Sophia Voigtmann & Augustin Speyer

2022

Information Density as a Factor for Syntactic Variation in Early New High German

in

Robin Hörnig, Sophie von Wietersheim, Andreas Konietzko & Sam Featherston (eds.),

Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data

pp. 113–133

Tübingen: University of Tübingen

https://publikationen.uni-tuebingen.de/xmlui/handle/10900/119301

In R. Hörnig, S. von Wietersheim, A. Konietzko & S. Featherston (eds.) (2022), *Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data*. Tübingen: University of Tübingen. https://publikationen.uni-tuebingen.de/xmlui/handle/10900/119301

Information Density as a Factor for Syntactic Variation in Early New High German[1]

Sophia Voigtmann & Augustin Speyer
Saarland University
sophia.voigtmann@uni-saarland.de, a.speyer@uni.saarland.de

1 Introduction

In contrast to other languages like English, German has certain liberties in its word order. Different word orders do not influence the proposition of a sentence. The frame of the German clause are the sentence brackets (the left (LSB) and the right (RSB) sentence brackets) over which the parts of the predicate are distributed in the main clause, whereas in subordinate clauses, the left one can host subordinate conjunctions. But apart from the sentence brackets, the order of constituents is fairly variable, though a default word order (subject, indirect object, direct object for nouns; subject, direct object, indirect object for pronouns) exists. A deviation of this order can be caused by factors like focus, given-/newness, topicality, definiteness and animacy (Zubin & Köpcke, 1985; Reis, 1987; Müller, 1999; Lenerz, 2001 among others). Though, as mentioned before, the sentence brackets function as clause boundaries, these boundaries are often crossed. The most prominent example for this phenomenon is relative clauses (RC, underlined):

(1) a. Die Mutter [hat$_{LSB}$] *ein Kind*, <u>das ein Eis isst</u>, [gesehen$_{RSB}$].
the mother has a child who an ice.cream eats seen.

b. Die Mutter [hat$_{LSB}$] *ein Kind* [gesehen$_{RSB}$], <u>das ein Eis isst</u>.
the mother has a child seen who an ice.cream eats
'The mother has seen a child who eats an ice cream.'

Example (1a) presents a RC that is adjacent to its head noun. Both RC and antecedent are placed within the sentence brackets. In contrast to that, (1b) shows a RC behind the right sentence bracket, an extraposed RC. Both sentences convey the exact same meaning. So why do people use both variations equally frequently? Extraposition is said to be caused by different factors, such as the length of the extraposed material (Wasow, 1997; Zifonun et al., 1997; Uszkoreit et al., 1998 among others), the distance to the post field (Hawkins, 1994; Gibson, 1998) as well as a sentence's closeness to the oral discourse mode (Zifonun et al., 1997). For RCs, it is also important to note their type (Zifonun et al., 1997); restrictive relative clauses are more likely to be separated from their head noun than non-restrictive ones.

The extraposition of relative clauses is not only a phenomenon of present-day German. Since the Old High German language period, extraposition of relative clauses was a common

[1] 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 conference *Linguistic Evidence 2020.* Our thank goes to the organizers, and the audience of this conference 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.

phenomenon (Sapp, 2014; Coniglio & Schlachter, 2015; Sahel, 2015 among others), which has not changed remarkably over the centuries.

RC extraposition as described above is not completely adaptable for phrasal extraposition. As mentioned, phrases can be placed at different positions within the sentence brackets. But they are also found behind the right sentence bracket, a phenomenon that has been much more common in earlier language periods (Sahel, 2015 among others). Also today, native speakers produce clauses with phrasal extrapositions.

(2) Paul [hat$_{LSB}$] seiner Schwester Geld [gegeben$_{RSB}$] <u>für ein neues Fahrrad</u>.
Paul has his sister money given for a new bike
'Paul has given his sister money for a new bike.'

The reasons for putting phrases behind the right sentence bracket resemble those for relative clauses: influential factors are also the length of the phrase, the distance between the assumed base position of the phrase to the post field, closeness to the oral discourse mode and the phrase type. Preposition Phrases (PP), for example, are more often placed behind the right sentence bracket than Noun Phrases (NP), especially when these phrases are used as subjects or objects (Zifonun et al., 1997).

If we compare RC and phrasal extraposition, we see that they share two crucial factors: length and discourse mode. Another factor, not mentioned so far, is information management. Zifonun et al. (1997) state for RCs that "Informationsentflechtung" (= information disentanglement) influences the position of an RC, without giving a detailed explanation for what they mean by that.

In this paper, we want to concentrate on this aspect of extraposition that can explain both phrasal and RC extraposition. In former research, information was often understood in terms of newness and givenness or focus (Chafe, 1976; Prince, 1981; Krifka, 2007 among others). We understand information here in terms of Shannon's (1948) *information density* (ID). ID can be defined as the "amount of information per unit comprising the utterance" (Levy & Jaeger, 2007: 1). With surprisal measurements, one can define how likely the occurrence of a word is, given a context of n other words. Frequent combinations show lower surprisal values than rare combinations. Since surprisal is connected to perceiving difficulties (Hale, 2001), the impact of frequent combinations with minor surprisal values on the working memory is lower than it is for rare combinations with higher surprisal values (Hale, 2001; Levy & Jaeger, 2007). To improve text comprehension, producers therefore distribute information as evenly as possible across a discourse ("Uniform Information Density Hypothesis (UID)", Levy & Jaeger, 2007). UID also provides evidence for the influence of ID on syntax.

Gibson et al. (2019) describe the influence of ID on all aspects of language. So, we assume that it is also relevant for the placement of linguistic material in positions after the points where the clause can in fact be called finished. Besides UID, findings from Genzel & Charniak (2003) also influence our current study. They found a stable entropy distribution over texts from different genres and languages as well as an increasing entropy rate for sentences that occur later in a context because formerly mentioned information makes it easier to process these parts. We take these findings into account for our first hypothesis H1:

(H1) Higher cumulative surprisal values favor extraposition.

Since former research has shown changes in the frequency of extraposition and changes in the style of scientific writing, we aim to take a look at these changes. The decreasing frequency of extraposition can be related with a decreasing influence of ID on that phenomenon. The right sentence bracket as a boarder is fully established up to the 19th century, so that extraposition over this position might not facilitate comprehension but make it more difficult. This assumption leads to the following hypothesis H2:

(H2) The influence of surprisal values declines over time.

This paper aims to cast light on both hypotheses. Therefore, we first present a more detailed description of Information Density and former research on Extraposition (Section 2). Section 3 deals with our Early New High German (ENHG) corpus and the methods we used to test the hypotheses. In Section 4, we present our results, before discussing them in Section 5. The paper closes with a conclusion in Section 6.

2 Theoretical Background

2.1 Information Density

Information density can be defined as the amount of information per utterance (Levy & Jaeger, 2007: 1). The concept of its calculation goes back to Shannon (1948). He investigated how coded messages can be transported through a channel in which interference is present and decoded again at the receiver without loss of information (Shannon, 1948: 379). Despite different coding of the message, its meaning usually remains the same and must be understood by the receiver. According to Shannon (1948), variations are not due to differences in meaning, but due to the quantity and nature of the information to be transmitted. Detailed research on this phenomenon was done by Levy & Jaeger (2007), Jaeger (2010) and Hale (2001), among others. Their research confirms that the information in texts tends to be distributed evenly. Shannon (1948) also assumes this distribution and therefore calculates the probability of variation with a logarithmic function, which can also take up these variations (Shannon, 1948: 380).

Another important aspect of Shannon's calculations is the interpretation of human communication as a 'discrete system', which allows a one-to-one transmission of meaning to signs, as is the case with words. With the help of this system, it can be shown that in an undisturbed channel only a certain amount of information per unit of information (clause or time span) can be transmitted without risking a loss of information. Each channel therefore has a limited capacity. To what extent this capacity is used depends on the source and the transmitter. If a transmitter wants to make communication efficient, it must encode the information in such a way that it makes maximum use of the channel's bandwidth.

However, if there is interference in this channel that prevents the recipient of the message from receiving all the information, or if there is ambiguity in the message sent, it is useful to be able to make predictions about the next word in the communication. For this purpose, however, it is necessary to know the context of the message, which is made up of the preceding words (Shannon, 1948: 383).

To make this prediction, so-called n-gram models are used (Shannon, 1948: 383). A 1-gram model (unigram) simply determines the frequency of words in a text (Shannon, 1948; Genzel & Charniak, 2002), a 2-gram model (bigram model) yields the probability of a word occurring after a certain other word, a trigram model calculates this probability when two other words precede it, etc.

Mathematically, this is expressed by conditional probabilities. If n is increased, the complexity of the conditions also increases.

(3) Peter sees a blue [chair].

So with the n-gram models, the probability that "chair" occurs when "blue", "a blue" or "sees a blue" is in front of it can be calculated. This example shows that with increasing complexity of the conditions, i.e. with an increasing number of words in front of the lexeme in question, a greater mathematical effort is required to calculate the surprisal values.

To achieve the feasibility of the mathematical condition, the Markoff assumption is used. This states that the probability of a future unit can be predicted without looking too far into history (Mürmann, 2014). For languages, this means that not every linguistic utterance ever produced has to be included in the calculation, but a part of the linguistic utterances is sufficient

to make acceptable statements. By calculating the probabilities, the 'surprisal' value of a word is obtained. The surprisal value indicates how unexpectedly a word appears in its context. It is high if the combination occurs rarely and low if it occurs frequently. The surprisals can be used to create a 'Language Model', which Hale (2001) defines as "a probability of every string of words from the lexicon" (Hale, 2001: 1). However, the lexicon varies depending on the type of text, the authors, the age and the content of the texts. For this reason and due to the effort of calculation, there is (still) no language model of the entire German language, but only models for parts of the language. The model is trained using a training corpus, which should be similar to the test corpus in terms of content and language. For instance, a Middle High German training corpus should be used to study Middle High German texts. This is done to achieve a better prediction probability. In order to be able to compare the quality of test sets, a standardized measure is necessary. For this, entropy is used.

Entropy (H) is "a measure of how high the degree of order of the system under consideration is" (Tipler et al., 2015: 621). It can thus be used to calculate how likely it is that an event can be guessed correctly (Genzel & Charniak, 2002: 199).[2]

For each possible state of a source or context, there are probabilities with which a symbol or a word can occur (Shannon 1948: 393). The more symbols a source contains, the more precisely H can be determined. If the entropy of a source could be determined, the ratio of this entropy to the maximum value that can be reached indicates the relative entropy (Shannon, 1948: 394). In the case of sender and receiver, Shannon assumes a limited working memory which can influence the understanding of the messages. The understanding of information can also be influenced by interference in the channel. The calculations of Shannon showed that the effects of the interference can be circumvented by a constant transmission rate of information.

This observation also applies to human communication (Genzel & Charniak, 2002: 199). Genzel & Charniak (2002) apply Shannon's (1948) concept to texts and calculate the entropy of words in a given context. They assume that the average entropy of words in a text is equal (Genzel & Charniak, 2002: 199). Both, models that calculate only n-grams and models that, taking syntactic information into account, are closer to the true probable distribution, show entropy increasing with the number of sentences, if the context was not taken into account. However, when looking at the average values, an even distribution of entropy is visible. In Genzel & Charniak (2003), they expand their findings to other languages (Spanish and Russian) and other genres than newspaper texts (literature).

The Uniform Information Density (UID) Theory of Levy & Jaeger (2007) refers to an even distribution of information and the avoidance of peaks which would overburden the channel capacity and endanger an optimal communication. In spoken language, speakers tend to adapt the duration of a phonetic utterance to the predictability of the unit, as shown by Aylett & Turk (2004) in their Smooth Signal Redundancy Hypothesis. The unexpected is thus pronounced longer than the expected. The principle of UID, that followed Aylett & Turk's (2004) Hypothesis, can be applied to all linguistic levels. To prove this for the syntactic design of utterances, 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). Jaeger (2005, 2010) as well as Wasow & Jaeger (2011) investigated this effect

[2] Various general conditions apply to the entropy value: If all events are equally probable and given, the entropy value corresponds to 0 (Shannon, 1948: 390). The entropy becomes larger the more similar the probability of occurrence of the events is. This means that the result is more difficult to predict because there are more probable choices (Shannon, 1948: 391). The entropy of two combined events is less than or equal to the sum of the individual uncertainties of the events (Shannon, 1948: 391). If two random events are dependent on each other, this is called 'conditional entropy' (Shannon, 1948: 392) and is calculated with the average of the entropy values of the two events. The probability of occurrence of one of the events must be known. The entropy of event y will not increase if another event x related to y is known (Shannon, 1948: 392).

for relativizers in corpora as well as in experiments and found an effect for optional "that"-realization in relative clauses. RCs tend to not express the relativizer when their occurrence is necessary. Jaeger (2005) and Frank & Jaeger (2008) furthermore find evidence for an incremental production of words in a sentence, influencing the addition of relativizers and therefore also the whole placement of larger parts of the sentence such as phrases or RCs.

Hale (2001) investigated a similar phenomenon and attributed it to the linear construction of language. Both investigations use n-gram models to calculate the probability of words in a context. Levy & Jaeger (2007) found that their theory of UID actually holds true and that English relative clauses are actually more likely to be introduced by a pronoun when entropy is higher. This also shows the influence of information density on the syntax of sentences, which in turn was also shown by Hale (2001) by calculating the probable sequence of purely syntactic information such as determiner phrases before NPs.

2.2 Extraposition

In this paper, extraposition is treated in terms of the topological field model (Wöllstein, 2010). This model is used to describe sentence types dependent on the position of the finite verb. It is characterized by the sentence brackets, the left and right sentence bracket (LSB, RSB). In main clauses and *wh*-questions, the finite part of the verb form is placed in the LSB, while the infinite part of the verb form or verb particles, if existent, are put into the RSB. In subordinate clauses, both the finite and infinite part are found in the RSB, as the LSB is filled with the subordinate conjunction. Independent of the clause type, the RSB marks the end of the clause. There is however a position behind the RSB, the postfield (PoF; "Nachfeld", Wöllstein, 2010), which is usually occupied by other clauses, mostly subordinate clauses. Other formally integrated, embedded material like comparative elements (4a), prepositional phrases (PP, 4b) or relative clauses (RC, 4c) can also be placed there.

(4) a. Paul hat *mehr* [gearbeitet$_{RSB}$] als Susanne.
Paul has more worked than Susanne
'Paul has done more work than Susanne.'

b. Ich habe dich gestern [gesehen $_{RSB}$], in dem Kino neben dem Einkaufscenter.
I have you yesterday seen in the cinema next.to the mall
'I have seen you in the cinema next to the mall yesterday.'

c. Hast du *den Film* [gesehen $_{RSB}$], von dem ich dir erzählt habe?
have you the movie seen about which I you told have
'Have you seen the movie I have told you about?'

In Modern German, extraposition is never obligatory, but there are several factors influencing the movement of constituents or parts of them.

The possibility of occupying the PoF must be determined gradually. This means that not all elements that have their base position in the middle field (i.e. between the LSB and the RSB) are equally acceptable in the PoF. The PoF can host independent phrases, but also attributes and extending parts of constituents, e.g. RCs (Zifonun et al., 1997: 1650). It can be observed that complements are less often in the postfield than adjuncts. Overall, the tendency of an element to appear in the PoF is less dependent on syntactic and semantic factors than on "form related properties" (Zifonun et al., 1997: 1651). Nevertheless, it can be observed that in Modern German, certain expressions cannot appear in the PoF, such as e.g. particles or pronouns (Zifonun et al., 1997: 1651). The elements that can stand in the post field are clauses, adjunct phrases, prepositional phrases (PPs), noun phrases (NPs) and adjective or adverb phrases (AP). The order of the enumeration also reflects the frequency of positioning behind the RSB.

One of the main factors influencing extraposition is the size of the material. Distinctions must be made as to whether it is size in structural or numeric terms. Material containing a

predicate or consisting of one or more clauses is structurally extensive. The more structurally extensive the material is, the more likely it is to be placed in the PoF. This factor also explains the high frequency of clauses there. However, the pure length of the material is also an im-portant factor. Uszkoreit et al. (1998) found that RCs in the PoF are on average 10.3 words long, whereas adjacent RCs are only nine words long on average (Uszkoreit et al., 1998: 131). This allows a lower burden on short-term memory, provided the distance between antecedent and RC is not too great. In general, Uszkoreit et al. (1998), (also Hawkins, 1994; Gibson, 1998; more recently Levy, 2008; Gibson et al., 2019), see the sequence of phrases or generally words as closely related to the fastest possible processing of the information. The tendency described by Hawkins (1994) that not too many elements may be placed between the head of the relative clause and the clause bracket fits into this picture (see also Uszkoreit et al., 1998; Weber, 2019). In principle, the presence of a referring expression before the RSB facilitates the removal of material, which can also be observed in the accumulation of comparative elements at this posi-tion.

In addition to the length of the material and the distance over which the element has to be moved, information management is also important for extraposition, as already suggested by the relief of short-term memory. Vinckel-Roisin (2015) calls extraposition discourse-relevant when a new topic is introduced or a discourse-topic is established in the PoF. In addition, there are text type specific aspects of extraposition, for example when newspaper articles contain punch lines.

Text type specificity and proximity to oral discourse (according to the classification of Koch & Oesterreicher, 2007) favour the positioning of elements in the PoF. Averintseva-Klisch (2007), for example, describes extraposition as a way of not overburdening the listener and structuring the conversation. In addition, oral communication offers fewer opportunities to plan the structural presentation of a content in detail, which is why the influences of the incremental construction of clauses should be more noticeable.

With all these mentioned factors influencing extraposition, it should be noted that they pri-marily relate to the current Modern Standard German. In former periods of German, we find a somewhat different picture of extraposition. For once, extraposition used to be much more com-mon. This can be seen e.g. from the statistics by Schildt (1976), who investigated the develop-ment of the sentence bracket in German. He found for the time frame 1470 to 1530 that 68 % of the clauses in his corpus had no material in their PoF. In 1670 to 1730, it was already 81 % (Schildt, 1976: 271). Conversely, this means that around 1500, in about one third of the clauses there was material in the PoF. This fits nicely to the study reported in Speyer (2016) about PoF filling in narrative dialect texts. Here, the percentage of clauses with material in the PoF de-creased from 1450 to 1900, e.g. in Bavarian from 28 % to 8 % (Speyer, 2016: 145). Interest-ingly, this coincides with a decreasing information structural specialization of the PoF (in terms of givenness). We should expect the PoF to be used more frequently once it became more per-missive in information structural terms. So, obviously the decrease in frequency is a process independent of the information structural development. We see it as a reflex of the developing prose style in which processing constraints are less and less taken into consideration.

3 Corpus and Method

For our corpus, we use texts from the Deutsches Textarchiv (DTA). The DTA consists of a collection of different genres and periods to provide an overview over the development of the German language from the 16th to 20th century with balanced samples from newspapers, novels,

literature for a specific purpose ("Gebrauchsliteratur") and scientific texts. The texts are digitalized and preprocessed for corpus linguistic analysis, that is normalized, tokenized, lemmatized and POS-tagged.[3]

The DTA's advantage for us are high variety of genres, and, most relevant for our analysis, the preprocessing step of lemmatization. We are currently not interested in the problems of grammatical perception of RC and phrasal extraposition, but in the lexical perception. Therefore, we need to do the ID analysis on the basic word forms without any bias by different inflections or orthographic inconsistencies typical for ENHG. The preprocessing steps provided by the DTA eliminate the problem of different spellings of the same word well enough to use the existing lemmatized version of the texts.

The variety of text types is another key feature of the DTA. The current study considers scientific texts, a genre developing in ENHG times. German scientists published their findings mostly in Latin until the 17th century. When they started writing in German, their writing style resembled letters since this was the first medium non-Latin scientific writing was published in. The style of letters is settled between oral and written discourse mode (Koch & Oesterreicher, 2007), therefore typical variations of orality such as extraposition could be found there. However, the former Latin tradition might also provide more complex sentence periods, so authors used to this style might incorporate it into the new German writing, resulting in a style that is more complex than an average orally designed sentence. So, scientific texts are in the field of tension between oral and written style. Longer and more complex sentence periods put a strain on the working memory that could be reflected in ID. In this respect, different topics of scientific texts also influence the style, since the authors were exposed to different traditions and address a different kind of audience with respect to their previous experience in the topic of concern. We assume that texts meant for a specific audience like physicians are even in this early period of German scientific writing closer to the written discourse mode than texts with a broader audience design like theological texts (on a similar finding for English concerning the distinction between arts and natural science, see Degaetano-Ortlieb, 2017).

Our whole corpus from the DTA consists of 33 medical and theological texts or text parts from 1600 to 1900 with 2 883 109 tokens. The texts themselves were chosen arbitrarily as long as they were originally written in German and not translated. We manually annotated the corpus for several grammatical factors. These include the annotation of the extraposed or embedded phenomena in question (RC, PP, NP, comparative elements, APs), if existent, their antecedent as well as the left and right sentence brackets (see Section 2.2). These parameters are necessary for the distinction between the embedded and extraposed variants. When an element in question is put between the brackets or in front of the LSB, it is classified as 'embedded'. When an element is placed behind the RSB, it is called 'extraposed'. These classification mechanisms are linked to the existence of a filled RSB. For cases without a distinct RSB, the phenomena are classified as 'ambiguous'. In case of RCs, an RC is also classified as 'extraposed' when there is any other material between the antecedent and the RC itself (5), since we assume that movement has taken place to separate the RC from its head noun. Whether it is place in the PoF or still part of the middlefield cannot surely be determined.

[3] The DTA-Project has been started more than ten years ago and is therefore not on a standard we are used to by newer projects like the *Referenz Korpus Frühneuhochdeutsch (ReF)*. We are aware of updates on the data, but have not included possible improvements on the annotations because we manually annotated relevant information on downloaded versions of the texts. The time of the download was 2018, so this is the version of the corpus used here.

(5) Er sah endlich *den Film* mit seinen Kindern, <u>den er sehen wollte</u>.
he watched eventually the movie with his kids that he watch wanted.to
'He eventually watched the movie with his kids that he wanted to watch.'

Though example (5) does not have a filled RSB, 'mit seinen Kindern' (with his kids) intervenes between the antecedent (italic) and the RC (underlined).

The annotation had to be done manually because the Tag-set of the DTA does not allow a successful automatic annotation or detection of relevant material due to a high orthographic variation in ENHG and an all-automatic annotation of POS-Tags in the DTA which resulted in a rather poor tagging, especially for earlier periods. Our manual annotation was made using WebAnno (Eckart de Castilho et al., 2016).

In a following step, we calculated a language model (Hale, 2001) and created a bigram language model with Kneser-Ney smoothing (Chen & Goodman, 1999) for a period of 50 years on lemmata using an SFB-internal tool. As a training corpus, we used the remaining medical and theological texts from the DTA.

This paper concentrates on the first period (1650-1700) and the last period (1850-1900). For the first time span, the training data consisted of 2 107 590 tokens, the test data presented here of 51 943 tokens, the corpus part from 1650 to 1700 in total consisted of 204 142. We are aware of the rather small amount of training data, but we intended to keep our language model as close as possible to the test data, to prevent a high amount of out-of-vocabulary-tokens. The results of the language model speak in favor of our method. The in-vocabulary score is 4 219 / 4 358, that means the out-of-vocabulary percentage is only 3.19 %. We can cope with this bias on our analysis and ignored OOV-lemmata for the ID calculation presented below.

For the second time span, the training data consists of 1 270 561 tokens, the test set of 36 262 tokens. The OOV-score for this period is not as good as it is for the other time span: the in-vocabulary rate is 34 415 / 36 262, that is an OOV-score of 5.09 %. In both cases, punctuation marks were eliminated before the Language Model was calculated.

In a next step, we used R (The R-Core-Team, 2018) to clean the data. We removed all sentences not including a phenomenon in question and all annotation steps that we are currently not interested in, like POS-tags or the normalized word forms. Then, we divided the corpus into the different kinds of extraposition: RC, PP, NP, Adverbial Phrases, Adjective Phrases and comparative elements and continued the analysis.

Due to the formerly mentioned incremental sentence production, it can be assumed that cognitive load is created when the surprisal value of each word is added to the following word. We call this 'cumulative' or 'summed up surprisal' and estimate that it is an adequate measure for the cognitive load of complex expressions computed as a whole. The cumulative surprisal is of course connected to the length of a sentence. This measure is still suitable for a study which is concerned with a phenomenon regularly connected with length (see Chapter 2.2 on Extraposition). However, we also checked whether length, measured in words, has an influence on extraposition of RCs or phrases. We calculated the mean surprisal for every RC and every extraposed and embedded phrase to test for an influence of that value instead or in combination with the summed up surprisal values, too.

Though it would be interesting to test the whole sentence including a phenomenon in question, this is not possible because the sentence boundaries in our corpus were often marked incorrectly, so that cognitively coherent clauses are spread over two or more sentences or independent sentences were combined. We have yet not been able to solve this problem. Therefore, only the RC or phrases themselves can be tested for surprisal values.

4 Results

For this paper, we only analysed a part of our corpus. We chose to include texts from the first period (1650 to 1700) and from the last period (1850 to 1900) to show the range within the corpus data and to be able to answer Hypothesis two, which is concerned with the change of ID influence over time. For both periods, we included medical and theological texts. The following, arbitrarily chosen texts were used for the evaluation:

Table 1. Texts used for the evaluation

Time span	Text	Genre	Total number of sentences	Total number of tokens
1650-1700	Purmann, Matthäus Gottfried: Der rechte und wahrhafftige Feldscher. Halberstadt, 1680.	Medicine	2 537	46 971
1650-1700	Abel, Heinrich Kaspar: Wohlerfahrner Leib-Medicus der Studenten. Leipzig, 1699.	Medicine	3 335	49 972
1650-1700	Spener, Philipp Jakob: Pia Desideria. Frankfurt (Main), 1676.	Theology	1 814	58 680
1650-1700	[Rotth, Albrecht Christian]: Eylfertiges Bedencken über M. August Hermann Franckens [...] Seine Schutz-Predigt. Halle, 1692.	Theology	197	7 934
1850-1900	Kraepelin, Emil: Ueber die Beeinflussung einfacher psychischer Vorgänge durch einige Arzneimittel. Jena, 1892.	Medicine	3 341	81 738
1850-1900	Koch, Robert: Untersuchung über die Aetiologie der Wundinfectionskrankheiten. Leipzig, 1878.	Medicine	1 083	24 246
1850-1900	Egger, Augustinus: Der christliche Vater in der modernen Welt. Erbauungs- und Gebetbuch. Einsiedeln u. a., [1895].	Theology	5 073	86 146
1850-1900	Hasak, Max: Die Predigtkirche im Mittelalter. Berlin, 1893.	Theology	446	10 193

4.1 Relative Clauses

If both time periods are considered together, 869 extraposed and 767 embedded RCs are found, which indicates an overall even distribution of RC positions. Table 1 also lists the ambiguous RCs excluded from further analyses. A total of 1 675 records were excluded due to their ambiguous assignment to PoF or the middle field.

Separated according to the genres examined, we see that the medical texts contain more RCs than the theological ones (medical texts: 1008 RCs (61.61 %), theological texts: 628 RCs (38.39 %)). This could be explained by a different distribution of the verbs and tenses used, which do not always require RSB, over which extraposition is still more likely than over other, non-verbal material (Uszkoreit et al., 1998).

Table 2. Results for relevant values of ID and length for RCs

Variable	Minimum	Maximum	Mean
Mean surprisal value	1.094	4.42	2.67
Cumulative surprisal value	4.38	202.497	29.58
Length in words	2	81	11.47

4.1.1 Correlation between Relative Clauses Extraposition and Information Density

To answer the first hypothesis, which deals with whether there is a connection between the information density of a relative clause and its position, a look at the distribution within the positions is necessary.

If we now differentiate the two positions of the RC, the following picture emerges: In the medical texts, there are 457 (45.33%) embedded and 551 (54.66%) extraposed RCs, in the theological texts, the picture is more balanced, with 310 (49.4%) embedded and 318 (50.6%) extraposed RCs. The distribution of the cumulative surprisal values on the two RC positions examined can be clearly seen in the box plot Figure 1, as well as the distribution of the outliers and the average values in relation to the positions.

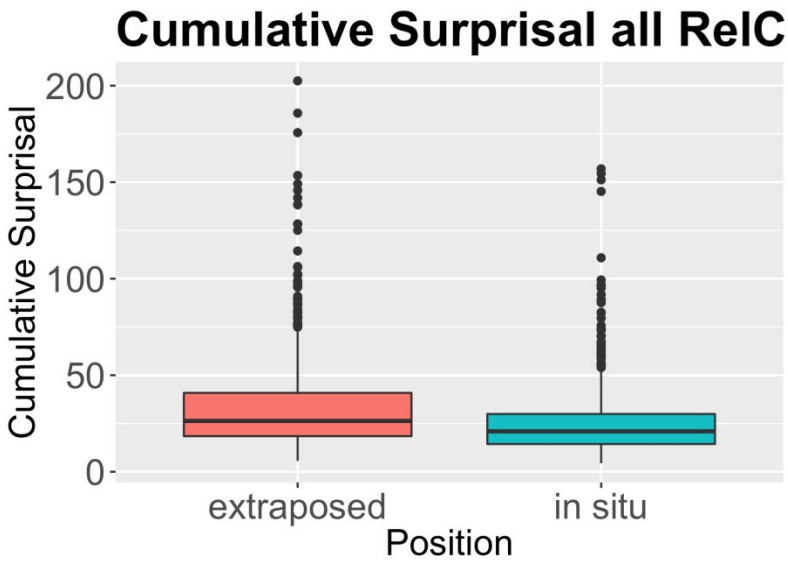

Figure 1. Contrast of extraposed and in situ RC

The embedded RCs (like 6) have lower cumulative surprisal values than the extraposed RCs. The former fluctuates between 4.38 and 156.9, their mean is 25.36. Their counterpart, on the other hand (7), fluctuates between 5.6 and 202.5 bits with a mean of 33.3. Differences can also be seen in the length. The in situ RCs vary between two and 60 words. The smallest extraposed RC has three words, the longest 81, their mean length is 9.83. The results for the mean surprisal values as well as for the cumulative surprisal values and the length can be found in Table 2, though this table does not differentiate between embedded and extraposed RCs.

(6) Alles aber auff die art, [daß] *solche leute*, <u>mit denen man handelt</u>, selbst
all still in the way that such people with whom one trades themselves
[sehen koennen$_{RSB}$], daß man alles thue
see can that one everything does

'But everything in a way that such people with whom one trades can see themselves that one does everything.'

(Spener, 1676)

(7) Fragst du […] welches denn unsere Teutsche Thee ſey ſo wiſſe daß es *alle Kräuter*
ask you […] what the our German tea is so know that it every herbs
[ſeyn$_{RSG}$], die ein liebliches flüchtiges reines und ſtarckes Saltz haben
are that a lovely fleeting pure and strong scent have
als zum Exempel Ehren-Preiß […]Wachholder Beer ꝛc
as for example Veronica […] Juniper etc.
'If you ask which was our German tea, know, that it is every herb that has a lovely, fleeting, pure and strong scent as for example Veronica… Juniper and so on.'

(Abel, 1699)

Whether these apparent differences actually influence the positioning of the relative clauses or not is determined by means of logistic regression. For this, we used R (The R Core Team, 2018) and the lme4 package (Bates, 2015). This analysis shows that the differences in extraposition do not depend on the authors of the texts ($s^2 = 0.14$). Also, the variance of the mean surprisal values (average 2.68 for extraposed and 2.67 for embedded) has in fact no influence on the position of the RC ($z = 0.1$, $p = .92$). The cumulative surprisal value, on the other hand, is significant ($z = 1.7$, $p = .048$).

4.1.2 The Change of the Influence of Information Density in RCs over the Centuries

In order to test the second hypothesis, namely that the influence of information density weakens over the centuries, it is important to differentiate between the two time periods studied. It should be noted that only two periods are available for consideration. Whether there has been a fluctuation in the influence of information density in the meantime cannot yet be answered in our current study. An expansion of the material is planned, however.

In the first investigated period (1650-1700), we find 1 334 RCs. These are divided into 401 extraposed and 397 embedded RCs. The remaining 536 ambiguous RCs were sorted out due to the problems described in Chapter 3 and are not considered in the analysis. In the sentences that have a clearly identifiable RSB, we see an even distribution of the two types of relative clauses.

The distribution of relative clauses between the two genres is less even. It should be noted here that 373 RCs (46 % of occurrences) occur in the medical texts and 425 in the theological texts (54 % of occurrences). The two medical articles show a more uniform distribution of relative clauses. Abel records 181 RCs, Purmann 192 RC. These values are to be understood in each case after deduction of the ambiguous RCs.

The shortest RC is formed by three words, the longest, embedded RC measures 81 words. On average, the relative clauses are 11.73 words long. With regard to the surprisal measurements, the following picture emerges:

Table 3. Distribution of variables in question in RCs for the 17th century

Variable	Minimum	Maximum	Mean
Mean surprisal value	1.5	4.14	2.75
Cumulative surprisal value	4.8	202.497	31.11

Again, logistic regression was used to find out what has the greatest influence on extraposition. The distribution is again not attributable to the authors ($s^2 = 0.068$). The differences in mean surprisal values are also due to chance ($z = 0.7$, $p = .48$). The influence of the cumulative surprisal value is highly significant ($z = -4.47$, $p < .0001$). Since there is a direct relationship and

a direct influence of length in words of a relative clause and the cumulative surprisal value, the influence of length will also be investigated. This value is also indicated as significant (z = -4.42, p < .001), but the influence of length is only slightly stronger than that of the cumulative surprisal. A significantly higher influence of length on extraposition compared to the information density could not be found. Thus, we find strong evidence for the first hypothesis for the 17th century.

In the second period (1850-1900), we see a somewhat different picture. On the one hand, much more is extraposed: 468 (55.8 %) to 370 embedded RCs (44.1 %). There are also differences between the genres. In the medical texts, more than twice as many are extraposed (635, 75.7 %) than in the theological texts (203, 24.3 %), which is partly due to the text by Kraepelin (medical article, 471 relative clauses) and that of Hasak (theology), which contains only 65 relative clauses. The length of the RCs also differs slightly from that of the 17th century. The shortest RC is two, the longest one 66 words long. On average, authors of the late 19th century formulated RCs with a length of 11.23 words.

As Table 4 shows, the mean surprisal values vary between 1.09 and 4.42, the average is 2.6. These values are also very close to those of the 17th century. The cumulative surprisal values diverge from the 17th century. The lowest value is 4.38, the highest is 185.76, the mean is 28.12 (for comparison: in the other period examined, these values were 4.08, 202.5 and 31.11).

Table 4. Distribution of variables in question in RCs for the 19th century

Variable	Minimum	Maximum	Mean
Mean surprisal value	1.09	4.42	2.6
Cumulative surprisal value	4.38	185.76	28.12

For the 19th century, the influence on extraposition was also examined by means of a logistic regression. Here, the same picture emerges as in the 17th century: the variation of extraposition is not caused by the authors and their style of writing (s^2 = 0.21). However, a difference can be observed in the influence of the mean surprisal value. In the 19th century, it is significant (z = 2.32, p = .02). However, the cumulative surprisal value is highly significant (z = -6.955, p < .0001), which again confirms the first hypothesis in this period. The influence of length, measured in words, is also significant here (z = -7.104, p < .0001).

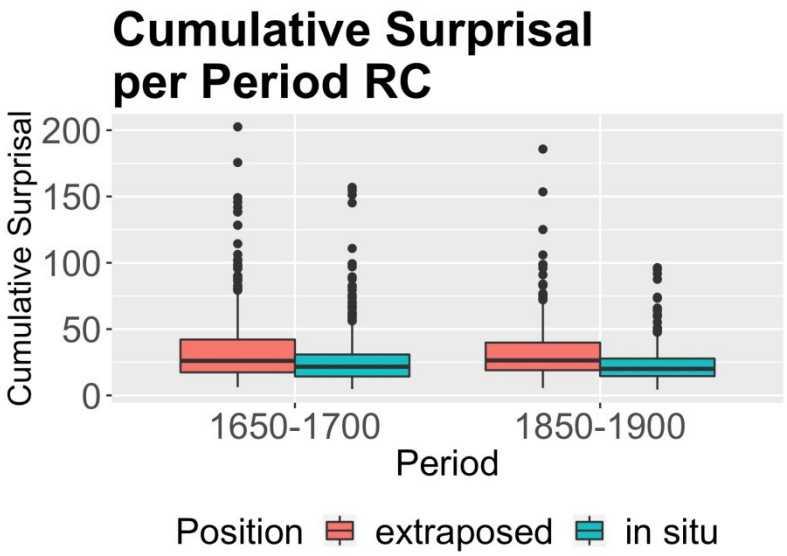

Figure 2. Cumulative RC surprisal values per period

With regard to the second hypothesis, it must be said that it is not confirmed for relative clauses. In both investigated time periods of the ENHG, a highly significant influence of information density on the positioning of RCs is shown, which is even stronger in the 19th century than in the 17th century. Possible reasons for this will be discussed in the last chapter.

4.2 Prepositional Phrases

For the investigation of PP extraposition, both periods should first be considered. The most important value here is the number of extraposed PPs. This is 317 in total divided into 179 extrapositions in the medical and 138 in the theological subcorpus. In order to create the minimal pair "extraposed – in situ", a total of 245 in situ PPs were sought. Care was taken to ensure that the prepositions matched in both cases, i.e. if a PP was extraposed starting with "mit" (*with*), an in situ "mit"-PP was chosen for comparison (8a, b).

(8) a. Also laſſet uns dann alle mit hertzlicher Andacht einander helffen [kåmpffen$_{RSB}$]
so let us then all with kind devotion eachother help fight

mit Gebet und Flehen.
with prayer and plea
'So let us help eachother to fight with kind devotion, with prayer and with plea.'

(Spener, 1676: 25)

b. Aber er hat mit dir den bund [gemacht$_{RSB}$]…
but he has with you the bond made…
'But he has bonded with you...'

(Spener, 1676: 116)

This was to avoid a possible influence of the preposition. In addition, an attempt was made to find an equivalent with the same preposition for PPs containing an RC. Under the condition that both sentence brackets are unambiguously filled, this was not always possible. Therefore, the 316 extraposed PPs are only matched by 245 in situ PPs.

Remarkable is the distribution of the evidence over the two time periods. In the 17th century, for example, three quarters of all records (426, 75.9 %) are found. Consequently, only 135 PPs (24.1 %) remain to be considered for the 19th century. The distribution between the genres also differs greatly: the medical texts contain 63 phrases (312 PPs) more than the theological ones (249 PPs).

The following picture emerges for the variables examined:

Table 5. Distribution of variables in question in PPs

Variable	Minimum	Maximum	Mean
Mean surprisal value	0.93	4.57	2.67
Cumulative surprisal value	0.99	210.26	18.99
Length in words	2	81	7.35

4.2.1 Correlation between PP Extraposition and Information Density

For this phenomenon, too, the values of the extraposed PPs should be compared with those of the embedded PPs. The medical texts contain 133 in situ PPs, the theological ones 112. The distribution over the two periods of investigation again shows a large difference. Between 1650 and 1700, 194 in situ PPs were annotated, two hundred years later only 51. The length of the extrapositions varies between two and 28 words with a mean of 4.63.

The cumulative and mean surprisal values are presented in Table 6:

Table 6. Distribution of variables in in situ PPs

Variable	Minimum	Maximum	Mean
Mean surprisal value	0.93	4.44	2.67
Cumulative surprisal value	0.99	72.46	11.75

We are interested foremost in extraposed PPs in contrast to in situ PPs. With those, too, we see a higher proportion in the medical texts (179 to 138), even if this does not differ significantly. This is in contrast to the distribution over the time periods. Again, the proportion is significantly higher in the early period, at 232, than in the late period, at 85 (73 % to 27 %). The length also differs significantly compared to the embedded phrases. It fluctuates between two and 81 words and is 9.45 words on average. The surprisal values also differ from the in situ variants (Table 7).

Table 7. Distribution of variables in extraposed PPs

Variable	Minimum	Maximum	Mean
Mean surprisal value	1.45	4.57	2.69
Cumulative surprisal value	2.9	210.26	24.59

Apparently, the difference between the extraposed and the in situ PPs in terms of their cumulative surprisal values (24.6 vs. 11.75) is very clear and possibly significant. Whether this is actually true is again clarified by the logistic regression.

As was the case with the RCs, we see that neither the author, nor the text type, nor the time period has any effect on extraposition. Also, the mean surprisal values, whose differences, as can be seen in the Table 7, are already very small even at first glance, do not influence the position of the PP. On the other hand, the cumulative surprisal values are again highly significant (z = -6.78, p < .001). The higher this value, the more likely it is that the PP is behind the RSB. Only the length has a slightly stronger influence on the extraposition (z = -6.95, p < .001), but this is also related to the fact that cumulative surprisal values by definition become larger when the set is longer. Nevertheless, we see the first hypothesis confirmed by the data.

4.2.2 The Change of the Influence of Information Density on PP Extraposition over the Centuries

In a second step, the data from the two time periods are again contrasted. In the 17th century, 232 extraposed and 194 embedded PPs were annotated. Again, the occurrence in the medical texts (273) is more frequent than in the theological texts (153). The length of the extraposed phrases is between two and 81 words, while the average length of the examined phrases in the 17th century is 7.7 words. The surprisal values are distributed as in Table 8:

Table 8. Distribution of variables in 17th century PPs

Variable	Minimum	Maximum	Mean
Mean surprisal value	0.3	4.57	2.76
Cumulative surprisal value	0.99	210.26	20.43

For the second period (1850-1900), there are again significantly fewer PPs behind the RSB. In the first period, there were 232, now there are only 84, which are opposed by 51 embedded PPs. The distribution between the medical texts and the theological ones is for the first time distributed in favour of theology. In this genre, there are 96 PPs, in the medical ones only 39 extraposed PPs (71 % to 29 %). Also, the length is clearly shorter than in the previous period. Again, the smallest extraposed PP, due to the fact that a PP must consist of at least two words, is also two words long, whereas the longest extraposed PP contains only 29 words. On average,

the annotated PPs in this time period are 6.2 words long. The surprisal values are given in Table 9.

Table 9. Distribution of variables in 19th century PPs

Variable	Minimum	Maximum	Mean
Mean surprisal value	1.01	3.99	2.38
Cumulative surprisal value	2.03	70.45	14.42

Again, a logistic regression was carried out on the basis of these values. In the 17th century, again no influence of the different authors and none of the mean surprisal values ($z = 0.47$, $p = .64$) is shown. A clearly different picture emerges for the cumulative surprisal values, which show a highly significant result ($z = -6.29$, $p < .001$). There is therefore a strong correlation between extraposition and cumulative surprisal values during this period. Only the length of the material has a slightly greater influence ($z = -6{,}78$, $p < .001$). We consider the first hypothesis to be confirmed when considering this period of time.

In the first two points, the results of the 19th century do not differ from those of the 17th century. Again, neither the authors nor the mean surprisal values have any influence on the position of the PP. For the mean surprisal values, the *z*-value is 1.51 and the *p*-value is .25. Again, the influence of the cumulative surprisal values is confirmed. However, it is significantly lower than in the previous period ($z = -2.548$, $p = .01$). The result for length is also significant ($z = -2.59$, $p = .009$).

The deterioration with respect to the *p*-values strengthens our second hypothesis, which speaks of a decrease in the influence of the ID on extraposition.

4.3 Noun Phrases

For the NPs, the picture is comparable with the difficulties encountered with PPs. There are 84 in situ NPs and 116 extraposed NPs, since care was also taken to ensure that the selected in situ NPs correspond to the extraposed NPs. This means that attention was paid to whether attributes were present or the NP was combined by relative clauses (9) or sequences of NPs.

(9) Den Anfang [sol$_{LSB}$] [machen$_{RSB}$]
the beginning shall make

[mein rothes Wundpflaster [dessen Beschreibung folgende (ist) $_{RC}$] $_{NP}$].
my red wound.plaster the description of which the following is
'The start shall be made by my red wound plaster which is described the following way.'

(Purmann, 1680)

Since an exact fit of the minimal pairs was not always possible, the result is a somewhat different picture. Also, the distribution of the NPs across the genre is remarkable: Again, in the medical texts, more extraposition is done than in the theological texts (119 NPs in the medical texts versus 81 in the theological ones). But the greatest difference is noticeable in the time periods. In the 17th century, we find 145 documents, in the 19th century only 55. The practice of putting nominal elements into the PoF obviously decreases drastically, which also corresponds to the linguistic perception of today's recipients. The examined ID values and the length of the NP shall be presented in tabular form here as well (Tables 10 and 11):

4.3.1 Correlation between NP Extraposition and Information Density

In order to answer the first hypothesis, in situ and extraposed NPs must be distinguished. We start with the NPs that clearly stand before the RSB. In the medical texts, 50 NPs appear, in the theological texts, 34 NPs were annotated. The distribution of the time periods also reflects the picture that can already be seen from the comparison of the time periods. In the first time period,

66 in situ NPs were annotated (78.6 %), in the last time period, 18 (21.4 %). The length of the NPs studied varies between two and 39 words, with an average length of 8.2 words.

Table 10. Distribution of variables in question in in situ NPs

Variable	Minimum	Maximum	Mean
Mean surprisal value	1.3	4.66	2.79
Cumulative surprisal value	2.67	105.14	18.99

Since the annotation of the data was attempted to have an equivalent in situ NP follow each extraposed NP, the distribution of the extraposed NPs is largely consistent with that of the in situ NPs. We find 47 NPs behind the RSB in the theological texts and 69 NPs behind the RSB in the medical texts. Again, NPs are more frequently extraposed in the 17th century than in the 19th century: 79 NPs vs. 37 NPs, which is a percentage of 68.1 % to 31.2 %, and which again suggests that the willingness to extrapose decreases over the centuries. The length of the material behind the RSB varies from one word to 50 words. The average length of an extraposed NP is 11.78 words. This distribution can be seen in the values assigned to information density (Table 11):

Table 11. Distribution of variables in question in extraposed NPs

Variable	Minimum	Maximum	Mean
Mean surprisal value	1.29	5.05	2.91
Cumulative surprisal value	1.29	134.18	30.89

Compared with the cumulative surprisal values of the embedded NPs, the surprisal value of the extraposed NPs is significantly higher on average. The mean surprisal value means, on the other hand, hardly differs between the two position variants. The difference in the length of the material of about three words is also rather small. Which of these variables has the greatest influence should again be clarified by the logistic regression.

Here, it is shown again that the cumulative surprisal value is significant ($z = 32.845$, $p < .01$). This confirms above all the influence of information density on the extrapositional process, which also confirms the first hypothesis for this phenomenon. Even the length of the extraposition, which is closely related to the cumulative surprisal value, reveals only a slightly lower p-value ($p = .0036$), so that it cannot be said to be the most important component of the extraposition.

4.3.2 The Change of the Influence of Information Density on NP Extraposition over the Centuries

In order for us to be able to test the second hypothesis as well, we need to take another look at the two time periods. In the 17th century, we find, as expected, significantly more extrapositions (145) than in the 19th century (55), which is also the reason for the different values in the distribution of NPs. Between 1650 and 1700, 79 NPs were placed behind the RSB and 66 before it. For the distribution by genre, the medical texts contain 78 NPs and the theological ones 67 NPs.

Table 12. Distribution of variables in 17th century NPs

Variable	Minimum	Maximum	Mean
Mean surprisal value	1.29	5.05	2.88
Cumulative surprisal value	1.29	134.18	26.36

The logistic regression shows only a weak significance for the 17th century in the summed surprisal values (z = -2.403, p = .0162). The differences in mean surprisal values can be attributed to chance ($p > 0.1$) and the different authors are also not decisive for the differences. Again, the close relationship between length and cumulative surprisal values is evident. The length of the NP is also reported as significant (z = -2.437, p = .0148), whereby the difference is again largely negligible. All in all, the first hypothesis is confirmed by the differentiated observation for the earliest period examined.

The 19th century shows clearly smaller values overall. Only 37 phrases are extraposed, 18 phrases that meet the criteria for embedding were annotated. These phrases are distributed very unevenly among the genres. We see 41 annotated NPs in the medical texts (about 75 %) and only 14 in the theological ones. A differentiated picture also emerges for the relevant factors compared to the other period under study. We see, for example, that the highest value of the summed up surprisal is almost half as large as that of the 17th century (63.7 to 134.17). On average, however, the extraposed phrases have higher cumulative surprisal values in the late period, but this is not necessarily reflected in the length. This varies between two and 28 words, averaging 11 words.

Table 13. Distribution of variables in 19th century NPs

Variable	Minimum	Maximum	Mean
Mean surprisal value	1.8	4.09	2.81
Cumulative surprisal value	3.60	63.71	28.14

In the 19th century, we see a clearly changed picture with regard to the influence of ID on extraposition. For none of the parameters, we find significant values. As before, the different authors do not have any influence on the position of the examined NPs. Also, the result of the mean surprisal value ($z < 1$) is not surprising after the previous investigations. However, the cumulative surprisal value also only reaches a p-value of .17 in the regression analysis, which means that the different position of the NP is due to chance with a probability of 83 %. The length of the extraposed material is also not significant for the position of the NP (p = .14).

This marked difference from the 17th century reinforces our second hypothesis, which states that the influence of the ID decreases over time. In the 19th century, other factors seem to influence NP extraposition.

5 Discussion

Let us summarize the results of the investigation again at this point. Using a corpus of scientific texts from the late 17th and 19th centuries, we have investigated whether there is a relationship between the position of relative clauses, prepositional phrases and NPs and the Shannon information density (1948). At the same time, the factor of the length of the extraposed material, which is not based on information density, but is often referred to as relevant, was examined for its effect.

The first hypothesis investigated whether RCs, PPs and NPs are more likely to be extraposed if they have high surprisal values. The surprisal value, which was mainly used for this purpose, is the cumulative surprisal value, i.e. the sum of all surprisal values of the lexical

elements in the investigated syntactically complex expression. If the values are not considered separately by time periods, the hypothesis can be confirmed for all phenomena. The higher the cumulative surprisal values are, the more likely linguistic material is pushed into the PoF. The *p*-values resulting from the investigation are usually highly significant, but not more significant than the results calculated based on the length of the material. However, this distinction is also due to the calculation of the cumulative surprisal values. Since all surprisal values are added together, the number of surprisal values that exactly corresponds to the number of words, that is our way of measuring length, favors a larger surprisal value. Furthermore, despite the low OOV-values when creating the language model, it is still possible that particularly high surprisal values may not be calculated because their bigram does not appear in the training corpus. The smoothing can only partially compensate for this problem. With this in mind, it may be worthwhile to eliminate the out-of-vocabulary problem or to use other types of smoothing.

The second hypothesis deals with the fact that the correlation between extraposition and ID decreases over the centuries. This is generally accompanied by the observation that at least phrasal extrapositions have become rare in contemporary German, especially in written discourse modes. In this respect, the analysis of the data reveals a confirmation of this hypothesis for PPs and NPs. In the 17th century, the correlation between cumulative surprisal values and the position of phrases was highly significant; in the 19th century, however, it was at most weakly significant. In contrast, this change is not evident in the relative clauses. We did not find evidence for our hypothesis there.

One reason for this is the clause status of the RC. By default, the PoF is the position in the sentence that is intended for clausal elements. This is already evident in Early New High German, although the variability of the subordinate clause position was even greater here. The position at the edge of the sentence thus seems to release a larger cognitive capacity, which is then available for understanding even syntactically complex elements, as can be seen from the fact that we also find a combination with sentences, especially RCs, in phrasal extrapositions.

The consistently high influence of the ID on the position of the RC in both time periods indicates a constant influence of the ID on this range. This could be extended in further studies by extending the study period to Middle or Old High German. In this case, the limitation to the genre of scientific texts would have to be removed in order to obtain a sufficiently large amount of data.

A possible explanation, which includes both the confirmation of the second hypothesis in the case of phrasal extraposition and the lack of evidence for this hypothesis with regard to relative clauses, is the establishment of the RSB in ENHG. In the earlier period, the RSB may not have been as firmly established in the consciousness of the authors in both the medical and theological texts as it was in the later period. This is evident from the overall high quota of clauses with material in the PoF (see Section 2.2). Extraposition into the PoF may therefore simply have occurred more easily than in the 19th century, when this sentence barrier was more firmly anchored in the writing practice and the consciousness of the people. An indication of this potential explanation can also be found in the fact that sentence boundaries in ENHG were generally not yet as fixed as in modern German. This characteristic makes the analysis of the UID in our complete sentences and a meaningful evaluation of the matrix sentence and RCs significantly more difficult in the early periods of time, which is also due to the sentence boundaries based on modern German, which cannot be directly transferred to ENHG.

Considering this type of explanation for phrasal extraposition, a more global scope of the ID is conceivable. If one assumes that the RSB has still not fully established itself as the conclusion of the sentence, it is more expectable and less surprising as a grammatical construction also in other places, so that an overall even curve would result, which corresponds to the UID (Levy & Jaeger 2007). This means that a language model based on POS tags would have to give lower surprisal values for the prepositions behind a POS tag marking the RSB in the earlier

periods than in the later periods. There should be no differences for the RCs. However, to pursue this idea, the POS tags of the DTA would have to be significantly better than they are.

In general, the study struggles with some corpus-related problems; the factors of record length have already been mentioned, as well as the faulty POS-tagging and the lemmatization, which is partly too much based on Modern German, influence the results presented here. It must also be emphasized that the analysis of the data presented is not yet complete. In order to be able to represent a real change over time, the time periods between 1700 and 1850 still have to be analyzed. This is the only way to confirm the second hypothesis. Nevertheless, we feel that information density can make a valuable contribution to the question of why elements are placed in the PoF.

Corpus

[Rotth, Albrecht Christian]: *Eylfertiges Bedencken über M. August Hermann Franckens [...] Seine Schutz-Predigt*. Halle, 1692.

Abel, Heinrich Kaspar: *Wohlerfahrner Leib-Medicus der Studenten*. Leipzig, 1699.

Egger, Augustinus: *Der christliche Vater in der modernen Welt. Erbauungs- und Gebetbuch*. Einsiedeln u. a., [1895].

Hasak, Max: *Die Predigtkirche im Mittelalter*. Berlin, 1893.

Koch, Robert: *Untersuchung über die Aetiologie der Wundinfectionskrankheiten*. Leipzig, 1878.

Kraepelin, Emil: *Ueber die Beeinflussung einfacher psychischer Vorgänge durch einige Arzneimittel*. Jena, 1892.

Purmann, Matthäus Gottfried: *Der rechte und wahrhafftige Feldscher*. Halberstadt, 1680.

Spener, Philipp Jakob: *Pia Desideria*. Frankfurt (Main), 1676.

References

Averintseva-Klisch, M. (2007). Anaphoric properties of German right dislocation. In M. Schwarz-Friesel, M. Consten & M. Knees (Eds.), *Anaphors in Text: Cognitive, formal and applied approaches to anaphoric reference* (pp. 165–182). Amsterdam: John Benjamins.

Aylett, M., & Turk, A. (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.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. Li (Ed.), *Subject and Topic* (pp. 27-55). New York, NY: Academic Press.

Chen, S., & Goodman, J. (1999). An empirical study of smoothing techniques for language modeling. *Computer Speech & Language, 13*(4), 359-393.

Coniglio, M., & Schlachter, E. (2015). Das Nachfeld im Deutschen zwischen Syntax, Informations- und Diskursstruktur. In H. Vinckel-Roisin (Ed.), *Das Nachfeld im Deutschen* (pp. 141-163). Berlin: De Gruyter.

Degeatano-Ortlieb, S. (2017). Variation in language use across social variables: A data-driven approach. In *Proceedings of the Corpus and Language Variation in English Research Conference (CLAVIER).* Bari.

Eckart de Castilho, R., Mújdricza-Maydt, É., Yimam, S., Hartmann, S., Gurevych, I., Frank, A., & Biemann, C. (2016). A web-based tool for the integrated annotation of semantic and syntactic structures. In *Proceedings of the Language Technology Resources and Tools for Digital Humanities (LT4DH) workshop at COLING* (pp. 76-84).

Frank, A., & Jaeger, F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Societey* (pp. 939-944).

Genzel, D. & Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics - ACL '02* (pp. 199-206). Philadelphia, PA: University of Pennsylvania.

Genzel, D. & Charniak, E. (2003). Variation of entropy and parse trees of sentences as a function of the sentence number. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing - EMNLP '03* (pp. 65–72). Stroudsburg, PA: ACL.

Gibson, Edward. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition 68*(1), 1-76.

Gibson, E., Futrell, R., Piantadosi, S., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences, 23*(5), 389-407.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies 2001 - NAACL '01*.

Hawkins, J. (1994). *A Performance Theory of Order and Constituency*. Cambridge, MA: Cambridge University Press.

Jaeger, T. F. (2005). Optional *that* indicates production difficulty: Evidence from disfluencies. In *Proceedings of DiSS'05. Disfluency in Spontanous Speech Workshop* (pp. 103-108). Aix-en-Provence.

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

Koch, P., & Oesterreicher, W. (2007). Schriftlichkeit und kommunikative Distanz. *Zeitschrift für germanistische Linguistik, 35*(3), 346-375.

Krifka, M. (2007). Basic notions of information structure. In C. Féry & M. Krifka (Eds.) *Interdisciplinary Studies on Information Structure (Volume 6)* (pp. 13-55). Potsdam: Universitätsverlag Potsdam.

Lenerz, J. (2001). Word order variation: Competition or co-operation. In G. Mueller & W. Sternefeld (Eds.), *Competition in Syntax* (pp. 249-281). Berlin: De Gruyter.

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

Levy, R., & Jaeger, F. (2007). Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems, 19*, 849-856.

Müller, G. (1999). Optimality, markedness, and word order in German. *Linguistics, 37*(5), 777-818.

Mürmann, M. (2014). *Wahrscheinlichkeitstheorie und Stochastische Prozesse*. Berlin: Springer Spektrum.

Prince, E. (1981). Towards a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics* (pp. 223-255). New York, NY: Academic Press.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reis, M. (1987). Die Stellung der Verbargumente im Deutschen: Stilübungen zum Grammatik:Pragmatik-Verhältnis. *Sprache und Pragmatik, 5*, 139-178.

Sahel, S. (2015). Zur Ausklammerung von Relativsätzen und Vergleichsphrasen im frühen Neuhochdeutschen (1650-1800). In H. Vinckel-Roisin (Ed.), *Das Nachfeld im Deutschen: Theorie und Empirie* (pp. 165-183). Berlin: De Gruyter.

Sapp, C. (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 (Ed.), *Zur Ausbildung der Norm in der deutschen Literatursprache auf der syntaktischen Ebene* (pp.235-284). Berlin: Akademie-Verlag.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379-423.

Speyer, A. (2016). Die Entwicklung der Nachfeldbesetzung in verschiedenen deutschen Dialekten. In A. Speyer & P. Rauth (Ed.), *Syntax aus Saarbrücker Sicht 1. Zeitschrift für Dialektologie und Linguistik, Beiheft 165* (pp. 137-157). Stuttgart: Steiner.

Tipler, P., Mosca, G., Basler, M., Kersten, P., Dohmen, R., Wagner, J., Heinisch, C., Schleitzer, A., & Zillgitt, M. (2015). *Physik*. Berlin: Springer Spektrum.

Uszkoreit, H., Brants, T., Duchier, D., Krenn, B., Konieczny, L., Oepen, S., & Skut, W. (1998). Studien zur performanzorientierten Linguistik. *Kognitionswissenschaft, 7*(3), 129-133.

Vinckel-Roisin, H. (2015). Facetten des Nachfelds im Deutschen. In H. Vinckel-Roisin (Ed.) *Das Nachfeld im Deutschen: Theorie und Empirie* (pp. 1-22). Berlin: De Gruyter.

Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change, 9*(1), 81-105.

Wasow, T., & Jaeger, T. (2011). Lexical variation in relativizer frequency. In H. J. Simon & H. Wiese (Eds.), *Expecting the Unexpected: Exceptions in Grammar* (pp.175-196). Berlin/ New York, NY: De Gruyter Mouton.

Weber, S. (2019). The acceptability of extraposition of PPs out of NP in German. In A. Gattnar, R. Hörnig, M. Störzer & S. Featherston (Eds.) *Proceedings of Linguistic Evidence 2018: Experimental Data Drives Linguistic Theory* (pp. 63 – 84).

Wöllstein, A. (2010). *Topologisches Satzmodell*. Heidelberg: Universitätsverlag Winter.

Zifonun, G., Hoffmann, L., & Strecker, B. (1997). *Grammatik der deutschen Sprache, Vol. 2*. Berlin: De Gruyter.

Zubin, D., & Köpcke, K. (1985). Cognitive constraints on the order of subject and object in German. *Studies in Language, 9*(1), 77-107.