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

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.