JS Journal of Historical Syntax
Volume 7, Article 19: 1–48, 2023

WHERE TO PLACE A PHRASE? AN INFORMATIONAL AND GENERATIVE APPROACH TO PHRASAL EXTRAPOSITION*

Sophia Voigtmann
Saarland University

Augustin Speyer
Saarland University

Abstract In the following paper, we aim to cast light on the placement of prepositional phrases (PPs) in the so-called postfield, the position behind the right sentence bracket. Our focus is on the period of early New High German from 1650 to 1900. In a first step, extraposition will be correlated with Information Density ('ID', Shannon 1948). ID is defined as "amount of information per unit comprising the utterance" (Levy & Jaeger 2007: 1). It can be calculated as surprisal. The higher the surprisal values the higher the impact on working memory and the more likely perceiving difficulties become (e.g. Hale 2001). We expect PP with such high surprisal values to be more likely to be placed in the postfield where more memory capacities are available than in the middle field. We test this hypothesis on a corpus of scientific articles and monographs dealing with medicine and theology and taken from the Deutsches Textarchiv (DTA, BBAW 2019). We only find evidence for the hypothesis in the timespan from 1650 to 1700 and for the rare case that attributive PPs are placed in the postfield. Since this has already been shown for attributive relative clauses (Voigtmann & Speyer 2021), we want to take this up and argue for a similar generative analysis for attributive PP and relative clauses in a second step.

* This paper is based on material collected in the context of the CRC ‚Information Density and Linguistic Encoding', funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102. Most of the material was presented at the Workshop Diachronic Generative Syntax 22 (DiGS 22) in 2021. Our thank goes to the organizers, and the audience of this workshop for many helpful comments, furthermore to Katharina Wilhelm, Nils Dörr and Katrin Ortmann for helping with the corpus creation and tools. All remaining errors are of course ours.

1 Introduction

Since its final establishment in the Early New High German (ENHG) period, the sentence frame has been one of the key characteristics of the German language. The sentence frame describes the distribution of verbal material over two positions, the so-called left and right sentence brackets. The left sentence bracket (LSB) hosts the finite verbal material or subordinate conjunctions whereas the right sentence bracket (RSB) holds non-finite verbal parts or the whole predicate in verb-last clauses. From a generative perspective, the LSB corresponds to C° (e.g. Dürscheid 1989) and the RSB is the base position of the verb at the right edge of the clause. The RSB is consistent with the end of the clause. Nevertheless, this clause boundary is frequently crossed especially by other and mostly subordinate clauses, but also by phrases as the example in table 1 shows. Their landing position is the so-called Postfield (PoF).

Prefield	LSB	Middle Field	RSB	Postfield
Ich	habe	ihm Geld	gegeben	**für ein Eis.**
I	have	him money	given	for an ice.cream.

Table 1 Example of an extraposed PP.

The PoF is not usually filled with prepositional phrases (PPs). Unfortunately, the literature on this topic is rather vague about the triggers for PP extraposition in most cases. Exceptions are given by Hawkins (1992) and Weber (2019), who tested her assumptions experimentally and found a correlation between length and distance for PP extraposition out of other phrases. Grammars and theoretical literature agree that PPs are the most common phrase type found in the PoF. The reason for their extraposition is their (great) length and complexity, followed by other frequently given explanations of extraposition like avoiding an overstrain of the sentence frame, discourse, or information management (Zifonun, Hoffmann & Strecker 1997). In summary, those explanations mostly claim that PP extraposition ensures successful communication, but they hardly ever check this claim empirically, especially not for historical data.

We want to fill that gap and continue the research made by Voigtmann & Speyer (2021) by adding the periods between 1650 and 1850 and theological data (see section 3) and explain the phenomenon with information theory (Shannon 1948). But we will go beyond the scheme of information theory and use the results to argue for a similar generative analysis of attributive

PPs and relative clauses, and a mingling of information theory and generative approaches.

We apply Information Density as a theoretical foundation of extraposition. Information Density describes information as "the amount of information per unit comprising the utterance" (Levy & Jaeger 2007: 1). This information is also called surprisal and is understood as the likelihood of a word given a context of a certain number of words. Frequent word combinations receive lower surprisal values as they are more expected in this particular context whereas rare combinations gain higher surprisal values and, thus, contain more information. Hale (2001) and Levy (2008) have shown a connection between surprisal values and processing difficulties. The higher the surprisal value, the more strain is put on the working memory to process the information. We take up this idea and combine it with the approaches of Gibson (1998) and Hawkins (1992), which claim that more memory capacity is free in the PoF. Therefore, information that is harder to process is more likely to be placed there. This leads us to our first hypothesis i:

i. Higher surprisal values of phrases favor phrasal extraposition.

We test the hypothesis on a corpus of medical and theological texts from 1650 to 1900. They are taken from the Deutsches Textarchiv (DTA, BBAW 2019) divided into 50-year time spans to account for a possible language change. We manually annotated the extraposed PPs and for each an embedded counterpart. Language Models were used to calculate the surprisal values which were in turn used to estimate the processing difficulties of the constituent. We found a decreasing influence of ID over the centuries and a strong influence of length for the whole data. When we split the data into independent constituents and attributive PP, the results change and we see an effect of the surprisal values on extraposition (see section 4).

We take this as a basis for the generative approach and argue that placement in the Postfield is mainly due to PF considerations (see section 5).

The paper is structured into the following two parts: First, we will present the information theoretic approach and results. This includes a general overview of PP extraposition (section 2.1) and Information Density (section 2.2). Then, we present our corpus and the methods applied for the information theoretic investigation (section 3) before we show the results (section 4). They will be discussed subsequently (section 4.3). The second part of the paper consists of the generative considerations about the PoF and PP extraposition there (section 5). The conclusion closes the paper (section 6).

2 Theoretical Background

This section gives an overview of previous research on the topic of the placement of PP in the postfield.[1] To do so, we will first give a detailed overview of the "Topologisches Feldermodell" (topological field model) to classify the postfield and to introduce termini used in the other sections. Second, we will talk about reasons for extraposition in modern German and show that these explanations have not changed over the centuries despite the decreasing tendency to extrapose phrases. The first and more descriptive part of this section (2.1) is followed by information about Information Density (2.2) because we understand information in terms of Information Density in this paper.

2.1 PP extraposition

Topological Field Model As already mentioned, the landing point of PP extraposition is the postfield (PoF), i.e. the area behind the right sentence bracket. The division of the German sentence into different fields, which may have further subfields, goes back to linguists of the 19th century, and is generalised by Höhle (1981) and later by Wöllstein (2014) with a somewhat modified terminology. We refer to the model by Wöllstein (2014) and use own translations for the names of the individual fields. Older models (e.g. Engel 1970) use a more complicated and less generalisable structure, which would also restrict the evidence we found and further limit the already small amount of data (see section 3.1).

Pre-Pref.	**Prefield**	**LSB**	**Middle Field**	**RSB**	**Postfield**
Aber	gestern	habe	ich von Paul	erfahren	...
But	yesterday	was	I by Paul	told	...
		dass	wir bald	reden	über das Buch.
		that	you soon	reden	about the book.

Table 2 'But yesterday I was told by Paul that we talk about the book today' Example of a complex sentence in the 'Topologisches Feldermodell'.

As can be seen in table 2, a German sentence can be divided into six fields: the pre-prefield (Pre-Pref.), containing, for example, appellatives, and the prefield, which may be occupied by only one constituent of any size, the left sentence bracket (LSB) with either subordinating conjunctions or the finite verb,

1 We will call this extraposition for the sake of simplification and independent of possible movements for now.

the middle field in which a theoretically unlimited number of constituents can be found, the right sentence bracket (RSB), which contains either infinite verbs, verbal particles or, in subordinate clauses, the whole verbal complex, and the postfield (PoF), which is usually filled with subordinate clauses, parts of comparative phrases, but also with phrases. In various sources, we find approaches to further subdivide the different fields. Engel (1970), for example, distinguishes 12 positions in the middle field. For the postfield, different positions have also been described by Altmann (1981), depending on the extraposed material, possible prosodic effects, and the presence of resumptives in the midfield. He describes the placement of complements in the PoF as "exclusion". These complements do not have a resumptive in the middle field, but they do have a syntactic function.[2] He counts prepositional objects among the elements most frequently placed in the PoF.

PP extraposition in Modern German In Zifonun et al. (1997), PPs are also cited as the phrases that are among those which are most frequently placed in the PoF in contemporary German. However, it must be stressed that the position of PPs in PoF is not the normal case and that they are far more often found in the prefield or middle field, and that they are rated more naturally there (e.g. Weber 2019). So, their positioning must be motivated by other factors. The literature (e.g. Zifonun et al. 1997, Wöllstein 2014) on this point can be summed up as follows; the length of a PP increases the likelihood of extraposition, and by placing material in the postfield successful communication is ensured, and extraposition is, thus, a more pragmatic feature of language.

Extraposition shortens the distance between the LSB and RSB. Distance and memory capacities have not just recently been found to play a crucial role in processing. Futrell, Gibson & Levy (2021) and before them, for example, Gibson (1998) showed that an increase in the distance between cohesive linguistic material also increases the likelihood of forgetting the first part of this material because the strain on the memory becomes too high. The result is information loss and, thus, failed communication (see section 2.2 for a more detailed account of information loss and communication failure). Length is often used synonymously with complexity. The more words a phrase contains, the more likely it is to be moved to the adjunct field. This increase in length can happen through coordinating clauses, adding attributes of any

2 Although Altmann (1981: 67) differentiates between optional and obligatory arguments, he fails to make a statement about adjuncts. A distinction is made overall between elements that have a syntactic function in the sentence, which would also apply to adjuncts, and attributes or predicatives (Altmann 1981: 70), which are extraposed as so-called "Nachträge". We assume that "free indications" can also be counted among the elements that (Altmann 1981: 67) describes with "echter Nachfeldbesetzung" ('genuine post-field occupation')

kind, or increasing the complexity of the attributes themselves (example (1)).

(1) *Ich habe$_{LSB}$ [auf dem Bahnhof, wo wir uns damals getroffen*
I have [at the station where we us then met
haben, als du zum ersten Mal in meiner Stadt warst,]
have when you for.the first time to my city was]
gestern lange auf dich gewartet$_{RSB}$.
yesterday long for you waited.
"I have waited for you for a long time [at the station where we met at the time when you have been to my city for the first time."

(2) *Ich habe$_{LSB}$ gestern lange auf dich gewartet$_{RSB}$ [auf dem*
I have yesterday long for you waited at the
Bahnhof, wo wir uns damals kennengelernt haben, als du
station where we us then met have when you
zum ersten Mal in meiner Stadt warst.]
for.the first time to my city was.
"I have waited for you for a long time [at the station where we met at the time when you have been to my city for the first time."

When we look at the examples, we also see that in example (2) the distance between the LSB and the RSB is much shorter than in example (1). The middle field is kept as compact as possible to prevent the aforementioned strain on the processing capacities. These are more likely to be impeded in oral than in written communication (Brooks 2006, Zahn 1991). In oral discourse, information is fleeting due to the medium. If the information has not been understood it must be tediously re-asked. In written communication, on the other hand, information is available for longer; it can simply be re-read. Therefore, extraposition is often called a characteristic of the oral discourse mode (Koch & Österreicher 2007). When it is found in modern German written texts, it is also often used to introduce a new topic for the following paragraph (e.g. Petkova-Kessanlis 2015).

Both in written and oral communication, extraposition stands between the conflicting priorities of two different kinds of distance. As shown above, one goal of extraposition is to shorten the distance between the sentence brackets, but also the distance between the possible position in the middle field where the phrase would otherwise be and the PoF.[3]

3 Although there are preferences for the position of adjuncts and complements in the middle field, this distance can only be measured accurately for attributes moved from other constituents since there is a wide margin for the sequence of objects and adjuncts in the middle field (Frey & Pittner 1998 and, to some extent, Lenerz 1977).

To briefly sum up: Extraposition, in the sense of material standing in the postfield, is said to be caused by the length of the extraposed material, but only if the distance between the former position of the PP and the PoF is not too large. The main goal behind this is the prevention of information loss and ensuring successful communication.

PP extraposition in German diachrony Surprisingly, this does not change much in the history of German. Since we concentrate on Early New High German (ENHG) and early Modern German (eMG), we will just give a brief overview of extraposition in Old High German (OHG, 750–1050) and Middle High German (MHG, 1050–1350).

To talk about extraposition in OHG is rather difficult. Though scholars like Hinterhölzl (2010) and Borter (1982) describe the phenomenon and find it in texts from this period, it must be made clear that OV and VO linearizations were allowed and unmarked in OHG (Hinterhölzl 2010: 128). Still, tendencies towards a VO order are visible and linked to weight and informational aspects of the object in question. Hinterhölzl's 2010 study is mostly about NP extraposition, but his findings agree with those of Paul (2007) in that there are differences between the extraposition out of main and subordinate clauses - extraposition is more frequent out of main clauses. At least for Notker's psalms, Borter (1982) describes a correlation between the position of a phrase in the Latin original and the extraposition in the German translation. He finds only 57 exceptions (Borter 1982: 122). Furthermore and in accordance with evidence from Modern German and MHG, he also observes that complex phrases, especially when extended by a relative clause, are more likely to be extraposed (Borter 1982: 122).

That the RSB can be regarded as the end of the clauses is already evident in MHG, but it is still not fully established (Bergmann 2019). According to Paul (2007), already half of all propositional clauses have the RSB as a boundary symbol in MHG, with the occupation of the marginal fields occurring still more frequently than in later periods. Already at that time, PPs were increasingly extraposed in cases of long matrix clauses (Morlicchio 1991) and in cases where an extension of the phrase by a sentential attributive like relative clauses (Paul 2007) is given. Paul (2007) also finds that the linearization "inf > fin" of the verbal material in the RSB is connected to extraposition. Moreover, main clauses have higher rates of extraposition than subordinate clauses (Paul 2007).

In Early New High German (ENHG, 1350–1700, Polenz, Moulin & Harion 2013), the sentence frame is said to be fully established. Eichinger (1995) relates this to the extralinguistic factors like changes in society and higher

demands made by scientific discoveries as well as the need to communicate most efficiently. This is achieved because the sentence brackets direct the recipient's attention to grammatical information, mostly found in the LSB, and lexical information, mostly found in the RSB (Dal & Eroms 2014: 202). Dal & Eroms (2014) postulate that this is connected with the increasing distribution of information on independent words since OHG. Still, the principle of an unbroken sentence frame is a phenomenon of written rather than oral discourse mode (Dal & Eroms 2014: 203). Extraposition of phrases was highly common in ENHG, as shown by e.g. Schildt (1976), but decreases in the course of this period. We have already described for MHG and Modern German that the length of the matrix clause and of the phrase itself influences extraposition. For ENHG, Ebert (1980) confirms this in his corpus of texts from Nuremberg. Sapp (2014) also confirms these results and, thus, supports Zifonun et al. (1997)'s findings on Modern German that PPs are the most frequently extraposed phrases, followed by NP, several elements of various types, adverbial phrases, and pronouns. Using logistic regression to determine which of the following factors is most influential for extraposition (among them: type of constituent, length of the extraposition in syllables and words, focus, clause type, period, dialect, and genre of the text in the corpus), he finds that length and contrastive focus or focus signaling new information are the most important factors for extraposition, followed by genre (Sapp 2014). He, thus, confirms the distinction between conceptually oral and written sources (Koch & Österreicher 2007); extraposition is more frequent in texts that are closer to orality. To sum up Sapp's (2014) findings; in texts from 1350–1650, grammatical and pragmatic factors influence extraposition.

This is also reasoned by scholars from the late ENHG period. Their description of the sentence frame does not vary much from the one we have given above for modern German (Takada 1998). Regarding extraposition, Takada (1998) contrasts conflicting statements of grammarians from the late 17th century. In general, they accept the sentence frame as a major characteristic of German, but they are unsure about the reasons for breaking it. One reason to do so is to be able to retain the meaning of the sentence which might be lost otherwise because of the increasing distance between the LSB and RSB. On the other hand, a complete sentence frame is called prestigious (Takada 1998: 229) and applied for sociological reasons (Takada 1998: 230). These statements built the transition to the period we focus on in our current research.

The period from 1700 to 1900 provides us with the difficulty that systematic research is rare. An exception is Admoni (1990), who gives an overview of the whole German language, but is also rather non-systematic and pro-

vides a poor methodological quality. In the period 1500–1700, extraposition is applied less frequently and mostly for "subordinate clauses, infinitive construction, comparative elements and prepositional constructions"(Admoni 1990: 201, translation made by us). The same holds for the 18th century extraposition while he claims for the 19th century that sentence structures become shorter, easier and clearer (Admoni 1990: 220). Again, it is important to notice that PPs are among the most common phrase type for PoF placement. This is also confirmed by Konopka (1996) who investigates the 18th century. It does not matter whether PPs are adjuncts or arguments. Both types can be extraposed (Konopka 1996: 167). The reason given by scholars from that period resembles those that we stated above: Extraposition is used to keep the sentence frame more compact and to ensure successful communication in doing so (Konopka 1996: 178). Thus, it is not surprising that phrases containing a relative clause are more often found in the PoF than in single PPs. Furthermore, he summarizes several pragmatic factors for extraposition like either adding less important information or the contrary and building a second area of emphasis in the sentence. More important information is said to occur later in the sentence (Konopka 1996: 171).

Research about later periods is even more scarce or not focused on extraposition. Engel (1970: 55) summarizes this as follows: A consistent tendency towards extraposition cannot be found. It varies between text types. We intend to fill this gap at least for scientific texts and for the period 1650–1900.

2.2 Information Density

General Concept of Information Density The previous section showed that PP extraposition is regularly correlated with an improvement in information transmission and, thus, successful communication. The latter is one of the main goals also pursued by Information Density (ID, Shannon 1948) theory. Based on the communication model which includes a speaker, a recipient, a message consisting of a string of words, and a channel of a limited capacity through which this message is transmitted, Shannon (1948) presents an objective and mathematical model to explain how successful communication is achieved in this model. Successful communication takes place when the recipient understands the information the speaker has sent. In an ideal setting, that means that no information is lost. In ID, information is not just understood as some vague concept but measurable as the probability that one element follows others. It is further defined as follows: Given a certain context, unexpected words carry more information than expected ones. A negative correlation between expectability and information is observed. Since most languages offer flexibility in the alignment of words (example (3)) and be-

cause language users are aware of this set of codes, both speaker and recipient can make use of these variabilities and knowledge to (usually unconsciously) choose the most suitable code from this set during the transmission of a message.

(3) **Tomorrow,** he will go to the dentist.
He will go to the dentist **tomorrow.**

All possible choices from this set of codes are equally probable according to Shannon (1948). Because of that, information is calculated using the logarithmic function on the basis 2 and measured as bits. The value we obtain with this formula is called surprisal value:[4] $P(word) = -log_2\ (word|context)$.

Context and its influence on Information Density Though the choice of different codes is equally probable, the words within the sets can have different surprisal values. Thus, the signals and the coding must be adapted to the kind of transmission without exceeding the limits of the channel through which the message is sent and its specific capacity. Every message is transferred to a language that already limits the possible set of codes because of its specific syntax. It further defines a natural frequency of elements. Speaker and recipient know these boundaries. That leads either to time-saving in the transmission of the message or to a reduced load on the channel when the signal sequence correctly encodes the message (Shannon 1948: 384). The transmission of the symbols is incremental and dependent on the previous symbol and the symbol itself. A stochastic process and probability theory can describe how subsequent symbols are selected (Shannon 1948). $p_i(j)$ adequately describes the probability of j following i (Shannon 1948: 384). When only the element itself is considered in its frequency, it is called unigram frequency. Unigram frequency is the simplest way to approach the stochastic process of element selection. However, this simple approach does not even approximate existing languages. Bigram, trigram... n-gram take more context than just the current word into account. In the case of bigram-surprisal values, the probability of occurrence of the element before the considered element is included, in the case of trigram-surprisal values the two preceding elements are included, and so on. Thus, they can describe real languages better: "A sufficiently complex stochastic process will give a satisfactory representation of a discrete source." (Shannon 1948: 386). All considerations about ID have so far referred to a "noiseless channel" (Shannon 1948: 398),

4 We will use the term *surprisal* instead of conditional entropy as suggested by a reviewer though the terms can be used synonymously.

that is, to a channel where no disturbances are present. This is, however, not only rarely the case, but also not the core question of Shannon (1948)'s ID theory. He focused on the transmission of messages through a noisy channel. A noisy channel is for example present when there are actual background noises or information is for some other reason not received in exactly the way it was sent. Nevertheless, most communication is successful even under the circumstances of a noisy channel (Shannon 1948: 398). That the recipient still understands the message is highly dependent on context, because certain words are so expectable in their context that they can easily be guessed and that they do not carry much information in this context. Another way to ensure successful communication is to use redundancy. Speakers use this in the hope of having information loss on the redundant parts instead of highly informative parts. Certain words are more expectable in their context than other words (example (4)):[5]

(4) You may now kiss the **bride**.

This sentence is extraordinarily common in wedding scenarios. Thus, the recipient expects *bride* to follow "kiss the". *bride* does not carry much information and should receive a small surprisal value. Still, we must distinguish between the possible sizes of context. In a narrow context, even small changes can have a high impact on the surprisal value of a word. If we change the predicate to *lecture* in example (4), *bride* would gain a much higher surprisal value because we do not frequently see the combination of these two words.

However simplified these examples are, they give a first impression of ID theory and a problem of strong focus of classical surprisal calculation on single words. Ostermann (2021), among others, could show that script and world knowledge also strongly influence the expectability of a word and its processing. However, the strong intralinguistic orientation of classical ID approaches is highly useful for historical data because researchers can only have limited access to the actual world knowledge of the period in question and even less knowledge about the script and world knowledge of an individuum of this time. But because Shannon's (1948) calculation methods of ID are purely orientated towards written sources and provide suitable measurements, it facilitates objective data evaluation.

Other linguistic sources and processing A further advantage of ID is that "the relationship between the predictability of linguistic material and efficient communication exists at all linguistic levels (Gibson, Futrell, Pianta-

5 The following examples are taken from Voigtmann & Speyer (2021).

dosi, Dautriche, Mahowald, Bergen & Levy 2019) and a relationship between processing effort, i.e., psycholinguistic reality, and information density could be shown as well (Levy 2008: among others)" (Voigtmann & Speyer 2021: 6). Levy (2008) describes a probabilistic, expectation-based theory of syntactic understanding. Human resources for language processing are limited so they face problems when they have to process syntactic structures which are known to consume more resources than others. If this is the case, then the channel is overloaded and information is lost. To prevent this, it gains importance that information is understood based on different sources: structural, lexical, pragmatic, and discourse-based (Levy 2008: 1128). Thus, similar analyses compete with each other because the aforementioned sources are combined for understanding (Jurafsky 2003). These efforts correspond to the surprisal of a word. Here, an interface is built between the linguistic representation during the sentence comprehension and processing difficulties on a certain word within a sentence (Levy 2008: 1128). The recipient thereby keeps in mind the complete set of the different, probable, partially processed constituents from the already seen or heard input. S*he assigns to it a possible probability distribution over the complete structure to which the already received constituents can expand. Surprisal is, thus, seen as the difficulty of replacing an old distribution with a new one (Levy 2008: 1132).

Uniform Information Density The core element of successful communication is to distribute information as evenly as possible at all linguistic levels. Levy & Jaeger (2007), and before them Aylett & Turk (2004), found that speakers arrange their utterances in a way to prevent peaks and troughs in the information profile of all linguistic levels. On the phonetic level, speakers vary the duration of phones. They take longer to pronounce unexpectable or less frequent words (Aylett & Turk 2004). Levy & Jaeger (2007) investigate syntactic reductions, "a phenomenon in which speakers have the choice of either marking a phrase with an optional word or leaving it unmarked" (Levy & Jaeger 2007: 2). Their research topic is the optional *that* in English Relative Clauses (RCs). In their corpus study, they find that the insertion of *that* can lower the surprisal value of the first word in the RC which would otherwise exceed the channel capacity and cause information loss. This first evidence results in the *Uniform Information Density Hypothesis*. The usage of *that* to reduce the surprisal value at the relative clause onset was found across registers and standard varieties (Jaeger 2010: 163). UID can also be integrated into existing processing approaches and preferences. "Dependency processing accounts", which assume that preference is given to variants that have shorter dependency relationships, were taken up by UID, as well as Behaghel

(1932)'s "Gesetz der wachsenden Glieder" (law of increasing constituents). Furthermore, it concerns "alignment accounts" which regard access to referents as a major factor for linguistic preferences. These accounts rely on the conceptual accessibility and mentioning of referents and can be combined with "availability accounts", which focus more on the referent and claim that cognitively available material appears earlier in the sentence (Jaeger 2010: 165). Incremental speech production is also related to this; what is available earlier can be expressed earlier, which in turn can be combined with the other approaches mentioned above.

In summary, Shannon (1948)-ID determines the information content of a word in a certain context and links this information content to the likelihood of the word in the context. The surprisal value is calculated by the logarithmic function and expressed in bits. ID theory aims to provide a descriptor for the optimal encoding of a message and thus to be able to demonstrably describe how information loss can be prevented. In the classical method of calculation with n-grams, all words, that is content and function words, are considered in the calculation of the surprisal values, whereas in classical information-structural studies often only content words are considered. Thus, even the absence of positional changes can already lead to visible effects.

3 Corpus and Methods

3.1 Corpus

We built our corpus on texts from the Deutsches Textarchiv (DTA, BBAW 2019). It was created as a project funded by the Deutsche Forschungsgemeinschaft (DFG) to provide an overview of the German language from the 16th to the 20th century. Newspaper articles, novels, "Gebrauchsliteratur", that is literature written for a specific purpose, and scientific articles and monographs are used to give this overview. Besides this broad basis of different genres, the texts are tokenized, normalized, lemmatized, and also POS-tagged.[6] This and the diversity in genres are the major advantages of the DTA. There is no other collection of texts, at least to our knowledge, that includes pre-processed scientific texts.

On the one hand, this is not surprising because German scientists used Latin as their lingua franca until the 17th century. The German scientific text tradition begins rather late in the ENHG period. In their beginnings, the texts resembled letters instead of the articles we know today. These letters were even addressed to interested colleagues at first and published later. This puts

6 Unfortunately, especially the POS-tagging is of low quality, which has prevented us from using automatic annotations or language models based on these annotations.

the genre in the field of tension between the two poles of written and oral discourse mode, which were presented by Koch & Österreicher (2007). These early scientific letters are assumed to be planned as conceptually oral. They address, for example, the addressee or present shorter middle fields. We find these characteristics in the early texts. Nevertheless, we are also aware that the writers might also stand in the field of tension between the former Latin tradition and this more oral style of (letter) writing. The Latin tradition is interesting and worth mentioning because Latin has free word order, but is highly influenced by informational considerations. Furthermore, long dependencies are very common in Latin but not in German. When the early writers of German scientific texts try to imitate this style, these long dependencies might put a strain on German native speakers' parsing capacities and might, therefore, contradict an optimal distribution of information because the sentences are too long or too intertwined. Although the Latin tradition stays present, we still expect its influence to decrease over the centuries. 19th century texts are more likely to favour style with shorter clauses and fewer extrapositions. We want to show and include these possible changes in our account which is another reason for choosing the DTA as a corpus basis.

Since we focus on lexical instead of grammatical predictability, the lemmatization of the data is most important for our analysis. Orthography has not been standardized in the early NHG period, so normalization is necessary as a first step to processing the data. However, in normalized data, case marking is included in the language model which would cause further data sparsity. To have a broader data basis for the language model, we use lemmata for our language model to capture only the lexical likelihood of the tokens.

Our text sample from the DTA consists of 12 medical texts with 1,083,720 tokens and 15 theological texts with 778,976 tokens from 1650 to 1900. We chose the texts arbitrarily and excluded translations. As already mentioned in the introduction (1), we divided the time span into five parts of 50 years to account for language change. For a better overview of the corpus, a table with the texts included in the corpus is provided in table 3.[7]

7 The annotated data set is available at https://github.com/SFB1102/C6Samples.

Author	Year	Title
Abel	1699	Wohlerfahrner Leib-Medicus Der Studenten [...]
Bengel	1751	Abriß der so genannten Brüdergemeine[...]
Benner	1739	Christliches Bedencken von dem vorsetzlichen Meineid[...]
Bräuner	1714	Pest-Büchlein[...]
Carus	1820	Lehrbuch der Gynäkologie. Bd. 1.
Daumer	1895	Die dreifache Krone Rom's. [...]
Egger	1895	Der christliche Vater in der modernen Welt[...]
Gall	1791	Philosophisch-Medicnische Untersuchungen über Natur und Kunst im kranken und gesunden Zustande des Menschen
Hanssen	1731	Achtzig erläuterte Grund-Fragen.[...]
Hasak	1893	Die Predigtkirche im Mittelalter
Koch	1878	Untersuchung über die Aetiologie der Wundinfectionskrankheiten
Kraepelin	1892	Ueber die Beeinflussung einfacher psychischer Vorgänge durch einige Arzneimittel.[...]
Löhe	1847	Prediget das Evangelium aller Creatur! Predigt am Missionsfeste zu Nürnberg den 17. Juni 1847
Ludwig	1856	Lehrbuch der Physiologie des Menschen
Niviandts	1708	Güldenes Schwerd[...]
Pahl	1799	Leben und Thaten des ehrwürdigen Paters Simpertus
Reil	1803	Rhapsodieen über die Anwendung der psychischen Curmethode auf Geisteszerrüttungen
Rotth	1692	Eylfertiges Bedencken über M. August Hermann Franckens [...]
Spener	1676	Pia Desideria[...]
Strauß	1835	Das Leben Jesu, kritisch bearbeitet
Unzer	1746	Gedanken vom Einfluß der Seele in ihren Körper

Table 3 Texts used for the corpus.

3.2 Annotation

When annotating extraposed PPs, we first searched for those prepositional phrases that can be found in the PoF (we follow Wöllstein's (2014) understanding of the PoF here). The RSB must necessarily be present to label a PP

"extraposed". In many cases, independent PPs have been placed in the PoF. The PP was annotated inclusive of all attributes and other extensions dependent on them. These include adjective attributes, relative clauses, and other dependent subordinate clauses (example (5)a). When two nominal phrases depend on the same preposition and are coordinately linked, they are annotated as one PP (example (5)b). If they depend on different prepositions, they are annotated as two (or more) phrases.

(5) (a) *Es werden die Menschen auch von der Pest*
It are those people also by the plague
angestecket$_{RSB}$ [*durch die Bett / auf welchen jemand an*
infected [in the bed on which someone of
der Pestilentz gestorben / oder kranck gelegen.]
the plague died or sickly laid.
"People are infected by the plague in lying on beds on which someone else died of plague or has been lying sick." (Braeuner 1714)

(b) *Darnach kommt er auch her*$_{RSB}$ *von Verstopffung des*
Afterwards results it also *RSB* of constipation of.the
gantzen Unterleibes / von sitzen und Nacht trincken /
whole underbelly of sitting and night drinking
nicht selten auch von dem verderbten und versaurten
not rarely also of the spoiled and sour
Magen.
stomach
"Afterwards, it also results of constipation of the whole underbelly, of sitting and drinking at night, not rarely also of spoiled and sour stomach." (Abel 1699)

After annotation of the extraposed PP, matching minimal pairs are searched. These embedded PP should fulfill the following criteria:

- It must be initiated by the same preposition as the extraposed PP to control for possible semantic effects of this preposition.
- It must have a comparable length and complexity. If we only take, for example, only short embedded PPs into account, they naturally have a lower cumulative and mean surprisal value than more complex ones (for the surprisal calculation used here, please see section 3.3), so our data might be skewed.

- To be annotated as embedded the PP must either stand in the middle or the prefield. To be embedded in the middle field, either the RSB must be filled lexically or there must be at least one more constituent between the end of the clause and the prepositional phrase in question. Since it was especially hard to find PPs fulfilling that criterion and the first two, we decided for a better operationalisation to take the prefield as an embedded position as well. We are aware that the prefield is also a boundary position with restrictions. But there should be at least as many cognitive capacities free as in the middle field so we expect this position not to change our results significantly.
- In the first step, no distinctions were made as to whether the PP is an independent constituent or an attribute. In the second step, we divided the data accordingly.

However, it was still not always possible to find a matching PP that fulfilled all criteria. In these cases, we graded the criteria. The latter, the position, must be matched in any case. When it was not possible to find an embedded PP with the same preposition as the extraposed one, a phrase with a different preposition was sought. Thus, if a prepositional phrase was found in the postfield with, for example, "by", but no other prepositional phrase with this preposition could be found in the text that met the embedding criteria, one with a different preposition was searched for that was similar in complexity to the extraposed one. Other properties such as length and complexity were of secondary importance. See example (6) for a minimal pair:

(6) (a) *Ihr habt fleissig zu merken, daß sie durch aus*
You have diligently to note that they indeed
keine Verwandtschaft haben$_{RSB}$ [*mit dem alten Buch, das*
no relationship have [with the old book that
man die Bibel heisset:]$_{PP}$
one the bible calls]
"You must note diligently that they do not have relations with the old book which is called 'bible'." (Bengel 1751, extraposed)

(b) *Siegfrieds Beleuchtung, das Creuzreich selbs, und die*
Siegfried's explanations the realm itself and the
Reflexionen [*mit ihren Beylagen, unter denen sich*
reflexion [with their additions among which itself
auch die Gewissens rueget]$_{PP}$, *zielen*$_{LSB}$ *auf*
also the conscience is] aim to
Verantwortungen, und gehen auf Historien.
responsibility and refer to histories.

> "Siegfrieds' explanations, the old realm itself and the reflection with their additions, among which the conscience is also present, aims to responsibility and refer to histories."
>
> (Bengel 1751, embedded)

The search for minimal pairs was usually done in the immediate context of the extraposed PP, but the entire text is also considered. In some cases, we still could not find a suitable embedded counterpart, and this explains the slightly imbalanced data (see section 4). For every annotated PP, we calculated the mean surprisal value, the cumulative surprisal value, and the length (see section 3.3).

3.3 Language Model and Surprisal Value Calculation

Language Model and Skipgram Surprisal We have chosen to use conditional surprisal as a measure of information density. For its calculation, a language model (LM) is needed. LMs are mostly calculated using training data. This data is necessary to gain estimated values over the following word given its context, resulting in a LM. The process uses a hidden Markov model. It states that the probability of a future unit can be predicted without looking too far into history (Mürmann 2014). It is not necessary to include every linguistic utterance ever made in the calculation and it is sufficient to incorporate only a part of these utterances to create acceptable statements. In calculating the probability of a word occurring in the context of the training data, we can map that likelihood to the test data via Maximum Likelihood Estimate: "The maximum likelihood estimate is so-called because it is the choice of parameter values which gives the highest probability to the training corpus. [...] It does not waste any probability mass on events that are not in the training corpus, but rather it makes the probability of observed events as high as it can subject to the normal stochastic constraints." (Manning & Schütze 1999: 198). Since it is unlikely to see every possible combination of the test data in the training data, smoothing methods are necessary to assign surprisal values to those unseen combinations of tokens.

There are several ways to calculate conditional surprisal values, one of the most famous being bigram or trigram surprisal. However, our data set is not large enough to calculate trigram surprisal values and bigram surprisal does not capture enough context. Thus, we use a skip-gram Language Model with a 2-skip-bigram (Guthrie, Allison, Liu, Guthrie & Wilks 2006). This model does not only take the immediately adjacent words into account but allows tokens to be skipped. In doing so, the number of possible n-grams per sentence is increased, which results in a larger coverage of possibly important

Period	Training Data (in tokens)	Test Data (in tokens)
1650-1700	2107590	48169
1700-1750	1481259	39251
1750-1800	2572263	26325
1800-1850	998639	16757
1850-1900	1270561	29060

Table 4 Language Model Data.

n-grams.[8] Guthrie et al. (2006) argue that their model captures context better and achieves a better coverage of the data in cases of small training data. Although some combinations do not make sense, they do not worsen the coverage of the model (Guthrie et al. 2006: 1225). Furthermore, the positive effect Guthrie et al. (2006: 1225) found can otherwise only be achieved by increasing the size of the training corpus. This is the main argument for the usage of this kind of language model as early NHG data pre-processed in a feasible way for this study is still sparse. We trained the model on all remaining scientific medical and theological texts in the DTA which we did not include in the test data itself. For our model we used lemmata and excluded punctuation marks because their usage has not been standardized in early NHG and is, thus, neither meaningful for the distribution of information nor can it be generalized over authors and centuries. All model calculations are done by an SFB-internal tool (Kusmirek, Greenberg, Oualil & Klakow 2023). For every 50 years, we calculate a separate LM. Usually, it is useful that we have a higher amount of training than test data. We were able to fulfill this criterion, as table 4 shows.

Cumulative and Mean Surprisal Values We calculated two different surprisal measures; the cumulative and the mean surprisal value. The cumulative surprisal value is the sum of all surprisal values of the annotated part (see for the annotation process section 3.2), while the mean surprisal value is the arithmetic mean of all these values. Cumulative surprisal values are justified by the parallel processing of information (McClelland & Elman (1986),

8 Guthrie et al. (2006: 1222) explain this with the example "Insurgents killed in ongoing fighting.". In a classic bigram-approach, the model would see the following four bigrams: "insurgents killed, killed in, in ongoing, ongoing fighting". When a 2-skip-bi-gram-model is used, the number of combinations increases to nine: "insurgents killed, insurgents in, insurgents ongoing, killed in, killed ongoing, killed fighting, in ongoing, in fighting, ongoing fighting".

among others). In Shannon (1948)'s information theoretic approach, only incremental processing is implemented. The likelihood of a word is, as described above, only dependent on its immediate context. Depending on the n-gram model used, this context is more or less narrow. For most experimental methods such as reading time studies, it is sufficient to focus on a specific position in the sentence but when regarding a wider context the method is no longer feasible. When a whole clause or phrase is concerned, we must take into account the parallel processing of grammatical structures or the inclusion of different sources (e.g. Cutler 2008). We claim that the cumulative surprisal value calculation is a suitable approximation to capture the processing of constituents or even complete clauses. We do not expect all cognitive capacities to be free after having processed one word. Instead, we assume some kind of legacy to remain present while the next word is processed. The cognitive load, thus, increases until the construction is processed completely. Only then can the recipient start the process anew.

Using summation still comes with problems. First, it is highly affected by length. In Voigtmann & Speyer (2021), it is argued that an increase in length must not necessarily result in an increasing surprisal value but some correlation cannot be ruled out. To factor out this influence, we also use the mean surprisal. Because the arithmetic mean is naturally influenced by outliers it reflects processing difficulties on highly unexpected words sufficiently and is less influenced by the length of the material. We also calculate the length separately by adding the number of tokens of the respective PP using R (R Core Team 2018).

4 Results

To find evidence for our hypothesis, we apply both descriptive and inferential statistics. R (R Core Team 2018) is used for all analyses. We performed linear regressions (glm, Base-Package R Core Team 2018) to determine the best predictors for PP extraposition. The model includes cumulative surprisal values, mean surprisal values, length, genre, and period and their two-way-interactions.[9] The variable "genre" was sumcoded; -0.5 was used for medical texts and 0.5 for theological texts. To include the period in the model, we changed the variable into an ordinal variable. The period 1650 to 1700 is coded as 1, the period 1700 to 1750 as 2, and so on.

The following section will show that it is worthwhile to separate the data into three parts; all PPs, independent PPs, and attributive PPs. For each sub-

9 The formula used is the following: Position~(cumulative surprisal + mean surprisal + length + genre + period)^2.

corpus, we have checked whether a transformation of the variables is necessary. When all PPs are regarded, no transformations are needed. In the subcorpus consisting of independent PPs, we log-transformed length; in the subcorpus of attributive PPs, the mean skipgram values were also log-transformed to fit the model better.[10] To determine which model explains the data best, we used backward model selection (Gries 2021). In this procedure, one interaction and one independent predictor, respectively, are excluded one after the other (Gries 2021: 366). We started with the interactions and excluded there first those with the highest non-significant p-value (Gries 2021: 366). To find out whether the exclusion of a predictor or an interaction actually led to an improvement of the model, a likelihood ratio test with the anova function is performed.[11] The figures displayed here are created with ggplot 2 (Wickham 2016).

4.1 Results of all PPs

In total, we annotated 726 PPs. Among them were 414 extraposed and 312 embedded ones. Divided by genre, we found 239 extraposed PPs in the medical texts and 175 extraposed PPs in the theological articles and monographs. The differences between the two genres are relatively small, though our findings contradict the notion that extraposition might be more common in genres that should resemble oral more than written discourse mode, that is in our theological part of the corpus. When we divide the data into different periods, we find the following distribution (see table 5).

Period	**All Texts**	**Medical Texts**	**Theological Texts**	**embedded and extraposed PPs**
1650-1700	168	154	14	298
1700-1750	102	4	83	183
1750-1800	22	10	12	37
1800-1850	70	35	35	122
1850-1900	52	21	31	87

Table 5 Number of extraposed PPs in different periods and by genre. The last column contains the number of all annotated PPs per period, including embedded PPs.

10 The base of the logarithmic function is 10 in all cases.

11 If the p-value of the anova is not significant, the more complex model does not explain the data significantly better. This process is repeated until only significant effects or main effects involved in a significant interaction are included in the model.

Table 5 shows that PP extraposition is more common in older periods of German. From 1650 to 1750, we find over 100 extraposed PPs. This number is never even approached in the later periods, especially not in the period from 1750 to 1800. Here we have only been able to annotate 37 PPs in total, which cannot be attributed to a decrease in test data (see table 4). This suggests a preference for embedding in this period. In the 19th century, however, more phrases are extraposed, though their number decreases towards the end of this decade. It is also interesting to note the large difference between medical and theological texts from 1650 to 1750. In the first period, only 14 phrases were found in theological texts, but 154 in medical ones. This might be attributed to the author whose text we used for the corpus because we could only include one medical writer in this period. The other explanation would be that this period is orientated towards a more pragmatically influenced word order. However, to make a definite statement about the latter, more medical texts must be analysed, which will not be done in this paper.

When the whole data included in this paper is examined, the cumulative surprisal values range from 3.22 to 468.44, with a mean of 30.56 and a median of 19.37, a standard error of 1.3, and a standard deviation of 35.25. The maximum cumulative surprisal value belongs to a phrase that is very complex in its attribution of material to the PP which is in itself complex and displays a high amount of other embedded clauses. This PP corresponds to the PP with the greatest length (133). The length of the PPs starts at 2 and displays a mean length of 8.65 and a median of 5, a standard error of 0.37, and a standard deviation of 9.92. The mean surprisal values vary between 2.89 and 4.39, with a mean of 3.7 and a median of 3.69 and a standard error of 0.008, and a standard deviation of 0.24 (see table 6).

	Cum. Surprisal	**Mean Surprisal**	**Length**
Minimum	6.42 (3.22)	3.21 (2.89)	2 (2)
Mean	39.6 (18.61)	3.7(3.69)	11.24 (5.24)
Median	27.09 (14.36)	3.69 (3.96)	8 (4)
Maximum	468.44 (95.54)	4.39 (4.38)	133 (28)
Standard Error	1.31	0.008	0.37
Standard Deviation	35.25	0.023	9.92

Table 6 Cumulative and mean Surprisal values and Length of all extraposed PPs. The values in brackets are the results for the embedded PPs. The standard error and the standard deviation have not been distinguished by embedded or extraposed PPs.

	Estimae	Standard Error	z-value	p-value
cum. surprisal	0.04	0.03	1.07	0.29
length	-0.33	0.13	-2.66	<0.01 **
Genre	0.86	0.37	2.36	<0.05 *
Period	0.01	0.06	0.14	0.89216
cum. surprisal & Length	0.0004	0.0001	2.94	<0.01 **
Genre & Period	-0.35	0.13	-2.78	<0.01 **

Table 7 Logistic Regression all PPs.

We applied logistic regression to all PPs regardless of the subperiod of our corpus they belong to and determined the best fitting model with a backward model selection with anova (R Core Team 2018). Due to the model selection the main predictor "mean surprisal" and several interactions have been eliminated in the final model which is presented in table 7.

This final model shows that the surprisal measures did not have any significant influence on extraposition. Only in interaction with length, does it show a significant influence, which is in contrast to the hypothesis; an increase in length and surprisal increases the likelihood of embedding according to the model. Independent of the regression model, we calculated the correlation between length and cumulative surprisal and found that they are highly correlated ($r = 0.99$). Thus, it must be assumed that this interaction is not the most reliable one to predict extraposition and that the high informativeness of the extraposed phrase is due to its length. Here, the most significant predictor for extraposition is length. The longer a PP, the more likely it is positioned in the postfield. This is in line with previous research on this topic (Weber 2019, Wasow 1997). The other significant main effect is found for the genre. Theological texts seem to favor embedding, which does not mirror the interaction between genre and period. Younger, theological texts favour extraposition.

There are two things that need to be taken into account when regarding these results. First, the different periods display a strong variation in the number of extraposed PPs found (see figure 1). The different number of PPs makes it worthwhile to look at the periods separately. Thus, the following section (4.1.1) will show the variation of influential factors per period. The methods used do not change.

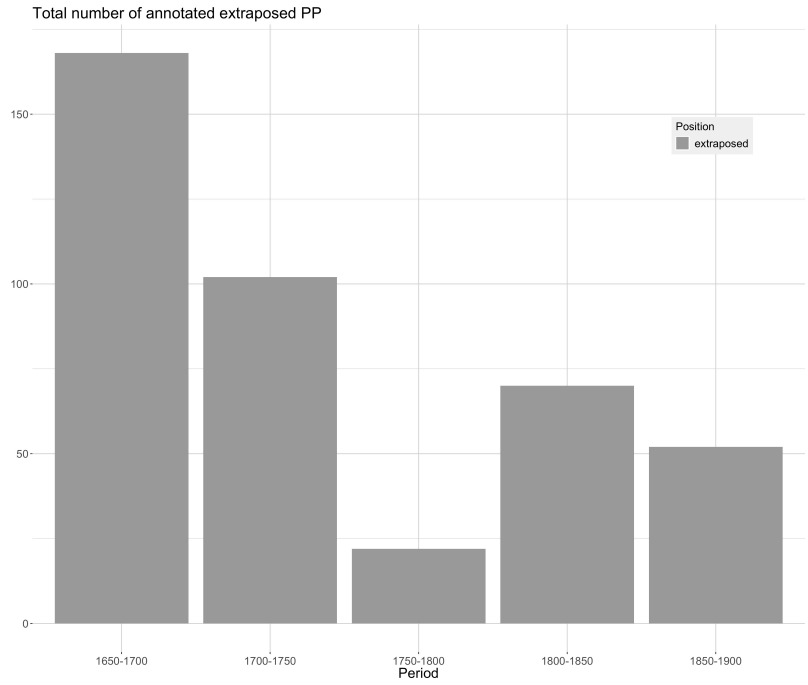

Figure 1 Distribution of extraposed and embedded PPs in 50-year-periods.

Second, the analysed data does not differentiate between independent and attributive PPs. To distinguish between them might, however, be a good idea as Voigtmann & Speyer (2021) show that the position of *attributive* relative clauses is strongly connected to high surprisal values. The attributive characteristics might result in a different analysis of extraposed material. This assumption is tested in section 4.2.

4.1.1 Analysis of PP extraposition per period

The following section is concerned with the analysis of PP extraposition in the various periods of the corpus. The analysis of the whole data has shown an influence of time on extraposition. However, it was not able to show changes within the different periods. To fill this gap is the purpose of this section. For the analysis, we use logistic regression and backward model selection (glm, base-package, R Core Team 2018). Every model contains the variables cumulative skipgram surprisal, mean skipgram surprisal, length, and genre (sum-coded) as well as every two-way interaction between these variables.

	Estimate	Standard Error	z-value	p-value
cum. surprisal	-0.08	0.01	-5.72	<0.001 ***
genre	0.308	0.47	0.65	0.51

Table 8 Logistic Regression of PPs from 1650 to 1700.

Since there is only one period under investigation for each model, the variable "period" is excluded from the analysis. It is indicated in the different sections whether any transformations for the variables have been done.

PP extraposition in 1650 to 1700 For the first time span under investigation, the best-suited model includes cumulative surprisal and genre. We do not have to log-transform any variables. The results differ enormously from the model of the whole data. Here, cumulative surprisal values present a highly significant value ($p<0.001$). Length is not significant for extraposition at all and has been removed from the model during the backward model selection. The same holds for all interactions (see table 8).

PP extraposition in 1700 to 1750 Though the figure 2 seems to display a similar result for the following periods - we can see higher surprisal values in the extraposed column and lower values for the embedded PPs - the logistic regression contradicts this observation and reflects the observations made by regarding the figure which displays the mean surprisal values (figure 3), where the picture is no longer as clear. We do not have to apply any transfigurations to the variables.

However, the model that is claimed the most suitable by anova does not show any significant predictors for extraposition in this period (table 9). It is interesting to note that the direction of the mean skipgram surprisal and the length point differs from the one for cumulative surprisal, meaning that the model suggests that embedding might be preferred for variables with great length and/or high mean surprisal values. Nevertheless, the values are not significant, so we cannot make any implications for this period.

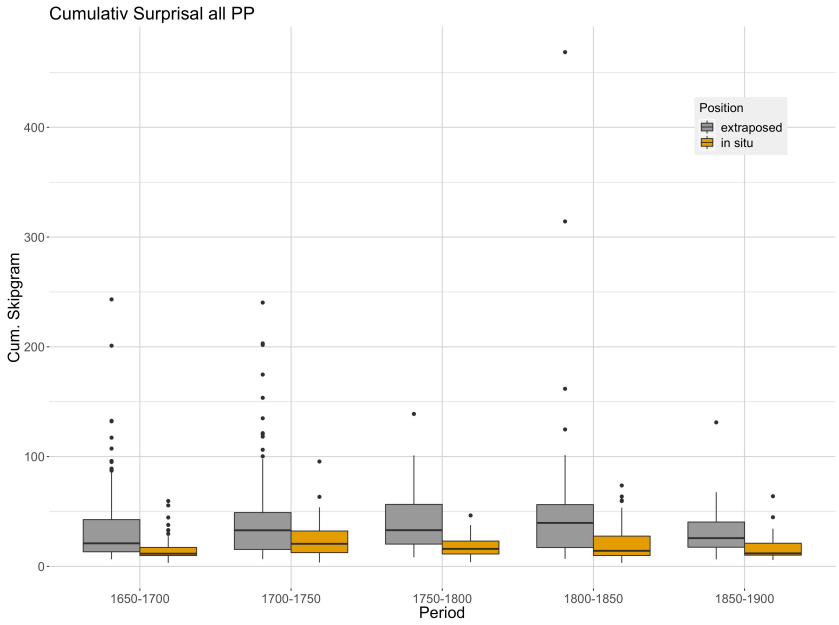

Figure 2 Cumulative Surprisal values PP.

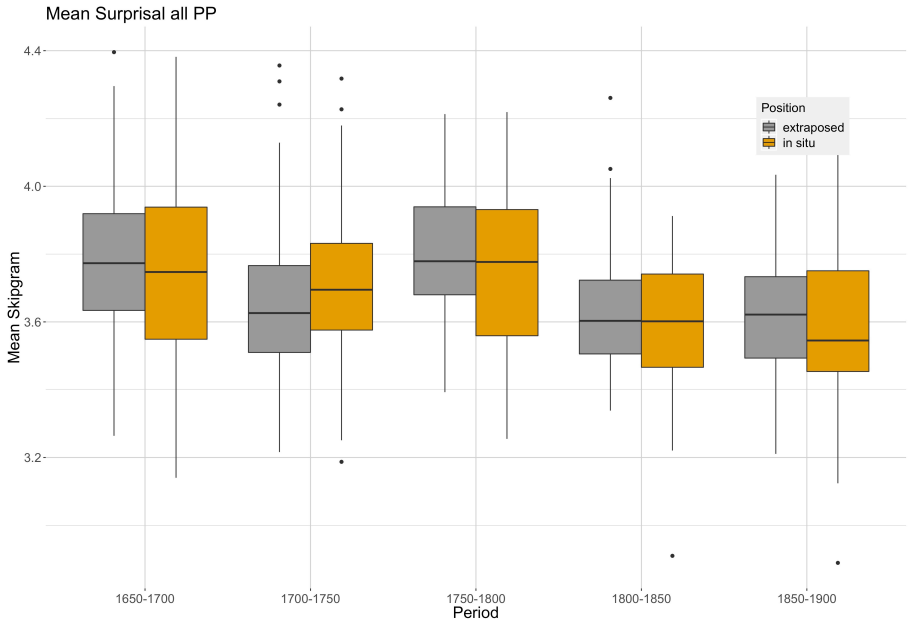

Figure 3 Mean Surprisal values of PPs.

	Estimate	Standard Error	z-value	p-value
cum. surprisal	-2.02	1.86	-1.08	0.28
mean surprisal	1.24	0.92	1.35	0.18
length	6.28	5.79	1.09	0.28
genre	-7.55	8.20	-0.92	0.36
cum. surprisal & genre	4.15	3.72	1.12	0.26
length & genre	-13.19	11.58	-1.14	0.25

Table 9 Logistic Regression of PPs from 1700 to 1750.

PP extraposition in 1750 to 1800 The results for the next periods change a lot compared to the period from 1700 to 1750. They are only similar in the fact that no transformations of any variables were necessary. The best-fitting model contains the variables mean skipgram surprisal, length, and the interaction between the two (table 10). Both length and mean skipgram surprisal are significant for extraposition ($p<0.05$); their interaction, however, is significant for embedding. It must be remembered that the results are gained by considering only 37 PPs, so they must be interpreted with care. It remains interesting, however, that not only length but also the mean surprisal values indicate a higher likelihood for extraposition.

	Estimate	Standard Error	z-value	p-value
mean surprisal	-12.36	6.06	-2.04	<0.05 *
length	-7.03	3.39	-2.08	<0.05 *
mean surprisal & length	1.75	0.87	2.01	<0.05 *

Table 10 Logistic Regression of PPs from 1750 to 1800.

PP extraposition in 1800 to 1850 In this period, we find a reliable number (122) of extraposed PPs and their respective counterparts, so we can make more certain statements about the regression. No transformations of any values have been necessary. The backward model selection leaves a model that includes length, genre, and the interaction between the two (table 11). In contrast to the last period, both mean and cumulative skipgram surprisal have been excluded from the model. Their influence has apparently decreased in this century. Length, however, is highly significant for extraposition ($p<0.001$).

The longer a PP, the more likely it is extraposed. The interaction between length and genre is also significant for extraposition which becomes more likely for longer phrases and theological texts ($p<0.05$).

	Estimate	Standard Error	z-value	p-value
length	-0.29	0.08	-3.53	<0.001 ***
genre	0.63	0.91	0.70	0.48
length & genre	-0.33	0.17	-1.99	<0.05 *

Table 11 Logistic Regression of PPs from 1800 to 1850.

PP extraposition in 1850 to 1900 In the last period under investigation, the main results do not change either. First, we do not achieve a better fit of the model by any transformations, second, the surprisal variables have been omitted in the process of the backward model selection. The final model includes length and genre (table 12). Length is highly significant for extraposition ($p<0.001$), and genre does not present a significant result. As in the paragraph on the corpus from 1800 to 1850, longer PPs tend to be extraposed without regard for their informativeness.

	Estimate	Standard Error	z-value	p-value
genre	-0.35	0.49	-0.71	0.48
length	-0.23	0.08	-3.05	<0.01 ***

Table 12 Logistic Regression of PPs from 1850 to 1900.

Summary To sum up these results, we see that information density measures are significant in the first period (1650–1700). This aligns itself with our hypothesis and shows that high surprisal is correlated to extraposition. The second time span (1700–1750), on the other hand, shows no influence of any of our variables on the position of the PP, probably signaling that the variables influence PP placement equally or that other predictors which have not been considered and which do not appear in the literature on that topic (see 2.1) are more influential. This changes in the next period (1750–1800). Here both mean surprisal and length are significant for PP extraposition. This marks the transition point because the surprisal values do not influence PP extraposition in the 19th century any more. In both periods, mean and cumulative surprisal have been omitted in the model and both final models show a

Period	All Texts	Medical Texts	Theological Texts
1650-1700	165	151	14
1700-1750	99	19	80
1750-1800	22	10	12
1800-1850	63	30	33
1850-1900	44	17	27

Table 13 Number of independent PPs, embedded and extraposed, in different periods and by genre.

highly significant result of length on extraposition. Overall, length is the best predictor since the 18th century.

As already mentioned, these results concern all PP and do not make a differentiation between attributive and independent PPs. To distinguish between them might, however, be a fruitful idea because in Voigtmann & Speyer (2021) it was shown that the position of attributive relative clauses is strongly connected to high surprisal values. The following section looks at the change in the results when attributive PPs and independent PPs are considered separately.

4.2 Results of attributive and independent PPs

Independent PPs The distinction between independent and attributive PPs showed that the majority of PPs are independent. 393 PPs are extraposed, 296s embedded. The number of PPs in medical texts (405) exceeds those in theological texts (284). There are also strong differences between the periods, which resemble those of all PPs. The following table (13) shows the distribution between periods and genres. It can, thus, be said that there is no period in which the extraposition of attributive PP would be more common than the extraposition of independent PPs. Furthermore, it is interesting that there are no attributive PPs in theological texts from 1650 to 1700 and non at all in the period from 1750 to 1800. Since only 37 PPs in total have been annotated as attributive, we do not expect the results of the regression analysis to differ from those of the analysis of all PP.

Before we present the regression analysis, we will give some descriptive statistics: The mean of the cumulative skipgram values is 30.68, and the median 19.37. Both do not vary significantly from the values presented for all PP. The standard error of the cumulative surprisal is 1.43 and the standard devi-

	Estimate	Standard Error	z-value	p-value
intercept	2.23	0.38	6.00	<0.001 ***
length (log)	-3.31	0.46	-7.25	<0.001 ***
genre	0.68	0.42	1.59	0.1111
period	0.01	0.08	0.12	0.91
genre & period	-0.32	0.15	-2.07	<0.05 *

Table 14 Logistic Regression of independent PPs.

ation 1.37. The same holds for the mean of the mean skipgram value, which is 3.7, and the median (3.69). The standard error, and the standard deviation are both 0.009. For length, we find a mean of 8.69 tokens and median of 5, a standard error of 0.38, and a standard deviation of 10.11. For all numeric values, the difference in the measures of both independent and attributive PPs is very small.

However, for the regression analysis, we had to log-transform length to have a better fit of the model. Again, we performed a general linear regression with R (R Core Team 2018), which included the variables cumulative skipgram, mean skipgram, length (log-transformed), genre (sum-coded as in the previous analyses), period (transformed to ordinal values) and the two-way-interactions between those variables, and used the backward model with anova (R Core Team 2018) to determine which interactions and possibly main effects can be omitted. The final model is shown in table 14.

As expected, the results do not differ much from the analysis of all PPs, presented in section 4.1. However, it becomes even clearer that surprisal measures don't seem to have an effect on the placement of PPs. Among the main effects, only length is highly significant for extraposition ($p<0.001$). Neither period nor genre is significant for the placement of PPs on their own, but their interaction is significant for extraposition ($p<0.05$) which means that the likelihood of extraposition is increased in the case the PP is found in a theological and younger text. The difference to the model with all PPs is that the interaction between length and cumulative surprisal was not significant and has been omitted. Due to that, cumulative surprisal itself has been omitted without impairing the model.

Attributive PPs In the whole corpus, we have found only 21 attributive, extraposed PPs. They are compared to 16 embedded PPs with matching prepositions, which results in a total of 37 PPs to be analysed here. This small num-

ber makes it hard to generalize the results, but this small amount is interesting in itself. The literature presented in section 2.1 does not differentiate between attributive and independent PPs and, therefore, claims that the same principles hold for their placement in the postfield. Thus, we also include the same variables in our analysis, but we will see a slightly different result here.

First, we will again describe the data. In the last section, we have already seen that there are no attributive PPs in the period 1750–1800, so the PPs must be distributed over the other periods. From 1650 to 1700 and from 1700 to 1750 three extraposed PPs have been found. This number doubles from 1800 to 1850 where 7 PPs have been annotated and increase once more from 1850 to 1900. Here we annotated 8 PPs. Note that the number of extraposed attributive PPs increases while the number of all PPs which are extraposed does not show this tendency and rather decreases.

Dividing the data into medical and theological texts, we find more PPs in the medical texts (12) than in theological ones (9). Because of the small sample, we have not divided these numbers by period.

The sum skipgrams of both embedded and extraposed PPs have a mean of 28.65, a median of 19.48, a standard error of 3.29, and a standard deviation of 20.06. The mean skipgram values have a mean of 3.59, a median of 3.6, a standard error of 0.029, and a standard deviation of 0.17. The mean of the length is 8.08, the median is 6, the standard error 0.88, and the standard deviation 5.38. Though the distribution of cumulative surprisal values according to the PP position has seemed clear in figure 2, the differences become even more obvious when only attributive PPs are concerned as in figure 4.

Except for the period, 1700 to 1750 the grey-coloured extraposed PPs have a much higher cumulative surprisal value than their embedded counterparts. This is partially reflected in the logistic regression analysis. The proceeding is the same as for all PPs and for the independent PPs. We had to log-transform the mean skipgram values to fit the model better, applied sum-coding for the genre, and changed the periods to ordinal values. However, the following model, presented in table 15, must still be handled with care because of the very small amount of data.

None of the predictor variables is significant, but there are several effects with significant results ($p<0.1$). Cumulative Skipgram values could not be excluded from this model and are significant for extraposition whereas length is significant for embedding, which is also the case for the period while the likelihood of extraposition increases in theological texts.

Several interactions are also significant. An increase of the cumulative surprisal value and the mean surprisal is connected to embedding, as is an increase in the mean surprisal value while having a theological text or a longer

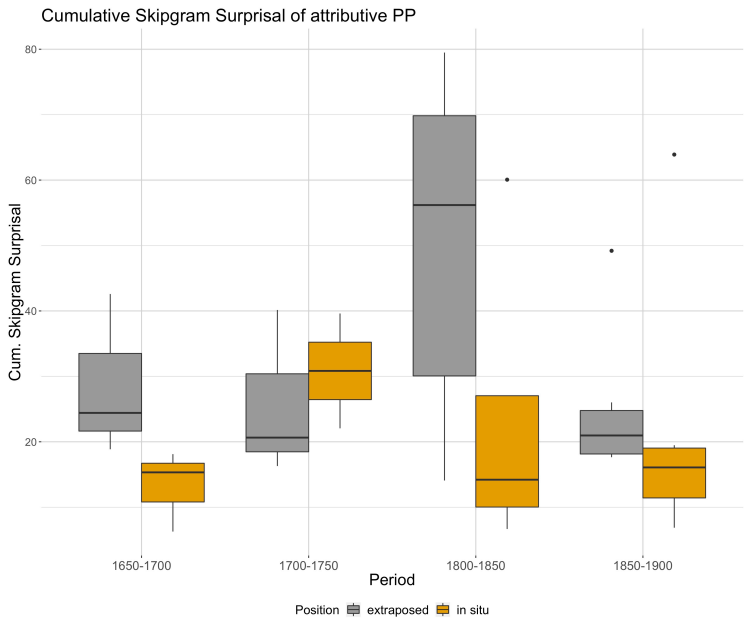

Figure 4 Cumulative Surprisal values of attributive PPs.

	Estimate	Standard Error	z-value	p-value
cum. skipgram	-35.42	19.32	-1.83	<0.1 .
mean skipgram (log)	40.88	124.25	0.33	0.74
length	100.335	60.37	1.66	<0.1 .
genre	-320.883	177.26	-1.81	<0.1 .
period	31.041	17.29	1.80	<0.1 .
cum. skipgram & mean skipgram	57.87	32.42	1.79	<0.1 .
cum skipgram & genre	-10.34	5.93	-1.74	<0.1 .
mean skipgram & length	-160.45	102.47	-1.57	0.12
mean skipgram & genre	570.39	314.92	1.81	<0.1 .
mean skipgram & period	-55.99	31.16	-1.80	<0.1 .
length & genre	38.08	21.90	1.74	<0.1 .

Table 15 Logistic Regression of attributive PPs.

PP that is simultaneously from a theological text. Extraposition on the other hand is favoured by an interaction of cumulative surprisal and genre, making extraposition more likely for high cumulative surprisal values and theological texts, and mean surprisal value and period increasing the likelihood of extraposition for high mean surprisal values and younger texts.

Regarding these results, the model is highly contradictory, especially for results like the favoured embedding in younger texts though there have been more extraposed PPs than in earlier centuries. However, the small data sample size must be remembered. What is noteworthy in this model is the fact that cumulative surprisal values show a significant result and that we could not exclude them from our model.

This strongly hints towards a different reason for extraposition than we have found for the independent PPs.

4.3 Discussion

In our corpus, we have found PP extraposition in every time period under investigation. We observe changes in frequencies and in influential factors on extraposition over time. Furthermore, there is also a change in the kind of PP that is extraposed; independent PPs are placed 17 times more often in the postfield than attributive PPs.

All PPs Considering all PPs independent of their status as attributive or independent and of their period, we find that length is the most influential factor for extraposition. The same holds for the interaction between genre and period indicating that PPs in younger and theological texts are more likely to be extraposed. We do not find an influence of our surprisal measures except in an interaction that is likely to be based on the strong correlation between cumulative skipgram surprisal and length. Though we argued in section 3.3 that surprisal might be independent of length, our data suggests the opposite; longer phrases can contain more information. The influence of length is, basically, in line with previous work on PP extraposition (see section 2.1). Our data does not allow the conclusion that high lexical information is relevant for the position of PPs, but their length is. This weight effect (Wasow 1997) still facilitates processing and production, and can, thereby, be linked to information density even if the claim cannot be made as strong as it would have been when surprisal measures would have been significant. This effect is created because extraposing a PP shortens the distance between the sentence brackets. In most cases in German, the lexical verb only appears in the RSB as the left sentence bracket is either filled with subordinate conjunctions

or with an auxiliary verb. Thus, especially listeners need many cognitive resources to find out which phrases are requested by the lexical verb and have to appear and which are just accessories following ideas of linguistic valency. Levy (2008) has argued that receivers keep possible lexical and syntactic continuations of the sentence in mind. This leads to high entropy at the beginning of sentences which is reduced over the course of the clause as choices limit themselves. This effect will be even stronger for German VL-clauses because until the RSB is processed there are more choices and, therefore, higher entropy which leads to more resources being consumed and needed. Presenting the RSB earlier leads to facilitated processing. Extraposing long and in many cases complex PPs leads to this result and can explain why we find a significant result for length overall PPs. We will return to this point at the end of section 5.

Influence of Genre In the model of all PPs, the genre has yielded a significant result as well. This might hint at an explanation we cannot pursue in this paper, namely the closeness to orality. Former research (e.g. Degaetano-Ortlieb & Teich 2018) has shown that there is a difference between natural scientific writing and writing in the field of humanities as the latter is closer to oral discourse mode (Koch & Österreicher 2007). Closeness to orality is linked to more extraposition which can explain why theological texts have a higher tendency to extraposition. However, to fully determine the influence of orality one must capture the grade of orality of a text. This will be accomplished by future research.

PP extraposition over time When we do not look at attributive PPs separately, we find a decreasing influence of information density on extraposition. In the earliest period, cumulative skipgram surprisal is the only influential factor for extraposition. Then, our model cannot detect any influential variables for PP placement from 1700 to 1750, though the mean skipgram surprisal value indicates a (non-significant) preference for embedding. In the second half of the 18th century, both length and mean surprisal are linked to extraposition. This is the last period in which surprisal measures must even be considered in the best-suited model of the linear regression analysis. In the 19th century length is the most important predictor for extraposition.

So, apparently, there is a change from the 18th to 19th century which might be linked to the consolidation of the sentence frame. We have argued for its importance for processing above. Though the sentence frame was established in the Early New High German period from 1350 to 1650 (e.g. Admoni 1990), there is still research about the violation of this frame in the early Modern

German period which is considered here. Thus, the sentence frame gains importance over time, and, consequently, keeping the distance between the two brackets becomes crucial the stronger the sentence frame is implemented. That is in line with our findings that a PP must not necessarily be lexically challenging to be extraposed in general.

However, for the earlier periods (1650–1800) we find - except for the period 1700 to 1750 - that it is not necessarily enough for a phrase to be long to be placed in the Postfield. In the first 150 years of our corpus, PP extraposition is triggered by high surprisal values. This can be explained as follows:

Especially for the 1650 to 1700 period, an influence of the so-called "Kanzleistil" (chancery style) with very long middle fields must be assumed. This stylistic choice was highly prestigious though also questioned by scholars at the time (Konopka 1996, Takada 1998). It is possible that the authors were familiar with this style which should in fact contradict good processing based on lexical choices. Being used to certain grammatical constructions and seeing them frequently has recently been discovered as a major factor for people's ability to deal with constructions that contradict processing principles like dependency length (Futrell et al. 2021). Futrell et al. (2021) found this for example for the sentence frame in German in contrast to English where such long distances between grammatically related elements are considered nearly ungrammatical. Consequently, the authors of the periods where surprisal plays a role in extraposition need more than just long PPs to trigger extraposition because they are used to long middle fields in written texts. Thus, they might not find them hard to understand. However, when the lexical verb has not yet occurred, uncertainty remains and processing might be impeded largely when readers have to cope with this uncertainty and highly surprising phrases. Though the increased importance of length can indicate a general language change away away from overly large middle fields and might show a more sophisticated writing style, for these earlier periods the high surprisal values function as a kind of additional catalyst for extraposition.

Independent and attributive PPs This is also the distinction between independent and attributive PPs. We have already stated above that there are 17 times more independent PPs than attributive ones. Thus, attributive PPs seem to need an additional trigger to be extraposed compared to independent PPs. The regression analysis of the independent PPs does not suggest an interpretation different from the one given for all PPs; length is still the most significant predictor for extraposition while surprisal does not have an effect.

This is different for attributive PPs. In contrast to independent PPs which have a certain freedom in placement (e.g. Lenerz 1977), the position of attributive PPs is set. They must stand adjacent to their head phrase. Placing them in the postfield disrupts the alignment of head phrase and attribute and should be harder to process in any case. However, if their surprisal values are very high and are, consequently, putting a high strain on the working memory, the disadvantage of splitting the phrase can be overwritten by better processing of the whole clause by extraposing attributive PPs. Voigtmann & Speyer (2021) have found similar evidence for the extraposition of attributive relative clauses. Their extraposition is strongly influenced by surprisal as relative clauses with higher surprisal values are more likely to be placed in the postfield. The same seems to hold for attributive PPs. Though the results must be regarded with caution due to the small data set, we find evidence for a link between high surprisal values and extraposition for these PPs.

These similarities will be also considered in the now-following general approach to the topic of PP extraposition (see section 5).

5 Generative Approach

After having considered information theoretic constraints on relative clauses in the postfield, we will now investigate whether these constraints lead to consequences for the analysis of the German postfield. There have been brought forward several approaches to analyzing the postfield:

- A. Rightward extraposition: Material is moved to a position that is adjoined rightwards to the verb phrase or a higher projection in the functional overlay of the verb phrase, thus forming the postfield (see e.g. von Stechow & Sternefeld 1988, Müller 1995, Büring & Hartmann 1997). In order to make sure that the landing site controls the launching site of the movement, it should be adjunction to some functional projection in the I- or C-complex (see figure 5).
- B. Variable base: The verb phrase is not obligatorily right-headed in German, but can switch to left-headedness (e.g. Haider 2010). The complement of a left-headed verb phrase appears as postfield in these cases.
- C. Post-syntactic phenomenon: The postfield is not part of the generation in syntax proper but is formed only after spell-out at PF (e.g. see figure 5b, and Truckenbrodt 1995).
- D. Mixed approach: Not all postfields are the outcome of the same structural configuration, but there are several mechanisms leading to

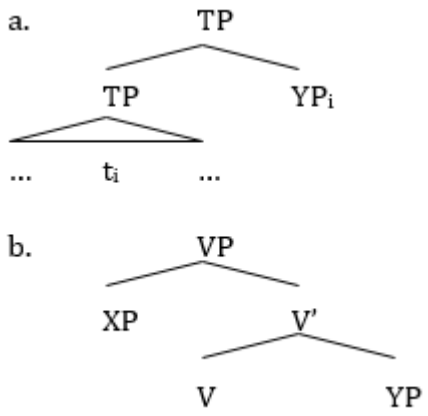

Figure 5 Approach to the Postfield.

a postfield. Inaba (2007), concentrating on clauses in the postfield, distinguishes between the postfield as the result of extraposition (A) for non-clausal postfields, as the outcome of a left-headed verb phrase (B) for subordinated clauses that are not in attributive function to a noun phrase in the matrix clause and as a post-syntactic phenomenon (C) in the case of adjunct clauses. Frey (2015) basically adopts this approach, but does away with the clausal/non-clausal distinction. If a phrase YP that has the property [+ root], meaning that it is directly a complement or adjunct of VP, appears in the postfield, it is an instance of a left-headed VP. If a phrase that has the property [- root], meaning that it is a complement or adjunct to some XP which is a complement or adjunct of VP, appears in the postfield, this is the result of a post-syntactic process. The property [-root] would translate in more traditional terms to being an attribute.

The variable base approach has its advantages in explaining why complement clauses are ungrammatical left of the verb (example (7)).

(7) (a) *Gwendolyn hat gewusst, dass Uller sich darüber aufregen würde.*
Gwendolyn has known that Uller *reflex. pron.* about.that upset.be would.

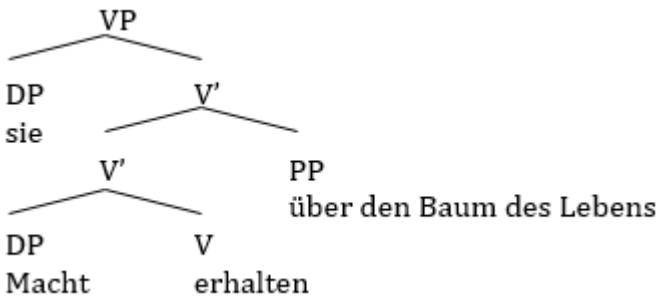

Figure 6 Possible structure of example (8).

(b) ***Gwendolyn hat, dass Uller sich darüber*
Gwendolyn has that Uller *reflex. pron.* about.that
aufregen würde, gewusst.
upset.be would known.
Gwendolyn has known that Uller would be upset about that.

A problem is, however, that it is unclear how this analysis accommodates adjuncts. If we look at non-clausal postfield-filling, it is most often adjuncts that appear there. We might now investigate whether there is some evidence that might give hints as to which approach is to be preferred. We concentrate on attributive material, as this is the main issue in this paper.

Let us begin with the variable-base approach B. In cases in which only attributes (Attr(NP)) to some noun phrase are in the postfield it is not possible to implement the variable-base approach in its pure form, as a sequence of the form NP_i – V – Attr(NP_i) presupposes a right-headed VP if the NP_i is the complement of V. Still, the verb phrase could be of a form like figure 6, illustrated with the original example from the corpus given in example (8).

(8) *auf daß sie Macht erhalten über den Baum des Lebens*
on that they power receive over the tree of-the life
'so they receive power over the tree of life' (Daumer 1859)

But this is excluded on independent grounds. Phrases for which [- root] holds show dependency which is visible in e.g. case assignment or coreference relations. Non-clausal attributes of noun phrases or adjective phrases are assigned the morphosyntactic form (e.g. being a PP headed a particular prepo-

sition, (9)a,b) or abstract genitive case by their nominal or adjectival head ((9)c,d).

(9) (a) [*NP Macht* [*PP über/*auf etc./*∅ den Baum des Lebens*]]
[NP power [PP over/*on etc./*∅ the tree of life]]
"Power over the tree of life"

(b) [*AP* [*PP auf/*über etc./*∅ den Enkel*] *stolz*]
[AP [PP of/*over etc./*∅ the grandchild] proud]
"proud of the grandchild"

(c) [*NP Absperrung* [*DP des Weges*]]
[NP barrier [DP of.the way]]
"barrier of the way"

(d) [*AP* [*DP des Englischen*] *mächtig*]
[AP [DP in English] fluent]
"fluent in English"

With relative clauses, there is a coreference relation between the antecedent noun and the relative pronoun in relative clauses which are introduced by a relative pronoun.

The assignment of case or the general morphosyntactic shape of linguistic elements is usually considered to happen via c-command. For this, the element that receives case should be in a position in which it can be c-commanded by the governing element. If the postfield is however a position higher in the tree – and it does not play a role whether it is adjoined within the VP, as in (8), or a position adjoined to the I- or C-architecture –, there is no c-command relation between the governing noun and its logical complement in the postfield, if the noun is at the complement position of VP. This can be demonstrated on the original example (8): The morphosyntactic shape of *über den Baum des Lebens* ('over the tree of life'), which is the logical complement of *Macht* ('power'), is idiosyncratic to this lexeme. This means that it cannot be assigned by default but must be assigned by c-command from the nominal head *Macht*. This is however impossible if it is base generated at some high position like for instance as an adjunct to vP or TP (Culicover & Rochemont 1990). *Macht* is the complement of VP and therefore lower in the tree (Frey 2015: 73). This indicates that attributes, even non-clausal ones, cannot be base-generated at some position outside the noun phrase they are related to. So analysis (B) is discarded. What about analysis A, the extraposition approach? This approach faces several problems that are discussed e.g. in Frey (2015). One exemplary counterargument against the movement of postfield

phrases, in general, is the following (examples and argumentation from Frey 2015: 56f.): Consider the sentences in (10).[12]

(10) (a) *Max hat heute t_1 t_2 gesprochen [mit wem]$_1$ [über was]$_2$*
Max has today spoken with someone about something
"Max talked today with someone about something."

(b) *?? Max hat heute t_1 t_2 gesprochen [über was]$_2$ [mit wem]$_1$*
Max has today spoken about something with someone
"Max talked today with someone about something."

(c) *Max hat heute t_1 t_2 gesprochen [mit Maria]$_1$ [über den Euro]$_2$*
Max has today spoken with Maria about the Euro.
"Max talked today with Mary about the Euro."

(d) *Max hat heute t_1 t_2 gesprochen [über den Euro]$_2$ [mit Maria]$_1$*
Max has today spoken about the Euro to Maria.
"Max talked today with Mary about the Euro."

The judgments in (10a,b) indicate that with [+wh]-expressions, the serialization in the postfield must follow the base order in the middlefield. This ban on crossing dependencies is however not visible with [-wh]-phrases, as (10c,d) shows. If it is a matter of extraposition, we should expect that it does not matter whether the phrase to be extraposed is a [+wh]- or a [-wh]-phrase. The postfield is not a position that is marked with respect to the [wh]-feature.

So, the conclusion, which Frey (2015) also draws, is that attributes in the postfield must be a postsyntactic phenomenon. We can conclude this however not only by excluding the other possibilities, but also because there is evidence that attributes in the postfield are sensitive to prosody. So an intervening Phonological Phrase between the attribute and the noun phrase it is related to leads to compromised acceptability (example (11), Frey 2015: 70).

12 The wh-phrases in example (10) have an indefinite reading and are consequently supposed to be inert with respect to information-structurally-induced movement. They sound more natural if *wem* is replaced by *jemandem* and *was* by *etwas*. The acceptability distance is relatively weak, but noticeable.

(11) (a) *?? Maria hat* [Φ *dem Kollegen*$_1$] [Φ *die Räume*]
Mary has [the colleague]$_{dat}$ [the rooms]$_{acc}$
gezeigt [Φ *mit den graumelierten Haaren*]$_1$
shown with the greying hair

(b) *Maria hat* [Φ *die Räume*] [Φ *dem Kollegen*$_1$
Mary has [the rooms]$_{acc}$ [the colleague]$_{dat}$
gezeigt] [Φ *mit den graumelierten Haaren*]$_1$
shown with the greying hair
"Mary showed the rooms to the colleague with the greying hair"

This account held also for the older stages of German which we investigated. In all examples for attributes in the postfield, there was only the infinite verb form intervening. In one case (12) particles and other obligatorily unstressed elements were intervening as well, but they, being unable to bear stress, do not form Phonological Phrases on their own.

(12) [Φ *einige Unterscheide*]$_1$ *kan man noch machen*] / [Φ
some difference can one still make
von dem Orte / wo der Bruch geschehen]$_1$
regarding the place where the fracture happened
"Some differences can still be made regarding the place where the fracture occurred." (Purmann 1680)

If the attributes were extraposed within syntax proper, we would assume that they are freely interchangeable just like the [-wh]-phrases in (10c,d).

There are other phenomena that indicate that attributes in the postfield are not extraposed within syntax proper. Frey (2015) discusses them at length and convincingly, so we do not repeat them here. What is more, at issue for the aims of the present paper is in what ways the analysis pursued here (attributes in the postfield as a postsyntactic phenomenon) relates to questions of information density. We have seen in the previous sections that two types of clearly [-root] or attributive material, relative clauses, and attributive PPs, share sensitivity to information density. This might distinguish them from non-attributive PPs. For answering that question, we first have to ask another question: At which point of the syntactic derivation is information density relevant at all? As information density is a measurement that gives an approximation to real processing cost in language generation or comprehension, we might assume that it plays a role quite early in the derivation. The point at which it is relevant is when the numeration is formed, that is, when referents and predicates are retrieved from the mental lexicon (in the case

of information that is new in the discourse) or the common ground, that is, information already in the buffer (in the case of given information). Pieces of information, be they predicates or referents, came with a 'price tag', so to speak, which can be approximated via information density measurements. If there are many high-density elements in the numeration, this may overtax the working memory capacity of the producer. So the point when the processing costs of the elements to be expressed play a role is at the very beginning of the derivation. For the following building of the syntactic structure it is irrelevant what lexical processing cost is associated with the elements undergoing syntactic derivation. It gains relevance again only at PF, when Lexical Insertion (Harley & Noyer 1999) takes place, as here the loop back to the lexicon is made in order to associate the abstract (semantic) feature bundles with the word forms deposited in the mental lexicon. Sheer length, on the other hand, is irrelevant for the items in the numeration, as long as the mere number of elements in the numeration does not overtax the working memory capacity. The length or complexity is not at issue in phrase generation. This is a recursive process that could theoretically apply ad infinitum. Lengthy constituents are avoided rather for prosodic reasons (cf. Féry 2015). So length plays a role again only at PF, but after Lexical Insertion has taken place.

So it appears as if PF in itself is to be thought of as a procedural module. Narrow Syntax feeds via Spell-out input into PF. On PF, relatively early Lexical Insertion takes place. In a second step, movement operations apply in order to accommodate the information to optimal information packaging. Subsequently, Prosody is assigned and if prosodic assignment is not possible (e.g. because the unit to which a prosodic constituent should be assigned is too long), we may assume that further operations apply in order to accommodate the utterance to optimal prosodic well-formedness.

Another aspect related to information density is the following: We have seen (section 4) that extraposing material has a processing advantage because the main verb is presented earlier and thus the entropy is lowered. This is an advantage mainly on the side of the listener. For the speaker, it should not play a role to such an extent in the course of the syntactic derivation. So we assume that syntactic planning in principle is less dependent on processing constraints, at least until Spell Out. As soon as PF is entered, however, processing begins to play a role as this is the interface at which the language producer has to juggle with prosodic constituency and pragmatic constraints such as information packaging.

The very case of information packaging shows that the speaker is not ‚egoistical', but takes the optimal reception by the listener into account in the course of language production. As these are context-sensitive aspects of the

utterance, we cannot expect them before Spell Out. In the end, any strategy that eases the assumed processing cost on the listener's side is advantageous. Reducing entropy by extraposition is such a strategy.

6 Summary

This paper looked at the placement of prepositional phrases in the so-called Postfield, the position behind the right sentence bracket, in Early New High German and early New High German scientific texts. It applied surprisal measurements to detect whether positioning PPs in the Postfield is correlated to a high informativeness of those phrases, which was measured according to Information Density Theory. In a second step, these results were connected to a generative approach that aimed to give further explanations about the Postfield itself.

Analysing the whole data, we find a strong effect for length on placing relative clauses in the Postfield, which is in line with previous findings, but none of the applied surprisal measures shows an effect for the position of PPs, followed by the genre and revealing that embedding of PPs is more likely in theological texts. However, when analysing the different 50-year-periods from 1650 to 1900, we find an effect of surprisal in the early periods. It vanishes over the centuries, leaving length to be the strongest predictor for extraposition in the 18^{th} and 19^{th} centuries.

In a further step, the data was divided into attributive PPs and independent PPs. This analysis showed that length is most influential for the 'extraposition' of independent PPs, but that cumulative surprisal has a significant effect for the 'extraposition'. These findings about attributive PP was similar to findings made in a study looking at relative clause adjacency (Voigtmann & Speyer 2021).

This approach was used for the generative analysis. Having excluded various alternatives, we also come to the conclusion that the placement of attributives is a postsyntactic phenomenon. Information density is already crucial at the point of the retrieval of referents from the mental lexicon or the common ground. Still, its structuring becomes relevant at PF level. To see extraposition as a PF phenomenon has the advantage that it also explains the high influence of the PP length which was visible in the PPs when we do not distinguish between attributive and independent PPs.

References

Admoni, Vladimir G. 1990. *Historische Syntax des Deutschen*. Tübingen: Niemeyer.

Altmann, Hans. 1981. *Formen der Herausstellung im Deutschen: Rechtsversetzung, Linksversetzung, freies Thema und verwandte Konstruktionen* Linguistische Arbeiten 106. Tübingen: Niemeyer. Zugl.: München, Univ., Habil.-Schr., 1979.

Aylett, Matthew & Alice Turk. 2004. The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech* 47(1). 31–56.

BBAW. 2019. Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. Berlin-Brandenburgische Akademie der Wissenschaften; http://www.deutschestextarchiv.de/.

Behaghel, Otto. 1932. *Deutsche Syntax: eine geschichtliche Darstellung*, vol. 4: *Wortstellung, Periodenbau*. Germanische Bibliothek. Heidelberg: Winter.

Bergmann, Rolf. 2019. Alt- und Mittelhochdeutsch: Arbeitsbuch zur Grammatik der älteren deutschen Sprachstufen und zur deutschen Sprachgeschichte. https://elibrary.onixtools.de/Cover/9783838551944.jpg.

Borter, Alfred. 1982. *Syntaktische Klammerbildung in Notkers Psalter.*, vol. 7 Das Althochdeutsche von St. Gallen. Berlin [u.a.]: de Gruyter.

Brooks, Thomas. 2006. *Untersuchungen zur Syntax in oberdeutschen Drucken des 16. - 18. Jahrhunderts* Schriften zur deutschen Sprache in Österreich ; 36. Frankfurt am Main: Lang.

Büring, Daniel & Katharina Hartmann. 1997. Doing the right thing. *The Linguistic Review* 14. 1–42.

Culicover, Peter & Michael Rochemont. 1990. Extraposition and the Complement Principle. *Linguistic Inquiry* 21. 23–47.

Cutler, Anne. 2008. The abstract representations in speech processing. *The Quarterly Journal of Experimental Psychology* 61(11). 1601–1619.

Dal, Ingerid & Hans-Werner Eroms. 2014. *Kurze deutsche Syntax auf historischer Grundlage* Sammlung kurzer Grammatiken germanischer Dialekte. Berlin [u.a.]: de Gruyter 4th edn.

Degaetano-Ortlieb, Stefania & Elke Teich. 2018. Using relative entropy for detection and analysis of periods of diachronic linguistic change. In *In proceedings of the 2nd joint sighum workshop on computational linguistics for cultural heritage, social sciences, humanities and literature at coling2018*, Santa Fe, NM, USA. ACL. https://www.uni-saarland.de/fileadmin/upload/lehrstuhl/teich/degaetano/latech2018_final.pdf.

Dürscheid, Christa. 1989. *Zur Vorfeldbesetzung in deutschen Verbzweit-Strukturen*. Trier: Wissenschaftlicher Verlag Trier.

Ebert, Robert Peter. 1980. Social and stylistic variation in Early New

High German word order: the sentence frame (›Satzrahmen‹). *Beiträge zur Geschichte der deutschen Sprache und Literatur* 102(Jahresband). 357–398. doi:doi:10.1515/bgsl.1980.1980.102.357. https://doi.org/10.1515/bgsl.1980.1980.102.357.

Eichinger, Ludwig M. 1995. Syntaktischer Wandel und Verständlichkeit: zur Serialisierung von Sätzen und Nominalgruppen im frühen Neuhochdeutschen. In Heinz Leonhard Kretzenbacher & Heinz L. Kretzenbacher (eds.), *Linguistik der Wissenschaftssprache* Forschungsbericht ; 10, Berlin: de Gruyter.

Engel, Ulrich. 1970. *Studie zur Geschichte des Satzrahmens und seiner Durchbrechung* Studien zur Syntax des heutigen Deutsch: Paul Grebe zum 60. Geburtstag (Sprache der Gegenwart ; 6). Düsseldorf: Schwann.

Frey, W. & K. Pittner. 1998. Zur Positionierung der Adverbiale im deutschen Mittelfeld. *Linguistische Berichte* 176. 489–534.

Frey, Werner. 2015. Zur Struktur des Nachfelds im Deutschen. In Hélène Vinckel-Roisin (ed.), *Das Nachfeld im Deutschen. Theorie und Empirie*, 53–76. Berlin: de Gruyter.

Futrell, Richard, Edward Gibson & Roger Levy. 2021. Lossy-Context Surprisal: An Information-Theoretic Model of Memory Effects in Sentence Processing. *Cogn Sci.* 45(5).

Féry, Caroline. 2015. Extraposition and prosodic monsters in German. In Lyn Frazier & Edward Gibson (eds.), *Classics in linguistics*, vol. 5, 11–38. Amsterdam, Philadelphia: Benjamins 2nd edn.

Gibson, Edward. 1998. Linguistic Complexity: Locality of Syntactic Dependencies. *Cognition* 68. 1–76.

Gibson, Edward, Richard Futrell, Steven Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen & Roger Levy. 2019. How Efficiency Shapes Human Language. *Trends in Cognitive Sciences* 23(5). 389–407.

Gries, Stefan Th. 2021. *Statistics for Linguistics with R. A Practical Introduction*. Berlin, Boston: De Gruyter Mouton. doi:doi:10.1515/9783110718256. https://doi.org/10.1515/9783110718256.

Guthrie, D., B. Allison, W. Liu, L. Guthrie & Y. Wilks. 2006. A Closer Look at Skip-gram Modelling. *Proceedings of the Fifth International Conference on Language Resources and Evaluation* .

Haider, Hubert. 2010. *The Syntax of German*. Cambridge: CUP.

Hale, John. 2001. A probabilistic Early parser as a psycholinguistic model. *Proceedings of the second meeting of the North American chapter of the Association for Computional Linguistics* .

Harley, Heidi & Rolf Noyer. 1999. Distributed morphology. *Glot International* 4. 3–9.

Hawkins, John A. 1992. Syntactic Weight Versus Information Structure in Word Order Variation. *Informationsstruktur und Grammatik. Linguistische Berichte Sonderhefte* 4. 196–219.

Hinterhölzl, Roland. 2010. Zur Herausbildung der Satzklammer im Deutschen. In Arne Ziegler (ed.), *Historische Textgrammatik und Historische Syntax des Deutschen*, 121–138. Berlin: De Gruyter.

Höhle, Tilman N. 1981. Beitrage zur deutschen Grammatik: Gesammelte Schriften von Tilman N. Höhle. In Marga Reis, Frank Richter, Stefan Müller & Tilman N. Höhle (eds.), *Classics in linguistics*, vol. 5, Berlin: Language Science Press Zweite durchgesehene Auflage edn.

Inaba, Jiro. 2007. *Die Syntax der Satzkomplementierung*. Berlin: Akademie Verlag.

Jaeger, Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61(1). 23–62.

Jurafsky, Dan. 2003. Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In *Probabilistic linguistics*, Cambridge, MA: R. Bod, J. Hay, S. Jannedy.

Koch, Peter & Wulf Österreicher. 2007. Schriftlichkeit und kommunikative Distanz. *Zeitschrift für germanistische Linguistik* 35(3). 346–375.

Konopka, Marek. 1996. *Strittige Erscheinungen der deutschen Syntax im 18. Jahrhundert*, vol. 173 Reihe germanistische Linguistik. Tübingen: Niemeyer.

Kusmirek, Adam, Clayton Greenberg, Youssef Oualil & Dietrich Klakow. 2023. Lm-gui: A graphical user interface for n-gram language models. https://github.com/uds-lsv/languagemodels.

Lenerz, J. 1977. *Zur Abfolge nominaler Satzglieder im Deutschen*. Tübingen: Narr.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177.

Levy, Roger & Florian Jaeger. 2007. Speakers optimize information desity through syntactic reduction. In B. Schölkopf & J. C. Platt & T. Hoffman (ed.), *Advances in neural information processing systems 19*, 849–856. MIT Press.

Manning, C. & H. Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

McClelland, James L. & Jeffrey L. Elman. 1986. The trace model of speech perception. *Cognitive Psychology* 18(1). 1–86.

Morlicchio, Elda. 1991. Der Satzrahmen: die Länge des Satzes und die Mitteilungsperspektive. Eine Analyse anhand deutscher Urkunden des 13. Jahrhunderts. *Zeitschrift für deutsche Philologie*. 1–11.

Müller, Gereon. 1995. On extraposition and successive cyclicity. In Uli Lutz

& Jürgen Pafel (eds.), *On extraction and extraposition in German*, 213–243. Amsterdam: Benjamins.

Mürmann, M. 2014. *Wahrscheinlichkeitstheorie und stochastische Prozesse*. Berlin: Springer Spektrum.

Ostermann, Simon. 2021. *Script knowledge for natural language understanding*. Saarbrücken: Saarland University dissertation.

Paul, Hermann. 2007. Mittelhochdeutsche Grammatik. In Thomas Klein, Hans-Joachim Solms, Klaus-Peter Wegera, Ingeborg Schöbler & Heinz-Peter Prell (eds.), *Sammlung kurzer Grammatiken germanischer Dialekte*, Tübingen: Niemeyer 25th edn.

Petkova-Kessanlis, Mikaela. 2015. Nachfeldbesetzung und ihre kommunikative Funktion in wissenschaftlichen Texten. In Hélène Vinckel-Roison (ed.), *Das Nachfeld im Deutschen. Theorie und Empirie*, 211–288. de Gruyter.

Polenz, Peter von, Claudine Moulin & Dominic Harion. 2013. *Deutsche Sprachgeschichte vom Spatmittelalter bis zur Gegenwart*, vol. 2: *17. und 18. Jahrhundert.* De Gruyter Studienbuch. Berlin: De Gruyter 2nd edn.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Sapp, Christopher D. 2014. Extraposition in Middle and New High German. *The journal of comparative Germanic linguistics* 17(2). 129–156.

Schildt, J. 1976. Zur Ausbildung der Satzklammer. In G. Kettmann & J. Schildt (eds.), *Zur Ausbildung der Norm in der deutschen Literatursprache auf der syntaktischen Ebene*, Akademieverlag.

Shannon, Claude E. 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27(3). 379–423.

von Stechow, Armin & Wolfgang Sternefeld. 1988. *Bausteine syntaktischen Wissens*. Opladen: Westdeutscher Verlag.

Takada, Hiroyuki. 1998. *Grammatik und Sprachwirklichkeit von 1640-1700: zur Rolle deutscher Grammatiker im schriftsprachlichen Ausgleichsprozeß*, vol. 203 Reihe Germanistische Linguistik. Tübingen: Niemeyer.

Truckenbrodt, Hubert. 1995. *Phonological phrases: their relation to syntax, focus, and prominence*: MIT dissertation.

Voigtmann, S. & A. Speyer. 2021. Information Density and the Extraposition of German Relative Clauses. *Frontiers in Psychology* .

Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change* 9. 81–105.

Weber, Sabrina. 2019. The Acceptability of Extraposition of PPs out of NP in German. *Proceedings of Linguistic Evidence 2018: Experimental Data Drives Linguistic Theory* 27. 63–84.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. http://ggplot2.org.

Wöllstein, Angelika. 2014. *Topologisches Satzmodell*. Heidelberg: Winter 2nd edn.

Zahn, Günther. 1991. *Beobachtungen zur Ausklammerung und Nachfeldbesetzung in gesprochenem Deutsch* Erlanger Studien ; 93. Erlangen: Palm und Enke.

Zifonun, Gisela, Ludger Hoffmann & Bruno Strecker. 1997. *Grammatik der deutschen Sprache*, vol. 2. Berlin: Niemeyer.

Sophia Voigtmann
Universität des Saarlandes
Campus A2 2 · 3.24–3.25
D-66123 Saarbrücken
sophia.voigtmann@uni-saarland.de
https://nds.uni-saarland.de/voigtmann/

Augustin Speyer
Universität des Saarlandes
Campus A2 2 • 3.13
D-66123 Saarbrücken
a.speyer@mx.uni-saarland.de
https://nds.uni-saarland.de/speyer/