general past. A pattern spread is followed by a "pattern adaptation", making it more adapt to its new uses, for example through reduction or condensation. A perfect is less grammaticalized than a perfective or a general past, and further less frequent, hence less reduced.

Next to redundancy management, Dahl discusses "prominence management", which, briefly put, implies that if a part of a message is difficult to recover, or if it is critical to the message, then it should be enhanced, and likewise, it should be reduced if it is easy to deduce or less important. Both redundancy and prominence management hence regard what the listener can infer and what he or she expects to hear, and are considered by Dahl to be the main forces when it comes to sound change. By "sound change", Dahl intends "adaptive sound change", which is "a reaction to the changed role of an

[14] Dahl (2004) uses the term "grammatical maturation" as a broader notion for grammaticalization.

expression, when it is trapped in a complex construction, or when the construction of which it is a part expands its territory" (2004: 158). It is hence a reaction to the expression becoming more frequently used or acquiring new meanings. So, an adaptive phonetic reduction is intended as "a response to a decrease in the informational or rhetorical value of the expression" (2004: 159). Dahl further stresses that the informational value of an item does not directly depend on it acquiring grammatical meaning or not, but the change of its role in general.

Frequency is, according to Dahl, connected to grammaticalization mainly in two ways: a frequent item is more resistant to new patterns, but, on the other hand, is also more probable to be subject to reductive change. The latter regards an element with little informational value as well. In grammatical maturation, there is often a "split" of the uses of an item: it will become reduced only in one of them. An example if this is the English indefinite article from the numeral "one". Dahl further argues that:

> [i]f reduction were directly dependent on the token frequency of an item, it ought to follow that all tokens of the item would be equally reduced, which is clearly not the case (2004: 159-160)

The maturation process is, according to Dahl, more "dialectic" than cyclic. As patterns spread, the redundancy increases, which is counteracted by "smart redundancy", and not simply by going back to the initial state. In this way, later states bear information of previous states. According to Dahl, linguistic patterns, perhaps rather than languages themselves, can be more or less complex, and can further become more complex, more mature, as time passes. A bound expression is more mature than a periphrastic expression, and Dahl, referring to his previous study from 1985, states that there is less variation among the more mature elements. The sources are many, but the resulting gram-types are cross-linguistically similar.

2.6 Bybee 2006: The role of repetition and inference

In *Language change and universals* (2006) by Bybee, the importance of a diachronic aspect is further emphasized: each synchronic pattern bears a diachronic dimension and "true universals of language are not synchronic patterns at all, but the mechanisms of change that create these patterns" (2006: 179).

> Linguistic theories that just compare synchronic states are like a hypothetical biological theory that compared the seeds of various plants with no regard to what type of development trajectory a seed is meant to embark upon. Such a theory would also study a seedling without considering its past and future development (Bybee 2006: 183)

Bybee focuses on the mechanisms which lie as a basis for the paths that lead to grammatical meaning. The role of repetition and the role of pragmatic inference[15] are especially stressed. In short, the idea is that there are universal mechanisms of change, which are operative as language is being used: repetition leads to automatization, habituation and conventionalization of pragmatic inferences, which, in turn, produce universal paths of change. The role of these mechanisms will now be further deepened.

2.6.1 Repetition: grammaticalization as automatization of frequently occurring sequences

As mentioned in previous sections, Bybee and Dahl (1989) argue that as the developing gram generalizes in meaning there is an increase of its frequency. Bybee (2006) further states that "[t]he frequency increase is both a result of the process and a contributor to it, as repetition has certain effects on neuromotor and cognitive representations" (2006: 187). According to Bybee, the capacity of language is, in fact, to be considered in correlation to the rest of the neuromotor and cognitive abilities, and, therefore, regards processes such as automatization, habituation and categorization. On this note, she refers to a study by Joyce Tang Boyland in 1996, in which a parallel between the grammaticalization process and the automatization of non-linguistic skills is drawn.

The repetition of linguistic sequences, Bybee argues, takes place in all languages of the world at all time, and leads to automatization and habituation of the sequences. The repetition of a sequence may finally lead to the sequence being perceived as a single unit: its form and identity reduce as its components reduce and overlap, just like when learning how to drive a car. The sequence of actions, whether it be articulatory gestures or driving, becomes automatized. Repetition also leads to the meaning becoming more general, "bleached", since the force of the word or linguistic sequence is gradually lost. A more general term can be used in a larger amount of contexts, leading to a frequency increase, which, in turn, was why it became more general in first place. Repetition, hence, explains both why grammatical morphemes tend to be phonologically reduced and dependent on lexical material, and why grammatical meaning tends to be highly general.

---

[15] The role of inference is also treated in Bybee and Dahl (1989) regarding the development of evidential markers in perfects.

2.6.2 The role of pragmatic inference

We do not always overtly express everything we wish to communicate, a fact which is partly compensated by the use of inferences. The role of pragmatic inference in grammaticalization has been well studied by Elizabeth Traugott, as noted by the author. Bybee further refers to *Grammaticalization* (1993) by Paul Hopper and Traugott, regarding how the meaning of "be going to" already conveyed more meaning of purpose than of movement in this example from *Two Gentlemen of Verona* (1595, III.i.51) by William Shakespeare:

(3) *Duke* Sir Valentine, whither away so fast?
Val. Please it your grace, there is a messenger
That stays in to bear my letters to my friends.
And I am going to deliver them.

Bybee continues, "[w]hen the same pattern of inference occurs frequently with a particular grammatical construction, those inferences can become part of the meaning of the construction" (2006: 189). Even though the Duke asks where Valentine is going, and even though Valentine answers in terms of movement, the underlying message here regards Valentine's intention. The connection between the lexical source "movement towards a goal" and intention is part of one of Bybee and Dahl's three major universal paths of change, as discussed in 2.4, which according to Bybee is "evidence that speakers in different cultures tend to infer intentions in the same context" (2006: 189).

Following Joseph H. Greenberg, Bybee refers to universals of language as "the mechanisms of change that propel the constant creation and re-creation of grammar" (2006: 179-180).

CHAPTER 3

The Haspelmath Theory

In this chapter I present the theory which follows from *Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability* (2020) by Martin Haspelmath, according to which certain cross-linguistic regularities are best explained through a functional-adaptive approach. This is further complemented by previous papers such as *Can cross-linguistic regularities be explained by constraints on change?* (2019).

3.1 Grammatical coding asymmetries

Languages can be more or less efficient in their coding, that is, they can allow the speaker to be more or less explicit in their utterances, and leaving, therefore, more or less to be inferred by the listener. In the paper from 2020, Haspelmath divides the coding systems of the world's languages into efficient and non-efficient coding,[16] or asymmetric and symmetric coding, respectively. What differentiates the asymmetric coding is that it takes advantage of "the additional efficiency that is provided by asymmetric coding" (2020: 5), for example, by letting only the nominal plural and not the singular form be overtly coded, as can be seen in English regular nouns: "book" vs. "book-s". Symmetric coding, on the other hand, can be divided into "uniformly explicit", in which two constructions are equally coded, and into "uniformly parsimonious", in which both constructions are uncoded. In symmetric coding, explicit expression of grammatical meanings or saving coding energy is hence prioritized to efficient coding. As an example of uniformly explicit coding Haspelmath mentions Modern Greek, and of uniformly parsimonious coding, Mandarin Chinese (2020: 5):

(1) Modern Geek (symmetric overt)
SG *vivlí-o* ("book")
PL *vivlí-a* ("books")

[16] Each one, however, described as "optimal in their own way" (2020: 2).

(2) Mandarin Chinese (symmetric zero)
SG *shū* (“book”)
PL *shū* (“books”)

Haspelmath, as can be evinced by the title of his paper, focuses on grammatical coding asymmetries, which, in the author’s words, can be described as a “minimal grammatical opposition […] where one member is typically zero-coded (or shorter), while the other member has an overt coding element or contains more segments” (2020: 1). The minimal difference is a semantic opposition, such as the nominal difference between singular and plural, or the predicational one between present and future tense. The existence of asymmetric coding was, as Haspelmath notes, first treated by Joseph H. Greenberg in 1966. Greenberg considered these asymmetries to be manifestations of markedness. According to Haspelmath, however, markedness cannot provide an adequate explanation to these asymmetries, since “there is no unitary and generally recognized markedness concept” (2020: 18). Instead of the term “markedness”, Haspelmath (2006) argues for referring to notions such as “overtly coded”, “frequency of use”, “phonetic difficulty”, etc., depending on the sense in which “markedness” is used.

To illustrate how grammatical coding asymmetries can be expressed, Haspelmath inserts a table with ten different examples, regarding both the nominal and the predicational domain, in English or Spanish, and are as follows:

Table 1*: Examples of universal grammatical coding asymmetries.* (2020: 2)[17]

singular	plural	(*book — book-**s***)
nominative (A/S)	accusative (P)	(*he — him-**s***)
allative	ablative	(to — **from**)
positive	comparative	(*small — small-**er***)
present	future	(*go — **will** go*)
affirmative	negative	(*go — **don’t** go*)
inanimate patient	animate patient	(Spanish *Ø la casa — **a** la mujer*)
3rd person	2nd person	(Spanish *canta.*3.SG / *canta-**s.***2.SG ‘sing(s)’)
2nd person imperative	3rd person imperative	(*praise! — **let** her praise!*)
attributive adjective	attributive verb	(*small — play-**ing***)

[17] Following Haspelmath (2020), the element of the overt coding have been made bold throughout the current chapter.

These are not only language-specific examples of coding asymmetries, but Haspelmath argues that they are also examples of universal grammatical coding asymmetries, or "strong universal tendencies" (2020: 2). To justify this claim, he refers to existing literature, starting from the first universals which were formulated by Greenberg (1963; 1966).

Based on the claim that the ten above-mentioned examples illustrate universal tendencies, Haspelmath urges the need of a general explanation, which is also the purpose of the paper. The explanation proposed is based on previous work by Greenberg, William Croft and John A. Hawkins. What Haspelmath argues for can, in John Haiman's (1983) words, be described as an "economic" explanation. Haiman argued that "reduction of form is an economically motivated index of familiarity" (Haiman 1983: 802). That is, if something is familiar or expected, it can be expressed in a reduced form, saving coding energy without sacrificing clarity. The two authors differ, however, regarding the role of iconicity, since Haiman considers iconicity and economy to be "competing motivations" when it comes to linguistic expressions. In a previous paper, *Explaining alienability contrasts in adpossessive constructions: Predictability vs. iconicity* (2017), Haspelmath argues that, regarding adpossessive constructions, "[m]y explanation is quite different from Haiman's in that it makes no reference to iconicity, but only to frequency of use, hearer expectations (= predictability) and length of coding" (2017: 194). I will come back to Haspelmath's view on iconicity in 3.3. Both Haiman and Haspelmath refer to George K. Zipf. Zipf famously stated that "the length of a word tends to bear an inverse relationship to its relative frequency" (1935: 38), Zipf's Law of Abbreviation. From this follows Haspelmath's form-frequency correspondence principle: "[l]anguages tend to have shorter forms for more frequent meanings" (2020: 2). What Haspelmath presents in the paper from 2020 is an extended version, which regards grammatical constructions instead of word length in general, and goes as follows (2020: 2):

> **The grammatical form-frequency correspondence hypothesis**
> When two grammatical construction types that differ minimally (i.e. that form a semantic opposition) occur with significantly different frequencies, the less frequent construction tends to be overtly coded (or coded with more segments), while the more frequent construction tends to be zero-coded (or coded with fewer segments), if the coding is asymmetric

Haspelmath thereafter proposes the following causal chain: "frequency of use —> predictability —> shortness of coding" (2020: 2). In other words, a frequently used construction is more predictable and, therefore, necessitates less explicit coding. This

casual chain is further extended to include context as well, found at the base of the chain along with frequency, since there is, of course, meaning which can be predicted based on the context. Contextual predictability can be related to something which has just occurred, and Haspelmath illustrates this with the following examples: “The girl went to the river, **she** looked for fish” and “The girl went to the river and **Ø** looked for fish” (2020: 19). Instead of repeating “the girl”, it can be replaced with a pronoun or omitted. The “counter-efficient pattern”, on the other hand, with the short form “she” at the beginning, is not possible.

To strengthen his hypothesis, 25 asymmetric pairs of different grammatical constructions from languages of the world are analysed regarding length and frequency. The constructions can be divided into “simple meaning pairs”, as the ones seen above in (1) and (2), in which there are two contrasting grammatical meanings, such as singular and plural, and into “differential-coding pairs”. In the latter, the different expression of a single grammatical meaning depends on its grammatical context or on the lexical subclass. Both types will now, for the sake of clarity, be more carefully illustrated.

3.1.1 Example of a simple meaning pair

In the paper from 2020, 13 different simple meaning pairs are illustrated, of which seven are from the nominal domain and six from the predicational domain. Here below is the one regarding present tense compared to future tense (2020: 8):

(3)	English	Latin	Kiribati[18]
PRS	*they praise*	*lauda-nt*	*e taetae* (“he speaks”)
FUT	*they **will** praise*	*lauda-**b**-unt*	*e **na** taetae* (“he will speak”)

The three languages in (3) show the universal tendency according to which the present tense is zero-coded, or at least shorter compared to the future tense.[19] Present tense-forms are also more frequent,[20] in accordance with the form-frequency correspondence explanation. The other twelve examples show the same pattern: the more frequent form is also the shorter one.

[18] Kiribati is an Austronesian language.

[19] Haspelmath refers to Bybee (1994).

[20] Haspelmath refers to Greenberg (1966).

3.1.2 Example of a differential-coding pair

As mentioned above, differential-coding pairs express the same meaning but in different contexts or in different lexical subclasses. There are hence two types of differential-coding: split-coding and subclass-conditioned coding, respectively. Haspelmath names differential object marking as a well-known example of differential coding and illustrates it with the following examples in Spanish (2020: 10):

(4) a. *Veo la casa*
I.see the house
"I see the house"

b. *Veo **a** la mujer*
I.see ACC the woman
"I see the woman"

The patients in the two examples, the house and the woman, respectively, differ in being inanimate or animate, and the claim is that inanimate patients are more frequent than animate patients, and therefore require less coding.

The idea is that the form-frequency correspondence when it comes to differential-coding pairs can be seen as a "USUAL ASSOCIATION of a grammatical meaning with a grammatical context or a lexical subclass" (2020: 11), as opposed to more unusual associations which are in need of more coding.

3.2 Language change

Haspelmath argues for a functional-adaptive explanation to the universal coding asymmetries. Accordingly, language asymmetries, and language change in general, are the result of adaptation. Languages change over time, as speaker utterances vary and novel constructions are created, or old markers eliminated. What Haspelmath argues is that "language systems favour efficient coding" (2020: 2-3), where coding energy is spent only when necessary. This, however, does not exclude the possibility that a certain language might do differently, as seen with the symmetric coding in Modern Greek and Mandarin Chinese. Haspelmath specifies that the hypothesis is not to be applied to particular languages, but is to be applied at the level of language universals, since "language histories are subject to a large number of contingencies, and the adaptive forces are relatively weak" (2020: 6). Counter-efficient coding, on the other hand, in which

the less frequent construction is zero-coded or shorter and the more frequent one is the construction which contains more segments, are not expected, and are “virtually unattested” (2020: 23).

One might wonder how a language can adapt to the need of the language-users. To answer this, Haspelmath points to one of the defining qualities of language, its malleability, and how “system-external functional-adaptive forces can have minute effects on the behaviour of individual speakers in individual utterances, which eventually have the cumulative result of creating efficient systems” (2020: 3). The notion “adaptation” is hence intended in the sense of “user-friendly” constructions, which facilitate the mutual understanding between speaker and listener without being too costly. This is, therefore, a result-oriented explanation, in which different paths of change can “lead to uniform outcomes” (2020: 21), namely strong universal tendencies.

Haspelmath, accordingly, lists different ways in which frequent linguistic material is shortened. One mechanism is clipping, and as Haspelmath states, it was first discussed by Zipf (1935). Clipping is, however, not a “regular diachronic process” (2020: 16). Another relevant mechanism is phonetic reduction, studied by Bybee, as Haspelmath notes.[21] However, neither of these two are part of the mechanisms which, according to the author, can be held responsible for the majority of the asymmetric coding patterns, which “in most cases […] are the result of differential development of a new construction” (2020: 16-17), and the expansion of the new construction in turn depends on predictability. Haspelmath illustrates three main ways in which asymmetric coding may arise. The two first regard the creation of a new construction resulting from the creation of a novel element, which can either arise only for one member of the pair, whilst the other does not take on a corresponding marker, or it can simply spread to one member and not to the other. The first case is illustrated with the English future-tense marker “will” (2020: 17):

(5) *they will praise* (future, will from “want”)
they Ø praise (present tense)

The second case can be illustrated with the German definite article, which is a relatively new development, and which did not spread to possessed nominals. A possessed noun is usually also something which is definite, its definiteness is hence predictable (2020: 17):

[21] See chapter 2 for a more detailed description of the role of phonetic reduction.

(6) *das Buch* (“the book”)
mein Ø Buch (“my book”)

Thirdly, Haspelmath mentions the elimination of a marker for one of the members if it is no longer “needed”, as can be seen with the German masculine nouns of the *-n* class. In Middle High German, these nouns were overtly coded for the difference between nominative and accusative. In Modern German, on the other hand, only animate nouns mark this case difference (2020: 17):

(7)		NOMINATIVE	ACCUSATIVE	
(a)	Middle High German	*knote-Ø*	*knote-n*	(“knot”)
		affe-Ø	*affe-n*	(“ape”)
(b)	Modern German	*Knoten*	*Knoten*	(“knot”)
		Affe-Ø	*Affe-n*	(“ape”)

Haspelmath further indicates a second possible way to explain universal tendencies, namely through “mutational constraints”. “Mutational constraint” in the linguistic field is a term introduced by Haspelmath (2019), and can be defined as constraints on language change. The causal factor of the change is hence found in the change itself. Haspelmath (2019) compares mutational constraints to functional-adaptive constraints, by defining the former as source-oriented, and the latter as result-oriented: a mutational constraint regards “what is preferred or necessary in language change” (2019: 6), and a functional-adaptive constraint “what facilitates communication” (2019: 6). Haspelmath (2019) argues that these constraints on change are stronger evidence than the evidence of recurrent pathways of change when it comes to explaining universal tendencies. The author states that, for example, even though the pathways regarding the development of perfectives proposed by Bybee (2006)[22] are widespread, “nobody knows how widespread they are” (2019: 10). Other sources to perfectives, possibly also more common, than perfects[23] can, therefore, not be excluded. Another source would indicate a different path.

[22] Originally proposed by Bybee and Dahl (1989).

[23] Referred to as “anterior” by Haspelmath.

3.3 Criticism and alternative theories

Haspelmath proposes, as has been illustrated, functional-adaptive explanations to the existence of most of the universal coding asymmetries. He, however, makes it clear that this is not to be confused with traditional functional explanations, such as the ones in which the coding asymmetries would have a meaning-form related explanation. The crucial factors are predictability and efficiency, not iconicity. In Haspelmath (2006), the author gives examples of language phenomena which can be explained through a predictability focused explanation, and not through iconicity of distance as intended by Haiman: "[t]he linguistic distance between expressions corresponds to the conceptual distance between them" (Haiman 1983: 782). Haspelmath discusses possessive markers together with inalienable and alienable nouns: the length of possessive person forms tend to be shorter with inalienable nouns, such as body part-terms, compared to alienable nouns, such as "house" or "car" (2006). According to Haimain's iconicity explanation "the additional element in alienable constructions should occur in the middle between the possessor and the possessed noun" (Haspelmath 2006: 209-210). This means that the existence of languages which differentiate the possessive marking of inalienable and alienable nouns, and in which the extra element in the alienable construction is not found in the middle would contradict Haiman's hypothesis. Haspelmath gives four examples from four different languages of this phenomenon. The following example from Puluwat[24] was also discussed by Haiman (1983) (Haspelmath 2006: 210):

(8) a. alienable construction
nay-*iy* *hamwol*
POSS-1.SG chief
"my chief"

b. inalienable construction
pay-iy
hand-1.SG
"my hand"

The predictability hypothesis, on the other hand, only regards length and not the linear order of the elements, and is hence not contradicted by the above-mentioned example.

A criticism to functional-adaptive explanations is the lack of evidence for the existence of the functional-adaptive constraints, which, unlike the mutational constraints,

[24] An Oceanic language.

are not observable through language change. Sonia Cristofaro (2017) argues that a functional explanation, such as economy or processing ease, to an implicational universal, such as the distribution of overt coding regarding singular and plural noun, implies that there is a dependency relationship between the singular and plural form. Namely that the overt coding of the more frequent one, in this case the singular form, is only realized if the less common one is also overtly coded. The problem, Cristofaro argues, is that there is no diachronic evidence for such a relationship, and, since she considers diachronic evidence necessary to explain language universals, this is crucial. Cristofaro (2017), furthermore, argues that another problem with functional explanations regards the fact that the phonological reduction which leads to one of the two forms of a pair being zero or less coded can also affect the less frequent one. Moreover, the reinterpretation of an already existing element leading to an overt marker, can affect the less frequent as well as the more frequent one. May it be so, but Haspelmath (2020), on the other hand, argues for a strong tendency regarding the difference in relative frequency of asymmetric pairs, and according to this tendency the patterns in which there is overt coding only of the more frequent one are "virtually unattested" (2020: 23). Cristofaro finally states that there is no evidence that functional principles actually do play a role in the conformation of language, in contrast to adaptive traits in evolutionary biology.

Haspelmath (2019) motivates the functional-adaptive explanation based on the existence of "multi-convergence", that is, different pathways leading to a uniform result. He argues that these can only be explained with functional-adaptive explanations. Cristofaro (2017; 2019), as Haspelmath notes, recognizes that different processes can have the same result, and she further stresses the possibility that these processes may have reached that same result for different reasons, and further that these differences must be taken into account. Cristofaro argues for a source-oriented explanation to universals, in which the focus is the source constructions and the developmental processes. According to Haspelmath, if different processes lead to the same result for different reasons, then it is no longer a matter of universal tendencies, it is simply a matter of accident.

Haspelmath (2019) refers to Bybee (1988) and Cristofaro (2017) regarding their common criticism against a functional explanation to universal tendencies. According to Bybee (1988) and Cristofaro (2017), for a functional explanation to be valid, its role in the final result must also be explained. According to Haspelmath, this is not necessary if the explanation to a universal tendency is believable, much like in other human fields: "when

the result is preferred, any kind of change can give rise to the result, and we do not need to understand the nature of the change" (2019: 15). He further draws a parallel with evolutionary biology to illustrate how functional-adaptive changes are not directly linked to the adaptive change *per se*, but a result of functional selection: favourable features are spread and survive.

Finally, regarding what could be the "ultimate cause", Haspelmath addresses the possibility that it might not be frequency. It could be the other way around: certain constructions could be more frequent because they are shorter. However, if this were the case, the author argues, there should be no frequency differences in languages with symmetric coding, which is not the case. This is option is thus excluded. Along the same line, there is also the possibility that there is something else which causes the frequency in first place, an unknown "factor X", but since the author cannot think of a good candidate this possibility is abandoned.

In the following chapter, the Haspelmath theory will be put to test by looking at the token frequency of present perfect and perfective past or simple past constructions in four languages.

CHAPTER 4

The Corpus Research

4.1 Methodology and aim

To test Haspelmath's idea, according to which the more frequently used constructions tend also to be the shorter ones, this chapter will be focused on corpus-based research. The relative frequency of present perfect constructions to perfective past constructions will be compared, or that of present perfect constructions to simple past constructions, depending on the language in question. The focus here is the form and the length in general, that is, whether the construction is bound or periphrastic. I have chosen 30 verbs which have similar meaning in each language, even though it is, of course, a very complicated task to find perfect synonyms. I looked at regular and irregular verbs in the affirmative first person singular form. The verbs are of varying aspectual meaning, following *Linguistics in Philosophy* (1967) by Zeno Vendler. Verbs with possible readings of activity, accomplishment, state and achievement have hence been included. This is, of course, a matter of *Aktionsart*, that is, it regards the inherent aspectual meaning of a verb, and not grammatical aspect, the focus of the current paper. It is, however, relevant, since grammatical aspect and *Aktionsart* interact. Certain *Aktionsarten* are more frequently used in one aspect compared to in another. For example, verbal constructions of achievement, such as "reach the top" (Vendler 1967: 146), can describe punctual and telic situations, and are therefore more compatible with perfective constructions, as opposed to an imperfective aspect.

The languages which will be examined are English, Swedish, Spanish and Portuguese. This is a rather limited language sample, since they are all Indo-European languages from either the Germanic or the Romance group. The main characteristics regarding the present perfect and perfective past expressions and uses of each language have already been described in chapter 1. The matter will, therefore, not be treated in a detailed manner here. I will, however, describe which measures I took for each language for the results to be as realistic as possible.

I used the Mark Davies' corpora of English, Spanish and Portuguese. The intention was to look at a spoken subcorpus for each language, since spoken language, compared to written language, is less conservatory and gives a better idea of which changes are going on in a language. I was, however, able to use spoken corpora only for English. Regarding Spanish and Portuguese, I had to prioritize examining only Peninsular Spanish and European Portuguese, respectively, because of the variation regarding the way in which present perfect and perfective past are used in Spain and Portugal, respectively, compared to in other countries where the languages are spoken. This matter is more closely discussed in chapter 1. I used Språkbanken (The Swedish Language Bank) for the Swedish corpus research. The only existing spoken corpus was, however, too limited. The different corpora and the method will be discussed more in detail in the following sections, divided for language.

The results will be presented comparing the occurrences of each verb expressed in present perfect and perfective past, or in present perfect compared to simple past, depending on the language examined. I am hence interested in the relative token frequency, and not in the absolute frequency, which would be related to the total corpus. Different verbs have different frequencies for extralinguistic reasons, regardless of the tense or aspect they are expressed in. People eat more frequently than they applaud, which is reflected in the absolute token frequency of each verb. I will include tables showing the three verbs which were most commonly expressed in present perfect and in perfective past constructions, or in present perfect and in simple past constructions, for each language. The total percentage regarding the sum of the 30 verbs was, of course, not calculated based on token numbers per verb, but on each verb's frequency regarding the two constructions. A more detailed account of the results including all the verbs can be found in the appendix.

4.2 English

There is no grammatical expression of the perfective in English. The frequency of verbs expressed in Present Perfect was, therefore, compared to the frequency of Simple Past. The former has a periphrastic construction and the latter is bound. Since there are differences in American and British English regarding this matter, as noted in 1.4.1, the two varieties were tested separately. The method was the same: expressions of Present Perfect and Simple Past were searched for in first person singular. A pronoun was added

to be able to differentiate Simple Past from Present Perfect, since the English expression of Present Perfect "contains" Simple Past when the past participle and Simple Past have the same expression. That is, by searching for "worked", the results will refer to both Simple Past and Present Perfect. I chose first person singular since third person singular in the contracted form of Present Perfect can also refer to a passive construction. "She's hit" could have an active reading, "She has hit the ball", or a passive one, "She was hit by the ball". Both the contracted and the non contracted form of Present Perfect were included. Finally, I also included searches with an intervening adverb, between the auxiliary and the past participle in the Present Perfect construction, and between the pronoun and the verb in Simple Past. Six separate searches were thus amde for each verb. For example, with the verb "work" I searched for: "I have worked", "I've worked", "I have ADV worked", "I've ADV worked", "I worked" and "I ADV worked". The four separate searches for each verb in Present Perfect were then summed, and compared to the sum of the two separate ones in Simple Past.

4.2.1 American English

I used the Corpus of Contemporary American English (COCA henceforth), which is a corpus of mixed genres. It dates from 1990 to 2019 and contains an estimate of 1 billion words. The spoken subcorpus, which I used, contains over 127 million words of mostly unscripted conversations from more than 150 different TV and radio programmes, it is hence only broadcast English.[25]

Among the 30 examined verbs, only one of the Present Perfect/Simple Past pairs turned out to be more common in the Present Perfect, "admit". It should also be added that the difference between the occurrences of the Present Perfect construction and of the Simple Past construction of "admit" was minimal: 51% compared to 49%. Except for four other verbs together with "admit", the remaining 25 verbs were found to be expressed in the Present Perfect in 70% of the cases, or more. The other four verbs are: "live", "hear", "learn" and "applaud".

A note should be made regarding the verb "learn", since it can be conjugated both as a regular verb, "to learn, learned, learned", and as an irregular verb, "to learn, learnt, learnt". As expected, few occurrences of the irregular conjugation were found in COCA, all in the Simple Past. The irregular conjugation of "learn" is, in fact, more common in British English.

[25] https://www.english-corpora.org/coca/help/coca2020_overview.pdf

Table 1: *The three verbs most commonly expressed in American English Simple Past and the three verbs most commonly expressed in American English Present Perfect, out of the verbs in the sample.*[26]

	Tokens in Present Perfect	Tokens in Simple Past	Percentages
open	10	455	2% vs. 98%
knock	2	75	3% vs. 97%
walk	56	1464	4% vs. 96%
.....			
hear	4162	7048	37% vs. 63%
applaud	5	8	38% vs. 62%
admit	28	27	51% vs. 49%

The total percentage of verbs expressed periphrastically, in the Present Perfect, was 17%, compared to 83% in the Simple Past.[27]

4.2.2. British English

I used the British National Corpus (BNC henceforth), which is not as recent as COCA is, dating from the 1980s to 1993, and it is smaller, containing 100 million words of mixed genres. I used the spoken subcorpus, which unlike COCA, is made up of both formal speeches and informal private conversations. The spoken corpus is 10% of the total BNC, so it contains approximately 10 million words, which is a tenth of the spoken subcorpus of COCA.[28] The numbers are thus less trustworthy.

One verb, out of 30, resulted to be more common in the Present Perfect, “include”, which was found in a periphrastic construction in 60% of the cases. This verb, however, occurred only for a total of five times, counting both Present Perfect and Simple Past constructions. Furthermore, six more verbs were found in the Present Perfect in 40% or more of the hits, and two of these, “reach” and “lose”, were split 50/50 between the Present Perfect and the Simple Past. “Reach”, however, occurred only for a total of eight times. The other four verbs are: “eat”, “hear”, “check” and “receive”. More than two thirds

[26] The verbs which altogether only had a total number of less than 10 tokens were excluded from the table.

[27] The irregular conjugation of “learn” was not included when calculating the total percentage, since it was close to non-existent and would have skewed the result.

[28] http://www.natcorp.ox.ac.uk/corpus/index.xml

of the verbs, however, were expressed in the Simple Past 67% of the occurrences, or more. “Stroll” and “applaud” were not found either in the Present Perfect or in the Simple Past, and “swim” was found only once, in the Simple Past.

Table 2: *The three verbs most commonly expressed in British English Simple Past and the three verbs most commonly expressed in British English Present Perfect, out of the verbs in the sample.*[29]

	Tokens in Present Perfect	Tokens in Simple Past	Percentages
walk	8	127	6% vs. 94%
understand	6	53	10% vs. 90%
fall	5	60	11% vs. 89%
.....			
eat	25	31	45% vs. 55%
hear	416	469	47% vs. 53%
lose	132	133	50% vs. 50%

The total percentage of verbs expressed in the Present Perfect is approximately 26%, and hence 74% in the Simple Past. Comparing British and American English, one notices that the Present Perfect is slightly more common in British English, as expected. See 1.4.1 for details on this note.

4.3 Swedish

The Swedish “Språkbanken”, created by the University of Gothenburg, has a spoken subcorpus, “Talbanken”, but it is too small for this purposes, with only approximately 100 000 tokens. It is, furthermore, not exclusively composed of spoken language. I decided, for these reasons, to use the “social media” corpora, since the language used on social media tends to be closer to spoken language. The social media corpora is composed of blog texts, forum texts and messages from the micro-blog Twitter. I used the former two. The Twitter corpus was excluded since the amount of characters one can use is limited,

[29] The verbs which altogether only had a total number of less than 10 tokens were excluded from the table.

which has an effect on the formulations one uses, possibly also regarding the tense-aspect field.[30]

The “Bloggmix” (“blog mix”) is composed of 21 subcorpora, from 1998 to 2017, with the addition of blog posts from unknown date, and contains approximately 600 million tokens. The “Diskussionsforum” (“discussion forum”), is composed of 39 subcorpora which can be divided into two main groups: “Familjeliv” (“family life”) and “Flashback”, which is another Swedish forum. The Diskussionsforum contains almost 7,5 billion tokens.[31]

There is no grammatical expression of the perfective in Swedish, so the Swedish Present Perfect was compared to the Swedish Simple Past. The former has a periphrastic construction and the latter is bound. Unlike the English Simple Past, the Swedish one is different from the supine form used to construct Present Perfect. The supine does, however, partly coincide with past participle forms used in adjectival constructions. A first person singular pronoun was, therefore, included also in the Swedish corpus research.

In the Simple Past no adverbs intervene between the pronoun and the verb in main clauses. This is, however, possible in a subordinate clause. I, therefore, allowed for a word to intervene between the pronoun and the verb. This was done with all verbs except for one, *äta* (“eat”), which in the Simple Past has the same form as the preposition *åt* (“towards”). This homograph only causes problems if the search allows for a word to intervene between the pronoun and “åt”. In the Present Perfect form adverbs can occur between the auxiliary and the supine form, I hence allowed from zero to three words to intervene.

Four verbs resulted to be more frequent in the Present Perfect: *höra* (“hear”), *uppnå* (“achieve”; “reach”), *förstå* (“understand”) and *lära* (“learn”; “teach”). Out of these, the first one was expressed in the Present Perfect in 69% of the occurrences. One could imagine that this high frequency depends on the high relevancy of the moment of speech in many cases when using a past tense of “hear”, one tells something they have heard. *Bygga* (“build”), like *uppnå, förstå* and *lära*, was closer to being split between the two constructions.[32] 24 of the verbs were used in the Simple Past in 70% of the cases, or

---

[30] It should be noted that the new communication technologies, such as text messages, may have an effect on language. This is, however, not the focus of this paper and I hence decided not to include the corpus from Twitter.

[31] http://www.lrec-conf.org/proceedings/lrec2012/pdf/248_Paper.pdf

[32] *Uppnå, förstå* and *lära* were expressed in Present Perfect in 55-58% of the occurrences, and *bygga* was expressed in Simple Past in 58% of the occurrences.

more. *Veta* (“know”) and *knacka* (“knock”) were only used in 2% and 7%, respectively, of the Present Perfect constructions.

A note should be made regarding the verb *säga* (“say”). Its Simple Past form can be expressed in two ways: *sa* or sade. Both expressions were included in the research. *Sade* is the older form and is today mostly used in more formal texts, and, as a matter of fact, turned out to be less common in my research which included mostly unedited texts.

Table 3: *The three verbs most commonly expressed in Swedish Simple Past and the three verbs most commonly expressed in Swedish Present Perfect, out of the verbs in the sample.*

	Tokens in Present Perfect	**Tokens in Simple Past**	**Percentages**
veta (“know”)	3402	175 842	2% vs. 98%
knacka (“knock”)	41	567	7% vs. 93%
älska (“love”)	2747	29 098	9% vs. 91%
.....			
förstå (“understand”)	101 841	83 199	55% vs. 45%
uppnå (“achieve”; “reach”)	480	354	58% vs. 42%
höra (“hear”)	196 859	88 252	69% vs. 31%

The total percentage of the 30 verbs expressed in the Present Perfect was 25%, compared to 75% in the Simple Past.

As mentioned in 1.4.2, the auxiliary can be omitted in subordinate clauses, and the Swedish Simple Past form, unlike in English, does not correspond to the supine form. This difference is illustrated in the following examples, with the Simple Past form *arbetade* in (1) and the supine form *arbetat* in (2):

(1) *Jag tror att han arbetade där i tio år* (“I think that he worked there for ten years”)

(2) *Jag tror att han (har) arbetat där i tio år* (lit. “I think that he (has) worked there for ten years”)

For this reason a smaller manual corpus research was also executed. The purpose of this smaller corpus research was to see if a significant number of verbs in the Present Perfect had been excluded from my original corpus research. The results are presented in the following section.

4.3.1 Swedish, smaller corpus research

I used the Bloggmix subcorpus of 2017 which contains approximately 1,7 million tokens and searched for the same 30 verbs as in the bigger corpus research. However, only the 16 verbs which were found in at least ten constructions, counting both Simple Past and Present Perfect, will be discussed. The complete account of the results, including also the rarer verbs, can be found in the appendix. The 16 verbs here examined are: *prata* ("speak"), *förstå* ("understand"), *äta* ("eat"), *bo* ("live"), *öppna* ("open"), *höra* ("hear"), *kolla* ("check"; "look"), *veta* ("know"), *älska* ("love"), *gå* ("walk"; "function"), *lära* ("learn"; "teach"), *göra* ("do"; "make"), *säga* ("tell"), *få* ("receive"; "get"; "may"), *ta* ("take") and *glömma* ("forget"). I repeat a note regarding *säga*: its Simple Past can be expressed in two ways: *sa* or sade. Both expressions are included this time as well.

I looked at first person singular once again. Since the goal was to include results in which the auxiliary could be omitted I searched for the pronoun followed by the supine form, allowing, however, from zero to three words to intervene. The research was made manually, mainly so that I could exclude Past Perfect, which can also be expressed with the omission of the auxiliary. It was, furthermore, a way to exclude sentences in which the auxiliary preceded the subject, or would have preceded it, if it had not been omitted. This was necessary since the word order in the Simple Past search was *jag* ("I") followed by the verb, with the possibility of one word to intervene. The manual research also allowed me to exclude some other irrelevant results, such as constructions in which the subject was not first person singular, but, for example *Jag och du* (lit. "I and you"). The last kind of constructions, however, made up a very limited number of the total result.

In the main Swedish research, four verbs resulted to be more frequent in the Present Perfect. This smaller research showed the same result for three of these verbs. The fourth verb, *uppnå* ("achieve", "reach"), was excluded from the smaller research since only one occurrence was found of it. Furthermore, two more verbs, *äta* ("eat") and *glömma* ("forget"), were more frequent in the Present Perfect in the smaller research. *Göra* ("do"; "make") was close to being evenly split between the two constructions.

The remaining ten verbs were expressed in the Simple Past in 63% to 93% of the occurrences. This reflected the results of the main Swedish corpus research. However, with the exception of two verbs, the Present Perfect constructions in the smaller research had in general a higher token frequency compared to the Present Perfect constructions in the main research, as can be seen in table 4. This is the result of the smaller research being more inclusive, allowing also Present Perfect constructions without the auxiliary.

The total percentage of the Present Perfect constructions turned out to be slightly higher, compared to in the main Swedish research, with 32% of the verbs expressed in the Present Perfect, compared to 68% in the Simple Past. As mentioned in 1.4.2, the auxiliary can typically be omitted from Present Perfect constructions in more formal written texts. This smaller corpus research confirms that this tendency concerns unedited texts as well.

Table 4: *The token frequency of Swedish Present Perfect compared to that of Swedish Simple Past, including also Present Perfect constructions in which the auxiliary has been omitted. The results of the Swedish main research is also displayed.*

	Tokens in Present Perfect	**Tokens in Simple Past**	**Percentages**	**Percentages (Main research)**
veta (“know”)	2	26	7% vs. 93%	2% vs. 98%
ta (“take”)	32	107	23% vs. 77%	15% vs. 85%
säga (“tell”)	12	37	24% vs. 76%	21% vs. 79%
(sa)	-	35	-	-
(sade)	-	2	-	-
få (“receive”; “get”; “may”)	140	362	28% vs. 72%	18% vs. 82%
gå (“walk”; “function”)	33	73	31% vs. 69%	20% vs. 80%
prata (“speak”)	6	12	33% vs. 67%	26% vs. 74%
älska (“love”)	5	10	33% vs. 67%	10% vs. 90%
öppna (“open”)	3	6	33% vs. 67%	14% vs. 86%
kolla (“check”; ”look”)	9	16	36% vs. 64%	25% vs. 75%
bo (“live”)	7	12	37% vs. 63%	18% vs. 82%
göra (“do”; “make”)	108	130	45% vs. 55%	26% vs. 74%
förstå (“understand”)	13	11	54% vs. 46%	58% vs. 42%
höra (“hear”)	19	15	56% vs. 44%	71% vs. 29%
äta (“eat”)	26	19	58% vs. 42%	25% vs. 75%
glömma (“forget”)	11	8	58% vs. 42%	32% vs. 68%
lära (“learn”; “teach”)	20	6	77% vs. 23%	57% vs. 43%

4.4 Spanish

The Mark Davies’ Spanish corpus consists of 5 macro corpora. I focused on the “Web/ Dialects” corpus, which contains about 2 billion words from 21 Spanish-speaking

countries. I, however, limited it one more step, and looked only at Spanish spoken in Spain. This corpus contains texts from 25 500 web sites and approximately 460 million words. The data were collected between 2013 and 2014.[33]

Spanish verbs are conjugated for person and number, and, in analogy with the English and Swedish research, I searched for verbs in first person singular, however, without the pronoun. The unmarked position of Spanish adverbs is at the beginning or at the end of the sentence (Butt et al. 2019: 225), so for this reason no extra searches with adverbs positioned between the auxiliary and the past participle for the Present Perfect were made.

The perfective aspect, the Preterite, is expressed through a periphrastic construction, whilst the Spanish Perfect has a bound expression.

Table 5: *The three verbs most commonly expressed in Spanish Preterite and the three verbs most commonly expressed in Spanish Present Perfect, out of the verbs in the sample.*

	Tokens in Present Perfect	**Tokens in Preterite**	**Percentages**
saber (“know”)	1181	4332	21% vs. 79%
caer (“fall”)	455	1372	25% vs. 75%
caminar (“stroll”, “walk”)	50	151	25% vs. 75%
.....			
incluir (“include”)	339	218	61% vs. 39%
oír (“hear”)	3285	2073	61% vs. 39%
comprobar (“check”, “verify”)	1182	649	65% vs. 35%

As can be seen in table 5, the relative frequency difference between the two constructions is not as marked as it resulted to be in the other languages. As noted in 1.4.3, there is a tendency for the Spanish Present Perfect to be used also to indicate punctual past situations, when the time of the situation is of today, or possibly also yesterday. For example, *Me he despertado a las 7 de la mañana* (lit. “I have woken up at seven o’clock this morning”) can be uttered also in the afternoon. This expansion of the periphrastic construction could explain the less marked difference in frequency of the two constructions.

[33] https://www.corpusdelespanol.org/web-dial/

Five of the 30 verbs, *comprobar* (“check”; “verify”), *oír* (“hear”), *incluir* (“include”), *trabajar* (“work”) and *vivir* (“live”), were more common in the Present Perfect than in the Preterite, and of these *comprobar* was the most commonly used in Present Perfect constructions, which constituted 65% of the two constructions. Furthermore, eight verbs were rather evenly distributed between the two expressions.[34] These verbs are *hablar* (“speak”), *entender* (“understand”), *comer* (“eat”), *construir* (“build”), *aprender* (“learn”), *ganar* (“win”; “earn”), *perder* (“lose”) and *recibir* (“receive”). Twelve verbs were expressed in the Preterite in 65% of the cases or more, and are the following: *saber* (“to know”), *caminar* (“stroll”; “walk”), *caer* (“fall”), *golpear* (“to knock”; “hit”), *abrir* (“open”), *cortar* (“cut”), *aplaudir* (“applaud”), *nadar* (“swim”), *andar* (“go”; “walk”; “function”), *hacer* (“do”; “make”), *amar* (“love”) and *olvidar* (“forget”). The verb which was most commonly expressed in the Preterite was *saber*, in 79% of the cases. A general preference for Preterite constructions was thus found, but it was less marked than in the other three languages.

The total percentage of verbs expressed in the Present Perfect was 40%, compared to 60% in the Preterite.

4.5 Portuguese

The Mark Davies’ Portuguese corpus consists of 4 macro corpora, and, like with the Spanish research, I chose the “Web/Dialects” corpora, which contains approximately 1 billion words form Brazil, Portugal, Angola, Mozambique. I focused on European Portuguese corpus, which is a corpus with about 327 million tokens from 21 000 web sites. The data were collected between 2013 and 2014.[35]

Like Spanish verbs, Portuguese verbs are conjugated for person and number. I examined verbs in first person singular, without the pronoun, in the Present Perfect (*Pretérito Perfeito Composto*) and in the Preterite (*Pretérito Perfeito*). The former is the compound form, and the latter the bound form.

As noted in 1.4.4, the Portuguese compound Present Perfect has a limited present perfect meaning, denoting only past durative or iterative situations which are relevant for the moment of speech. This limited use is clearly reflected in its lack of frequency, as can

[34] By “evenly distributed” I mean verbs that were expressed in the Preterite between 50-55% of the occurrences.

[35] https://www.corpusdoportugues.org/web-dial/

be seen in table 6. It is the least common compound Present Perfect out of the four languages examined in the present paper. Since its frequency was so low, I also tried adding an adverb between the auxiliary and the past participle, even though the adverbs are usually found outside the verbal complex. This, however, barely had any effect on the final result, and these numbers were, therefore, not included.

Table 6: *The three verbs most commonly expressed in Portuguese Preterite and the three verbs most commonly expressed in Portuguese Present Perfect, out of the verbs in the sample.*

	Tokens in Present Perfect	Tokens in Preterite	Percentages
incluir (“include”)	0	17 230	0% vs. 100%
aplaudir (“applaud”)	0	49	0% vs. 100%
compreender (“understand”)	1	1144	0% vs. 100%
.....			
passear (“stroll”)	7	131	5% vs. 95%
verificar (“check”, “verify”)	115	1481	7% vs. 93%
caminhar (“walk”; “progress”)	134	264	34% vs. 66%

The relative percentage numbers are displayed in whole numbers, seven of the examined verbs resulted to occur through bound morphology in 100% of the cases, even though most of them did occasionally also occur in periphrastic constructions. These seven verbs are: *incluir* (“include”), *aplaudir* (“applaud”), *compreender* (“understand”), *abrir* (“open”), *cair* (“fall”), *chegar* (“arrive”, “reach”) and *amar* (“love”).

Caminhar (“walk”; “progress”) was the only verb which had relatively high frequency of Present Perfect constructions, probably since it is a verb which is semantically compatible with durative meaning. Next to c*aminhar* there were eight more verbs which had a slightly higher token frequency in the Present Perfect. They were, however, far from being frequent, each making up 4-7% of the occurrences. These verbs are: *verificar* (“check”, “verify”), *passear* (“stroll”), *trabalhar* (“work”), *receber* (“receive”), *reconhecer* (“to recognize”; “admit”), *ouvir* (“hear”), *aprender* (“learn”) and *construir* (“build”).

The total percentage of verbs expressed in the Present Perfect was only 3%, compared to 97% in the Preterite.

CHAPTER 5

Discussion

In the previous chapter the results of the corpus researches in English, Swedish, Spanish and Portuguese were presented. I examined the relative frequency of 30 verbs in each language's expression of present perfect compared to perfective or simple past, depending on the different tense-aspect systems of the languages. The aim with this corpus research was to test Martin Haspelmath's theory, according to which the more frequent form of a grammatical expression is also the shorter one. This follows from his form-frequency correspondence hypothesis, discussed in chapter 3. The hypothesis can briefly be summarized as follows: the less frequent of two constructions bearing a semantic opposition is expected to be overtly coded, whilst the more frequent one tends to be zero-coded or composed of fewer segments. The reader should be reminded that, according to Haspelmath, the frequent form is not more frequent because it is the shorter one, but the other way around. Predictability and efficiency are driving forces.

According to Haspelmath's idea then, I expected to find a higher frequency of verbs expressed in simple past or perfective forms, which are bound in the four above mentioned languages, compared to when expressed periphrastically, as present perfects. The five[36] corpus researches confirmed this expectation, although on different levels. The clearest results came from Portuguese: 97% of the examined verbs were found in bound constructions. In the American English research 83% of the verbs were expressed through bound morphology. In the British English and in the Swedish researches 74% and 75%, respectively, of the verbs were expressed through bound morphology. The Swedish smaller corpus research showed a similar tendency, even though less marked, with 68% of the verbs in the Simple Past. Finally, the less convincing data came from the Spanish research, in which only 59% of the verbs were expressed through bound morphology. Each result is discussed more in detail in the following sections.

[36] Or six, counting also the smaller Swedish corpus research.

The main question is, however, do these results truly show a correlation between frequency and form, so that frequency can be considered the ultimate cause? In a quest to answer this question, different aspects of the gathered results will now be investigated.

5.1 The English results

As stated above, the numbers from the English researches are coherent with Haspelmath's idea that the more frequent is also the shorter. In 1.4.1, the differences between American English and British English regarding Present Perfect and Simple Past were briefly discussed. In short, there is a general decline of the Present Perfect, and this decline is more marked in American English, although present in British English as well. The corpus research seems to confirm this picture, since the Present Perfect, in fact, turned out to be less common in American English than in British English, which is further confirmed by most corpus-based studies (Bao et al. 2018). Chenyao Bao et al. (2018), like Yao and Collins (2012), describe the development as a general decline of the Present Perfect, and highlight the fact that this is different from what is happening in other Indo-European languages, such as French and German.

There are, however, also some divergent results from another study. Marianne Hundt and Nicholas Smith (2009) report an opposite trend in spoken English during the second half of the 20th century, parallel with the general decline of the Present Perfect: the use of the Present Perfect together with past time adverbials is becoming more frequent. For example, "And Roberts has played for us last season" (Hundt and Smith 2009: 46). The authors explain this phenomenon through highlighting pragmatic functions of this use of the Present Perfect, and refer to Jim Miller (2004). According to Miller (2004), the addition of the past time adverbial is simply a way to allow the listener to better locate the referred moment, a function which would go back to Middle English.[37] This is in line with Hundt and Smith's general view of the development, according to which "the proportion of the two constructions is stable over time" (2009: 57).

Next to the American English (and British English) tendency to decrease the use of the Present Perfect, for example using the Simple Past also with adverbs indicating current relevance, such as "yet" and "already", there is hence another trend, in which the Present Perfect can be used together with past time adverbials, such as "yesterday". The latter is, however, described as "rare and usually 'locally' triggered" (Hundt and Smith

[37] See 5.2.2 for a possible parallel in Swedish.

2009: 57). Jim Walker (2011) discusses the increase of a similar phenomenon in British English, in which the Present Perfect is used with definite past reference, but without past time adverbials. It is the context that sets the time: the initial sentences of the utterances are in the Simple Past. Walker gives the following example from an interview with the manager of the Newcastle United football team: “Well, I thought that apart from the goal we were the better team today, I mean Shay’s made a fantastic save in the first half […], and we’ve had some really good opportunities […]” (Walker 2011: 73). Walker proposes this as a fifth category of present perfect, and next to resultative, continuative, existential and “hot news”, there would hence be a “narrative perfect”. In this reading, it is the relaxation of the importance of the current relevance-meaning that allows this function, and not the fact that this has been a possible pragmatic function of the Present Perfect since Middle English (Walker 2011). As discussed many times, according to Bybee and Dahl’s grammatical paths theory, the development of a perfect into a perfective past or past marker involves that the moment of speech becomes irrelevant. The narrative perfect, according to Walker, might develop out of narrating situations, in which “the distance between Now and Then becomes blurred” (Walker 2011: 75), as the speaker and listener envision the narrated events.

There seem thus, both in American English and in British English, to be tendencies of increase and decrease for both constructions. This might appear contradictory. However, the expansion of the Present Perfect can be seen as a semantic change, with the loss of the necessity of current relevance of the past event. The increase of the Present Perfect would then be in line with was is happening in French, Northern Italian and German. The expansion of the Simple Past, on the other hand, could, following Elsness (1997) be a change first and foremost in form, in which the auxiliary of the Present Perfect is omitted. According to Elsness (1997) the omission of the auxiliary together with the similarity between the past participle and the Simple Past form for most verbs enable this development. The deletion of the auxiliary is a common way for a periphrastic construction to become bound (Bybee and Dahl 1989). Elsness hypothesis, however, only works for verbs in which the past participle and the Simple Past forms coincide.

If Elsness (1997) is correct in that (part of) the increase of the English Simple Past actually is a formal reduction of the Present Perfect, in which the auxiliary is dropped, then we would expect to find expressions such as “I done it”, which we generally do

not.[38] This means that “finished“ in (1) is not a Present Perfect “in disguise”, in which the auxiliary was first frequently reduced and then finally omitted, but a Simple Past:

(1) - Have you read *Hamlet*?
- Yes, I just finished it.

A possible interpretation is that in (1), there is less importance given to the current relevance of having read a certain book, and more to the act of having finished it, hence viewing the action primarily as concluded and not as relevant. For this reason then, the Simple Past is preferred.

5.2 The Swedish results

Like the English results, also the Swedish research showed a strong preference for the Simple Past.

5.2.1 The omission of the auxiliary

The relative preference for the Simple Past, however, decreased partly with the smaller, manual search in which also Present Perfect constructions without an auxiliary were allowed. The omission of the auxiliary, as noted in 1.4.2, is a common and not recent phenomenon in more formal written Swedish. The data here analysed was, however, unedited blog texts and forum texts, and hence closer to spoken language. A corpus-based study using PAROLE corpus of Present-Day Swedish and the Swedish Spoken Language Corpus by Maia Andréasson et al. (2002) showed that the auxiliary of the Present Perfect construction is omitted in 83% of the newspaper material, 67% of the literary texts and 10-15% of the spoken Swedish.[39] This is further in accordance with a previous study by Sven-Göran Malmgren (1985).

The auxiliary omission goes back to Late Old Swedish, and, as mentioned, it is debated whether the High German influence was the only factor, or if the High German influence rather reinforced an already existing phenomenon. Its frequent omission in 17th century prose and diaries makes it plausible to think that it was also already part of the

[38] From the spoken subcorpus of COCA: 83 occurrences of “I done”, excluding “have I done”, compared to 48 787 occurrences of “I did”. From the Corpus of Historical American English: 1335 occurrences of “I done”, excluding “have I done” compared to 91 868 occurrences of “I did”.

[39] Both corpora are part of Språkbanken (The Swedish Language Bank), see 4.3 for more information.

spoken language (Larsson 2009). Swedish is the only Scandinavian language which allows complete deletion of the auxiliary in Perfect constructions. The auxiliary omission is, however, possible only in subordinate clauses,[40] but both for the Present Perfect and the Past Perfect. There are several theories regarding what enables the omission of the auxiliary. According to Malmgren (1985) it is first and foremost possible because of the existence of the main clause, since the subordinate tends to be in the same tense as the main clause. According to Marit Julien (2000) the omission is possible if there is a subject or subordinating conjunction. There are, however, examples which contradict both hypotheses, as illustrated by Ida Larsson (2009). According to Larsson (2009), it is the past tense participle which makes it possible to identify the auxiliary, and, therefore, to omit it. Andréasson et al. further stress the role of the general context. It is, for example, the context which enables the listener or reader to understand whether the omitted auxiliary is of present or past tense in the following example (Andréasson et al. 2002: 70):

(2)	*Vilken*	*snögubbe*	*du*	(*har/hade*)	*byggt!*
	what	snowman	you	have/had	built
	"What a snowman you have/had built!"				

Several factors seem to be at work.

It is possible to draw a parallel with Elsness's (1997) analysis of the increased use of the English Simple Past with adverbs of current relevance, in which the phenomenon is interpreted as an omission of the auxiliary in the Perfect construction. The main difference, however, is that the Swedish supine and the Simple Past form do not coincide. This difference in form is also what helps the listener to understand which of the two constructions is being used, even if the auxiliary is missing. Little, or no, information is lost, yet less energy is spent on the expression. That is, the predictability of the information contained in the message remains the same even though a shorter form is used. The omission could hence be the result of efficient coding, as predicted by Haspelmath, since the context allows you to. It could further be interpreted as a bound form.

[40] The auxiliary *ha* ("have") can also be omitted in main clauses if the main clause contains *kanske* ("maybe"). *Kanske* is composed of *kan* ("can", "may") and *ske* ("happen").

5.2.2 Present perfect with past time adverbial

As noted in 1.4.2, there are (rare) occurrences of Present Perfect constructions together with a past time adverbial, both in written and spoken Swedish. A common interpretation of these constructions is that, since they are not part of standard Swedish, they are the result of "afterthought" constructions (Dahl 1985). Dahl (1985), however, argues that not all of them can be analysed as such, and that the time adverbial in those cases has a particular pragmatic function, denoting "new information", highlighted by the accompanying full stress. A typical definition of a perfect is that the point of reference and the point of the event are separated (Reichenbach 1947), and, Dahl continues, this is possible "only if the E[41] is not already 'given' in the context" (1985: 138). A time adverbial which is classified as denoting new information is, per definition, not given in the context. It is possible to draw yet another parallel with English, this time with the phenomenon mentioned in 5.1 regarding present perfect constructions together with past time adverbials. As a matter of fact, Miller (2004), as well, analyses the corresponding English constructions as bearing a particular pragmatic function.

5.3 The Spanish results and the Standard Average European

As mentioned in 4.4, a possible explanation to the rather similar relative frequency of the Present Perfect and the Preterite constructions, is the semantic change which involves the Peninsular Spanish Present Perfect at the present. In certain varieties of Peninsular Spanish[42] (henceforth Spanish), and especially in younger age groups, the Present Perfect can also be used for hodiernal or even hesternal pasts (Howe and Schwenter 2003).[43] This means that the Present Perfect is expanding its possible functions, and hence its frequency.

As discussed in chapter 2, perfects which develop functions bearing perfective meaning is not an uncommon phenomenon cross-linguistically, on the contrary, Bybee and Dahl (1989) argue that it is one of the most frequent semantic paths involving perfects. If a perfect changes into a perfective past, or a general past, the relevance of

[41] Point of the event.

[42] For example the Spanish spoken in Madrid, Alicante and Seville.

[43] A similar, yet not the same, tendency can be seen in certain varieties of Latin-American Spanish such as the Spanish spoken in Peru, Bolivia, Paraguay, northwestern Argentina and part of Central America (Howe and Schwenter 2003; Ariolfo 2019).

the moment of speech disappears. In *Ayer he ido al cine* (lit. “Yesterday I have gone to the cinema”) there is no link to the moment of speech, and hence only the Preterite “should” be used in accordance with standard Spanish. Nevertheless, utterances of this type are becoming more frequent (Howe and Schwenter 2003).

A possible outcome of this development is that the Present Perfect will continue to take over functions of the Preterite, the bound form. Present perfects acquiring perfective meaning have already occurred in other Romance varieties, as in French, or is in the process of happening, as in Northern Italian. In Southern German dialects a similar change is taking place, this time involving the present perfect and the past. The reason the Southern German Present Perfect has not taken on a specifically perfective meaning, but a general past meaning, is that there is no grammatical expression of imperfective in German, unlike in French and in Italian (Bybee and Dahl 1989).

French, Northern Italian and (Southern) German also share the fact that they are all part of the core of the linguistic area called Standard Average European (henceforth SAE), coined by Benjamin Lee Whorf (Haspelmath 2001). A linguistic area involves geographically close languages, which due to mutual relations have developed common features over time, independently of their genetic relations. The exact boundaries of a linguistic area are, of course, difficult to trace, but Portuguese, Spanish, English and Swedish are generally not considered to be part of the innermost core of SAE. This also corresponds to the geographical distances. The closer to the “core” of the area a language is, the more typical features it is expected to have. One of these features is the “have”-perfect, the transitive perfects formed by “have” plus a passive participle, which all Romance and Germanic languages share (Haspelmath 2001). The languages differ, however, as we have seen, when it comes to which meanings the “have”-perfect has: in the core of the area it has a perfective past or general past meaning, and in the less central areas it has a more typical present perfect meaning.

Anna Giacalone Ramat argues that next to “universal tendencies and internally motivated changes” (2008: 1), language contact can also trigger grammaticalization processes. She further argues that the European “have”-perfects have developed out of language contact, and more specifically originally from Latin. What has happened in French, Northern Italian and in Southern German is simply a “further step”, thus in accordance with Bybee and Dahl (1989).[44] The reason why the meaning of present

[44] Giacalone Ramat (2008) considers Standard Italian and Standard German as intermediate stages.

perfects of English, Swedish and Spanish still denote present relevance is, according to Giacalone Ramat, that they are geographically on the outskirts of the area and hence more conservative. However, as mentioned above, the Spanish Present Perfect does not only denote present relevance anymore. This could, as well, be the effect of language contact. On the other hand, the Portuguese Present Perfect is close to non-existent.

This is a good moment to remind the reader that Haspelmath's functional-adaptive explanation is not intended for studies of individual languages, but for universal tendencies. Particular cultural and historical contexts may have strong effects on a language (Haspelmath 2020), and language contact is, of course, an example of this.

5.4 The Portuguese results

The clearest preference for the bound form over the periphrastic expression came from the Portuguese research, in which only 3% of the verbs were found in a periphrastic expression, that is, in the Present Perfect. This is partly explained by the limited semantics of the Portuguese Present Perfect, which is only used to refer to past durative or iterative situations with relevance for present time, and, therefore, not to single situations.

As mentioned in 1.4.4, the Preterite, the bound form, is used not only to express perfective meaning, but can further express certain perfect meanings, such as resultative and experiential. The use of *já* ("already", "yet", "ever"), typical in bound present perfect constructions (Alego 1976), can be a way to emphasize a present perfect meaning of the bound form, since present perfect and perfective meanings can both be expressed through a bound construction. This can be illustrated with the following examples:

(3) *Voce esteve em Salvador?* (lit. "You were in Salvador?")

(4) *Voce já esteve em Salvador?* (lit. "You were already in Salvador?")

Both sentences are in the Preterite, and the only formal difference is the use of *já* in (4), which hence could be interpreted as a periphrastic construction. (3) would typically be translated with "Were you in Salvador?", denoting a specific moment in the past, that is, perfective. The other example, (4), is most frequently translated with a Present Perfect in English, "Have you ever (/already) been to Salvador?". The use of a word with a meaning similar to "already" to convey perfect meaning in languages lacking a grammatical perfect

can be found in different languages of the world, such as Russian (Dahl 1985). This adverb, together with a main verb, is further mentioned by Bybee and Dahl (1989) as a typical diachronic source for perfects. The question is then, whether this construction is (becoming) another example of a periphrastic perfect in Portuguese.

5.5 The concept of Simultaneity

There are, as we have seen so far, temporal and aspectual differences regarding the perfect in the four investigated languages. For example, in certain varieties of Spanish, unlike in the other three languages,[45] the Present Perfect can be used also with adverbials referring to prehodiernal time, such as *ayer* ("yesterday"). Furthermore, the Portuguese compound Present Perfect cannot refer to single situations in the past, but only to iterative or durative situations.

Susana Azpiazu (2018), in an attempt to explain these differences, proposes an analysis of the perfect through the concept of "simultaneity", that is, "the temporal coincidence between the event and the speech act" (2018: 117). This time span, involving the moment of reference (and utterance) and the event in the past, would be conceived differently in different languages, and thus resulting in language-particular variations of perfect. Azpiazu's idea of simultaneity is based on previous works by Guillermo Rojo (1974) and Rojo and Alexandre Veiga (1999), and, furthermore, on the concept of *presente ampliado* by Emilio Alarcos (1947) and "Extended Now" by Robert W. McCoard (1978), and finally on the concept of "Perfect Time Span" by Arnim Von Stechow (1999), Sabine Iatrodou et al. (2003), Roumyana Pancheva and Arnim Von Stechow (2004), Teresa M. García Xiqués (2015). These concepts will now be presented, following Azpiazu 2018.

Alarcos's (1947) concept of the time span of the Spanish Present Perfect, *presente ampliado* ("Increased Present", henceforth IP), views the present as an abstract period of time, which extends into the past, hence "increased" in Present Perfect constructions. Azpiazu points out that the IP does not have precise temporal boundaries and, most importantly, that the left boundary is set according to the speaker's subjective view. It is thus the speaker who decides.

---

[45] With the exception of what Walker (2011) calls the "narrative perfect" in English, the observations of Hundt and Smith (2009) regarding the increased use of Present Perfect together with past time adverbials and the occasional use of the Swedish Present Perfect together with past time adverbials. The English constructions will be discussed in 5.6.

McCoard's "Extended Now" (henceforth XN) is similar to IP in the importance given to the subjective temporal view. XN can, however, only refer to unspecific moments, which is why the English Present Perfect cannot be combined with specific past time adverbials, such as "yesterday". The same goes for the Portuguese Present Perfect.

The revision of the XN theory by Von Stechow (1999), Iatridou et al. (2003), Pancheva & Von Stechow (2004) and Xiqués (2015) resulted in the Perfect Time Span (henceforth PTS) theory, in which the "Perfect implies a time interval whose left boundary is the speech act moment (which is also the reference time)" (Azpiazu 2018: 124). The concept of "interval" is central in this theory, and replaces "point", used in Reichenbach's theory (1947). Another alteration allows the existence of different cross-linguistic properties related to the time span, which were ignored in the XN theory. An example of this is the existence, or not, of a constraint regarding specific moments, which differentiates the English Present Perfect from the Spanish Present Perfect. As a result of this revision, a "weak PTS theory" is laid out by Pancheva and von Stechow (2004), in which two different time intervals are considered, compared to the original theory in which there is only one interval. In the weak PTS theory, with two time intervals, the reference time does not necessarily completely overlap with the PTS and the PTS can even precede the reference time, as can be seen with German and French perfects in which the present perfect can be used together with the past time adverbial such as "yesterday" (Azpiazu 2018: 124)[46]:

(5) German

Gestern	*bist*	*du*	*nicht*	*gekommen*
Yesterday	BE.2.SG	YOU.SG	NEG	COME.PTCP

"You didn't come yesterday"

(6) French

Tu	*n'est*	*pas*	*venu*	*hier*
You	NEG-BE.3.SG	NEG	COME.PTCP	yesterday

"You didn't come yesterday"

A consequence of this, according to Azpiazu (2018), is that, in examples like (5) and (6), there is no real competition between the simple and the compound past constructions.

[46] I further add an example of this in Northern Italian: *Ieri non sei venuto* (lit. "Yesterday you are not come").

The use of the compound form is here in the function of a “modal mark” (Azpiazu 2018: 131), and not simply a direct substitute of the simple form.

In comparison, the Spanish varieties which allow Present Perfect together with a past time adverbial are seen as “a growing trend in this direction” (Azpiazu 2018). The Spanish PTS is, however, not considered to be different from the reference time, but seen simply as “broader”. The English and the Portuguese present perfects,[47] like the Spanish, consider the PTS and the reference time as a single time interval, with the difference that the PTS is not broadly conceived, and their present perfects thus do not allow events located at specific moments, and hence neither definite past time adverbials. What emerges is a picture according to which it is the possible extension of the PTS which varies among the four languages, but that they all share the view that the beginning of the time interval is the reference (and utterance) time, which extends towards an indefinite previous moment.

In Azpiazu’s hypothesis, PTS is replaced with the concept of “simultaneity”, to highlight the different cross-linguistic semantic features of present perfect, without having to refer to temporal boundaries. The advantage of this, Azpiazu argues, is that, since the speaker “cannot determine the beginning or the end of their present world” (2018: 120), the lack of real-time boundaries in the concept of simultaneity is conform with this. The English and Portuguese (and Swedish) present perfects focus on the persistence of a past event to present time, the simultaneity hence begins in the past. The English (and Swedish) Present Perfect, furthermore, allows single past events, as long as they are not time specific. The simultaneity in the Spanish Present Perfect, on the other hand, begins in the present tense and extends into the past, highlighting its subjectivity. The Portuguese and the English present perfects are described as “special kinds of present tense” (Azpiazu 2018: 129), and the Spanish as a combination of two tenses. The Spanish is further depicted as being more developed and moving towards a perfective meaning, and the Portuguese as a possible previous stage of the Spanish one, resembling Classical Spanish of the 16th and 17th century (Azpiazu 2018). This is in accordance with Bybee and Dahl’s view of grammatical paths, in which the semantic development of a perfect into a perfective is, as mentioned, the gradual loss of present relevance (Bybee and Dahl 1989).

[47] And, I add, the Swedish Present Perfect.

5.6 Multiple forces

On the one hand, there is an increase of the use of periphrastic pasts, as can be seen in French, Italian, German, but also in Spanish, and further possibly also in English. On the other hand, the bound forms are are clearly predominant in Portuguese, English and Swedish. However, a note should be made regarding the Portuguese present perfect construction with the Preterite, which together with *já* can be interpreted as a sort of periphrastic construction. If this is a correct interpretation, then the result of the large frequency difference regarding periphrastically expressed past and inflectional past in Portuguese must be reconsidered. The occurrences of the Swedish Present Perfect constructions without an auxiliary in subordinates, on the other hand, can technically no longer be counted as periphrastic constructions.

If the tendencies observed by Hundt and Smith (2009) regarding the use of the English Present Perfect together with past time adverbials, and the "narrative perfect" by Walker (2011) are correct, then this could be a sign that a further step along the grammaticalization path has been taken, as the importance of the current relevance of the past event is becoming weaker. However, since there is also a more general increase of the bound form, this speaks for a different development. What we are dealing with hence appears to be the result of several forces at work. On the one hand, there is the grammaticalization process involving the English Present Perfect, which is moving along one of the cross-linguistic paths, according to which a perfect has a tendency to develop into a perfective or a past as the importance of the current relevance diminishes. On the other hand, there are instances of efficient coding, according to which only the energy necessary for a successful communication is spent. In a sentence such as "I just ate", "just" conveys the meaning of current relevance, making "I have just eaten" superfluous.

With Azpiazu's concept of simultaneity, the differences of these European present perfects result being different stages along the same path. This is hence much like what Bybee and Dahl (1989) argue in general regarding the development of grammatical morphemes in different languages. The order along this path, among the discussed languages, seems to be the following: Portuguese, English and Swedish, followed by Spanish, and then Northern Italian and Southern German, and finally French. Comparing these seven languages, there does not seem to be one single general trend in act, but more than one. However, there seem to be recurring forces: cross-linguistic grammaticalization paths, features shared through language contact, preference for efficient coding.

With time an expression, such as a perfect construction, can lose its explicitness and become more general. As a result, new constructions are necessary. This is similar to what, according to Giacalone Ramat, has happened with the "have"-perfects of the Romance and Germanic languages, which she describes as "renewals of categories that were already present in the languages involved but had been overused or had undergone formal reduction" (2008: 36). However, the Romance and Germanic "have"-perfects were introduced to different tense-aspects systems. As mentioned, the Perfect in Latin could express both simple past and present perfect. In the Germanic languages, on the other hand, there was no morphological expression for perfect before the "have"-perfect. The new constructions were thus inserted in different systems, which could partly explain the different spread of the present perfects in the modern Romance and Germanic languages. The South German dialects, being Germanic, however, do not fit into this picture. This could be an example of effects of language contact, being situated in the very "core" of the linguistic area SAE. If this is so, it could explain why there seems to be a general tendency for the expansion of the periphrastic present perfect among Romance languages. The exception seems to be Portuguese, in which the use of the Present Perfect is highly constrained by its limited semantics.

Perfects based on a possessive construction, like the "have"-perfects, appear to be a typically European construction, judging by its rarity on a cross-linguistic level (Dahl 1990: 7). In general, the use of a transitive verb to express possession is cross-linguistically a "relatively infrequent construction" (Dahl 1990: 7). However, the question whether it is only of Latin origin and thus the result of language contact, as Giacalone Ramat (2008) argues, is not as clear. Accordingly also the Germanic languages would have adopted the Late Latin/Romance resultative construction with "have".[48] Considering the fact that Latin was the official language and of great cultural influence during the Middle Ages, that the oldest examples of this construction are found in Latin documents (Giacalone Ramat 2008), and further that periphrastic constructions are more easily borrowed than bound ones (Dahl 2004), the hypothesis is plausible. According to Haspelmath (2001), the spread of the construction is probably of older origins, dating back to the time of the great migrations, between late antiquity and early Middle Ages. In that case, Latin rather reinforced already existing structures.

[48] Even though the Latin *habere* ("have") appears to be very similar to its Germanic equivalent, they are of different Indo-European origins, as noted by Giacalone Ramat (2008).

Conclusion

I have investigated whether perfects cross-linguistically tend to be longer because they are younger or because they are less frequently used. The biggest challenge has been to try to isolate the frequency effect, which, according to Martin Haspelmath (2020), is the ultimate cause regarding how grammatical material is coded, leading to shorter forms for more frequently used elements.

The shorter forms expressed through bound morphology are, according to Joan L. Bybee and Östen Dahl (1989), not only the older forms, but also the more general ones. For the two authors, semantic generalization and formal reduction go hand in hand. When a linguistic item has a more general meaning, it can be used in a wider range of contexts, and will thus increase in frequency. From this follows that the shorter, older, forms tend to be more frequently used. Based on the languages here examined, it is difficult to deny the existence of cross-linguistically common paths which involve the development of grammatical constructions, as argued by Bybee and Dahl. These were, however, not the only tendencies which I have observed. The Swedish phenomenon, for example, in which the auxiliary of the present perfect constructions can be omitted in subordinate clauses, can be interpreted as an example of efficient coding, since the listener or the reader can rely on other cues for the correct interpretation of the message. It could, however, further be interpreted as the next step along the grammaticalization path, in which the auxiliary is often dropped. Several forces are at play when it comes to what determines how a language is shaped, such as the strive for efficient coding, the role of predictability, grammaticalization paths, language contact and the already existing language-specific structures and features. Can frequency, that is, shortening by efficient coding, be considered the ultimate cause, the driving force?

The corpus research did not reach a homogenous result. I found that the bound construction, a perfective past or a general past, tends to have a higher relative frequency than the periphrastic construction, a present perfect. However, the Spanish Present Perfect is expanding its functions in some varieties, and a similar tendency appears to partially hold for the English Present Perfect as well. Glancing at other Indo-European languages, one can compare this to the extensive use of periphrastic constructions in

French, Northern Italian and Southern German. However, the periphrastic constructions in French and Northern Italian do not specifically denote present perfect, but have replaced, or are replacing, the older bound perfective. This fact does not fit very well with Haspelmath's idea regarding a preference for the shorter form. Haspelmath's form-frequency correlation does, however, not refer to specific languages, but regard general human tendencies. The French and Italian periphrastic perfectives, furthermore, go against Bybee and Dahl's theory that perfectives, like general pasts, are typically inflectional being older. That being said, these constructions are relatively young and could become inflectional over time. Only time will tell. The same goes for the Portuguese Present Perfect, with its limited semantics: it could either broaden its semantics, following Spanish, or it could continue to rely on the *já* construction together with the Preterite. Periphrastic constructions are more easily adopted, but also more easily lost (Dahl 2014). As stressed by Bybee (2006), a merely synchronic point of view does not give an adequate picture of how grammatical structures are established and interconnected.

The role of frequency is undeniable when it comes to grammatical coding, but the language sample I used was limited and it is clear that further research, including a much wider and varied sample of languages, is necessary to be able to draw more decisive conclusions.

Appendix

Table 1. The token number and percentage for each verb in first person singular in Present Perfect and Simple Past in American English (COCA: 127 million tokens).

	Tokens in Present Perfect	**Tokens in Simple Past**	**Percentages**
admit	28	27	51% vs. 49%
applaud	5	8	38% vs. 62%
build	40	163	20% vs. 80%
check	23	341	6% vs. 94%
cut	46	298	14% vs. 86%
do	3159	52178	6% vs. 94%
eat	53	221	19% vs. 81%
fall	46	676	6% vs. 94%
forget	144	628	19% vs. 81%
hear	4162	7048	37% vs. 63%
include	7	23	23% vs. 77%
knock	2	75	3% vs. 97%
know	1225	10333	11% vs. 89%
learn (total)	1019	2298	31% vs. 69%
learn (regular)	1019	2293	31% vs. 69%
learn (irregular)	0	5	0% vs. 100%
live	483	1026	32% vs. 68%
love	258	2935	8% vs. 92%
lose	463	1678	22% vs. 78%
open	10	455	2% vs. 98%
reach	62	280	18% vs. 82%
receive	182	525	26% vs. 74%
sing	27	200	12% vs. 88%
speak	577	2673	18% vs. 82%
stroll	0	7	0% vs. 100%
swim	3	28	10% vs. 90%
take	449	3579	11% vs. 89%
tell	490	5934	8% vs. 92%
understand	58	680	8% vs. 92%
walk	56	1464	4% vs. 96%
win	90	546	14% vs. 86%
work	842	2063	29% vs. 71 %
TOTAL COUNT	-	-	17% vs. 83%

Table 2. The token number and percentage for each verb in first person singular in Present Perfect and Simple Past in British English (BNC: 10 million tokens).

	Tokens in Present Perfect	**Tokens in Simple Past**	**Percentages**
admit	2	4	33% vs. 67%
applaud	0	0	-
build	2	10	17% vs. 83%
check	17	24	41% vs. 59%
cut	19	69	22% vs. 78%
do	782	6114	11% vs. 89%
eat	25	31	45% vs. 55%
fall	6	50	11% vs. 89%
forget	145	261	36% vs. 64%
hear	416	469	47% vs. 53%
include	3	2	60% vs. 40%
knock	5	13	28% vs. 72%
know	102	691	13% vs. 87%
learn (total)	29	73	28% vs. 72%
learn (regular)	5	25	17% vs. 83%
learn (irregular)	24	48	33% vs. 67%
live	28	64	30% vs. 70%
love	9	38	19% vs. 81%
lose	132	133	50% vs. 50%
open	8	65	11% vs. 89%
reach	4	4	50% vs. 50%
receive	27	35	44% vs. 56%
sing	1	6	14% vs. 86%
speak	69	145	32% vs. 68%
stroll	0	0	-
swim	0	1	0% vs. 100%
take	68	505	12% vs. 88%
tell	129	859	13% vs. 87%
understand	6	53	10% vs. 90%
walk	8	127	6% vs. 94%
win	31	65	36% vs. 64%
work	72	193	27% vs. 73%
TOTAL COUNT	-	-	26% vs. 74%

Table 3. The token number and percentage for each verb in first person singular in Present Perfect and Simple Past in Swedish (Språkbanken: 8,1 billion tokens).

	Tokens in Present Perfect	**Tokens in Simple Past**	**Percentages**
erkänna (“admit”)	623	2098	23% vs. 77%
applådera (“applaud“)	10	47	18% vs. 82%
bygga (“build”)	2445	3423	42% vs. 58%
kolla (“check”; “look”)	26139	83453	24% vs. 76%
skära (“cut”)	649	2904	18% vs. 82%
göra (“do”; “make”)	147307	478272	24% vs. 76%
äta (“eat”)	40175	122817	25% vs. 75%
falla (“fall”)	1393	11166	11% vs. 89%
glömma (“forget”)	22330	51515	30% vs. 70%
höra (“hear”)	196859	88252	69% vs. 31%
inkludera (“include”)	65	256	20% vs. 80%
knacka (“knock”)	41	567	7% vs. 93%
veta (“know”)	3402	175842	2% vs. 98%
lära (“learn”; “teach”)	30221	25152	55% vs. 45%
bo (“live”)	13683	64560	17% vs. 83%
älska (“love”)	2747	29098	9% vs. 91%
förlora (”lose”)	4946	15689	24% vs. 76%
öppna (“open”)	3019	19910	13% vs. 87%
uppnå (“achieve”; “reach”)	480	354	58% vs. 42%
få (“receive”; “may”)	304100	1545319	18% vs. 82%
sjunga (“sing”)	513	3210	14% vs. 86%
prata (“speak”)	27809	84441	25% vs. 75%
promenera (“stroll”)	715	2159	25% vs. 75%
simma (“swim”)	291	1208	19% vs. 81%
ta (“take”)	304100	1545319	18% vs. 82%
säga (“tell”)	86295	356696	19% vs. 81%
(*sa*)	-	336727	-
(*sade*)	-	20969	-
förstå (“understand”)	101841	83199	55% vs. 45%
gå (“walk”; “function”)	96833	413679	19% vs. 81%
vinna (“win”)	2831	19660	13% vs. 87%
arbeta (“work”)	4941	8869	36% vs. 64%
TOTAL COUNT	-	-	25% vs. 75%

Table 4. The token number and percentage for each verb in first person singular in Present Perfect and Simple Past in Swedish (Språkbanken: 1,7 million tokens).

	Tokens in Present Perfect	**Tokens in Simple Past**	**Percentages**
erkänna (“admit”)	0	1	0% vs. 100%
applådera (“applaud“)	0	0	-
bygga (“build”)	2	1	67% vs. 33%
kolla (“check”; “look”)	9	16	36% vs. 64%
skära (“cut”)	0	1	0% vs. 100%
göra (“do”; “make”)	108	130	45% vs. 55%
äta (“eat”)	26	19	58% vs. 42%
falla (“fall”)	2	7	22% vs. 78%
glömma (“forget”)	11	8	58% vs. 42%
höra (“hear”)	19	15	56% vs. 44%
inkludera (“include”)	0	0	-
knacka (“knock”)	0	0	-
veta (“know”)	2	26	7% vs. 93%
lära (“learn”; “teach”)	20	6	77% vs. 23%
bo (“live”)	7	12	37% vs. 63%
älska (“love”)	5	10	33% vs. 67%
förlora (”lose”)	1	2	33% vs. 67%
öppna (“open”)	3	6	33% vs. 67%
uppnå (“achieve”; “reach”)	0	1	0% vs. 100%
få (“receive”; “may”)	140	362	28% vs. 72%
sjunga (“sing”)	0	1	0% vs. 100%
prata (“speak”)	6	12	33% vs. 67%
promenera (“stroll”)	1	1	50% vs. 50%
simma (“swim”)	0	1	0% vs. 100%
ta (“take”)	32	107	23% vs. 77%
säga (“tell”)	12	37	24% vs. 76%
(*sa*)	-	35	-
(*sade*)	-	2	-
förstå (“understand”)	13	11	54% vs. 46%
gå (“walk”; “function”)	33	73	31% vs. 69%
vinna (“win”)	0	7	0% vs. 100%
arbeta (“work”)	2	1	67% vs. 33%
TOTAL COUNT	-	-	32% vs. 68%

Table 5. The token number and percentage for each verb in first person singular in Present Perfect and Preterite in Peninsular Spanish (Web/Dialects: 460 million tokens).

	Tokens in Present Perfect	**Tokens in Preterite**	**Percentages**
reconocer (“recognise”; “admit”)	194	339	36% vs. 64%
aplaudir (“applaud”)	20	50	29% vs. 71%
construir (“build”)	82	91	47% vs. 53%
comprobar (“check”, “verify”)	1182	649	65% vs. 35%
cortar (“cut”)	133	351	27% vs. 73%
hacer (“do”; “make”)	15789	29860	35% vs. 65%
comer (“eat”)	691	796	46% vs. 54%
caer (“fall”)	455	1372	25% vs. 75%
olvidar (“forget”)	739	1372	35% vs. 65%
oír (“hear”)	3285	2073	61% vs. 39%
incluir (“include”)	339	218	61% vs. 39%
golpear (“knock”)	24	67	26% vs. 74%
saber (“know”)	1181	4332	21% vs. 79%
aprender (“learn”)	4804	4854	50% vs. 50%
vivir (“live”)	3469	2465	58% vs. 42%
amar (“love”)	304	565	35% vs. 65%
perder (“lose”)	2690	2998	47% vs. 53%
abrir (“open”)	550	1485	27% vs. 73%
llegar (“reach”, arrive”)	5028	7553	40% vs. 60%
recibir (“receive”)	2648	3266	45% vs. 55%
cantar (“sing”)	105	171	38% vs. 62%
hablar (“speak”)	3332	3608	48% vs. 52%
caminar (“stroll”, “walk”)	50	151	25% vs. 75%
nadar (“swim”)	16	36	31% vs. 69%
tomar (”take”; “drink”)	2039	3200	39% vs. 61%
decir (“tell”)	17886	30832	37% vs. 63%
entender (“understand”)	2255	2742	45% vs. 55%
andar (“walk”; “go”; “function”)	149	287	34% vs. 66%
ganar (“win”; “earn")	841	911	48% vs. 52%
trabajar (“work”)	2352	1585	60% vs. 40%
TOTAL COUNT	-	-	41% vs. 59%

Table 6. The token number and percentage for each verb in first person singular in Present Perfect and Preterite in European Portuguese (Web/Dialects: 327 million tokens).

	Tokens in Present Perfect	**Tokens in Preterite**	**Percentages**
reconhecer (“recognise”; “admit”)	25	560	4% vs. 96%
aplaudir (“applaud”)	-	49	0% vs. 100%
construir (“build”)	13	335	4% vs. 96%
verificar (“check”, “verify”)	115	1481	7% vs. 93%
cortar (“cut”)	7	983	1% vs. 99%
fazer (“do”; “make”)	1589	47277	3% vs. 97%
comer (“eat”)	74	2969	2% vs. 98%
cair (“fall”)	2	1012	0% vs. 100%
esquecer (“forget”)	21	2725	1% vs. 99%
ouvir (“hear”)	646	15543	4% vs. 96%
incluir (“include”)	-	17230	0% vs. 100%
bater (“knock”)	14	828	2% vs. 98%
conhecer (“know”; “meet”)	110	8810	1% vs. 99%
aprender (“learn”)	422	10810	4% vs. 96%
viver (“live”)	184	6121	3% vs. 97%
amar (“love”)	7	2815	0% vs. 100%
perder (“lose”)	63	6885	1% vs. 99%
abrir (“open”)	4	2252	0% vs. 100%
chegar (“reach”, arrive”)	28	14792	0% vs. 100%
receber (“receive”)	442	9542	4% vs. 96%
cantar (“sing”)	3	383	1% vs. 99%
falar (“speak”)	232	8774	3% vs. 97%
passear (“stroll”, “walk”)	7	131	5% vs. 95%
nadar (“swim”)	1	64	2% vs. 98%
tomar (”take”; “drink”)	67	4971	1% vs. 99%
dizer (“tell”)	882	131530	1% vs. 99%
compreender (“understand”)	1	1144	0% vs. 100%
caminhar (“walk”; “progress”)	134	264	34% vs. 66%
ganhar (“win”; “earn")	36	3232	1% vs. 99%
trabalhar (“work”)	198	3712	5% vs. 95%
TOTAL COUNT	-	-	3% vs. 97%

Abbreviations

1	first person
2	second person
3	third person
ACC	accusative
ADV	adverb
FUT	future
GEN	genitive
NEG	negation, negative
NOM	nominative
PAST	past
PFV	perfective
POSS	possessive
PRS	present
PRT	particle
PTCP	participle
SG	singular

References

Alacros, E. 1947. “Perfecto simple y compuesto en español”. In *Revista de Filología Española* XXXI: 108-139.

Alego, J. E. 1976. “The Portuguese present perfect”. In *Luso-Brazilian Review* 13 (2): 194-208. University of Wisconsin Press.

Andréasson, M., S. Karlsson, E. Magnusson and S. Tingsell 2002. “Har/hade-bortfall i svenskan: Hur finit är ett naket supinum?”. In B. Melander, U. Melander Marttala, C. Nyström, M. Thelander och C. Östman (eds.), *Svenskans beskrivning 26. Förhandlingar vid Tjugosjätte sammankomsten för svenskans beskrivning. Uppsala den 25-26 oktober 2002*: 67-74. Uppsala: Hallgren och Fallgren.

Ariolfo, R. 2019. “The Use of the Present Perfect (*pretérito perfecto compuesto*) with Aoristic Value in the Speech of Latin American Students”. In *Languages* 4 (32).

Azpiazu, S. 2018. “Simultaneity and ‘increased present’ in the European Spanish perfect”. In *Catalan Journal of Linguistics* 17: 117-134.

Bao, C., X. Zhang, Y. Qu and Z. Feng 2018. “American English Perfect Construction Across Registers”. In *Journal of Quantitative Linguistics* 25 (4): 314-341. Routledge.

Boyland, J. T. 1996. *Morphosyntactic change in progress: A psycholinguistic approach*. Berkeley: University of California, Berkeley dissertation.

Butt, J., C. Benjamin and A. M. Rodríguez 2019. *A New Reference Grammar of Modern Spanish*. 6th ed. London and New York: Routledge.

Bybee, J. L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam and Philadelphia: John Benjamins.

Bybee, J. L. 1988. “The diachronic dimension in explanation”. In John A. Hawkins (ed.), *Explaining language universals*: 350-379. Oxford: Blackwell.

Bybee, J. L. 1995. “Spanish tense and aspect from a typological perspective”. In P. Hashemipour, R. Maldonado and M. van Naerssen (eds.), *Studies in language learning and Spanish linguistics in honor of Tracy Terrell*: 442-457. San Francisco: McGraw Hill.

Bybee, J. L. 2006. “Language change and universals”. In R. Mairal and J. Gil (eds.), *Linguistic Universals*: 179-194. Cambridge: Cambridge University Press.

Bybee, J. L. and Ö. Dahl 1989. “The creation of tense and aspect systems in the languages of the world”. In *Studies in Language* 13 (1): 51-103.

Bybee, J. L. and W. Pagliuca 1985. “Cross-linguistic comparison and evolution of grammatical meaning”. In J. Fisiak (ed.), *Historical Semantics, Historical Word Formation*: 59-83. The Hague: Mouton.

Bybee, J. L., R. Perkins and W. Pagliuca 1994. *The evolution of grammar: tense, aspect and modality in the languages of the world.* Chicago: University of Chicago Press.

Comrie, B. 1976. *Aspect*. Cambridge: Cambridge University Press.

Comrie, B. 1985. *Tense*. Cambridge: Cambridge University Press.

Cristofaro, S. 2017. “Implicational universals and dependencies”. In N. J. Enfield (ed.), *Dependencies in Language: On the Causal Ontology of Linguistic* Systems: 9-22. Berlin: Language Science Press.

Cristofaro, S. 2019. “Taking diachronic evidence seriously: Result-oriented vs. source-oriented explanations of typological universals”. In K. Schmidtke-Bode, N. Levshina, S. M. Michaelis and I. A. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*: 25-46. Berlin: Language Science Press.

Dahl, Ö. 1985. *Tense and aspect systems.* Oxford: Basil Blackwell.

Dahl, Ö. 1990. “Standard Average European as an exotic language”. In J. Bechert, G. Bernini and C. Buridant (eds.), *Toward a typology of European languages*: 3-8. Berlin: Mouton de Gruyter.

Dahl, Ö. 2004. *The Growth and Maintenance of Linguistic Complexity.* Amsterdam and Philadelphia: John Benjamins.

Deen, K. U. 2005. *The acquisition of Swahili*. Amsterdam and Philadelphia: John Benjamins.

Elsness, J. 1997. *The perfect and the preterite in contemporary and earlier English.* Berlin and New York: Mouton de Gruyter.

Giacalone Ramat, A. 2008. “Areal convergence in grammaticalization processes”. In M. J. López-Couso and E. Seoane (eds.), in collaboration with T. Fanego. *Rethinking Grammaticalization: New perspectives*: 129-167. Amsterdam and Philadelphia: John Benjamins.

Greenberg, J. H. 1963. “Some universals of grammar with particular reference to the order of meaningful elements”. In J. H. Greenberg (ed.), *Universals of Language*: 40-70. Cambridge, MA: MIT Press.

Greenberg, J. H. 1966. *Language universals: With special reference to feature hierarchies.* The Hague: Mouton.

Haiman, J. 1983. “Iconic and economic motivation”. In *Language* 59 (4): 781-819.

Haspelmath, M. 2001. “The European linguistic area: Standard Average European”. In: M. Haspelmath, E. König, W. Oesterreicher and W. Raible (eds.), *Language Typology and Language Universals. Sprachtypologie und sprachliche Universalien. Sprachtypologie und sprachliche Universalien: La typologie des langues et les universaux linguistiques: An International Handbook: Ein internationales Handbuch: Manuel international*: 1492–1510. Berlin/New York: Walter de Gruyter.

Haspelmath, M. 2006. “Against Markedness (And What to Replace It With)”. In J*ournal of Linguistics* 42 (1): 25-70.

Haspelmath, M. 2017. “Explaining alienability contrasts in adpossessive constructions: Predictability vs. iconicity”. In *Zeitschrift für Sprachwissenschaft* 36 (2): 193-231.

Haspelmath, M. 2019. “Can cross-linguistic regularities be explained by constraints on change?”. In K. Schmidtke-Bode, N. Levshina, S. M. Michaelis and I. A. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*: 1-23. Berlin: Language Science Press.

Haspelmath, M. 2020. “Form-frequency correspondences and predictability”. To appear in *Journal of Linguistics*.

Hopper, P.J. and E. C. Traugott 1993. *Grammaticalization*. Cambridge, Cambridge University.

Howe, C. and S.A. Schwenter 2003. “Present Perfect for Preterite across Spanish dialects“. In Tara Sánchez & Uri Horesh (eds.), *Penn Working Papers in Linguistics: Selected Papers from NWAV-31* 9 (2). Philadelphia, University of Pennsylvania Press: 61–76.

Hualde, J.I. and C. Mounole 2019. “Tense, Aspect and Mood”. In A. Berro Urrizelki, B. Fernández & J. Ortiz de Urbina (eds.), *Basque and Romance: Aligning grammars*. Leiden: Brill, 59-81.

Hundt, M. and N. Smith 2009. “The present perfect in British and American English: Has there been any changes, recently?”. In *ICAME Journal* 33: 45-63.

Iatridou, S., E. Anagnostopoulou and R. Izvorski 2003. “Observations about the form and meaning of the perfect”. In A. Alexiadou et al. (eds.), *Perfect explorations*: 153-204. Berlin: Mouton de Gruyter.

Iverson, M. and J. Rothman 2008. “Adverbial quantification and perfective/imperfective interpretative Nnances in L2 Portuguese”. In *Proceedings of the 9th Generative Approaches to Language Acquisition Conference*: 70-80. Sommerville: Cascadilla Press.

Julien, M. 2002. “Optional *ha* in Swedish and Norwegian”. In *Journal of Comparative Germanic Linguistics*: 67-95.

Kardos, É. 2016. “Telicity marking in Hungarian”. In *Glossa: A Journal of General Linguistics* 1(1) 41: 1-37.

Klein, W. 1992. “The Present Perfect Puzzle”. In *Language* 68 (3): 525-552. Linguistic Society of America.

Kornfilt, J. 1997. *Turkish.* London and New York: Routledge.

Larsson, I. 2009. *Participles in Time. The Development of the Perfect Tense in Swedish.* Ph.D. dissertation, Göteborg: Göteborgs Universitet.

Li, N., S. Thompson & R. Thompson 1982. “The discourse motivation for the perfect aspect: the Mandarin particle le”. In P. Hopper (ed.), *Tense-Aspect: Between semantics & pragmatics*: 19-44. Amsterdam and Philadelphia: John Benjamins.

Liu, M. 2015. “Tense and aspect in Mandarin Chinese”. In W. S-Y. Wang (ed.), *The Oxford Handbook Of Chinese Linguistics*: 274-289. Oxford University Press.

Malmgren, S.-G. 1985. “Om utelämning av hjälpverbet *ha* i förbindelse med supinum”. In Sture Allén et al. (eds.), *Svenskans beskrivning 15: förhandlingar vid Sammankomst för att dryfta frågor rörande svenskans beskrivning*: 347-358. Göteborg: Göteborgs Universitet.

Mauroni, E. 2013. “La difficile alternanza di imperfetto e passato prossimo in italiano: tempo, aspetto, azione”. In *ACME* 1-2: 247-294.

McCoard, R. W. 1978. *The English perfect: tense-choice and pragmatic inferences.* Amsterdam: North-Holland.

Miller, J. 2004. “Perfect and resultative constructions in spoken and non-standard English”. In O. Fischer, M. Norde and H. Perridon (eds.), *Up and down the cline - the nature of grammaticalization*: 229-246. Amsterdam and Philadelphia: John Benjamins.

Pancheva, R. and A. von Stechow 2004. “On the Present Perfect Puzzle”. In K. Moulton and M. Wold (eds.), *Proceedings of NELS* 34.

Reichenbach, H. 1947. *Elements of Symbolic Logic*. New York: The MacMillan Company.

Rojo, G. 1974. “La temporalidad verbal en español”. In *Verba* 1: 68-149.

Rojo, G. and A. Veiga 1999. “El tiemo verbal. Los tiempos simples”. In I. Bosque and V. Demonte (dirs.), *Gramática descriptiva de la lengua española* 2 (cap. 44): 2867-2934. Madrid: Espasa.

Rounds, C. 2001. *Hungarian: An Essential Grammar*. London and New York: Routledge.

Sangfelt, A. 2014. *Om uppkomst och spridning av ha-bortfall i äldre svenska: En studie i syntaktisk innovation och förändring*. Uppsala: Universitet.

Santos Lopes, dos C. R. and M. T. Brocardo 2016. “Main Morphosyntactic Changes and Grammaticalization Processes”. In W. L. Wetzels, J. Costa and S. Menuzzi (eds.), *The Handbook of Portuguese Linguistics*: 471-487. West Susex: Wiley Blackwell.

Shakespeare, W. 1595. *Two Gentlemen of Verona.*

Shannon, C. E. 1949. “The mathematical theory of communication” [1948]. In *The Bell System Technical Journal* 27: 379-423, 623-656.

Stechow, von, A. 1999. “Eine erweiterte Extended Now-Theorie für Perfekt und Futur”. In *Zeitschrift für Literaturwissenschaft und Linguistik* 113: 86-118.

Stevenson, C.H. 1970. *The Spanish Language Today.* London: Hutchinson University Library.

Stewart, M. 1999. *The Spanish Language Today*. London and New York: Routledge.

Strzelecka, E. 2003. *Svenska partikelverb med* in, ut, upp *och* ner: *En semantisk studie ur kognitivt perspektiv*. Diss. Uppsala: Universitet. Skrifter utgivna av Institutionen för nordiska språk vid Uppsala universitet: 62.

Vendler, Z. 1967. *Linguistics in Philosophy*. London: Cornell University Press.

Walker, J. 2011. “The emergence of the Narrative Present Perfect in British English: reality or illusion?”. In *Groninger Arbeiten zur germanistischen Linguistik* 53 (2): 71-87.

Xiao, R. and T. McEnery 2004. *Aspect In Mandarin Chinese: A Corpus-based Study*. Amsterdam and Philadelphia: John Benjamins.

Xiqués García, T. M. 2015. *Towards a Unified View of the Present Perfect. A Comparative Study on Catalan, English and Gĩkũyũ*. PhD Dissertation. Universitat Autònoma de Barcelona.

Yao, X. and P. Collins 2012. “The present perfect in world Englishes”. In *World Englishes* 31 (3): 386-403.

Zipf, G. K. 1935. *The psycho-biology of language: an introduction to dynamic philology*. Cambridge: MIT Press.

Corpora references

Språkbanken: http://www.lrec-conf.org/proceedings/lrec2012/pdf/248_Paper.pdf

The British National Corpus: http://www.natcorp.ox.ac.uk/corpus/index.xml

The Corpus of Contemporary American English corpus: https://www.english-corpora.org/coca/help/coca2020_overview.pdf

The Mark Davies' Corpus del Español: https://www.corpusdelespanol.org/web-dial/

The Mark Davies' Corpus do Português: https://www.corpusdoportugues.org/web-dial/

Borin, L., M. Forsberg and J. Roxendal 2012: Korp – the corpus infrastructure of Språkbanken. Available online at https://spraakbanken.gu.se/korp/.

Davies, M. 2004-. *British National Corpus* (from Oxford University Press). Available online at https://www.english-corpora.org/bnc/.

Davies, M. 2008-. *The Corpus of Contemporary American English (COCA): One billion words, 1990-2019*. Available online at *https://www.english-corpora.org/coca/*.

Davies, M. 2016-. *Corpus del Español: Two billion words, 21 countries*. Available online at http://www.corpusdelespanol.org/web-dial/. (Web / Dialects)

Davies, M. 2016-. *Corpus do Português: One billion words, 4 countries*. Available online at http://www.corpusdoportugues.org/web-dial/. (Web / Dialects)