Information structure and information theory

Edited by

Robin Lemke

Lisa Schäfer

Ingo Reich

Topics at the Grammar-Discourse Interface 10

language science press

Topics at the Grammar-Discourse Interface

Editors: Philippa Cook (University of Göttingen), Anke Holler (University of Göttingen), Cathrine Fabricius-Hansen (University of Oslo)

In this series:

1. Song, Sanghoun. Modeling information structure in a cross-linguistic perspective.
2. Müller, Sonja. Distribution und Interpretation von Modalpartikel-Kombinationen.
3. Bueno Holle, Juan José. Information structure in Isthmus Zapotec narrative and conversation.
4. Parikh, Prashant. Communication and content.
5. Balogh, Kata, Anja Latrouite & Robert D. Van Valin, Jr. (eds.) Nominal anchoring: Specificity, definiteness and article systems across languages.
6. Næss, Åshild, Anna Margetts & Yvonne Treis (eds.). Demonstratives in discourse.
7. Gergel, Remus & Jonathan Watkins (eds.). Quantification and scales in change.
8. Nakagawa, Natsuko. Information structure in spoken Japanese: Particles, word order, and intonation.
9. Same, Fahime. Referring expression generation in context: Combining linguistic and computational approaches.
10. Lemke, Robin, Lisa Schäfer & Ingo Reich (eds.). Information structure and information theory.

ISSN: 2567-3335

Information structure and information theory

Edited by

Robin Lemke

Lisa Schäfer

Ingo Reich

language
science
press

Robin Lemke, Lisa Schäfer & Ingo Reich (eds.). 2024. *Information structure and information theory* (Topics at the Grammar-Discourse Interface 10). Berlin: Language Science Press.

This title can be downloaded at:
http://langsci-press.org/catalog/book/465

ISBN: 978-3-96110-481-9 (Digital)
978-3-98554-110-2 (Hardcover)

ISSN: 2567-3335
DOI: 10.5281/zenodo.12784266
Source code available from www.github.com/langsci/465
Errata: paperhive.org/documents/remote?type=langsci&id=465

Cover and concept of design: Ulrike Harbort
Proofreading: Amy Amoakuh
Fonts: Libertinus, Arimo, DejaVu Sans Mono
Typesetting software: XeLaTeX

Language Science Press
xHain
Grünberger Str. 16
10243 Berlin, Germany
http://langsci-press.org

Storage and cataloguing done by FU Berlin

Contents

Chapter 1

Information structure and information theory: A short introduction

Ingo Reich[a], Robin Lemke[a] & Lisa Schäfer[a]

[a]Universität des Saarlandes

This introduction sets the ground for the contributions to this volume. In a first step, we will try to convince the reader that both information structural and information theoretical considerations are relevant to a deeper understanding of how we linguistically encode the message that we want to get across. Since this volume brings together two different strands of research, research on information structure and research on information theory, we will introduce the key notions of both approaches in a second step, and illustrate with the example of non-canonical word order in German how both approaches try to account for the observations in question. We take this as an opportunity to reflect more generally on the possible relationships between information structure and information theory. As is usual, this introduction concludes with short descriptions of each contribution.

1 Introduction

It is uncontroversial that communication is, at least to a large extent, about the exchange of information: If, in a café, I order an espresso with the words *An espresso, please*, then I inform the addressee that I would like to have an espresso, and not, for example, a chai latte. And if the waiter replies *2.60*, then he informs me that this is the price for an espresso (and that he wants me to pay it). Thus, communication is – again, at least to a large extent – about the world, and the information conveyed is (mostly) propositional.

The way we encode this propositional information crucially depends on what we think the addressee knows or takes for granted: If I return to the counter, and

Ingo Reich, Robin Lemke & Lisa Schäfer. 2024. Information structure and information theory: A short introduction. In Robin Lemke, Lisa Schäfer & Ingo Reich (eds.), *Information structure and information theory*, 1–22. Berlin: Language Science Press.
DOI: 10.5281/zenodo.13383783

ask the same waiter for another espresso, I will most likely use the words *Another espresso, please* simply because I know that the waiter knows that I already had an espresso before. But if I return to the counter, and ask a different waiter for another espresso, then I cannot be fully sure that this waiter knows that I had an espresso before, and there is a good chance that I place my new order with exactly the same words as before: *An espresso, please*. (Though by uttering *Another espresso, please* I might intend to inform the waiter that I already had an espresso before.) Thus, what happened and what was said before guides us as speakers in choosing a specific linguistic encoding for the information that we want to get across. And this is not only true for the choice between *a* and *another* but also for the choice between active and passive or between alternative word orders. In other words, the previous linguistic and non-linguistic context shapes the way the propositional information encoded in an utterance is presented, it determines the utterance's INFORMATION STRUCTURE.

Of course, in the same-waiter scenario above, I could just as easily say *Another one, please*, since, when hearing *Another* and at the same time knowing that I ordered an espresso before, the waiter can easily predict that the noun following *Another* will refer to espresso in one way or another. Thus, in this utterance, the noun *espresso* is obviously much less informative to the waiter as compared to its occurrence in my previous utterance *An espresso, please*, where my order of an espresso was not yet foreseeable and the explicit mention of the noun *espresso* was therefore crucial for the understanding of my order. However, the observation that different occurrences of one and the same word can differ in informativity cannot be (easily) accounted for in terms of propositional information: the propositional content of the word *espresso*, its denotation, is always the same. Rather, it is its predictability in context that varies from utterance to utterance, from occurrence to occurrence. This notion of predictability in context thus relates to a notion of information as developed in INFORMATION THEORY that is probabilistic in nature and orthogonal to a propositional understanding of information. But at the same time, it is exactly the predictability of the customer ordering espresso in the context of *Another* that made me choose *one* over *espresso* in the context at hand.

Thus, it is not only the relation of propositional information to a propositional common ground that guides us as speakers in choosing a specific linguistic encoding but also the degree to which an expression in an utterance is predictable from its linguistic and non-linguistic context. This raises the questions of how these two different notions of information relate to each other, whether they are in fact completely independent from each other, and if they are not, in what

way they interact in determining an actual linguistic encoding.[1] These are the questions this volume wants to address in one way or another.

Since these questions bring together two different strands of research, the linguistic tradition of propositional semantics and information structure and the computational, and to some extent also psycholinguistic, tradition of information theory, we first introduce the key notions of each strand in the following two sections on information structure and information theory. In a third section, we take up the above questions and illustrate the relevance of both notions with word order in German as an example. This introduction concludes, as is customary, with short summaries of the contributions to this volume.

2 Information structure

Information structure, also termed information packaging in Chafe (1976), deals "with the relation of what is being said [and how it is being said, *the authors*] to what has gone before in the discourse, and its internal organization into an act of communication" (Halliday 1967: 199). During this process, i.e., when "packaging" information, the speaker takes the temporary state of mind (the current cognitive state) of their addressee into account (Chafe 1976: 28). According to Krifka (2007) and many others, this can be modeled by building on Stalnaker's (1978) concept of COMMON GROUND (CG).[2] By assumption, the CG consists of the set of propositions that both speaker and hearer assume or believe to be true at a certain point in the discourse. The CG thus represents the shared background between the interlocutors in the form of presupposed propositions. From this perspective, the field of information structure is all about how the relation to the current CG impacts the actual linguistic encoding of the message that the speaker wants to get across. Traditionally, three concepts are taken to be crucial in this respect:

[1] That the two notions in question are in fact related in a non-trivial way is already suggested by the fact that what we know about the addressee and the utterance situation adds to the predictability of the customer ordering espresso in the context of *Another*. Thus one might alternatively propose, and in fact it has been proposed in the literature, that switching to the encoding *Another one, please* in the same-waiter scenario is simply licensed, because the noun *espresso* has already been mentioned in the discourse before (i.e., because it is GIVEN in the immediately preceding context; see Section 2 for details), and it is motivated in order to avoid redundancy. Because of this intricate relationship, it is actually hard to decide which approach is the more promising one. What needs to be looked into here, is, whether the determiner *another* substantially adds to the possibility of using *one*, whether we need a gradual notion of predictability/givenness, and how we operationalize a notion like redundancy.

[2] Stalnaker (1978) actually attributes the term to H. Paul Grice, who used it in the William James Lectures in the form of the common ground status of propositions (Grice 1989).

givenness, focus and topicality. In the following three subsections, we will try to flesh out the core ideas behind these concepts.

2.1 Givenness

We will start with the concept of givenness. Traditionally, given (also termed *old*, *known*) is defined categorically in opposition to new, but authors often assume further gradations such as inferrable or accessible for entities which are neither completely new nor completely given (e.g. Chafe 1976, Prince 1981, Lambrecht 1994). With Krifka (2007: 37), we can distinguish two types of givenness, which we can term referential givenness and givenness by entailment.

Referential givenness concerns all kinds of referential expressions like noun phrases with an indefinite or definite article, personal pronouns or clitics. These referential expressions typically come with morpho-syntactic features as part of their lexical specification that indicate whether the referent is or is not present in the immediate CG and, if so, to what degree. For these referential expressions, authors like Prince (1981), Ariel (1990) and Gundel et al. (1993) have developed hierarchies along which these expressions are ordered according to their degree of givenness in the CG. These hierarchies make use of concepts like familiarity or identifiability, which relate to the salience of referents in the CG and the presumed degree of their cognitive activation in the mind of the addressee. In order to model referential givenness, Stalnaker's notion of CG needs to be enriched in such a way that we can keep track of the referents that have been introduced into the discourse, and that allows for dynamic updates of their activation status as the discourse proceeds (see Krifka (2007) for discussion).

This is probably not much different with givenness by entailment. Givenness by entailment concerns virtually all kinds of sentential and sub-sentential expressions with the exception of referential and functional phrases (like conjunctions, complementizers or articles). The basic idea is that such expressions typically denote propositional information and that this information (and thereby the expression denoting this information) can be taken to be given at a certain point in the discourse, if the CG entails this propositional information at this point in the discourse (see Schwarzschild (1999) for an elegant implementation of this idea). However, also in the case of givenness by entailment, one might want to keep track of the dynamically changing salience of propositional information in the CG, since propositions that have just been added to the CG are arguably much more salient and relevant to givenness than propositions that are part of the CG but currently not activated in the mind of the addressee. Givenness by entailment

is typically marked by deaccentuation of the expression in question or even by its deletion.

Givenness is closely linked to the second key concept of information structure, FOCUS, for it is often taken for granted that all expressions that are not given in one way or another are in focus (see again, e.g., Schwarzschild 1999).

2.2 Focus

In terms of content, focus is often associated with the concept of new information and thus with the opposite pole to givenness (see, e.g., the notion of information or presentational focus in É. Kiss 1998). In the tradition of the Prague school (see e.g. Mathesius 1975), focus in the sense of new(er) information is also called the RHEME or rhematic information (in opposition to the THEME or thematic information). Building on Rooth (1985, 1992), however, Krifka (2007) and many others propose to interpret focus exclusively as a means to indicate (contextual) alternatives. Krifka (2007: 20) distinguishes two kinds of focus: first, focus that indicates alternatives to linguistic expressions (expression focus), as is illustrated with the correction of a pronunciation in (1). Second, and more importantly, focus that indicates alternatives to the semantic denotation of a linguistic expression (denotation focus). The latter use is illustrated with a question-answer pair in (2) where *John* is marked as one of several possible people who have been called.

(1) a. They live in BERlin.
    b. They live in [BerLIN]$_F$!

(2) a. A: Who did you call?
    b. B: I called [JOHN]$_F$.

In languages like English and German, focus is typically marked with prosodic prominence in the form of pitch accents (see e.g. Pierrehumbert & Hirschberg 1990), as is indicated through capitalization in (1) and (2). Importantly, words with such a pitch accent can signal not only that they themselves are focused, but also that larger constituents or phrases in which they are contained are in focus, as exemplified in (3). This is called FOCUS PROJECTION, for example in Höhle (1982), and can give rise to ambiguities (Selkirk 1984, 1995).

(3) a. A: What did Mary do?
    b. B: She [[praised]$_F$ [her [BROther]$_F$]$_F$]$_F$

Denotation focus can be used pragmatically and semantically. Semantic uses of denotational focus are typically associated with focus-sensitive expressions like focus particles (e.g. *only*, *also*, *even*) and affect the truth-conditional content of the utterance in question, and consequently, after updating, also the content of the immediate CG. Pragmatic uses of denotational focus, on the other hand, like highlighting the answer to a *wh*-question, do not affect the truth-conditional content of the utterance, and thus rather serve what Krifka (2007) calls the management of the CG, that is the way the CG content is conversationally organized, the way utterances relate to previous utterances and the way the CG content is supposed to develop according to the communicative needs of the participants. Another prominent case of CG management is CONTRASTIVE FOCUS in corrections like (4), where the utterance in (4b) directly relates to the one in (4a).

(4) a. They live in BerLIN.
    b. No, they live in [CoLOGNE]$_{F}$.

In this introduction, we can only highlight the most prominent distinctions with respect to focus. For related uses like *exhaustive* or *verum focus*, we must refer the reader to the relevant literature (e.g., the overview in Krifka 2007).

2.3 Topicality

The notions of givenness and focus are frequently taken to essentially partition a sentence into at least two (not necessarily continuous) parts. This is even more true of another dichotomy, the one between topic and comment. The term *topic* goes back to Hockett (1958: 201) and captures the intuition that the information conveyed in a sentence is typically information about some person or object:[3] "the speaker announces a topic and then says something about it." In the example sentence (5), it is *John* who is announced as topic and *ran away* is the comment, i.e., what is said about *John* (Hockett 1958: 201).

(5) John ran away. (Hockett 1958: 201)

Among the most influential definitions of topic in modern linguistics is the one proposed in Reinhart (1981), who follows Hockett (1958) in conceiving of topics

[3]The terms *topic* and *comment* have become established in modern linguistic research since Hockett (1958), but the concepts behind them, i.e., the separation between what the utterance is about and what is said about this element, are much older. They go back to Aristotle's distinction between subject and predicate with the predicate saying something about the subject (see, e.g., Lambrecht 1994) and were reintroduced into linguistics much later by von der Gabelentz (1868) and Paul (1919) as psychological subject and psychological predicate.

as entities (persons or objects) referred to in a sentence, about which something is said in the very sentence. But at the same time Reinhart relates this concept to Stalnaker's common ground (CG), and illustrates this relation through the metaphor of a library catalog (Reinhart 1981: 79–80): The ordered library catalog represents the CG and the book-entries in this catalog are the propositions contained in the CG. The process of entering a new book into the catalog corresponds to adding a proposition to the CG. Just like the book is stored under a specific entry, the proposition is stored under an entry which corresponds to the topic of the sentence in question. In (5), for example, the proposition *that John ran away* is stored in the CG under the referent of *John*.[4]

Based on this metaphor, Krifka (2007: 41) defines *topic* as follows: "The topic constituent identifies the entity or set of entities under which the information expressed in the comment constituent should be stored in the CG content". Krifka (2007: 41–42) objects against two equations that are occasionally made in the literature: He clarifies that even though topics are often given information they do not necessarily have to be and that there is no one-to-one correspondence between topic/comment and focus/background either. In specifying the concept further, he assumes that sentences typically have exactly one topic, but following Lambrecht (1994) and opposing Reinhart (1981), he states that there might also be sentences with more than one topic constituent or so-called thetic sentences with no topic constituent at all (Krifka 2007: 42–43). Another typical co-occurrence is that between the topic and the grammatical function subject. According to Lambrecht (1994: 132), subjects can be considered to be unmarked topics in many languages, even if there are of course topics that are not subjects and vice versa (see also Reinhart 1981: 62).

Since topics are typically given, they are typically not in focus and thus typically also not accented (or even deaccented). A prominent exception to this rule are CONTRASTIVE TOPICS (e.g., Büring 1997), see (6) for illustration. The question in (6a) sets Ann and Bill as topics for the following discourse, but in the answer (6b), both *Bill* and *Ann* still carry a (rising) accent.

(6) a. What grade did Ann and Bill get?
b. BILL got a C, but ANN got an A, of course.

[4]We saw in Section 2.1 that, in order to account for referential givenness, Stalnaker's notion of CG needs to be enriched in one way or another to keep track of the referents that have been introduced into the previous discourse. This extended notion of CG comes in handy, when we take Reinhart's metaphor seriously and implement it in a formal framework. Also, such an extension of Stalnaker's notion of CG allows us to model concepts like topic continuity or topic shift across discourse, as suggested, for example, in centering theory (Walker et al. 1998).

Contrastive topics, however, might not be that special after all, since "[t]hey arguably do not constitute an information-packaging category in their own right, but represent a combination of topic and focus, as indicated in the example, in the following sense: They consist of an aboutness topic that contains a focus, which is doing what focus always does, namely indicating an alternative. In this case, it indicates an alternative aboutness topic" (Krifka 2007: 44).

3 Information theory

Information-theoretic approaches explain optional linguistic variation without resorting to meaning from a purely probabilistic perspective. Originally, Shannon (1948) did not intend to apply information theory to the production and comprehension of natural language but to the transmission of signals in technical systems from an engineering perspective. Nevertheless, more recently, linguists have applied key ideas of information theory to actual language use. According to Shannon (1948), communication consists in a sender sending a message to a receiver across a noisy channel. For this purpose, the message is encoded by the sender into a linguistic signal, which is decoded by the recipient, and it can be corrupted by noise in the transmission process. In order to be efficient, the sender should convey as much information as possible in a given unit of time across the channel. The channel, however, has only a limited capacity and exceeding this capacity increases the noise rate above the expected efficiency gain by making the message more dense. In order to counterbalance noise, the sender can encode the message in a more redundant fashion, which increases the chance of getting the message across, but also results in a longer (and a priori less efficient) signal. As a consequence, an efficient sender will actively modulate the degree of redundancy in the signal so that the information communicated approximates the channel capacity without exceeding it.

3.1 A probabilistic notion of information

Now, how is the information conveyed by a signal measured? Shannon proposes a purely probabilistic notion, according to which the information conveyed by a signal (say, a word) is equivalent to the negative logarithm of the probability of the event that this signal occurs in a given context. According to the definition in equation (7), the information of a signal (measured in bits) is higher, the less likely the signal is, and would equal 0 if the signal were perfectly predictable.

(7) $$\text{Information}(\text{signal}) = -\log_2 P(\text{signal} \mid \text{context})$$

The occurrence of *dogs* in the context of *it rains cats and* is, for example, highly predictable, and its Shannon information thus very low. The occurrence of *dogs* in the context of *yesterday, I bit some*, on the other hand, is rather unlikely, therefore highly unpredictable (surprising) and highly informative.

3.2 Information theory and language use

Both the concept of communication across a noisy channel and the probabilistic definition of information can be straightforwardly applied to human language. Intuitively, we encode our utterance in a more redundant or explicit fashion if the noise ratio is high. This comprises acoustic noise, like a nearby construction site or departing train as well as other sources of noise that possibly impede communication, like a distracting task, the listener being a recent learner of the language etc. Defining information solely in probabilistic terms at first glance clashes with a semantic concept of information, e.g., a sentence communicating a proposition that is added to the common ground (CG), but it is actually somewhat related: Communicating a rather predictable sentence like (8a) invokes a smaller update in our mental representation of the world (given the reasonable assumption that there are more accessible worlds in which people order pizzas topped with mozzarella than pizzas topped with french fries) than a sentence containing an unpredictable one instead (8b). Therefore, an unpredictable word or sentence, which has a high Shannon information, will often also be perceived as conveying more semantic information.

(8) a. She ordered a pizza topped with mozzarella.
    b. She ordered a pizza topped with french fries.

3.3 Distributing information uniformly across utterances

Applying information theory to language thus predicts that speakers modulate the redundancy in the utterance to communicate efficiently by approaching but not exceeding channel capacity. This idea has been proposed for different levels of linguistic analysis under labels such as *Constant flow of information* (Fenk & Fenk 1980), *Smooth signal redundancy* (Aylett & Turk 2004) or *Uniform information density* (Levy & Jaeger 2007). Speakers can adapt their utterances in at least three ways to achieve this overall goal.

First, if lengthy utterances are assigned to predictable meanings, it is reasonable to make the signal denser by shortening them. This results in more information being communicated in the same amount of time, making the most efficient

use of the available channel capacity. Shortening can occur on different levels of linguistic analysis, for instance, by increasing the speech rate (Aylett & Turk 2004), contraction (*don't*, Frank & Jaeger 2008), pronominalization (Tily & Piantadosi 2009) or ellipsis (Levy & Jaeger 2007, Lemke 2021, Schäfer 2021).

Second, speakers should also avoid unpredictable words or expressions when they exceed the capacity of the channel. In this case, speakers can increase the redundancy of the utterance at the critical position by reducing their speech rate or choosing more redundant forms, like complete DPs instead of pronouns or non-elliptical variants of the utterance. Additionally, they can exploit the fact that the linguistic context contributes to the predictability of words (Levy 2008) when planning their utterance. For instance, inserting a determiner before a noun in newspaper headlines, where determiner omission is often possible but not required (Stowell 1991, Reich 2018), reduces the likelihood of an otherwise unpredictable noun just because it is very likely that any noun will appear after the determiner.

Third, changes in word order can affect the distribution of information, as the contrast in (9) illustrates. Since *to buy* does not require a recipient, *John* will be rather unexpected in (9a). In (9b), however, a recipient is much more likely if we know that Mary bought a present, and if *John* is introduced by a preposition (since this further boosts the likelihood of an animate recipient). Therefore, (9b) probably has the more uniform distribution of information.

(9) a. Mary bought John a present.
b. Mary bought a present for John.

For more on how this can be applied to various subfields of linguistics, see Crocker et al. (2016), Jaeger & Buz (2017) and the following case study on word order in German.

4 Non-canonical word order in German

In contrast to English, German has a comparably rich inflectional system, which permits a rather transparent marking of case. At the same time, German also allows, in principle, for different serializations of subjects and objects. In neutral contexts (in the context of the question *What happened?*) the subject typically precedes the indirect object (IO), and the indirect object typically precedes the direct object (DO), see example (10) from Lenerz (1977). This serialization of the

verb's arguments is usually considered to be the canonical word order, since it allows for focus projection from the DO (e.g., Höhle 1982).[5]

(10) [*Der Mann hat* [*dem* *Kassierer*]$_{IO}$ [*das* *GELD*]$_{DO}$ *gegeben.*]$_{F}$
The man has the-DAT cashier the-ACC money given
'The man gave the cashier the money'

As Lenerz (1977) observes, the canonical word order is consistent with the DO being focused narrowly (in the context of the question *What did the man give the cashier?*), see (11), but also with the IO being focused narrowly (in the context of the question *Who did the man give the money?*), see (12).

(11) *Der Mann hat* [*dem* *Kassierer*]$_{IO}$ [*das* *GELD*]$_{DO,\ F}$ *gegeben.*
The man has [the-DAT cashier] [the-ACC money] given

(12) *Der Mann hat* [*dem* *KasSIErer*]$_{IO,\ F}$ [*das* *Geld*]$_{DO}$ *gegeben.*
The man has the-DAT cashier the-ACC money given

However, as Lenerz (1977) also observed, this is different, when it comes to non-canonical word order in German. If the DO *das Geld* ('the money') precedes the IO *dem Kassierer* ('the cashier'), it is perfectly fine to put a narrow focus on the IO (in the context of the question *Who did the man give the money?*), see (13), but the result is degraded if we put a narrow focus on the DO (in the context of the question *What did the man give the cashier?*), see (14).

(13) *Der Mann hat* [*das* *Geld*]$_{DO}$ [*dem* *KasSIErer*]$_{IO,\ F}$ *gegeben.*
The man has the-ACC money the-DAT cashier given

(14) *? *Der Mann hat* [*das* *GELD*]$_{DO,\ F}$ [*dem* *Kassierer*]$_{IO}$ *gegeben.*
The man has the-ACC money the-DAT cashier given

Thus, a generalization emerges to the effect that in the case of non-canonical word order in German, the preceding DO must not be more rhematic (i.e., encode newer information) than the following IO (see Lenerz 1977: 45). Or if we want to put it somewhat more strongly, the preceding DO must be more given than the following IO. This shows that givenness is a relevant factor for non-canonical word order in German,[6] and it suggests, as a rule of thumb, that generally (more)

[5] Though this in fact depends on properties of the verb in question, see Höhle (1982).

[6] This is not to say, of course, that givenness is the only relevant factor for non-canonical word orders. There are quite a few other factors at stake here, some less related to givenness (e.g. the syntactic function of the arguments or animacy), and some more (e.g. the in/definiteness of the noun phrases in question), see Rauth (2020) for a recent overview.

given information tends to precede new(er) information.[7] Or if we want to put it in a nutshell: "Serialize given before new!"

This directive is confirmed if we consider the serialization of pronouns or the positioning of topics. In German, weak pronouns typically precede full noun phrases and are preferably positioned in the left periphery of the German middle field (Wackernagel 1892), as illustrated in (15).

(15) *Der Mann hat* [*es*]$_{DO}$ [*dem KasSIErer*]$_{IO}$ *gegeben (das Geld).*
The man has it the-DAT cashier given (the money)

In a similar fashion, topics also tend to be positioned at the left periphery of the middle field or the sentence, i.e., in the prefield. In the context of the question *What did the man do with the money?*, which arguably sets the money as the topic for the following utterance, the DO preferably precedes sentence adverbials like *vermutlich* ('presumably'), see (16) and the discussion in Frey (2000). And if the subject is less given than the topic, say in the context of the question *What happened to the money?*, the topic may even shift to the prefield, see (17).

(16) *Er hat* [*das Geld*]$_{DO}$ *vermutlich* [*dem KasSIErer*]$_{IO}$ *gegeben.*
He has the-ACC money presumably the-DAT cashier given

(17) [*Das Geld*]$_{DO}$ *hat* [*jemand*] [*dem KasSIErer*]$_{IO}$ *gegeben.*
The money-ACC has someone the-DAT cashier given

In the context of the "given before new" directive this is not very surprising: Both pronouns and topics are typically given to a rather high degree.

As we just argued, one of the crucial factors guiding non-canonical word order in German can be described in information-structural terms: given information typically precedes new information. This raises, of course, the question of why this should be so. Possibly, an answer to this question can be given in information-theoretic terms: In Section 3, we saw that in information theory, information is not defined in terms of denotations but based on the probability of an event *e*, like uttering a certain word *w* in a given context *c* (say, in the context of a string of other words). This is to say that, in principle, each expression that is part of the relevant context *c* contributes to the probability of *w* occurring in this very context. The string of words *it rains cats and*, for example, makes *dogs*

[7] It is important to keep in mind though that in the case of canonical word order, a more rhematic IO preceding a less rhematic DO is fine. That is, the givenness constraint does not necessarily trigger a reordering of the verb's arguments. This suggests that there is a kind of trade-off between the processing advantages of conforming to the given before new directive on the one hand, and the effort involved in the reordering process on the other hand.

highly predictable, the context *yesterday, I bit some*, on the other hand, does not. Now, since given expressions are easily accessible from the linguistic and/or non-linguistic context, they are also quite predictable in information-theoretic terms (i.e., in general, givenness contributes to predictability) and thus less informative (compared to contexts in which they are not given). Serializing given expressions before new(er) ones therefore makes sense, since this way, the given expressions are part of the context that predicts the new(er) ones, and thus serves to lower, in general, their informativity. The consequence is generally a smoother information profile, which is, as we saw above, advantageous for efficient and successful communication. Fenk-Oczlon (1989: 519) puts it as follows:

> It is precisely for this reason that what has already appeared in the preceding discourse, that is, what is 'old' and familiar in the textual context, or in the given context of action or situational context, bears less subjective information than a 'new' element in the same context (Fenk-Oczlon 1983a). In this context it is more expectable, its analysis requires fewer cognitive costs. It is in the interest of an economical and constant flow of information to place such informationally poor elements at the beginning of a sentence (and perhaps also of a phrasal conjunct?), because as the sentence or phrase progresses there is in any case a significant reduction of information (= a constriction of the permissible possibilities for continuing).

Let us illustrate this with the following example: Suppose someone asks *Why did you walk all the way to the university?*, and you respond with (18).

(18) *Ich habe* [*mein Auto*]$_{DO}$ [*meinem Nachbarn*]$_{IO}$ *verkauft.*
I have my car my neighbor sold
'I sold my car to my neighbor.'

The question implicitly raises another question, namely the question of why you did not take your car (as you probably usually do). This way, your car is made salient through the question (via bridging), and in this sense your car is given in the context of the question. This is different with your neighbor. Your neighbor is new information. But parsing the DO *my car* in the utterance makes all those things salient that one can do with cars, including selling them to somebody (which is, by the way, a good reason to walk to the university). This way, the utterance of *meinem Nachbarn verkauft* becomes far more predictable (and thus less informative) compared to an utterance in which *meinem Nachbarn* ('my neighbor') precedes *mein Auto* ('my car').

This approach to word order also makes testable predictions, which are at least partly independent of information-structural considerations: The less predictable the string following *my car* is, the stronger the pressure to position the DO *my car* before the IO. Consider (19) and suppose that it is known to everybody in the conversation that my daughter is 8 years old.

(19) *Ich habe* [*mein Auto*]$_{DO}$ [*meiner Tochter*]$_{IO}$ *verkauft.*
I have my car my daugther sold
'I sold my car to my daughter.'

Since selling cars to 8 year old children is not usually done, this is quite unexpected. And since it is unexpected, it is highly informative and creates a peak in the information profile. To lower this peak, it is even more advisable to first mention *my car* in order to reduce this peak to some extent.

On the other hand, there are also effects that are arguably mainly driven by information-structural considerations. Remember that according to information-theoretic reasoning, given should always precede new. Still, when answering the question *What did you give to the cashier?* in (20), it is perfectly fine to put the focused DO in the prefield at the left periphery of the sentence.

(20) [*Das GELD*]$_{DO,\,F}$ *habe ich* [*ihm*]$_{IO}$ *gegeben (aber nicht die Waffe).*
The money I have him given (but not the gun)
'The money, I gave to him (but not the gun).'

The reason for this is most probably that in (20), the focused DO *the money* is also contrastively used (contrasting with the contextual alternative *the gun*). And as has been observed in Speyer (2010), contrastively used expressions are in fact even more frequent in the German prefield than are topics.

This excursion into word order in German suggests that both information-structural and information-theoretic concepts are relevant to describe the facts and to get a deeper understanding of what exactly is going on in this area. At the same time, the two concepts do not appear to be completely independent of each other, but rather connected in an interesting way. In the following section, we therefore sketch more systematically what this relation could look like.

5 Information structure or information theory?

As we saw above, information-structural and information-theoretic approaches have been proposed to account for similar phenomena. Further examples are the

distribution of ellipsis and the prosodic realization of utterances. Even though the underlying reasoning between both families of approaches is fairly different (e.g., discrete categories like topic and comment or background and focus on the one hand and gradual predictability on the other hand), their predictions are often aligned. For instance, the focus of an utterance (whether understood as presenting new information or signaling alternatives) is probably less predictable on average than its background, which is given relative to the previous discourse, or topics, which we know the utterance is about. So we do not really know whether a word is deaccented, omitted or placed at the beginning of an utterance because it is given or because it is predictable.

As we have just sketched in the case study on non-canonical word order, the relationship between information-structural and information-theoretic notions is an intricate one. Therefore, let us briefly delimit the space of possible relationships between these two approaches. First, information-structural concepts could be crucial in determining the form of utterances and information-theoretic ones could just be an artifact of the average higher predictability of certain information-structural categories. Words might be accented or resist ellipsis because they belong to the focus of the utterance and not because they are unpredictable, and the lower predictability of foci is just an irrelevant co-occurrence. Second, it might be the other way round: Predictability is what determines encoding choices, and the overlap between information-structural concepts and predictability only suggests effects of information structure. A third option is that, as Levy (2008) argues, probabilistic information is a causal bottleneck to the choice between encodings. How predictable a word is in context is of course affected by many linguistic and extralinguistic factors, including information-structural ones. But, in the end, it is surprisal that triggers the actual encoding choice. And finally, there might be independent and possibly interacting contributions of information-theoretic and information-structural concepts, which cannot be traced back to the other theory or to surprisal being a causal bottleneck. For instance, Kehler & Rohde (2017) show that the interpretation of a pronoun depends on the Question under Discussion that comprehenders assume, which they predict probabilistically from the preceding context. Taking only an information-structural perspective (a QuD-based model of discourse) or only an information-theoretic one (predicting upcoming words) fails to explain the data.

Of course, not all of these possibilities are mutually exclusive, for instance, the theories might have independent effects on different levels of linguistic analysis even though surprisal functions as a causal bottleneck on others.

The goal of this volume is to bring together contributions which shed light on the relationship between information structure and information theory with

respect to different linguistic phenomena. In what follows, we will briefly summarize the contributions in order to give the reader a first idea of the papers, to reduce their surprisal, and thus to facilitate their reading and processing.

6 Overview of the contributions

In their contribution *The comprehension of broad focus: Probing alternatives to verb phrases* to this volume, Radim Lacina, Patrick Sturt and Nicole Gotzner concentrate on the empirical testing of the hypothesis that focus triggers mental representations of alternatives. While Gotzner et al. (2016) tested this hypothesis with respect to minimal focus on nouns, this contribution extends their approach to cases of broad focus (more concretely, to focused VPs that consist of a noun and a verb). Gotzner et al. (2016) conducted a probe recognition task and found that association of focus with focus particles like *only* results in longer reaction times in the case of related alternatives as compared to unrelated probes. This inhibition effect is interpreted as the result of a competition of alternatives, and it is, in principle, also to be expected in cases of broad focus. To test this, this contribution presents 3 probe recognition tasks, with experiment 1 testing alternatives to the noun, experiment 2 alternatives to the verb, and experiment 3 alternatives to the VP. While there is a main effect of relatedness across all three experiments, the expected inhibition effect is only observed in experiment 1.

The contribution *An information-theoretic account of constituent order in the German middle field* by Katrin Ortmann, Sophia Voigtmann, Stefanie Dipper and Augustin Speyer investigates to what extent information theory and information structure explain the preferred ordering of arguments in German and how these concepts are related empirically. In their corpus study, they operationalize the tendency to distribute Shannon information uniformly as deviation of the rolling mean between by-word surprisal across the utterance (Cuskley et al. 2021). Givenness is operationalized as definiteness. The data show that distributing information uniformly predicts two tendencies observed in the literature: Dative objects are preferably placed before accusatives, and given/definite objects before new/indefinite ones.

Word order is also addressed by Yvonne Portele and Markus Bader in their contribution *Choosing referential expressions and their order: Accessibility or Uniform Information Density?*. They discuss whether accessibility accounts or the information-theoretic uniform information density (UID) hypothesis are better suited to explain (i) which referential expressions are produced for specific discourse referents and (ii) in which order they are arranged. With respect to (i),

a sentence continuation task in German showed that speakers are more likely to pronominalize the topic of the previous sentence than the most expected discourse referent. Based on this result, Portele and Bader argue that the choice of expressions can be better explained by accessibility than by UID. As for (ii), Portele and Bader's results on word order from two picture description tasks are largely consistent with both the accessibility accounts and with UID. For example, the observed preference for patient-initial clauses after a narrow question asking for the patient can be explained both with topic continuity (accessibility) and with predictability (UID). As a result, Portele and Bader argue that accessibility and UID can both contribute to adequately describe word order preferences in German.

In their contribution *The role of information in modeling German intensifiers*, J. Nathanael Philipp, Michael Richter, Tatjana Scheffler and Roeland van Hout investigate intensifiers in a new German-language corpus of tweets and blog posts. They determined context free and context dependent information measures for these expressions and found, first, that both measures are highly correlated and, second, that they account for the distribution of the intensifiers in their data. They conclude that these findings support the assumption of a common word class "intensifier". Furthermore, Philipp et al. tested the hypothesis following from the uniform information density hypothesis that in stacked intensifiers, intensifiers with lower information content precede intensifiers with higher information content. This way, expressions with lower information should introduce those with higher information and thus facilitate processing for the recipient. By comparing the original sentences with stacked intensifiers to variants with either the intensifiers in the reverse order or with only the last intensifier, Philipp et al. found that the original sentences indeed exhibit on average more uniform information profiles.

Swantje Tönnis' contribution *Cleft sentences reduce information density in discourse* proposes an information-theoretic explanation for the hypothesis put forward in Tönnis (2021) that clefts address less expected questions under discussion (QUDs) while canonical sentences address relatively expected QUDs. The idea is that clefts can be a means to reduce information density and achieve an even distribution of information when the QUD addressed by the utterance is less predictable. Tönnis formalizes this with a theoretical model based on the new concept of QUD surprisal, which is inspired by Asr & Demberg's (2015) discourse relational surprisal. This model predicts the choice between a cleft and a canonical sentence based on the likelihood of the QUD that is answered by the corresponding sentence. Tönnis shows that in contrast to previous accounts, the model makes correct predictions for her discussed example.

While much of the previous research on information theoretic constraints on language investigates how speakers organize their utterance based on differences in predictability, the contribution *Tell me something I don't know: Speaker salience and style affect comprehenders' expectations for informativity* by Vilde R. S. Reksnes, Alice Rees, Chris Cummins and Hannah Rohde looks into what makes an expression predictable. In two production experiments, they find that listeners expect speakers to produce informative utterances rather than predictable ones. The first study shows that more diverse and surprising information is elicited when the salience of the speaker is increased in the experiment, e.g. by presenting a picture of them. When only a bare sentence is to be completed, subjects are more likely to produce typical material, like in a cloze task. The second experiment shows that listeners' expectations about a speaker's degree of informativity is adapted to the behavior of particular speakers: if somebody is known to frequently provide (un)informative utterances, listeners also expect this person to do so in the future.

The contribution *Prosodic factors do not always suppress discourse or surprisal factors on word-final syllable duration in German polysyllabic words* by Ivan Yuen, Bistra Andreeva, Omnia Ibrahim and Bernd Möbius investigates whether discourse factors such as information status, prosodic factors such as prosodic boundary type and accenting, and information-theoretic measures like surprisal contribute to the acoustic realization of the word-final syllable duration in polysyllabic words. To this effect, they extracted polysyllabic words from the DIRNDL corpus that occur at a phrase boundary, and which are annotated for lexical information status (given or new). The authors added surprisal estimates based on data from the deWaC. In long words (4 syllables or more), they only found a prosodic boundary effect. In short words (up to 3 syllables), however, the information status, the presence of a pitch accent and the log surprisal also significantly affected the duration of the word-final syllable. These results show that with respect to word-final syllable duration, both information status (given and new) and surprisal can influence the acoustic realization.

References

Ariel, Mira. 1990. *Accessing noun-phrase antecedents* (Croom Helm Linguistics Series). London & New York: Routledge. DOI: 10.4324/9781315857473.

Asr, Fatemeh Torabi & Vera Demberg. 2015. Uniform Information Density at the level of discourse relations: Negation markers and discourse connective omission. In Matthew Purver, Mehrnoosh Sadrzadeh & Matthew Stone (eds.), *Proceedings of the 11th International Conference on Computational Semantics*, 118–128. London: Association for Computational Linguistics.

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. DOI: 10.1177/00238309040470010201.

Büring, Daniel. 1997. *The meaning of topic and focus: The 59th Street Bridge Accent* (Routledge Study in German Linguistics 3). London: Routledge. DOI: 10.4324/9780203448014.

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 27–55. New York: Academic Press.

Crocker, Matthew W., Vera Demberg & Elke Teich. 2016. Information density and linguistic encoding (IDeaL). *KI – Künstliche Intelligenz* 30(1). 77–81. DOI: 10.1007/s13218-015-0391-y.

Cuskley, Christine, Rachael Bailes & Joel Wallenberg. 2021. Noise resistance in communication: Quantifying uniformity and optimality. *Cognition* 214. 104754. DOI: 10.1016/j.cognition.2021.104754.

É. Kiss, Katalin. 1998. Identificational focus versus information focus. *Language* 74(2). 245–273. DOI: 10.2307/417867.

Fenk, August & Gertraud Fenk. 1980. Konstanz im Kurzzeitgedächtnis – Konstanz im sprachlichen Informationsfluß? *Zeitschrift für experimentelle und angewandte Psychologie* 27(3). 400–414.

Fenk-Oczlon, Gertraud. 1989. Word frequency and word order in freezes. *Linguistics* 27. 517–556. DOI: 10.1515/ling.1989.27.3.517.

Frank, Austin F. & T. Florian Jaeger. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 939–944. Austin.

Frey, Werner. 2000. Über die syntaktische Position der Satztopiks im Deutschen. *ZAS Papers in Linguistics* 20. 137–172. DOI: 10.21248/zaspil.20.2000.82.

Gotzner, Nicole, Isabell Wartenburger & Katharina Spalek. 2016. The impact of focus particles on the recognition and rejection of contrastive alternatives. *Language and Cognition* 8(1). 59–95. DOI: 10.1017/langcog.2015.25.

Grice, H. Paul. 1989. *Studies in the way of words*. Cambridge; London: Harvard university Press.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307. DOI: 10.2307/416535.

Halliday, Michael Alexander Kirkwood. 1967. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3(2). 199–244. DOI: 10.1017/S0022226700016613.

Hockett, Charles F. 1958. *A course in modern linguistics*. New York, NY: Macmillan.

Höhle, Tilman N. 1982. Explikation für ‚normale Betonung' und ‚normale Wortstellung'. In Abraham Werner (ed.), *Satzglieder im Deutschen. Vorschläge zur syntaktischen, semantischen und pragmatischen Fundierung* (Studien Zur Deutschen Grammatik 15), 75–153. Tübingen: Narr.

Jaeger, T. Florian & Esteban Buz. 2017. Signal reduction and linguistic encoding. In Eva M. Fernández & Helen Smith Cairns (eds.), *The handbook of psycholinguistics*, 38–81. Hoboken, NJ, USA: John Wiley & Sons. DOI: 10.1002/9781118829516.ch3.

Kehler, Andrew & Hannah Rohde. 2017. Evaluating an expectation-driven question-under-discussion model of discourse interpretation. *Discourse Processes* 54(3). 219–238. DOI: 10.1080/0163853X.2016.1169069.

Krifka, Manfred. 2007. Basic notions of information structure. In Caroline Féry, Gisbert Fanselow & Manfred Krifka (eds.), *The notions of information structure* (Interdisciplinary studies on information structure 6), 13–55. Potsdam: Universitätsverlag Potsdam.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents* (Cambridge Studies in Linguistics 71). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511620607.

Lemke, Robin. 2021. *Experimental investigations on the syntax and usage of fragments* (Open Germanic Linguistics 1). Language Science Press. DOI: 10.5281/zenodo.5596236.

Lenerz, Jürgen. 1977. *Zur Abfolge nominaler Satzglieder im Deutschen* (Studien zur deutschen Grammatik). Tübingen: TBL-Verlag Narr.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177. DOI: 10.1016/j.cognition.2007.05.006.

Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In Bernhard Schölkopf, John Platt & Thomas Hofmann (eds.), *Advances in Neural Information Processing Systems 19*, vol. 19, 849–856. Cambridge: The MIT Press. DOI: 10.7551/mitpress/7503.003.0111.

Mathesius, Vilém. 1975. *A functional analysis of present day English on a general linguistic basis*. Josef Vachek (ed.). Berlin, Boston: De Gruyter Mouton. DOI: 10.1515/9783110813296. 10.1515/9783110813296.

Paul, Hermann. 1919. *Deutsche Grammatik. Bank III. Teil IV: Syntax (Erste Hälfte)*. Halle, Saale: Max Niemeyer Verlag. DOI: 10.1515/9783110929805.4.fm.

Pierrehumbert, Janet & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in communication*, 271–312. Cambridge, MA: MIT Press. DOI: 10.7551/mitpress/3839.003.0016.

Prince, Ellen F. 1981. Toward a taxonomy of given-new formation. In Peter Cole (ed.), *Radical Pragmatics*, 223–254. New York: Academic Press.

Rauth, Philipp. 2020. *Ditransitive Konstruktionen im Deutschen: Geschichte und Steuerung der Objektabfolge im Mittelfeld* (Studien zur deutschen Grammatik 100). Tübingen: Stauffenburg.

Reich, Ingo. 2018. Ellipsen. In Frank Liedtke & Astrid Tuchen (eds.), *Handbuch Pragmatik*, 240–251. Stuttgart: J.B. Metzler. DOI: 10.1007/978-3-476-04624-6_24.

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27. 53–94.

Rooth, Mats. 1985. *Association with focus (Montague grammar, semantics, only, even)*. University of Massachusetts Amherst. (Doctoral dissertation).

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116. DOI: 10.1007/BF02342617.

Schäfer, Lisa. 2021. Topic drop in German: Empirical support for an information-theoretic account to a long-known omission phenomenon. *Zeitschrift für Sprachwissenschaft* 40(2). 161–197. DOI: 10.1515/zfs-2021-2024.

Schwarzschild, Roger. 1999. Givenness, Avoidf and other constraints on the placement of accent. *Natural Language Semantics* 7(2). 141–177. DOI: 10.1023/A:1008370902407.

Selkirk, Elisabeth O. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge: MIT Press.

Selkirk, Elisabeth O. 1995. Sentence prosody: Intonation, stress and phrasing. In John A. Goldsmith (ed.), *The handbook of phonological theory* (Blackwell Handbooks in Linguistics 1), 550–569. Cambridge: Blackwell.

Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27(3). 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Speyer, Augustin. 2010. Filling the German Vorfeld in written and spoken discourse. In Sanna-Kaisa Tanskanen, Marja-Liisa Helasvuo, Marjut Johansson & Mia Raitaniemi (eds.), *Discourse in interaction*, 263–290. Amsterdam: John Benjamins. DOI: 10.1075/pbns.203.19spe.

Stalnaker, Robert C. 1978. Assertion. In Peter Cole (ed.), *Pragmatics* (Syntax and Semantics 9), 315–332. New York, San Francisco, London: Academic Press.

Stowell, Tim. 1991. Empty heads in abbreviated English. *GLOW Newsletter* 26. 56–57.

Tily, Harry & Steven Piantadosi. 2009. Refer efficiently: Use less informative expressions for more predictable meanings. In *Proceedings of the workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference*. Tilburg: Tilburg University.

Tönnis, Swantje. 2021. *German 'es'-clefts in discourse: A question-based analysis involving expectedness*. Graz University. (Doctoral dissertation).

von der Gabelentz, Georg. 1868. Ideen zu einer vergleichenden Syntax: Wort- und Satzstellung. *Zeitschrift für Völkerpsychologie und Sprachwissenschaft* 6. 376–384.

Wackernagel, Jacob. 1892. Über ein Gesetz der indogermanischen Wortstellung. *Indogermanische Forschungen* 1. 333–436. DOI: 10.1515/9783110242430.333.

Walker, Marilyn, Aravind K. Joshi & Ellen Prince. 1998. Centering in naturally-occurring discourse: An overview. In Marilyn A. Walker, Aravind K. Joshi & Ellen F. Prince (eds.), *Centering theory in discourse*, 1–28. Oxford: Oxford University Press.

Chapter 2

The comprehension of broad focus: Probing alternatives to verb phrases

Radim Lacina[a], Patrick Sturt[b] & Nicole Gotzner[a]

[a]Osnabrück University [b]University of Edinburgh

Recent research has shown that comprehenders represent alternative meanings to single focused words online (for a review, see Gotzner & Spalek 2019), consistent with Rooth's (1992) formal semantic account. However, focus can also take a scope over whole phrases such as the VP *read the manuscript*. We examined whether in these cases, too, alternatives are represented by testing for an interference effect of the particle *only*, which necessarily evokes alternatives. Using the probe recognition task, we first tested unmentioned alternatives to the constituent parts of VPs, to object nouns (Experiment 1, *letter* for *manuscript*) and verbs (Experiment 2, *wrote* for *read*). In Experiment 3, we tested alternatives to the whole phrase (*wrote the letter*). In all experiments, alternative probes were processed slower than unrelated ones. We found varying evidence of the interference effects of *only* with noun, verb and whole-phrase alternatives. Overall, this study does not provide support for the generalisation of the effects of *only* to larger units. Since our study was the first to use the probe recognition task with phrase-sized constituents, we discuss the methodological implications of our work – we found a relatedness effect for whole phrases and this shows that the probe recognition task can be used to test the representation of larger constituents.

1 Introduction

Speakers' utterances often contain presupposed parts as well as those that are emphasised as new or contrastive. This information must be rapidly integrated within an ever-evolving model of the discourse (Johnson-Laird 1983). Focus is a key category within this information structure, which has been likened, as a level

Radim Lacina, Patrick Sturt & Nicole Gotzner. 2024. The comprehension of broad focus: Probing alternatives to verb phrases. In Robin Lemke, Lisa Schäfer & Ingo Reich (eds.), *Information structure and information theory*, 23–53. Berlin: Language Science Press. DOI: 10.5281/zenodo.13383785

of linguistic structure, to the "packaging" of truth-conditional content that makes it fit within different contexts (Chafe 1976). Alternative meanings that speakers could have used but did not have been proposed as crucial to understanding focus (Krifka 2008). Recent research using psycholingustic methods has found that when faced with individual focused words, comprehenders actually consider contextually related alternatives to those words in their minds in real time (Braun & Tagliapietra 2010, Husband & Ferreira 2016, Gotzner & Spalek 2019). In the current research, we aim at expanding this investigation to cases of focus on larger units, i.e., situations where whole phrases as opposed to single words are focused, and test whether comprehenders also represent alternatives to these more complex elements in the course of online processing.

Firstly, we cover the theoretic treatment of the phenomenon of focus to provide a base for our understanding of meaning alternatives. Next, we review the experimental evidence regarding the activation, selection, and representation of alternatives in real-time processing. Then, we discuss the distinction between narrow and broad focus. From this, we derive the motivation and design of our three experiments, which follow in the next section. Finally, we discuss our results both in relation to the current psycholinguistic investigations of focus but also to some of the theoretical debates.

1.1 Focus and its theoretical explanations

Let us now turn to the theoretical approaches that have been proposed to account for focus. Many researchers have broadly associated it with the "new" or "contrastive" parts of utterances in opposition to the known and presupposed parts (Halliday 1967, Jackendoff 1972, Sgall et al. 1973, 1986). Here, we will mostly zoom in on an approach that sees meaning alternatives as crucial, as this is the basis of the current study. Additionally, we will also discuss a recent attempt to ground focus within a general framework of pragmatic reasoning, namely the Rational Speech Act model.

1.1.1 The Roothian approach

An influential semantic approach to focus claims that its main function is to introduce alternative meanings that could replace the focused element in the discourse (Krifka 2008). This theory, proposed by Rooth (1985), states that focus generates an additional level of meaning, which consists of a set of propositions derived by replacing the focused element with contextually plausible alternatives of the same semantic type. Take the following example:

(1) In the monastery, $[\text{Jane}]_{\text{F}}$ read the manuscript.

According to Roothian alternative semantics (Rooth 1992), the focused subject noun *Jane* in (1) gives rise to alternatives of the same semantic type as the focused element, here let us assume type *e*. This formal set containing alternatives is then subject to a further process, namely contextual restriction. Suppose that there is a world containing four individuals – Jane, William, George, and Whiskers the cat. The focus value of sentence (1) is then said to be the set of propositions that are derived by systematically replacing the focused element with its contextually appropriate alternatives. In our example, the set would be {Jane, William, George}. Even though Whiskers is also of type *e*, it cannot, being a cat, read anything and thus is not contextually appropriate in (1). Equally, should the word *manuscript* be focused in (1), alternatives such as *letter* or *scroll*, i.e. words or concepts that could be the objects of *read*, would be generated.

One of the strengths of Rooth's (1992) theory is that it can also account for various focus-related effects in addition to bare focus. This includes the behaviour of focus-sensitive particles such as *only*, *also*, or *even*. For example, if inserted into (1) before *Jane*, *only* would assert that the proposition expressed in (1) is true for the focused element, but that none of the propositions obtained by replacing *Jane* with its alternatives are. Therefore, the communicated content is that neither William nor George read the manuscript. The analysis of *only* is that it is a particle that takes a proposition to combine with and asserts that no other contextually relevant proposition is true (von Fintel 1997). Focus is thus seen as one of the many phenomena in language where alternative utterances play a crucial role (see Gotzner & Romoli 2022 for a review).

1.1.2 Explanations within the RSA framework

Another explanation of the phenomenon of focus in language is provided by the Rational Speech Act framework (Goodman & Frank 2016, Franke & Jäger 2016), which has become a standard modelling tool for a variety of semantic and pragmatic phenomena in recent years. This framework provides a general model of speaker-hearer interactions and human cognition. Within this approach, non-literal meaning is modelled as successive Bayesian calculations on the side of the hearer about the intended meaning of the speaker's utterances. Hearers are said to compute probability distributions over the possible worlds in which any given utterance would be true. Crucially for the purposes of giving an account for focus, the RSA approach incorporates the variable of utterance cost into the computation that the hearer conducts. Within the RSA framework, pragmatic

enrichment occurs due to some utterances being more costly to produce than others.

Bergen & Goodman (2015) attempt to combine the RSA approach with noisy channel theories (Stevens 2016, Stevens & Roberts 2019). These noisy channel approaches see language as essentially a solution to the problem of transferring information between individuals in a situation where some of it can be lost or distorted (Shannon & Weaver 1949). They start their account by appealing to the prosodic marking of focus in languages such as English or German. What this prosodic prominence is said to achieve is to lower the probability that the word in question is misheard by the listener and potentially misinterpreted for one of its plausible alternatives, these may be previously given in the context or constructed by the comprehender. Given a word, this prominence also carries a certain cost on the part of the speaker. Within the RSA framework then, Stevens (2016) argues, the way focus alternatives are introduced is by means of the following chain of iterative Bayesian reasoning. In essence, upon hearing a word with prosodic prominence marking focus, the listener reasons that the speaker must have intended to expend extra production costs with the goal of improving the chances that the correct word would make it across the noisy channel of communication to the listener. If this were not done, the listener might take the word to be not the intended one, but one of its contextually plausible alternatives. Therefore, it must be that these other alternatives are in fact false according to the beliefs of the speaker given that they were willing to give the extra cost associated with the prominence. In this way then, the listener arrives at the conclusion that a prosodically prominent focused word carries with it an exhaustivity implicature, i.e. that its alternatives are false.

This is how the RSA model proposes to explain the phenomenon of focus alternatives. However, as has been perceptively pointed out by an anonymous reviewer of this chapter, there are several issues with this proposal that ought to be acknowledged. Whether or not something is a plausible alternative to a focused word is dependent, unsurprisingly, on the words' meaning. However, similarity in meaning does not, in the majority of cases, necessitate similarity in the phonological form. For example, should *sheep* be focused in a particular sentence, its alternatives would presumably be *goat* or *cow*, words very different from *sheep* phonologically and not likely to be misheard. *Ship* on the other hand would in most cases not be a plausible alternative, yet this word is much more likely to be confused with *sheep*.

Notice that the account is not mutually exclusive with that of Rooth (1992). Rather, it embeds the effects of focus within a computational model of pragmatic reasoning. It could therefore be seen as complementary to the formal semantic

theories that were presented earlier in the introduction. There are however differences, since, arguably, the RSA-based account relies on the notion of prosody playing a crucial role in the computation of focus alternatives (this is at least the case in Germanic languages). It could be argued that where there is no prosodic prominence in the marking of focus, alternatives should not be entertained by comprehenders. We will return to the role of prosody and the associated RSA-based explanation when discussing both the predictions and results of the current study.

1.1.3 Processing focus

What we have discussed so far was concerned with the interpretations and the associated formalisations of sentences with focused elements. We will now turn to how these interpretations are arrived at in the minds of comprehenders when perceiving and parsing such sentences in real time. This is the algorithmic level of analysis of cognitive phenomena (Marr 2010) argued to be distinct from formal analyses, yet complementary to them and capable of influencing them (Love 2015).

As far as online comprehension is concerned, both focus and focus particles have been linked to distinctive effects with research suggesting that focused information is processed "more deeply" (Sturt et al. 2004, Ward & Sturt 2007), that focus enhances anaphor integration (Klin et al. 2004, Sanford et al. 2009), that it affects ellipsis processing (Frazier et al. 2007, Carlson 2015), and that it exerts an influence on parsing (Filik et al. 2005). Focus particles have been found to affect syntactic attachment (Carlson & Potter 2022) and allow comprehenders to predict upcoming contrasts (Carlson 2013).

Crucially for the purposes of the current study, the past decade has seen evidence suggesting that Rooth's (1992) semantic approach can also be applied to the online processing of focus by comprehenders (for an overview, see Gotzner & Spalek 2019). This was arrived at by examining the patterns of the activation, selection, and representation of alternatives in comprehenders when exposed to either spoken or written stimuli with focus marked either prosodically or syntactically. When we speak of *activation* or *representation*, what we wish these terms to refer to are phenomena within the real-time processing of language.

Many experimental paradigms have been used in focus alternative research. However, two in particular have been implemented extensively, namely the lexical decision (Braun & Tagliapietra 2010, Byram Washburn et al. 2011, Husband & Ferreira 2016, Gotzner et al. 2016, Yan & Calhoun 2019, Yan et al. 2023) and probe recognition tasks (Gotzner et al. 2016, Gotzner & Spalek 2017, Jördens et al.

2020, Spalek & Oganian 2019). Below, we review some of the studies using these two tasks that have been instrumental in establishing the processing reality of Roothian-style alternatives.

Husband & Ferreira (2016) examined whether focus alternatives are activated in processing and what the time-course of this activation is. In their study, English speakers heard sentences in which a particular word (here *sculptor*) was pronounced either with contrastive (L+H*) or non-contrastive (H*) prosody:

(2) The museum thrilled the *sculptor* when they called about his work.

Contrastively (*painter*) and noncontrastively (*statue*) associated probe words were subsequently presented. In their two experiments, they found that when the probe word was presented immediately after the prime word (0ms SOA) both contrastively and non-contrastively associated words were activated (in contrast to unrelated words)[1]. However, when there was a 750ms delay relative to the prime word, only the contrastively associated probes remained activated. Husband & Ferreira (2016) measured this activation by means of a lexical decision task where their participants had to judge whether a given probe was an existent word of English or not. This pattern suggested that comprehenders were generating sets of focus alternatives in real time, since conditional upon prosodic contrastive focus marking, only those words that could replace the focused element were activated. It also appears that in the course of processing focus, the mechanism first activates broadly associated words and only afterwards selects the final set of focus alternatives.

Regarding focus particles, it has been found that they have additional effects on the processing of alternatives. Gotzner et al. (2016) conducted a study, in which their German participants had the task of indicating whether a given probe word appeared anywhere in discourses such as these:

(3) In the fruit bowl, there are peaches, cherries, and bananas.

(4) I bet Carsten has eaten cherries and bananas.

(5) No, he _/only/even ate the [peaches]$_F$.

In this study, the noun was always spoken with a contrastive pitch accent (L+H*). The final sentence also included either one of two focus particles, *only* or *even*, or neither was present. They used the probe recognition task, in which

[1]The researchers also presented non-words as probes in their filler items to balance the ratio of yes-no responses.

participants have to judge whether a given word appeared anywhere in the previously presented stimulus, while their accuracy and response times are measured. This task has been argued to tap into the mental representation of the discourse (Gernsbacher & Jescheniak 1995) as opposed to the immediate activation within the lexical-semantic network, which is often studied with the lexical decision task (Meyer & Schvaneveldt 1971). Their results revealed that the presence of focus particles interfered with the recognition of mentioned alternatives (*cherries*) as well as the rejection of unmentioned yet plausible alternatives (*melons*). Both types of probes were reacted to even more slowly (compared to unrelated words) when either *only* or *even* was present. This additional slowdown caused by focus particles was later shown to be specific to contrastive alternatives by Gotzner & Spalek (2017), who found that when general associates of focused nouns were presented as probes, no interference effects of focus particles were observed.

Gotzner (2017) interprets these results as an indication of a competition process taking place between the mentioned alternatives (*cherries*, *bananas*), the unmentioned alternatives (*melons*) and the focused element (*peaches*). Comprehenders are said to create a place holder upon processing a focus sensitive particle (Gotzner 2017), since in order to successfully parse the exhaustive or additive meaning conveyed by the particle, alternatives to the focused element must be represented. This place holder then is sensitive to what could replace it. Since both the mentioned and unmentioned alternatives match the place holder, interference occurs. The results of both Gotzner et al. (2016) and Gotzner & Spalek (2017) show that focus particles create additional effects compared to bare intonational focus and that they are specific to alternatives appropriate within the utterance's context. As such, these focus particle interference effects can be said to be litmus test of the representation of focus alternatives in cases where focus itself is not manipulated.

Further studies have replicated the effects exhibited by focus alternatives either in languages other than English (e.g. see Braun & Tagliapietra 2010, for Dutch; Jördens et al. 2020, for German; Yan & Calhoun 2019, Yan et al. 2023, for Mandarin Chinese; Calhoun et al. 2022, for Samoan; Tjuka et al. 2020, for Vietnamese; Káldi et al. 2021, for Hungarian; Lacina et al. 2023, for Czech) or using different paradigms (see Kim et al. 2015; Braun et al. 2019 for eye-tracking), and have examined how focus influences the memory recall of alternatives (Fraundorf et al. 2010, Lee & Fraundorf 2017, Norberg & Fraundorf 2021). What the evidence converges on is that focus alternatives indeed play an important role during online comprehension, in addition to being a valuable tool in the domain of formal semantic theory.

1.1.4 Broad focus

However, the alternative-based account of the processing of focus has so far only been tested on cases of narrow focus, that is on sentences where a single word is focused. It has long been known that focus scope can vary and that it can encompass a constituent larger than a single word (Selkirk 1995, Gussenhoven 1999, Erteschik-Shir 2007). For example, a whole VP including the verb and its direct object may be focused. Consider the contrast between the two following sentences:

(6) In the monastery, Jane read the [manuscript]$_F$.

(7) In the monastery, Jane [read the manuscript]$_F$.

Cases such as (7) are known as broad focus and contrast with narrow focus constructions seen in sentence (6). This phenomenon is readily accommodated under Rooth's (1992) formal account. The principle of substituting elements of the same semantic type is maintained. The only difference from narrow focus is the type being substituted. Let us take (7) as an example. Here, the phrase *read the manuscript* is in focus and its meaning is standardly analysed as being type <*e,t*> (Kratzer & Heim 1998). According to Rooth (1992), its alternatives should also be of this type. The set of alternatives for (7) might then consist of, for example, λ.x [x sealed the letter]; λ.x [x wrote the scroll].

One feature of broad focus that ought to be mentioned is that these structures are often ambiguous with narrow focus ones. This is the phenomenon of focus projection (Rochemont 1986, von Stechow & Uhmann 1986, Selkirk 1995), where in languages such as English, prosodic prominence marking focus is placed at the right edge of the broad-focused phrase. As a consequence, it has been claimed that this is prosodically indistinguishable from narrow focus on the right-most element (Ladd 2008). This debate has not been conclusively settled with some evidence suggesting that speakers tend to give larger prenuclear prominence to verbs under broad focus compared to narrow focus situations (e.g. Breen et al. 2010). What is known, however, is that the disambiguation between broad and narrow focus can be made by means of the preceding context (Büring 2007), for example by an explicit question-under-discussion (Roberts 2012) or by means of what is given in the context, according to the Givenness Principle (Schwarzschild 1999).

1.1.5 The question of the processing of broad focus

Broad focus is arguably crucial to our understanding of how alternatives operate. Given the alternatives' more complex structure, new questions arise both on the

level of formal analysis (e.g. Fox & Katzir 2011) and processing regarding their generation and representation. However, no study to-date has tested whether the comprehension results presented in the previous section obtain for these broad focus cases as well. While there has been a processing study on broad focus, namely that of Bishop (2017), it only examined the above-mentioned role of the prenuclear accent on the verb in marking broad focus and did not test either verb or whole-phrase alternatives.

This question is crucial for the study of how people process information structure and incorporate focused elements into their evolving interpretation of meaning. We believe this to be the case, given that there is a clear prediction from Rooth's (1992) formal semantic theory – broad focus is semantically analogous to narrow focus and therefore, comprehenders ought to exhibit the same patterns of the representation of alternatives that research has shown for narrow focus when they encounter broad focus. In essence, we are asking whether the Roothian-inspired processing theory truly generalises to all the cases that the formal theory of Rooth (1992) applies to. Below, we report our attempt to test this prediction by means of three experiments.

1.2 The current study

In the current study, we aimed to test whether the alternative-based approach to the processing of focus generalises to cases of broad focus. We put forward the following hypothesis:

(8) Alternative Representation Hypothesis
Comprehenders create representations of contextually appropriate alternatives to the focused element concordant with its semantic type.

As far as our account of the processing of focus is concerned, the predictions of this Roothian-inspired (1992) processing approach, with the central claim that alternatives are being entertained in the minds of comprehenders in real-time processing, are clear. In case a comprehender encounters a sentence in which the whole VP is focused, they ought to activate, select, and represent alternatives to this larger constituent. Likewise, the constituent parts of these alternatives ought to give rise to an enhanced activation and representation of their constituent parts. In the case of focused VPs with transitive verbs, these would be both the verbs and nouns within the alternative phrase.

As for what the RSA-based approaches might give us for predictions in the case of broad focus, these depend on the question of whether broad and narrow

focus are prosodically distinct. Should only the right-most element receive prominence, then it could be argued that alternatives should only arise for this element and not for the whole VP. Therefore, if we do not see evidence for VP-level alternatives in processing, this could be seen as consistent with the RSA approach. Should, however, the two have different prosodic profiles, the RSA model could incorporate whole VP alternatives.

We conducted three probe recognition experiments aimed at testing the Alternative Representation Hypothesis. As our starting point, we took the interference effect of focus particles that has been identified by the research of Gotzner et al. (2016). Should the previous results obtained for narrow focus fully generalise to cases of broad focus, we would expect the same interference pattern. We reasoned that if this effect were to be found when rejecting unmentioned alternatives to focused VPs, this would constitute evidence in favour of these alternatives being activated, selected, and represented by comprehenders. We also predicted that the alternatives would exhibit the interference effect of *only*, this should also be the case for their constituent parts. We thus predicted that if the Alternative Representation Hypothesis is true, we ought to see *only* interfering with alternatives to both nouns, verbs as well as whole phrases. Remember that this is because the interference effect of focus particles has been interpreted in the literature as a sign of the additional unmentioned alternatives being activated and competing for selection (Gotzner et al. 2016). Should we not find this interference, this would go against the straightforward generalisability of the Roothian-inspired (1992) processing approach to broad focus.

We constructed discourses designed to elicit a broad focus interpretation in the final sentence presented. Given that we were working with texts, no explicit prosody could potentially distinguish between different focus structures. To this end, we used context sentences preceding the critical one. These context sentences first set up an assertion concerning a particular event. Then, a sentence, contrasting both in the action performed and in the patient affected, follows, and either includes the focus particle *only* or not. What is then probed are unmentioned yet plausible alternatives to broadly focused phrases.

Below, we report three probe recognition experiments. In all, the stimuli were presented in the rapid serial visual presentation mode (RSVP). The results of Byram Washburn et al. (2011) showed that written stimuli can induce focus alternative effects when *only* is present. The experiments were web-based; the first two were hosted on the IbexFarm platform (Drummond et al. 2016), whilst Experiment 3 was conducted using PCIbex (Zehr & Schwarz 2018). The first two experiments probed alternatives (i.e. tested the speed of rejection of unmentioned plausible alternative words compared to unrelated ones) to the constituent nouns

(Experiment 1) and verbs (Experiment 2) within focused VPs, whilst Experiment 3 tested whole phrase alternatives to focused VPs.

2 Experiments 1 & 2

In Experiment 1, we aimed to test alternatives to the nouns within the focused phrase (e.g. *letter* for *manuscript*). Experiment 2 tested verbal alternatives (*wrote* for *read*). This means that we presented participants with stimuli, in which we set up a broad-focused phrase. This phrase was then either bare or preceded by the focus particle *only*. Detecting the established interference of *only* when plausible alternatives to either the noun or verb are presented to comprehenders would be evidence for their representation in real-time focus processing.

Both experiments were set up according to a 2 × 2 factorial design. Firstly, we manipulated particle presence, i.e. whether the focus particle *only* was present in the critical sentence of the stimuli, i.e. in the sentence where a broad-focused phrase was present. The second factor was alternative status, in which we manipulated the type of probes that participants were reacting to. Either alternative or unrelated probes were shown. This means words that could plausibly replace the corresponding part of the focused phrase within the given context or words whose meaning is contextually incompatible with this replacement.

Given the discussion of our hypothesis and general predictions, we expect to detect the following effects. Firstly, we predict a main effect of the alternative status manipulation, since a general relatedness effect has been reported by many studies that used the probe recognition task with focus alternatives (Gotzner et al. 2016, Jördens et al. 2020). We predict that related probes will be rejected more slowly compared to unrelated ones. Critically, we predict an interaction between particle presence and alternative status. We expect *only* to interact with the plausible unmentioned alternatives and not with unrelated probes and this to be evidenced in the response time measure. If seen, this would be evidence of an interference effect caused by the focus particle *only* and thus consequently, an indication that the constituent part of the alternative is being represented by comprehenders within the mechanisms responsible for processing focus.

2.1 Method

2.1.1 Participants

62 (mean age 22.9) native English speaker participants were recruited to take part in Experiment 1. The participants did not receive any monetary compensation for

their time. In Experiment 2, 60 (mean age 32.9) native English speakers took part. These were recruited on the Prolific platform and received £5 for their time. Full demographic information can be found on an OSF project entry (https://osf.io/uvbdr/).

2.1.2 Materials

We constructed 40 experimental items based broadly on the stimuli used by Gotzner et al. (2016). They can be found on the OSF project page mentioned above. Each item consisted of four sentences which introduced a context together with an agent, then alternative actions that the agent could perform, an assertion of one of them and then finally a negation of this action and a correction. Take the following example:[2]

(9) Harry is a butcher.

(10) At the butcher's shop, Harry could smoke and carve the ham and the brisket.

(11) Harry surely carved the brisket.

(12) No, he _/only smoked the ham.

The items were constructed in such a way to induce a broadly focused interpretation of the VP in the final sentence. There, we also manipulated the presence or absence of the focus particle *only*, i.e. the particle presence manipulation. Next, the items differed in the probe words that were presented after the final sentence. These were either plausible yet unmentioned alternatives or unrelated words. We also included mentioned alternatives in the second context sentence (10). These were constructed in such a way to elicit in the mental model of the comprehenders the possible actions that the agent of the scenario could take with the aim that any permutation of the conjoined verbs and nouns would be included. In Experiment 1, we probed alternatives to the object noun (*ham*) within the VP (alternative: *sirloin*; unrelated: *mastiff*). In Experiment 2, the verbs (*smoked*) within those phrases had their alternatives tested (alternative: *salted*; unrelated: *distanced*). This was the alternative status manipulation. These probe words were controlled for letter length and the log-frequency of word forms in

[2]During the review process, it has been pointed out to us that some of our items might have the issue that the third sentence contains a presupposition of the other alternative and that the final sentence denies it. However, since what we are testing are unmentioned alternatives that are not presupposed, we do not believe this to be a detrimental issue of the design.

the British National Corpus. We also conducted a latent semantic analysis (Landauer & Dumais 1997) to measure the degree of association between the probe words and the focused ones with the goal of maximising it for alternative probes and minimising it for unrelated ones. The descriptive statistics and models computed for these purposes can be accessed on the OSF project entry.

Since in the current study, the correct answer to the probe recognition task was always *no* for the experimental items, we created a set of 80 filler items, out of which 60 required the participant to answer *yes*.

2.1.3 Procedure

Every trial consisted of four sentences presented in the RSVP mode. Words appeared individually on the screen for 300ms followed by 100ms of a blank screen. When the final word was reached, 2000ms of a blank screen followed. Next, a probe word appeared in capital letters together with a forced choice of *yes* or *no* that was made by pressing *j* for the former and *k* for the latter. In the instructions, participants were told that their task was to indicate whether the given probe word appeared anywhere in the preceding four sentences. They were told that "any form of the word" would count as the word having appeared. For example, if the form *dogs* was found in the stimulus and the probe word was *dog*, the correct response was to be *yes*. The participants were instructed to answer as quickly as possible. There was no timeout for their answers and no feedback was given. Before the experiment proper, the participants were given example items with their correct responses and completed a practice part. Each participant saw 40 experimental items and 80 fillers with no two experimental trials in immediate succession organised in a list according to the Latin Square design.

We collected participants' responses (*yes* or *no*) as well as the associated response times. These were measured from the moment of the appearance of the probe on the screen.

2.1.4 Analysis

We fitted Bayesian hierarchical models to the log-transformed response time data. Only those trials where participants correctly rejected the probe were included in the analysis. We used the `brms` package (Bürkner 2017) in the R programming language (R Core Team 2022). We included the factors of alternative status, particle presence and their interaction, as well as mean-centred ordinal trial position as fixed effects and the three-way interaction between the factors. The factors of alternative status and particle presence were sum coded. In our random effects

structure, we included the full structure justified by the design of the experiment (Barr et al. 2013), i.e. random intercepts for both participants and items as well as random slopes for both. For the full specification of the models used here, consult the preregistration entry for Experiment 3 on OSF (https://osf.io/cf36w).

Below, we report the posterior distributions of the sizes of the main effects and interactions given the data and the priors, along with their 95% credible intervals (CrI). In cases where the credible interval of the posterior distributions of the size of an effect does not include zero, we will consider this to be compelling evidence for the hypothesis that the size of the effect is different from zero (Franke & Roettger 2019).

2.2 Results

Firstly, we report the observed response times of the correct rejections of unmentioned probes in Figure 1 (Experiment 1) and 2 (Experiment 2). The data are divided by block, i.e. into the first and second half of trials for each participant. The data are reported following outlier removal. This was done to show the variability of data observed in Experiment 1, but not in Experiment 2.

The Bayesian models fitted to the data produced posterior distributions of the parameters given both the priors and the data (see Figures 3 and 4). In both Experiment 1 and 2, the models fitted to the response time data provided compelling evidence for the main effect of alternative status being larger than zero (Exp1: $\beta = 0.082$, CrI [0.055, 0.110]; Exp 2: $\beta = 0.104$, CrI [0.070, 0.139]). This means that alternative probes were being rejected more slowly compared to unrelated ones. In neither experiment did we see compelling evidence for the main effect of particle presence, since in both experiments, the credible intervals included zero (Exp1: $\beta = -0.008$, CrI $[-0.023, 0.007]$; Exp 2: $\beta = -0.006$, CrI $[-0.028, 0.015]$). As for the interaction of alternative status and particle presence, the posterior distributions also included zero within their 95% CrIs (Exp1: $\beta = 0.001$, CrI $[-0.013, 0.015]$; Exp 2: $\beta = 0.0004$, CrI $[-0.023, 0.022]$). Where the results differed between the experiments was in the three-way interaction of the two manipulations and centred trial order. In Experiment 1, there is compelling evidence for this three-way interaction effect being larger than zero ($\beta = 0.0012$, CrI [0.0001, 0.0024]). This means that there was more interference at the beginning of the experiment and that the effect seemed to have been evolving over the course of the experimental session. In Experiment 2, however, the model did not compellingly show this three-way interaction ($\beta = 0.0007$, CrI $[-0.0007, 0.0022]$).

We find some evidence that replicates the previously identified interference effects of *only* in the case of noun alternatives (Experiment 1). These were not

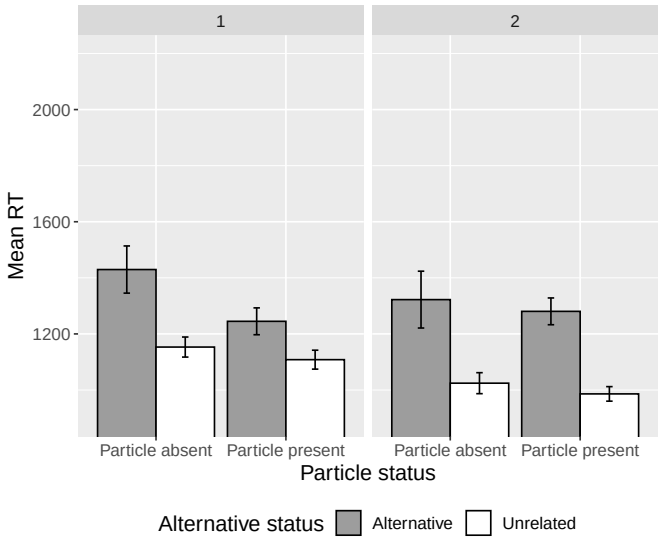


Figure 1: Mean response times and standard errors of correct rejections after outlier removal by condition and block in Experiment 1 (Nouns)

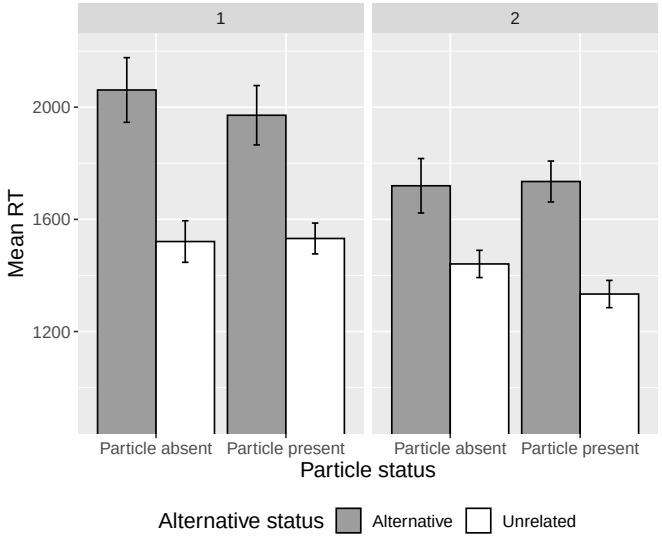


Figure 2: Mean response times and standard errors of correct rejections after outlier removal by condition and block in Experiment 2 (Verbs)

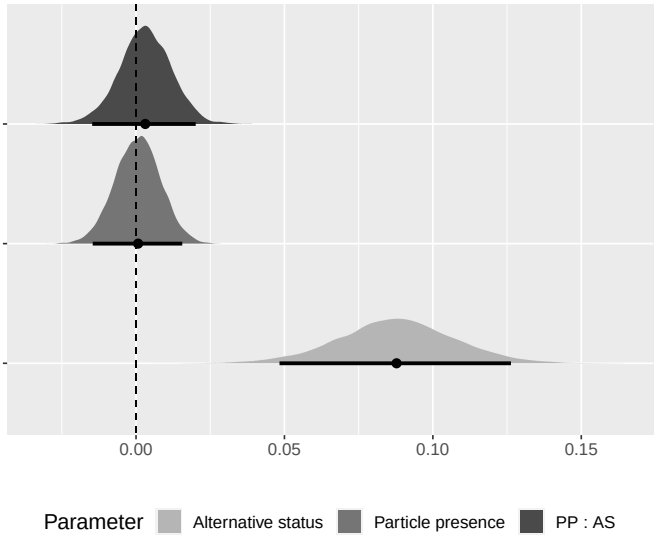


Figure 3: Posterior probabilities and 95% CrIs for the parameters of interest in Experiments 1 and 2

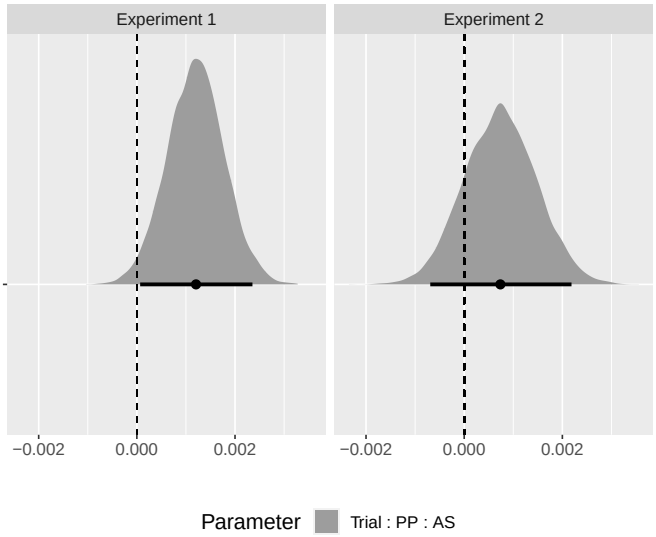


Figure 4: Posterior probabilities and 95% CrIs for the three-way interaction between particle presence, alternative status, and trial position in Experiments 1 and 2

extended to the case of verbs (Experiment 2), for which we saw no evidence of interference. Therefore, these results are on the whole inconsistent with our predictions, since we expected both the noun and verb constituent parts of focused VPs to show the interference of *only*. However, Experiments 1 and 2 cannot provide a conclusive falsification of the hypothesis. This is due to it being possible that the constituent parts of focused VPs could be represented differently depending on their word class and that they could give rise to different patterns of interactions with focus particles while whole VP alternatives are being represented. In other words, the interference pattern of *only* might still hold for the whole phrases while being present in the alternatives to only some constituent parts of the focused phrases. Therefore, in order to fully test the hypothesis, we conducted Experiment 3.

3 Experiment 3

Since Experiments 1 and 2 only tested probes that were alternatives to the constituent parts of focused VPs rather than to the VPs as a whole, we ran Experiment 3, in which we examined the representation of whole alternative phrases. This experiment was pre-registered and the time-stamped entry can be found on the OSF platform (https://osf.io/cf36w). The design and predictions were the same as those in Experiments 1 and 2.

3.1 Method

3.1.1 Participants

121 participants aged 20 to 30 (mean age 24.6) recruited on the Prolific platform (£3.13 in compensation) took part in the experiment (see OSF for more information). From this set, we excluded one participant for not being a native speaker of English.

3.1.2 Materials

The materials used in Experiment 3 had the same structure as the first two experiments. However, we reduced the number of items to 24 in order to lessen the load on participants and shorten the length of the experimental session. Take the following example of a critical sentence:

(13) Lily is a tailor.

(14) At the workshop, Lily could sew and stitch the shirt and the skirt.

(15) Lily surely stitched the skirt.

(16) No, she _/only sewed the shirt.

We probed either whole alternative phrases (*knitted the scarf*) or unrelated ones (*published the study*). Following the creation of the probes for this experiment by combining the noun and verb probes from the first two experiments, we ran a naturalness rating study with this set of stimuli. We found a statistically significant difference in naturalness ratings between the alternative and unrelated probes with the latter being rated as less natural. Believing this to be a potential confound for the main experiment, we replaced the lowest rated probes with highly rated filler items that were a part of the rating study. Finally, we conducted a new set of analyses of naturalness ratings, log-frequencies and letter-length of the nouns and verbs, and the LSA measures of relatedness of the noun probes to the object noun and verb probes to the main transitive verbs found in the focused phrase and no confound was found (see the OSF entry for more information).

3.1.3 Procedure

The procedure was nearly identical to that of Experiment 2 with only the response keys changed to *y* for *yes* and *n* for *no*.

3.1.4 Analysis

Firstly, two items (14 and 15) were excluded from further analyses, since an error in the construction of their probes was discovered after data collection.

As per our pre-registration, the log-transformed response times of correct trials were analysed by a Bayesian hierarchical model with trial order, alternative status, particle presence, the two-way interaction of alternative status and particle presence, and the three-way interaction of trial order, alternative status and particle presence as fixed effects. The full specification of the model can be viewed on the OSF platform.

3.2 Results

The reader can see the observed response times for Experiment 3 summarised in Figure 5. We report the means of correct rejections after outlier removal by condition and block.

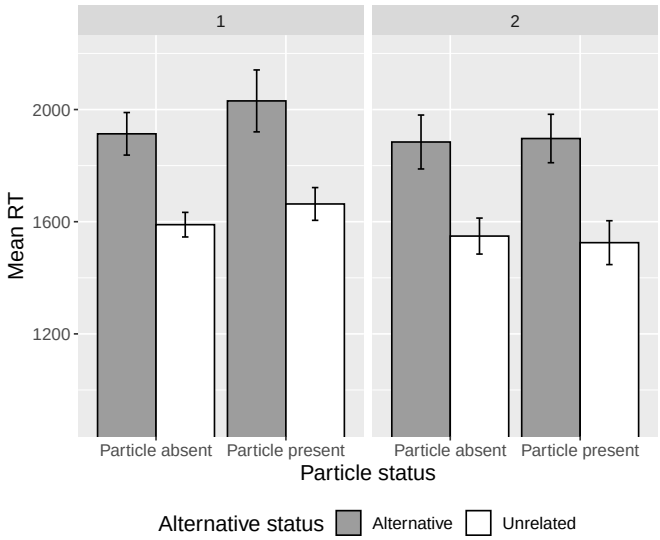


Figure 5: Mean response times and standard errors of correct rejections after outlier removal by condition and block in Experiment 3 (Phrases)

In Figure 6, we display the distributions of the posterior probabilities of the parameters of interest from the Bayesian model fitted to the log-transformed response time data. These are the main effects of particle presence and alternatives status, and their interaction. The model showed compelling evidence that alternative probes are rejected substantially more slowly compared to unrelated probes (β = 0.088, CrI [0.048, 0.126]). On the other hand, the maximum likelihood estimate for the size of the effect of the presence of *only* is close to zero (β = 0.0006, CrI [−0.0146, 0.0156]). Crucially for our hypothesis, the model together with the data and given our priors does not provide us with compelling evidence to assert that the effect of the interaction between alternative status and particle presence is different from zero (β = 0.00, CrI [−0.01, 0.02]). Neither was there evidence (see Figure 7) for a three-way interaction of particle presence, alternative status, and centred trial order (β = −0.0001, CrI [−0.0021, 0.0019]).

These results are again not in line with our predictions, since they do not give us evidence for *only* causing further interference in the rejection of whole alternative probes. Neither do we see this effect emerging through an interaction with trial order.

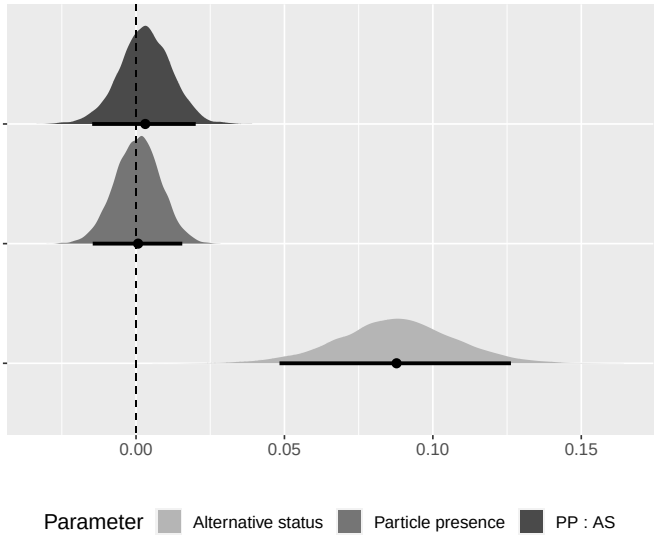


Figure 6: Posterior probabilities and 95% CrIs for the parameters of interest in Experiment 3

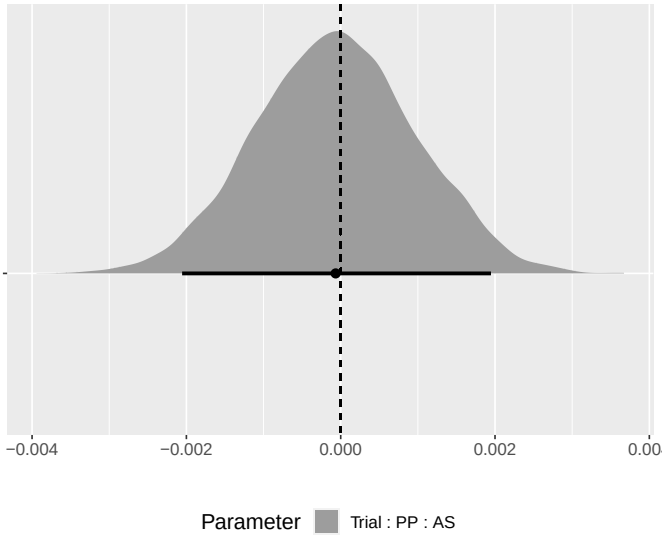


Figure 7: Posterior probabilities and 95% CrIs for the three-way interaction between particle presence, alternative status, and trial position in Experiment 3

4 General discussion

4.1 Results summary

We conducted three probe recognition experiments aimed at testing whether the Alternative Representation approach to the processing of focus based on Rooth's (1992) semantic theory, the Alternative Representation Hypothesis, can be generalised to cases other than narrow focus, namely broadly focused VPs. We investigated this by means of using the interference effect of *only* on the rejection of unmentioned alternatives (Gotzner et al. 2016, Gotzner & Spalek 2017) as a litmus test of whether focus alternatives were being activated, selected and represented by comprehenders.

Experiments 1 and 2, which targeted alternatives to the constituent nouns and verbs respectively, revealed a pattern that was not predicted. In neither experiment did we observe the simple interaction of our two manipulations, alternative status and particle presence. However, in the case of Experiment 1 only, the model and the data provided compelling evidence for a three-way interaction of the factors with trial order. We take this to be indicative of an interference effect, which additionally seems to be evolving over the course of the experimental session. Therefore, we have evidence for the interference by the presence of *only* for alternatives to nouns, but not for verbs. Finally, Experiment 3 with entire phrases also did not provide compelling evidence for the presence of the predicted interference effect. Neither was there any evidence for a three-way interaction.

4.2 Implications for the Alternative Representation account

The results of our experiment are mixed and overall speak against the straightforward extrapolation of previous research and the Alternative Representation model to cases of broad focus. This is because this model predicted that the interference effects of *only* would be present in broad focus. This would be equally the case in the experiment which tested constituent parts of the alternatives, as well as when whole phrases would be tested. The current research tentatively suggests that our credence in the presence of this interference effect on the rejection of plausible alternatives ought to be lowered in cases of broad focus. While the current results fail to conclusively support the Alternative Representation Hypothesis, it remains nevertheless plausible that comprehenders represent alternatives to broadly focused phrases without the emergence of the interference pattern of focus sensitive particles identified in the cases of narrow focus.

We take the fact that in all three experiments, especially in Experiment 3, the main effect of relatedness was observed to be indirect evidence in favour of our

comprehenders representing phrasal alternatives. That we observed inhibition in the case of phrases suggests that comprehenders were in fact representing them – adopting the view of Gotzner et al. (2016) for single nouns, this might possibly have been due to competition with the mentioned alternatives. This evidence would have been more conclusive had the interaction with the focus particle *only* been present. Since it remains possible that the behaviour of focus particles in processing differs between narrow and broad focus, a direct test of the presence of alternatives in broad focus constructions would give us even more credence. We put forward ways in which this could be done by explicitly manipulating focus in Section 4.4.

What requires discussion is the fact that as far as Experiments 1 and 2 are concerned, the interference was observed only in the case of alternatives to the constituent nouns within broadly focused phrases and that this effect was modulated by trial order in the case of Experiment 1.

The first plausible explanation assumes that alternatives to broadly focused phrases are being represented by comprehenders and postulates that it is due to the peculiar processing features of nouns that these effects only occur while not being present in either verbs or whole phrases. In other words, the source of the divergence may be sought in the general processing profile that has been claimed to exist between nouns and verbs. For instance, Soloukhina & Ivanova (2018) studied the comprehension of verbs and nouns both in healthy individuals as well as in people with aphasia. They found differences in reaction times between verbs and nouns in a matching picture task in both populations, with verbs being reacted to more slowly.

The second explanation lies in the interpretation given by our participants to the presented stimuli. Even though the discourses were designed to elicit broad focus interpretations in the final sentence of each item, an interpretation with narrow focus on the direct object might be possible due to the phenomenon of focus projection (Selkirk 1995, Gussenhoven 1999) mentioned in the introduction, which refers to the direct object receiving focus prosody both in cases of narrow and broad focus, making the scope of focus ambiguous, according to some researchers. Furthermore, for reasons of comparability, we modelled our stimuli on the study of Gotzner et al. (2016), which uses corrective focus. While corrective and contrastive uses of focus have been equated in the literature (e.g., Zimmermann 2008), the use of negation might have biased our participants towards a narrow focus interpretation. Our results could therefore be due to participants in fact interpreting the final sentences as encoding narrow focus on the noun.

The influence of trial order could also be accommodated within this explanation, since while comprehenders might have started with the broad focus in-

terpretation, the presentation of single noun probes might have caused them to change their preferred reading of the sentences towards narrow focus. This would be consistent with the interference effect being absent both in the case of probing whole phrases as well as in probing the verbal constituents in Experiment 2 and only present in Experiment 1. Since verb alternatives could have therefore been completely bypassed from the comprehenders' processing of our stimuli, they would not be expected to be interfered with by *only*. The interaction found in Experiment 1 (nouns) would therefore be consistent with the rest of the literature on the effects of *only* in the processing of narrow focus (Gotzner et al. 2016, Gotzner & Spalek 2017).

There are several potential reasons for the lack of an interaction effect in verbs. One could be that NPs and VPs are differently restricted. While the object NPs are faced with selectional verb restrictions, this is not the case for the whole VP, which is only restricted by (non-linguistic) context. This could have been one of the sources of the difference found between nouns and verbs in Experiments 1 and 2. Another reason for why interference effects were not found in the rejection of verbal probes might have been the greater recency of the direct object noun, as one reviewer pointed out. Likewise, processing strategies could have influenced the results. We believe these issues ought to be addressed in further research by the employment of a cross-modal presentation method with auditory stimuli and textual probes. This would also address the issue of the lack of explicit prosody that makes it possible for comprehenders to assign implicit prosody to different elements and obscure the results.

Overall, the above-mentioned methodological issues mean that as far as the implications for theories of alternative representation in comprehension go, caution is to be taken. However, the presented evidence suggests that the probe recognition methodology can be extended to larger constituents and provides some initial evidence that comprehenders represent alternatives to broad-focused phrases.

4.3 Implications for the RSA-based approaches

The above mentioned explanation of the pattern of our results warrants further discussion in connection to the approach taken by the proponents of the RSA framework to focus. Given the phenomenon of focus projection, the distinction between broad and narrow focus in English is often not possible to make, given that only the rightmost element receives prosodic marking. In the case of our stimuli, had they been presented auditorily, this prosodic marking would be on the direct object noun. Therefore, prosodic prominence is arguably realised only on the noun and not on the verb (under the assumption that broad and narrow

focus are prosodically indistinguishable). Assuming that this translates to implicit prosody employed by our participants when reading our sentences, this could be seen as supporting a view of focus being exclusively tied to prosodic prominence via the mechanism proposed within the RSA framework (Bergen & Goodman 2015, Stevens 2016, Stevens & Roberts 2019).

The reasoning is as follows – given the ambiguity of the scope of focus, listeners do not perceive a difference in the expended effort on part of the speaker on anything but the focused object noun. Consequently, the pragmatic reasoning employed by the listener would be the same as in the cases of narrow focus. This would then be consistent with our findings, which showed the predicted pattern only in the case of alternatives to the nouns within broad-focused phrases. This, however, would also lead to the RSA approach needing to be amended to be able to deal with the effects associated with broad focus, such as its felicity given certain QUDs.

All in all, the results of the current study are not decisive with regards to the RSA approach to focus. We have highlighted here how our experiments may be interpreted in light of these proposals, yet we are also of the opinion that this theory needs to be spelled out more for researchers to be able to test it properly.

4.4 Future directions

The possibility that the comprehenders in this study interpreted the stimuli with narrow focus necessitates further studies that would use either explicit questions-under-discussion to elicit broad focus or unambiguous broad focus constructions such as pseudoclefts:

(17) What Jane did was read the manuscript.

If seen, differential patterns of representation or activation between appropriate and inappropriate alternatives dependent on the type of the clefted constituent would provide evidence in favour of the Roothian-inspired Alternative Representation approach.

The observed differences between the noun and verb alternatives also highlight a gap in the literature. The research on the comprehension of focus has mostly examined focused nouns, leaving other word classes understudied (but see Ito & Speer 2008, Fraundorf et al. 2010, Kurumada et al. 2014). To our knowledge, there has not yet been a study examining the processing of alternatives with narrow focus on either finite verbs or infinitives. Furthermore, the issue of how narrowly focused verbs differ in their processing profile from narrowly focused nouns remains to be examined.

4.5 Methodological implications

Moving finally onto further implications of this research, the unambiguous evidence for alternative probes in all three experiments being rejected more slowly compared to unrelated ones is, we believe, methodologically significant. While the probe recognition task has been conducted with either single words as probes (e.g. Cowan & Saults 2013) or with entire sentences (e.g. Radvansky et al. 2005), there has been, to our knowledge, no research that used phrases in the probe recognition task. That we observed the main effect of alternative status in Experiment 3 suggests that probing whole phrasal constructions in the probe recognition paradigm can be done. The pattern of observed results is in line with what has been observed in this task in the situation when single words are used for probes. We observed an interference of semantically associated phrases on the rejection of unmentioned probes in the expected direction and thus, our research constitutes the first piece of evidence in favour of this extension of the probe recognition. This then opens novel possibilities for the use of this paradigm in psycholinguistic research.

5 Conclusion

This study reported the first attempt to test the generalisability of the Alternative Representation theory of the processing of focus based on the semantic approach of Rooth (1992) to larger focused phrases. Regarding constituent parts, the results of Experiments 1 and 2 were contrary to our predictions. We have some evidence that focus particles operate on noun alternatives, but we do not have evidence that they operate on verb alternatives. Finally, Experiment 3 showed that while whole-phrase alternatives were being rejected more slowly, there was no compelling evidence in favour of the interference effect of *only*. This main effect provides some initial evidence that comprehenders represent phrasal alternatives. Yet overall, the results do not allow us to support the generalisability of the Alternative Representation model to cases of broad focus as the relatedness effect cannot be solely attributed to focus in our design. On the methodological side, this paper provides the contribution of the evidence in favour of using the probe recognition task with constituents larger than a single word yet smaller than a whole sentence.

Acknowledgements

Firstly, we thank Hannah Rohde for her advice in the initial planning stages of this research. We would like to thank Henrik Discher for his help with the creation of the stimuli for Experiment 3. Next, we are very grateful for the feedback E. Matthew Husband provided to us on an earlier draft of this article. We are also thankful to the audiences of AMLaP2021, HSP2022, and DGfS2022 conferences for their valuable comments and suggestions. Finally, we would like to thank the two anonymous reviewers of this chapter and Ingo Reich, one of the editors of this volume, who helped us improve it substantially. The first and third author were supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the Emmy Noether project awarded to Nicole Gotzner (Grant Nr. GO 3378/1-1).

References

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. DOI: 10.1016/j.jml.2012.11.001.

Bergen, Leon & Noah D. Goodman. 2015. The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science* 7(2). 336–350. DOI: 10.1111/tops.12144.

Bishop, Jason. 2017. Focus projection and prenuclear accents: Evidence from lexical processing. *Language, Cognition and Neuroscience* 32(2). 236–253. DOI: 10.1080/23273798.2016.1246745.

Braun, Bettina, Yuki Asano & Nicole Dehé. 2019. When (not) to look for contrastive alternatives: The role of pitch accent type and additive particles. *Language and Speech* 62(4). 751–778. DOI: 10.1177/0023830918814427.

Braun, Bettina & Lara Tagliapietra. 2010. The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes* 25(7-9). 1024–1043. DOI: 10.1080/01690960903036836.

Breen, Mara, Evelina Fedorenko, Michael Wagner & Edward Gibson. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25(7-9). 1044–1098. DOI: 10.1080/01690965.2010.504378.

Büring, Daniel. 2007. Semantics, intonation, and information structure. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford handbook of linguistic interfaces*, 445–474. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199247455.013.0015.

Bürkner, Paul-Christian. 2017. Brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software* 80(1). 1–28. DOI: 10.18637/jss.v080.i01.

Byram Washburn, Mary, Elsi Kaiser & Maria Luisa Zubizarreta. 2011. Focus facilitation and non-associative sets. In *Semdial 2011: Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue*, 94–102.

Calhoun, Sasha, Mengzhu Yan, Honiara Salanoa, Fualuga Taupi & Emma Kruse Va'ai. 2022. Focus effects on immediate and delayed recognition of referents in Samoan. *Language and Speech* I(27). 175–201. DOI: 10.1177/00238309221101396.

Carlson, Katy. 2013. The role of 'only' in contrasts in and out of context. *Discourse processes* 50(4). 249–275. DOI: 10.1080/0163853X.2013.778167.

Carlson, Katy. 2015. Clefting, parallelism, and focus in ellipsis sentences. In Lyn Frazier & Edward Gibson (eds.), *Explicit and implicit prosody in sentence processing*, 63–83. Cham: Springer.

Carlson, Katy & David Potter. 2022. Focus Attracts Attachment. *Language and Speech* 65(2). 491–512. DOI: 10.1177/00238309211033321.

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 27–55. New York: Academic Press.

Cowan, Nelson & Scott Saults. 2013. When does a good working memory counteract proactive interference? Surprising evidence from a probe recognition task. *Journal of Experimental Psychology: General* 142(1). 12. DOI: 10.1037/a0027804.

Drummond, Alexander, Titus von der Malsburg, Michael Yoshitaka Erlewine & Mahla Vafaie. 2016. *Ibex farm*. https://github.com/addrummond/ibex.

Erteschik-Shir, Nomi. 2007. *Information structure: The syntax-discourse interface*, vol. 3. Oxford: Oxford University Press.

Filik, Ruth, Kevin B. Paterson & Simon P. Liversedge. 2005. Parsing with focus particles in context: Eye movements during the processing of relative clause ambiguities. *Journal of Memory and Language* 53(4). 473–495. DOI: 10.1016/j.jml.2005.07.004.

Fox, Danny & Roni Katzir. 2011. On the characterization of alternatives. *Natural Language Semantics* 19. 87–107. DOI: 10.1007/s11050-010-9065-3.

Franke, Michael & Gerhard Jäger. 2016. Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft* 35(1). 3–44. DOI: 10.1515/zfs-2016-0002.

Franke, Michael & Timo B. Roettger. 2019. Bayesian regression modeling (for factorial designs): A tutorial. *PsyArXiv*. DOI: 10.31234/osf.io/cdxv3.

Fraundorf, Scott H., Duane G. Watson & Aaron S. Benjamin. 2010. Recognition memory reveals just how contrastive contrastive accenting really is. *Journal of Memory and Language* 63(3). 367–386. DOI: 10.1016/j.jml.2010.06.004.

Frazier, Lyn, Charles Jr. Clifton & Katy Carlson. 2007. Focus and VP ellipsis. *Language and Speech* 50(1). 1–21. DOI: 10.1177/00238309070500010l0.

Gernsbacher, Morton Ann & Jörg D. Jescheniak. 1995. Cataphoric devices in spoken discourse. *Cognitive Psychology* 29(1). 24–58. DOI: 10.1006/cogp.1995.1011.

Goodman, Noah D. & Michael C. Frank. 2016. Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Science* 20(11). 818–829. DOI: 10.1016/j.tics.2016.08.005.

Gotzner, Nicole. 2017. *Alternative sets in language processing: How focus alternatives are represented in the mind*. London: Palgrave Macmillan.

Gotzner, Nicole & Jacopo Romoli. 2022. Meaning and alternatives. *Annual Review of Linguistics* 8(1). 213–234. DOI: 10.1146/annurev-linguistics-031220-012013.

Gotzner, Nicole & Katharina Spalek. 2017. Role of contrastive and noncontrastive associates in the interpretation of focus particles. *Discourse Processes* 54(8). 638–654. DOI: 10.1080/0163853X.2016.1148981.

Gotzner, Nicole & Katharina Spalek. 2019. The life and times of focus alternatives: Tracing the activation of alternatives to a focused constituent in language comprehension. *Language and Linguistics Compass* 13(2). e12310. DOI: 10.1111/lnc3.12310.

Gotzner, Nicole, Isabell Wartenburger & Katharina Spalek. 2016. The impact of focus particles on the recognition and rejection of contrastive alternatives. *Language and Cognition* 8(1). 59–95. DOI: 10.1017/langcog.2015.25.

Gussenhoven, Carlos. 1999. Discreteness and gradience in intonational contrasts. *Language and speech* 42(2-3). 283–305. DOI: 10.1177/00238309990420020701.

Halliday, Michael Alexander Kirkwood. 1967. *Intonation and grammar in British English*. The Hague: Mouton.

Husband, E. Matthew & Fernanda Ferreira. 2016. The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience* 31(2). 217–235. DOI: 10.1080/23273798.2015.1083113.

Ito, Kiwako & Shari R. Speer. 2008. Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 58(2). 541–573. DOI: 10.1016/j.jml.2007.06.013.

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge: MIT Press.

Johnson-Laird, Philip Nicholas. 1983. *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Harvard University Press.

Jördens, Kim A., Nicole Gotzner & Katharina Spalek. 2020. The role of non-categorical relations in establishing focus alternative sets. *Language and Cognition* 12(4). 729–754. DOI: 10.1017/langcog.2020.21.

Káldi, Tamás, Ágnes Szöllösi & Anna Babarczy. 2021. Hungarian structural focus: Accessibility to focused elements and their alternatives in working memory and delayed recognition memory. *Frontiers in Psychology*. 1–17. DOI: 10.3389/fpsyg.2021.514886.

Kim, Christina S., Christine Gunlogson, Michael K. Tanenhaus & Jeffrey T. Runner. 2015. Context-driven expectations about focus alternatives. *Cognition* 139. 28–49. DOI: 10.1016/j.cognition.2015.02.009.

Klin, Celia M., Kristin M. Weingartner, Alexandria E. Guzmán & William H. Levine. 2004. Readers' sensitivity to linguistic cues in narratives: How salience influences anaphor resolution. *Memory & Cognition* 32(3). 511–522. DOI: 10.3758/BF03195843.

Kratzer, Angelika & Irene Heim. 1998. *Semantics in generative grammar*, vol. 1185. Oxford: Blackwell.

Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3-4). 243–276.

Kurumada, Chigusa, Meredith Brown, Sarah Bibyk, Daniel F. Pontillo & Michael K. Tanenhaus. 2014. Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition* 133(2). 335–342. DOI: 10.1016/j.cognition.2014.05.017.

Lacina, Radim, Radek Šimík & Nicole Gotzner. 2023. "Czech" the alternatives: A probe recognition study of focus and word order. *Proceedings of Sinn und Bedeutung* 27. 365–380. DOI: 10.18148/sub/2023.v27.1075.

Ladd, D. Robert. 2008. *Intonational phonology*. Cambridge: Cambridge University Press.

Landauer, Thomas K. & Susan T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104(2). 221–240. DOI: 10.1037/0033-295X.104.2.211.

Lee, Eun-Kyung & Scott H. Fraundorf. 2017. Effects of contrastive accents in memory for L2 discourse. *Bilingualism: Language and Cognition* 20(5). 1063–1079. DOI: 10.1017/S1366728916000638.

Love, Bradley. 2015. The algorithmic level is the bridge between computation and brain. *Topics in Cognitive Science* 7(2). 230–242. DOI: 10.1111/tops.12131.

Marr, David. 2010. *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge: MIT Press.

Meyer, David E. & Roger W. Schvaneveldt. 1971. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology* 90(2). 227. DOI: 10.1037/h0031564.

Norberg, Kole A. & Scott H. Fraundorf. 2021. Memory benefits from contrastive focus truly require focus: Evidence from clefts and connectives. *Language, Cognition and Neuroscience* 36(8). 1010–1037. DOI: 10.1080/23273798.2021.1901944.

R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna. https://www.R-project.org/.

Radvansky, Gabriel A., Rose T. Zacks & Lynn Hasher. 2005. Age and inhibition: The retrieval of situation models. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 60(5). 276–278. DOI: 10.1093/geronb/60.5.P276.

Roberts, Craige. 2012. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5. 1–69. DOI: 10.3765/sp.5.6.

Rochemont, Michael. 1986. *Focus in generative grammar*, vol. 4. Vancouver: John Benjamins Publishing.

Rooth, Mats. 1985. *Association with focus (Montague grammar, semantics, only, even)*. University of Massachusetts Amherst. (Doctoral dissertation).

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116. DOI: 10.1007/BF02342617.

Sanford, Alison J. S., Jessica Price & Anthony J. Sanford. 2009. Enhancement and suppression effects resulting from information structuring in sentences. *Memory & Cognition* 37(6). 880–888. DOI: 10.3758/MC.37.6.880.

Schwarzschild, Roger. 1999. GIVENness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics* 7(2). 141–177. DOI: 10.1023/A:1008370902407.

Selkirk, Elisabeth O. 1995. Sentence prosody: Intonation, stress, and phrasing. In John Goldsmith (ed.), *The handbook of phonological theory*, 550–569. Blackwell.

Sgall, Petr, Eva Hajičová & Eva Benešová. 1973. *Topic, focus, and generative semantics*. Kronberg: Scriptor.

Sgall, Petr, Eva Hajičová & Jana Panevová. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. Jacob Mey (ed.). Springer Science & Business Media.

Shannon, Claude E. & Warren Weaver. 1949. *The mathematical theory of communication*. Champaign: University of Illinois Press.

Soloukhina, Olga A. & Maria V. Ivanova. 2018. Investigating comprehension of nouns and verbs: Is there a difference? *Aphasiology* 32(2). 183–203. DOI: 10.1080/02687038.2017.1396572.

Spalek, Katharina & Yulia Oganian. 2019. The neurocognitive signature of focus alternatives. *Brain and language* 194. 98–108. DOI: 10.1016/j.bandl.2019.04.007.

Stevens, Jon S. 2016. Focus games. *Linguistics and philosophy* 39(5). 395–441. DOI: 10.1007/s10988-016-9192-5.

Stevens, Jon S. & Gareth Roberts. 2019. Noise, economy, and the emergence of information structure in a laboratory language. *Cognitive Science* 43(2). 1–33. DOI: 10.1111/cogs.12717.

Sturt, Patrick, Anthony J. Sanford, Andrew Stewart & Eugene Dawydiak. 2004. Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic bulletin & review* 11(5). 882–888. DOI: 10.3758/BF03196716.

Tjuka, Annika, Huong Thi Thu Nguyen & Katharina Spalek. 2020. Foxes, deer, and hedgehogs: The recall of focus alternatives in vietnamese. *Laboratory Phonology* 11(1). 1–29. DOI: 10.5334/labphon.323.

von Fintel, Kai. 1997. Bare plurals, bare conditionals, and 'only'. *Journal of Semantics* 14(1). 1–56. DOI: 10.1093/jos/14.1.1.

von Stechow, Arnim & Susanne Uhmann. 1986. Some remarks on focus projection. In Meij Sjaak & Abraham Werner (eds.), *Topic, focus, and configurationality*, 295–320. Amsterdam: Benjamins.

Ward, Peter & Patrick Sturt. 2007. Linguistic focus and memory: An eye movement study. *Memory & Cognition* 35(1). 73–86. DOI: 10.3758/BF03195944.

Yan, Mengzhu & Sasha Calhoun. 2019. Priming effects of focus in Mandarin Chinese. *Frontiers in Psychology* 10. 1–16. DOI: 10.3389/fpsyg.2019.01985.

Yan, Mengzhu, Sasha Calhoun & Paul Warren. 2023. The role of prominence in activating focused words and their alternatives in Mandarin: Evidence from lexical priming and recognition memory. *Language and Speech* 66(3). 678–705. DOI: 10.1177/00238309221126108.

Zehr, Jeremy & Florian Schwarz. 2018. *PennController for internet based experiments (IBEX)*. DOI: 10.17605/OSF.IO/MD832.

Zimmermann, Malte. 2008. Contrastive focus and emphasis. *Acta Linguistica Hungarica* 55(3-4). 347–360.

Chapter 3

An information-theoretic account of constituent order in the German middle field

Katrin Ortmann[a], Sophia Voigtmann[b], Stefanie Dipper[a] & Augustin Speyer[b]

[a]Ruhr-Universität Bochum [b]Universität des Saarlandes

This paper proposes a novel approach to explain object order in German. Although the order of constituents is relatively free in modern German, there are clear preferences for the order dative before accusative (nominal) objects and for the order given before new objects. A range of influential factors have been described in the literature, most prominently givenness and length. We assume processing-related reasons and use information-theoretic measures, in particular surprisal and DORM (Cuskley et al. 2021), to explore the interplay of information structure and information density as factors for object order. We propose a measure called $\text{DORM}_{\text{diff}}$ and the *corpus of variants* method for comparing information profiles between different plausible constituent orders. Our investigations show that language users follow information-theoretic principles (UID, Levy & Jaeger 2007) in choosing the object order that leads to a more uniform distribution of information. We argue that this preference also explains deviations from the unmarked object order (i.e., accusative preceding dative and new preceding given) if it is associated with smoother information profiles.

1 Introduction

In contrast to languages with a fixed word order, the order of constituents in a language like German is relatively free. Nevertheless, there still exist clear preferences for certain word and constituent orders in German. One such preference

Katrin Ortmann, Sophia Voigtmann, Stefanie Dipper & Augustin Speyer. 2024. An information-theoretic account of constituent order in the German middle field. In Robin Lemke, Lisa Schäfer & Ingo Reich (eds.), *Information structure and information theory*, 55–86. Berlin: Language Science Press. DOI: 10.5281/zenodo.13383787

concerns the relative order of nominal dative and accusative object. For example, sentence (1a) is generally preferred over sentence (1b), even though both constituent orders are possible and occur in natural data.

(1) a. *Ich werde* [*einem Jungen*]$_{\text{DAT}}$ [*ein Buch*]$_{\text{ACC}}$ *geben.*
I will [a boy]$_{\text{DAT}}$ [a book]$_{\text{ACC}}$ give
'I will give a boy a book.'

b. *Ich werde* [*ein Buch*]$_{\text{ACC}}$ [*einem Jungen*]$_{\text{DAT}}$ *geben.*
I will [a book]$_{\text{ACC}}$ [a boy]$_{\text{DAT}}$ give
'I will give a book to a boy.'

There are numerous works on this phenomenon which try to capture the observed preferences. Among the known influential factors are animacy, familiarity, givenness, salience and length (cf., e.g., Lenerz 1977, Speyer 2011, Behagel 1932, and for English, Bresnan 2007). However, these factors cannot *explain* the preferences but only describe them. In this paper, we try to go beyond a mere description and attempt to explain this phenomenon based on the cognitive processing effort of the constructions (cf., e.g., Fenk-Oczlon 1983).

Some of the factors mentioned above certainly have an influence on processing effort, e.g., givenness as illustrated in (2). These sentences all have the marked case order accusative before dative, but differ with respect to givenness. Regarding givenness, the order given before new represents the common order (Section 5), so (2a) should be easier to process than the other examples since it is the most common object order, and familiarity can facilitate processing (cf., e.g. Futrell et al. 2021). However, such factors, and in fact all of the factors mentioned above except length, are difficult to quantify and thus hard to operationalize.

(2) a. *Ich werde* [*das Buch*]$_{\text{ACC, given}}$ [*einem Jungen*]$_{\text{DAT, new}}$ *geben.*
I will [the book]$_{\text{ACC, given}}$ [a boy]$_{\text{DAT, new}}$ give
'I will give the book to a boy.'

b. *Ich werde* [*ein Buch*]$_{\text{ACC, new}}$ [*dem Jungen*]$_{\text{DAT, given}}$ *geben.*
I will [a book]$_{\text{ACC, new}}$ [the boy]$_{\text{DAT, given}}$ give
'I will give a book to the boy.'

c. *Ich werde* [*das Buch*]$_{\text{ACC, given}}$ [*dem Jungen*]$_{\text{DAT, given}}$ *geben.*
I will [the book]$_{\text{ACC, given}}$ [the boy]$_{\text{DAT, given}}$ give
'I will give the book to the boy.'

In the present study, we explore the application of information-theoretic concepts to objectively quantify and approximate the effects of processing effort on

object order in the middle field of the German sentence. We expect that a certain constituent order is used to assure an optimal information flow and to avoid processing difficulties. As a measure of processing difficulties, we use information density (Shannon 1948). In this framework, information is derived from the probability of a word in context. Information theory has been widely used to relate the probability of linguistic material occurring in an utterance (measured as surprisal: $S(unit) = -\log_2 P(unit|context)$, Hale (2001)) to the effort required to process that utterance. Lower predictability (probability) correlates with higher processing effort (e.g., Hale 2001). Also, very high surprisal values or an uneven information profile are correlated with information loss, as (Cuskley et al. 2021) argue. Therefore, speakers aim to keep the information flow as uniform as possible to ensure optimal communication ("Uniform Information Density Hypothesis", UID, Levy & Jaeger 2007, Aylett & Turk 2004).

Since the predictability of a word depends strongly on its context, the order of words and constituents has a high impact on the uniformity of the utterance (Cuskley et al. 2021). Changing the order can thus lead to more successful communication and, based on this assumption, we propose that changes in object order in the German middle field can be described and even explained by information density. We test our hypothesis in a pilot study based on a large corpus of modern German.

The remainder of this paper is structured as follows: Section 2 gives an introduction of the theoretic background and explains the different factors that are known to influence constituent order in the German middle field. Section 3 describes the data selection for this study, and Section 4 details the methods used for analysis, including the calculation of constituent surprisal and information profiles. In Section 5, the results are presented and the effects of information-theoretic principles on constituent order are evaluated. Possible problems and enhancements of the methodology are discussed in Section 6. The paper concludes with a summary of the findings in Section 7.[1]

2 Constituent order in the German middle field

As already mentioned, German is a language with a relatively free constituent order. This means that constituent order is not exclusively governed by structural factors such as grammatical function (subject, direct object, etc.) as is the

[1]The statistical data and the R script used in this study as well as the list of light verb constructions applied in data preparation are available at https://gitlab.ruhr-uni-bochum.de/comphist/c6dormdiff.

case, e.g., in English. Instead, constituent order in German is influenced by several factors, many of which are non-syntactic factors but rather of a semantic or pragmatic nature (see, e.g., Lenerz 1977, Rauth 2020). This goes for historical stages of German as well (Speyer 2011, 2013, Rauth 2020).

The point of interest for our study is the so-called middle field in the German clause. The term *middle field* has its origin in the topological field model of the German clause (for a recent overview, see, e.g., Wöllstein 2010, 2014). We introduce the model using the terminology of Telljohann et al. (2017).

Word order in German sentences is best described not by notions such as SVO (subject > verb > object[2]) or the like, but rather by relating the constituents relative to the verb positions. Verb forms tend to be distributed over the German (matrix) clause in such a way that the finite part stands relatively early in the clause (linke (Satz-)Klammer ('left sentence bracket'), abbreviated LK) and the remainder of the verb form at the end or close to the end of the clause, in a position often referred to as the *right sentence bracket* ('rechte Satzklammer'). In the scheme of Telljohann et al. (2017), this position is called *VC* (for verb complex). The positions of the nonverbal constituents of the clause can be described relative to these verbal positions. Nonverbal constituents can be located:

- either before the LK, i.e., in the *Vorfeld* (VF, 'initial field'); this position is normally restricted to one constituent;
- or after the VC position, i.e., in the *Nachfeld* (NF, 'final field'); this position is often not filled;
- or between the two brackets LK and VC, i.e., in the *Mittelfeld* (MF, 'middle field'); it is this field that is in the focus of this paper.

A sample German declarative main clause with its topological structure is given in Table 1.

The middle field is the relevant area for our investigations because most constituents of the clause cluster in this field. For example, the example given in Table 1 shows four basic constituents: the subject *Uller*, the temporal adverbial *heute*, the indirect object *einem Freund* (in German usually in the dative case) and the direct object *ein Buch* (in German usually in the accusative case). Three of these constituents are located in the middle field.

[2]We use the notation *a* > *b* for denoting the order *a before b*.

Table 1: Example for the topological structure of a German declarative main clause

VF	LK	MF					VC	NF
Heute	hat	Uller	einem	Freund	ein	Buch	empfohlen	
today	has	Uller	a	friend	a	book	recommended	

'Today, Uller recommended a book to a friend.'

As already mentioned, the relative order of the constituents in the middle field is subject to different syntactic, semantic and pragmatic factors. In short, syntactic factors, such as grammatical function (subject > objects) or case (dative object > accusative object, in the following DAT > ACC) and the like are at play, but they can be easily overridden by non-syntactic factors (cf. the seminal study by Lenerz 1977). In this paper, we focus on the relative order of nominal objects in the German middle field. The unmarked order is DAT > ACC (Lenerz 1977).[3]

Semantic factors that have proven to be quite prominent are definiteness and animacy. The effect of definiteness is such that definite referents tend to precede indefinite referents (Lenerz 1977). It is questionable whether this definite > indefinite constraint is an effect of definiteness by itself or whether this is an epiphenomenon of other constraints. We will touch on this question later in this section.

Animacy has been identified as an important factor for the ordering of constituents in the German middle field by, e.g., Hoberg (1981). Here, the unmarked order is animated referent > unanimated referent, see (3). In the prehistory of German, this ordering principle might have been quite prominent and in the end might have led to the development of DAT > ACC as the unmarked order (see Speyer 2015) because the dative is correlated with the semantic role of recipient in the classical case of verbs with three arguments that instantiate the agent–patient–recipient scheme, such as *geben* ('give'), *übermitteln* ('convey'), or *anbieten* ('offer'). The recipient is usually animated whereas the patient is normally not.

(3) *Heute hat* [*die Lehrerin*]$_{\text{NOM, anim}}$ [*der Schülerin*]$_{\text{DAT, anim}}$ [*das Buch*]$_{\text{ACC, inanim}}$ *gegeben.*
today has the teacher the student the book given
'The teacher gave the book to the student today.'

[3] Interestingly, the unmarked order of *pronominal* objects is ACC > DAT. In this study, we focus on nominal objects, excluding pronominal objects.

We concentrate here on pragmatic factors, especially those that have traditionally been described in terms of *information structure* (Féry & Krifka 2008). Information-structural notions that have been found to play a role are, for example, the given > new constraint (Lenerz 1977) and the topic > comment constraint (Frey 2004). In our investigation, we focus on the given > new constraint. Basically, this ordering constraint says that knowledge that is assumedly familiar to the hearer is positioned before material that is new to the hearer. These constraints are not to be read as *given information always stands before new information* but rather as constraints that can override the unmarked constituent order DAT > ACC in certain cases, as in (4). In this example, the accusative object represents given information, whereas the dative object refers to a person that has not yet been introduced to the discourse.

(4) [Context: discussion about a certain mystery novel]
Und dann hat sie [*den Krimi*]$_{\text{ACC, given}}$ [*einer Freundin*]$_{\text{DAT, new}}$ *geschenkt.*
and then has she the novel a friend presented
'And then she gave the novel to a friend of hers as a present.'

We see in (4) that the objects bear different articles. A constraint that is correlated with given > new is the constraint that definite noun phrases precede indefinite noun phrases (Lenerz 1977, Rauth 2020). The correlation is as follows: Definite reference normally implies that the entity referred to is known to the speaker and hearer (hence given information). Using a definite determiner is felicitous only if the hearer can uniquely identify the referent, and this is only possible if it is known to the hearer or can be inferred by them (Prince 1981). In contrast, in conveying new information, speakers tend to refer via indefinite noun phrases, indicating that the referent is not yet part of the discourse universe. This comes in handy, as it allows us to use definiteness as a proxy for givenness and indefiniteness as a proxy for newness in our pilot study, when dealing with data that is not annotated for givenness or *information status*.

German is not the only language that allows for variable orders of the direct and indirect objects. In other closely related languages such as Dutch and English, the relative linearization of the direct object (DO) and the indirect object (IO) are subject to variation as well. An example is the phenomenon of dative alternation in English: The indirect object can be realized as a noun phrase preceding the direct object (5a), or as a prepositional phrase following the direct object (5b). The phenomenon of Heavy NP shift provides another example: long (i.e., heavy) direct objects can be put after the prepositional indirect object (5c).

(5) a. *Then she gave* [*her friend*]$_{\text{IO, NP}}$ [*the new mystery novel*]$_{\text{DO, NP}}$.

b. *Then she gave* [*the new mystery novel*]$_{\text{DO, NP}}$ [*to her friend*]$_{\text{IO, PP}}$.
c. *Then she gave* [*to her friend*]$_{\text{IO, PP}}$ [*the new mystery novel about the murderer from Dartmoor*]$_{\text{DO, NP}}$.

The factors governing these variations are partly of a different nature. While the length of the respective objects seems to be a governing factor, given-/newness does not seem to play a primary role here. Engel et al. (2022) found evidence that definiteness is a good predictor also for the English dative alternation (if the indirect object is indefinite, it is more often realized as prepositional phrase, but this effect is strongest in spoken informal texts). So it looks as though something similar to the German definite > indefinite constraint is at play in English as well, and the fact that the effect is strongest in orally produced texts indicates that it is a matter of constraints on language processing.

In our investigations, we focus on sentences with ditransitive verbs whose objects are located in the middle field. In our study, we compare the two objects in their original order with a generated, reversed order (see Section 4). In this direct comparison, we want to investigate whether the role of givenness for word order can be quantified with the help of information-theoretic measures such as surprisal. Hence, as described in Section 3, we exclude all cases where the objects are either both definite or both indefinite (i.e., where givenness does not play a role) and keep the mixed cases only so that the two variants differ with regard to definiteness, our proxy for givenness. Moreover, other factors that could influence the order of constituents should be excluded when comparing the two variants. Hence, we control for object length because variations in length are known to have an impact on the order of constituents in the sentence ("Gesetz der wachsenden Glieder", or *law of increasing constituents*, Behagel 1932).

3 Data

We use the SdeWaC corpus (Faaß & Eckart 2013)[4] as the source of data for our analysis. The corpus consists of 44M sentences with more than 845M tokens from German webpages. It has been automatically tokenized, tagged, lemmatized, and parsed with Bohnet (2010)'s dependency parser.[5] Using the dependency annotation, we select all sentences from the corpus that contain at least one ditransitive

[4] https://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/sdewac, accessed 2022/12/01.

[5] Bohnet (2010)'s dependency parser was trained on the TIGER corpus (Brants et al. 2004: release August 2007) which had been converted to dependency structures by Wolfgang Seeker.

verb with a dative and an accusative object, labeled DA (= DAT) and OA (= ACC), respectively, in the dependency annotation.[6] In addition, the objects must meet the following criteria:

(i) Both objects have a nominal head. This means that the word forms labeled with the dependency relation OA and DA must be tagged with the STTS tag NN for "normal noun" (Schiller et al. 1999). For example, in (6), the accusative object *ein Buch* ('a book') and the dative object *dem Jungen* ('the boy') in (6a) are recognized as having a nominal head. In contrast, the pronominal dative object *ihm* ('him') in (6b) is tagged as PPER for personal pronoun and the sentence would be excluded from the sample.

(6) a. *Ich werde* [*dem Jungen*/NN/DA]$_{\text{DAT}}$ [*ein Buch*/NN/OA]$_{\text{ACC}}$ *geben.*
I will the boy a book give
'I will give a book to the boy.'

b. *Ich werde* [*ihm*/PPER/DA]$_{\text{DAT}}$ [*ein Buch*/NN/OA]$_{\text{ACC}}$ *geben.*
I will him a book give
'I will give him a book.'

(ii) To draw conclusions about the givenness of the objects, the object noun phrases must differ with regard to definiteness, one being definite, the other being indefinite. That is, the head nouns of one of the objects must directly dominate a definite article (def) and the head noun of the other object must directly dominate an indefinite article (indef). Definite articles are word forms that are tagged with the STTS tag ART and are lemmatized as *der* ('the'). Indefinite articles are word forms tagged as ART with the lemma *ein* ('a'). Examples (2a) and (2b) from the introduction would thus be included, while (1a), (1b), and (2c) with two given or two new objects would be excluded. This criterion also entails that sentences with an indefinite plural object, like *Bücher* ('books') in (7), are rejected because they do not have a determiner in German.

(7) *Ich werde* [*dem Jungen*/DA]$_{\text{DAT}}$ [*Bücher*/OA]$_{\text{ACC}}$ *geben.*
I will the boy books give
'I will give books to the boy.'

[6] The label DA is also used for free datives, see Brants et al. (2004). However, free datives occur mainly in pronominal form, which are excluded from the present study.

(iii) To control for effects of length, the objects must contain the same number of words (ignoring punctuation). Example (8a) with two objects of length two would be accepted, but not (8b) with objects of different lengths (two vs. three words).

(8) a. *Ich werde* [*dem Jungen*]$_{\text{DAT}}$ [*ein Buch*]$_{\text{ACC}}$ *geben.*
I will [the boy]$_{\text{DAT}}$ [a book]$_{\text{ACC}}$ give
'I will give a book to the boy.'

b. *Ich werde* [*dem Jungen*]$_{\text{DAT}}$ [*ein gutes Buch*]$_{\text{ACC}}$ *geben.*
I will [the boy]$_{\text{DAT}}$ [a good book]$_{\text{ACC}}$ give
'I will give a good book to the boy.'

(iv) Both objects must be located within the same middle field (`MF`).[7] We only keep sentences in which the same `MF` node dominates both objects, as in (9a). If one object is located in another field, for example, in another `MF` or in the initial field `VF` as in (9b), the sentence is excluded.

(9) a. *Ich werde* ⟦[*das Buch*]$_{\text{ACC}}$ [*einem Jungen*]$_{\text{DAT}}$⟧$_{\texttt{MF}}$ *geben.*
I will the book a boy give
'I will give the book to a boy.'

b. ⟦[*Das Buch*]$_{\text{ACC}}$⟧$_{\texttt{VF}}$ *werde* ⟦*ich* [*einem Jungen*]$_{\text{DAT}}$⟧$_{\texttt{MF}}$ *geben.*
the book will I a boy give
'I will give the book to a boy.'

(v) Finally, we exclude light verb constructions, in which a semantically faded ("light") verb establishes one fused meaning with its object. For instance, the phrase *einer Prüfung unterziehen* ('submit a check') in (10) is an example of such a construction: *(to) submit a check* corresponds to *(to) check*. In these constructions, there is a clear bias for the order in which the fused object is directly adjacent to the light verb. This even holds for cases where the fused object is the dative object, resulting in the fixed (otherwise marked) object order ACC > DAT, as in (10).

[7] For determining the topological structure, we parse the sentences with the Berkeley parser (Petrov et al. 2006) and a constituency model from Ortmann (2021) trained on the TüBa-D/Z treebank, a corpus that has been annotated with syntactic and topological categories (Telljohann et al. 2017). We use the `News1` model from https://github.com/rubcompling/konvens2021, which was trained on 80% of the TüBa-D/Z corpus. The model annotates constituents and topological fields at the same time.

(10) *Wir werden* [*die neuen Daten*]$_{\text{ACC}}$ [*einer genauen Prüfung*]$_{\text{DAT}}$
we will the new data a thorough check
unterziehen.
give
'We will submit the new data to a thorough check.'

We compiled a list of 120 light verb constructions from Eisenberg (2020) and ProGram2.0 (2018).[8] If the lemmas of the verb and of the head nouns of the objects are included in the list, the object pair is removed.

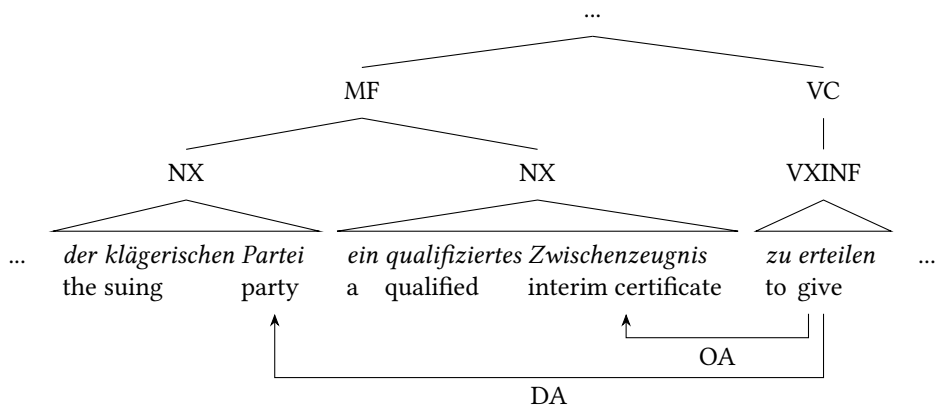


Figure 1: Excerpt from an example sentence (engl. 'to give the suing party a qualified interim certificate') with a ditransitive verb and its two objects, along with a constituency (top) and dependency (bottom) analysis

Figure 1 shows an example object pair with the corresponding dependency and constituency analysis. On top of the text, the constituency tree is displayed, consisting of noun phrases (labeled as `NX`, following the TüBa-D/Z annotation scheme, Telljohann et al. 2017), an infinitive (`VXINF`), and nodes representing topological fields (`MF`, `VC`). Below the text, the relevant dependency relations are shown. As required, the verb dominates a nominal dative (`DA`) and accusative (`OA`) object pair within the same middle field (`MF`) and with the same number of words. The dative object has a definite article (*der* ('the')) and the accusative object an indefinite one (*ein* ('a')).

[8]The list is available at https://gitlab.ruhr-uni-bochum.de/comphist/c6dormdiff.

For our analysis, the selected sentences are split into constituents based on their constituency parse. For each terminal token (ignoring punctuation), we choose as the constituent node the highest dominating phrasal node below the next topological field node. (11) shows an example constituency analysis from the data set.

(11) [*Sie*]$_{\texttt{NX}}$ [*sind*]$_{\texttt{VXFIN}}$ [*zudem*]$_{\texttt{PX}}$ [*ein wichtiges Stilmittel*]$_{\texttt{NX}}$, [*um*]$_{\texttt{C}}$ [*dem Film*]$_{\texttt{NX}}$ [*eine Struktur*]$_{\texttt{NX}}$ [*zu verleihen*]$_{\texttt{VXINF}}$
they are moreover an important stylistic.device to the film a structure to give
'Moreover, they are an important stylistic device to give the film a structure.'

The SdeWaC corpus contains approximately 1.8M ditransitive verbs. Among those, 13,472 object pairs in 13,458 sentences meet the aforementioned criteria. Table 2 gives a summary of the data. It shows that in 95.87% of the cases, the dative object precedes the accusative object and 87.61% of the definite objects precede an indefinite object. Only 5.32% of the objects in the original data are longer than three words, so we decided to only include objects of length two and three in our final data set.[9]

The above constraints concerning case and definiteness result in a total of four possible combinations of object pairs:

(i) DAT.DEF > ACC.INDEF (i.e., the definitive dative object precedes the indefinite accusative object)

(ii) DAT.INDEF > ACC.DEF

(iii) ACC.DEF > DAT.INDEF

(iv) ACC.INDEF > DAT.DEF

Examples (12–15) show one sentence per group from the sample.

[9]This decision was also made because data processing proved to be error-prone for objects with more than three words. This could be solved by filtering as described above.

Table 2: Summary of the selected sentences and object pairs from the SdeWaC corpus, for the original complete data and the final data set with objects of length two and three only

	Original data		Final data set	
	n	%	*n*	%
Sentences	13,458		12,742	
Object pairs	13,472		12,756	
Sentences with >1 pair	14	0.10	14	0.11
Dative before accusative (DAT>ACC)	12,916	95.87	12,253	96.06
Definite before indefinite (def>indef)	11,803	87.61	11,171	87.57
(i) DAT.DEF>ACC.INDEF	11,601	86.11	10,999	86.23
(ii) DAT.INDEF>ACC.DEF	1,315	9.76	1,254	9.83
(iii) ACC.DEF>DAT.INDEF	354	2.63	331	2.59
(iv) ACC.INDEF>DAT.DEF	202	1.50	172	1.35
Min. object length (in words)	2		2	
Max. object length (in words)	13		3	
Avg. words per object	2.35		2.22	
Avg. constituents per sentence	12.23		12.28	

(12) Group (i): DAT.DEF > ACC.INDEF
Beim Zeichnen des eigenen Gesichts kann man [*dem Schüler*]$_{\text{DAT, def}}$
when drawing the own face can one the student
[*einen Spiegel*]$_{\text{ACC, indef}}$ *geben, aber man kann die Unterrichtseinheit auch*
a mirror give but one can the lesson also
mit der Fotografie beginnen.
with the photography start
'When drawing your own face, you can give the student a mirror, but you can also start the lesson with photography.'

(13) Group (ii): DAT.INDEF > ACC.DEF
Ich fühle mich jetzt viel sicherer, schlafe nachts ruhig, weil
I feel myself now much safer sleep at.night peacefully because
ich mir keine Sorgen darüber machen muß, wie ich [*einem*
I me no worries about make must how I a

Geldverleiher]$_{\text{DAT, indef}}$ [*das Geld*]$_{\text{ACC, def}}$ *zurückzahlen soll.*
money.lender the money pay.back shall
'I feel much safer now, sleep peacefully at night because I don't have to worry about paying back a money lender.'

(14) Group (iii): ACC.DEF > DAT.INDEF
Ein paar Tage später zeigte ich [*den Film*]$_{\text{ACC, def}}$ [*einem Freund*]$_{\text{DAT, indef}}$
a few days later showed I the film a friend
und sah ihn noch einmal mit der gleichen Begeisterung.
and watched it once more with the same enthusiasm
'A few days later, I showed the film to a friend and watched it again with the same enthusiasm.'

(15) Group (iv): ACC.INDEF > DAT.DEF
Wegen der geänderten Zuständigkeiten im Grundgesetz
because.of the changed responsibilities in.the constitution
müsse der Bund [*eine Neukonzeption*]$_{\text{ACC, indef}}$ [*den*
would.have.to the federal.government a redesign the
Ländern]$_{\text{DAT, def}}$ *überlassen.*
states leave
'Because of the changed responsibilities in the constitution, the federal government would have to leave a redesign to the states.'

The vast majority follows the unmarked order of definite dative before indefinite accusative (group (i)), cf. Figure 2.[10] The example in Figure 1 is also an instance of the unmarked order DAT.DEF > ACC.INDEF.

4 Methods

We propose information density and, more specifically, the uniform distribution of information in the sentence as an explanation of object order. In the information-theoretic framework, information can be derived from the predictability of a word in context (Shannon 1948), with lower predictability causing higher processing effort (Hale 2001, Levy 2008).

We use language models to estimate the probability $p(w)$ of individual tokens w from bigram lemma frequencies in the SdeWaC corpus. To keep the

[10]The plots have been created with the R package ggplot2, https://github.com/tidyverse/ggplot2.

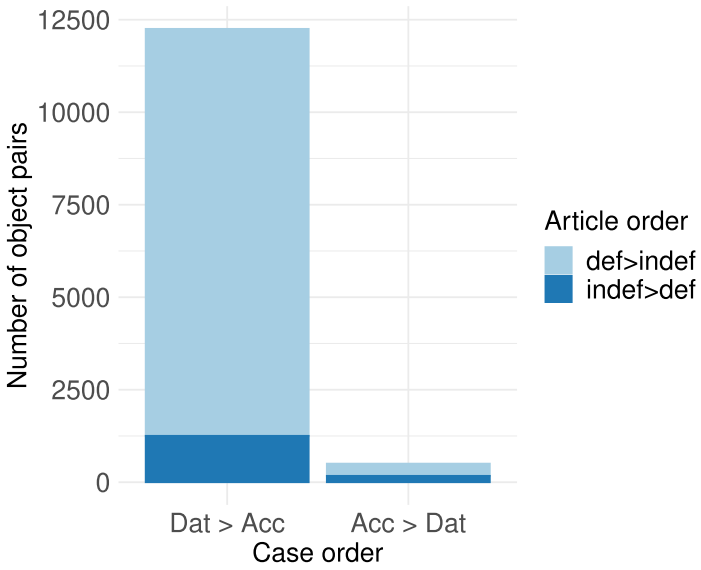


Figure 2: Frequencies of case and article order in the final data set. The majority of object pairs follow the unmarked order of definite dative before indefinite accusative (group (i); upper part of the left bar).

data size manageable, we include only bigrams with ≥50 occurrences and apply Jeffreys-Perks smoothing with $\lambda = 0.5$ (Jeffreys 1946), yielding a total amount of approximately 1M bigrams with 100K distinct lemma types. Punctuation is ignored as we assume that it does not provide any additional information about processing efforts in the German middle field.

As we are interested in the order of constituents, we measure predictability not at the word level but at the level of whole constituents. We calculate the mean surprisal $Surpr_{\text{mean}}$ of a constituent $c = w_1, \ldots, w_n$ by adding up the individual surprisal values of all the words in the constituent and averaging them, see equation (16).

$$\text{Surpr}_{\text{mean}}(c) = \frac{1}{n} \sum_{i=1}^{n} -\log_2(p(w_i)) \tag{16}$$

The information profile of a sentence, which indicates whether information is distributed uniformly and smoothly across the sentence, is composed of the surprisal values of all the constituents in the sentence, which are simply concatenated. Figure 3 shows an example: The fragment marked as *original* consists of the constituents [*um*] ('in order'), [*dem Film*] ('the film'), [*eine Struktur*] ('a struc-

ture'), [*zu verleihen*] ('to give') (see example (11) for the complete sentence).[11] The corresponding information profile is displayed in the second row ($\text{Surpr}_{\text{mean}}$(c)): For instance, the lemma-based mean bigram surprisal of the dative object (*dem Film*) is 5.921 bits, and the surprisal of the accusative object (*eine Struktur*) is 9.879 bits. The resulting profile of this fragment is the sequence [16.701, 5.921, 9.879, 8.348].

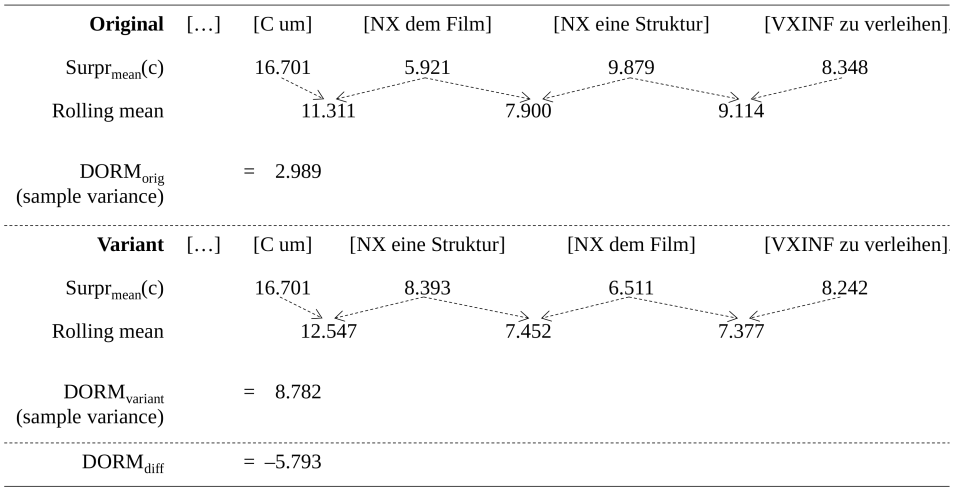


Figure 3: Example calculation of rolling means and DORM values for a part of sentence (11)

We then compare this information profile with the profile of a competing variant, i.e., a generated alternative sentence that looks like the original sentence, except that the two objects are swapped. In Figure 3, the variant sentence with the two swapped objects is displayed below the original sentence. The upper part of Figure 3 shows the original constituent order, the variant is displayed in the lower part. Note how the surprisal values change because of the swapped objects. As the original order has a lower DORM value (i.e., a smoother profile) than the generated variant, $\text{DORM}_{\text{diff}}$ is negative for this fragment.

We call this approach the *corpus of variants* method because it allows us to inspect the differences between the observed word order and a plausible alternative order, while keeping other factors constant. The variant generation causes a change of bigram surprisals at the edges of the swapped objects, so we recalculate the surprisal values on the basis of the language model that was also

[11] One could argue that the phrase *eine Struktur verleihen* is a light verb construction because it can be replaced by *strukturieren* ('(to) structure'). However, it is not part of our list of light verb constructions (see Section 3) and is therefore not excluded from the data.

used for the original sentence and the information profile for the generated variant sentence, see Figure 3: The dative object now has a mean surprisal value of 6.511 bits and the accusative object a surprisal of 8.393 bits.

For comparing the information profiles of the original sentence and the generated sentence, we use measures called DORM and $\text{DORM}_{\text{diff}}$, as explained in the next sections.

4.1 DORM

DORM (Deviation of the Rolling Mean), which has been proposed by Cuskley et al. (2021), is a measure that allows us to quantify the uniformity of a sentence's information profile. Cuskley et al. (2021: 9) describe DORM as an "easily interpretable summary of how uniform or clumpy a particular utterance is". DORM is calculated as follows: Given the sequence of surprisal scores of all constituents in a sentence, we first compute the rolling means RM_i of each adjacent pair of surprisal scores s_i, s_{i+1} as in equation (17).

$$\text{for } i \text{ in } (1 \ldots n-1) : RM_i = \frac{s_i + s_{i+1}}{2} \tag{17}$$

For instance, the first mean RM_1 in Figure 3 (original sentence) is the mean of 16.701 (= [*um*]'s surprisal) and 5.921 (= [*dem Film*]'s surprisal):

$$\frac{(16.701 + 5.921)}{2} = 11.311 \tag{18}$$

We next compute DORM, which corresponds to the sample variance of the rolling means and serves us as a measure of the overall smoothness, as shown in equation (19).

$$\text{DORM} = s^2 = \frac{\sum_{i=1}^{n} (RM_i - \bar{x})^2}{n-1} \tag{19}$$

A lower DORM value indicates less variance, i.e., a smoother information signal, while a higher DORM value points at a less uniform information profile. This is usually achieved by placing linguistic units, in our case constituents, with similar surprisal values next to each other since extreme differences would no longer result in a low DORM value (Cuskley et al. 2021). Extreme surprisal values should, thus, be spread evenly across a sentence.

In Figure 3, the original sentence has a DORM value of 2.989, and the variant sentence has a DORM value of 8.782. This means that the original object order results in a smoother profile.

As we show in the next section, we use the DORM values for pairwise comparing information profiles of original sentences and their variants and introduce a new measure, $\text{DORM}_{\text{diff}}$, for measuring the difference between the original and the variant sentence.

4.2 $\text{DORM}_{\text{diff}}$

DORM values are directly comparable only for sequences that contain the same (number of) elements. Hence, the absolute DORM values can only be compared between the original constituent order ($\text{DORM}_{\text{orig}}$) and the swapped variant ($\text{DORM}_{\text{variant}}$) of the same sentence.

In order to compare values from different sentences, we use the difference between DORM value pairs, as defined in equation (20). That is, we collect the individual differences between all original and variant pairs of the sample and use these scores in our investigations.

$$\text{DORM}_{\text{diff}} = \text{DORM}_{\text{orig}} - \text{DORM}_{\text{variant}} \tag{20}$$

$\text{DORM}_{\text{diff}}$ allows us to investigate the difference between the observed information profile and the profile of the variant constituent order. If there was no connection between object order and information profile, $\text{DORM}_{\text{diff}}$ should be zero. In contrast, if speakers aimed at a smooth information profile in accordance with the UID hypothesis (Levy & Jaeger 2007), DORM should be lower for original sentences than for the variants. If the information profile of the variant sentences was more uniform, there would have to be other explanations for the observed object order.

Our hypothesis is therefore that, in general, $\text{DORM}_{\text{diff}}$ should be negative (as in the example in Figure 3) – because this would mean that the original sentence has a smoother profile than its variant and, hence, that constituent order can be traced back to information-theoretic principles.

4.3 $\text{DORM}_{\text{case}}$ and DORM_{giv}: Case and givenness order

We use logistic regressions to investigate the effects of information profile, case, and givenness on object order. If any of these factors significantly influenced the order of dative and accusative or given and new object, they should help to predict which order will occur in the sentence.

However, we cannot simply use $\text{DORM}_{\text{diff}}$ as defined in equation (20) to predict case and givenness order because the order is encoded in the score. If the original sentence order is DAT > ACC, $\text{DORM}_{\text{diff}}$ is calculated as $\text{DORM}_{\text{DAT > ACC}}$ –

$DORM_{ACC > DAT}$. And if the original sentence order is ACC > DAT, $DORM_{diff}$ is calculated as $DORM_{ACC > DAT} - DORM_{DAT > ACC}$. The same applies analogously to def > indef.

Hence, $DORM_{diff}$ as a predicting factor must not be calculated with reference to *orig* and *variant*. Instead, it must abstract away from the actually occurring order and always use the same order of minuend and subtrahend, as shown in the equations (21) and (22).[12]

(21) $$DORM_{case} = DORM_{DAT > ACC} - DORM_{ACC > DAT}$$

(22) $$DORM_{giv} = DORM_{DEF > INDEF} - DORM_{INDEF > DEF}$$

Based on equations (21) and (22), we can predict if and how the order of the two objects will be influenced by a change in the uniformity of the information profile resulting from a change in case order ($DORM_{case}$) or givenness order ($DORM_{giv}$).

$DORM_{case}$ is smaller than zero if the order of DAT > ACC has a more uniform information profile than ACC > DAT, and greater than zero otherwise. Similarly, a negative $DORM_{giv}$ indicates a more uniform information profile for DEF > INDEF, while a positive value shows a more uniform distribution for INDEF > DEF.

As DAT > ACC and DEF > INDEF are considered the unmarked constituent order (cf. Section 2), they can be expected to be easier to process for language users since they are more familiar with this conventionalized order. However, if the information profile for ACC > DAT or INDEF > DEF was smoother than for the default order, this could potentially lead to an inverse, marked order of objects to reduce processing difficulty. If this is true, a higher $DORM_{case}$ (i.e., a less optimal information profile for DAT > ACC) should increase the likelihood of ACC > DAT. And, along the same lines, a higher $DORM_{giv}$ (i.e., a smoother information profile for INDEF > DEF) should increase the probability of INDEF > DEF.

5 Results

5.1 $DORM_{diff}$: Object order and the information profile

To explore the relevance of information-theoretic principles for object order in the German middle field, we inspect the information profiles of the original sen-

[12] Note that $DORM_{case} = -(DORM_{ACC > DAT} - DORM_{DAT > ACC})$, and, similarly, $DORM_{giv} = -(DORM_{INDEF > DEF} - DORM_{DEF > INDEF})$. Hence, as long as it is used consistently, the order of minuend and subtrahend is irrelevant, and we arbitrarily decided for the orders DAT > ACC and DEF > INDEF as the minuends.

tences and their generated variants. For the data described in Section 3 and their corresponding variants, $\text{DORM}_{\text{diff}}$ lies between −33.16 and 23.90, with slightly more than half of the values (52.7%) being smaller than zero. On average, the DORM value of the original constituent order is significantly lower than the DORM value of the generated variants: $\text{DORM}_{\text{diff}} = -0.17$ ($t = -6.88$, $p < 0.001$).[13] The effect size (Cohen's $d = 0.06$) is smaller than 0.2, which is traditionally assumed to indicate a small effect (Winter 2020), but the result suggests that natural language indeed follows information-theoretic principles, as writers tend to produce sentences with information profiles smoother than the ones that would result from an also plausible, but inverse object order.

Table 3: Mean $\text{DORM}_{\text{diff}}$ values for different object orders; (i)–(iv) refer to the four groups of possible combinations (*** $p < 0.001$; for the complete statistics, see Table 4).

	DEF>INDEF	INDEF>DEF	all
DAT>ACC	(i) −0.12 ***	(ii) −0.68 ***	−0.18 ***
ACC>DAT	(iii) 0.10	(iv) −0.23	−0.01
all	−0.12 ***	−0.63 ***	−0.17 ***

As Table 3 shows, this observation holds independently of the observed order in the original sentence of dative and accusative or definite and indefinite object.[14] Looking first at the right-most column (*all*), we see that for the unmarked order DAT > ACC (first row), which appears in the majority of sentences of the original data set (Section 3), the mean $\text{DORM}_{\text{diff}}$ is −0.18. For ACC > DAT, the mean $\text{DORM}_{\text{diff}}$ is also negative (−0.01) even though it is not significantly different from zero. Looking at the bottom row (*all*), we see that for the marked order INDEF > DEF, $\text{DORM}_{\text{orig}}$ is on average −0.63 lower than $\text{DORM}_{\text{variant}}$. For the unmarked order DEF > INDEF, the difference (−0.12) is negative, too, and also significantly different from zero.

Regarding the four possible combinations of case and givenness order (i.e., groups (i)–(v) in the inner part of Table 3), we see that three out of four groups show negative $\text{DORM}_{\text{diff}}$ values on average, the order ACC_{def} > $\text{DAT}_{\text{indef}}$ being an exception with a $\text{DORM}_{\text{diff}}$ of 0.10. Only the two groups with default case order DAT > ACC result in highly significant differences, both for the unmarked

[13] Statistical calculations have been performed with R (R Core Team 2018). We used two-tailed Welsh *t*-tests for these calculations.

[14] The complete statistics are presented in Table 4.

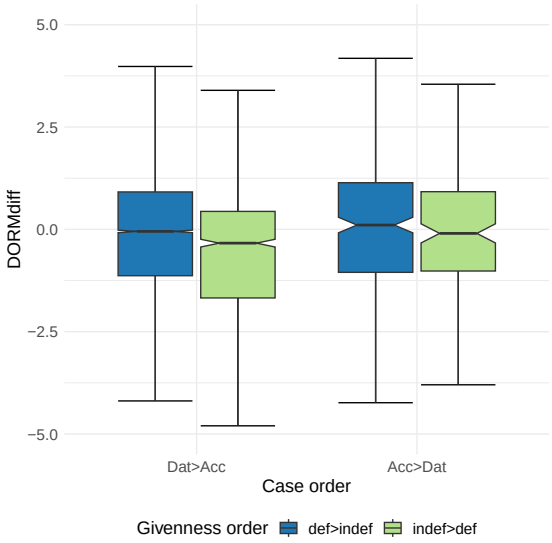


Figure 4: $DORM_{diff}$ by object order and givenness order (not displaying outliers). The boxes show the interquartile range from first to third quartile, with a black line for the median $DORM_{diff}$. The notches indicate the confidence intervals for the median. The four boxes correspond, from left to right, to the groups (i)–(iv), respectively.

givenness order, DEF > INDEF, with a mean of −0.12 as well as for the marked givenness order with a mean of −0.68. In the two cases where the original order is ACC > DAT, no significant differences are found between the $DORM_{diff}$ values. This can possibly be attributed to the small amount of data that is available in these groups (cf. Table 2).

Figure 4 shows additional details about the distribution of the four combinations of case and givenness. If t is negative, $DORM_{orig}$ is lower on average than $DORM_{variant}$, which indicates a more uniform information profile for the original sentence. Traditionally, values of $0.2 \leq d < 0.5$ are interpreted as a small effect. In three out of four conditions, the majority of values lie below zero. However, this difference is significant only in the left group (DAT > ACC) and, in particular, for the marked order INDEF > DEF (green box).

We can interpret the observed trends as follows: In many cases, the information profil of the original sentence and its variant are rather similar, which is shown by many values close to zero and the small effect sizes (see Table 4). However, if sentences show the default case order (DAT > ACC), this is associated with a more uniform information profile, which may explain the large preponderance of this order in modern German (cf. Section 3). At the same time, original sentences generally show a more uniform distribution of information than possible

Table 4: Results of two-sided one sample t-tests for $DORM_{diff}$

	$DORM_{diff}$	t	df	Cohen's d	p
all	−0.17	−6.88	12755	0.06	<0.001 ***
DAT > ACC	−0.18	−6.98	12252	0.06	<0.001 ***
ACC > DAT	−0.01	−0.11	502	0.00	0.91
DEF > INDEF	−0.12	−4.37	11329	0.04	<0.001 ***
INDEF > DEF	−0.63	−8.26	1425	0.22	<0.001 ***
(i) DAT.DEF > ACC.INDEF	−0.12	−4.55	10998	0.04	<0.001 ***
(ii) DAT.INDEF > ACC.DEF	−0.68	−8.20	1253	0.23	<0.001 ***
(iii) ACC.DEF > DAT.INDEF	0.10	0.63	330	0.03	0.53
(iv) ACC.INDEF > DAT.DEF	−0.23	−1.36	171	0.10	0.17

variant sentences — even if the realized order violates the unmarked order of case or givenness though the effect is only significant in the DAT > ACC order. So the preference of language users for smooth information profiles, as predicted by the UID hypothesis, may license deviations from the default case or givenness order.

5.2 $DORM_{case}$ and $DORM_{giv}$: Case and givenness order

To inspect possible effects of the information profile on the order of dative and accusative object and definite and indefinite object, we use logistic regression analyses in R (R Core Team 2023). We start with case order and run a logistic regression with $DORM_{case}$, givenness status, and the number of constituents in the sentence as well as all two-way-interactions as predictors.[15]

Case order is sum-coded: DAT > ACC received the coding 1 and ACC > DAT was sum-coded as −1. Thus, positive estimates in the main effects indicate the order DAT > ACC. As givenness status, we use the definiteness of the dative, which was also sum-coded to increase the precision of the model (Gries 2021).[16] A definite dative was coded as −1, an indefinite dative as +1. While we control for

[15] `glm(formula = Dat > Acc ~(`$DORM_{case}$ `+` $Dat_{definiteness}$ `+ n_Constituents)`^{2}`, family = binomial(), data = constituents_sample)`; for the complete final regression model, see Table 5. Furthermore, we include the two-way interactions of the three factors. Since $DORM_{case}$ and $DORM_{giv}$ are strongly correlated (r = 0.73), we choose to only include one of them as a predictor in each regression analysis.

[16] The objects always exhibit opposing definiteness (cf. Section 3). If the dative object is definite, the accusative object is indefinite, and vice versa. We arbitrarily selected the definiteness of the dative object as predictor. With the accusative as predictor, results would simply be reversed.

object length in that both objects consists of the same number of words, the number of constituents varies between sentences. It seems plausible that a long sentence with a high number of constituents is harder to process than a sentence with fewer constituents. When the amount of information in a sentence already threatens to strain the working memory, the default order DAT > ACC might be preferred to ease overall sentence processing. There might also be an interaction between the information profile of the sentence and its length. However, the order ACC > DAT only occurs once in sentences that have more than 40 constituents (cf. Figure 5). We, consequently, run the logistic regression on a sample of the whole data excluding sentences with more than 30 words.

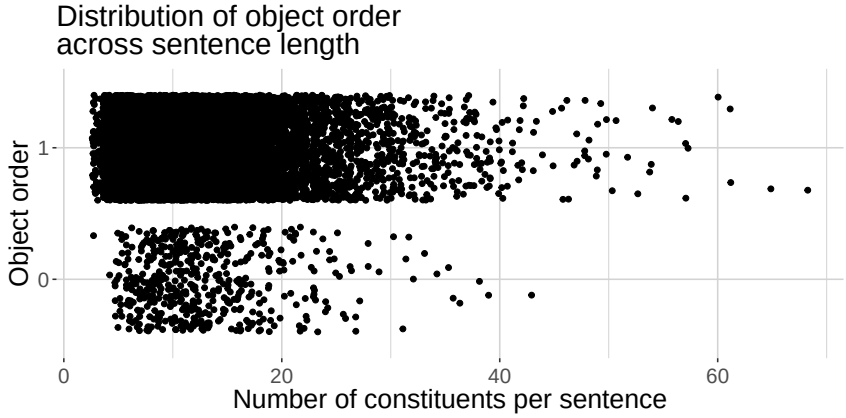


Figure 5: Distribution of object order in the sentences of various length, shown by the number of constituents

Then, we perform *backward model selection* (Gries 2021), excluding one interaction or one main effect at a time, depending on the *p*-value of the predictor. We start with the interactions and first exclude those with the highest non-significant *p*-value. To find out whether the exclusion led to an improvement of the model, a *likelihood ratio* test with the `anova` function in R (R Core Team 2018) is performed. It allows model comparison by capturing how well the model explains the data (Winter 2020). This process is repeated until only significant effects or main effects involved in a significant interaction remain in the model. As soon as the *likelihood ratio* test shows a significant difference between the models, the process of backward model selection is completed. The final model then corresponds to the model before the exclusion of the last predictor and is used to interpret the results.

5.2.1 Case order

Table 5: Logistic regression with $\text{DORM}_{\text{case}}$, definiteness of the dative object and the number of constituents in the sentence to predict case order dative > accusative

Variable	Estimate	SE	z	p	
Intercept	2.90	0.12	23.34	<0.001	***
$\text{DORM}_{\text{case}}$	−0.06	0.02	−3.58	<0.001	***
DAT_{def}	−1.83	0.12	−14.695	<0.001	***
Constituents	−0.014	0.009	−1.60	0.11	
DAT_{def}:Constituents	0.03	0.01	3.64	0.001	***

Table 5 shows the results of the regression analysis for case order. According to the model, $\text{DORM}_{\text{case}}$ ($z = -3.58$, $p < 0.001$) has a highly significant influence on case order.[17] A higher $\text{DORM}_{\text{case}}$ reduces the likelihood of observing DAT > ACC. An increase of $\text{DORM}_{\text{case}}$ means that the information profile of ACC > DAT is smoother than that of DAT > ACC. Hence, a more uniform, smoother distribution for the order ACC > DAT increases the likelihood of observing this marked order in the sentence. And vice versa, a more uniform distribution of DAT > ACC increases the likelihood of this default order.

The second predictor, the definiteness of the dative, also significantly influences case order ($z = -14.695$, $p < 0.001$). In accordance with information structure, an indefinite dative reduces the likelihood of observing the order DAT > ACC. If the dative object is indefinite, it is more likely to follow the accusative (when controlling for other factors, including information density). This result can also hint at an explanation for the positive $\text{DORM}_{\text{diff}}$ value in Table 4 as the influence of the givenness seems to be stronger than the influence of the $\text{DORM}_{\text{case}}$.

The raw number of constituents in the sentence does not significantly influence the order of objects. In the interaction with an indefinite dative ($z = 3.64$, $p < 0.001$), we can see that a definite dative is still a significant predictor for the DAT > ACC constituent order. However, in long sentences, the likelihood of an indefinite dative preceding a definite accusative increases slightly.

We conclude from these results that the information profile, indeed, influences object order as we hypothesized. Language users are more likely to produce the

[17] The model comparison with `anova` showed a p-value of 0.11. However, we cannot reduce the model any further because the number of constituents interacts with the definiteness of the dative.

order of objects that results in the more uniform distribution of information. This holds independently of general preferences for the unmarked order DAT > ACC: If placing the accusative before the dative object smoothes the information profile, language users are more likely to produce the marked order ACC > DAT.

What we also see from the regression analysis is that an indefinite dative tends to trigger the order ACC > DAT, i.e., it favors maintaining the default order DEF > INDEF (our proxy for given before new). This finding provides evidence for the influence of information status on object order, as described in Section 2. The effect is larger than for information density, though, which may explain violations of givenness order if that is associated with a more uniform information profile. Also, the interaction of definiteness and the number of constituents shows that, potentially, the importance of givenness decreases with increasing sentence length.

5.2.2 Givenness order

In addition to the investigation of case order, we run a second logistic regression analysis to inspect the effects of information distribution on givenness order. Similar to above, we include DORM_{giv}, givenness status, and the number of constituents as predictors[18] and performed a backward model selection, as described above. As shown in Table 6,[19] DORM_{giv} is a significant predictor for givenness order ($z = -3.58$, $p < 0.001$). An increase in DORM_{giv} reduces the likelihood of the DEF > INDEF order. A high DORM_{giv} indicates that the information profile of the DEF > INDEF order is less smooth than the information profile of the INDEF > DEF order. Hence, similar to above, a more uniform, smoother information profile for INDEF > DEF increases the likelihood of observing this marked order. And vice versa, a more uniform distribution of DEF > INDEF increases the likelihood of this default order.

These results may provide insights into the relationship between information theory and information structure. In general, we expect both concepts to make similar predictions regarding the order of objects. Placing a given object before a new object, as preferred by information structure, could help to ease processing of the new object by lowering its surprisal and smoothing the information

[18] `glm(formula = def > indef ~(`DORM_{giv}` + `$\text{Dat}_{\text{definiteness}}$`+n_Constituents)`^{2}`, family = binomial(), data = constituents_sample)`; for the complete final regression model, see Table 6. The DEF > INDEF order is coded as 1, the INDEF > DEF order as −1. A definite dative was coded as −1, an indefinite dative as +1. Since $\text{DORM}_{\text{case}}$ and DORM_{giv} are strongly correlated (r=0.73), we choose to only include one of them as a predictor in each regression analysis.

[19] The model comparison with `anova` had a p-value of 0.11.

Table 6: Logistic regression with DORM_{giv} and number of constituents in the sentence to predict givenness order definite > indefinite

Variable	Estimate	SE	z	p
Intercept	1.81	0.12	14.68	<0.001 ***
DORM_{giv}	−0.06	0.02	−3.58	<0.001 ***
Dat_{def}	−2.72	0.05	−54.45	<0.001 ***
Constituents	−0.03	0.01	−3.40	<0.001 ***

signal. Indeed, we find that language users prefer placing definite, i.e., given objects before indefinite, i.e., new objects if that is associated with a more uniform distribution of information. If, however, the information profile of INDEF > DEF is smoother, this can license a deviation from the default information structure.

Surprisingly, the number of constituents in a sentence ($z = -3.40$, $p < 0.001$) also influences the givenness order: An increase in sentence length predicts the marked order INDEF > DEF. Above, we argued that longer sentences should favor an unmarked object order to counterbalance the effort required for processing the high amount of information in the sentence. Instead, in long sentences, the less frequent givenness order seems to be preferred. In the first regression analysis, we already found this effect for the interaction of an indefinite dative and the number of constituents (Table 5). Here, the effect is predicted independently of case order, which was excluded during *backward model selection*.

Perhaps there are other influences on givenness order in longer sentences. As explained in Section 2, definiteness is only a proxy for givenness that we selected because it does not require complex additional annotations. However, our operationalization is independent of the context in which a constituent occurs, whereas givenness, as defined by Prince (1981), Gundel et al. (1993), and Riester & Baumann (2017), can only be determined from the actual context. The longer the sentence, the more context is given in the sentence itself, probably leading to discrepancies between definiteness and givenness. In particular, longer sentences may include more referents and, therefore, require finer increments of givenness than a binary distinction of *given/definite* vs. *new/indefinite*. In future work, we will explore such effects with a more advanced annotation of givenness.

6 Discussion

The results from the previous section can be interpreted as a confirmation of our assumption that information-theoretic features influence the order of objects in the German middle field (cf. Section 5.1): Small but significant effects of $\text{DORM}_{\text{diff}}$ show up within the groups (i) and (ii) with default case order DAT > ACC. No significant effects occur within the groups (iii) and (iv), possibly due to the small size of these groups. Independent of the group size, we could show in Section 5.2 that DORM, i.e., the smoothness of the information profile, can indeed predict the object order in the middle field. Speakers choose the order that results in the most uniform information profile. This holds both for the case order and for the givenness order (cf. Tables 5 and 6).

As we saw in Figure 2, there is a clear preponderance of the unmarked order DAT > ACC. Even though recipients are not consciously aware of the default order, it seems reasonable that they will unconsciously expect the most frequent order of dative preceding accusative. So, if the sentence exhibits the default case order, less cognitive capacity would be consumed for processing the grammar (i.e., case order), according to Futrell et al. (2021). Instead, this capacity would then be free, for example, to process deviations from default givenness order. Similarly, facing the default givenness order (given before new, as reflected by the determiner) would facilitate processing of the unusual case order ACC > DAT. This view is supported by the fact that there is a tendency towards the order DEF > INDEF in sentences with ACC > DAT (cf. Figure 2).

In future work, we want to extend and refine the approach from this pilot study. In particular, we plan to develop improved language models. So far, we used lemma-based bigram models to estimate the probability of observing specific words (and constituents) in different possible orders. Such models reflect lexical or content-based surprisal and can reveal whether a change in object order results in processing advantages on the lexical level. Compared to language models based on word forms, the use of lemmas has the advantage of reducing data sparsity by mapping different word forms to the same lemma. However, this also comes at the price of lemmas being less informative than word forms. In the context of our investigation, this especially concerns case information, which is overtly realized by German determiners but has not been included in our language models. A model based on word forms instead of lemmas could capture the fact that during reading (or listening), the case of objects can already be recognized on the basis of the determiner, helping to reduce entropy early on. Especially in sentences that violate the default order, this could be particularly relevant for processing.

In this pilot study, we have also excluded indefinite plural noun phrases, as they do not have an explicit article in German – and, as a consequence, are shorter than equivalent definite noun phrases, which makes it difficult to compare $\text{DORM}_{\text{diff}}$ values across different types of noun phrases. Integrating indefinite plurals into the analysis may give additional insights into the relevance of information-theoretic concepts for object order. For example, we have seen in exemplary observations that the proportion of the default order given>new seems to be even higher for object pairs with an indefinite plural object. Following our aforementioned considerations, this might be due to the missing determiner: because case is marked at the determiner, the recipient cannot easily infer the case of an object realized as an indefinite plural noun phrase without a determiner. In these cases, the meaning of the word and its grammatical case must be processed simultaneously, which might increase the strain on the working memory. Maintaining the default order could be especially beneficial for processing such cases.

Besides the mentioned enhancements, we plan to experiment with language models beyond *n*-grams. Depending on the sentence, the main verb can be located in the left or right sentence bracket, i.e., before or after the objects in the middle field. We assume that it makes a difference whether the main verb was already uttered or not, and that this should affect expectations and, thus, object surprisal. Overall, the majority of verbs in Geman are simple transitive verbs, requiring an accusative object only. In contrast, ditransitive verbs or verbs requiring a dative object are less frequent. If the main verb is located in the left sentence bracket, it is evident at an early stage whether a dative object is to be expected in the sentence. Hence, a dative object located in the middle field should be processed rather easily. In contrast, auxiliaries in the left bracket do not set up any expectations for a dative object. In this case, it might help the recipient to narrow down possible expectations of the verb in the right sentence bracket if the dative object (which is less frequent than an accusative object) occurs first. Due to the limited context, simple bigram models cannot capture such effects, and we plan to experiment with skip-gram models or models based on content words only. Implementing dependency-based models that take into account the relations between object head nouns and full verbs could also shed light on the direct influence of verb valency.

A topic related to the issue of language models is the calculation of surprisal and DORM values. We proposed to investigate the effects of information density on object order by comparing information profiles of original sentences and variant sentences in which we swapped the two objects. We call this the *corpus of variants* method because it allows us to directly inspect the differences between

plausible alternative word orders, while keeping other factors constant. However, swapping two objects creates only punctual changes in the information profile of the entire sentence, leading to rather small $\text{DORM}_{\text{diff}}$ values. Calculating DORM values only for the local context of the modified parts of the sentence (e.g., as in Figure 3) may return different results and, perhaps, reflect more closely the unfolding of the information flow and resulting effects on local decisions between different structures.

We calculate the $\text{DORM}_{\text{diff}}$ values by subtracting the variant DORM values from the original DORM values. We argued in Section 4.2 that a negative $\text{DORM}_{\text{diff}}$ value indicates a smoother information profile for the original variant. Since the $\text{DORM}_{\text{diff}}$ values are influenced by the length of the sentence, as stated above, the most relevant part of the resulting figures is the algebraic sign, i.e., whether the $\text{DORM}_{\text{diff}}$ value is negative or positive. Thus, it should be possible to interpret and use the $\text{DORM}_{\text{diff}}$ values as a categorical variable instead of a numerical variable (though we would sacrifice the visibility of gradual changes in doing so).

One area where this study could be further enhanced is by exploring alternative measures for givenness, instead of relying on definiteness as a proxy. We chose this operationalization because it does not require additional complex annotations. However, the binary distinction of *given/definite* vs. *new/indefinite* may not be accurate enough, especially in longer sentences or longer contexts in general. We plan to work on creating more nuanced annotations of givenness and inspect how this influences the order of objects in the middle field. Furthermore, we intend to also include objects with the same givenness status in the investigation to confirm that the information profile has an influence on the object order without being also influenced by the givenness.

Finally, it is yet an open question how the current order preferences have been established. In future work, we want to extend the experiments to historical German. In historical language stages of German, the word order was generally more flexible than in modern German. Crucially, this also holds for dative and accusative objects, which showed much more variation with respect to their relative order than nowadays. However, similar factors as in modern German already played a role, in particular givenness (Rauth 2020). Hence, in the long term, we are interested in investigating how information density relates to object order variation in historical German. Furthermore, a diachronic analysis could provide insight into the historical development of object order and reveal which role information density might have played diachronically, ultimately resulting in the clearly-preferred order of objects (dative before accusative) as we observe them for modern German.

Using the proposed methods, we will investigate how the object order in historical data can be explained. In a second step, we will trace the development to modern German and inspect relevant factors that contributed to the formation of modern standard object order. A prerequisite is that we can control for other factors besides length, in particular animacy, which plays an important role in language and cognitive processing.

7 Conclusion

In this paper, we have motivated the order of dative and accusative objects in the German middle field with information-theoretic concepts, while controlling for the factor length.

Overall, the corpus data shows an exceedingly strong bias for the unmarked orders (DAT > ACC in 96% and DEF > INDEF in 88% of the cases). As we hypothesized, the corpus sentences are in general characterized by a more uniform information profile than the generated swapped variants. This is true for corpus sentences with the default order DAT > ACC. This observation is confirmed by logistic regression models in which lower $\text{DORM}_{\text{case}}$ and DORM_{giv} values increase the likelihood of the marked orders (accusative before dative, new before given). We thus argue that deviations from the default orders can be explained by more uniform information profiles, which improve overall sentence processing.

In future work, we will extend the proposed approach to historical data. We plan to investigate how the modern order preferences have been established and which role information-structural and information-theoretical factors may have played in this process.

Acknowledgements

The work reported in this paper has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project ID 232722074 – SFB 1102 “Information Density and Linguistic Encoding”.

Our thanks goes to the reviewers and editors of this volume for many helpful comments, furthermore to Magdalena Meiser for helping with the light verb constructions. All remaining errors are of course ours.

References

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. DOI: 10.1177/00238309040470010201.

Behagel, Otto. 1932. *Deutsche Syntax: Eine geschichtliche Darstellung*, vol. 4: Wortstellung, Periodenbau (Germanische Bibliothek). Heidelberg: Winter.

Bohnet, Bernd. 2010. Top accuracy and fast dependency parsing is not a contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, 89–97. Beijing, China.

Brants, Sabine, Stefanie Dipper, Peter Eisenberg, Silvia Hansen-Schirra, Esther König, Wolfgang Lezius, Christian Rohrer, George Smith & Hans Uszkoreit. 2004. TIGER: Linguistic interpretation of a German corpus. *Research on Language and Computation* 2(4). 597–620. DOI: 10.1007/s11168-004-7431-3.

Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots. Linguistics in search of its evidential base*, 75–96. Berlin, New York: De Gruyter Mouton. DOI: 10.1515/9783110198621.75.

Cuskley, Christine, Rachael Bailes & Joel Wallenberg. 2021. Noise resistance in communication: Quantifying uniformity and optimality. *Cognition* 214. 104754. DOI: 10.1016/j.cognition.2021.104754.

Eisenberg, Peter. 2020. *Grundriss der deutschen Grammatik*. 5., aktualisierte und überarbeitete Auflage. Berlin: J. B. Metzler.

Engel, Alexandra, Jason Grafmiller, Laura Rosseel & Benedikt Szmrecsanyi. 2022. Assessing the complexity of lectal competence: The register-specificity of the dative alternation after give. *Cognitive Linguistics 2022* 69. 727–766. DOI: 10.1515/cog-2021-0107.

Faaß, Gertrud & Kerstin Eckart. 2013. SdeWaC: A corpus of parsable sentences from the web. In Iryna Gurevych, Chris Biemann & Torsten Zesch (eds.), *Language processing and knowledge in the web*, 61–68. Berlin, Heidelberg: Springer.

Fenk-Oczlon, Gertraud. 1983. Ist die SVO-Wortfolge die natürlichste? *Papiere zur Linguistik* 29. 23–32.

Féry, Caroline & Manfred Krifka. 2008. Information structure. Notional distinctions, ways of expression. In Piet van Sterkenburg (ed.), *Unity and diversity of languages*, 123–136. Amsterdam: Benjamins.

Frey, Werner. 2004. A medial position for topics in German. *Linguistische Berichte* 198. 153–190.

Futrell, Richard, Edward Gibson & Roger Levy. 2021. Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science* 44(3). e12814. DOI: 10.1111/cogs.12814.

Gries, Stefan Th. 2021. *Statistics for linguistics with R. A practical introduction.* Berlin, Boston: De Gruyter Mouton. DOI: 10.1515/9783110718256.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69. 274–307.

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 (NAACL'01)*. 159–166. DOI: 10.3115/1073336.1073357.

Hoberg, Ursula. 1981. *Die Wortstellung in der geschriebenen deutschen Gegenwartssprache.* München: Hueber.

Jeffreys, Harold. 1946. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London Series A: Mathematical and Physical Sciences* 186. 453–461.

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

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177. DOI: 10.1016/j.cognition.2007.05.006.

Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In Bernhard Schölkopf, John Platt & Thomas Hofmann (eds.), *Advances in neural information processing systems 19: Proceedings of the 2006 conference*, 849–856. Cambridge: The MIT Press. DOI: 10.7551/mitpress/7503.003.0111.

Ortmann, Katrin. 2021. Automatic phrase recognition in historical German. In *Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021)*, 127–136. Düsseldorf. https://aclanthology.org/2021.konvens-1.11.

Petrov, Slav, Leon Barrett, Romain Thibaux & Dan Klein. 2006. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, 433–440. Sydney, Australia. DOI: 10.3115/1220175.1220230.

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In P. Cole (ed.), *Radical pragmatics*, 223–254. New York: Academic Press.

ProGram2.0. 2018. *Liste der Funktionsgefüge.* https://program.idf.uni-heidelberg.de/fvg/liste.

R Core Team. 2018. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

R Core Team. 2023. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Rauth, Philipp. 2020. *Ditransitive Konstruktionen im Deutschen: Geschichte und Steuerung der Objektabfolge im Mittelfeld*. Tübingen: Stauffenburg.

Riester, Arndt & Stefan Baumann. 2017. The RefLex scheme – annotation guidelines. *SinSpeC* 14.

Schiller, Anne, Simone Teufel, Christine Stöckert & Christine Thielen. 1999. *Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset)*. http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf.

Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27(3). 379–423.

Speyer, Augustin. 2011. Die Freiheit der Mittelfeldabfolge im Deutschen – ein modernes Phänomen. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 133. 14–31.

Speyer, Augustin. 2013. Mündlichkeitsnähe als Faktor für die Objektstellung im Mittel- und Frühneuhochdeutschen. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 135. 1–36.

Speyer, Augustin. 2015. Object order and the thematic hierarchy in Older German. In Jost Gippert & Ralf Gehrke (eds.), *Historical corpora: Challenges and perspectives*, 101–124. Tübingen: Narr.

Telljohann, Heike, Erhard W. Hinrichs, Sandra Kübler, Heike Zinsmeister & Kathrin Beck. 2017. *Stylebook for the Tübingen treebank of written German (TüBa-D/Z)*. Seminar für Sprachwissenschaft, Universität Tübingen. https://www.sfs.uni-tuebingen.de/resources/tuebadz-stylebook-1707.pdf.

Winter, Bodo. 2020. *Statistics for linguists: An introduction using R*. New York: Routledge, Taylor & Francis Group.

Wöllstein, Angelika. 2010. *Topologisches Satzmodell*. Heidelberg: Winter.

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

Chapter 4

Choosing referential expressions and their order: Accessibility or Uniform Information Density?

Yvonne Portele[a] & Markus Bader[a]
[a]Goethe University Frankfurt

Choosing referential expressions as well as fixing word order are among the central tasks of speakers when producing language. *Accessibility* accounts focus on the fact that the accessibility status of referents involved in the current discourse is known to influence both these tasks. *Uniform Information Density* (UID) accounts highlight the role of information transmission when producing language, with consequences for choosing referential forms as well as word order. In the current article, we compare accessibility based and UID accounts and evaluate them based on linguistic findings obtained in our language production experiments on German. In line with the fact that neither one of the accounts claims to account for all aspects of referential as well as word order choices, we found that both accounts offer relevant insights. Whereas the pattern found for referential choices seems to be explained more straightforwardly in terms of accessibility compared to UID accounts, we identify helpful aspects offered by UID accounts in the domain of syntactic options. We suggest to include both perspectives to gain a more comprehensive picture in future production studies.

1 Introduction

Suppose that a speaker of German wants to inform some interlocutors that a journalist saw a former teacher in a café and that the journalist greeted the teacher warmly. For the first part of this message, the speaker will probably use indefinite NPs for introducing the two referents in a sentence starting with the subject NP, as shown in (1).

Yvonne Portele & Markus Bader. 2024. Choosing referential expressions and their order: Accessibility or Uniform Information Density? In Robin Lemke, Lisa Schäfer & Ingo Reich (eds.), *Information structure and information theory*, 87–116. Berlin: Language Science Press. DOI: 10.5281/zenodo.13383789

(1) Ein Journalist sah einen ehemaligen Lehrer in einem Café.
'A journalist saw a former teacher in a café.'

For the second part of the message, the speaker then has a whole range of options available, some of which are shown in (2).

(2) a. *Er/Dieser/Der Journalist hat ihn/diesen/den Lehrer*
he/this-one/the journalist.NOM has him/this-one/the teacher.ACC
herzlich begrüßt
warmly greeted
'He/The journalist greeted him/the teacher warmly.'

b. *Ihn/Diesen/Den Lehrer hat er/dieser/der Journalist*
him/this-one/the teacher.ACC has he/this-one/the journalist.NOM
herzlich begrüßt
warmly greeted
'He/The journalist greeted him/the teacher warmly.'

c. *Er/Dieser/Der Lehrer wurde von ihm/diesem/dem Journalist*
he/this-one/the teacher.NOM was by him/this-one/the journalist
herzlich begrüßt
warmly greeted
'He/The teacher was greeted by him/the journalist warmly.'

Referents already introduced into the discourse can be realized by a variety of referential expressions differing in terms of explicitness, including pronouns (*er* 'he'/*ihn* 'him'), demonstratives (*dieser* 'this one.NOM'/*diesen* 'this one.ACC') and definite NPs (*der*/*den Lehrer*/*Journalist* 'the.NOM/the.ACC teacher/journalist'), among others. With respect to the linear order and the syntactic functions of the two referents, the speaker also has several options, defined in terms of the linear position of the subject (sentence-initial or not) and voice (active or passive). Thus, as shown in (2), the speaker has a choice, among others, between active subject-before-object (SO) sentences (2a), active object-before-subject (OS) sentences (2b), and passive sentences (2c).[1]

Within the subfield of psycholinguistics concerned with language production, choosing referential expressions and determining word order and voice are usually considered separate topics that are investigated independently from each

[1]When we talk of passive clauses in the following, we always mean passive clauses with a *by*-phrase and the linear order subject before *by*-phrase. The *by*-phrase can also precede the subject, but passive clauses with this order are exceedingly rare in German; see Bader & Meng (2018: Table 1) for corpus evidence.

other. Despite this, a single notion has turned out to play a dominant role in both areas – the notion of conceptual accessibility. Conceptual accessibility refers to the activation of referents in short- and long-term memory.[2] Referents that are in the focus of attention of speaker or hearer are highly activated and therefore easily accessible whereas referents that are outside of the focus of attention are more difficult to access. With regard to the choice of referential expressions, less explicit expressions (e.g., pronouns) are typically used for more accessible referents whereas less accessible referents are referred to by more explicit expressions (e.g., proper names or definite NPs), as captured in various referential hierarchies (e.g., Gundel et al. 1993, Ariel 2001). With regard to the choice of word order, it has often been found that more accessible referents are produced before less accessible referents (e.g., Bock & Warren 1985, McDonald et al. 1993, Ferreira 1994, Prat-Sala & Branigan 2000), which indicates that referents are produced in the order in which they become available in memory.

Accessibility-based accounts do not claim that accessibility is the only factor governing the choice of referential expressions and the choice of word order. For example, production experiments investigating various languages have repeatedly found that speakers produce almost exclusively active subject-initial sentences when the subject referent is animate and the object referent inanimate (*The student bought the book*), in accordance with the claim that animate referents are more accessible than inanimate referents. When the subject referent is inanimate and the object referent is animate, the percentages of passives clauses increase (*The student was impressed by the book*), as expected given that an animate object is realized as a sentence-initial subject in a passive clause. Importantly, however, even in this case, participants produce a substantial number of active sentences (*The book impressed the student*), sometimes even outnumbering the production of passive sentences (e.g., McDonald et al. 1993, Ferreira 1994). Thus, almost all sentences are produced in the active voice when active is favored by animacy, whereas many, but far from all, sentences are produced in the passive voice when passive is favored by animacy. This asymmetry is usually explained with recourse to the fact that active sentences are structurally less complex and much more frequent than passive sentences. Active sentences are therefore produced by default, whereas non-active sentences are produced only when favored by sufficiently strong reasons (see the principle of Plan Reuse proposed in MacDonald 2013).

[2] Conceptual accessibility contrasts with lexical or lemma accessibility, which refers to the activation status of lexical entries in the mental lexicon (see Ferreira & Dell 2000). The latter type of accessibility can also affect the order of elements in a sentence, but for the sake of brevity, we only consider conceptual accessibility in the following.

The last fifteen years have seen the rise of information theoretic approaches to language production in which the distribution of information takes the role that the notion of accessibility has in accessibility-based accounts. The information associated with a linguistic unit is defined in terms of the unit's probability – the information conveyed by a unit is the higher the less probable the unit is. The formal definition of information, also called surprisal, is given in (3) (from Crocker et al. 2016). Surprisal is therefore inversely related to predictability: Lower surprisal equals higher predictability and vice versa.

(3) $$\text{Surprisal}(\text{unit}_i) = \log_2 \frac{1}{P(\text{unit}_i \mid \text{context})}$$

The variable *context* in formula (3) is to be understood in a wide sense. For example, when considering the surprisal associated with each word in a sentence, the context includes the words preceding the current word within the sentence, but also the linguistic and non-linguistic context preceding the current sentence.

For language comprehension, surprisal has been claimed to capture the word-by-word complexity of human parsing (Levy 2008, see Levy 2013 for critical discussion). In particular, when a word in a sentence is extremely surprising, as for example in the case of the disambiguating word of a garden-path sentence, processing breakdown may result. For smooth communication, the speaker should therefore avoid sentences with extreme information peaks. On the other hand, the speaker should also avoid sentences containing words with extremely low information value, otherwise the hearer's resources are wasted. Taken together, the speaker should strive for sentences in which information takes neither extremely low nor extremely high values. This idea is captured in the Hypothesis of Uniform Information Density (UID) given in (4) (see also Fenk-Oczlon 1989 and Levy & Jaeger 2006).

(4) Uniform Information Density (UID, Jaeger 2010: 25)
Within the bounds defined by grammar, speakers prefer utterances that distribute information uniformly across the signal (information density). Where speakers have a choice between several variants to encode their message, they prefer the variant with more uniform information density (ceteris paribus).

Models based on UID are also called *rational* models (e.g., Levy & Jaeger 2006, Arnold & Zerkle 2019, Orita et al. 2021, see also Frank & Goodman 2012 for the *Rational Speech Act Model*). The rationale behind this term is that speakers try to optimize information transmission to be efficient. Efficiency is usually driven by

two main principles: the goal to be informative and and the goal to reduce speech cost. With regard to the choice of referential expressions, speakers can be more efficient by using shorter words and phrases (e.g., a pronoun compared to a definite description) to refer to expected referents. Assuming that a referent becomes more probable (in context) with higher accessibility, UID predicts shorter word forms to be chosen for more accessible referents (e.g., Tily & Piantadosi 2009). With regard to the choice of word order, UID predicts that speakers should use word order options that allow for a more uniform distribution of information compared to other options. Depending on the differing accessibility statuses of the referents involved in a sentence, different options, for example in terms of the order of grammatical functions or the (non-)inclusion of optional elements within the sentence (e.g., Jaeger 2010), might be favored in terms of UID. Like accessibility-based accounts, UID-based accounts of language production do not claim that UID is the only factor governing the choice between competing variants.

To sum up so far, accessibility and UID have been proposed as alternative overarching influences on speakers' choices of referential expressions and word order. Since there is general agreement that speakers' choices can only be explained in a multifactorial way, language production can well be governed by accessibility and by UID, although not necessarily to the same extent. Thus, depending on which of these two notions is considered most important, we get either accessibility-based or information-based accounts.[3] In this paper, we discuss a range of recent findings from our lab that bear on the question of what is more important – accessibility or UID. Section 2 is devoted to the choice of referential expressions and Section 3 to the issue of word order and voice. The paper ends with a summary and an outlook into future research in Section 4.

2 Referential expressions

A vast literature on the production of referential expressions has shown that speakers' and writers' choices of referential forms are influenced by the accessibility of the various referents that a message is about.[4] As pointed out in the introduction, the underlying idea of accessibility accounts is that accessibility is

[3] Our distinction between accessibility-based accounts and information-based accounts corresponds to Arnold & Zerkle's (2019) distinction between *pragmatic selection models* and *rational models*.

[4] Instead of accessibility, the terms *discourse* or *cognitive status* as well as *salience* and *prominence* are often used in the literature more or less interchangeably. We use *accessibility* as an umbrella term for these different notions.

reflected by a particular referential expression chosen by the speaker which in turn helps the listener to identify the respective referent – based on the chosen form; the more accessible a referent, the more reduced the chosen referential form (i.e., pronouns or null forms).

To investigate referential form choices, participants are usually presented with story continuation tasks. They are shown sentences or sentence fragments and are asked to complete the sentence or write a natural continuation to the story. In the following, we review some of the linguistic factors identified to contribute to a referent's accessibility and consequently influencing the choice of referential expressions using story continuation tasks. In line with the literature, we focus on the most studied phenomenon within this domain – the choice of pronouns instead of definite descriptions or proper names. We will then turn to a controversial factor in the choice of referential expressions – a referent's predictability. After a short discussion of the role of predictability for accessibility and UID accounts as well as previous results, we will continue by reinspecting some of our own data with regard to the question of predictability effects on pronoun choice.

2.1 Factors governing the choice of pronouns

Several studies have shown that speakers are more likely to use pronouns for referents mentioned recently (e.g., in the same or previous sentence) compared to referents that have been mentioned at an earlier time (e.g., Givón 1983, Ariel 1990, Arnold 1998). Furthermore, there is considerable evidence that pronominalization rates are higher when referents were mentioned last in sentence-initial subject position compared to sentence-final object position (e.g., Stevenson et al. 1994, Arnold 2001, Fukumura & van Gompel 2010).

Stevenson et al. (1994) used different linguistic contexts (5) to investigate the choice of referential expressions in written production. They found that in all conditions, pronouns were produced more often for the first-mentioned subject referent, whereas names were preferred for the second mentioned referents.

(5) a. John seized the comic from Bill. … (goal-source verb)
b. Joseph hit Patrick. … (agent-patient verb)
c. Ken admired Geoff. … (experiencer-stimulus verb)
d. Simon ran towards Richard. … (agent-goal verb)

Arnold (2001: 141) used three-sentence stories including source-goal (6a) or goal-source verbs (6b).

(6) a. There was so much food for Thanksgiving, we didn't even eat half of it. Everyone got to take some food home. Lisa gave the leftover pie to Brendan. …

b. I hate getting sick. It always seems like everyone gets sick as soon as it's vacation. Marguerite caught a cold from Eduardo two days before Christmas. …

In an oral continuation task, she found that in their continuations, participants used pronouns more frequently for subject referents compared to object referents. This finding was replicated by Fukumura & van Gompel (2010) for stimulus-experiencer (7a) and experiencer-stimulus verbs (7b) in written production.

(7) a. Gary scared Anna after the long discussion ended in a row. This was because…

b. Gary feared Anna after the long discussion ended in a row. This was because…

Rohde & Kehler (2014) proposed that the main determinant of pronominalization is a referent's topic status. Following classic definitions of the information-structural concept of sentence topic, they assume that the topic of a sentence is the entity the sentence makes a statement *about* (e.g., Reinhart 1981, Lambrecht 1996). Note that in English, grammatical subject and information-structural topic are highly correlated and therefore usually confounded in linguistic materials, which might also hold for the previous studies investigating pronominalization rates mentioned above. To tease apart the influence of the subject versus topic status on the choice of referential forms, they presented participants with active and passive sentences, followed by a blank line (the continuation prompt).

(8) a. Amanda amazed Brittany. …

b. Brittany was amazed by Amanda. …

The comparison of active and passive sentences is based on the assumption that whereas the subject is the default topic in active sentences such as (8a), the passive structure (8b) establishes the subject as topic more strongly – promoting the non-agent to the subject position. If topichood instead of subject status is the main force boosting a referent's accessibility and consequently the choice to pronominalize this referent, pronoun rates in continuations should be higher in passive compared to active contexts. This is indeed what Rohde & Kehler (2014) found with a written continuation task. Subjects of passive sentences were pronominalized more often than subjects of active sentences, whereas non-subject referents were pronominalized at a similar rate for the two structures.

2.2 The predictability controversy

A longstanding controversy with regard to the use of referential expressions concerns the question of whether the *predictability* of a referent has an influence on the choice of a referential expression. Several studies have shown that certain verbs establish strong expectations with regard to re-mentioning one of the referents involved in the event denoted by the verb. Implicit causality biases are among the best studied phenomena in the literature on thematic role expectations (e.g., Garvey & Caramazza 1974, Au 1986, Bott & Solstad 2014).

Following sentence fragments such as (9a) (see also example (5c) of Stevenson et al. 1994, and example (7) of Fukumura & van Gompel 2010), participants systematically re-mention Kathy (the previous subject; e.g., *…because she/Kathy was so helpful*) when completing fragments such as (9a), whereas they re-mention Tom (the previous object; e.g., *…because he/Tom was so helpful*) when completing fragments such as (9b) in continuation tasks (see also Holler & Suckow 2016, Bittner 2019, Portele & Bader 2020 for German).

(9) a. Kathy impressed Tom because…
b. Kathy admired Tom because…

The re-mentioning preferences have been ascribed to the stimulus of the event being regarded as causing the psychological state of the experiencer. People are more likely to refer to the implicit cause of the event, especially in the context of explicit causal discourse connectives such as *because*. Similarly, continuation experiments investigating transfer of possession contexts such as (10) have shown that participants start their completions by re-mentioning the goal – Kathy in (10a), Tom in (10b) – more often than referring back to the source (e.g., Stevenson et al. 1994, Arnold 2001, Kehler & Rohde 2013).

(10) a. Kathy got a present from Tom. …
b. Kathy gave a present to Tom. …

The prevalence of certain thematic arguments to be mentioned again due to verb meaning and discourse connectives has been subsumed under the notion of *semantic biases*. Of interest for the discussion at hand is the fact that several researchers have ascribed the preference for a certain referent to be mentioned again, its higher predictability, to being more accessible than the alternative referent. Assuming that predictability affects the discourse status of referents, we might expect there to be influences on the choice of referential expressions. There

is, however, mixed evidence from previous work, leading to an inconclusive pattern. Whereas thematic preferences in terms of likelihood of reference have been replicated for both implicit causality as well as transfer of possession verbs, studies have found differing results with regard to the choice of referential forms. Studies involving transfer of possession verbs provide evidence that the choice to pronominalize can indeed be influenced by predictability or likelihood of reference. Arnold (2001) found that participants are more likely to pronominalize goal characters compared to source characters. This finding was also replicated by Rosa & Arnold (2017). In implicit causality contexts, however, several studies did not find effects of the more likely referent to be mentioned again on the choice of referential expressions (Stevenson et al. 1994, Kehler et al. 2008, Fukumura & van Gompel 2010), but in recent work, Weatherford & Arnold (2021) could reveal influences of semantic predictability on the choice of referential forms in implicit causality contexts. They found that participants were more likely to pronominalize stimulus referents compared to experiencer referents, but this effect was limited to object referents. The mixed results found in terms of predictability effects on pronoun production have been ascribed to the different verb types investigated (transfer of possession vs. implicit causality verbs) as well as to task variation (sentence continuation tasks vs. production tasks encouraging stronger discourse representations).

The question whether predictability influences the production of referential expressions is of utmost importance for the current discussion, since this is exactly what UID accounts (see Introduction) predict. The higher the predictability of a certain word, the less information the respective word carries. Jaeger (2010: 48), for example, states: "Speakers should be more likely to produce pronouns (e.g., *she*) instead of full noun phrases (e.g., *the girl*) when reference to the expression's referent is probable in that context".

In their information theoretic study, Tily & Piantadosi (2009) had participants guess upcoming referents in authentic contexts. In a second analysis, they investigated whether writers were influenced by the predictability of the referents (participants' guesses) when choosing referential forms. The authors found that pronouns were indeed used more often when the referent was predictable, suggesting an influence of predictability on the choice of referential expressions in line with UID accounts. Tily & Piantadosi (2009) conclude that "[p]ronouns in particular provide language with context-dependent code that allows more predictable nouns to be referenced with a shorter word. These results align with recent production theories such as Uniform Information Density (Genzel & Charniak 2002, Jaeger 2006, Levy & Jaeger 2006)".

In a recent study, Orita et al. (2021) compared an accessibility based model with an informativity driven model of referential choices. The accessibility based model (called the *topicality* model by Orita et al.) was created under the assumption that the topichood of a referent is the main determinant of choosing referential expressions (as discussed above). The informativity model (called the *rational* model by Orita et al.) was based on the assumption that referential choices reflect the amount of information words carry in discourse, on speakers' speech cost, and on the predictability of referents. The results of their simulations suggest that both the referent's topichood status as well as word informativity are important factors in the choice of referential forms.[5]

To investigate the question whether referent predictability affects the choice of referential expressions, we returned to a free sentence continuation study we conducted some years ago, investigating the interpretation and production of pronouns in German. The respective study was published in Bader & Portele (2019).

2.3 A new look at Experiment 3 of Bader & Portele (2019)

Experiment 3 of Bader & Portele (2019) used a free sentence continuation task in order to obtain the production data necessary to test the Bayesian theory of pronoun resolution proposed by Kehler et al. (2008) (see also Kehler & Rohde 2013). According to this theory, pronoun resolution is based on two production probabilities – the probability of which referent to mention next and the probability to use a pronoun for a given referent. These two probabilities, which are assumed by the Bayesian theory to be independent of each other, are combined using Bayes formula to predict the most likely referent of a referentially ambiguous pronoun. Although Experiment 3 was run without having UID in mind, the two probabilities needed to apply the Bayesian theory are exactly the two probabilities needed to test the prediction based on UID accounts that pronoun production is influenced by referent predictability. Free sentence continuation allows us to measure surprisal/predictability values by analyzing which of the referents mentioned in the previous context was taken up again in the written continuation. By looking at the rate of different referential expressions used, we can measure pronoun rates for the respective referents. The predictions based on information theoretic models, such as UID, is that we should find higher pronoun rates for predictable/less surprising referents. Importantly, the experiment that we discuss used verbs without strong semantic biases with regard to which

[5]The speech cost factor of preferring shorter word forms, on the other hand, did not turn out to be an influential factor to the choice of referential form.

referent is mentioned next. This avoids potential pitfalls due to (strong) semantic biases of differing verb types (e.g., psych verbs, transfer-of-possession verbs). These biases have been discussed as one of the reasons for the mixed results in terms of predictability effects on pronoun production.

The materials for this experiment were a German adaption of sentences used in a study conducted by Kaiser & Trueswell (2008) in Finnish. A complete item is shown in Table 1.[6] Experimental items consisted of three context sentences followed by a blank line, the free continuation prompt. The first context sentence (C1 in Table 1) constituted a scene-setting sentence. This sentence always introduced a female character in the form of a proper name. The second context sentence (C2 in Table 1) re-mentioned the female character in the form of a personal pronoun and introduced a male character by using an indefinite NP. The third context sentence (C3 in Table 1) re-mentioned this male character by using a definite NP and introduced a second male character in the form of an indefinite NP. A relative clause modified the indefinite NP object. The last context sentence was followed by a blank line, the prompt eliciting participants' continuations.

Table 1: Sample stimulus for Experiment 1 (from Bader & Portele 2019).

C1	Sabine (C_1) war am Sonntag im Zirkus. 'Sabine (C_1) visited a circus on Sunday.'
C2	Bevor die Aufführung begann, hatte sie schon einen Clown (C_2) herumlaufen sehen. 'Before the show began, she saw a clown (C_2) walking around.'
C3	Der Clown (C_2) umarmte einen Mann (C_3), der ganz wirre Haare hatte. 'The clown (C_2) hugged a man (C_3) with completely tousled hair.'

Forty-four students of the Goethe University Frankfurt read 16 experimental and 24 filler contexts and wrote a sensible continuation sentence for each context. Participants were asked to provide a complete continuation sentence, but there were no further restrictions in terms of form or content of the continuation. For the continuations given by participants, we scored which characters were mentioned again (female referent, first male NP, second male NP) and which referential expression was chosen to refer back to the referent (pronoun, proper name, definite NP). For continuations containing more than one of the human referents, the first referent given in the continuation was counted. Exemplary

[6]We limit ourselves to one of the four conditions of the original experiment.

continuations together with their respective scoring information are shown in (11).

(11) a. Re-mentioned: female character – Form: pronoun
Sie hoffte, dass der Clown sie in Ruhe ließ.
'She hoped the clown would leave her alone.'

b. Re-mentioned: female character – Form: proper name
Sabine war froh, dass es nicht sie getroffen hatte.
'Sabine was glad it was not her.'

c. Re-mentioned: first male NP – Form: definite NP
Der Clown hatte ebenso wirre Haare.
'The clown also had tousled hair.'

d. Re-mentioned: first male NP – Form: pronoun
Er war später im Programm sehr lustig.
'He was very funny later on in the show.'

e. Re-mentioned: second male NP – Form: definite NP
Der Mann saß in der ersten Reihe und freute sich über jeden Auftritt.
'The man was sitting in the front row and was happy about every appearance.'

f. Re-mentioned: second male NP – Form: pronoun
Er war der Bruder des Clowns.
'He was the clown's brother.'

To estimate referent predictability/surprisal, we looked at the re-mention rates of the three human characters. Percentages of referents from the context occurring as the only or first referent in participants' continuations and the derived surprisal levels are shown in Table 2. The female character (introduced in the first context sentence) was taken up again in 53% of the continuations provided by participants, followed by the second male character (introduced in context sentence three) with 23% and the first male character (introduced in the second context sentence) with 13% of references. We think that the majority of continuations refers back to the female character because she is interpreted as the discourse topic of the narrational contexts (e.g., Asher 2004). Based on these percentages, the female character is the most predictable or least surprising referent, followed by the second male character with a medium predictability/surprisal level, whereas the first male character is the least predictable/most surprising character. If the use of pronouns is indeed driven by information theoretic considerations as discussed above, pronoun rates should be inversely related to the

surprisal levels of the referents. We would therefore expect to find the highest pronominalization rate for the female character, whereas the first male character should be pronominalized least often and the pronoun rate for the second male character should fall between these two.

Table 2: Percentages of referents from the context occurring as the only or first referent in participants' continuations

Referent	% Re-mention	
C_1 (Female NP)	53	→ high predictability/low surprisal
C_2 (1. male NP)	13	→ low predictability/high surprisal
C_3 (2. male NP)	23	→ medium predictability/medium surprisal

Table 3: Percentages of referents from the context occurring as the only or first referent in participants' continuations and percentages of referential forms for each referent

		Referential expression		
Referent	% Re-mention	Pronoun	Proper name	Definite NP
C_1 (Female NP)	53 → low surprisal	26	74	
C_2 (1. male NP)	13 → high surprisal	67	–	33
C_3 (2. male NP)	23 → medium surprisal	11	–	72

Table 3 shows percentages of referential forms for each referent. Contrary to expectations based on UID accounts to pronominalization, the highest pronoun rate (67% pronouns vs. 33% definite NPs) was found for reference to the first male character (the subject of the preceding clause), the most surprising or least expected referent based on re-mention rates. The second highest pronoun rate (26%) was found for the least surprising referent (the female character). Note, however, that overall the female character was re-mentioned more often with a proper name (74%) compared to a personal pronoun. The lowest pronoun rate was found for the medium expected/surprising second male character (11% vs. 72% definite NPs).

2.4 Discussion

In order to shed further light on the question whether referent predictability guides the choice of referential expressions, we looked at data from a former experiment (Experiment 3 of Bader & Portele 2019) that used a sentence continuation task to measure referent predictability (Who is mentioned again in the continuation?) as well as pronominalization rates (How often did participants use a pronoun to talk about the referent mentioned again?). UID accounts (e.g., Fenk-Oczlon 1989, Levy & Jaeger 2006, Tily & Piantadosi 2009) predict that pronoun rates should be higher for predictable and thus less surprising referents. However, the re-mention ranking found in our continuation results, which was used to derive surprisal levels, was not mirrored by participants' choice of referential expressions. Pronoun use was not highest for the least surprising referent, as might be expected under an information-theoretic account of referring expressions. To the contrary, we found the highest pronominalization rate for the most surprising/less expected referent to be taken up again in the continuation. This referent, however, was the (subject) topic of the previous sentence. In line with previous studies (e.g., Fukumura & van Gompel 2010, Rohde & Kehler 2014) we therefore suggest that accessibility based accounts are better suited to account for the referential pattern found in our study.

A potential argument against this conclusion might be the fact that we used contexts involving alternative referents or *competitors*. In our materials, the contexts introduced one female character and two male characters. Whereas the choice of the feminine third person singular pronoun (*sie* 'she') should not be influenced by the other two male referents, the choice of using a potentially ambiguous masculine third person singular pronoun (*er* 'he') might be influenced by the fact that there are two characters of the same gender participating in the context, therefore establishing two potential referents for one and the same pronoun. Both accessibility as well as UID accounts of the choice of referential expressions acknowledge a role of potential competitors in contexts. Within accessibility accounts, *competition* (e.g., Ariel 1990) or *interference* (e.g., Givón 1984) are taken to influence a referent's accessibility status. In contexts involving more than one potential referent for the pronoun, speakers should therefore use more explicit expressions (i.e., definite NPs or proper names). This prediction was borne out in studies investigating the production of referential expressions in English (e.g., Arnold & Griffin 2007). Within UID accounts, contextual competitors are taken into account since the informativity of a referential expression, the personal pronoun in this case, decreases with an increase of contextual competitors (see also Orita et al. 2021 for a formalization of competitors within an information theoretic model evaluation).

We are well aware of the fact that there are potential competitors present in our experimental materials. However, we want to bring up two thoughts for further discussion. The highest pronoun rate we found in our data was for re-mentioning one of the two male characters, which was the most surprising referent based on the re-mention rates. If the presence of potential competitors for the use of a specific referential expression (in this case the male personal pronoun) was an influential factor in the current study, it is not clear why we should find the overall highest pronoun rate (67% pronouns vs. 33% definite NPs) for one of the two male characters, since the use of a potentially ambiguous pronoun still leads to a decrease in terms of informativity of the pronoun. On the other hand, if the highest pronoun rate found in this study is already influenced by the fact that there is a competitor present in the context in the sense that without this competitor, we might even expect higher pronoun rates for this referent, this does not alter the fact that we are still talking about the referent having the lowest predictability value to be taken up again in the continuation. Taking competitors into account therefore does not remove the argument against UID accounts.

Furthermore, we should also focus on the referent with the lowest surprisal level, or in other words, the highest predictability – the female referent. We have attributed the highest re-mention rate for the female referent to the fact that she establishes the discourse topic in the contexts. The female character is introduced in the first (scene-setting) context sentence. She is re-mentioned in the form of a personal pronoun in the second context sentence (which also introduced the first male character) and picked up most often in the continuation following the third context sentence (that mentioned both male characters). However, when looking at the referential expressions chosen to refer to the female character in the continuations, we again do not see a prevalence of pronouns to refer back to this character. Whereas pronouns were chosen in 26% of cases, the majority of cases (74%) is constituted of proper names. Although this pattern might be explained in terms of (non-)recency of the respective referent (there is one sentence intervening between the last mention of the female character and the continuation) or stylistic reasons, this is not what we would expect based on information theoretic considerations. This is at least what holds for the singular pronoun. However, note that the feminine third person singular pronoun *sie* ('she') is identical in form with the German third person plural pronoun. The personal pronoun could in principle refer to, for example, the two male characters mentioned in the sentence preceding the continuation prompt. When extending the notion of potential competitors to referents differing in number, one might therefore argue that an avoidance of the personal pronoun might be due to preventing a potential ambiguity. Since participants were asked to write complete

sentences, this potential ambiguity, however, would be limited to the pronoun itself in most cases, since the finite verb in second position would disambiguate the pronoun immediately.

3 Word order

We now turn to the question of how language producers determine in which order they mention the referents of a message. Since its introduction by K. Bock and colleagues (e.g., Bock & Warren 1985), the notion of conceptual accessibility has played a dominant role with regard to this question. The accessibility of referents depends on their inherent features, like animacy, and on features temporarily derived from the context in which the referents appear, like givenness or contextual salience. Both inherent and derived accessibility have been shown to influence how speakers order the various referents of a message (e.g., McDonald et al. 1993, Ferreira 1994, Prat-Sala & Branigan 2000) – the most accessible referent tends to start the sentence, subject to independent constraints like Plan Reuse (MacDonald 2013).

The question of how speakers choose the order of words has also received some attention within the UID framework. For ease of reference, we repeat the UID hypothesis in (12).

(12) Uniform Information Density (UID, Jaeger 2010: 25)
Within the bounds defined by grammar, speakers prefer utterances that distribute information uniformly across the signal (information density). Where speakers have a choice between several variants to encode their message, they prefer the variant with more uniform information density (ceteris paribus).

When encoding a given message, speakers can typically choose from a range of different syntactic structures that vary along several dimensions – besides having different word order, the variants can differ in terms of optional elements (e.g., the complementizer *that*) and in terms of elements that can be produced in more or less reduced forms (e.g., proper names versus pronouns). In contrast to phenomena involving optional and/or reduced items, we are not aware of much research applying the UID approach to the issue of word order variation. Furthermore, the research we are aware of (Maurits et al. 2010, Collins 2014, Jain et al. 2018, Rubio-Fernández et al. 2021) has yielded mixed results, sometimes supporting the UID approach to word order and sometimes contradicting it.

In the following, we focus on German main clauses with an agent and a patient argument. Such clauses can have an active SO structure, in which case the agent argument precedes the patient argument. Alternatively, the patient argument can be moved in front of the agent argument either by producing a passive clause or by producing an active OS clause. Depending on the particular structure, information will be more or less evenly distributed. In the corpus study of Hoberg (1981: 162), about 58% of all German main clauses started with the subject and only about 3.5% with an object, the rest being made up of sentences starting with an adverbial or a predicate nominal. According to these data, encountering an object as the first phrase of a sentence should be surprising whereas encountering a subject should be unremarkable. Of course, the surprisal value of a phrase is not only a function of its syntactic function. As defined in (3), the surprisal value of a linguistic unit is inversely related to its predictability in context. For a phrase in sentence-initial position, the preceding sentences constitute the context; for later parts of a sentence, the context is made up of the preceding sentences and the initial part of the sentence.

Consider first the sentence-initial phrase, which for German main clauses is the phrase occupying the so-called prefield. The probability that either the agent or the patient will be mentioned next – the so-called next-mention bias – depends on semantic and pragmatic properties of the preceding context. For example, if the last context sentence contains an implicit causality verb, chances are high that the stimulus will be mentioned next whereas with verbs that are not associated with strong semantic biases, which argument is mentioned next is less predictable a priori. However, whatever referent of the preceding context is selected to start the next sentence (if any is selected, because starting a sentence with an adverbial is also common), it will not be surprising if the referent occurs with the syntactic function of subject whereas it is highly surprising if it occurs as object.

With regard to the surprisal values of elements following the sentence-initial phrase, general statements do not seem to be possible, not least because of the verb-second nature of German. In a declarative main clause, the initial phrase is necessarily followed by the finite verb in German in verb-second position, but the identity and thus the predictability of the finite verb can vary widely, even when taking into account the first phrase. For illustration, consider the example in (13). Here, the sentence-initial position is occupied by the subject. Besides predicting that a finite verb form comes next, a sentence-initial subject imposes few constraints.

(13) Der Bussard … 'The buzzard'

a. … ***attackierte*** *den Läufer / **hat** den Läufer attackiert.*
attacked the runner / has the runner attacked
'… attacked the runner/has attacked the runner.'

b. … ***wurde*** *von dem Läufer erschreckt. / **ist** von dem Läufer erschreckt*
was by the runner frightened / is by the runner frightened
worden.
been
'… was frightened by the runner / has been frightened by the runner.'

c. … ***hat*** *ein großes Nest / **ist** ziemlich gefährlich / **wurde** immer*
has a large nest / is rather dangerous / became always
lauter.
louder
'… has a large nest / is rather dangerous / became louder and louder.'

On the one hand, the sentence can be an active sentence with a transitive verb, as in (13a). In this case, the lexical verb serves as finite verb itself, or it appears in a non-finite form, for example, as a participle in which case the finite verb is a form of the auxiliary *haben* ('have'). Any particular main verb has a low predictability (because there are so many different main verbs) unless there are strong contextual constrains. A perfect auxiliary, in contrast, is much more predictable and thus less surprising for the simple reason that perfect tense is a frequent tense form and *haben* is the perfect auxiliary of a very large number of verbs. Note furthermore that the predictability of the following object also strongly depends on the finite verb in verb-second position. A transitive main verb in verb-second position makes an upcoming object predictable and thus less surprising whereas an auxiliary in verb-second position is also compatible with intransitive verbs, making the appearance of an object less predictable. As shown in (13b), a subject in sentence-initial position can also start a passive clause, which can be signaled by a passive auxiliary as finite verb. In a composite tense form, however, the passive auxiliary appears clause-finally and a form of the perfect auxiliary *sein* ('be') occupies the verb-second position. Since this is also the perfect auxiliary for a certain subset of intransitive verbs, it is only mildly predictive of a passive clause. The situation is even further complicated by the fact that all auxiliaries also have other uses – as main verb or as copula verb, as illustrated in (13c).

In sum, although it seems impossible to make general statements about the distribution of information across sentences, it seems safe to conclude that information is typically relatively uniformly distributed when a sentence starts with

the subject whereas a sentence-initial object leads to a considerable information peak at the sentence beginning. Thus, all else being equal, a speaker adhering to UID will prefer to produce subject-initial sentences. This still leaves open the choice between a sentence in the active voice and a sentence in the passive voice. Whether the fact that active SO sentences are much more common than passive sentences also follows from UID must be considered an open question. If not, an independent principle like Plan Reuse would have to be invoked, too.

The question now is whether conditions favoring patient-initial orders remove or at least reduce the disadvantage of patient-initial sentences with regard to the distribution of information. In principle at least, this seems to be the case. One recurrent finding in word order research has been that animate referents tend to precede inanimate referents. This is compatible with UID accounts given that animate referents (and human referents in particular) are typically more predictable than inanimate referents, both at the conceptual and at the lexical level. A further recurrent finding, namely that contextually more salient referents tend to occur earlier than contextually less salient referents, is also compatible with UID. If a referent is contextually more salient, it is more likely that the referent is mentioned again, that is, the predictability of the referent increases, making it more likely that the referent will occur in initial position.

To sum up so far, basic findings concerning the choice between SO and OS order can – at least in principle – be explained both in terms of accessibility and UID. In the rest of this section, we discuss a range of more specific findings from our lab concerning word order in German main clauses with two arguments. These findings show, among others, that the relationship between topichood and word order is more complicated than sketched above: Contrary to common assumptions, topics precede non-topical referents only under specific conditions. We first present these findings and afterwards discuss how they may be accounted for in terms of accessibility or in terms of UID.

3.1 Producing agent-patient sentences in context

In order to further explore the role of accessibility for the choice of referential expressions and the choice of word order, we have run a series of picture description experiments in the last few years. For reasons of space, we concentrate here on pictures depicting an event in which an animate agent acts on an animate patient, for example, a doctor examining a teacher. In two experiments reported in Bader et al. (2017), pictures were preceded by contexts consisting of a sentence introducing an agent and a patient and a question which varied across experiments. In Experiment 2 of Bader et al. (2017), the question was a wide-focus one.

In Experiment 3 of Bader et al. (2017), it was a narrow-focus question asking for further information about the patient argument, thereby establishing the patient as topic of the following picture description. For the picture of a doctor examining a teacher, the two kind of questions are shown in (14).

(14) a. Wide-focus question
Hier geht es um einen Arzt und einen Klavierlehrer. Was ist zu sehen?
'This picture involves a doctor and a piano teacher. What can be seen?'

b. Narrow-focus question with patient as topic
Hier geht es um einen Arzt und einen Klavierlehrer. Was lässt sich über den Klavierlehrer sagen?
'This picture involves a doctor and a piano teacher. What can one say about the piano teacher?'

With regard to word order, these two experiments revealed a striking effect of the context question: With a wide focus question, almost all sentences were produced with SO order (*Der Arzt hat den Lehrer untersucht.* 'The doctor examined the teacher.'). With a narrow focus question that established the patient as topic, the large majority of sentences were passive sentences with the patient argument in initial position (*Der Lehrer wurde von dem Arzt untersucht.* 'The teacher was examined by the doctor.').

The contexts in (14) are admittedly somewhat artificial and situations where contexts of this kind are produced may be rare (although not impossible). We therefore ran a further set of experiments in which pictures were embedded in an evolving story. To this end, the contexts preceding the pictures consisted of a header introducing either the agent or the patient as the topic followed by three sentences. The contexts thus formed the beginning of a story that was continued by the event depicted in the picture. In Experiment 2 of Bader & Portele (2023), the picture was preceded by an agent or a patient context as shown in (15).

(15) a. Topic = Agent
Der beste Arzt
In unserem Viertel gab es einen sehr guten Arzt.
Dieser Arzt konnte fast immer helfen.
Einmal musste er einen scheinbar schwerhörigen Klavierlehrer behandeln.
'The best doctor – A very good doctor was practicing in our quarter. This doctor could help almost always. Once he had to treat a seemingly hearing-impaired piano teacher.'

b. Topic = Patient
Sorgen eines Klavierlehrers
In unserem Viertel gab es einen guten Klavierlehrer.
Dieser Klavierlehrer hatte eine Zeit lang Probleme beim Hören.
Einst suchte er einen angesehenen Ohrenarzt auf.
'Sorrows of a piano teacher – A good piano teacher was living in our quarter. This piano teacher was having hearing problems for quite a while. Once he visited a respected ear specialist.'

24 students of the Goethe University Frankfurt, who participated for course credit, read each context and then described the picture using the verb shown above the picture. Participants' descriptions were digitally recorded and transcribed. In this experiment, participants produced almost only sentences with SO order, but topichood still had a significant effect. When the patient was the topic, the percentage of SO sentences was only slightly below 100%. When the agent was the topic, the rate of SO sentences was about 90%, which was significantly lower than the almost 100% percent in the condition with a patient topic. Thus, with an agent as topic, about 10% patient-initial sentences were produced, which means that when patient-initial sentences were produced, they had the non-topic in first and the topic in second position. A final finding of this experiment was that the object in OS sentences was realized as a demonstrative pronoun in most cases.

A comparison of the findings for question contexts as in (14) and story contexts as in (15) reveals two puzzling differences. First, the experiment in which a question set the patient as topic (*What can one say about the piano teacher?*) found a high rate of patient-initial sentences. With a story-like context, in contrast, almost only agent-initial sentences, that is, SO-sentences, were produced when the patient was the topic whereas a small number of patient-initial sentences was observed when the agent was the topic. Second, whereas the patient-initial sentences in the experiment with a context question were produced as passive sentences in most cases, in the experiment with story-like context mainly OS sentences were observed. In the next two sections we discuss how far accessibility (Section 3.2) and UID (Section 3.3) take us in accounting for these differences.

3.2 Accessibility-based accounts of word order in German main clauses

The finding of mainly agent-initial sentences following a wide-focus question (see 14a) and mainly patient-initial sentences following a narrow-focus question

with the patient as topic (see 14b) is clearly compatible with accessibility-based accounts of grammatical encoding. First, prior research has shown that the momentary accessibility of a referent increases when it is assigned a thematic role that is high on the thematic hierarchy (e.g., Ferreira 1994). The accessibility of an agent argument is therefore higher than the accessibility of a patient argument so that the agent argument will precede the patient argument when both arguments are on a par in other respects, like animacy and givenness. This was the case in the experiment with a wide-focus question. In the experiment with a narrow-focus question, in contrast, the question explicitly established the patient referent as topic. Given that in this experiment most sentences started with the patient topic, we can conclude that the topic-setting question increased the accessibility level of the patient to a larger degree than the accessibility increase brought about by being the agent. The only finding which accessibility does not account for is that the patient was almost always inserted in sentence-initial position by means of passivization whereas OS sentences, which allow fronting the patient without a concomitant change in voice, were almost never produced.

Accounting for the findings of the experiment in which the picture was preceded by a story-like context is less straightforward in terms of accessibility. First, when the patient was the topic, almost only agent-initial SO sentences were produced. This is in stark contrast to the finding of a majority of patient-initial passive sentences when the patient was explicitly established as topic by means of a question. Second, when patient-initial OS sentences were produced at all, they were produced in the condition with the agent as topic, that is, when the agent was more accessible than the patient.

As stressed several times before, there is general agreement that determining word order during language production is subject to a multitude of factors. By itself, the fact that some findings cannot be accounted for in terms of accessibility is therefore not worrisome. The question of course is whether there are plausible and independently motivated factors that can explain those findings not accounted for by accessibility.

Consider first the finding that the patient was fronted when a question explicitly established it as topic but not when it was implicitly established as topic by the preceding context. With regard to this finding, we can follow Rohde & Kehler (2014) and assume that the boost in accessibility brought about by topicalization varies with the degree of explicitness of establishing the topic. Independent reasons also account for the additional finding that the patient was preposed mainly by means of passivization when it was explicitly set as topic by the preceding question. As proposed in Bader (2020), this finding follows from the preference

of topics to be realized as subjects. When the patient argument is the topic, passivization achieves the preferred association of topichood and subjecthood.

Consider finally the finding that a small number of OS sentences was observed when the agent was the topic and the object was realized as a demonstrative. Research has found that demonstrative objects are especially prone to occur sentence-initially (e.g., Bader & Portele 2021). This preference can tentatively be ascribed to a violable constraint requiring demonstrative NPs to occur as close to their antecedent as possible. In the case of a demonstrative object, this can be achieved by fronting the object, which decreases the distance between object and antecedent. Since demonstratives are typically confined to non-topical referents, it also follows that OS sentences with a demonstrative object only occur when the agent is the topic and the patient (= the object) is not the topic.

3.3 UID-based accounts of word order in German main clauses

We now consider how the UID account fares with regard to the findings from picture description summarized above. As discussed before, subject-initial sentences exhibit a more uniform information density than object-initial sentences. This is in agreement with the finding that participants basically produced only SO sentences in the experiment with a wide-focus question, where the context introduced both arguments without making one of them more salient or prominent than the other. When a narrow-focus question established the patient as topic, it seems plausible to assume that the patient becomes more predictable. After all, when one is asked to provide information about a particular referent, it is highly likely that the answer will contain a reference to this referent. Because topics are preferentially realized as subject in sentence-initial position, a passive structure is the optimal choice because it allows the sentence to start with a subject. Thus, like the accessibility account, the UID account has no difficulty accounting for word order in contexts with wide or narrow focus question.

Consider next the finding that following a story-like context with the patient introduced as topic, participants produced almost only SO sentences, that is, sentences with the non-topical agent in first position. This seems to be in conflict with the high number of patient-initial passive clauses when the patient was established as topic by means of a narrow-focus question. However, under a UID perspective, there is not necessarily a conflict between these two results. Under a UID perspective, what counts is the predictability of referents. With an explicit topic-setting question, the predictability of the topic referent is high, as argued above. Whether the same is true without an explicit topic-setting question is not so clear. Although it is often claimed that continuing with the topic referent has

a privileged status (see, for example, the preferred transition types of Centering Theory, Grosz et al. 1995), free continuation experiments do not in general confirm this claim. To determine whether the observed preference for agent before patient is compatible with UID, it is therefore necessary to determine experimentally which referent is more likely to be re-mentioned and thus more predictable – the topic or the non-topic. Running the necessary free continuation experiments must be left as a task for future research.

The final finding to consider is that active sentences with OS order were rarely produced and if so, only in contexts where the object was not the topic and the object was realized by a demonstrative pronoun. Does the observed association between the form of the object and its position follow from UID? Since sentence-initial objects are unexpected, a more explicit form should be chosen for them. Assuming the simplified referential hierarchy "def NP > demonstrative pronoun > personal pronoun" (modified from Kaiser & Fedele 2019: 313), definite NPs are most explicit and personal pronouns least explicit, with demonstrative pronouns in between. The inverse correlation between explicitness and predictability therefore explains the rareness of OS sentences when the object is realized as a personal pronoun. On the other hand, it remains unclear why demonstrative pronoun objects show a preference for OS order whereas definite NPs do not, although the latter are more explicit and should thus be especially appropriate for the highly surprising sentence-initial objects. Thus, like the accessibility account, the UID account needs an independent explanation for this finding. As in the case of the accessibility account, a constraint favoring short distances between demonstratives and their antecedents could be hypothesized for this purpose.

4 General discussion

In this article, we discussed two central notions within the domain of choosing referential expressions and word order during language production: accessibility and UID. Accessibility accounts focus on the differing accessibility status of referents involved in the current discourse. Accessible referents are highly activated and more easily retrieved from memory. When choosing referential expressions, less explicit referential forms (e.g., pronouns) are chosen for more accessible referents. When determining word order, more accessible referents are produced earlier than less accessible referents. UID accounts, on the other hand, focus on information transmission. The underlying assumption is that speakers prefer choices leading to uniform information density. When choosing referential

expressions, speakers are expected to choose shorter expressions (e.g., pronouns) for more predictable referents, since the higher the predictability of a word, the less information it carries. With regard to word order, UID prefers orders in which constituents are neither highly predictably nor highly unpredictable.

Note that neither accessibility nor UID accounts claim that the respective factors constitute the sole determinants of referential forms as well as word order during language production. In order to adjudicate between accessibility and UID accounts, we (re-)evaluated some data that we gathered in language production experiments over the last few years. To evaluate the question whether referent predictability influences the choice to pronominalize, we revisited Experiment 3 of Bader & Portele (2019). The sentence continuation task used in this experiment allowed us to derive surprisal/predictability values based on data on which referent was mentioned next, following a context that introduced several referents. By looking at the rates of different expressions chosen to refer back to this referent, we were able to investigate the question whether more predictable referents were pronominalized more often compared to less predictable referents, as expected within UID accounts. The surprisal levels derived for referents were, however, not reflected by participants' referential expressions. The highest pronominalization rate was indeed found for the most surprising referent. In the materials presented, this referent was the topic/subject of the preceding sentence. Thus, for the data at hand, accessibility accounts for pronominalization, claiming that distance and grammatical function/topic status are the main force behind pronoun choice (e.g., Fukumura & van Gompel 2010, Rohde & Kehler 2014), are superior to UID accounts.

With regard to word order, our discussion was more open-ended. We presented a range of findings of an ongoing series of picture description experiments. Most of the results turned out to be compatible with both accessibility accounts and with UID accounts. For some findings, however, additional assumptions were necessary for accessibility based-accounts. One question was whether these additional assumptions could be avoided under an UID perspective. The answer to this question was mixed. On the one hand, the presence or absence of topic effects on word order may follow when looking at predictability instead of topichood. On the other hand, the association between referential form and word order (OS with demonstrative object) turned out to be difficult to reconcile with UID.

It's in the nature of things that the factors discussed in this article are not exhaustive. Previous research within both domains, for example, highlights communicative aspects in terms of speaker vs. listener, potential ambiguity avoidance, audience-design, etc. In the experiments discussed here, we looked at writ-

ten data gained in monologue lab settings. Comparing these results to patterns gained by using different, more interactive tasks as well as within different linguistic contexts are just a few of the many options to continue this line of work. We hope that the current article helps to identify some fruitful aspects for the investigation of accessibility vs. UID accounts to language production, thereby offering a starting point for future research. Including and comparing considerations from both accounts will certainly be of crucial value for investigations within the domain of language production.

Acknowledgements

The authors would like to thank the organizers of the workshop *Discourse obligates* at the 2022 annual meeting of the German Linguistic Society, the audience at the workshop, as well as two anonymous reviewers for helpful discussions and comments.

References

Ariel, Mira. 1990. *Accessing NP antecedents*. London: Routledge.

Ariel, Mira. 2001. Accessibility Theory: An overview. In Ted Sanders, Joost Schilperoord & Wilbert Spooren (eds.), *Text representation: Linguistic and psycholinguistic aspects*. Amsterdam: Benjamins. DOI: 10.1075/hcp.8.04ari.

Arnold, Jennifer E. 1998. *Reference form and discourse patterns*. Stanford University. (Doctoral dissertation).

Arnold, Jennifer E. 2001. The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes* 31(2). 137–162. DOI: 10.1207/S15326950DP3102_02.

Arnold, Jennifer E. & Zenzi M. Griffin. 2007. The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language* 56(4). 521–536. DOI: 10.1016/j.jml.2006.09.007.

Arnold, Jennifer E. & Sandra A. Zerkle. 2019. Why do people produce pronouns? Pragmatic selection vs. rational models. *Language, Cognition and Neuroscience* 34(9). 1152–1175. DOI: 10.1080/23273798.2019.1636103.

Asher, Nicholas. 2004. Discourse topic. *Theoretical Linguistics* 30. 163–201. DOI: 10.1515/thli.2004.30.2-3.163.

Au, Terry Kit-fong. 1986. A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language* 25(1). 104–122. DOI: 10.1016/0749-596x(86)90024-0.

Bader, Markus. 2020. Objects in the German prefield: A view from language production. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 15–39. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198844303.003.0002.

Bader, Markus, Emilia Ellsiepen, Vasiliki Koukoulioti & Yvonne Portele. 2017. Filling the prefield: Findings and challenges. In Constantin Freitag, Oliver Bott & Fabian Schlotterbeck (eds.), *Two perspectives on V2: The invited talks of the DGfS 2016 workshop "V2 in grammar and processing: Its causes and its consequences"*, 27–49. Konstanz: University of Konstanz.

Bader, Markus & Michael Meng. 2018. The misinterpretation of noncanonical sentences revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44(8). 1286–1311. DOI: 10.1037/xlm0000519.

Bader, Markus & Yvonne Portele. 2019. The interpretation of German personal and d-pronouns. *Zeitschrift für Sprachwissenschaft* 38(2). 155–190. DOI: 10.1515/zfs-2019-2002.

Bader, Markus & Yvonne Portele. 2021. Discourse and form constraints on licensing object-first sentences in German. *Languages* 6(2). 82. DOI: 10.3390/languages6020082.

Bader, Markus & Yvonne Portele. 2023. *Ranking animacy and discourse status as determinants of pronoun use.* Talk presented at the Conference on Architectures and Mechanisms for Language Processing (AMLaP), Basque Center on Cognition, Brain and Language, San Sebastián.

Bittner, Dagmar. 2019. Implicit Causality in younger and older adults. *Linguistics Vanguard* 5(s2). DOI: 10.1515/lingvan-2018-0023.

Bock, J. Kathryn & Richard K. Warren. 1985. Conceptual accessability and syntactic structure in sentence formulation. *Cognition* 21. 47–67. DOI: 10.1016/0010-0277(85)90023-X.

Bott, Oliver & Torgrim Solstad. 2014. From verbs to discourse: A novel account of implicit causality. In Barbara Hemforth, Barbara Mertins & Cathrine Fabricius-Hansen (eds.), *Meaning and understanding across languages*, 213–251. Chicago: Springer. DOI: 10.1007/978-3-319-05675-3_9.

Collins, Michael Xavier. 2014. Information density and dependency length as complementary cognitive models. *Journal of psycholinguistic research* 43(5). 651–681. DOI: 10.1007/s10936-013-9273-3.

Crocker, Matthew W., Vera Demberg & Elke Teich. 2016. Information density and linguistic encoding (IDeaL). *KI – Künstliche Intelligenz* 30(1). 77–81. DOI: 10.1007/s13218-015-0391-y.

Fenk-Oczlon, Gertraud. 1989. Word frequency and word order in freezes. *Linguistics* 27. 517–556. DOI: 10.1515/ling.1989.27.3.517.

Ferreira, Fernanda. 1994. Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language* 33. 715–736. DOI: 10.1006/jmla.1994.1034.

Ferreira, V. S. & Gary S. Dell. 2000. Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* 40. 296–340. DOI: 10.1006/cogp.1999.0730.

Frank, Michael C. & Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336(6084). 998–998. DOI: 10.1126/science.1218633.

Fukumura, Kumiko & Roger P. G. van Gompel. 2010. Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language* 62(1). 52–66. DOI: 10.1016/j.jml.2009.09.001.

Garvey, Catherine & Alfonso Caramazza. 1974. Implicit causality in verbs. *Linguistic Inquiry* 5(3). 459–464.

Genzel, Dmitriy & Eugene Charniak. 2002. Entropy rate constancy in text. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 199–206. DOI: 10.3115/1073083.1073117.

Givón, Talmy. 1983. Topic continuity in discourse: An introduction. In Talmy Givón (ed.), *Topic continuity in discourse: A quantitative cross-language study*, 1–42. Amsterdam: John Benjamins.

Givón, Talmy. 1984. *Syntax. A functional–typological introduction. (Volume I)*. Amstersam: John Benjamins. DOI: 10.1075/z.syn1.

Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21. 203–225. DOI: 10.21236/ada324949.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69. 274–307. DOI: 10.2307/416535.

Hoberg, Ursula. 1981. *Die Wortstellung in der geschriebenen deutschen Gegenwartssprache*. München: Hueber.

Holler, Anke & Katja Suckow. 2016. How clausal linking affects noun phrase salience in pronoun resolution. In Anke Holler & Katja Suckow (eds.), *Empirical perspectives on anaphora resolution*, 61–85. Berlin: de Gruyter. DOI: 10.1515/9783110464108-005.

Jaeger, T. Florian. 2006. *Redundancy and syntactic reduction in spontaneous speech*. Stanford: Stanford University. (Doctoral dissertation).

Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61(1). 23–62. DOI: 10.1016/j.cogpsych.2010.02.002.

Jain, Ayush, Vishal Singh, Sidharth Ranjan, Rajakrishnan Rajkumar & Sumeet Agarwal. 2018. Uniform information density effects on syntactic choice in Hindi. In *Proceedings of the Workshop on Linguistic Complexity and Natural Language Processing*, 38–48. Santa Fe, New-Mexiko: Association for Computational Linguistics. https://aclanthology.org/W18-4605.

Kaiser, Elsi & Emily Fedele. 2019. Reference resolution: A psycholinguistic perspective. In Jeanette Gundel & Barbara Abbott (eds.), *The Oxford handbook of reference*, 309–336. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199687305.013.15.

Kaiser, Elsi & John C. Trueswell. 2008. Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes* 23(5). 709–748. DOI: 10.1080/01690960701771220.

Kehler, Andrew, Laura Kertz, Hannah Rohde & Jeffrey L. Elman. 2008. Coherence and coreference revisited. *Journal of Semantics* 25(1). 1–44. DOI: 10.1093/jos/ffm018.

Kehler, Andrew & Hannah Rohde. 2013. A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics* 39(1-2). 1–37. DOI: 10.1515/tl-2013-0001.

Lambrecht, Knud. 1996. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511620607.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177. DOI: 10.1016/j.cognition.2007.05.006.

Levy, Roger. 2013. Memory and surprisal in human sentence comprehension. In Roger G. van Gompel (ed.), *Sentence processing*, 78–114. New York: Psychology Press.

Levy, Roger & T. Florian Jaeger. 2006. Speakers optimize information density through syntactic reduction. In *Proceedings of the 19th International Conference on Neural Information Processing Systems*, 849–856.

MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4:226. DOI: 10.3389/fpsyg.2013.00226.

Maurits, Luke, Dan Navarro & Amy Perfors. 2010. Why are some word orders more common than others? A uniform information density account. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel & A. Culotta (eds.), *Advances in Neural Information Processing Systems*, vol. 23. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2010/file/0c74b7f78409a4022a2c4c5a5ca3ee19-Paper.pdf.

McDonald, Janet L., Kathryn Bock & Michael H. Kelly. 1993. Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology* 25. 188–230. DOI: 10.1006/cogp.1993.1005.

Orita, Naho, Eliana Vornov & Naomi H. Feldman. 2021. Informativity, topicality, and speech cost: Comparing models of speakers' choices of referring expressions. *Discourse Processes* 58(8). 743–765. DOI: 10.1080/0163853X.2021.1878449.

Portele, Yvonne & Markus Bader. 2020. Coherence and the interpretation of personal and demonstrative pronouns in German. In Anke Holler, Katja Suckow & Israel de la Fuente (eds.), *Information structuring in discourse*, 24–55. Leiden: Brill. DOI: 10.1163/9789004436725_003.

Prat-Sala, Mercé & Holly P. Branigan. 2000. Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language* 42. 168–182. DOI: 10.1006/jmla.1999.2668.

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosphica* 27(1). 53–94. DOI: 10.21825/philosophica.82606.

Rohde, Hannah & Andrew Kehler. 2014. Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience* 29(8). 912–927. DOI: 10.1080/01690965.2013.854918.

Rosa, Elise C. & Jennifer E. Arnold. 2017. Predictability affects production: Thematic roles can affect reference form selection. *Journal of Memory and Language* 94. 43–60. DOI: 10.1016/j.jml.2016.07.007.

Rubio-Fernández, Paula, Francis Mollica & Julian Jara-Ettinger. 2021. Speakers and listeners exploit word order for communicative efficiency: A cross-linguistic investigation. *Journal of Experimental Psychology: General* 150(3). 583. DOI: 10.1037/xge0000963.

Stevenson, Rosemary J., Rosalind A. Crawley & David Kleinman. 1994. Thematic roles, focus and the representation of events. *Language and Cognitive Processes* 9(4). 519–548. DOI: 10.1080/01690969408402130.

Tily, Harry & Steven Piantadosi. 2009. Refer efficiently: Use less informative expressions for more predictable meanings. In *Proceedings of the workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference*.

Weatherford, Kathryn C. & Jennifer E. Arnold. 2021. Semantic predictability of implicit causality can affect referential form choice. *Cognition* 214. 104759. DOI: 10.1016/j.cognition.2021.104759.

Chapter 5

The role of information in modeling German intensifiers

J. Nathanael Philipp[a,b], Michael Richter[b], Tatjana Scheffler[c] & Roeland van Hout[d]

[a]Sächsische Akademie der Wissenschaften zu Leipzig [b]University of Leipzig
[c]Ruhr University Bochum [d]Radboud University

In this study, context-free and context-dependent information measures are applied to a new corpus of tweets and blog posts. The aim is to account for the expressive meaning and characterize the variability of available intensifying items. It comes to light that context-free and context-dependent information measures are highly correlated and account for the distribution of intensifiers in the data, giving credence to the notion that intensifiers form a common word class, even across syntactic and semantic differences.

Both information measures show that stacked intensifiers tend to be ordered from least to most expressive within a phrase, i.e., the information tends to increase. We explain this fact using the Uniform Information Density Hypothesis: The first, less expressive intensifier is used to introduce the phrase, ease the reader's processing load, and smooth the information flow.

1 Introduction

Intensifiers such as as *very* and *so* add little to the referential content of a sentence. Instead, these expressions serve to increase the expressive value of an utterance by indicating that some property applies to a higher degree (1), or even merely by signaling a heightened emotional state (2):

J. Nathanael Philipp, Michael Richter, Tatjana Scheffler & Roeland van Hout. 2024. The role of information in modeling German intensifiers. In Robin Lemke, Lisa Schäfer & Ingo Reich (eds.), *Information structure and information theory*, 117–145. Berlin: Language Science Press. DOI: 10.5281/zenodo.13383791

(1) *Der Sommer war so schön!*[1]
The summer was so pretty
'The summer was so wonderful!'

(2) *Ich freu mich so! :-)*
I be-happy me so :-)
'I am so happy! :-)'

Intensifying expressions are subject to immense variation in a language and can often be added to an utterance in informal discourse without adding much to its core denotation. Their use and choice are therefore governed by expressivity as well as information structure and general communicative needs. We argue that inserting an intensifier in an utterance can ease processing by evening out the information density in a sentence to avoid large heights or troughs.

In previous empirical studies (Richter & Van Hout 2020, Scheffler et al. 2023b), expressivity and stacking of intensifiers in Dutch and German Twitter data are modeled and operationalized by information theory (Shannon 1948) and information measures derived from it. Richter & Van Hout (2020) and Scheffler et al. (2023b) distinguish a paradigmatic information value that represents expressivity (and represents the strength of an intensifier) and a syntagmatic information value representing transitional information (i.e., its ability to combine with various targets). The authors observe that in stacks of intensifiers, expressivity increases from left to right, leading them to conjecture that blander, less expressive intensifiers precede the more expressive ones and thus smooth information flow in the sentence. In addition, they argue that intensifiers may lose their expressive value over time, while new and expressive intensifiers carrying high surprisal come into vogue, and that the expressivity of intensifiers can be boosted by capitalization, lengthening (*soooooo*), and repetitions (*very, very nice*). In addition, it has been shown that the choice of a longer (and thus typically more unusual) intensifier such as *atemberaubend* ('breathtakingly') over a short intensifier like *sehr* ('very') can further increase the expressivity of an utterance (Bennett & Goodman 2018).

In this paper, we follow the previous line of research by Richter & Van Hout (2020) and Scheffler et al. (2023b) and represent the expressivity of intensifiers through the amount of information encoded in an expression (Shannon 1948). That is, our model operationalizes the measurement of expressivity. We also draw on the work of Bennett & Goodman (2018) mentioned above, who related costs, i.e. information, to the strength and expressivity of intensifiers.

[1] All German examples in this paper are attested in our corpus, except where marked otherwise.

The idea of determining expressivity and strength of intensification is already found in the work of Zadeh (1972) couched in a fuzzy set theoretic model, and in work by Potts (2007), who proposes a separate expressive dimension of expressions. None of these works, however, include an empirically based determination of the strength or expressivity of particular expressions. Our proposed model does not give a definition of expressivity itself, but rather provides a methodology for precisely measuring the amount of expressivity. We use semantic surprisal and Shannon information content (abbreviated as IC in the remainder of this paper) as lexical features of an intensifier *w*. Surprisal and IC are negative log-probabilities of *w* that indicate the amount of information in bits. Surprisal is calculated in the sentential context of *w*, i.e., within the sentence in which *w* occurs and within its extra-sentential context (Levy 2008), that is, contexts in preceding and following sentences. IC, in contrast, is context-free, i.e., it is simply derived from the relative frequency of *w*. Context plays a key role in the calculation of surprisal, and in the present study, we define semantic surprisal in the context of the topics in the environment of *w*. This interpretation of the semantic context of a word is introduced by Kölbl et al. (2020, 2021), and previously used by Philipp et al. (2022, 2023). We will address our definition of surprisal and the topic contexts it is based on in more detail in Section 3 below.

We apply these information theoretic analyses to a new dataset of manually annotated intensifiers in two media, tweets and blog posts, in order to confirm the predictions established in the previous research. The data consists of texts from blogs and tweets from the same set of 44 authors (Scheffler et al. 2023a). Including blog post texts enables us to extend the empirical picture of intensification by observing their behavior in a new medium. Our results show that while the intensifier word class indeed shows high variability, as indicated by previous studies, the class conforms to the information theoretic predictions established in prior work. These results hold even though the data we report on here is no longer restricted to intensification of predicative adjectives. We compare the two media tweets and blogs to characterize the media related variability of intensification in our German corpus. We show that both context-free and context-dependent (semantic) information measures exhibit high correlation and can account for the distribution of intensifiers.

Finally, Scheffler et al. (2023b) argue that the order of stacked intensifiers (when multiple intensifiers follow each other in the same phrase) is determined by Uniform Information Density (Fenk & Fenk 1980, Aylett & Turk 2004, Levy & Jaeger 2007, Jaeger 2010): They observe that intensifiers tend to occur from least to most expressive and conjecture that the less expressive first intensifier serves to introduce the phrase and "bridge" the information flow from copula

verb to the expressive intensifier and adjective. In this paper, we test this conjecture on our entire corpus of sentences with intensifiers. Since we now employ contextual information values which we can compute for each word in a text, we can systematically compare sentences with stacked intensifiers with the communicatively equivalent alternatives (i) containing only the last, most expressive intensifier, or (ii) containing both intensifiers in the opposite order. We compute information profiles for both alternatives and compare them to the actually attested variants. Our results show that the attested variants exhibit a smoother information contour with a slightly higher uniformity of the wordwise information density. This supports the Uniform Information Density hypothesis, which states that smoother information flow is preferred keeping other factors equal. This analysis may serve to shed light on the puzzle why writers bother to add further, redundant intensifiers beyond the most expressive one: Their function is to introduce the phrase, alert the reader to the word class which follows, and smooth the information flow.

In the following, we first discuss the word class of intensifiers and why they are of particular interest for information theoretic approaches to language. We then discuss the empirical base of our annotation study for intensifiers in two written social media, blogs and tweets, before providing more details on the information measures we use for our analysis. Finally, we present our quantitative results.

2 Corpus study: Intensification in blogs and tweets

2.1 Intensifiers as a word class

Gradable properties such as the body size expressed by the adjective *fat* can be strengthened or toned down by certain fixed words or phrases. These words or phrases are called intensifiers because they increase (3) or reduce (4) the intensity to which the gradable property applies. In common usage, non-gradable properties are also often the target of intensification (5).

(3) This seal is extremely fat. (constructed)

(4) This seal is pretty fat. (constructed)

(5) This seal is completely round. (constructed)

Finally, it is well-known that although adjectives are the most common targets of intensification, other words such as verbs (2) or nouns (6) can also be intensified (see Bolinger 1972), as in the following two examples from our social media dataset (see Section 2.2):

(6) a. *Was ganz Neues!*
Some completely new-thing
'Something completely new!'

b. *Und warum zur Hölle habe ich damit so ein Problem, mich von den Kindern lösen zu können?*
And why in hell have I with-that so a problem, me from the kids separate to be-able-to
'And why in hell do I have such a problem with separating from the kids?'

In this study, we analyze German intensifiers using measures from information theory. We assume that all intensifiers share a common semantic core, the basic notion of intensification, though they may differ in other connotations (e.g., level of formality, strength, or semantic domain restrictions). We focus on a semantically relatively uniform subset of the more broader set of "intensity particles" (Breindl 2007) sometimes discussed, namely only those that increase the extent to which a property applies, as exhibited in all previous examples (1–6). This notion includes both so-called maximizers such as *total* ('totally'), which indicate only the end point of a scale, as well as boosters such as *sehr* ('very'), which restrict a property to a higher section of a scale (Bolinger 1972).

This means that we expressly exclude downtoners, which lower the grade to which a property applies rather than increasing it (e.g., *ein bisschen* 'a little', *etwas* 'somewhat'), from further consideration. Accordingly, we do not consider intensifiers that can only occur with negated properties (e.g., *gar* 'at all'), since the negation implies a lowered intensity of the modified property. We further exclude words which look like intensifiers but which should better be analyzed as focus adverbs (*einfach* 'simply', *echt*, *wirklich* 'really'; see Beltrama 2022, Scheffler et al. 2023b). In the following, we reserve the term *intensifier* for our more narrow definition and call the broad category of gradability modifiers (including downtoners as well as some other expressions) *grade indicators*.

Given that all remaining intensifiers share a common semantic contribution, it remains a mystery why a language should provide so many different variants (for German, see e.g. Claudi 2006, Stratton 2020). In addition, more intensifiers within the same phrase do not add much semantic value. Prior work proposes that longer intensification phrases (i.e., longer intensifiers or additional intensifiers within one phrase) increase the strength and thus the expressivity of intensification (Fortin 2011, Bennett & Goodman 2018, Fortin & Rainer 2022). Note though that this only explains the presence of intensifier stacks, but not the

clear preferences for intensifier ordering that have been observed. In addition, establishing that the addition of an intensifier in an already intensified phrase increases expressivity is merely restating the facts, if no mechanism can be proposed why additional intensifiers are more expressive.

In the following, we propose, based on our previous work (Richter & Van Hout 2020, Scheffler et al. 2023b), that both the variance between intensifiers as well as their stacking behavior can be modeled using information theoretic measures, which explain the different contributions of different intensifying expressions. We argue that the class of intensifiers is relatively uniform with respect to their core semantic contribution (increasing the degree of a property), but that they differ in expressive value. While the notion of amount of information which a linguistic item contributes has been characterized as that item's core propositional semantics in the past (and operationalized rather simply by (contextual) frequency), we extend this notion to include the non-propositional, expressive content of the item as well, for example the strength of intensification or involvement of the speaker when using an intensifier. In our view, it is this non-propositional, expressive difference between intensifiers that can be seen in their differing information values.

2.2 Dataset of blogs and tweets

Intensifiers are a highly variable word class (Tagliamonte 2008, Stratton 2020), which is undergoing constant innovation and with which speakers/authors adapt flexibly to different registers and media (Scheffler et al. 2022). They are typical of speech, but also occur frequently in social media and even increase in frequency in newspaper corpora (Schmidt 2022). In this research, we use a novel multi-medium dataset to confirm previous results and advance the state of the art on the role of information measures in modeling intensifiers. Previous research has rarely addressed the way individuals change their use of intensifiers when switching media or text type. The data we study here consists of a corpus of tweets and blog posts from 44 parenting bloggers, collected in 2017. The corpus consists of 81,440 tweets (~1.2m tokens) and 468 blog posts (~360k tokens) in total, and is available for academic research by request.[2] Each author (N = 44) is represented with about 5–10 blog posts and up to 3200 tweets (for more details, see Scheffler et al. 2023a). Any differences between the subcorpora can be attributed to special properties of the two media formats, since the authors and topics stay the same.

[2] http://tiny.cc/twiblocop

The corpus was automatically split into sentences and tokenized, and manually pseudonymized. The tweet subcorpus consists of 137,914 sentences, while the blog posts contain 24,981 sentences in total. In all texts, all grade indicators have been manually marked by student research assistants with the help of the software WebAnno (Eckart de Castilho et al. 2016). To facilitate annotation, the set of frequent intensifiers identified by Scheffler et al. (2023b) was automatically preselected. Annotators were asked to both add additional intensifiers not included in the previously known list, as well as disambiguate all occurrences of the preannotated words regarding their status as an intensifier. The annotation was based on a short guideline specifying examples and counterexamples. Contrary to some previous work, the annotation did not restrict the class of grade indicators (i.e., downtoners were included along with boosters; and intensity indicators modifying verbs and nouns were included along with adjective intensifiers). We aimed directly for the semantic/pragmatic function of intensity modification of a gradable property or relation. The annotation thus covered a broader range of items than are analyzed in this paper. Difficult cases were discussed with the students and the annotators did not raise major problems in identifying the grade indicators. To prevent any misunderstanding, we call this class of words grade indicators in the remainder of this paper.

Three of the authors individually evaluated all grade indicators annotated in the corpus manually in order to identify true intensifiers in the more narrow sense which we defined above, i.e., expressions which indicate that a property is strengthened or increased. Of the three evaluations, between a more conservative and a more lenient evaluation, we ultimately selected the intermediate one as our data for the following analysis.

3 Modeling the information structure of intensification

We use three information theoretic measures to model the intensifiers in our data: first, local information content (IC_{local}), which captures the paradigmatic information of a word. This measure was proposed by Scheffler et al. (2023b) to capture the variability of adjectival intensifiers and, actually, is the common Shannon Information, i.e., the logarithm of the unigram frequency of a linguistic unit w. In our case, the frequency is calculated (i) separately for the blog posts and for the tweets corpus and (ii) for the combined blog posts and tweet corpus. The definition is given in (7):

(7) $$IC_{local}(w) = -\log_2 P(w)$$

Second, we introduce here a semantic notion of information, that is, semantic surprisal based on the Topic Context Model (TCM) which incorporates the document context of a word. Third, we compare sentence variants using the average information change per word in a sentence, $\text{UID}_{\text{wordwise}}$.

3.1 Surprisal

Surprisal was introduced by Tribus (1961) in an engineering science context, and later adapted to communication theory and psycholinguistics (Hale 2001, Levy 2008). The surprisal of a word is its contextualized information and quantifies how unexpected it is, or among how many alternatives it was chosen. The occurrence of a linguistic unit causes an amount of surprisal for the language processor which is low when the linguistic unit is expected and high when it is unexpected. The amount of surprisal is proportional to the effort that is necessary to process it (Hale 2001, Levy 2008). One can say that surprisal expresses the discrepancy between what a language processor expects in a sentence and what actually occurs. For example, given the string *the old man*, a language processor might expect a verb as the next word. The surprisal is high if a determiner occurs, as in the garden path sentence *the old man the boat*.

Hale (2001) and Hale et al. (2015) define the surprisal of a word formally as its negative log-probability, which is equivalent to the definition of Shannon information IC. However, the calculation of surprisal requires conditional probabilities while the calculation of IC does not. Here, the probabilities $P(w_1, w_2 ... w_n)$ are calculated using the chain rule. Consider the example sentence from above *the old man the boat*. The second determiner is extremely unexpected in this context, and its surprisal is thus high.

Equation 8 gives the definition of surprisal as a conditional probability of a word w_i given the prior context $w_1 \dots w_{i-1}$ (Levy 2008), where $w_1 \dots w_{i-1}$ represent co-occurrences of any kind of the target word w_i within a sentence, and CONTEXT represents extra-sentential context of any kind, e.g., lexical, syntactic or semantic structures.

$$\text{surprisal}(w_i) = -\log_2(P(w_i \mid w_1, \dots, w_{i-1}, \text{CONTEXT}) \tag{8}$$

3.2 Topic context model

The Topic Context Model (TCM)[3] (Kölbl et al. 2020, 2021, Philipp et al. 2022, 2023) outputs the semantic surprisal of a word given the topics in its environment. Simply put, in this operationalization, semantic surprisal expresses how surprising

[3] https://github.com/jnphilipp/tcm

a word is to the language processor in the context of the topics in the words' environment. For example, the word *chocolate* is likely to have a low probability in the context of topics such as *information* or *algorithms*, and would probably cause a high degree of semantic surprisal.

TCM surprisal is thus specific to its environment, which can be the entire corpus, or documents within the corpus, paragraphs and even single sentences (in the remainder of the article, surprisal as information from the Topic Context Model is symbolized by TCM, and we use documents (blog posts, tweets) as the environments).

Like generative probabilistic topic models, TCM is based on the assumption that contexts/discourses are thematically structured, that documents are generated by topics, and that topics are characterized by words. In the present study, TCM employs Latent Dirichlet Allocation (LDA, Blei et al. 2003) for topic detection. LDA treats documents as probability distributions of topics and topics as probability distributions of words.

The definition of the average semantic surprisal (of a word w in a document d) from topics is given in Equation 9. $P(w \mid t_i)$ is the probability of a word w given a topic t_i, see Equation 10 where $c_d(w_d)$ is the frequency of a word w in a document d, $|d|$ is the total number of words in a document d, and WT is the normalized word topic distribution of the LDA.

$$\overline{\text{surprisal}}(w_d) = -\frac{1}{n}\sum_{i=1}^{n}\log_2 P(w \mid t_i) \tag{9}$$

$$P(w_d \mid t_i) = \frac{c_d(w_d)}{|d|} WT_{w_d,t_i} P(t_i \mid d) \tag{10}$$

For example, suppose a word w occurs in n documents in a corpus. LDA yields a topic distribution t for every document d. This distribution takes the form of a k-dimensional vector of probabilities, which must sum to 1. TCM calculates the surprisal of w by applying the formulas 9 and 10 (Philipp et al. 2022, 2023).

Suppose the word *mega* occurs two times in a document d with 42 words in total, further suppose we train a LDA with three topics. The three probabilities for $P(t_i \mid d)$ are 0.732, 0.183 and 0.085 and the three values for WT_{w_d,t_i} are 0.261, 0.101 and 0.096. This results in three probabilities for $P(w_d \mid t_i)$, that is, 0.00910, 0.00088 and 0.00039, respectively. Utilising Formula 9, we derive from these probabilities the average surprisal value of 9.41812.

In total, we trained three different LDAs and therefore have three different TCM structures. The first LDA was trained only on the blog posts, the second

one only on the tweets, and the last one on the complete corpus. All LDAs were trained with 32 topics.[4] The training corpora contained 30,111 words of blog data, 26,656 words from Twitter and 45,754 from the complete corpus. All three TCM values behave similarly in our analyses. However, since each tweet constitutes its own document, and the tweets are generally very short, we expect the tweet-based models to be less reliable than the models based on the blog data, where each document consists of an entire blog post.

3.3 Wordwise uniform information density

Following Collins (2014), we determine local Uniform Information[5] as a measure of the average squared information change per linguistic unit in sequences of linguistic units. For this measure, Scheffler et al. (2023b) use the term $\text{UID}_{\text{wordwise}}$. Its definition is given in (11), where *id* is the information value of a unit, n is the number of units in a sequence (for instance in a sentence or in a stack on intensifiers), and the index $i = 2$ indicates the second unit in the sequence (counting from left to right).

$$\text{UID}_{\text{wordwise}} = -\frac{1}{n}\sum_{i=1}^{n}(id_i - id_{i-1})^2 \tag{11}$$

Note that $\text{UID}_{\text{wordwise}}$ is negative by definition, and therefore a $\text{UID}_{\text{wordwise}}$ value close to zero indicates a high uniformity of the information density distribution, that is, smaller information differences from unit to unit. Information density is an important principle in linguistic communication, since information peaks and troughs must not be too extreme, so as not to make it too difficult for the recipient of a message to process it (Fenk & Fenk 1980, Aylett & Turk 2004, Levy & Jaeger 2007, Jaeger 2010).

4 Results

4.1 Grade indicators and intensifiers

We found 3446 grade indicators in the blog posts and 2034 in the tweets. Given the token size of these two subcorpora (blogs ~360k, tweets ~1.2m), grade indicators turn out to be more frequent in the blogs. This result is even more pronounced when looking at the sentence level: The blog subcorpus contains 24,981

[4] After some experiments, this appeared as a good number to us.

[5] https://github.com/jnphilipp/uid

sentences, yielding an average of 0.139 grade indicators per sentence. In contrast, the tweets, with 137,914 utterances, contain about ten times fewer grade indicators, only 0.015 per sentence. Sentence segmentation was carried out automatically in prior work (Scheffler et al. 2023a). Our database distinguishes single grade indicator tokens and stackings (sequences of more than one grade indicator). After lemmatization (combining variants), we obtained the top 10, which are presented in Table 1 (both single indicators and stacks are included here).

Table 1: Frequencies of the 10 most frequent grade indicators, per medium and in the total dataset

Lemma	Translation	N blogs	N tweets	N total
so	'so'	641	551	1192
ganz	'completely'	486	263	749
sehr	'very'	424	259	683
gar	'at all'	200	134	334
zu	'zu'	180	68	248
wirklich	'really'	159	61	220
echt	'really'	30	79	109
total	'totally'	55	49	104
ziemlich	'pretty'	80	20	100
etwas	'somewhat'	76	22	98

The list in Table 1 shows grade indicators that are not an intensifier in the strict sense, like *zu* ('too'), *ziemlich* ('pretty'), and *etwas* ('somewhat'). It also shows three highly frequent intensifiers, *so*, *ganz*, *sehr*, in both media. Compared to the extensive tweet database studied by Scheffler et al. (2023b), the high frequency of *ganz* ('completely') in the blogs is remarkable, but the most important observation is that a comparable set of frequent intensifiers is present in this new dataset, as well. In the following analyses, we exclude all grade indicators that are not intensifiers according to our definition, including downtoners (*ziemlich* 'pretty', *etwas* 'somewhat'), and the excessivity marker *zu* ('too').

New intensifiers that were not present in the extensive tweet database by Scheffler et al. (2023b) all occurred with a frequency of only 1, e.g., *affengeil* ('(monkey) awesome'), *bombenfest* ('bomb proof'), *entsetzlich* ('terribly'), *gewaltig* ('hugely'), *höllisch* ('hellish'), *riesen* ('giant'). The occurrence of new intensifiers in fact demonstrates the openness of the word class of intensifiers and its innova-

tive power, as pointed out earlier in comparing different inventories of German intensifiers (Scheffler et al. 2023b).

Our database of blogs and tweets contains single occurrences and stacks. We start with analyzing the single occurrences. Next, we present the results on the stacks.

4.2 Single intensifiers

We selected only those intensifiers that were also present in the large Twitter database studied previously (Scheffler et al. 2023b) for further analysis. In that study, 89,358 intensified predicative phrases of the form pronoun + *is/was* + … + adjective (e.g., *Frankfurt ist so arsch weit* 'Frankfurt is so damn far') from a corpus of over 6 million German tweets were extracted. These intensified phrases contained 124 different frequent intensifiers, 68 of which also occur in our new blog and tweet corpus. They include *echt* and *wirklich* ('really'), which can express an intensifying function, but are in fact focus adverbs. We proceed with these 68 items as they qualify as validated, frequent intensifiers in German.

For all 68 intensifiers, two information measures were computed, one for frequency (IC_{local}, see above) and one for the transitional probability between the intensifiers and the following adjective (IC_{trans}). In the extensive tweet database (Scheffler et al. 2023b) the anticorrelation between these two information measures was extremely high, −0.916. This still applies to the 68 shared intensifiers in our new database. Their anticorrelation is even a bit higher at −0.944, indicating that there is no selective bias in our new corpus compared to the previous large Twitter database. IC_{local} in the previously studied large Twitter database is correlated with the log frequencies in both blogs and tweets of the new corpus (respectively, $r = 0.735$, $r = 0.838$). Interestingly, the correlation with the Twitter part of our new corpus is significantly higher than with blogs ($t = 3.06$, $p < 0.001$), as is to be expected since the previous database also consisted of tweets. The register distinction between blogs and tweets is confirmed by data on capitalization and lengthening in the tokens of our 68 remaining intensifiers (respectively only 1 and 3 capitalizations, and 14 (0.6%) and 50 lengthenings (3.1%) in blogs vs. tweets); these numbers are much lower than in the extensive tweet database collected earlier. In particular, *so* was the most popular intensifier for lengthening, similar to the findings in Scheffler et al. (2023b).

In this paper, we, in addition to IC, introduced the contextualized information measure TCM. This measure computes the information value of a word depending on its context (in our case, a document). These two information values are

related to each other, insofar as the simple IC value captures some of the information content of a word, while TCM modulates this value to a specific context. The scattergrams of these two measures are given in Figure 1 and Figure 2. In the blog posts, the TCM value of single intensifiers correlates linearly with the IC values (Figure 1), whereas the relation is more fuzzy in tweets (Figure 2) in the higher regions of IC and TCM.

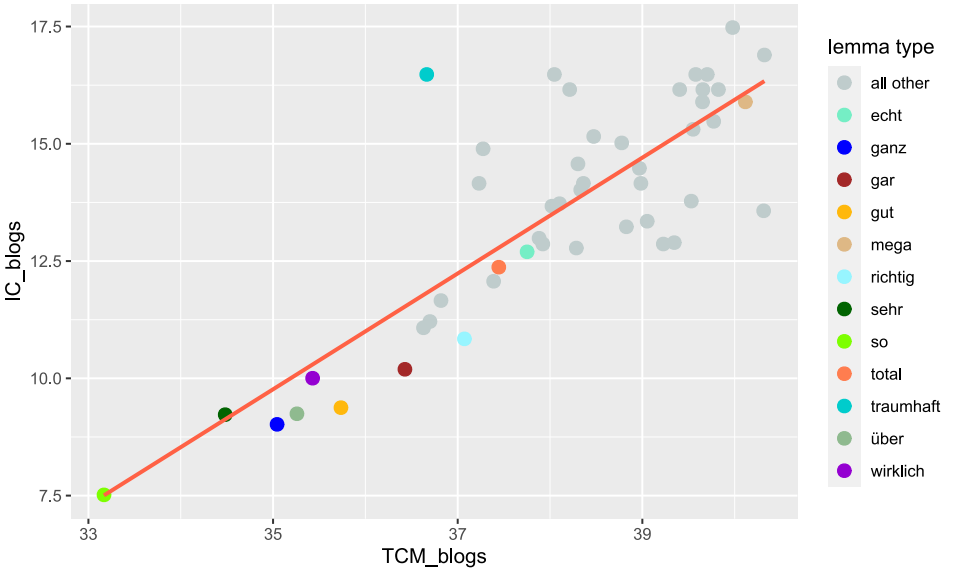


Figure 1: Information (TCM by IC) of single intensifiers in blogs

The TCM values are considerably higher in general for blogs than for tweets. This is due to the different context sizes. While the context in the tweets data set is restricted to only a single tweet in each case, the context for the blogs consists of several sentences. This means that the topic model on which TCM is based has more information available for training and can thus be more specific.

Despite these differences in TCM values, the two patterns in Figure 1 and Figure 2 resemble each other. The same frequent intensifiers can be found in their left parts. The patterns get more diffuse moving to the less frequent intensifiers in the second position in the stack. We also included *wirklich* ('really') that we considered a focus particle. This status as a focus particle seems to be supported by the different relative position it has in the two scattergrams, although they fit the linear pattern.

The scatter graphs indicate that the two measures are less closely related in the tweets subcorpus. It might be that the more diffuse pattern in tweets is (partly)

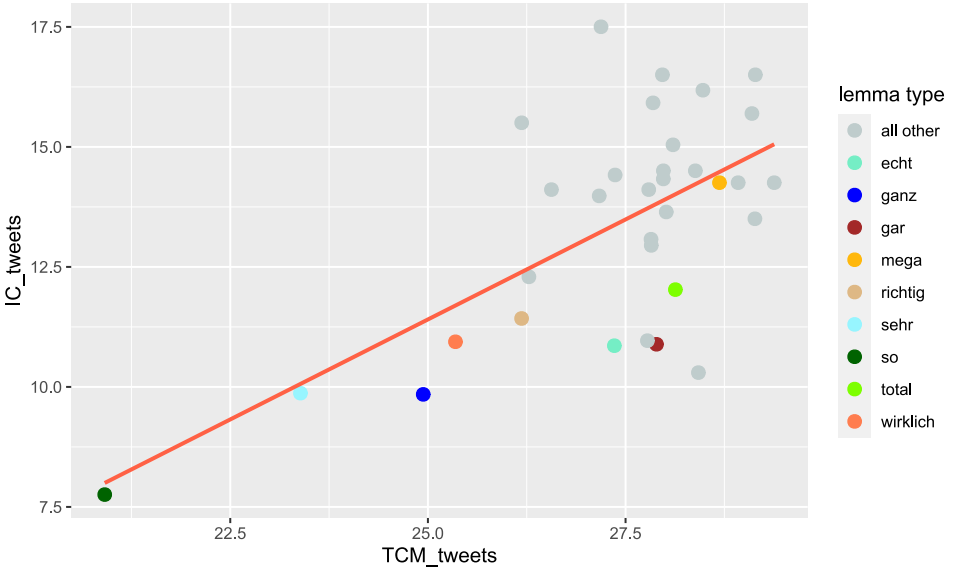


Figure 2: Information (IC by TCM) of single intensifiers in tweets

related to the size of the database, a smaller database means more noise, in particular since the number of intensifiers is also lower in the tweets. To investigate the relation between IC and TCM more closely, we not only need more data, we also have to estimate the additional value of the TCM index more precisely. We start doing so in the next subsection by looking at the stacks.

4.3 Information values in intensifier stacks

Both IC and TCM capture the information, and in our understanding, the expressiveness, of an intensifier. The first question we asked ourselves was whether the two types of information values TCM and IC correlate in the intensifiers occurring in the stacks. Since stacks with three intensifiers are extremely rare in our data sets (only three occurrences), we focused solely on stacks with two intensifiers.

We determined the correlation between the two types of information values of the first intensifier in a stack (TCM1–IC1) and the values of the second intensifier in the stack (TCM2–IC2). Table 2 shows that the correlations between the information values for both intensifier positions are positive; the correlations (Pearson) in the Twitter data set are higher than in the blogs.

In addition, the correlation is much lower across the entire data set, indicating a clear register difference between the blogs and tweets. This is due to higher

Table 2: Pearson correlation between TCM and IC information values

	Pearson correlation	
Data	TCM1–IC1	TMC2–IC2
all	0.46	0.52
blogs	0.66	0.72
tweets	0.83	0.71

TCM values in the blog subcorpus, since the TCM values depend on the number of different contexts and instances. This becomes clear in the scatter plots in Figures 3 and 4, which show the data points of TCM compared to IC for the first intensifier in blogs (Figure 3) and the data points of TCM and IC for the second intensifier (Figure 4).

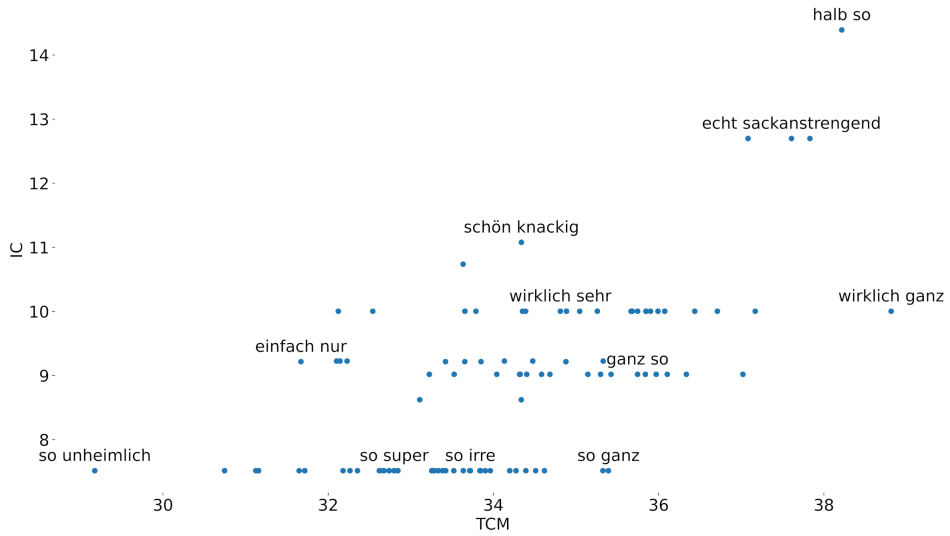


Figure 3: Scatterplot of the blog data set, using the *IC*/*TCM* values from the first intensifier in the stack

Figures 3 and 4 clearly demonstrate the differences in the calculation of the actual IC and TCM values. Figure 3 shows the IC and TCM values for the first intensifier in each stack in the blog subcorpus, Figure 4 shows the values for the second intensifier. All occurrences of *so* have the same IC value, seen in the bottom row of dots in both figures, as this value depends only on the overall

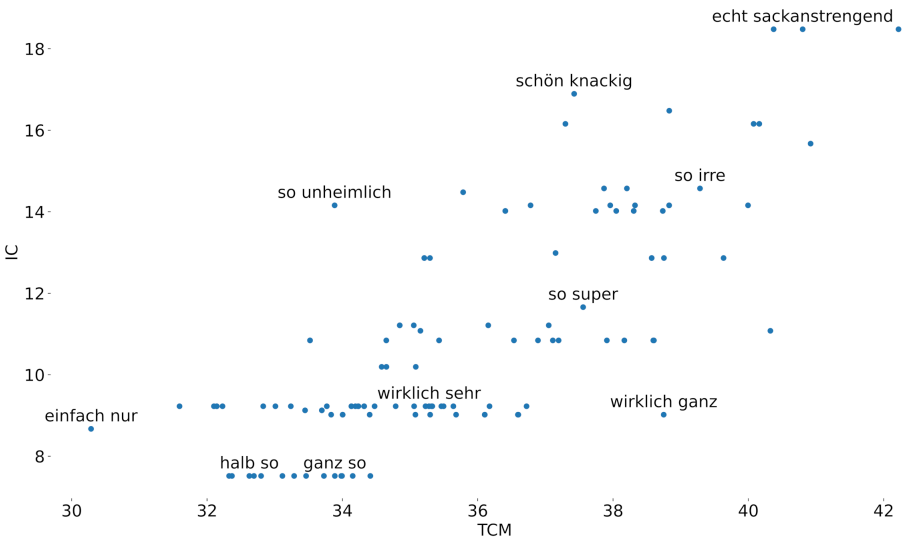


Figure 4: Scatter plot of the blog data set, using the IC/TCM values from the second intensifier in the stack

frequency of *so*, independent of its context. In contrast, the TCM surprisal values differ greatly between the instances, see Figure 3, with *so unheimlich* having the lowest and *so ganz* having the highest value. This is due to the fact that the topic modeling part of the TCM allows for a more fine-grained approach, enabling us to calculate a surprisal value that is text specific. The different TCM values for the same token indicate that it contributes different amounts of information to the context. It is thus not surprising that the correlation between the two information values IC and TCM is not perfect, as shown in Table 2, and that the correlation is lower for blogs, which have a greater context, when calculating the surprisal values.

Figures 3 and 4 also make it possible to compare the spread of TCM for an intensifier like *so* as the first in a stack and *so* as the second intensifier in a stack. This shows that the spread, when *so* occurs as the first intensifier in a stack, is much greater, compared to when it is only in the second position. One reason for this is that the word, being a low information intensifier, can be the start of many different intensifying phrases. This bridging function of the first intensifier towards another intensifier is discussed in Section 4.5.

4.4 Stacking intensifiers

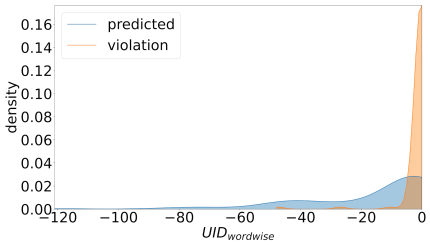
Confirming previous studies using information theory to model intensifiers in Dutch and German (Richter & Van Hout 2020, Scheffler et al. 2023b), we observed the tendency that in stacks of intensifiers, the information values increase from left to right. Table 3 depicts the number of stacks that behave according to the predictions, i.e., with information values increasing from the first to the second intensifier for both information types TCM and IC, and of stacks with violations of these predictions (mere repetitions such as *so so* are excluded). Table 3 illustrates the tendency towards increasing information in the blog, Twitter and the combined data sets. The proportions are almost identical when using TCM and IC.

Table 3: Counts of intensifier stacks following the predicted, increasing order of TCM and IC information values, compared to violations of this predicted order

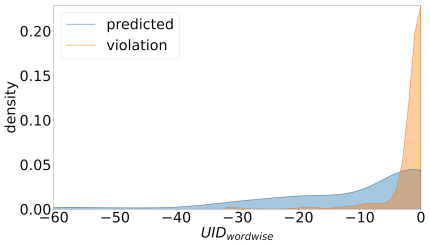
Data		Predicted	Violations	% Pred.	% Viol.
all	TCM	92	51	0.64	0.36
	IC	92	51	0.64	0.36
blogs	TCM	78	32	0.71	0.29
	IC	69	41	0.63	0.37
tweets	TCM	26	7	0.79	0.21
	IC	25	8	0.76	0.24

We follow the argumentation by Scheffler et al. (2023b) that some number of violations of the predicted stacking order are expected due to the creative choices of the authors, but also due to the inherent uncertainty of our estimates of the information values. Both information values are computed within our relatively small corpus and thus may not fully reflect the true underlying information values that drive authors' choices when selecting intensifiers. Scheffler et al. (2023b) propose that this view can be confirmed by comparing the difference in information values between predicted stacks and violations: Violations of the stacking order should tend to exhibit small differences in information values between the two intensifiers, indicating that the involved intensifiers are similar to each other, and variation in their order may be due to chance or to our estimation error. On the other hand, when there are large differences in information value in an intensifier pair, authors should be much more likely to choose the predicted order.

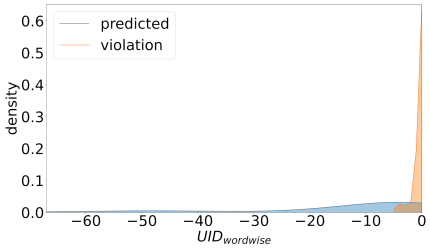
In the current study, we again find this picture confirmed, as Figure 5 illustrates. The figure depicts the density distributions of the $UID_{wordwise}$-values, i.e., the distribution of the average information change between the two intensifiers in a stack. The x-axis shows the $UID_{wordwise}$-values, the y-axis the normalized relative frequency of each value. Let *min* be the minimum of $UID_{wordwise}$ in a data set and the maximum = 0. The area under the curve is determined by the integral $\int_{min}^{0} x \, dx = 1$.



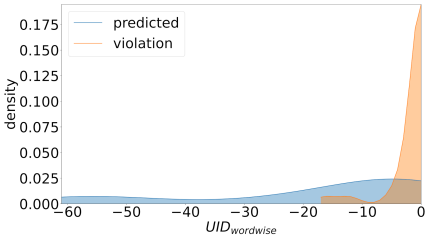
(a) Information Density of IC in the blog data



(b) Information Density of TCM in the blog data



(c) Information Density of IC in the Twitter data



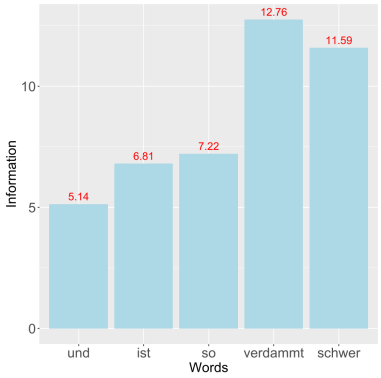
(d) Information Density of TCM in the Twitter data

Figure 5: $UID_{wordwise}$ density plot of predicted information structures (TCM values and IC values) (blue) and of violations (orange)

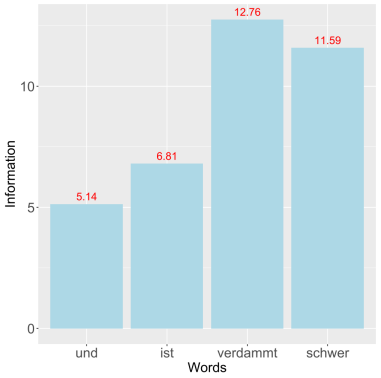
In the stacks with violations, the change in information values as captured by $UID_{wordwise}$ tend to be closer to zero than in stacks with the predicted order. This indicates closer information values within violations than in the stacks with predicted increasing information. The differing information density between predictions and violations holds both for the surprisal values TCM and for the IC values. This means that violations of the predicted increase in information values are more likely when the word-wise information changes are rather small. This replicates the previous finding by Scheffler et al. (2023b) on our new dataset in this study, for both information values. When the difference between the information values of two intensifiers in a stack is very small, the preference for their ordering is not very strong and they may thus occur in reverse order.

4.5 Can UID explain intensifier choice and order?

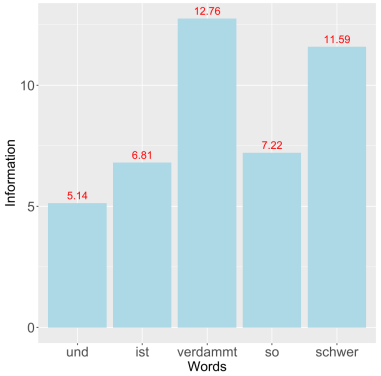
Scheffler et al. (2023b: 14) proposed "to interpret established intensifiers as the 'glue' that holds the whole set together: in stacks, established intensifiers prepare the language processor for more expressive, innovative intensifiers to follow". The conjecture was that adding a frequent intensifier with little own information contribution smooths the information flow within a sentence by decreasing the overall change in $\text{UID}_{\text{wordwise}}$. In this paper, we compute information values for all words in a sentence and can thus systematically test this conjecture, by comparing the attested version of a sentence with two intensifiers with two communicatively equivalent variants: (i) the shortened sentence where the first



(a) Flow of IC in the phrase *und ist so verdammt schwer*



(b) Flow of IC when *so* is omitted



(c) Flow of IC when the order of *so* and *verdammt* is reversed

Figure 6: Example flow of IC in (a) complete, (b) shortened, and (c) reversed stacks, for the phrase *und ist so verdammt schwer* ('and is so damn hard')

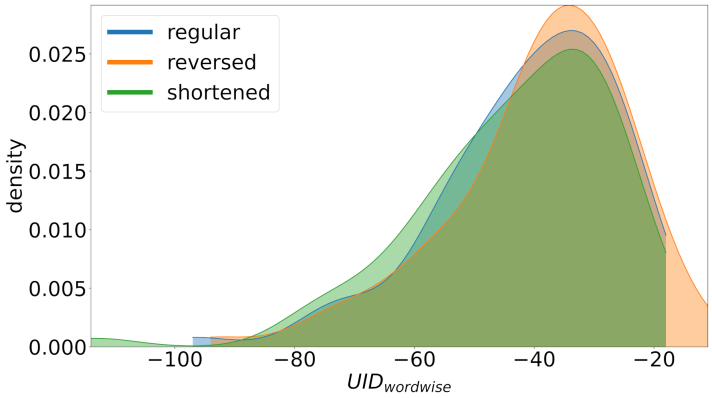
intensifier is omitted (shortened), and (ii) the sentence where the two intensifiers are reversed in order (reversed).

To give an example, in Figure 6, the flow of information in a stack with the extremely common intensifier *so* is depicted in Figure 6a. Figure 6b shows how the flow of information changes when *so* is omitted: the gap between the copula verb and the remaining second intensifier is somewhat larger, but the effect is not very pronounced. Finally, Figure 6c shows the flow of information when both intensifiers are reversed, which is much more bumpy and thus less optimal with respect to the UID hypothesis.

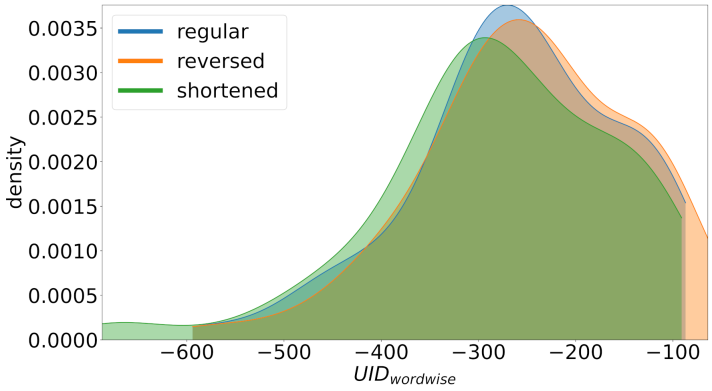
We computed the wordwise information density for all original sentences in the corpus, as well as in the shortened and reversed variants. Figure 7 shows the information density distribution over all sentences in the blog subcorpus, compared to the shortened and reversed variants, for both the IC (top) and TCM (bottom) models. We use the blog data because it provides fully coherent texts which allow for better estimates of the information values. However, the Twitter data shows similar behavior and the corresponding graphs for the Twitter subset and the complete data set are shown in the Appendix.

We applied an ANOVA to compare the three types of outcomes. For IC we obtained a significant result ($F(2, 158) = 13.037, p < 0.001$) as Figure 7a illustrates. Posthoc comparisons (Bonferroni) showed a significant difference between regular and reversed (higher scores) versus shortened (lower scores). For TCM we obtained a significant result as well ($F(2, 158) = 34.712, p < 0.001$), see Figure 7b. Moreover, all posthoc comparisons (Bonferroni) were significant ($p < 0.01$), going from regular (higher scores) via reversed to shortened (lower scores). The density plots for the Twitter data and the Twitter and blogs together (complete data) can be found as Figure 9 and Figure 10, respectively, in the Appendix. The number of observations in the Twitter data is rather low ($n = 24$). The result is thus not significant for IC ($F(2, 46) = 2.752, p = 0.076$). The result for TCM is significant ($F(2, 46) = 5.312, p = 0.013$). The only posthoc significant difference is between regular (higher scores) and shortened stacks (lower scores). The results for the complete data are similar to the results of the blog data. For IC we obtained a significant result ($F(2, 206) = 16.074, p < 0.001$), with posthoc outcomes similar to the blog data. For TCM we obtained a significant result ($F(2, 206) = 32.0293, p < 0.001$), as well, with the posthoc outcomes similar to the blog data.

It can be seen that the average change in density is very small, since most words stay the same within all variants of each sentence, and the change by removing or reordering the frequent intensifier is divided by the number of words



(a) Information Density of IC in the blog data, compared to shortened and reversed stacks



(b) Information Density of TCM in the blog data, compared to shortened and reversed stacks

Figure 7: Average per-word information density in complete sentences compared to their shortened and reversed variants, in the blog subcorpus

in the sentence. However, the observable small change is in the expected direction, at least for the shortened version: The alternative utterances show a decrease in the smoothness of the information flow compared to the actually attested utterances. Note that the information values used here are based on the relatively small Twitter and blog corpus. This means that the context for semantic surprisal TCM is very limited. We expect thus a more precise estimate of the information values based on a larger corpus to lead to a clearer result.

4.6 A quick note on ad-DP intensifiers

An interesting observation about German intensifiers is that they can occur in a DP-external position when modifying an attributive adjective or a noun, as shown in the constructed examples in (12), and discussed by Gutzmann & Turgay (2014), Willich & Politt (2023)[6]. Note that the English translations exhibit the intensifiers in the expected, pre-adjectival DP-internal position.

(12) a. *Das war voll das schöne Wochenende!*
That was INT the beautiful weekend
'That was a super nice weekend!'

b. *Ich habe voll das schöne Bild gemalt.*
I have INT the beautiful picture drawn
'I have drawn such a nice picture.'

In this construction, the intensifier *voll* ('fully') precedes the determiner of the DP; this determiner is typically definite (but indefinites are also possible). In terms of meaning, the sentences in (12) are denotationally equivalent to their canonical variants in (13) (Gutzmann & Turgay 2014). Gutzmann & Turgay (2014) make the important observation that the noun is interpreted as indefinite, even if a definite determiner is present.

(13) a. *Das war ein voll schönes Wochenende!*
That was a INT beautiful weekend
'That was a super nice weekend!'

b. *Ich habe ein voll schönes Bild gemalt.*
I have a INT beautiful picture drawn
'I have drawn such a nice picture.'

The DP-external intensifier variant in (12) is understood as more expressive and stronger than the canonical, DP-internal variant (13). We believe that the information theoretic measures introduced in this paper can explain this difference. Comparing the TCM-surprisals, that is, the information values derived from topic contexts, for the sentences in (12) with their corresponding (13) variants, we observe that the canonical versions have a much smoother information profile (see Figure 8). Thus, the speaker seems to be encoding additional surprisal by using the marked ad-DP intensifier construction. Our blog/tweets corpus contains no instances of this construction, but it would be very worthwhile to carry out a larger corpus study to corroborate this observation.

[6]We thank an anonymous reviewer for encouraging us to discuss this phenomenon.

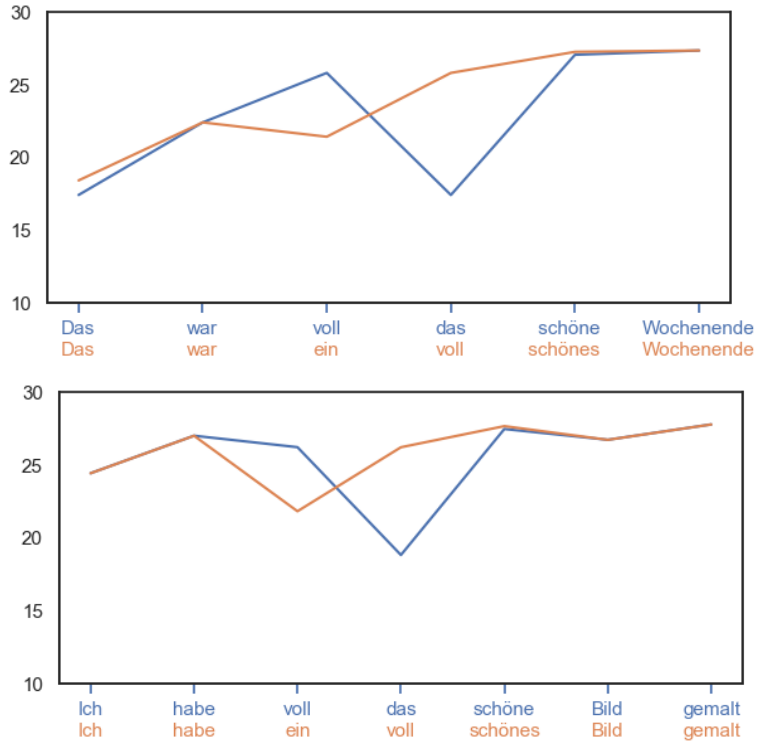


Figure 8: TCM information profiles for ad-DP and canonical sentences (12–13)

5 Discussion

In this paper, we apply information theoretic analyses to a new dataset of manually annotated intensifiers in two media, tweets and blog posts, to account for their expressive meaning and characterize the variability of available intensifying items. We confirm previous information theoretic predictions about the distribution of intensifiers in a new medium, blog posts. These results hold even though the data we report on here is no longer restricted to intensification of predicative adjectives, but includes intensifiers of adjectives in other positions, as well as nominal and verbal intensification. We show that the context-free information value IC and the context-dependent information measure TCM are highly correlated for the entire dataset. Both measures can account for the distribution of intensifiers in our data, as well as their tendency to occur in order of increasing information in a phrase, giving credence to the notion that intensifiers form a common word class, even across syntactic and semantic differences.

Since both values measure the surprisal of a word, it is expected that they are related. However, they are not identical, since IC is a static value computed based on overall word frequency, while TCM is dependent on the context of a specific instance (here, dependent on the document and its topic distribution). The TCM values are more reliable for the blog posts than the tweets. This is most likely due to the size of the contexts, which for the tweets only include the single tweet, which means an extremely small context for topic detection. We assume that the relationship between TCM and IC will change further when the context is increased again, i.e., larger than for the blogs and, for example, includes longer documents or encompasses an entire subcorpus. The prediction is that for large coherent contexts, TCM will yield a "better", that is more accurate, information value for words, which could manifest itself in a lower number of violations, i.e., deviations from the expected increase in information values in stacks of intensifiers.

We explicitly tested the conjecture from previous work, that stacked intensifiers tend to be ordered from least to most expressive within a phrase, on our entire corpus of sentences with intensifiers. We systematically compared sentences with stacked intensifiers with the communicatively equivalent alternatives (i) containing only the last, most expressive intensifier, or (ii) containing both intensifiers in reverse order. We computed information profiles for both variants and compared them to the actually attested variant. Our results show that the attested sentences exhibit a smoother information contour (higher uniformity of information density). This supports the Uniform Information Density hypothesis, which states that smoother information flow is preferred in communication, when keeping other factors equal. This analysis may serve to shed light on the puzzle why writers bother to add further, redundant intensifiers beyond the most expressive one: Their function is to introduce the phrase, alert the reader to the word class which follows, and smooth the information flow.

Acknowledgements

We would like to thank the annotators, in particular Imge Yüzüncüoglu, for their assistance in preparing the data. We are grateful for the detailed comments from two anonymous reviewers, as well as the editors of this volume.

References

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. DOI: 10.1177/00238309040470010201.

Beltrama, Andrea. 2022. Just perfect, simply the best: An analysis of emphatic exclusion. *Linguistics and Philosophy* 45. 321–364. DOI: 10.1007/s10988-021-09326-x.

Bennett, Erin D. & Noah D. Goodman. 2018. Extremely costly intensifiers are stronger than quite costly ones. *Cognition* 178. 147–161. DOI: 10.1016/j.cognition.2018.05.011.

Blei, David M., Andrew Y. Ng & Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3. 993–1022.

Bolinger, Dwight. 1972. *Degree words*. De Gruyter Mouton. DOI: 10.1515/9783110877786.

Breindl, Eva. 2007. Intensitätspartikeln. In Ludger Hoffmann (ed.), *Handbuch der deutschen Wortarten*, 397–422. De Gruyter. DOI: 10.1515/9783110217087.397.

Claudi, Ulrike. 2006. Intensifiers of adjectives in German. *Language Typology and Universals* 59(4). 350–369. DOI: 10.1524/stuf.2006.59.4.350.

Collins, Michael Xavier. 2014. Information density and dependency length as complementary cognitive models. *Journal of Psycholinguistic Research* 43(5). 651–681. DOI: 10.1007/s10936-013-9273-3.

Eckart de Castilho, Richard, Éva Mújdricza-Maydt, Seid Muhie Yimam, Silvana Hartmann, Iryna Gurevych, Anette Frank & Chris Biemann. 2016. A web-based tool for the integrated annotation of semantic and syntactic structures. In *Proceedings of the workshop on language technology resources and tools for digital humanities (LT4DH)*, 76–84. Osaka: The COLING 2016 Organizing Committee. https://www.aclweb.org/anthology/W16-4011.

Fenk, August & Gertraud Fenk. 1980. Konstanz im Kurzzeitgedächtnis – Konstanz im sprachlichen Informationsfluß. *Zeitschrift für experimentelle und angewandte Psychologie* 27(3). 400–414.

Fortin, Antonio. 2011. *The morphology and semantics of expressive affixes*. Oxford University. (Doctoral dissertation).

Fortin, Antonio & Franz Rainer. 2022. Evaluative suffixes. In Adam Ledgeway & Martin Maiden (eds.), *The Cambridge handbook of Romance linguistics*, 434–458. Cambridge: Cambridge University Press. DOI: 10.1017/9781108580410.017.

Gutzmann, Daniel & Katharina Turgay. 2014. Expressive intensifiers and external degree modification. *Journal of Comparative Germanic Linguistics* 17(3). 185–228. DOI: 10.1007/s10828-014-9069-3.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL*. https://aclanthology.org/N01-1021.

Hale, John, David Lutz, Wen-Ming Luh & Jonathan Brennan. 2015. Modeling fMRI time courses with linguistic structure at various grain sizes. In Tim O'Donnell & Marten van Schijndel (eds.), *Proceedings of the 6th workshop on cognitive modeling and computational linguistics*, 89–97. Denver: Association for Computational Linguistics. DOI: 10.3115/v1/W15-1110.

Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61(1). 23–62. DOI: 10.1016/j.cogpsych.2010.02.002.

Kölbl, Max, Yuki Kyogoku, J. Nathanael Philipp, Michael Richter, Clemens Rietdorf & Tariq Yousef. 2020. Keyword extraction in German: Information-theory vs. deep learning. *Proceedings of the 12th International Conference on Agents and Artificial Intelligence – Volume 1: NLPinAI*. 459–464. DOI: 10.5220/0009374704590464.

Kölbl, Max, Yuki Kyogoku, J. Nathanael Philipp, Michael Richter, Clemens Rietdorf & Tariq Yousef. 2021. The semantic level of Shannon information: Are highly informative words good keywords? A study on German. In Roussanka Loukanova (ed.), *Natural language processing in artificial intelligence – NLPinAI 2020*, vol. 939 (Studies in Computational Intelligence (SCI)), 139–161. Springer International Publishing. DOI: 10.1007/978-3-030-63787-3_5.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177. DOI: 10.1016/j.cognition.2007.05.006.

Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt & T. Hoffman (eds.), *Advances in neural information processing systems 19 (NeurIPS 2006)*.

Philipp, J. Nathanael, Max Kölbl, Yuki Kyogoku, Tariq Yousef & Michael Richter. 2022. One step beyond: Keyword extraction in German utilising surprisal from topic contexts. In Kohei Arai (ed.), *Intelligent computing*, 774–786. Cham: Springer International Publishing. DOI: 10.1007/978-3-031-10464-0_53.

Philipp, J. Nathanael, Michael Richter, Erik Daas & Max Kölbl. 2023. Are idioms surprising? *Proceedings of the 19th Conference on Natural Language Processing (KONVENS 2023)*.

Potts, Christopher. 2007. The expressive dimension. *Theoretical Linguistics* 33(2). 165–198. DOI: 10.1515/TL.2007.011.

Richter, Michael & Roeland Van Hout. 2020. Ranking Dutch intensifiers: A usage-based approach. *Language and Cognition* 12(2). 343–359. DOI: 10.1017/langcog.2019.50.

Scheffler, Tatjana, Lesley-Ann Kern & Hannah Seemann. 2022. The medium is not the message: Individual level register variation in blogs vs. tweets. *Register Studies* 4(2). 171–201. DOI: 10.1075/rs.22009.sch.

Scheffler, Tatjana, Lesley-Ann Kern & Hannah Seemann. 2023a. Individuelle linguistische Variabilität in sozialen Medien. In M. Kupietz & T. Schmidt (eds.), *Neue Entwicklungen in der Korpuslandschaft der Germanistik: Beiträge zur IDS-Methodenmesse 2022* (Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache (CLIP) 11). Tübingen: Narr.

Scheffler, Tatjana, Michael Richter & Roeland Van Hout. 2023b. Tracing and classifying German intensifiers via information theory. *Language Sciences* 96. DOI: 10.1016/j.langsci.2022.101535.

Schmidt, Jessica. 2022. Do intensifiers lose their expressive force over time? In Remus Gergel, Ingo Reich & Augustin Speyer (eds.), *Particles in German, English, and beyond.* Amsterdam: John Benjamins. DOI: 10.1075/slcs.224.03sch.

Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27(3). 379–423.

Stratton, James M. 2020. Adjective intensifiers in German. *Journal of Germanic Linguistics* 32(2). 183–215. DOI: 10.1017/S1470542719000163.

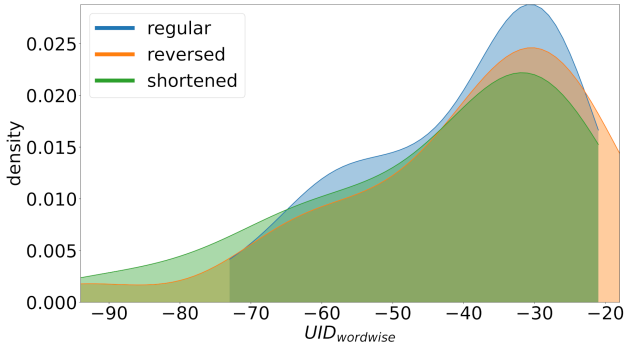
Tagliamonte, Sali A. 2008. So different and pretty cool! Recycling intensifiers in Toronto, Canada. *English Language and Linguistics* 12(2). 361–394. DOI: 10.1017/S1360674308002669.

Tribus, Myron. 1961. Information theory as the basis for thermostatics and thermodynamics. *Journal of Applied Mechanics* 28(1). 1–8. DOI: 10.1115/1.3640461.

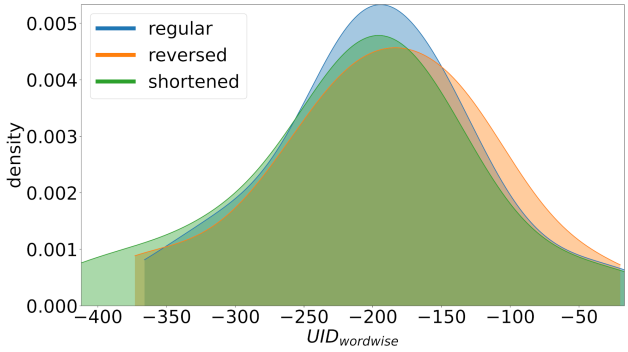
Willich, Alexander & Katja Politt. 2023. *"Du hast eh voll den guten Musikgeschmack!" – Underdetermination and grammaticalization of NP-external intensifying constructions in German.* Talk presented at ICCG12. http://tiny.cc/18v7vz (30 November, 2023).

Zadeh, Lotfi A. 1972. A fuzzy-set-theoretic interpretation of linguistic hedges. *Journal of Cybernetics* 2(3). 4–34. DOI: 10.1080/01969727208542910.

Appendix A Information density in shortened and reversed stacks

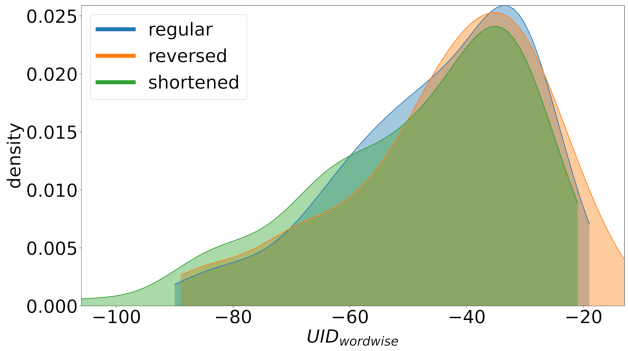


(a) Information Density of IC in the Twitter data, compared to shortened and reversed stacks

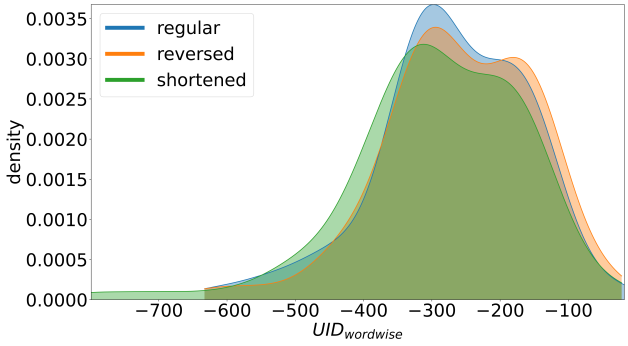


(b) Information Density of TCM in the Twitter data, compared to shortened and reversed stacks

Figure 9: Average per-word information density in complete sentences compared to their shortened and reversed variants, in the Twitter subcorpus.



(a) Information Density of IC in the complete data, compared to shortened and reversed stacks



(b) Information Density of TCM in the complete data, compared to shortened and reversed stacks

Figure 10: Average per-word information density in complete sentences compared to their shortened and reversed variants, in the complete corpus (tweets and blogs)

Chapter 6

Cleft sentences reduce information density in discourse

Swantje Tönnis
University of Stuttgart

This paper develops a novel theoretical analysis of clefts as a discourse structuring device in written German (following Destruel & Velleman 2014, Tönnis 2021). The analysis is based on the assumption of an expectation-driven Question Under Discussion (QUD) model where addressees form a probability distribution over questions that an (ensuing) utterance is likely to answer (cf. Kehler & Rohde 2017). Tönnis (2021) argued that clefts are used to address relatively less expected QUDs, in contrast to canonical sentences, which address relatively expected QUDs. In this paper, I propose an information-theoretic take on the expectedness approach and combine it with the cleft's function to disambiguate focus, following the uniform information density (UID) hypothesis (Levy & Jaeger 2007). I hypothesize the following: The cleft in written German is used to reduce information density in order to achieve UID at the discourse level. The additional material in the cleft, compared to the canonical sentence, explicitly marks the addressed QUD (i.e., it disambiguates focus). This way it takes over information that is otherwise carried by the words of the canonical sentence and, thus, reduces information density. I argue that this reduction is only necessary when a less predictable QUD is addressed. Following Asr & Demberg (2015) and Demberg & Keller (2008), I define QUD surprisal in order to integrate the expectedness of the addressed QUD into a model that predicts the choice of the conveyed message (cleft vs. canonical sentence). For the discussed example, and similar ones, the model makes correct predictions in contrast to previous analyses. Furthermore, aiming for UID in discourse provides a reason for why clefts tend to address relatively less expected QUDs, which was missing from Destruel & Velleman (2014) and Tönnis (2021).

Swantje Tönnis. 2024. Cleft sentences reduce information density in discourse. In Robin Lemke, Lisa Schäfer & Ingo Reich (eds.), *Information structure and information theory*, 147–175. Berlin: Language Science Press. DOI: 10.5281/zenodo.13383793

1 Introduction

When deciding how to realize the next discourse move in a written text, a German speaker has the choice between a canonical sentence, as in (1), and an *es*-cleft, as in (2), among other options.[1]

(1) *Bo hat die Kekse gegessen.*
Bo has the cookies eaten
'Bo ate the cookies.'

(2) *Es war Bo, der die Kekse gegessen hat.*
it was Bo who the cookies eaten has
'It was Bo who ate the cookies.'

This paper aims to predict and explain the choice of an author between a cleft and its canonical variant while taking the discourse context into account. The analysis is motivated by an example inspired by Tönnis (2021) which illustrates that the preference between the cleft and its canonical variant varies with the discourse context. In a discourse context such as (3), the cleft in (3b) is preferred over the canonical sentence in (3a).[2]

[1]This paper is only concerned with written German. In the following, I will, thus, refer to the *author* of a text instead of the *speaker* of an utterance. In principle, the analysis to be presented should also apply to spoken German, but it would have to be adapted, which I leave as a topic for further research. For instance, one would have to incorporate the effect of overt focus marking by intonation in spoken German, which is more flexible than implicit prosody in written German. More details on implicit prosody are presented in Section 2.2.

[2]A further continuation which is frequently judged as equally acceptable as the cleft in (3) is the canonical sentence with a past perfect verb form instead of the present perfect, as in (i). These judgments have not yet been systematically tested in an experimental setting, though.

(i) *Bo hatte die Kekse gegessen.*
Bo had the cookies eaten
'Bo had eaten the cookies.'

Tönnis & Tonhauser (2022: 678), however, provided the example in (ii) to show that clefts are not only acceptable in anteriority contexts. In this context, they judged the past perfect canonical sentence as equally dispreferred as the present perfect canonical sentence.

(ii) *Lena hat auf der Party mit einem Typen geflirtet. Sie hatte sehr viel Spaß. Anna war eher gelangweilt.*
'Lena flirted with some guy at the party. She had a lot of fun. Anna was rather bored.'
a. Es war Peter, mit dem Lena geflirtet hat.
'It was Peter Lena flirted with.'
b. ? Lena hat mit Peter geflirtet. / ? Lena hatte mit Peter geflirtet.
'Lena flirted with Peter.' / 'Lena had flirted with Peter.'

(3) *Als Lena in die Kaffeepause kam, war der Keksteller schon leer. Sie fand auch keinen weiteren Keksteller. Also entschied sie sich zum Bäcker zu gehen.*
'When Lena joined the coffee break, the plate of cookies was already empty. She couldn't find any other cookies, either. So she decided to go to the bakery.'

a. ? Bo hat die Kekse gegessen.
'Bo ate the cookies.'

b. Es war Bo, der die Kekse gegessen hat.
'It was Bo who ate the cookies.'

In contexts such as (4), there is no clear preference between the cleft and the canonical sentence, or possibly a slight tendency towards the canonical sentence.

(4) *Als Lena in die Kaffeepause kam, war der Keksteller schon leer.*
'When Lena joined the coffee break, the plate of cookies was already empty.'

a. Bo hat die Kekse gegessen.
'Bo ate the cookies.'

b. (?) Es war Bo, der die Kekse gegessen hat.
'It was Bo who ate the cookies.'

Based on my theoretical model (Tönnis 2021), I predicted the contrast for the cleft and the canonical sentence in (3) and (4) by referring to the expectedness of the question under discussion (QUD) (Roberts 2012) addressed by the cleft/canonical sentence (here *Who ate the cookies?*). She argued that clefts are used in German to address relatively less expected QUDs in discourse while canonical sentences can only be used to address relatively expected QUDs.

Tönnis & Tonhauser (2022) provided empirical evidence for this contrast for 16 context pairs like (3) and (4) depending on question expectedness. In a norming study, they measured the expectedness rating for the question Q (here *Who ate the cookies?*) to be addressed next by the author in two conditions: 1-sentence contexts, like (4), which gave rise to question Q, and 3-sentence contexts, like (3), which contained two interfering sentences after the Q-raising sentence, giving rise to new, more prominent questions. They found that Q was significantly more expected to be addressed next in the 1-sentence contexts compared to the 3-sentence contexts. Furthermore, they collected relative preference ratings between the cleft and the canonical sentence in the two context conditions. They found that there was a significantly stronger preference towards the cleft in the

3-sentence condition, i.e., when the addressed question was less expected, compared to the 1-sentence condition, where they found no clear preference between the cleft and its canonical variant. The analysis presented in this paper will be illustrated by examples (3) and (4), but it is assumed to apply equally well to other (similar) examples, given Tönnis & Tonhauser's empirical evidence.

The analysis in Tönnis (2021) correctly predicts the judgments as found by Tönnis & Tonhauser (2022). However, it does not provide an explanation for why the cleft fulfills the function of addressing a less expected QUD. Hence, I propose to extend her analysis by referring to information density (e.g., Shannon 1948, Levy & Jaeger 2007) relative to the predictability of the addressed QUD. Employing UNIFORM INFORMATION DENSITY (UID) (e.g., Jaeger 2010), I hypothesize that a cleft is used to reduce information density in discourse in the case of addressing a less predictable/expected QUD. Furthermore, the cleft's function of disambiguating focus (as claimed by, e.g., De Veaugh-Geiss et al. 2015, Tönnis et al. 2018) plays an important role in predicting the author's choice between the canonical sentence and its cleft variant. Assuming that the additional words of the cleft, compared to the canonical sentence, explicitly mark the QUD and, thus, take over the information of focus marking, clefting contributes to distributing information more uniformly and reduces information density per discourse move. Moreover, I argue that the information of focus marking only needs to be distributed further if the QUD that is addressed is relatively surprising. I implement this approach by defining QUD SURPRISAL, which measures the surprisal of a certain QUD to be addressed.

The information-theoretic approach to clefts can account for the (dis)preference of clefts in contexts such as (3) and (4). Importantly, it does not require any other effects of the cleft, such as exhaustivity or contrast, but still allows for such effects to occur. At the same time, the approach explains why the cleft is a suitable candidate to mark the reduced expectedness of the addressed QUD.

The paper is structured as follows: Section 2 presents prior analyses of clefts. Section 3 introduces a first take on an information-theoretic analysis of clefts in discourse. In Section 4, advantages, implications, and some possibilities of extensions of the proposed information-theoretic analysis are discussed. Section 5 concludes this chapter.

2 Prior analyses of clefts

Different purposes for choosing a cleft over a canonical sentence have been proposed in the literature, ranging from expressing exhaustivity (e.g., Horn 1981,

Büring & Križ 2013), contrast (e.g., Rochemont 1986), or a violation of expectations (e.g., Destruel & Velleman 2014, Tönnis 2021) to information structural functions, such as disambiguating focus (e.g., De Veaugh-Geiss et al. 2015, Tönnis et al. 2018). The main focus of this paper lies on the latter two approaches, given that the information-theoretic take on clefts, presented in Section 3, is based on those. Furthermore, Tönnis & Tonhauser (2022) argued that the other approaches to clefts cannot account for the contrast in contexts such as (3) and (4).

2.1 Inferences of clefts

I will first present those analyses that are concerned with the different kinds of inferences that are conveyed by a cleft. The cleft is commonly claimed to have the meaning components in (5) (e.g., Velleman et al. 2012, Krifka & Musan 2012, De Veaugh-Geiss et al. 2018, Destruel et al. 2019).

(5) Es war Bo, der die Kekse gegessen hat. ('It was Bo who ate the cookies.')
   a. Prejacent: Bo ate the cookies.
   b. Indication of question: Who ate the cookies?
   c. Existential inference: Somebody ate the cookies.
   d. Exhaustivity inference: Nobody other than Bo ate the cookies.

The prejacent in (5a) is the at-issue content of the cleft that is assumed to be identical to the at-issue content of its canonical variant (see De Veaugh-Geiss et al. 2018). Furthermore, several approaches, though for different reasons, have argued for the cleft indicating an implicit question (5b). According to Velleman et al. (2012), the cleft structure involves a cleft operator which requires a question to be present in the discourse. In the case of focus-background clefts, in which the clefted element, *Bo* in (5), is focused, focus marking also indicates the same implicit question (Krifka & Musan 2012). More details on the issue of focus marking in the cleft are given in Section 2.2. The existential inference, as in (5c), is commonly analyzed as a presupposition (e.g., Halvorsen 1978, Prince 1978, Percus 1997). The role of the exhaustivity inference of clefts, exemplified in (5d), is still debated. Some approaches (e.g., Szabolcsi 1981, Percus 1997, Büring & Križ 2013, Pollard & Yasavul 2015) analyzed the exhaustivity inference of clefts as a presupposition, i.e., a semantic inference. However, there are also approaches which analyzed it as a pragmatic inference (e.g., Horn 1981, De Veaugh-Geiss et al. 2015, Onea 2019). De Veaugh-Geiss et al. (2018) provided empirical evidence for the exhaustivity inference of German clefts being stronger than exhaustivity in canonical sentences with a focus on the constituent that forms the cleft pivot

in the cleft, such as (6). A stronger inference meant that violations of exhaustivity were less frequently accepted, and the truth of the exhaustivity inference was more frequently required for judging the respective sentence as true. Furthermore, the exhaustivity inference of clefts was found to be weaker than for exclusives, as in (7).

(6) *BO_F hat die Kekse gegessen.*
BO_F has the cookies eaten
'BO_F ate the cookies.'

(7) *Nur Bo hat die Kekse gegessen.*
only Bo has the cookies eaten
'Only Bo ate the cookies.'

De Veaugh-Geiss et al. (2018) concluded that exhaustivity in clefts is a not-at-issue pragmatic inference. Following Tönnis & Tonhauser (2022: 663), I argue that exhaustivity of clefts does not fully account for the preference between the cleft and its canonical variant in contexts such as (3) and (4). Example (8) represents a violation of exhaustivity which, nevertheless, is acceptable in contexts (3) and (4). More importantly, the cleft in (8) is still preferred over the canonical sentence (with or without exhaustivity violation) in context (3).

(8) Es war Bo, der die Kekse gegessen hat, und Lou auch.
'It was Bo who ate the cookies and Lou as well.'

A further function which is frequently assigned to clefts is marking contrast. Rochemont (1986), for instance, argued that a cleft necessarily expresses contrastive focus while its canonical variant can express both contrastive and informational focus. According to Tönnis & Tonhauser (2022), however, the preference between the cleft and the canonical sentence in the above contexts (3) and (4) cannot be accounted for by referring to contrastivity. In particular, they claim that in the cleft in context (3) (English version repeated in (9)) there is no explicit alternative provided by the context (e.g., *Lou ate the cookies* or *Lou didn't eat the cookies*) to establish a contrast with Bo eating the cookies. Accordingly, the cleft would be predicted to be dispreferred.

(9) *When Lena joined the coffee break, the plate of cookies was already empty. She couldn't find any other cookies, either. So she decided to go to the bakery.*
   a. It was Bo who ate the cookies.
   b. ? Bo ate the cookies.

Anticipating the information-theoretic approach to be presented in Section 3, I argue that the fact that Bo ate the cookies does not have to be particularly surprising for the cleft to be acceptable/preferred in (9), which is in line with the cleft not being used contrastively. Even if the author and the reader knew that Bo frequently finishes the cookies, the cleft (9a) would still be preferred over the canonical sentence (9b) in this context.

2.2 Clefts as a focus-disambiguating device

Some approaches to clefts proposed that clefts are used to disambiguate focus (e.g., De Veaugh-Geiss et al. 2015, Tönnis et al. 2018). In written German, where intonation cannot be used to mark focus, a canonical sentence is ambiguous with respect to focus. Example (10) illustrates some of the possible focus assignments for the canonical sentence (focus is marked by $[...]_F$ and the main accent is marked with capital letters).

(10) a. Bo hat [die KEKse gegessen]$_F$.
b. Bo hat [die KEKse]$_F$ gegessen.
c. [BO]$_F$ hat die Kekse gegessen.
d. *[Bo hat die KEKse gegessen]$_F$.*
Bo has the cookies eaten
'Bo ate the cookies.'

Fodor (2002), among others, pointed out that, even though intonation cannot be marked in written language, there is still evidence for implicit prosody during silent reading (for a comprehensive overview of implicit prosody, see Féry 2017: ch. 9.4). According to her implicit prosody hypothesis, the reader would assume the default prosody, which is "identical to the overt prosody for that sentence in a comparable context (i.e., same illocutionary force, focus structure, etc.)" (Fodor 2002: 115). Hence, when the sentences in (10) are interpreted in their discourse context the ambiguity is usually resolved, as in Krifka & Musan's (2012: 11) examples, given in (11a) and (11b).

(11) a. And then something strange happened. [A MEteorite fell down.]$_F$
b. Mary sat down at her desk. She [took out a pile of NOTES]$_F$.

The discourse context in these examples (represented by the respective first sentence) makes the focus assignments and focus markings, which are indicated in the second sentence, the only reasonable ones.

The discourse context has often been claimed to affect focus assignment (e.g., Beaver & Clark 2008, Krifka & Musan 2012, Simons et al. 2017, Tönnis 2021). Krifka & Musan (2012), for instance, analyzed focus as marking (implicit) questions on the basis of the context. In other words, focus marking in an utterance U helps the reader to identify the question which is addressed by U. If focus is marked in an ambiguous way, as in the case of the canonical sentence in written German, this implicit question needs to be accommodated by the reader. This process and, thereby, also the implicit prosody strongly depend on contextual cues.

Focus-background clefts, in contrast, simplify the accommodation process of the implicit question: De Veaugh-Geiss et al. (2015) proposed that clefts structurally mark focus by backgrounding the content of the relative clause. Furthermore, they assumed that focus cannot project out of the cleft pivot. This leads to the unambiguous narrow focus marking in (12).[3]

(12) Es war $[\text{Bo}]_{\text{F}}$, der die Kekse gegessen hat.
'It was $[\text{Bo}]_{\text{F}}$ who ate the cookies.'

Tönnis et al. (2018) supported their claim by an extensive corpus study on written German, in which they annotated the grammatical function of the cleft pivot in German clefts. They found that there were relatively more subject clefts, such as (12), than non-subject clefts, such as the object cleft in (13), even when the generally higher subject frequency was taken into account.

(13) Es waren die Kekse, die Bo gegessen hat.
'It was the cookies that Bo ate.'

Tönnis et al. (2018) argued that the subject preference in the pivot results from the fact that intonation cannot be used in written German to mark focus. Hence, the reader has to use cues from the context to accommodate the focus marking of each sentence. If there is no strong contextual cue, the reader will accommodate default focus marking, namely object focus or wide focus. However, if the author

[3] In some cases, focus is even ambiguous in clefts. Velleman et al. (2012: 442) presented the example in (i), where only part of the cleft pivot is focused.

(i) It was John's $[\text{eldest}]_{\text{F}}$ daughter who liked the movie.

Similar clefts can be found in German. In such cases, clefts only reduce the possible focus readings compared to the canonical sentence in written German. The examples discussed in this paper always contain cleft pivots which consist of one word, which always leads to a clear disambiguation. I leave the investigation of possible effects due to narrow focus inside of the pivot to future research.

wants to express narrow subject focus, she could use a cleft to shift default focus to the subject position (e.g., Reinhart 1995, Szendrői 1999). Since this extra marking is necessary for subjects but not for objects, Tönnis et al.'s (2018) approach correctly predicted a higher frequency of subject clefts compared to object clefts, and concluded that clefts are used to disambiguate focus in written German.[4]

The information-theoretic analysis proposed in this paper adopts the idea that clefts disambiguate focus, i.e., they explicitly indicate the addressed question. The question arises when focus disambiguation is necessary, and accordingly, why there is not a much higher cleft frequency compared to canonical sentences in written German. When is it not enough to assume default implicit prosody? I argue that the need to disambiguate focus depends on the expectedness of the question under discussion (QUD) that is addressed by the respective sentence, which will be the topic of the next subsection.

2.3 Expectation-based analyses to clefts

Destruel & Velleman (2014) claimed that clefts can not only be used to mark a contrast to some content mentioned in the discourse context (as in Rochemont 1986), but that they can also be used to mark that the discourse develops into an unexpected direction. Spelling out this idea, I argued in Tönnis (2021) that German cleft sentences, unlike their canonical variants, are used to address relatively less expected questions under discussion (QUDs) (following Roberts 2012). Canonical sentences, I claimed, address relatively expected QUDs. Empirical support for this claim is given in Tönnis & Tonhauser (2022), also presented in Section 1 of this paper.

The underlying assumption is that, at each point of a discourse, the interlocutors have certain expectations about which QUD is likely to be addressed next. These expectations are modeled as a probability distribution over questions to be addressed by the ensuing utterance/sentence (following Kehler & Rohde 2017), which assigns a probability to each possible question with respect to how likely it is to be addressed next. This probability distribution is affected by each new utterance/sentence of a text. For instance, a sentence containing an implicit causality verb, such as *admire*, would raise the probability mass on certain *why*-questions,

[4]Note that Tönnis et al.'s (2018) analysis is only applicable to written German since speakers can freely use intonation on most words and syntactic positions in German. Using intonation is preferred over the more complex cleft structure to mark focus or prominence in spoken German. The consequence is, as Tönnis et al. (2018) claimed, that clefts are less frequently used in spoken German compared to written German, a claim for which, to my knowledge, there is no thorough empirical evidence yet.

as exemplified in (14). Empirical evidence for this claim was provided by Westera & Rohde (2019). In a question elicitation experiment, they showed that significantly more *why*-questions were elicited for implicit causality contexts compared to other contexts.

(14) Lou admired Bo.
⇝ *Why did Lou admire Bo?* becomes more likely to be addressed in the ensuing utterance than before (14) was uttered.

In Tönnis (2021), I argued that, based on the distribution of QUDs to be addressed next, the addressee needs to decide whether to accept a sentence as a relevant discourse move (as described by Roberts 2012). Note that the definition of QUD in Tönnis (2021) diverges from Roberts' (2012) with respect to the assumed hierarchy between QUDs. In Roberts' (2012) approach, QUDs are organized on a stack, and in most cases, only addressing the top-most question or a sub-question thereof constitutes a relevant discourse move. In Tönnis (2021), QUDs are not strictly organized hierarchically and a wider variety of QUDs are acceptable to be addressed. Whether addressing a QUD constitutes a relevant discourse move mainly depends on the expectedness/probability of this QUD to be addressed next. Tönnis (2021: 286) introduced a threshold of expectedness for QUDs. A discourse move is only relevant if, among other conditions, the expectedness of the question it addresses exceeds this threshold. What the actual value of this threshold is is still to be determined empirically. For this paper, I assume such a threshold exists and that certain constellations push the expectedness value of a question above or below this threshold.

In the following, I present some examples illustrating the approach in Tönnis (2021). In context (14), for instance, the expectedness value of the question Q1: *Why did Lou admire Bo?* is assumed to exceed the threshold. Accordingly, the second sentence in (15) would be accepted as a relevant discourse move because it addresses Q1.

(15) Lou admired Bo. She loved the way he sung Queen's Bohemian Rhapsody.

Relatively expected questions are assumed to remain above the threshold until answered. However, an intervening sentence, which itself gives rise to a question, can lower the expectedness value of a previously raised, unanswered question, as illustrated in example (16).

(16) Lou admired Bo. But Bo had a secret.

The second sentence in (16) strongly increases the probability mass on the question Q2: *What was Bo's secret?*, which automatically reduces the expectedness of the previous question Q1: *Why did Lou admire Bo?* (because of the probabilities

of all questions adding up to 1). Therefore, the continuation in (17a), which addresses Q2, should be more acceptable than the continuation addressing Q1 in (17b), which is the case in my judgment. For similar examples, Tönnis & Tonhauser (2022) presented empirical evidence showing that the expectedness of an unanswered question that was raised by a sentence decreased after intervening sentences raised new questions (see Section 1 of this paper).

(17) *Lou admired Bo. But Bo had a secret.*
   a. His famous cover version of Queen's Bohemian Rhapsody was fake.
   b. Lou loved the way he sung Queen's Bohemian Rhapsody.

As mentioned above, I argued in Tönnis (2021) that a cleft sentence addresses relatively less expected QUDs in discourse. This means that it requires a lower threshold of question probability, compared to an unclefted sentence, in order to qualify as an acceptable discourse move. In particular, I assumed an expectedness value for the question addressed by a cleft which lies between this lower threshold for clefts and the threshold for canonical sentences.

The contrast between the cleft and the canonical sentence in the two contexts, repeated in (18), is correctly predicted by this approach.

(18) *Als Lena in die Kaffeepause kam, war der Keksteller schon leer. (Sie fand auch keinen weiteren Keksteller. Also entschied sie sich zum Bäcker zu gehen.)*
'When Lena joined the coffee break, the plate of cookies was already empty. (She couldn't find any other cookies, either. So she decided to go to the bakery.)'
   a. Bo hat die Kekse gegessen.
'Bo ate the cookies.'
   b. Es war Bo, der die Kekse gegessen hat.
'It was Bo who ate the cookies.'

The canonical sentence in (18a) is only acceptable in the shorter context, i.e., without the sentences in brackets. This context sentence evokes the question *Who ate the cookies?*, which plausibly raises the probability of this question to be addressed above the threshold. In this case, (18a) is predicted to be an acceptable discourse move and the cleft to be dispreferred.[5]

[5]Empirical evidence by Tönnis & Tonhauser (2022) showed that the cleft was not dispreferred, but that there was no clear preference between the cleft and the canonical sentence in contexts such as the short context in (18). Tönnis & Tonhauser explained this by referring to Tönnis' (2021) extended definition of acceptability of clefts/canonical sentences which specifies an overlapping region of expectedness, where both the cleft and its canonical variant are acceptable. In this case, Tönnis' account would correctly predict that both the cleft and the canonical sentence would be acceptable.

In the longer context (full context in (18)), the cleft in (18b) is preferred since the intervening two sentences give rise to new questions, such as *What did Lena get at the bakery?*. Such questions reduce the expectedness of the question addressed by the cleft (*Who ate the cookies?*), just as in example (17). If we assume that the expectedness is pushed below the threshold for canonical sentences but not below the threshold for clefts, the cleft can correctly be predicted to be acceptable while the canonical sentence is predicted to be dispreferred.

Subsection 2.2 pointed out that focus marking is ambiguous in canonical sentences for written German. The German cleft structure, however, makes focus explicit. I argue that it is not enough to analyze the cleft as a focus-disambiguating device in order to explain the preferences between German clefts and their canonical variants in discourse. What also needs to be explained is when it is necessary to disambiguate focus. I argue that the author's wish to disambiguate focus is only present when she wants to address a relatively less expected QUD. Disambiguating focus should only be necessary if it was not yet obvious in any way which QUD could be addressed next.

3 Towards an information-theoretic analysis of German clefts

Summing up previous insights, an analysis of clefts as addressing a relatively less expected QUD correctly predicts the preference between the cleft and its canonical variant in written German. What is still missing is an explanation for why the expectedness threshold is lower for clefts compared to canonical sentences. In the following, I present a proposal of an information-theoretic approach which aims to provide this explanation by combining the idea of clefts disambiguating focus and the idea of clefts addressing less expected QUDs. Note that this proposal still needs to be tested empirically, which exceeds the scope of this paper.

At first glance, the cleft just seems to be the syntactically more marked structure, which is an indication of an additional or a more complex function on some linguistic level. For instance, a more complex definite description, such as *the neighbor's dog*, is usually used to refer to a less salient antecedent in discourse than a less complex pronoun, such as *it* (e.g., Gundel et al. 1993). In the same way, I argue, that the more complex cleft addresses a less salient QUD than the less complex canonical sentence, and it does so by explicitly marking this QUD (via focus marking).

In the information-theoretic approach to language, language production is assumed to be efficient within the bounds of grammar (Jaeger 2010). The most

efficient way involves (i) distributing information uniformly across the speech signal, and (ii) keeping information density (i.e., the amount of information per unit, e.g., per word) close to the channel capacity (Genzel & Charniak 2002, Levy & Jaeger 2007). Jaeger (2010) spelled out (i) as the UNIFORM INFORMATION DENSITY (UID) hypothesis:

> Within the bounds defined by grammar, speakers prefer utterances that distribute information uniformly across the signal (information density). Where speakers have a choice between several variants to encode their message, they prefer the variant with more uniform information density (ceteris paribus). (Jaeger 2010: 25)

The two variants to encode the same message that will be relevant for the analysis of this paper are the cleft and the canonical sentence. The channel capacity in (ii) represents the information rate, i.e., a fixed amount of information per unit, that no unit should strongly deviate from (see Genzel & Charniak 2002). In other words, no unit, for instance, a word, should convey much more or much less information than the other units of the same category.

Information is understood in the sense of Shannon information (Shannon 1948), also called surprisal, for a unit of a signal. The information I of a unit, such as a word or a sentence, is defined as in (19).

(19) $$I(\text{unit}) = \log \frac{1}{p(\text{unit})} = -\log p(\text{unit})$$

This means that the higher the probability p of a unit, the lower is the information I (or the surprisal) of that unit. For instance, the more expected a word is the less new information it conveys. The information or surprisal of a unit often involves the conditional probability of the unit, for example, conditioned by the probability of the preceding units (Levy & Jaeger 2007), or the probability of possible syntactic trees (Demberg & Keller 2008) or discourse relations (Asr & Demberg 2015). The reasoning for an information-theoretic approach to clefts follows the reasoning used in Levy & Jaeger (2007) for a case of syntactic reduction, and builds on Asr & Demberg's (2015) approach, who defined discourse relational surprisal.

Levy & Jaeger (2007) predicted the syntactic variation observed for relative clauses with respect to the presence/absence of the relative pronoun. They hypothesized that the relative pronoun *that* is usually omitted when it would otherwise precede a relatively expected word, such as *you* in their example, repeated in (20).

(20) How big is the family (that) you cook for? (Levy & Jaeger 2007: 851)

They argue that the syntactic reduction is a consequence of UID at the sentence level. Both versions of (20), with and without the relative pronoun, express the same informational content, but the information is distributed differently. When the relative pronoun is not present, the first word of the relative clause, here *you*, fulfills two functions: It conveys its semantic content and it marks the onset of the relative clause. When the relative pronoun is present, these two functions are split up between the relative pronoun and the noun phrase *you*. According to the UID hypothesis, the relative pronoun is predicted to be dropped to avoid a trough in information density in case the surprisal/information of the word *you* is low, while it should be inserted to avoid a peak on *you* in case it is relatively surprising in its context. Levy & Jaeger (2007) found empirical evidence for this claim in a corpus study. I employ the same reasoning for the variation between the cleft and the canonical sentence, and hypothesize that the cleft is used to reduce information density at the discourse level when a relatively less expected/more surprising QUD is addressed.

First of all, note that the cleft and the canonical sentence express the same information when the same constituent is focused. Example (21) illustrates this for a subject cleft and a canonical sentence with subject focus marked by intonation.

(21) BO ate the cookies./It was Bo who ate the cookies.
   a. At-issue content/prejacent: Bo ate the cookies.
   b. Indication of question (focus): Who ate the cookies?
   c. Existential inference: Somebody ate the cookies.
   d. Exhaustivity inference: Nobody other than Bo ate the cookies.

As mentioned in Section 2.1, the meaning components are weighted differently for the two sentences. De Veaugh-Geiss et al. (2018), for instance, showed that the exhaustivity inference is stronger for clefts than for canonical sentences. For the analysis presented in this paper, these gradual differences will not be taken into account. Instead, it uses the simplification that clefts and their canonical variants express the same informational content. Here, I focus on the semantic content (21a) and the information structural contribution (21b).

As mentioned in Section 2.2, focus, and thereby the implicit QUD, is not overtly marked in many sentences of written German since the author cannot indicate intonation.[6] The focus is ambiguous and the implicitly indicated question must,

[6]In some types of text, for example, chat messages, the author can use capital letters to mark intonation/emphasis. In such cases, this analysis does not apply. I consider cases of written German where using capital letters for emphasis is not common.

thus, be inferred from contextual cues, as example (11) showed. Clefts, in contrast, disambiguate focus and, thus, explicitly indicate the QUD.

From an information-theoretic perspective, all words of the canonical sentence carry both the semantic content as well as the focus.[7] In the cleft, the same information is distributed onto more words. I argue that, in an information-theoretic sense, the words introduced by clefting (*es* 'it', *war* 'was', and *der* 'who') take over the information of focus (indicating the QUD) since clefting creates a syntactic structure that explicitly separates the focus from the background (see also É. Kiss 1998).[8]

The question arises why authors do not always want to disambiguate focus in written German, which would lead to a much higher frequency of clefts than actually observed in written German. I argue that the reason is efficiency, which information theory is well-suited to capture. For the choice between the cleft and the canonical sentence, I argue that reduction of information density by clefting is only necessary if the focus is difficult to identify, i.e., difficult to accommodate. This is the case if the QUD which the author wants to address is relatively less expected or more surprising. In this case, using a canonical sentence would exceed the channel capacity, i.e., too much information per word. The author is predicted to use a cleft. If the QUD the author intends to address was strongly expected, the words of the canonical sentence would not have to carry much extra information, and no extra marking by clefting would be necessary. Hence, the canonical sentence would be the preferred option.

In order to implement the influence of conveying the implicit QUD on information density, the discourse context needs to be incorporated into the calculation of information/surprisal. In particular, information density must be measured depending on the probability of the addressed QUD. Asr & Demberg (2015) presented a similar approach in their definition of DISCOURSE RELATIONAL SURPRISAL. Discourse relational surprisal describes the effect of a word on the belief distribution of discourse relations by comparing the belief distribution before and after the word. Asr & Demberg (2015) were particularly interested in the relational surprisal of discourse connectives, such as *because* or *therefore*. Relational surprisal is small if the connective did not have a strong effect on the distribution of discourse relations, i.e., the relation marked by the connective was likely

[7]The analysis does not hinge on focus to be conveyed by all of the words in a sentence. If focus is just conveyed by parts of the sentence, the same reasoning applies.

[8]The words *es war* also have a local effect on surprisal within the cleft sentence, in the sense of preparing the reader for what is to come in the cleft pivot (thanks to Lisa Schäfer for this comment). I discuss one such example in (42) in Section 4. For my main analysis, I make the simplification of assuming the words needed for clefting jointly have the function to convey focus.

even before the connective was uttered. In example (22), the connective *because* does not strongly change the distribution over discourse relations because the implicit causality verb *admire* already raised the probability mass attributed to the discourse relation CAUSE.

(22) Lou admired Bo, (because) he was such a good singer.

As indicated by the brackets in (22), the connective can be dropped in such a case. This can be explained by information theory: If the discourse relational surprisal of a connective is small, the connective should be dropped in order to avoid a trough in information density (Levy 2008, Demberg & Keller 2008, Asr & Demberg 2012, 2015).

I propose to adjust Asr & Demberg's (2015) approach and define QUD SURPRISAL, which affects the author's choice of how to encode her next message based on the expectedness of the addressed QUD. QUD surprisal captures this by comparing the two question distributions D_0 and D_w. D_0 is the previous question distribution, which speaker and addressee share given their previous conversation. It is based on the linguistic discourse context, prior probabilities for certain questions to be addressed, and the common ground.[9] D_w is the question distribution after the first word(s) of the next utterance.

For illustration, assume the simplified question distribution D_0, given in (23), in a discourse context.

(23) D_0 =
$Q1 \rightarrow 0.1$
$Q2 \rightarrow 0.2$
$Q3 \rightarrow 0.7$

Consider (24) as an example discourse context. Then, Q3 could be *Why did Lou admire Bo?*, given that it is a relatively expected question in this context.

(24) Lou admired Bo.

If the author's next word in (24) was *because*, the probability of question Q3 would increase. Accordingly, a possible question distribution D_w after (24) + *because* is given in (25).

(25) $D_w = D_{\text{because}}$
$Q1 \rightarrow 0.05$

[9]This is compatible with different versions of common ground (management) (e.g., Chafe 1976, Krifka 2008), which affects the probabilities of D_0. In this paper, I will not be concerned with how exactly D_0 is affected by the common ground.

$Q2 \rightarrow 0.1$
$Q3 \rightarrow 0.85$

If the next word of the author in (24) was *nevertheless* instead, the probability of question Q3 would decrease. In this case, a possible question distribution D_w after (24) + *nevertheless* could be (26).

(26) $D_w = D_{\text{nevertheless}}$
$Q1 \rightarrow 0.3$
$Q2 \rightarrow 0.6$
$Q3 \rightarrow 0.1$

In order to measure the difference between the previous distribution D_0 and the distribution D_w after one or more additional words, I define QUD surprisal S_{QUD} of a unit w as the Kullback-Leibler divergence D_{KL}, also called relative entropy, between the two distributions D_0 and D_w as in (27) (following Demberg & Keller 2008: 195). The set possQ is the set of all possible QUDs that could be addressed, which has no restrictions apart from each question being syntactically well-formed. In our simplified example, possQ is the set containing the questions Q1, Q2, and Q3.

(27) $$S_{\text{QUD}}(w) = D_{KL}(D_0 || D_w) = \sum_{q \in \text{possQ}} D_0(q) \log \frac{D_0(q)}{D_w(q)}$$

S_{QUD}(w) yields a relatively high value if the previous distribution D_0 differs strongly from the distribution D_w. S_{QUD}(w) yields a relatively low value if the two distributions are similar. In my example, S_{QUD}(*nevertheless*) is higher than S_{QUD}(*because*) because the previous distribution D_0 in (23) differs more strongly from the distribution $D_{\text{nevertheless}}$ in (26) than from the distribution D_{because} in (25). This is shown in (28), where possQ is the set {Q1,Q2,Q3} and D_0, D_{because}, and $D_{\text{nevertheless}}$ are the respective distributions presented in (23–26).

(28)
$$S_{\text{QUD}}(\text{nevertheless}) = \sum_{q \in \text{possQ}} D_0(q) \log \frac{D_0(q)}{D_{\text{nevertheless}}(q)}$$
$$> S_{\text{QUD}}(\text{because}) = \sum_{q \in \text{possQ}} D_0(q) \log \frac{D_0(q)}{D_{\text{because}}(q)}$$

The consequence of this outcome would be that the connective *nevertheless* should be inserted in our example case since it strongly affects the previous question distribution. The connective *because* could be dropped since it does not have a

strong effect on the previous question distribution. In their corpus study, Asr & Demberg (2015) found that *nevertheless* was more frequently expressed explicitly than *because*.

Coming back to the author's decision between the cleft and the canonical sentence, I assume that the author considers that the addressee has some uncertainty about which question the author wants to address with an utterance. From the addressee's perspective, a canonical sentence in written German always leads to some uncertainty about which question the author wants to address given focus ambiguity (see (10) for some examples of different possible focus assignments for the same sentence). The cleft, in contrast, reduces or, in most cases, eliminates this uncertainty because of a more explicit focus marking, i.e., more explicitly marking the QUD.

Analogously to the discourse connectives above, I assume that QUD surprisal of clefting is low when the cleft addresses a relatively expected question. Hence, the words used for clefting should be dropped, in order to avoid a trough in information density. When the cleft addresses a relatively unexpected question, the QUD surprisal of clefting is relatively high and the words used for clefting should not be dropped, in order to distribute surprisal more uniformly.

For calculating the QUD surprisal of clefting, I make the simplification of treating clefting (i.e., the words *it*, *was*, and the relative pronoun) as an operator $\mathrm{op}_{\mathrm{cleft}}$ that applies to the canonical sentence, following approaches like the one by Velleman et al. (2012). The QUD surprisal of $\mathrm{op}_{\mathrm{cleft}}$, as illustrated in (29), compares the distribution $\mathrm{D}_{\mathrm{can}}$ after having encountered the canonical sentence to the distribution $\mathrm{D}_{\mathrm{cleft}}$ after adding the cleft operator to the canonical sentence.

$$S_{\mathrm{QUD}}(\mathrm{op}_{\mathrm{cleft}}) = D_{KL}(D_{\mathrm{can}} \| D_{\mathrm{cleft}}) = \sum_{q \in \mathrm{possQ}} D_{\mathrm{can}}(q) \log \frac{D_{\mathrm{can}}(q)}{D_{\mathrm{cleft}}(q)} \tag{29}$$

The difference between these two distributions is relatively small when the addition of the cleft operator to the canonical sentence in the context does not have a strong effect on the question distribution. This would mean that the question marked by clefting was also a rather likely one in the context. If clefting affected the distribution to a stronger degree, $\mathrm{S}_{\mathrm{QUD}}(\mathrm{op}_{\mathrm{cleft}})$ would be relatively large.

Using the above example (English translations repeated in (30) and (32)), I will demonstrate how this approach can explain the preference between the cleft and the canonical sentence. Consider first the previous question distributions in the two contexts, which describe the expectedness values of each question before the cleft/canonical sentence is added. I argue that this question expectedness is one of the two crucial aspects one needs to incorporate to explain the choice between

the cleft and the canonical sentence (the other aspect being the cleft's function of focus-disambiguation). In the shorter context, repeated in (30), the QUD intended to be addressed by the author (Q1: *Who ate the cookies?*) is relatively expected, i.e., easy to accommodate for the addressee.

(30) *When Lena joined the coffee break, the plate of cookies was already empty.*
   a. Bo hat die Kekse gegessen. ('Bo ate the cookies.')
   b. ? Es war Bo, der die Kekse gegessen hat. ('It was Bo who ate the cookies.')

Hence, the previous distribution D_0, which is based on the non-linguistic context and the context sentence in (30), can be assumed to assign a relatively large amount of the probability mass to Q1. A plausible, but simplified, D_0 for context (30) is provided in (31).

(31) D_0 =
Q1: Who ate the cookies? → 0.25
Q2: What did Bo eat? → 0.05
Q3: What happened then? → 0.3
Q4: What did Lena eat? → 0.3
⋮
Qn: ...

The situation looks different in the slightly longer context, repeated in (32).

(32) *When Lena joined the coffee break, the plate of cookies was already empty. She couldn't find any other cookies, either. So she decided to go to the bakery.*
   a. ? Bo hat die Kekse gegessen. ('Bo ate the cookies.')
   b. Es war Bo, der die Kekse gegessen hat. ('It was Bo who ate the cookies.')

Example (33) illustrates a plausible and simplified question distribution D_0, given the context in (32).

(33) D_0 =
Q1: Who ate the cookies? → 0.05
Q2: What did Bo eat? → 0.05
Q3: What happened then? → 0.4
Q4: What did Lena eat? → 0.4
⋮
Qn: ...

The examples (34–36) illustrate how the question distribution D_0 changes for the addressee when (i) the canonical sentence is added to context (30) (D_{can}), and (ii) the cleft operator is then added to the canonical sentence (D_{cleft}). The examples (37–39) illustrate this for context (32). All the distributions are simplified, but aim to represent reasonable probability ratios between the questions Q1–Q4. Example questions Q1–Q3 are chosen to represent questions that can be addressed by the canonical sentence *Bo ate the cookies* (with the matching focus), Q1 is, furthermore, chosen because it is addressed by the cleft *It was Bo who ate the cookies*. The question Q4 is an example of a question that is likely to be addressed in both contexts, but could not be addressed by the cleft/canonical sentence. In examples (34–39), every step is illustrated in more detail.

Examples (34) and (37) repeat the previous distributions introduced above for the two contexts. After having read the canonical sentence, I assume that only the questions which are associated with one of the possible focus markings of the canonical sentence, here Q1–Q3, receive probability mass. The previous probability ratio between these questions is maintained while the probability of all the other questions drops to 0.[10] This step is illustrated in the pairs (34)/(35) and (37)/(38), which exemplify how the previous distribution D_0 in each context differs from the distribution D_{can} after having read the canonical sentence.

The pairs (35)/(36) and (38)/(39) illustrate how the distributions change after the cleft operator has been applied to the canonical sentence. Since the cleft disambiguates focus, only one question is left to be addressed by it, Q1 in our examples. All the other questions receive a probability of 0 (or close to 0).

(34) Distribution after context (30) (one sentence)
D_0 =
Q1: Who ate the cookies? → 0.25
Q2: What did Bo eat? → 0.05
Q3: What happened then? → 0.3
Q4: What did Lena eat? → 0.3
⋮
Qn: …

(35) Distribution after canonical sentence
D_{can} =
Q1: Who ate the cookies? → 0.42
Q2: What did Bo eat? → 0.08

[10] Strictly speaking, the probably would be close to 0, not identical to 0. For reasons of simplicity, we assume it to be 0.

Q3: What happened then? → 0.5
Q4: What did Lena eat? → 0
⋮
Qn: …

(36) Distribution after clefting
D_{cleft} =
Q1: Who ate the cookies? → 1
Q2: What did Bo eat? → 0
Q3: What happened then? → 0
Q4: What did Lena eat? → 0
⋮
Qn: …

(37) Distribution after context (32) (three sentences)
D_0 =
Q1: Who ate the cookies? → 0.05
Q2: What did Bo eat? → 0.05
Q3: What happened then? → 0.4
Q4: What did Lena eat? → 0.4
⋮
Qn: …

(38) Distribution after canonical sentence
D_{can} =
Q1: Who ate the cookies? → 0.1
Q2: What did Bo eat? → 0.1
Q3: What happened then? → 0.8
Q4: What did Lena eat? → 0
⋮
Qn: …

(39) Distribution after clefting
D_{cleft} =
Q1: Who ate the cookies? → 1
Q2: What did Bo eat? → 0
Q3: What happened then? → 0
Q4: What did Lena eat? → 0
⋮
Qn: …

According to the definition in (29), the QUD surprisal of the cleft operator, $S_{QUD}(op_{cleft})$, in contexts (30) and (32) is calculated by comparing (35) to (36) and (38) to (39), respectively. The example values already indicate that the change from (38) to (39) is more drastic than from (35) to (36). In other words, the calculated value $S_{QUD}(op_{cleft})$ is higher in context (32) than in context (30). This means that, given the effect of clefting on the question distribution, it could be dropped in (30) because it is less surprising. The canonical sentence is sufficient. In (32), clefting should be inserted in order to avoid a peak in information density on the words of the canonical sentence and, thus, to distribute information more uniformly. This result is in line with what was observed by Tönnis & Tonhauser (2022) for the choice between the cleft and the canonical sentence in contexts such as (30) and (32).

These examples illustrated how the QUD surprisal of clefting is affected by the cleft's function of explicitly marking the QUD (i.e., disambiguating focus) as well as by previous expectedness values of questions. The respective last step in (36) and (39) shows the effect of focus disambiguation on the question distribution, i.e., assigning probability 1 to question Q1. However, this step only strongly affected the question distribution D_{can} when the question addressed by the cleft was not already relatively likely in D_{can}. Hence, I argue both of these aspects are relevant in order to capture the preference between a canonical sentence and its cleft variant in a context.

4 Discussion

The information-theoretic take on clefts in written German makes the correct predictions for the author's choice between the cleft and its canonical variant in contexts such as (30) and (32), just as previous discourse-dependent analyses did (e.g., Destruel & Velleman 2014, Tönnis 2021). By introducing QUD surprisal, it provides a formal analysis that can account for the discourse context dependency of this choice: A higher QUD surprisal of the clefting operator leads to a cleft while a lower QUD surprisal leads to a canonical sentence.

A huge advantage of the information-theoretic approach presented in this paper is that, besides predicting the choice between the cleft and the canonical sentence, it also provides an explanation for why the cleft is a good candidate to address relatively less expected QUDs: Clefting contributes to establishing uniform information density in discourse. In case of addressing a relatively less expected QUD, the cleft makes this question explicit and, thereby, distributes the information onto more words compared to the canonical sentence.

This explanation was missing from previous discourse-dependent analyses. Those analyses struggled to explain which aspect of the cleft caused it to behave differently in discourse than plain canonical sentences. It is not the pragmatic inferences (exhaustivity inference, existential inference), which were not affected differently in the two discourse contexts in example (30) and (32), and can, therefore, not affect the preferences, as discussed in Section 2.1 above.

One might argue that assuming that clefts are used for the purpose of focus disambiguation is already sufficient to explain the preferences between clefts and canonical sentences in discourse. I argue that a cleft does indeed help the reader to accommodate the QUD. However, it must be explained when exactly focus disambiguation is necessary, and QUD surprisal provides a measure for that: Focus disambiguation is only necessary when the author intends to address a QUD that is still relatively less expected once the canonical sentence is added. This can only occur if it was also relatively less expected in the previous distribution (before the canonical sentence was added). In such a case, focus disambiguation, modeled by assigning probability 1 to the respective question, has a strong effect on the question distribution after the canonical sentence. Therefore, the QUD surprisal of clefting is relatively high, and UID requires a more explicit marking of the QUD in order to avoid a peak in information density on the words of the canonical sentence. If the author wanted to address a relatively expected question, the QUD surprisal of the clefting operator would be relatively low given that focus disambiguation would not strongly affect the question distribution after the canonical sentence was added.

Another benefit of my approach is that it treats clefts on a par with other discourse structuring devices, which I claim could have the same effect of marking relatively less expected QUDs. One such device could be the discourse marker *übrigens* ('by the way'), as illustrated in (40).

(40) *When Lena joined the coffee break, the plate of cookies was already empty. She couldn't find any other cookies, either. So she decided to go to the bakery.*

Übrigens Bo hat die Kekse gegessen.
By the way Bo has the cookies eaten
'By the way, Bo ate the cookies.'

Adding the discourse marker *übrigens* ('by the way') also makes the continuation acceptable while the plain canonical sentence is unacceptable in this context. In this context, the marker *übrigens* ('by the way') makes explicit that a relatively less expected QUD is going to be addressed. This is another way to reduce information density at the discourse level, which should be analyzed parallel to the cleft (only that focus disambiguation does not play a role here). Previous approaches which treated clefts on a par with structurally similar constructions,

such as definite descriptions (Percus 1997), cannot account for this parallel behavior.

Moreover, the information-theoretic approach to clefts can predict why there are less clefts in spoken German than in written German, as claimed by Tönnis et al. (2018) (based on their own and informants' judgments). In spoken German, there is no or less focus ambiguity in the canonical sentence since intonation can freely be used in German to express focus. Instead of clefting, the speaker would, therefore, rather use the canonical sentence with the main accent on the subject, as in (41).

(41) *BO_F hat die Kekse gegessen.*
BO_F has the cookies eaten
'BO_F ate the cookies.'

Since subject focus is not ambiguous, a probability of 1 would be assigned to the question *Who ate the cookies?* after the canonical (41) already, and the QUD surprisal of clefting would then be very low. Hence, my analysis would frequently predict to drop clefting in spoken German.

So far, the analysis presented in this paper does not make any predictions about how each single word of the cleft or canonical sentence affects QUD surprisal, given that I assumed clefting to be just one operator. For discourse connectives, such as *because* or *übrigens* ('by the way'), QUD surprisal can be calculated equally well as, for example, relational surprisal by Asr & Demberg (2015). More complex discourse structuring devices such as the cleft are more challenging if one intends to calculate QUD surprisal incrementally. Of course, one would eventually want to be able to account for the fact that the cleft is not processed by first reading the canonical sentence and only afterwards encountering the cleft operator. I leave this issue for future research.

An anonymous reviewer pointed out that, instead of treating the addressed QUD as relatively less expected, it could be just the cleft pivot, *Bo* in the above example, that is surprising. As mentioned in Section 2.3, this does not apply to the kind of examples discussed in this paper. However, there might be other uses of clefts where UID does not apply at the discourse level but at the sentence level, as indicated in (42).

(42) *Gestern war ich in der Kirche. Es waren aber nicht nur die üblichen Verdächtigen da.*
'Yesterday I was at church. But not only the usual suspects were present.'

Es war der Papst, der uns begrüßt hat.
'It was the pope who greeted us.'

In this example, it is not particularly unexpected that the QUD *Who greeted you/us?* is addressed at this point. However, one could assume that the word *pope* is surprising in this context. Therefore, this example could be explained by assuming that the words used for clefting are inserted to reduce information density at the sentence level instead of the discourse level. Hence, the present analysis cannot be generalized to all uses of clefts. At least, it applies when the cleft is used in its discourse structuring function of marking a relatively less expected QUD.

Last but not least, I want to come back to the inferences discussed in Section 2.1. The current approach is not aiming to derive the existential or exhaustive inference of clefts. Nevertheless, it is not in conflict with the existence of such inferences for many occurrences of clefts. I see a potential for future research to investigate what would follow from the information-theoretic and question-based approach for the exhaustivity inference in particular (see Velleman et al. 2012, Pollard & Yasavul 2015 and De Veaugh-Geiss et al. 2018 for approaches to cleft exhaustivity using the QUD framework).

5 Conclusions

In this paper, I presented a new phenomenon, besides connectives, which requires information theory at the discourse level. Building on expectation-based accounts of clefts (e.g., Destruel & Velleman 2014, Tönnis 2021), I analyzed clefts in written German as a device to reduce information density in discourse by relying on its function of disambiguating focus.

The proposed analysis was based on the assumption that the expectations of the author and addressee can be modeled as a probability distribution over questions that could be addressed, which is updated after each new sentence of the text or conversation (following Kehler & Rohde 2017, Tönnis 2021). Accordingly, the proposed analysis incorporated the concept of QUD surprisal (inspired by Demberg & Keller 2008, Asr & Demberg 2015), which measures the difference between the question distribution after having read the canonical sentence in a context and the question distribution after the cleft operator is applied to the canonical sentence. If the QUD surprisal of clefting is high, the extra marking provided by the cleft is required in order to satisfy UID. Ideally, this extra marking could also be achieved by a different means than the cleft, for instance, by adding the discourse marker *by the way*. As a consequence, this analysis treats clefts on a par with other constructions that reduce information density in discourse.

Previous analyses of clefts, such as those focusing on the semantic/pragmatic inferences, were shown to have problems accounting for the examples discussed

in this paper, where the cleft is used to address a relatively less expected QUD. Furthermore, the information-theoretic approach to clefts does not only make the correct predictions for the choice between the cleft and its canonical variant, but it also provides an explanation for why the cleft has the discourse function of marking a relatively less expected QUD.

Acknowledgements

I thank the audience of the DGfS 2022 workshop *Discourse obligates*, the audience of the linguistic colloquium at the University of Stuttgart, Judith Tonhauser, Lisa Schäfer and Edgar Onea for valuable feedback on this work.

References

Asr, Fatemeh Torabi & Vera Demberg. 2012. Implicitness of discourse relations. In *COLING 2012*, 2669–2684. Mumbai: The COLING 2012 Organizing Committee. https://aclanthology.org/C12-1163.

Asr, Fatemeh Torabi & Vera Demberg. 2015. Uniform Surprisal at the level of discourse relations: Negation markers and discourse connective omission. In *The 11th international conference on computational semantics*, 118–128. London: Association for Computational Linguistics. https://aclanthology.org/W15-0117.

Beaver, David I. & Brady Z. Clark. 2008. *Sense and sensitivity: How focus determines meaning*. Susan Rothstein (ed.), vol. 5 (Explorations in Semantics). Oxford: Wiley-Blackwell. DOI: 10.1002/9781444304176.

Büring, Daniel & Manuel Križ. 2013. It's that, and that's it! Exhaustivity and homogeneity presuppositions in clefts (and definites). *Semantics and Pragmatics* 6(6). 1–29. DOI: 10.3765/sp.6.6.

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 27–55. New York: Academic Press.

De Veaugh-Geiss, Joseph P., Swantje Tönnis, Edgar Onea & Malte Zimmermann. 2018. That's not quite it: An experimental investigation of (non-)exhaustivity in clefts. *Semantics and Pragmatics* 11. DOI: 10.3765/sp.11.3.

De Veaugh-Geiss, Joseph P., Malte Zimmermann, Edgar Onea & Anna-Christina Boell. 2015. Contradicting (not-)at-issueness in exclusives and clefts: An empirical study. In *Semantics and Linguistic Theory (SALT)*, vol. 25, 373–393. DOI: 10.3765/salt.v25i0.3054.

Demberg, Vera & Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2). 193–210. DOI: 10.1016/j.cognition.2008.07.008.

Destruel, Emilie, David I. Beaver & Elizabeth Coppock. 2019. It's not what you expected! the surprising nature of cleft alternatives in French and English. *Frontiers in Psychology* 10. 1–16. DOI: 10.3389/fpsyg.2019.01400.

Destruel, Emilie & Leah Velleman. 2014. Refining contrast: Empirical evidence from the English *it*-cleft. *Empirical Issues in Syntax and Semantics* 10. 197–214. http://www.cssp.cnrs.fr/eiss10/.

É. Kiss, Katalin. 1998. Identificational focus versus information focus. *Language* 74(2). 245–273. DOI: 10.1353/lan.1998.0211.

Féry, Caroline. 2017. *Intonation and prosodic structure* (Key Topics in Phonology). Cambridge: Cambridge University Press. DOI: 10.1017/9781139022064.

Fodor, Janet Dean. 2002. Prosodic disambiguation in silent reading. In *North East Linguistic Society (NELS)*, vol. 32, 113–132. https://scholarworks.umass.edu/nels/vol32/iss1/8.

Genzel, Dmitriy & Eugene Charniak. 2002. Entropy rate constancy in text. In *40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 199–206. DOI: 10.3115/1073083.1073117.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307. DOI: 10.2307/416535.

Halvorsen, Per-Kristian. 1978. *The syntax and semantics of cleft constructions*. University of Texas. (PhD thesis).

Horn, Laurence. 1981. Exhaustiveness and the semantics of clefts. In *North East Linguistic Society (NELS)*, vol. 11, 125–142. https://scholarworks.umass.edu/nels/vol11/iss1/10.

Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61(1). 23–62. DOI: 10.1016/j.cogpsych.2010.02.002.

Kehler, Andrew & Hannah Rohde. 2017. Evaluating an expectation-driven question-under-discussion model of discourse interpretation. *Discourse Processes* 54(3). 219–238. DOI: 10.1080/0163853X.2016.1169069.

Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3–4). 243–276. https://www.jstor.org/stable/26190905.

Krifka, Manfred & Renate Musan. 2012. Information structure: Overview and linguistic issues. In Manfred Krifka & Renate Musan (eds.), *The expression of information structure*, vol. 5, 1–43. Berlin: De Gruyter. DOI: 10.1515/9783110261608.1.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177. DOI: 10.1016/j.cognition.2007.05.006.

Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In *Advances in Neural Information Processing Systems 19: The 2006 Conference*, 849–856. https://ieeexplore.ieee.org/document/6287477.

Onea, Edgar. 2019. Exhaustivity in *it*-clefts. In Chris Cummins & Napoleon Katsos (eds.), *The Oxford handbook of experimental semantics and pragmatics*, 401–417. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780198791768.013.17.

Percus, Orin. 1997. Prying open the cleft. In *North East Linguistic Society (NELS)*, vol. 27, 337–352. https://scholarworks.umass.edu/nels/vol27/iss1/25.

Pollard, Carl & Murat Yasavul. 2015. Anaphoric *it*-clefts: The myth of exhaustivity. In Ross Burkholder, Carlos Cisneros, Emily R. Coppess, Julian Grove, Emily A. Hanink, Hilary McMahan, Cherry Meyer, Natalia Pavlou, Özge Sarıgül, Adam Roth Singerman & Anqi Zhang (eds.), *Proceedings of the fiftieth annual meeting of the Chicago Linguistic Society*, 381–393. Chicago: The Chicago Linguistic Society.

Prince, Ellen F. 1978. A comparison of *wh*-clefts and *it*-clefts in discourse. *Language* 54(4). 883–906. DOI: 10.2307/413238.

Reinhart, Tanya. 1995. *Interface strategies*. Utrecht: OTS working papers in linguistics.

Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6). 1–69. DOI: 10.3765/sp.5.6.

Rochemont, Michael. 1986. *Focus in generative grammar* (Studies in Generative Linguistic Analysis). Amsterdam: John Benjamins. 1–231. DOI: 10.1075/sigla.4.

Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27(3). 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Simons, Mandy, David I. Beaver, Craige Roberts & Judith Tonhauser. 2017. The best question: Explaining the projection behavior of factives. *Discourse Processes* 54(3). 187–206. DOI: 10.1080/0163853X.2016.1150660.

Szabolcsi, Anna. 1981. The semantics of topic-focus articulation. In J. A. G. Groenendijk, T. M. V. Janssen & M. B. J. Stokhof (eds.), *Formal methods in the study of language*, 513–540. Amsterdam: Mathematisch Centrum.

Szendrői, Kriszta. 1999. A stress-driven approach to the syntax of focus. *UCL Working Papers in Linguistics* 11. 545–573.

Tönnis, Swantje. 2021. *German "es"-clefts in discourse: A question-based analysis involving expectedness*. Graz University. (Doctoral dissertation).

Tönnis, Swantje, Lea M. Fricke & Alexander Schreiber. 2018. Methodological considerations on testing argument asymmetry in German cleft sentences. In Eric Fuß, Marek Konopka, Beata Trawiński & Ulrich H. Waßner (eds.), *Grammar and Corpora 2016*, 231–240. Heidelberg: Heidelberg University Publishing. DOI: 10.17885/heiup.361.c4703.

Tönnis, Swantje & Judith Tonhauser. 2022. German clefts address unexpected questions. In John R. Starr, Juhyae Kim & Burak Öney (eds.), *Semantics and Linguistic Theory (SALT)*, vol. 32, 661–684. DOI: 10.3765/salt.v1i0.5359.

Velleman, Leah, David I. Beaver, Emilie Destruel, Dylan Bumford, Edgar Onea & Liz Coppock. 2012. *It*-clefts are IT (inquiry terminating) constructions. In *Semantics and Linguistic Theory (SALT)*, vol. 22, 441–460. DOI: 10.3765/salt.v22i0.2640.

Westera, Matthijs & Hannah Rohde. 2019. Asking between the lines: Elicitation of evoked questions in text. In *The 22nd Amsterdam Colloquium*, 397–406. https://archive.illc.uva.nl/AC/AC2019/Proceedings/index.html.

Chapter 7

Tell me something I don't know: Speaker salience and style affect comprehenders' expectations for informativity

Vilde R. S. Reksnes, Alice Rees, Chris Cummins & Hannah Rohde
University of Edinburgh

A comprehender's estimate of what events or situations are typical in the world is distinct from their estimate of what a speaker is likely to report on. Comprehension and production studies have shown contradicting preferences for which types of estimates are used by comprehenders and by speakers: Typicality is favoured in comprehension (e.g. real-world typical content is associated with processing ease), whereas speakers' production choices favour the inclusion of surprising or informative content (i.e. easily inferable or typical content is disfavoured). We posit that comprehenders are aware of and make use of speakers' production preferences when anticipating upcoming content. In two studies, we elicit sentence completions as an index of comprehenders' expectations about upcoming material and evaluate the informativity of these completions (their object typicality, presence of modification or negation, and information theoretic entropy and relative entropy scores). Experiment 1 manipulated the salience of the speaker and found that increased emphasis on the speaker led to an increase in informativity, showing that the more aware comprehenders are made of an intentionally communicating speaker, the more their expectations favour upcoming words that would yield an informative utterance. Experiment 2 further tested the malleability of this informativity bias by familiarising participants with two speakers who differ in the informativity of their utterances. When completing utterances from the two speakers, comprehenders provide more informative completions for the high-informativity speaker, showing that comprehenders are able to adapt their expectations for informativity to individual speakers' communicative styles. This sensitivity to speakers' production preferences highlights a role for informativity-driven reasoning about the speaker in models of language processing.

Vilde R. S. Reksnes, Alice Rees, Chris Cummins & Hannah Rohde. 2024. Tell me something I don't know: Speaker salience and style affect comprehenders' expectations for informativity. In Robin Lemke, Lisa Schäfer & Ingo Reich (eds.), *Information structure and information theory*, 177–214. Berlin: Language Science Press. DOI: 10.5281/zenodo.13383795

1 Introduction

The question of how comprehenders generate expectations regarding upcoming content has received much attention in the field (see e.g. Kuperberg & Jaeger 2016 for a review). In a discourse context, expectations about upcoming content can be understood to reflect what comprehenders believe speakers will choose to talk about. On the one hand, speakers can choose to use language to talk about the world that they encounter, producing utterances that describe the kinds of situations that arise in their daily lives. On the other hand, speakers can choose to convey content that is worth talking about, producing utterances that they think a listener will find interesting and newsworthy.

The first choice characterises what we will call *transparent* language use, whereby speakers use language to directly narrate what happens to them. If speakers use language transparently, comprehenders ought to come to expect language that describes the kinds of situations that arise frequently and plausibly in the real world. The second choice characterises *filtered* language, whereby speakers limit their utterances to those whose content is novel and informative, inducing listeners to expect descriptions of situations which are interesting by virtue of their infrequency. If speakers do indeed filter available content and do so in ways to achieve informativity-driven communicative goals, and if comprehenders are aware of a speaker's role in this filtering, a prediction arises that a context that increases the emphasis on the speaker (in their role as a filter of potential content to convey) should increase comprehenders' expectations for content about infrequent situations. We will argue that prior work on comprehenders' expectations often focuses on expectations about transparent language use, missing an opportunity to find evidence of expectations that are driven by comprehenders' awareness of the speaker as someone engaged in intentional communication with informativity-driven goals.

A long-standing claim is that when a comprehender makes guesses about what words are coming next, they rely on their knowledge about the world. Studies have shown that knowledge of what is typical of real-world situations is active during language processing and thus typicality has been linked to processing ease in comprehension. Here, we define typicality as the frequency of a situation or event given the context (e.g. trains are typically present at a train station, a beach typically has sand). As an example of comprehenders' anticipation of upcoming content, Kamide et al. (2003) use a visual world eye-tracking methodology and show that participants expect utterances to convey real-world typical content. For example, for an image depicting a man, a motorbike and a carousel,

participants hear *The man will ride the…* and look to the motorbike before the noun is uttered, anticipating a continuation about the most real-world typical object in the scene (the motorbike). That work is in keeping with findings about the processing costs associated with encountering surprising events. For example, Kutas & Hillyard (1980) show that strong semantic incongruity caused a peak in the N400 relative to moderate incongruity (*taking a sip from the transmitter* vs *taking a sip from the waterfall*). Measures of sentence recall (Marks & Miller 1964) and reading with eye-tracking (Morris 1994) corroborate these effects.

There are studies showing that real-world typicality can be overridden. For example, Nieuwland & Van Berkum (2006) showed processing ease for a real-world non-typical event, namely a peanut singing about being in love, compared to a real-world typical event, namely a peanut being salted (see also Troyer & Kutas 2020). However, the important thing to note here is that comprehenders adjusted their expectations to fit the constraints of a fictional world, one where peanuts sing about their new girlfriends and their amorous feelings towards them. In other words, they still show a bias towards the typical, but this bias is governed by what is typical in the new "real world". Overall, these studies suggest that comprehenders have a preference for typicality when they are processing language, i.e. a bias towards transparent language use.

Conversely, research on production shows that speakers do impose a filter on their language use: They tend to make production choices that favour the inclusion of surprising or otherwise informative content about a situation (e.g. Brown & Dell 1987, Lockridge & Brennan 2002). In doing so, they show a dispreference for real-world typical content: Speakers will often choose to omit altogether information that is inferable or typical, and instead highlight non-typical information. In Brown & Dell's (1987) classic production study, participants read short stories that involved either a typical instrument or an atypical instrument, for example, a stabbing that happened either with a knife or an icepick. Participants were then asked to retell the stories to test whether they chose to mention the instrument in their event description. The results showed a preference to omit mention of the instrument if it was typical (knife being a typical stabbing instrument) and to include a mention of the instrument if it was atypical (icepick being an atypical stabbing instrument). In keeping with this pattern whereby inferable or typical content can be omitted or reduced, there is evidence that high frequency words are produced with shorter acoustic duration than low frequency words (Aylett & Turk 2004; see also Levy & Jaeger 2007, Jaeger 2010, Kravtchenko 2014, Lemke 2021). In other contexts, speakers are shown to include colour adjectives more often when referring to an object if the colour is atypical, such as a pink banana, than if it is typical (Sedivy 2003, Westerbeek et al. 2015, Rubio-Fernández

2016, Degen et al. 2020), and to mention the material of an object when atypical, such as a wool bowl compared to a ceramic bowl (Mitchell et al. 2013). Notably, Degen et al. (2020) found that speakers are less likely to mention the colour of an object if it is typical (e.g. a yellow banana) even when the colour is necessary for disambiguating the intended referent (i.e. when a brown banana is also present). The authors take this to indicate that the speakers are relying on the listener to infer that the most typical object was intended. In contrast, if the target object was an atypical colour (a blue banana, in this case), the colour was always mentioned when a competitor object was present.

The pictures that emerge from these two bodies of literature thus seem to conflict. Evidence from comprehension studies strongly suggests that comprehenders estimate upcoming content with a bias in favour of sentences about typical situations (transparent language use), while results from production studies demonstrate a preference for informativity or newsworthiness in content selection (filtered language use). According to this literature, it would therefore appear as if comprehenders' expectations are out of line with the predicted behaviour of speakers according to Gricean accounts; that is, speakers' contributions are expected to be appropriately informative and relevant (Grice 1975). However, if comprehenders are rational and sensitive to speakers' production preferences, these preferences should also bear upon their estimates regarding upcoming content: That is, comprehenders should expect the kind of content that cooperative speakers are likely to mention, rather than the kind of content that is likely to be the case in the real world. For example, you would probably expect there to be trains in a train station but generally not expect someone to tell you that fact out of the blue. However, if you receive a call from a friend who is at your local train station and they tell you the Hogwarts Express has just pulled up, you would be surprised by the fact that this particular train has made a visit to your local station, but likely not by the fact that your friend chose to communicate the information, given that the situation occurred.

On this view, a comprehender's estimate of what is typical in the world is distinct from their estimate of what the speaker is likely to say (Rohde et al. 2021). In principle, either of these estimates might be activated when a comprehender makes guesses about an upcoming utterance. In the current study, we examine the nature of comprehenders' predictions by using a sentence completion paradigm (a so-called Cloze task; Taylor 1953). We use this task to elicit specific sentence completions to assess what words comprehenders predict a speaker will produce, rather than testing whether comprehension is speeded/slowed by available content, as in previous work. Although the Cloze task involves a participant

producing words, it can also be understood as a comprehension-oriented process, in that the task probes what a comprehender thinks is coming next. This approach follows work on coreference that often uses sentence completion tasks to estimate comprehenders' interpretation biases and their expectations about next mention (e.g. Stevenson et al. 1994). In fact, the original Cloze study describes the task in comprehension terms as an "attempt to reproduce accurately a part deleted from a 'message' (any language product) by deciding, from the context that remains, what the missing part should be" and notably describes the participants as readers/listeners (Taylor 1953: 416).

In Experiment 1, we manipulate the emphasis on the presence of a speaker to test whether participants' estimates are modulated by how salient the speaker is. If this manipulation is sufficient to make participants consider the speaker's production process, their guesses about upcoming content should be influenced by their inferences about what the speaker's goals are and why they are choosing to speak. For example, completing the sentence *I'm at the train station, and there's* ____ with the words *a train* yields a description of a likely situation, in that a scenario with a train would be a typical occurrence given the location. However, if comprehenders consider what a speaker would deem worth reporting on, they might not expect such an utterance, precisely given the high typicality of the scenario described. The presence of a train is so typical as to be inferable without mention, so a comprehender might assume that a speaker would not choose to mention it. Consequently, something more informative such as *street performers* or a similarly low-typicality occurrence might be a more likely completion when the presence of a speaker is emphasised.

An emerging line of research shows that comprehenders do keep track of speakers' production preferences. Rohde et al. (2021) established the presence of informativity-driven effects in comprehension, showing that sentences containing newsworthy content were preferred in certain contexts over those that contained real-world typical content. In particular, they observed processing ease for atypical content in a natural dialogue setting with no contextual manipulations; participants were faster reading a newsworthy message about socks that cost $100 than a message about more standardly priced $2 socks. Rohde et al. take this to demonstrate that a communicative context with an intentionally communicating speaker behind an utterance is sufficient to induce a comprehension preference in favour of newsworthy content. Their studies show that comprehenders experience difficulty in integrating inexplicably uninformative content, but do not test what content comprehenders specifically expect or what factors might modulate such expectations. The current study aims to examine the nature of this predicted content in a more fine-grained way.

Corroborating the findings in Rohde et al. (2021), Rohde et al. (2022) conducted a series of forced-choice tasks in which participants were asked to estimate what a character in a story would say or what they might think; they found that the character's choice to produce an utterance unprompted, as opposed to responding to a question or thinking uncommunicated thoughts, cued participants to estimate a more atypical meaning, demonstrating that comprehenders favour sentences that convey newsworthy information. In fact, when comprehenders come across content that does not meet this standard for newsworthiness, they try to reconcile the absence of sufficiently informative content by recasting the utterance in a way that makes it an informative contribution (Kravtchenko & Demberg 2022). In Kravtchenko & Demberg's experiment, participants read short stories about stereotyped activities. In general, comprehenders' script knowledge about such activities allows speakers to omit typical or inferable information. Upon encountering an utterance about an easily inferable event, e.g. *Mary ate* in the context of going to a restaurant, participants made inferences that this generally typical action was nonetheless unusual for the protagonist. In other words, the study highlights cases in which a discrepancy emerges between the understanding of a speaker as having goals to be cooperatively informative and the uncooperative typicality of the situation being communicated. This discrepancy can be resolved if comprehenders change their prior beliefs about what was typical in this context, thereby ensuring that the utterance meets expectations for informativity.

Relatedly, Lemke et al. (2021b) make the point that script knowledge provides a way to quantify the likelihood of upcoming content as it may approximate extralinguistic knowledge, which is used when language users make predictions about upcoming events and, thereby, content. They argue for the importance of such knowledge when modelling anticipation, specifically for the production of ellipsis and fragments (Lemke et al. 2021a, Schäfer et al. 2021). The extralinguistic context is particularly relevant discourse-initially when linguistic context is not available to inform comprehenders' anticipations. Venhuizen et al. (2019) similarly argue that world knowledge has not been hitherto sufficiently incorporated in computational models of comprehension. They develop a model that derives a preference for informative content in processing. The studies presented here expand on such work by assessing the effect of speaker awareness, another aspect of extralinguistic knowledge, on comprehenders' expectations.

Awareness of the speaker can be understood to underlie comprehenders' guesses about the reason behind a speaker's production choice. In particular, there is a longer standing body of work that assesses comprehenders' guesses about

the reason behind a speaker's production choice, and in particular what referential expressions speakers choose when tasked with picking out a specific object among a set of possible referents (see e.g. Davies & Arnold 2019 for a review). Many of these studies show comprehenders reasoning specifically about the informativity of a speaker's utterances (e.g. Grodner & Sedivy 2011, Pogue et al. 2016, Ryskin et al. 2019). For example, if a speaker reliably uses adjectives contrastively (e.g. produces *click on the tall glass* when there is another glass present), participants are able to use that information to identify the referent more quickly; however, participants stop relying on the adjective to indicate contrast if a speaker habitually uses adjectives non-contrastively (e.g. produces *click on the tall glass* even when there is only one glass; Grodner & Sedivy 2011). Although this work shows comprehenders adjusting their expectations to different speakers with regards to informativity, these are specifically referential expectations; the pragmatic reasoning does not tap into expectations in the way we outline above, namely expectations for the selection of sentence content. Instead, this work tends to focus on tasks in which the communicative goal is referential success – the comprehender knows that the speaker has an intended referent they need to talk about and the speaker's choice lies in what forms to use for describing the referent. A comprehender can thus use their awareness of the speaker's referential communicative goal when interpreting the message (Sedivy 2003).

In contrast, the studies presented here test expectations for message content when an utterance is conveyed out of the blue, i.e. when content newsworthiness is more likely to be the goal rather than referential identification within a larger utterance. In this way, we aim to assess comprehenders' expectations for informativity more generally: What content do comprehenders expect cooperative speakers to mention? The current study also departs from the reference resolution studies discussed above in how the effect of speaker identity is manipulated. Past work on reference has examined this variable by contrasting two types of speakers, those who are pragmatically reliable and those who are unreliable. While our Experiment 2 implements an analogous manipulation to this, contrasting an informative and an uninformative speaker, our Experiment 1 investigates a more fundamental contrast by comparing the presence vs. absence of a speaker in order to test whether comprehenders' awareness of the speaker can itself influence their expectations about content newsworthiness.

There are current models of language use that, in line with the Gricean approach, take into account listeners' expectations for speakers to make informative contributions. For example, the Rational Speech Act (RSA) framework models a speaker who makes production choices based on a listener who in turn is

able to make inferences about the speaker and their communicative goals (see e.g. Goodman & Frank 2016 for an overview). The RSA model captures the interplay between speakers and listeners and how they reason about each other's linguistic choices. It follows that the speaker plays a key role when a listener processes an utterance. This point is tested explicitly by Kreiss & Degen (2020), who build an RSA model based on empirically elicited referring expressions from speakers to predict listeners' behaviour in a subsequent referential identification experiment with or without contrast objects present (e.g. another banana when the target is a yellow banana). They show that when listeners hear partial referring expressions (e.g. *click on the yellow*) they make inferences about how likely a speaker is to use a modifier based not only on the presence of a contrast object, which is what other contrastive inference studies have consistently found, but also on the typicality of a competitor (e.g. a yellow strawberry). Their RSA model accurately predicts this listener behaviour. In other words, listeners identify the intended referent based not only on linguistic or cognitive factors related to the referents themselves, but also based on the listener's reasoning about the speaker's production probabilities. Note that again, this study is one where referential success is the goal of the communicative interaction. The current work is consistent with the principles of RSA, and adds to this framework by explicitly testing the salience of the speaker and its effect on comprehenders' expectations for content.

In sum, whereas a range of prior studies have emphasised the importance of (real-)world typicality, an emerging body of work aims to bring attention to the role informativity plays in processing. Our study adds to this informativity-driven account of processing, and shows that expectations for informativity also matter for the estimates comprehenders make regarding upcoming content. In anticipating what someone will say next, comprehenders may be tracking several distributions of probabilities about the speaker, such as the probability of particular situations that a speaker might encounter, the likelihood of choosing to formulate an utterance about that situation (as opposed to staying silent), and even the choice among available formulations for expressing that meaning. On this view, anticipating upcoming content must be understood to rely on (at least) two components: One concerns the situation and one concerns the speaker and how and whether they will filter the scene. We argue that in order to accurately capture the processes at play, there needs to be a role for informativity-driven reasoning about the speaker in models of language processing.

2 Experiment 1: Awareness of speaker intentions in comprehension

Our main thesis, in keeping with the findings discussed above, is that comprehenders can and do take into account speakers' production preference to be informative during processing. This in turn means that certain contexts may emphasise the relevance of speakers' production preferences because they highlight the presence of the speaker themselves. The prediction is that increasing the salience of an intentionally communicating speaker should lead to an increase in comprehenders' expectations for informative content. Experiment 1 tests this claim using a Cloze task paradigm to elicit sentence completions as an indicator of what the participant expects a speaker would say. Specifically, we ask participants to complete statements about what is present at a particular location, e.g. a train station. A total of 20 different locations were used.[1] In a between-participants design we manipulate the salience of the speaker across four conditions. As shown in Figure 1, all conditions use the *there's* ____ prompt. The BARE condition only mentions the location (e.g. *At the train station, there's* ____); the THIRD PERSON condition describes someone at the location (*They're at the train station, and there's* ____); the FIRST PERSON condition directly mentions the speaker (*I'm at the train station, and there's* ____); and the VISIBLE SPEAKER condition uses first person and adds a photograph of a person speaking, with the text prompt being identical to the FIRST PERSON condition. The manipulation is intended to vary the perceived communicative intent of the context, such that our most speaker-salient condition, VISIBLE SPEAKER, is the one where communication is most strongly foregrounded by emphasising the presence of a speaker who has uttered the sentence. The expectation is that such contexts will elicit more completions about the presence of non-typical or otherwise unexpected entities in the target location.

If speaker salience influences comprehenders' guesses, the least speaker-salient condition is predicted to reveal comprehenders' reliance on real-world knowledge[2] of typical entities or situations that are likely to be present at the different locations, in keeping with prior comprehension studies demonstrating comprehenders' reliance on real-world knowledge (e.g. *a train* for the train station location). We in turn expect an increase in informative completions across the four

[1]The locations are: Bakery, library, forest, living room, beach, petrol station, cinema, office, playground, mountain, bathroom, park, bedroom, post office, train station, garden, golf course, restaurant, hospital, and stationery shop.

[2]Barring any construction of a fictional world, we assume comprehenders tend to rely on their real-world knowledge for their typicality estimates.

Figure 1: Example target sentences for the train station location in each of the four conditions

conditions, with the most informativity expected in the VISIBLE SPEAKER condition (e.g. *a celebrity*). The possibility that the less speaker-salient conditions will be distinguished from each other is predicted by work in embodied cognition showing that people are sensitive to the use of different pronouns when taking perspective; people adopt an actor's perspective when first person is used, and an external perspective with the third person (Borghi & Cimatti 2010). Our goal in including a gradation of speaker salience conditions is to test whether – and via what properties – an emphasis on the speaker can be achieved. These manipulations themselves are less of interest than the possibility that, if a speaker-oriented perspective can be achieved, comprehenders' guesses of what that speaker will say next may favour utterances conveying more informative content. We will be using a set of measures to capture different senses of informativity (see measures and data analysis below). An increase in informativity would indicate that rather than simply invoking typicality when they anticipate upcoming content, comprehenders make use of their awareness of speakers' production preferences and filter the possible options through the lens of the speaker – in that way estimating what content a cooperative speaker would consider worth uttering.

2.1 Methods

2.1.1 Participants

400 participants with English as their first language, no known language impairments, location in the US or the UK, and of minimum 18 years of age (M = 36.09, SD = 11.73, range = 18–78), were recruited on Prolific (https://www.prolific.com/). We tested an equal number of participants in each of the four conditions, i.e. 100 participants per condition, each providing 20 data points. Participants who failed two or more attention checks or reported another language than English as their first language were excluded and more participants were recruited to reach the desired number per condition. Participants were paid £7.12/hour on average.[3] In addition, 22 participants were recruited through the University of Edinburgh's Experiment Volunteer Panel for a pre-test, used to assess the typicality of a variety of objects in the item locations (see measures). These participants were students who received course credits as compensation.

2.1.2 Design

Experiment 1 tests whether speaker salience influences comprehenders' expectations about sentence completions. In order to implement this speaker salience manipulation, each of the four conditions increasingly highlight the speaker, as exemplified in Figure 1. The BARE condition only mentions the location; the THIRD PERSON condition invokes a speaker talking about someone at the location; the FIRST PERSON condition directly mentions the speaker; and the VISIBLE SPEAKER condition uses first person and adds a photograph of a person speaking, with the sentence embedded in a speech bubble. This last condition is intended as the one where communicative intent and therefore awareness of speakers' production preferences is most emphasised.

Participants in the BARE, THIRD PERSON and FIRST PERSON conditions are not given any information about the communicative context, whereas participants in the VISIBLE SPEAKER condition are told that each utterance is the beginning of a phone call where the speaker has called someone to tell them something. In other words, the speaker has not been prompted by a question and is speaking out of the blue.

[3] The details of the cost were established by Prolific based on the median time taken to complete the experiment per submission for each group.

2.1.3 Materials

The 20 target items each mentioned a different location (see Footnote 1). Locations were chosen to be places that would have an adequate number of typical and therefore expected objects present (e.g. trains, platforms, passengers, etc. for train station) but also allowed for the possibility of less typical entities being mentioned without raising doubts about the speaker's reliability.

Each target sentence followed the same format within a condition, mentioning a location in the first clause and providing a sentence-final blank space for the participant to fill. The target sentence templates for each condition were as follows: BARE: *At the [location], there's* ____, THIRD PERSON: *They're at the [location], and there's* ____, FIRST PERSON and VISIBLE SPEAKER: *I'm at the [location], and there's* ____ (all exemplified with *train station* in Figure 1). The VISIBLE SPEAKER condition additionally displayed the target sentence in a speech bubble alongside a picture of a person talking on the phone. A total of 20 different speaker images were used, 10 portraying female speakers and 10 portraying male speakers. Speaker images were allocated to specific items, but participants only saw each speaker for a single target item (e.g. the man in Figure 1 is seen talking about the train station location). The speaker images were reused in the fillers, so that a given list in the VISIBLE SPEAKER condition contained the same speaker image a total of three times. The experiment included 40 fillers in each condition. For BARE, THIRD PERSON and FIRST PERSON conditions, fillers were adapted Cloze task items from Altarriba et al. (1996) with the blank space appearing either initially, in the middle or at the end of the sentence. For VISIBLE SPEAKER, fillers were created that were more natural-sounding as conversation-initial utterances. Two fillers in each condition served as catch trials as they had only one or two likely completions, e.g. *I'm going to the swimming* ____ *this afternoon* in VISIBLE SPEAKER condition, where *pool* is the expected completion. See OSF for a full list of materials for the VISIBLE SPEAKER condition.[4]

2.1.4 Procedure

The experiment was presented using Qualtrics (https://www.qualtrics.com/). For BARE, THIRD PERSON and FIRST PERSON conditions, participants were told they would read sentences where a part was missing and their task was to type in the word or words they thought should be in the sentence. We suggested an upper limit of 3 words in the instructions at the beginning of the experiment with the aim of discouraging participants from expanding the sentence beyond a

[4]https://osf.io/7h5qs/

single target noun phrase. For VISIBLE SPEAKER, the instructions also emphasised the communicative intent of the sentences by telling participants that they were seeing someone who had just called up a friend and their task was to fill in the missing part with what they thought the speaker could have said. They were also instructed that what they saw was the first thing said in the conversation, to further highlight that these are unprompted utterances. Items were presented in fully randomised order. Participants were presented with one item at a time with a text field below the item stating *Fill in the blank*, in which they typed in their completions. They then clicked an arrow at the bottom right of the screen to proceed to the next item. Any one participant saw only one condition, i.e. a total of 60 items (20 of which were target items). The median completion times in each condition were between 10 and 15 minutes.

2.1.5 Measures and data analysis

Prior to data analyses, responses were standardised, i.e. manually cleaned to have matching case, punctuation and spelling, initial articles (*a/an/the*) were removed, and any spelling differences were collapsed (e.g. *doughnut/donut*). Where participants listed more than one response (e.g. *doctors/nurses*), only the first item mentioned was included. For each response, the main noun was identified as well as any modification and negation used. A total of 10 continuations were non-meaningful and therefore excluded from analysis. The continuations included ranged from 1 word in length to 11 words, with a mean of 1.61.

The pre-test was a non-communicative task in which a separate group of participants (N = 22) were asked to list a minimum of 3 and maximum of 10 things they would be likely to find at each of the locations mentioned in the target sentences. The pre-test provides a measure of what is considered typical at the different locations. Main nouns were extracted from responses and collated into ranked lists, ranging from highly expected/common (e.g. *train* for *train station*) to less expected/common (e.g. *bustle, musician, pigeon*). Each noun in a given location was then given a typicality score estimated as the proportion of pre-test participants who mentioned that noun (e.g. for the train station location, 0.91 and 0.00 for *train* and *delay*, respectively).

The informativity of participants' completions is assessed with five measures. Each of these measures is intended to capture a different sense of how a continuation provided in the experiment may be informative:

1. Variability of responses (mean entropy score per condition to compare consistency versus unpredictability of responses)

2. Kullback-Leibler (KL) divergence of responses (i.e. relative entropy of responses with BARE condition as baseline)

3. Inclusion of modification (which may make otherwise typical content like a train more newsworthy, e.g. *steam train*)

4. Inclusion of negation to mark the absence of something (often something typical, e.g. *no train*)

5. Typicality of objects mentioned (compared to responses elicited in the pre-test; specifically, how many participants in the pre-test listed object X for location Y)

Entropy is an information-theoretic measure for quantifying the variation in a distribution of outcomes. As such, entropy provides a measure of the average amount of information needed to represent an outcome drawn from a given distribution. In a condition where participants' sentence continuations are strongly biased to a few possible outcomes, the entropy for that distribution of outcomes will be low (i.e. a given outcome is, on average, not very informative as it is highly predictable). Conversely, in a condition with more unique outcomes or more variability across outcomes, the entropy for the distribution will be high (i.e. a given outcome is, on average, very informative as it is difficult to predict).

Entropy scores were calculated using participants' full text responses.[5] First, an entropy score was calculated for each location in each condition (e.g. for train station in VISIBLE SPEAKER condition), from which we derived a mean entropy score per condition. By testing an equal number of participants (=100) per condition, we avoid a concern that the probability computations which contribute to the entropy scores would be distorted by unequal sample sizes. Such distortion could arise, for example, if one condition had fewer responses than another, such that the singleton responses (those produced only once) would be assigned higher probability in the small-sample-size condition than singleton responses in a condition with more responses.

KL divergence was also calculated using participants' full text responses. Relative entropy provides a measure of comparison between two probability distributions. Comparing the distribution of responses from each condition relative to the BARE condition can therefore be seen as a proxy for comparing comprehenders' estimates of what objects or events they expect a speaker to mention versus what objects or events are expected to be present in a location. Relative

[5]Where x is a particular response provided by one or more participants (e.g. *a train*), we computed entropy as follows for each location and each condition: Entropy = $\sum_x p(x) \log(p(x))$.

entropy complements the entropy measure described above, which only characterises a single distribution rather than drawing a comparison between the response distribution between conditions. Also, using the entropy measure alone may risk missing potential differences between distributions; two distributions could have similar entropy but very different properties (e.g. the BARE condition for the train station location might favour *train* with the same probability that another condition favours *delay*; in that case, if the remaining alternatives have similar probabilities, the two distributions would have similar entropy and any differences in distributions would remain undetected). Since all responses did not appear in all conditions, responses with a zero probability in BARE condition (i.e. responses that were not provided by any of the participants in this condition but did occur in one or more of the other conditions) were given a +1 smoothing. We opted to calculate relative entropy over the top 4 most frequent responses and bin the remaining responses in *Other*. This method avoids giving undue weight to any of the responses that were not observed in, for example, the BARE condition (i.e. responses that were smoothed), which may only have a very low probability of being observed in, for example, the VISIBLE SPEAKER condition. That is, the *Other* bin ensures that no non-occurring response would be given the same probability as a low-probability response.

A speaker's use of modification and negation in a continuation provides a way of adding more information. For the analyses of modification and negation, the presence of either element was manually coded as present (1) or not (0).

Typicality is about what is surprising and atypical for a given situation. For the typicality measure, the main noun in each response was checked against main nouns extracted in the pre-test data set and assigned a score based on the popularity rank of that response in the pre-test.

2.2 Results

As predicted, VISIBLE SPEAKER yielded the most informative completions on all five measures, as shown in Figures 2 to 5. For an illustration of the variation in responses across the 4 conditions, see OSF for figures displaying the distribution of responses in the train station location.[6]

The mean number of distinct responses per location in each of the 4 conditions were as follows: VISIBLE SPEAKER: 70.5, FIRST PERSON: 47.1, THIRD PERSON: 49.0, and BARE: 39.6. Entropy scores were analysed with paired Wilcoxon signed-rank tests across the items. Completions showed higher entropy for VISIBLE SPEAKER

[6] https://osf.io/7h5qs/

(mean: 4.00), FIRST PERSON (mean: 3.14) and THIRD PERSON (mean: 3.23) conditions compared to the baseline BARE (mean: 2.80; $p < 0.001$ for all three pairwise comparisons), and also the entropy of VISIBLE SPEAKER condition was higher than that of FIRST PERSON ($p < 0.001$), as seen in the left panel of Figure 2. Although there is a numeric difference between FIRST PERSON and THIRD PERSON conditions in the opposite direction to the prediction, this difference was not significant ($p = 0.17$). However, the difference between VISIBLE SPEAKER and the numerically adjacent THIRD PERSON is significant ($p < 0.001$). In order to test whether the difference in entropy was independent of use of modification or negation in participants' responses, we also conducted the analysis on the subset of responses not containing any modification and negation.[7] The above pattern is somewhat altered, with VISIBLE SPEAKER still showing the highest entropy (mean: 2.70) but the remaining conditions showing a different descending order across THIRD PERSON (mean: 2.60), followed by BARE (mean: 2.40), and then lowest entropy in FIRST PERSON condition (mean: 2.33). VISIBLE SPEAKER differs significantly from BARE and FIRST PERSON, but only numerically from THIRD PERSON. There is a significant difference between BARE and FIRST PERSON, but none between BARE and THIRD PERSON.

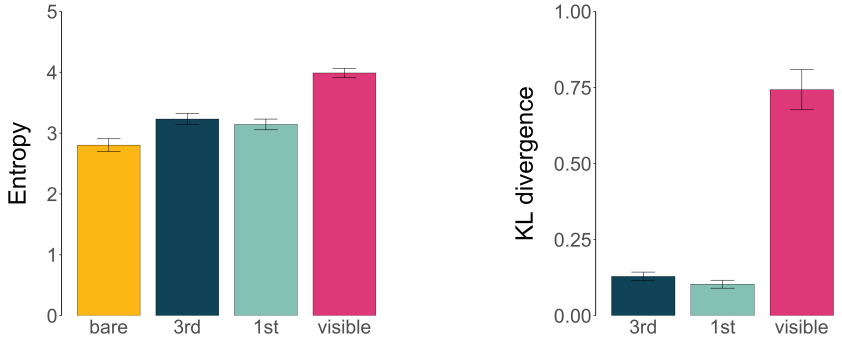


Figure 2: The left panel shows mean entropy for each condition; the right panel shows KL divergence as a comparison between the distribution in the listed condition and that of the baseline BARE condition.

[7]Thank you to an anonymous reviewer for suggesting this analysis. We note that while the subset analysis eliminates the matching number of responses across conditions, it eliminates more responses from the VISIBLE SPEAKER condition than any other, where the most use of modification and negation was registered. If higher entropy were to emerge simply due to the presence of more singleton unique responses, the subset analysis reduces the chance of seeing such an increase in VISIBLE SPEAKER. Hence, we do not believe this introduces a confound where analysing this subset would inadvertently yield higher entropy in the predicted VISIBLE SPEAKER condition.

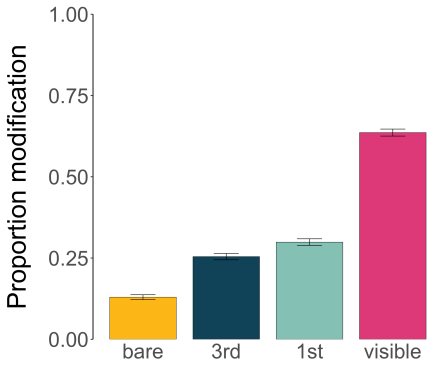


Figure 3: Proportion of use of modification across conditions

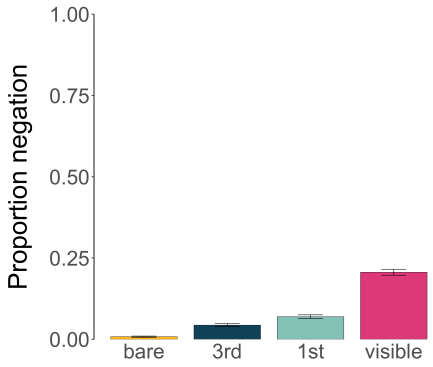


Figure 4: Proportion of use of negation across conditions

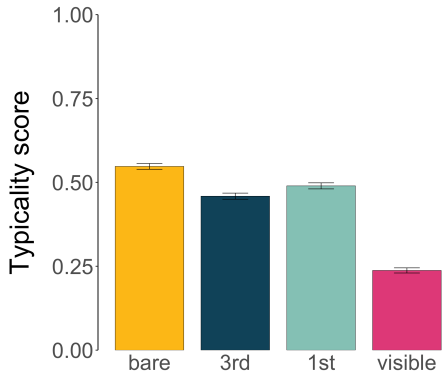


Figure 5: Mean typicality score of responses across conditions

The results for relative entropy confirm a role for speaker salience, as seen in the right panel of Figure 2[8]: Across items, the KL divergence between BARE and VISIBLE SPEAKER conditions (mean: 0.74) is greater than that between BARE and FIRST PERSON (0.10, paired t-test: $p > 0.001$) or between BARE and THIRD PERSON (0.13, $p > 0.001$). The comparisons between BARE and FIRST PERSON and between BARE and THIRD PERSON did not show a significant difference. As with entropy above, we also analysed the subset of responses that did not contain any modification or negation. In this case, the gradation of results stays the same: The KL divergence for BARE and VISIBLE SPEAKER (mean: 0.44) is higher than BARE and

[8]See OSF for examples of a list of the binned responses for each of the locations across conditions.

FIRST PERSON (0.08, $p > 0.001$), as well as than BARE and THIRD PERSON (0.12, $p > 0.001$). However, we also see a significant difference in the comparison between BARE and FIRST PERSON and between BARE and THIRD PERSON ($p > 0.05$). In sum, the distribution of participants' completions in VISIBLE SPEAKER condition is the one that is most clearly different from the distribution in BARE condition.

The binary outcomes of modification and negation were analysed with generalised mixed effects models[9] (GLMM: Jaeger 2008) using the lme4 package (Bates et al. 2015) in R (R Core Team 2022), treating condition as a fixed effect, with random slopes and intercepts of condition for participants and locations (Barr et al. 2013). In order to test whether we observe the predicted gradual increase in use of modification and negation across conditions, we forward coded the condition prior to analysis. This allows us to compare each condition to the adjacent one: BARE vs. THIRD PERSON, THIRD PERSON vs. FIRST PERSON, and FIRST PERSON vs. VISIBLE SPEAKER. To achieve convergence for the modification and negation models with maximal random effects structure, we used the optimiser `optimx` with the method `bobyqa` which led to convergence. For each fixed effect, we report a p-value as generated by the lmer logistic regression which is based on the Wald Z statistic (Agresti 2003). As seen in Figure 3, modification rates were highest in VISIBLE SPEAKER condition. VISIBLE SPEAKER was significantly higher than FIRST PERSON ($\beta = 2.81, \text{SE} = 0.38, z = 7.46, p < 0.001$), and THIRD PERSON was higher than BARE ($\beta = 2.46, \text{SE} = 0.47, z = 5.26, p < 0.001$). Although there is a difference between FIRST PERSON and THIRD PERSON in the predicted direction, this difference was not significant ($\beta = -0.13, \text{SE} = 0.41, z = -0.32, p = 0.75$). For negation (Figure 4), VISIBLE SPEAKER shows significantly higher proportions than FIRST PERSON ($\beta = 2.63, \text{SE} = 0.29, z = 9.01, p < 0.001$) and FIRST PERSON is higher than THIRD PERSON ($\beta = 3.78, \text{SE} = 0.85, z = 4.44, p < 0.001$); however, although numerically the predicted pattern is present for THIRD PERSON compared to BARE, this difference was not significant ($\beta = 11.18, \text{SE} = 7.60, z = 1.47, p = 0.142$).[10]

[9] modification~condition + (1 + condition | subjectID) + (1 + condition | location)
negation~condition + (1 + condition | subjectID) + (1 + condition | location)

[10] Using `bobyqa` produced a warning, but did not stop the models from converging. If we simplified the models by removing random slopes of condition for participant and for location, we achieve convergence without `bobyqa` and encountering no warnings. For modification the results stay the same (VISIBLE SPEAKER higher than FIRST PERSON ($\beta = 2.56, \text{SE} = 0.35, z = 7.29, p < 0.001$), no difference between FIRST PERSON and THIRD PERSON ($\beta = 0.19, \text{SE} = 0.36, z = 0.52, p < 0.60$), and THIRD PERSON higher than BARE ($\beta = 1.67, \text{SE} = 0.38, z = 4.44, p < 0.001$)), for negation we see significance in all three comparisons (VISIBLE SPEAKER higher than FIRST PERSON ($\beta = 2.08, \text{SE} = 0.30, z = 7.03, p < 0.001$), FIRST PERSON higher than THIRD PERSON ($\beta = 0.93, \text{SE} = 0.35, z = 2.66, p < 0.01$) and THIRD PERSON higher than BARE ($\beta = 2.54$, $\text{SE} = 0.72, z = 3.53, p < 0.001$)).

For the measure of typicality, we calculated a score for each main noun based on the number of pre-test participants who mentioned that object for the relevant location (meaning that a response given in Experiment 1 that was not provided in the pre-test receives a score of 0). Results are shown in Figure 5. To analyse these typicality scores we again forward coded the condition (comparing each condition to the adjacent one; BARE VS. THIRD PERSON, THIRD PERSON VS. FIRST PERSON, FIRST PERSON VS. VISIBLE SPEAKER)[11] and used a linear mixed effects model[12] with condition as a fixed effect and random slopes and intercepts of condition for participants and locations (Barr et al. 2013), as above. The significance of the fixed effect of condition was determined via a likelihood ratio test comparing the fit of the model to one with the same random effects structure but no fixed effect. Again, to achieve convergence with the models with maximal random effects structure, we used the optimiser `optimx` with the method `bobyqa`. We see a main effect of condition in the model comparison ($p < 0.001$), showing that the model with condition as a fixed effect significantly improved the model fit. Typicality was significantly lower in VISIBLE SPEAKER ($\beta = -0.25, \text{SE} = 0.03, t = -7.93$) compared to FIRST PERSON, and THIRD PERSON was lower than BARE ($\beta = -0.09, \text{SE} = 0.03, t = -2.86$). The difference between FIRST PERSON and THIRD PERSON was not significant ($\beta = 0.03, \text{SE} = 0.03, t = 1.20$).

2.3 Discussion

The results of Experiment 1 show that participants' expectations about informativity, as elicited by an utterance completion task, vary according to how and whether the speaker is made salient. In the conditions which were intended to make the speaker more salient, participants' completions score higher in informativity as shown by the five measures: Higher entropy, greater KL divergence for VISIBLE SPEAKER condition compared to the BARE condition, more use of modification and of negation, and fewer typical entities provided in the VISIBLE SPEAKER condition compared to the other conditions. We take this increase in informativity to reflect an increased awareness of the speaker, inducing participants to consider the speaker's production preferences when estimating possible utterance completions.

We find only partial support for the hypothesis that continuations gradually increase in informativity across the four conditions: The use of modification and

[11] Since the observed means do not pattern with the predicted order, we conducted an additional analysis setting the forward contrast coding to order the conditions to match the observed order; BARE VS. FIRST PERSON, FIRST PERSON VS. THIRD PERSON, THIRD PERSON VS. VISIBLE SPEAKER (as seen in Figure 5). The result remains the same.

[12] typicality~condition + (1 + condition | subjectID) + (1 + condition | location)

negation follows the expected pattern numerically with a gradual increase from the least to most speaker-salient conditions, but there is no significant difference between THIRD PERSON and FIRST PERSON conditions. For entropy, KL divergence and typicality, the pattern is somewhat less clear: The results pattern as predicted except for in THIRD PERSON and FIRST PERSON conditions, which are mirrored numerically compared to the predicted direction. However, these differences are also not significant. Overall, we interpret this to show that the manipulation generally works in the intended way; it affects a participant's probability of using the second estimate discussed above, i.e. one that takes into account a speaker's production preference to convey informative content. The gradual pattern observed in the data might thus reflect individuals' varying sensitivity to the manipulation: The VISIBLE SPEAKER condition most strongly cues participants to consider the speaker's production process, but evidently some participants seem to already be engaging their awareness of the speaker in the less speaker-salient conditions. The manipulation seems to affect the measures differently, suggesting that the mechanism for comprehenders to change perspectives in this way may be more complex than our experimental design can account for. Nevertheless, it seems that whatever the process may be that generates candidate expressions when anticipating content in a sentence completion task, this process is substantially modulated by comprehenders' awareness of the speaker.

3 Experiment 2: Awareness of speaker style in comprehension

Having established in Experiment 1 that comprehenders appear to use their awareness of speakers' production biases in generating expectations about what a speaker will say next, Experiment 2 tests whether such expectations also reflect properties of the speaker.[13] We ask whether comprehenders have fixed expectations for speakers to convey newsworthy content, or whether such expectations are malleable. This study is different from the studies on referential expectations discussed in the introduction, where the communicative goal is identifying the intended referent. Experiment 2 instead aims to test speaker-specific expectations for informativity regarding upcoming message content where the communicative goal is unspecified and likely to be communicative interest. For example, comprehenders may estimate that different speakers have different thresholds for what counts as an informative contribution. Comprehenders may in turn allow for variation in the way a speaker delivers newsworthy content depending on what they know about the speaker.

[13]This experiment was preregistered: https://osf.io/r2h7d

Studies on speaker adaptation in a range of linguistic domains show that comprehenders use cues from a speaker's previous discourse when anticipating upcoming content and when interpreting it. For example, a speaker's accent or dialect can lead comprehenders to shift their perception of phonemes (Hay et al. 2006, Hay & Drager 2010); when encountering temporarily ambiguous words comprehenders make anticipatory looks to a more frequently mentioned referent based on the speaker's previous behaviour (Creel et al. 2008); in the early stages of processing, comprehenders are influenced by a speaker's either literal or non-literal style when disambiguating metaphorical polysemous words (Davies et al. 2022); and comprehenders adapt their interpretations of scalar quantifiers such as *some* and *many* following exposure to different speakers (Yildirim et al. 2016). There are also formalised accounts of, e.g. phonetic adaptation (Kleinschmidt & Jaeger 2015) and semantic/pragmatic adaptation (Schuster & Degen 2020) which capture comprehenders' ability to track subtle characteristics of different speakers and adjust their interpretations accordingly.

Using the VISIBLE SPEAKER condition from Experiment 1, our second experiment tests whether speaker-specific effects are observable in comprehenders' expectations for the informativity of upcoming content. Two different speakers with different communication styles were introduced in an exposure phase: One who produces a mix of informative and uninformative utterances and one who produces only informative utterances. Following the exposure phase, participants completed sentences purportedly uttered by these two speakers. If comprehension reflects speaker-specific expectations for informativity, comprehenders are predicted to provide more informative completions for the HIGH-INFORMATIVITY speaker compared to the LOW-INFORMATIVITY speaker, showing that they can modify their expectation about what someone is going to say next after learning about the specific production preferences of the individual speakers. Such a finding would indicate that in anticipating upcoming content, we are able to use awareness of production dynamically as we get to know the communication style of particular speakers.

3.1 Methods

3.1.1 Participants

200 participants were recruited on Prolific with the same criteria as in Experiment 1: English as first language, no known language impairments, location in the US or the UK, and minimum 18 years of age (M = 36.07, SD = 14.20, range = 18–76). Additionally, people who had taken part in Experiment 1 were blocked

from participating in Experiment 2. As in Experiment 1, participants who failed two or more attention checks or reported another language than English as their first language were excluded and replaced by other participants to reach the desired number of 100 participants per condition. Participants were paid £7.60/hour on average.[14]

Figure 6: Example of a filler item for the LOW-INFORMATIVITY speaker in exposure phase. All items in exposure consisted of two parts, shown here in two boxes. Image from Freepik (https://freepik.com).

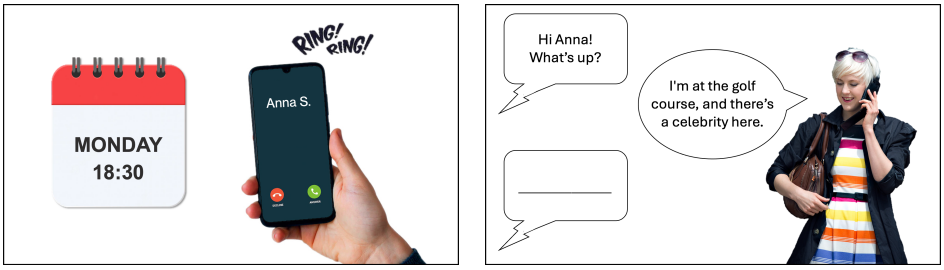


Figure 7: Example of a critical item for the HIGH-INFORMATIVITY speaker in exposure phase. All items in exposure consisted of two parts, shown here in two boxes. Image from Pixabay (https://pixabay.com).

3.1.2 Design

Experiment 2 tests whether comprehenders adapt their expectations for informativity based on the speaking style of two different speakers: One who produces a mix of informative and uninformative utterances and one who produces informative utterances only. The experiment has two conditions (HIGH-INFORMATIVITY vs. LOW-INFORMATIVITY speaker style), manipulated within participants. The task

[14]The details of the cost were established by Prolific based on the median time taken to complete the experiment per submission for each group.

consisted of two phases; an exposure phase and a test phase. In the exposure phase, participants were familiarised with two different speakers, Suzy and Anna, and their differing communication styles (see an example for each of the speakers in Figures 6 and 7). The test phase is a near replication of the VISIBLE SPEAKER condition in Experiment 1 in which participants write completions for the two different speakers (examples in Figures 8 and 9). We designed this paradigm to encourage participants to feel like they were engaged in a series of phone calls. In the exposure phase, the participant's task is to type in utterances that contribute their own turns in the dialogue. In the test phase, on the other hand, they are instructed to fill in parts of the caller's utterances. The exposure phase thus ensures that they witness the kind of utterances produced by the two callers, whereas the test phase asks them to indicate what they think each of the callers would be likely to say.

Figure 8: Example of a critical item for the LOW-INFORMATIVITY speaker in test phase

Completions for the HIGH-INFORMATIVITY speaker are expected to be more informative than those for the LOW-INFORMATIVITY speaker. Such an observed difference in the measures between LOW-INFORMATIVITY speaker condition and HIGH-INFORMATIVITY speaker condition would reflect a change in participants' evaluation of what counts as informative for each of the speakers, in line with our thesis that comprehenders are able to dynamically use their knowledge of speakers' communicative styles in their estimates of likely upcoming content.

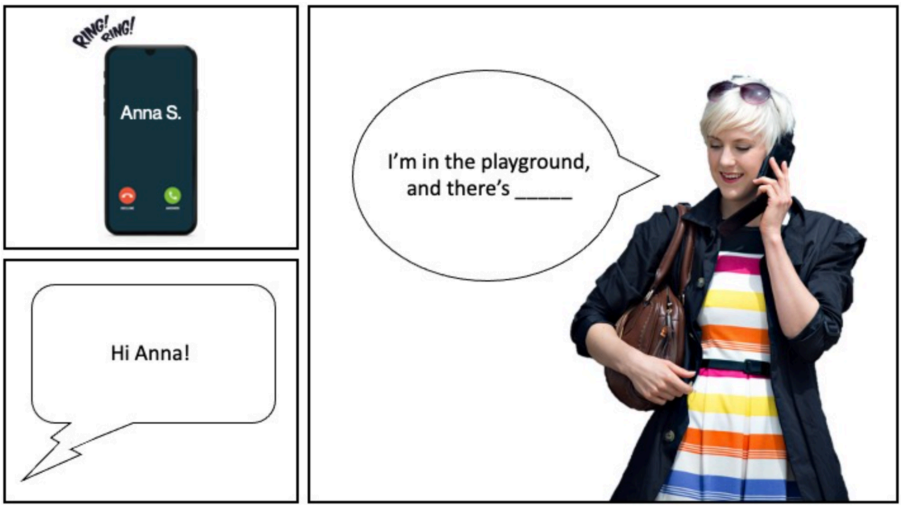


Figure 9: Example of a critical item for the HIGH-INFORMATIVITY speaker in test phase

3.1.3 Materials

Materials were constructed to depict a series of beginnings of phone calls, with a picture of the speaker (the person calling), and speech bubbles showing different turns of the conversation. The critical items (training items in the exposure phase and test items in the test phase) are in the form of *I'm at the [location], and there's* _____, as in Experiment 1, using the same 20 locations split across the two phases.

In the exposure phase, participants see the start of a phone call and a contribution from either the HIGH-INFORMATIVITY speaker or the LOW-INFORMATIVITY speaker. For the critical items in the exposure phase, the HIGH-INFORMATIVITY speaker only utters training sentences about non-typical situations (e.g. *I'm at the golf course, and there's a celebrity here.*), whereas the LOW-INFORMATIVITY speaker utters some training sentences about non-typical situations as well as some about typical situations (e.g. *I'm at the cinema, and there's popcorn.*). For the critical training items in the exposure phase, the speakers' utterances are taken from responses in Experiment 1: The HIGH-INFORMATIVITY speaker has 2 critical training items, both using completions from the VISIBLE SPEAKER condition, whereas the LOW-INFORMATIVITY speaker has 8 critical training items, 2 with completions from the VISIBLE SPEAKER condition and 6 from the BARE condition. Note that this setup ensures that both speakers provide the same number of interesting contributions overall, to avoid participants inferring that one speaker encounters more non-typical situations than the other. The different communication styles of the

speakers is then emphasised by the fact that the LOW-INFORMATIVITY speaker additionally reports on 6 typical situations. In addition to the critical training items, the exposure phase includes 18 filler items; 6 for the HIGH-INFORMATIVITY speaker and 12 for the LOW-INFORMATIVITY speaker. This setup means that overall the participant answers calls from the LOW-INFORMATIVITY speaker noticeably more often than the HIGH-INFORMATIVITY speaker (20 vs. 8 items). Filler items were either utterances requesting information or conveying information. These were constructed to match the communication style of the speakers, such that the fillers for the HIGH-INFORMATIVITY speaker targeted newsworthy information (e.g. *Do you know when Lisa is arriving next week?*), whereas the LOW-INFORMATIVITY speaker called about mundane events (e.g. *What did you have for dinner?*). Across the critical and filler items in the exposure phase, participants were encouraged to engage in the task and to see the setup as a communicative context, by requiring them to provide either a starting utterance for the conversations or a response to the speaker's utterance.

All the items in the exposure phase consisted of two images (see Figures 6 and 7). The first shows a calendar with a day and a time next to a hand holding a ringing phone with the caller's name clearly visible. The day and time was included to reinforce the manipulation that the LOW-INFORMATIVITY speaker calls more often; they call every day and occasionally several times a day, whereas the HIGH-INFORMATIVITY speaker calls a maximum of once a day and not every day. The second image shows the speaker and a series of speech bubbles indicating the first, second and third turn of the dialogue. The speaker's speech bubble, which is always the second turn, is always fully visible in order to expose the participant to each speaker's communicative style, whereas the speech bubbles for the callee (the participant) contain fill-in-the-blank spaces in either the first or the third turn.

For the test phase, the speakers' utterances include a fill-in-the-blank space to elicit participants' guesses about what each of the speakers is likely to say. All items in the test phase were similar to those in the VISIBLE SPEAKER condition of Experiment 1, showing the speaker and a speech bubble with an utterance that includes a blank; here in Experiment 2, the test items additionally depict a ringing phone and the callee's greeting in a speech bubble appearing to one side (see examples Figures 8 and 9). A participant sees an equal number of test items from each speaker (5 from the LOW-INFORMATIVITY speaker and 5 from the HIGH-INFORMATIVITY speaker), plus 10 filler items from each speaker as well. Filler items were the same as in Experiment 1, including the two that served as attention checks.

Critical training and test items (i.e. the 20 locations from Experiment 1) were counterbalanced across lists such that a location that appeared in the exposure phase in one list appeared as a test item in another list. Similarly, participants saw the HIGH-INFORMATIVITY (HI) and LOW-INFORMATIVITY (LI) speaker depicted as one of the two speaker images, counterbalanced across participants, to avoid any potential bias associated with the appearance of the speakers. Counterbalancing items and pictures of speakers resulted in 4 lists:

- LI speaker: dark-haired, HI speaker: blonde + exposure items: target set A
- LI speaker: blonde, HI speaker: dark-haired + exposure items: target set A
- LI speaker: dark-haired, HI speaker: blonde + exposure items: target set B
- LI speaker: blonde, HI speaker: dark-haired + exposure items: target set B

Under this counterbalancing, a given item in the exposure phase was always associated with the same speaker style (e.g. *train station* was always uttered by the LOW-INFORMATIVITY speaker in exposure), and likewise a given item in the test phase was always associated with the same speaker style (e.g. *restaurant* was always uttered by the HIGH-INFORMATIVITY speaker in test).

3.1.4 Procedure

The experiment was presented using Qualtrics. In the exposure phase, participants were instructed to imagine answering calls from each of two speakers, filling in an empty speech bubble from the participant's side of the conversation in each call. Each trial consisted of two parts (example trials in Figures 6 and 7). First, participants see a day and time and a phone showing the caller's name; second, they see the caller talking into a phone and the dialogue in speech bubbles, and a text box for filling in the missing content. The items varied between the participant having to complete the very first utterance (the greeting) or the response to the caller's first utterance. This variation was intended to ensure participants stayed engaged and read the utterances fully, as well as to distract them from the potentially unnatural sounding utterances from the caller by allowing them to influence the dialogue by contributing their own turns. Items in the exposure phase were presented in a fixed order, so that every participant within a list saw the same item order. As in Experiment 1, participants were presented with one item at a time with a text field below the item where they typed their completions, and they proceeded to the next trial by clicking the arrow on screen.

In the test phase, participants were again instructed that they would be seeing the beginnings of a series of phone calls from the same two speakers, but this time they would be completing the callers' utterances. The test phase was essentially a replication of the VISIBLE SPEAKER condition in Experiment 1, with the only difference being the edits explained above. The test phase used simpler visualisations of the phone calls compared to the exposure phase, as shown in Figures 6 and 7 versus 8 and 9; this simplification was intended to make the second phase of the experiment less tiresome for the participant by reducing the number of images they had to click through. Presentation of items in the exposure phase was fully randomised.

After the test phase, there were three questions intended to assess whether participants had paid attention and were sensitive to the different speaker styles. The first of these was a speaker line-up where participants had to choose which of five speakers they had talked to. The speaker line-up consisted of the images of the two speakers flipped horizontally, and three images of other speakers taken from Experiment 1. The second question asked if the participant noticed a difference between the two speakers, and if so, what that difference was. The third asked them to rate how interesting they found each of the two speakers on a scale from 1 (not at all interesting) to 7 (very interesting). The median completion times in each list were between 17 and 19 minutes.

3.1.5 Measures and data analysis

For Experiment 2, we used the same coding procedure and the same measures as in Experiment 1: Entropy as a measure of variability in participant responses, KL divergence to measure how the distribution of responses in the two conditions here compares to that of a baseline condition (the BARE condition from Experiment 1), modification and negation as measures of participants' enhancement of their responses, and typicality of the responses' main noun as compared to the nouns elicited in the Experiment 1 pre-test. A total of 2 continuations were non-meaningful and therefore excluded from analysis. The included continuations ranged from 1 word in length to 6 words, with a mean of 1.55.

3.2 Results

The mean number of distinct responses in the HIGH-INFORMATIVITY condition was 58.0, compared to 44.7 in the LOW-INFORMATIVITY condition. As predicted for the entropy measure, the HIGH-INFORMATIVITY speaker yielded a distribution of responses with higher entropy (mean: 3.60) compared to the LOW-INFORMATIVITY

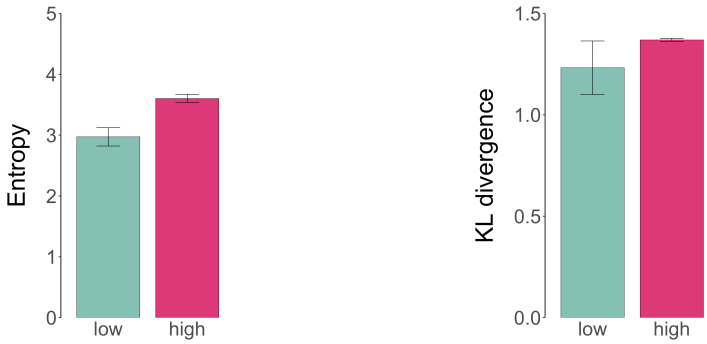


Figure 10: The left panel shows mean entropy for each speaker informativity condition; the right panel shows KL divergence as a comparison between the distribution in the listed condition and that of the baseline BARE condition.

speaker (mean: 3.00; Wilcoxon signed-rank test: $p < 0.005$), as seen in the left panel of Figure 10. For KL divergence, there was a numeric difference between HIGH-INFORMATIVITY vs. BARE (1.37) and LOW-INFORMATIVITY vs. BARE (1.23), but the difference was not significant (right panel, Figure 10). As in Experiment 1, we also calculated the entropy of the subset of responses not containing any modification or negation. For entropy, the pattern still holds, with HIGH-INFORMATIVITY (mean: 2.80) being significantly higher than LOW-INFORMATIVITY (mean: 2.20, $p < 0.05$). For KL divergence, the pattern above remained numeric only (HIGH-INFORMATIVITY vs. BARE (1.14) and LOW-INFORMATIVITY vs. BARE (1.10)).

As in Experiment 1, the binary variables of modification and negation were analysed with generalised mixed effects models treating condition (here, HIGH-INFORMATIVITY vs. LOW-INFORMATIVITY) as a fixed effect and with random slopes and intercepts of condition for participants and items. For the model for modification, we again used the optimiser `optimx` with the method `bobyqa` to achieve convergence with the full random effects structure. As predicted and as can be seen in Figure 11, modification was used more frequently in completions of utterances from the HIGH-INFORMATIVITY speaker compared to the LOW-INFORMATIVITY speaker ($\beta = -0.88$, SE = 0.19, $z = -4.57$, $p < 0.001$). The model for negation converged with full random effects structure and the default optimiser; however, although numerically the predicted pattern is observable (see Figure 12), the difference between HIGH-INFORMATIVITY/LOW-INFORMATIVITY speakers was not significant ($\beta = -0.58$, SE = 0.57, $z = -1.02$, $p = 0.306$).

Typicality was analysed with the same approach as in Experiment 1, and again the predicted pattern was confirmed (Figure 13). We constructed a linear mixed

effects model with condition (HIGH-INFORMATIVITY vs. LOW-INFORMATIVITY) as a fixed effect and random slopes and intercepts of condition for participants and items, with the optimiser `optimx` combined with the method `bobyqa` to achieve convergence. The model confirms that typical main nouns were used less often for the HIGH-INFORMATIVITY speaker compared to LOW-INFORMATIVITY speaker ($\beta = -0.13$, SE = 0.06, $t = -2.32$). To derive a p-value, we used model comparison via a likelihood ratio test and found a main effect ($p < 0.05$), showing that the model with condition as a fixed effect provides a significantly better fit than the model with same random effects structure but no fixed effect.

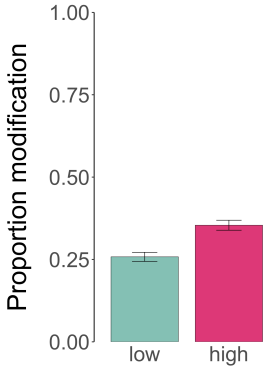


Figure 11: Proportion of use of modification across conditions

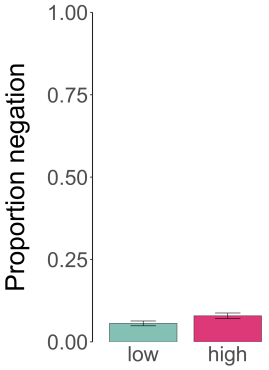


Figure 12: Proportion of use of negation across conditions

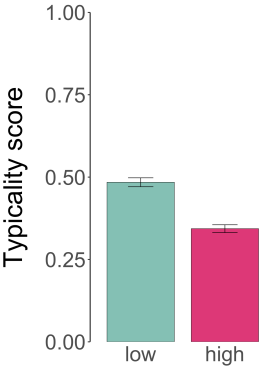


Figure 13: Mean typicality score of responses across conditions

3.3 Discussion

In line with the claim that comprehension reflects speaker-specific expectations for informativity, the results of Experiment 2 show that participants' completions were higher in informativity for the HIGH-INFORMATIVITY speaker compared to the LOW-INFORMATIVITY speaker. The HIGH-INFORMATIVITY speaker completions were more variable as shown by the higher entropy, and elicited more modification and lower typicality, showing that participants adapt their expectations to the individual speakers' communication styles as predicted. For KL divergence, there was a numeric difference in the predicted direction, but this was not significant. The results for the rate of negation were numerically in line with the prediction that the HIGH-INFORMATIVITY speaker is more likely to comment on the *lack* of something typical compared to the LOW-INFORMATIVITY

speaker, but this effect was also not significant. This may reflect the fact that none of the critical items in the exposure phase explicitly demonstrate how the two speakers use negation, whereas some of both filler and critical items in the exposure phase did use modification.

4 General discussion

The goal of this study was to investigate the effect of comprehenders' awareness of a speaker's production preferences and communication style on their anticipation of upcoming content. Previous work on comprehension and production tends to show contradicting preferences for comprehenders and speakers, such that typicality is favoured in comprehension (e.g. being associated with processing ease), whereas speakers tend to make production choices that favour the inclusion of surprising or otherwise informative content (meaning easily inferable or typical content is disfavoured). However, we posited that comprehenders are aware of and make use of speakers' production preferences in their expectations for upcoming content, such that they have a bias towards filtered language use rather than a simple expectation of transparent language use. The two Cloze task studies presented here tested this by tapping into the comprehender's expectations, when the comprehender does not have access to the scene being described.

Experiment 1 addressed the question of whether increasing the emphasis on the speaker would lead to an increased expectation for informative content. The higher entropy of completions in the VISIBLE SPEAKER condition and also FIRST PERSON and THIRD PERSON compared to BARE, show that participants' completions were more variable and thus less predictable in the conditions where the speaker was more salient. Furthermore, KL divergence showed that the distribution of participants' completions in VISIBLE SPEAKER condition was the most clearly different from the distribution in BARE condition. We take this as evidence that participants are tuning in to speakers' production preferences and expecting speakers to tell them interesting things by providing more unique completions in these conditions compared to the BARE condition. For example, while *train* was a common completion for the train station location in the BARE condition, it was much less frequent in the VISIBLE SPEAKER condition, suggesting that participants are taking into account what a speaker might consider worth talking about when they provide their completions.

Our predictions were upheld for the other three measures. Rather than looking at the overall variability in participants' completions, these measures aim to capture properties of the content of individual completions. Modification and

negation are taken to be strategies that participants can employ to make something that might otherwise be considered typical for a location more non-typical and therefore worthy of reporting. Participants made use of modification such as descriptive adjectives (e.g. *a cute dog*), quantifying expressions (e.g. to communicate unusual amounts of an entity: *so much choice*) or modification to specify a less typical subtype of an entity (e.g. *steam train*). Similarly, negation can be a way to mark the absence of something that one would typically expect to find at a location, such as a train at the train station, and reporting this absence (e.g. *no trains*) would therefore likely be considered informative. As for typicality, the results showed that participants produced completions mentioning typical entities most often in the BARE condition, and significantly less often in the VISIBLE SPEAKER condition. These measures and the progressive differences across conditions are taken to reflect comprehenders' sensitivity to the experimental manipulation of speaker salience and, more broadly, to reveal their expectation for speakers to convey non-typical and informative content.

These findings have potential repercussions for the use of Cloze tasks in other psycholinguistic research since the kinds of completions that participants provide are evidently malleable. Given the differences in response distributions that we observed across conditions, researchers using Cloze tasks may need to tune their experimental tasks to their precise research goals. This might require using standard Cloze task phrasing and procedures to elicit completions that reflect participants' estimates of what is typical in the world or using more situated tasks that depict a communicative context in order to elicit completions that reflect participants' estimates of what a speaker is likely to talk about.

Experiment 2 addressed the question of whether this expectation is further malleable depending on properties of the speaker, specifically whether comprehenders are sensitive to different communication styles. Using completions provided in Experiment 1, participants were exposed to two speakers whose utterances were either high or low in informativity. Similar to the speaker salience manipulation in Experiment 1, the manipulation of speaker properties in Experiment 2 allows us to test whether comprehenders take into account properties of the speaker in guessing upcoming content (transparent language use vs. filtered language use). Participants showed a bias to expect the HIGH-INFORMATIVITY speaker to produce utterances conveying more informative content than the LOW-INFORMATIVITY speaker. Entropy was higher for the HIGH-INFORMATIVITY speaker, showing more variability in completions. Although the KL divergence measure did not show a significant difference between the two conditions (when the distribution of responses in each speaker condition is compared to that of the BARE condition in Experiment 1), there was a numeric difference in the predicted

direction. Modification was higher for the HIGH-INFORMATIVITY speaker, and although not significant, we also saw an increase in use of negation with the HIGH-INFORMATIVITY speaker. Lastly, mention of typical entities was more frequent for the LOW-INFORMATIVITY speaker. We take this to show that comprehenders are able to estimate that different speakers have different thresholds for what counts as an informative contribution and adjust their expectations accordingly.

A challenge when discussing how we anticipate upcoming content is to clearly distinguish the relevant concepts involved. Typicality, plausibility and (im)possibility all seem interwoven, and although they all likely play a role in anticipating what someone is going to say next, we have focussed on typicality in this study. We have defined typicality as the frequency of an event or situation; the frequency with which something occurs is conceptually and empirically relatively straightforward to measure. As we observed in this study, participants expect utterances to convey less typical content when the speaker is emphasised. Importantly, however, no participant seemingly contributed responses that were so extremely non-typical that they crossed the threshold into being impossible (e.g. a steam train in the kitchen). One could imagine scenarios where utterances about highly non-typical situations would be felicitous (e.g. in descriptions of fictional or dream worlds; Foy & Gerrig 2014, Troyer & Kutas 2020), and perhaps even the most expected in some contexts, but there seems to be an intuition that utterances should stay within the realm of what is plausible.

One way of conceiving of the interplay between typicality and plausibility could be that the latter provides a range within which situations can happen, whereas whether or not those situations are typical will depend on their frequency. For instance, taking a sip from a transmitter is implausible; drinking from a waterfall, however, certainly is plausible, but significantly less typical than drinking from a tap. Since continuations in our current study all seemed to fall within the range of plausible utterances, our data cannot speak to the role of plausibility in processing and anticipating upcoming content. There is work attempting to tease apart the effects of impossibility and implausibility on processing (see e.g. Warren & McConnell 2007) and other work that makes a distinction between real-world plausibility and word-predictability (e.g. Albu et al. 2023). As future work looks at more fine-grained processing in relation to anticipation of upcoming content and speaker salience, it could become important to make clear distinctions between all these concepts; typicality, plausibility and possibility.

As discussed above, previous work has emphasised the importance of typicality in processing (e.g. Marks & Miller 1964, Kutas & Hillyard 1980, Morris 1994, Kamide et al. 2003). When looking at the psycholinguistic literature, one might

think that the default approach is to see sentences describing implausible or non-typical situations as anomalies that need to be reparsed (see e.g. Cai et al. 2022). However, such work sidesteps the role of pragmatic reasoning, failing to incorporate a comprehender's awareness of the possibility that such an utterance has been produced precisely because its content is surprising, in particular in contexts where the communicative goal of the interaction is unspecified. Of course, it is reasonable to assume that most language users will expect a contribution to stay within a certain range of plausibility – if a contribution is too implausible, one would expect processing to suffer because it may be hard for comprehenders to reconcile what the speaker has said with what they know about the world. However, the results presented here suggest that comprehenders do have expectations that speakers will talk about interesting and non-typical things, highlighting a role for informativity-driven reasoning about the speaker in models of language processing.

Acknowledgements

We thank editor Robin Lemke and two anonymous reviewers for very helpful comments.

References

Agresti, Alan. 2003. *Categorical data analysis*. 2nd edn., vol. 482 (Wiley Series in Probability and Statistics). Somerset: Wiley.

Albu, Elena, Carolin Dudschig, Tessa Warren & Barbara Kaup. 2023. Does negation influence the choice of sentence continuations? Evidence from a four-choice cloze task. *Quarterly Journal of Experimental Psychology*. 17470218231158109. DOI: 10.1177/17470218231158109.

Altarriba, Jeanette, Judith F. Kroll, Alexandra Sholl & Keith Rayner. 1996. The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition* 24(4). 477–492. DOI: 10.3758/BF03200936.

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. DOI: 10.1177/00238309040470010201.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language* 68(3). 255–278. DOI: 10.1016/j.jml.2012.11.001.

Bates, Douglas, Martin Mächler, Ben M. Bolker & Steve C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of statistical software* 67(1). 1–48. DOI: 10.18637/jss.v067.i01.

Borghi, Anna M. & Felice Cimatti. 2010. Embodied cognition and beyond: Acting and sensing the body. *Neuropsychologia* 48(3). 763–773. DOI: 10.1016/j.neuropsychologia.2009.10.029.

Brown, Paula M. & Gary S. Dell. 1987. Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology* 19. 441–472. DOI: 10.1016/0010-0285(87)90015-6.

Cai, Zhenguang G., Nan Zhao & Martin J. Pickering. 2022. How do people interpret implausible sentences? *Cognition* 225. 105101. DOI: 10.1016/j.cognition.2022.105101.

Creel, Sarah C., Richard N. Aslin & Michael K. Tanenhaus. 2008. Heeding the voice of experience: The role of talker variation in lexical access. *Cognition* 106. 633–664. DOI: 10.1016/j.cognition.2007.03.013.

Davies, Catherine & Jennifer E. Arnold. 2019. Reference and informativeness: How context shapes referential choice. In Chris Cummins & Napoleon Katsos (eds.), *The Oxford handbook of experimental semantics and pragmatics*, 474–493. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780198791768.013.6.

Davies, Catherine, Vincent Porretta, Kremena Koleva & Ekaterini Klepousniotou. 2022. Speaker-specific cues influence semantic disambiguation. *Journal of Psycholinguistic Research* 51(5). 933–955. DOI: 10.1007/s10936-022-09852-0.

Degen, Judith, Robert D. Hawkins, Caroline Graf, Elisa Kreiss & Noah D. Goodman. 2020. When redundancy is useful: A Bayesian approach to "overinformative" referring expressions. *Psychological Review* 127(4). 591–621. DOI: 10.1037/rev0000186.

Foy, Jeffrey E. & Richard J. Gerrig. 2014. Flying to Neverland: How readers tacitly judge norms during comprehension. *Memory & Cognition* 42(8). 1250–1259. DOI: 10.3758/s13421-014-0436-8.

Goodman, Noah D. & Michael C. Frank. 2016. Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences* 20. 818–829. DOI: 10.1016/j.tics.2016.08.005.

Grice, Herbert Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics: Speech acts*, 41–58. New York: Academic Press.

Grodner, Daniel & Julie C. Sedivy. 2011. The effect of speaker-specific information on pragmatic inferences. In Edward A. Gibson & Neal J. Pearlmutter (eds.), *The processing and acquisition of reference*, 239–272. Cambridge: The MIT Press. DOI: 10.7551/mitpress/9780262015127.003.0010.

Hay, Jennifer & Katie Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48(4). 865–892. DOI: 10.1515/ling.2010.027.

Hay, Jennifer, Aaron Nolan & Katie Drager. 2006. From fush to feesh: Exemplar priming in speech perception. *Linguistic review* 23(3). 351–379. DOI: 10.1515/TLR.2006.014.

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446. DOI: 10.1016/j.jml.2007.11.007.

Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61(1). 23–62. DOI: 10.1016/j.cogpsych.2010.02.002.

Kamide, Yuki, Gerry T. M. Altmann & Sarah L. Haywood. 2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language* 49(1). 133–156. DOI: 10.1016/S0749-596X(03)00023-8.

Kleinschmidt, Dave F. & T. Florian Jaeger. 2015. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review* 122(2). 148–203. DOI: 10.1037/a0038695.

Kravtchenko, Ekaterina. 2014. Predictability and syntactic production: Evidence from subject omission in Russian. *Proceedings of the annual meeting of the Cognitive Science Society* 36. 785–790.

Kravtchenko, Ekaterina & Vera Demberg. 2022. Informationally redundant utterances elicit pragmatic inferences. *Cognition* 225. 105159. DOI: 10.1016/j.cognition.2022.105159.

Kreiss, Elisa & Judith Degen. 2020. Production expectations modulate contrastive inference. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*. 259–265.

Kuperberg, Gina R. & T. Florian Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* 31. 32–59. DOI: 10.1080/23273798.2015.1102299.

Kutas, Marta & Steven A. Hillyard. 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207. 203–205. DOI: 10.1126/science.7350657.

Lemke, Robin. 2021. *Experimental investigations on the syntax and usage of fragments* (Open Germanic Linguistics 1). Language Science Press. DOI: 10.5281/zenodo.5596236.

Lemke, Robin, Ingo Reich, Lisa Schäfer & Heiner Drenhaus. 2021a. Predictable words are more likely to be omitted in fragments: Evidence from production data. *Frontiers in Psychology* 12. 662125–662125. DOI: 10.3389/fpsyg.2021.662125.

Lemke, Robin, Lisa Schäfer & Ingo Reich. 2021b. Modeling the predictive potential of extralinguistic context with script knowledge: The case of fragments. *PLoS ONE* 16. DOI: 10.1371/journal.pone.0246255.

Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In Bernhard Schölkopf, John Platt & Thomas Hofmann (eds.), *Advances in neural information processing systems*, 849–856. Cambridge: The MIT Press. DOI: 10.7551/mitpress/7503.003.0111.

Lockridge, Calion B. & Susan E. Brennan. 2002. Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review* 9. 550–557. DOI: 10.3758/BF03196312.

Marks, Lawrence E. & George A. Miller. 1964. The role of semantic and syntactic constraints in the memorization of English sentences. *Journal of Verbal Learning and Verbal Behavior* 3. 1–5. DOI: 10.1016/S0022-5371(64)80052-9.

Mitchell, Margaret, Ehud Reiter & Kees Van Deemter. 2013. Typicality and object reference. *Proceedings of the Annual Meeting of the Cognitive Science Society* 35. 3062–3067.

Morris, Robin K. 1994. Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory & Cognition* 20. 92–103. DOI: 10.1037/0278-7393.20.1.92.

Nieuwland, Mante S. & Jos J. A. Van Berkum. 2006. When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience* 18(7). 1098–1111. DOI: 10.1162/jocn.2006.18.7.1098.

Pogue, Amanda, Chigusa Kurumada & Michael K. Tanenhaus. 2016. Talker-specific generalization of pragmatic inferences based on under- and over-informative prenominal adjective use. *Frontiers in Psychology* 6. 2035–2035. DOI: 10.3389/fpsyg.2015.02035.

R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna. https://www.R-project.org/.

Rohde, Hannah, Richard Futrell & Christopher G. Lucas. 2021. What's new? A comprehension bias in favor of informativity. *Cognition* 209. 104491. DOI: 10.1016/j.cognition.2020.104491.

Rohde, Hannah, Jet Hoek, Maayan Keshev & Michael Franke. 2022. This better be interesting: A speaker's decision to speak cues listeners to expect informative content. *Open Mind* 6. 118–131. DOI: 10.1162/opmi_a_00058.

Rubio-Fernández, Paula. 2016. How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology* 7. DOI: 10.3389/fpsyg.2016.00153.

Ryskin, Rachel, Chigusa Kurumada & Sarah Brown-Schmidt. 2019. Information integration in modulation of pragmatic inferences during online language comprehension. *Cognitive Science* 43(8). e12769. DOI: 10.1111/cogs.12769.

Schäfer, Lisa, Robin Lemke, Heiner Drenhaus & Ingo Reich. 2021. The role of UID for the usage of verb phrase ellipsis: Psycholinguistic evidence from length and context effects. eng. *Frontiers in Psychology* 12. 661087–661087. DOI: 10.3389/fpsyg.2021.661087.

Schuster, Sebastian & Judith Degen. 2020. I know what you're probably going to say: Listener adaptation to variable use of uncertainty expressions. *Cognition* 203. 104285. DOI: 10.1016/j.cognition.2020.104285.

Sedivy, Julie C. 2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research* 32. 3–23. DOI: 10.1023/A:1021928914454.

Stevenson, Rosemary J., Rosalind A. Crawley & David Kleinman. 1994. Thematic roles, focus and the representation of events. *Language and Cognitive Processes* 9(4). 519–548. DOI: 10.1080/01690969408402130.

Taylor, Wilson L. 1953. "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly* 30(4). 415–433. DOI: 10.1177/107769905303000401.

Troyer, Melissa & Marta Kutas. 2020. Harry Potter and the Chamber of What?: The impact of what individuals know on word processing during reading. *Language, Cognition & Neuroscience* 35(5). 641–657. DOI: 10.1080/23273798.2018.1503309.

Venhuizen, Noortje J., Matthew W. Crocker & Harm Brouwer. 2019. Expectation-based comprehension: Modeling the interaction of world knowledge and linguistic experience. *Discourse Processes* 56(3). 229–255. DOI: 10.1080/0163853X.2018.1448677.

Warren, Tessa & Kerry McConnell. 2007. Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review* 14(4). 770–775. DOI: 10.3758/BF03196835.

Westerbeek, Hans, Ruud Koolen & Alfons Maes. 2015. Stored object knowledge and the production of referring expressions: The case of color typicality. *Frontiers in Psychology* 6. DOI: 10.3389/fpsyg.2015.00935.

Yildirim, Ilker, Judith Degen, Michael K. Tanenhaus & T. Florian Jaeger. 2016. Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language* 87. 128–143. DOI: 10.1016/j.jml.2015.08.003.

Chapter 8

Prosodic factors do not always suppress discourse or surprisal factors on word-final syllable duration in German polysyllabic words

Ivan Yuen, Bistra Andreeva, Omnia Ibrahim & Bernd Möbius
Saarland University

Predictability is known to influence acoustic duration (e.g., Ibrahim et al. 2022) and prosodic factors such as accenting and boundary-related lengthening have been postulated to account for this effect (e.g., Aylett & Turk 2004). However, it has also been shown that other factors such as information status or speech styles could contribute to acoustic duration (e.g. Baker & Bradlow 2009). This raises the question as to whether acoustic duration is primarily subject to the influence of prosody that reflects linguistic structure including predictability. The current study addressed this question by examining the acoustic duration of word-final syllables in polysyllabic words in DIRNDL, a German radio broadcast corpus (e.g. Eckart et al. 2012). We analysed polysyllabic words followed by an intermediate phrase or an intonational phrase boundary, with or without accenting, and with given or new information status. Our results indicate that the acoustic duration of the word-final syllable was subject to the effect of prosodic boundary for long host words, in line with Aylett & Turk (2004); however, we also observed additional effects of information status, log surprisal and accenting for short host words, in line with Baker & Bradlow (2009). These results suggest that acoustic duration is subject to the influence of prosodic (e.g., boundary and accenting) and linguistic factors (e.g., information status and surprisal), and that the primacy of prosodic factors impacting on acoustic duration is further constrained by some intrinsic durational constraints, for example word length.

Ivan Yuen, Bistra Andreeva, Omnia Ibrahim & Bernd Möbius. 2024. Prosodic factors do not always suppress discourse or surprisal factors on word-final syllable duration in German polysyllabic words. In Robin Lemke, Lisa Schäfer & Ingo Reich (eds.), *Information structure and information theory*, 215–234. Berlin: Language Science Press.
DOI: 10.5281/zenodo.13383799

1 Introduction

Information-theoretic measures have been used to account for variabilities in word length (e.g., Piantadosi et al. 2011), phrase duration (e.g., Arnon & Cohen-Priva 2013), word duration (e.g., Baker & Bradlow 2009, Seyfarth 2014), syllable duration (e.g., van Son & Pols 2003, Aylett & Turk 2004), vowel spectra (e.g., Aylett & Turk 2006, Brandt et al. 2021), vowel and consonant dispersion (e.g., Malisz et al. 2018), and lenition (e.g., Cohen-Priva 2017). These studies suggest that speakers' choice of phonetic forms is guided by informativity-based considerations, which include frequency (e.g., Gahl 2008, Arnon & Cohen-Priva 2013), or contextual predictability (e.g., Aylett & Turk 2004, 2006, Baker & Bradlow 2009, Seyfarth 2014, Piantadosi et al. 2011).

Adopting the information-theoretic perspective, Aylett & Turk (2004) postulated the smooth signal redundancy hypothesis (SSRH) to explain the acoustic variability of duration in English. According to this hypothesis, prosody directly affects speech acoustics through assignment of prosodic prominence or boundary. For the sake of robust optimal communication, this prosodic influence is inversely related to the influence from language predictability/redundancy factors in order to maintain smooth transmission of information (i.e., to avoid any abrupt surge or dip in information density). While language predictability will induce short acoustic duration, prominence will induce long acoustic duration. For instance, *nine* in *a stitch in time saves nine* will have shorter duration than the same word in *the winning number is nine*, because nine is more predictable and less prominent in the former than the latter. To test their hypothesis, they analysed syllable duration in the HCRC Map Task Corpus (Anderson et al. 1991), and found a significant inverse relationship between the language predictability factors (e.g., log word frequency, syllabic trigram probability and word mention) and syllable duration. They also observed significant influences on syllable duration from a range of prosodic factors (e.g., lexical stress, phrasal stress, different types of prosodic boundary). Further regression analysis revealed that the model with prosodic prominence structure accounted for most of the variance in the syllable duration, with little unique significant contribution from the language predictability factors. On such basis, they argued that prosody absorbs the effects from language predictability to influence speech acoustics. In other words, prosody mediates language predictability.

However, other studies showed that predictability can directly influence duration, rather than be mediated through prosodic prominence structure. For instance, Baker & Bradlow (2009) examined two predictability factors on reduction (i.e., word duration) in two speech styles: plain vs. clear. Plain style was defined as

one in which a speaker hypo-articulates because listeners do not have difficulty in perceiving one's speech; and clear style as one in which the speaker hyper-articulates because listeners might have difficulty in perceiving one's speech. The two predictability factors were first vs. second mention, and word frequency. As expected, word duration in clear speech is longer than that in plain speech for first and second mentions. Similarly, word duration is longer for first mention than second mention in both speech styles. These patterns remained, irrespective of the presence vs. absence of a prosodic break around the measured targets, or the presence vs. absence of accenting on the target stimuli. On the whole, word frequency is negatively correlated with first-mention word duration in plain and clear speech (short duration for high frequency words). However, this frequency-induced reduction effect is exaggerated on second mention duration in plain speech, not clear speech. Baker & Bradlow (2009) found a significant positive correlation between word frequency and second mention in plain speech (suggesting high frequency words undergo more second mention reduction), but not in clear speech. These findings then support the idea that other non-prosody factors such as discourse structure, lexicon-based frequency (predictability) and speech styles contribute to acoustic realization, with the implication that the connection between predictability and acoustics can be direct, in addition to being mediated through prosody.

Consistent with Baker & Bradlow (2009), a recent study reported independent effects of syllable-based surprisal (one type of predictability) and Lombard speech style (i.e., speech produced in a noisy environment) on syllable duration in German lab-speech (Ibrahim et al. 2022). Further evidence for the effect of surprisal was observed on word-final syllable duration in German preceding an intonational phrase (IP) boundary (Andreeva et al. 2020). Based on the analysis of the DIRNDL corpus (Eckart et al. 2012), the authors showed that the duration of a German word-final syllable with high surprisal was longer than that with low surprisal. Critically, they found an interaction between surprisal and the strength of an IP boundary, with the effect of surprisal being more pronounced in the presence of a strong IP boundary. The presence of such interaction indicates that both prosodic boundary and surprisal contributed to the acoustic duration of the word-final syllable in German. Interestingly, the duration of a word-final syllable preceding a strong IP boundary was shorter than that preceding a weak IP boundary.

However, this study did not differentiate between monosyllabic and polysyllabic words preceding an IP boundary. While the location of lexical stress does not vary for monosyllabic words, this cannot be said for polysyllabic words. The number of syllables may then be confounded with pitch accenting (associated

with lexical stress) in influencing the measured syllable duration. Pitch accenting is typically associated with focus or information status (e.g., Cooper et al. 1985, Cruttenden 1993, 2006), although the distribution and use of accenting can be language-specific (e.g., Swerts et al. 2002). For instance, Swerts and colleagues reported Dutch speakers accenting new and contrastive information, but not given information. Cruttenden (1993) showed that speakers of English de-accent repeated or old information; however, this tendency will be attenuated in the presence of a contrast in the discourse. As such, this raises further questions as to whether accenting or information status, or both, contribute to the measured duration and whether or not such effect(s) will interact with surprisal. Besides, it remains to be seen whether the effect of surprisal continues to be observed for other types of prosodic boundary, say intermediate phrase (ip) boundary.

To better understand how discourse-based structure (i.e., information structure), language predictability (i.e., surprisal), prosody (i.e., presence vs. absence of accenting or prosodic boundary types), and/or their interactions might account for the acoustic variability of duration, the current study used broadcast data from an annotated German corpus (DIRNDL) to examine any effects of information status (an aspect of discourse-based information structure) and syllable-based surprisal on word-final syllable durations adjacent to an intonational or an intermediate phrase boundary.

Given the previous finding from Aylett & Turk (2004), we expected prosody (through boundary-related lengthening or accenting) to largely account for the acoustic duration of word-final syllables, mediating any effects of surprisal or discourse (e.g., information status). However, according to observations from Baker & Bradlow (2009), Andreeva et al. (2020) and Ibrahim et al. (2022), we also expected prosodic, surprisal and discourse factors (or their interactions) to contribute to the acoustic duration.

2 Method

We extracted polysyllabic words from the DIRNDL corpus to empirically test whether information status and syllable-based surprisal moderate word-final syllable duration in two prosodic boundary types in German. The DIRNDL corpus (Discourse Information Radio News Database for Linguistic analysis) consists of 5 hours of audio news recordings in German from 9 speakers (5M, 4F) with prosodic annotations for pitch accent types and boundaries according to the GToBI(S) framework (Mayer 1995). The accompanying written scripts were annotated for information status (see Eckart et al. 2012 for details of corpus construction and segmentation).

2.1 Data selection criteria

A total of 3716 polysyllabic words were identified to occur before either an intermediate phrase (as denoted by ip) or an intonational phrase (as denoted by IP) boundary in the DIRNDL corpus. The word-final syllable constitutes the target syllable because it occurs immediately adjacent to a prosodic boundary. As not all of the identified polysyllabic words before a prosodic boundary were annotated for lexical information status, the data set was further trimmed to include only those with a clearly specified information status. Note that information status was grouped into two levels: given vs. new. Items annotated as "accessible" in the DIRNDL corpus were classified as "given" in the current study. This procedure reduced the data set to a total of 2907 items for statistical analysis. Information related to the host word containing the target syllable were extracted from the DIRNDL corpus: speaker identity, speaker gender, identity of the orthographically transcribed host word, phonemic transcription of the target syllable, pitch accent type (if present) for the host word, prosodic boundary (i.e., intermediate or intonational), and lexical information status.

2.2 Language modelling

We estimated the syllable-based surprisal measure in the current study from language models based on the deWaC (deutsches Web as Corpus) corpus (Baroni et al. 2009). The corpus is a collection of web-crawled data containing about 1.7 billion word tokens and 8 million word types from a diverse range of genres such as newspaper articles and chat messages. The corpus was first pre-processed and normalized using German Festival (Möhler et al. 2000). This procedure consisted of removing unnecessary/irrelevant/duplicate document information, for example, web-specific structures such as HTML structures or long lists. After pre-processing, the normalized corpus was divided into a training set (80%) and a test set (20%). Syllable-based trigram language models including word boundary as a unit were trained on the training set using the SRILM toolkit (Stolcke 2002). All language models underwent Witten-Bell smoothing (Witten & Bell 1991). The best-performing trained language model was then used as the default to calculate the conditional probability of a syllable, given the preceding context, i.e.,

$$S(\text{unit}_i) = -\log_2 P(\text{unit}_i|\text{context}) \tag{1}$$

where S = surprisal and P = probability (Hale 2016). The context consisted of two units/states: syllable or/and word boundary. The conditional probability constituted the syllable-based surprisal measure for the target syllables.

2.3 Analysis

Prior to the main analysis, we first checked the estimated surprisal values of the target syllables in the polysyllabic words preceding the two prosodic boundaries and observed two patterns: (a) target syllables preceding an intonational phrase boundary (IP) had overall higher surprisal values than those preceding an intermediate phrase boundary (ip) when the host word contained 2 or 3 syllables, (b) target syllables preceding an intermediate phrase boundary had overall higher surprisal values than those preceding an intonational phrase boundary when the host word contained 4 to 8 syllables. In (a) higher surprisal values were associated with an intonation phrase boundary (IP); whereas in (b) higher surprisal values were associated with an intermediate phrase boundary (ip) (see Figure 1). Because of this, we divided the full data set of 2907 polysyllabic words into two separate data sets to de-confound the effect of surprisal from that of prosodic boundary: 2317 words with no more than 3 syllables and 590 words with no more than 8 syllables (but at least 4 syllables). The former was referred to as *short words*, and the latter as *long words* hereafter. A custom Python script was then used to extract durations of the word-final target syllables from the DIRNDL corpus.

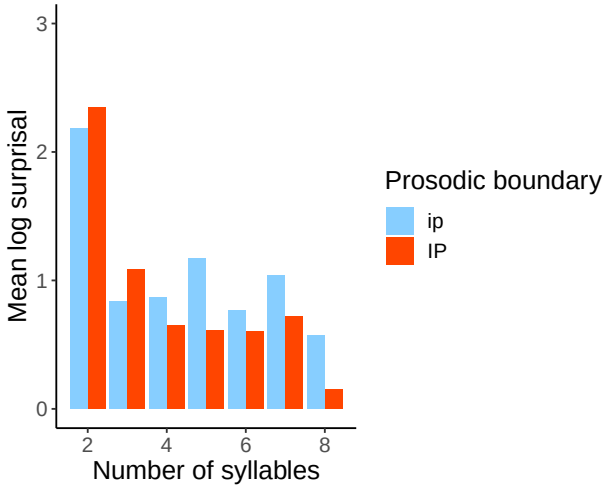


Figure 1: Mean syllable-based log surprisal values of the final syllables according to the number of syllables in the polysyllabic words and prosodic boundary

The duration of the word-final syllable was the dependent variable. Predictors included prosodic boundary type (intermediate vs. intonational), information status (given vs. new), word length (short vs. long), presence vs. absence of a pitch

accent, log surprisal of the word-final syllable, prosodic boundary * information status interaction, presence vs. absence of a pitch accent * information status interaction, log surprisal * prosodic boundary interaction, and log surprisal * information status interaction. Random factors included speaker identity and syllable identity.

3 Results

Linear mixed effects models were then fit to the dependent variable, namely word-final syllable duration, using the lme4 package (Bates et al. 2015) in R (R Core 2022). Multiple random structures were first constructed and compared, using AIC (Alkaike Information Criterion) to determine the optimal random structure as the baseline model. The baseline random structure included by-speaker and by-item intercepts, with prosodic boundary by-speaker slope. Predictors were then included in the baseline model to construct simple and interactive models, which were compared using AIC in order to determine the optimal predictive model. In case of singularity, the complexity of a model structure was reduced to minimize overfitting. Statistical significance of the predictors in the optimal model was then evaluated using the anova() function with Satterthwaite to approximate degrees of freedom. The same procedures were followed in all analyses below. The factors included prosodic boundary, information status, log surprisal of the word-final syllable, presence vs. absence of a pitch accent for the host word, word length, and their interactions. The omnibus analysis revealed significant effects of prosodic boundary, presence vs. absence of a pitch accent, log surprisal, with a significant 2-way prosodic boundary * word length interaction and a significant 3-way prosodic boundary * information status * word length interaction (Table 1). To better understand the 3-way interaction, we analysed short and long words separately.

3.1 Final syllable duration in short words

Figure 2 illustrates the word-final syllable duration in accented or unaccented host words preceding an intermediate or intonational phrase boundary, labelled as having the information status of either “given” or “new”. The word-final syllable duration was increasingly longer when the prosodic boundary immediately following the host word became stronger (i.e., intermediate vs. intonational phrase) and this pattern was magnified when the host word was accented, as well as when the host word was labelled as “new” information. Table 2 summarizes the mean and standard deviation of word-final syllable durations in the 2 prosodic

Table 1: Statistical results of linear mixed effects modelling on word-final syllable durations in all polysyllabic words. *The model: ~ prosodic boundary * information status * word length + presence vs. absence of accent * information status + prosodic boundary * log surprisal + information status * log surprisal + (prosodic boundary | speaker) + (prosodic boundary | syllable item)*

Factors	F	df	*p*
Prosodic boundary (PB)	17	1	<0.001 ***
Information status (IS)	2.6	1	0.11
Log surprisal (S)	46.9	1	<0.0001***
Word length (WL)	1.61	1	0.21
Presence vs. absence of pitch accent (PA)	77.9	1	<0.0001***
PB * IS	1.9	1	0.17
PB * WL	4.9	1	0.03 *
IS * WL	0.5	1	0.5
IS * PA	0.5	1	0.5
PB * S	3	1	0.08
IS * S	1.1	1	0.3
PB * IS * WL	7.5	1	0.006 **

boundary (i.e. intermediate vs. intonational) × 2 information status (i.e. given vs. new) × 2 accenting (i.e. with vs. without) conditions.

Linear mixed effects models were then fit to the dependent variable: word-final syllable duration in short words. The factors in the optimal predictive model included prosodic boundary, information status, log surprisal of the word-final syllable, presence vs. absence of a pitch accent for the host word, and prosodic boundary * information status interaction.

The structure of the final model was ~ prosodic boundary * information status + log suprisal + presence vs. absence of pitch accent + (prosodic boundary | speaker) + (prosodic boundary | syllable item), with significant main effects of prosodic boundary, information status, log surprisal, presence vs. absence of pitch accent and the significant prosodic boundary * information status interaction (Table 3).

As expected, the word-final syllable duration was longer preceding an intonational phrase than an intermediate phrase boundary (prosodic boundary effect). It was longer when the host word contained "new" rather than "given" information (information status effect). It was longer when the host word was accented as opposed to unaccented (presence vs. absence of a pitch accent effect). It was

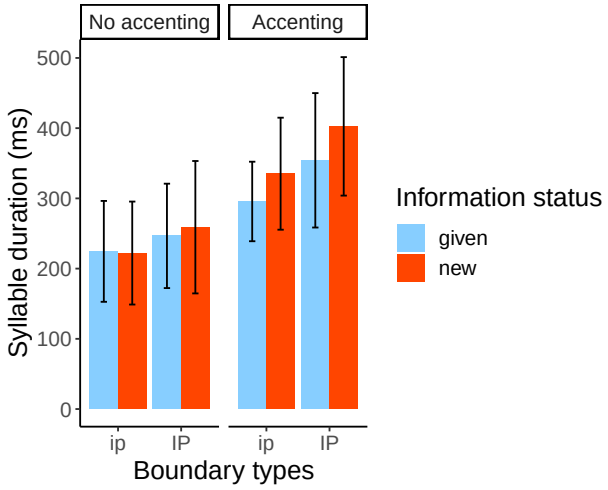


Figure 2: Mean word-final syllable durations in short words, with the following intermediate (ip) or intonational phrase (IP) boundary, with given or new information status, and with or without accent, with +/- 1 SD

Table 2: Mean (SD) word-final syllable durations in short words, with the following intermediate (ip) or intonational (IP) phrase boundary, with given or new information status, and with or without accenting

Accenting	Information status	Prosodic boundary	No. of items	Mean in ms (SD)
No	Given	ip	230	225 (72)
		IP	166	247 (74)
	New	ip	724	222 (73)
		IP	736	259 (94)
Yes	Given	ip	64	296 (57)
		IP	40	354 (96)
	New	ip	173	335 (80)
		IP	184	403 (99)

Table 3: Statistical results of linear mixed effects modelling on word-final syllable duration in all short words. *The model: ~ prosodic boundary * information status + log suprisal + presence vs. absence of pitch accent + (prosodic boundary | speaker) + (prosodic boundary | syllable item)*

Factors	F	df	*p*
Prosodic boundary (PB)	18.3	1	0.001 **
Information status (IS)	4.3	1	0.04 *
Log surprisal (S)	64.5	1	<0.0001***
Presence vs. absence of pitch accent (PA)	93.3	1	<0.0001***
PB * IS	17.7	1	<0.0001***

also longer when the word-final syllable had high log surprisal (log surprisal effect). However, the significant prosodic boundary * information status interaction suggests that the effect of information status on word-final syllable duration was magnified when the immediately adjacent boundary constitutes an intonational phrase. These results were more in line with the predictions from Baker & Bradlow (2009) than those from Aylett & Turk (2004). Counter to our expectations, log surprisal did not interact with information status or prosodic boundary.

To further investigate this, we sub-divided our data into words with vs. without accenting and separately analysed them. The optimal model for the data with accenting included predictors: prosodic boundary, information status, log surprisal, prosodic boundary * information status interaction, information status * log surprisal interaction, and prosodic boundary * log surprisal interaction. The model structure was ~ prosodic boundary * information status + prosodic boundary * log surprisal + information status * log surprisal + (prosodic boundary | speaker) + (prosodic boundary | syllable item), with log surprisal and information status * log surprisal interaction reaching statistical significance (Table 4). Unlike the analysis of all short words, we observed the effect of log surprisal and the log surprisal * information status interaction.

The optimal model for the data without accenting included the following predictors: prosodic boundary, information status, log surprisal, prosodic boundary * log surprisal interaction, prosodic boundary * information status interaction and information status * log surprisal interaction. The structure of the final model for the data without accenting was ~ prosodic boundary * information status +

Table 4: Statistical results of linear mixed effects modelling on word-final syllable duration in short words with accenting. *The model: ~ prosodic boundary * information status + prosodic boundary * log surprisal + information status * log surprisal + (prosodic boundary | speaker) + (prosodic boundary | syllable item)*

Factors	F	df	*p*	
Prosodic boundary (PB)	0.9	1	0.34	
Information status (IS)	0.7	1	0.4	
Log surprisal (S)	20.4	1	<0.0001	***
PB * IS	2.5	1	0.12	
S * PB	1.4	1	0.24	
S * IS	3.9	1	0.05	*

prosodic boundary * log surprisal + information status * log surprisal + (prosodic boundary | speaker) + (prosodic boundary | syllable item), with significant effects of prosodic boundary, information status, log surprisal and the prosodic boundary * information status interaction (Table 5). These results were consistent with the results in the analysis of all short words, suggesting that the overall pattern might be driven primarily by the data without accenting (which contained more items overall).

Table 5: Statistical results of linear mixed effects modelling on word-final syllable duration in short words without accenting. *The model: ~ prosodic boundary * information status + prosodic boundary * log surprisal + information status * log surprisal + (prosodic boundary | speaker) + (prosodic boundary | syllable item)*

Factors	F	df	*p*	
Prosodic boundary (PB)	25.9	1	<0.001	***
Information status (IS)	5.9	1	0.02	*
Log surprisal (S)	40.2	1	<0.0001	***
PB * IS	13.3	1	<0.001	***

The interaction of log surprisal and information status on word-final syllable duration is illustrated in Figure 3 for accented words and the lack of interaction in Figure 4 for unaccented words. In Figures 3 and 4, the word-final syllable duration was lengthened when the log surprisal value was high, as reflected in the positive correlation. However, in Figure 3, the slope between the word-final

syllable durations and the log surprisal values was conspicuously steeper when host words contained new rather than given information. That is, the effect of surprisal on a word-final syllable duration was attenuated when the host word contained "given" information and accenting. Unlike Figure 3, the differences in the slope between given vs. new information in Figure 4 were less obvious.

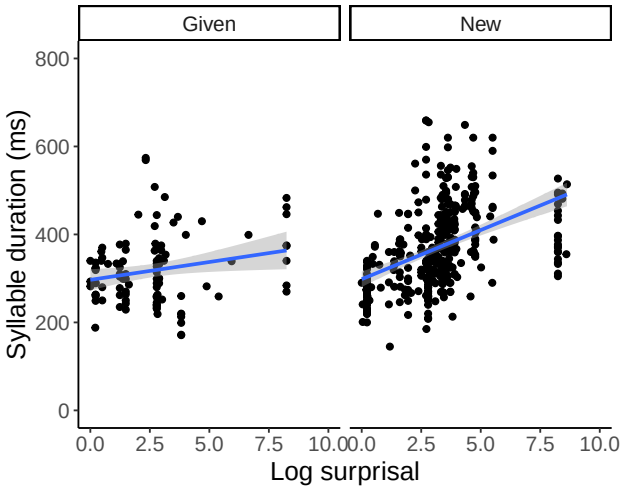


Figure 3: Scatterplots relating word-final syllable durations on the y-axis to syllable-based log suprisal values on the x-axis in accented short host words labelled with given or new information status

3.2 Final syllable duration in long words

A total of 590 long words were included in this analysis. Figure 5 illustrates the patterns of word-final syllable duration preceding two different prosodic boundaries, with either given or new information status, and with or without accenting. Generally, the word-final syllable duration with new information status was longer than that with given information status. This held for unaccented words preceding an intermediate phrase or an intonational phrase boundary; however, the pattern was not as consistent for accented words. Table 6 summarizes the mean duration with SD.

The optimal model included prosodic boundary, information status, log surprisal, presence vs. absence of pitch accent and prosodic boundary * information status interaction as predictors. The model structure of the optimal model was ~ prosodic boundary * information status + log surprisal + presence vs. absence

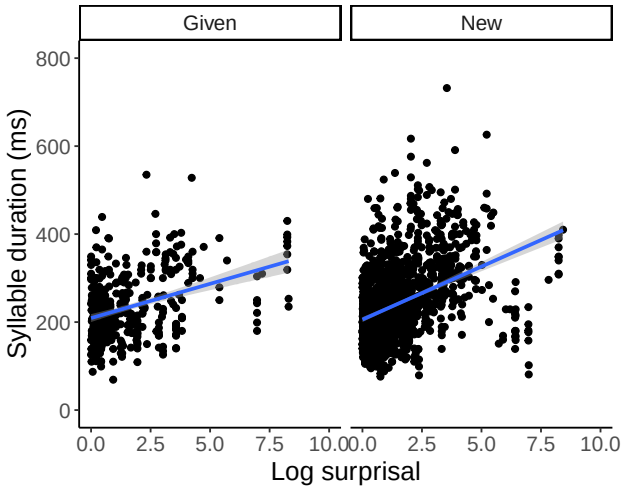


Figure 4: Scatterplots relating word-final syllable durations on the y-axis to syllable-based log suprisal values on the x-axis in unaccented short host words labelled with given or new information status

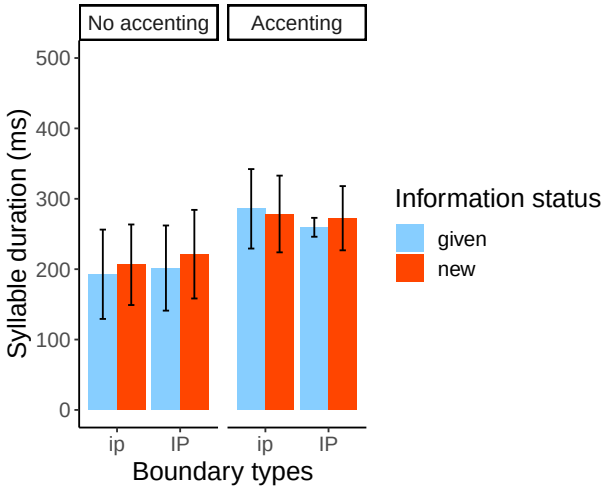


Figure 5: Mean word-final syllable durations in long words, followed by an intermediate (ip) or intonational phrase (IP) boundary, with given or new information status, and with or without accenting, with +/− 1 SD

Table 6: Mean (SD) word-final syllable duration in long words, followed by an intermediate (ip) or an intonational (IP) phrase boundary, with either given or new information status, and with or without accenting

Accenting	Information status	Prosodic boundary	No. of items	Mean in ms (SD)
No	Given	ip	64	193 (63)
		IP	60	202 (61)
	New	ip	232	206 (57)
		IP	213	221 (63)
Yes	Given	ip	4	286 (56)
		IP	2	260 (13)
	New	ip	10	278 (55)
		IP	5	272 (46)

Table 7: Statistical results of linear mixed effects modelling on the final syllable duration in long words. *The model: ~ prosodic boundary * information status + log surprisal + presence vs. absence of pitch accent + (prosodic boundary | speaker) + (1 | syllable item)*

Factors	F	df	*p*
Prosodic boundary (PB)	15.3	1	0.002**
Information status (IS)	2.3	1	0.13
Log surprisal (S)	0.05	1	0.82
Presence vs. absence of pitch accent (PA)	2.5	1	0.11
PB * IS	2.2	1	0.14

of pitch accent + (prosodic boundary | speaker) + (1 | syllable item), with the significant effect of prosodic boundary (Table 7).

Counter to the prediction from Baker & Bradlow (2009), neither the effect of log surprisal nor its interaction with other predictors were observed on the word-final syllable duration (as exemplified by the lack of log surprisal * information status interaction in Figure 6). These results are more in line with the predictions from Aylett & Turk (2004).

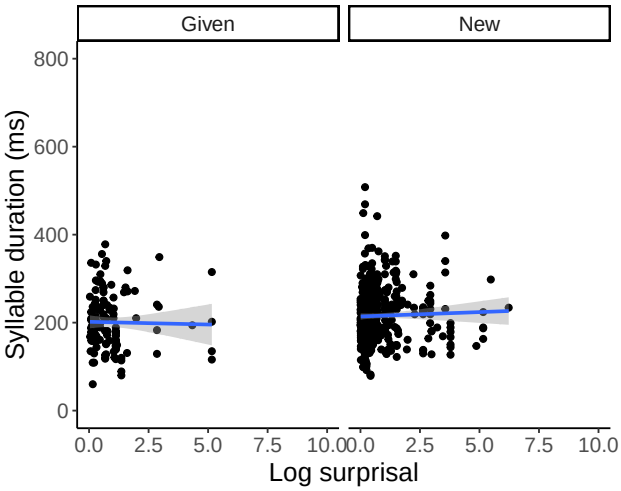


Figure 6: Scatterplots relating word-final syllable durations on the y-axis to syllable-based log surprisal values on the x-axis in all long words with given or new information status

4 Discussion

The goal of the current investigation was to empirically test whether discourse factors such as information status, prosodic factors such as prosodic boundary type, accenting, and surprisal (or their interactions) would contribute to the acoustic realization of the word-final syllable duration in polysyllabic words. We expected the word-final syllable duration with new information status to be longer than that with given information status (e.g., Fowler & Housum 1987, Lam & Watson 2010). We also expected the word-final syllable duration to be longer when followed by an intonational phrase boundary rather than an intermediate phrase boundary (e.g., Wightman et al. 1992). We also expected word-final syllable duration in an accented word to be longer than that in an unaccented word (e.g., Turk & White 1999). We further hypothesized that the word-final syllable duration with high log surprisal would be longer than that with low log surprisal (e.g., Ibrahim et al. 2022). We postulated that prosodic factors would largely account for the acoustic duration without unique contribution from other factors, in line with Aylett & Turk (2004); however, other factors in addition to the prosodic factors might contribute to the measured duration, in line with Baker & Bradlow (2009).

Our overall results on short polysyllabic words are consistent with the effects of boundary-related lengthening, accentual lengthening, surprisal and informa-

tion status on the acoustic realization of syllable duration in previous studies. These results are more in line with Baker & Bradlow (2009) rather than Aylett & Turk (2004), because of additional contributions (including interactions) from prosodic and language predictability factors. However, our overall results on *long* polysyllabic words are more in line with Aylett & Turk (2004) rather than Baker & Bradlow (2009), because we observed only the prosodic boundary effect. Perhaps, polysyllabic shortening constrains the extent to which these various factors can modify syllable duration (i.e., a word length constraint). Despite such a constraint, the durational adjustment is primarily attributable to prosodic boundary type. This suggests a strong tendency for duration to maintain information about a major prosodic boundary. But this interpretation has to be taken with caution because of the low statistical power arising from a relatively smaller data set containing long words (a total of 590 items) than that containing short words (a total of 2317 items).

Depending on whether or not host words were accented, different results were revealed in the data set containing short words. When host words were accented, only the effect of log surprisal and the log surprisal * information status interaction reached statistical significance. However, when host words were unaccented, prosodic boundary, information status, log surprisal and prosodic boundary * information status interaction significantly predicted the word-final syllable durations. As the syllable duration in an unaccented word is shorter than that in an accented word (as reflected in the different y-intercepts in Figure 4 vs. Figure 3), the former might have more room than the latter to accommodate durational increases from multiple sources, resulting in more reported effects and interactions. This interpretation may account for the different results between the unaccented vs. accented words on the assumption that there is an upper duration limit, which seems to be the case, because the majority of the data for word-final syllable duration fell below 600ms (Figure 3 and Figure 4). An alternative interpretation may be related to the statistical power of the relatively small sample size for accented words (461 items) to detect multiple effects as compared to that for unaccented words (1856 items).

Recall that one of our questions is whether information status might interact with the log surprisal effect, presence vs. absence of accenting effect and/or the prosodic boundary effect. Since information status did not have any effect on long words, our discussion focuses on short words. In short *accented* words, information status interacts with log surprisal. The interaction occurs because the effect of log surprisal on word-final syllable duration was stronger for words with new information status than those with given information status. However, no such interaction was observed in *unaccented* words. In that case, information

status interacts with prosodic boundary instead. This interaction is due to the effect of prosodic boundary on pre-boundary syllable duration which is stronger for words with new information than those with given information. In other words, information status additionally exaggerates the effect of log surprisal in accented words, and the effect of prosodic boundary in unaccented words. These results suggest that information status (a discourse-based factor) cannot be subsumed under surprisal (a language predictability factor) or prosodic boundary (a prosodic factor).

These observations based on the corpus data raise further questions as to whether a speaker will differentially weigh these factors (language predictability, information structure and prosody) according to speech styles, e.g., scripted vs. spontaneous or formal vs. informal speech. Speech styles could affect how utterances will be structured, because a speaker might adopt different production planning strategies to cope with time pressure for example. It is possible that less scripted styles might induce more pause breaks. Since these structural considerations can also affect speech acoustics (e.g., Watson & Gibson 2004), it remains to explore in future questions as to how other acoustic cues such as pause might relate to pre-boundary syllable duration.

5 Conclusion

Our overall results showed that information status and surprisal do not encode the same type of linguistic information and that neither information status nor surprisal are redundant. Both can influence the acoustic realization of word-final syllable durations. Critically, our results showed that prosodic factors such as prosodic boundary type could largely account for the acoustic duration as predicted by Aylett & Turk (2004) for long words on the one hand, but that factors other than prosody also contribute to the measured acoustic duration as predicted by Baker & Bradlow (2009) for short words. In other words, information status as a discourse factor can interact with language predictability and prosodic factors to influence the measured acoustic syllable duration, but these interactions are subject to some duration constraint(s) arising from word length.

Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-ID 232722074 (SFB 1102 Information Density and Linguistic Encoding).

References

Anderson, Anne, Miles Bader, Ellen Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry Thompson & Regina Weinert. 1991. The HCRC map task corpus. *Language and Speech* 34(4). 351–366. DOI: 10.1177/002383099103400404.

Andreeva, Bistra, Bernd Möbius & James Whang. 2020. Effects of surprisal and boundary strength on phrase-final lengthening. In *Proceedings in 10th International Conference on Speech Prosody*, 146–150. DOI: 10.21437/SpeechProsody.2020.

Arnon, Inbal & Uriel Cohen-Priva. 2013. More than words: The effect of multiword frequency and constituency on phonetic duration. *Language and Speech* 56(3). 349–371. DOI: 10.1177/0023830913484891.

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. DOI: 10.1177/00238309040470010201.

Aylett, Matthew & Alice Turk. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America* 119(5). 3048–3058. DOI: 10.1121/1.2188331.

Baker, Rachel & Ann Bradlow. 2009. Variability in word duration as a function of probability, speech style and prosody. *Language and Speech* 52(4). 391–413. DOI: 10.1177/0023830909336575.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky Wide Web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3). 209–226. DOI: 10.1007/s10579-009-9081-4.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software* 67. 1–48. DOI: 10.18637/jss.v067.i01.

Brandt, Erika, Bernd Möbius & Bistra Andreeva. 2021. Dynamic formant trajectories in German read speech: Impact of predictability and prominence. *Frontiers in Communication* 6. DOI: 10.3389/fcomm.2021.643528.

Cohen-Priva, Uriel. 2017. Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition* 160. 27–34. DOI: 10.1016/j.cognition.2016.12.002.

Cooper, William, Stephen Eady & Pamela Mueller. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America* 77. 2142–2155. DOI: 10.1121/1.392372.

Cruttenden, Alan. 1993. The de-accenting and re-accenting of repeated lexical items. In *Working papers*, vol. 41, 16–19. Lund: Department of Linguistics & Phonetics. https://journals.lub.lu.se/LWPL/article/download/17147/15524/.

Cruttenden, Alan. 2006. The de-accenting of given information: A cognitive universal? In Giuliano Bernini & Marcia Schwartz (eds.), *Pragmatic Organization of Discourse in the Languages of Europe*, 311–356. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110892222.311.

Eckart, Kerstin, Arndt Riester & Katrin Schweitzer. 2012. A discourse information radio news database for linguistic analysis. In C. Chiarcos, S. Nordhoff & S. Hellmann (eds.), *Linked data in linguistics: Representing and connecting language data and language metadata*, 65–75. Heidelberg: Springer-Verlag Berlin. DOI: 10.1007/978-3-642-28249-2_7.

Fowler, Carol & Jonathan Housum. 1987. Talkers' signalling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26. 489–504. DOI: 10.1016/0749-596X(87)90136-7.

Gahl, Susanne. 2008. Time and thyme are not homophones: The effect of lemma frequency on word duration in spontaneous speech. *Language* 84(3). 474–496. DOI: 10.1353/lan.0.0035.

Hale, John. 2016. Information-theoretical complexity metrics. *Language and Linguistic Compass* 10(9). 397–412. DOI: 10.1111/lnc3.12196.

Ibrahim, Omnia, Ivan Yuen, Marjolein van Os, Bistra Andreeva & Bernd Möbius. 2022. The combined effects of contextual predictability and noise on the acoustic realization of German syllables. *Journal of the Acoustical Society of America* 152(2). 911–920. DOI: 10.1121/10.0013413.

Lam, Tuan & Duane Watson. 2010. Repetition is easy: Why repeated referents have reduced performance. *Memory and Cognition* 38. 1137–1146. DOI: 10.3758/MC.38.8.1137.

Malisz, Zofia, Erika Brandt, Bernd Möbius, Mi Oh Yoon & Bistra Andreeva. 2018. Dimensions of segmental variability: Interactions of prosody and surprisal in six languages. *Frontiers in Communication* 3(25). DOI: 10.3389/fcomm.2018.00025.

Mayer, Jörg. 1995. Transcription of German intonation: The Stuggart system. In *University of Stuggart Technical Report*. Institute of Natural Language Processing. www.ims.uni-stuttgart.de/documents/arbeitsgruppen/ehemalig/ep-dogil/STGTsystem.pdf.

Möhler, Gregor, Antje Schweitzer, Mark Breitenbücher & Martin Barbisch. 2000. *IMS German Festival (version 1.2-os)*. University of Stuggart: Institut für Maschinelle Sprachverarbeitung (IMS).

Piantadosi, Steven, Harry Tily & Edward Gibson. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108(9). 3526–3529. DOI: 10.1073/pnas.1012551108.

R Core, Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133. 140–155. DOI: 10.1016/j.cognition.2014.06.013.

Stolcke, Andreas. 2002. SRILM – an extensible language modeling toolkit. In *Proceedings of interspeech*, vol. 2, 901–904. Denver. DOI: 10.21437/ICSLP.2002-303.

Swerts, Marc, Emiel Krahmer & Cinzia Avesani. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30(4). 629–654. DOI: 10.1006/jpho.2002.0178.

Turk, Alice & Laurence White. 1999. Structural influences on accentual lengthening. *Journal of Phonetics* 27(2). 171–206. DOI: 10.1006/jpho.1999.0093.

van Son, Rob & Louis Pols. 2003. Information structure and efficiency in speech production. In *Proceedings of the 8th Conference on Speech Communication and Technology, Eurospeech*. Geneva. DOI: 10.21437/Eurospeech.2003-63.

Watson, Duane & Edward Gibson. 2004. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes* 19. 713–755. DOI: 10.1080/01690960444000070.

Wightman, Colin, Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91(3). 1707–1717. DOI: 10.1121/1.402450.

Witten, Ian & Tim Bell. 1991. The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression. *IEEE Transactions on Information Theory* 37(4). 1085–1094. DOI: 10.1109/18.87000.

Name index

Information structure and information theory

This volume results from the workshop ”Discourse obligates – How and why discourse limits the way we express what we express” at the 44th Annual Meeting of the German Linguistic Society in Tübingen, Germany. The workshop brought - and this book brings - together information-structural and information-theoretic perspectives on optional variation between linguistic encodings. Previously, linguistic phenomena like linearization, the choice between syntactic constructions or the distribution of ellipsis have been investigated from an information-structural or information-theoretic perspective, but the relationship between these approaches remains underexplored.

The goal of this book is to look more in detail into how information structure and information theory contribute to explaining linguistic variation, to what extent they explain different encoding choices and whether they interact in doing so. Using experimental and corpus-based methods, the contributions investigate this on different languages, historical stages and levels of linguistic analysis.