Information structure
Jaakko Leino

1.   Introduction

The term *information structure* was introduced by M. A. K. Halliday (1967). Halliday distinguishes between three "areas of syntactic choice" which affect the realization of clauses: transitivity, mood, and theme. Theme, he states, "is concerned with the information structure of the clause; with the status of the elements not as participants in extralinguistic processes but as components of a message; with the relation of what is being said to what has gone before in the discourse, and its internal organization into an act of communication". Thus, the study of information structure focuses on the positioning of clause elements in discourse, rather than e.g. the positioning of their referents in the events described by clauses.

Many of the central ideas behind work on information structure by Halliday as well as more recent works owe much to previous work by members of the Prague School (notably Daneš, e.g. 1964, and Firbas, e.g. 1962, 1964). Not all linguists employ the term *information structure,* however; e.g. Daneš speaks of *Functional Sentence Perspective,* other members of the Prague School (e.g. Firbas and Mathesius) of *Communicative Dynamism* and *organization of utterance,* and Chafe (1976, 1994) and Prince (1981) of *information packaging.*

Information structure is without any doubt the pragmatically oriented phenomenon or subject area that has received the most attention in the context of Construction Grammar (CxG). In the context of and in relation to CxG, Information Structure has been most studied by Knud Lambrecht (e.g. Lambrecht 1981, 1984, 1994, 2000, 2001, 2004; Michaelis and Lambrecht 1994, 1996). Lambrecht, like a number of other linguists working with information structure, has worked extensively on clefted and dislocated structures and their functions.

Lambrecht defines the concept of information structure as follows

> "**Information structure:** That component of sentence grammar in which propositions as conceptual representations of states of affairs are paired with lexicogrammatical structures in accordance with the mental states of interlocutors who use and interpret these structures as units of information in given discourse contexts." (Lambrecht 1994: 5)

Lambrecht's own paraphrase for this definition is that the information structure of a sentence is the formal expression of the pragmatic structuring of a proposition in a discourse. Lambrecht lists the following as "the most important categories of information structure" (id. 6):

1) Presupposition and assertion, which have to do with the structuring of propositions into portions which a speaker assumes an addressee already knows or does not yet know
2) Identifiability and activation, which have to do with a speaker's assumptions about the statuses of the mental representations of discourse referents in the addressee's mind at the time of an utterance
3) Topic and focus, which have to do with a speaker's assessment of the relative predictability vs. unpredictability of the relations between propositions and their elements in given discourse situations.

These characterizations are very abstract, and applying them to authentic sentences and utterances is not always unproblematic. I will return to each one in more detail in sections s 3.1–3.3. Put simply, what Lambrecht means by (pragmatic) presupposition is the "old" information contained or evoked by the sentence, and what he means by (pragmatic) assertion is the "new" information expressed or conveyed by the sentence (id. 52).[1] Thus, for example the sentence in (1) presupposes several propositions:[2]

(1)   Mary heard yesterday that her brother has finally quit smoking.

---

[1] I shall return to Lambrecht's notions of *pragmatic presupposition* and *pragmatic assertion* in some more detail in section 3.1.
[2] For a more thorough discussion of a similar example, see Lambrecht (1994: 55–56).

The sentence evokes (at least) the following propositions:

1. the addressee is able to identify the person referred to as Mary
2. the person referred to as Mary has a brother
3. the person referred to as Mary's brother previously had a habit of smoking (tobacco)
4. the person referred to as Mary has expected (or wished) that her brother will quit smoking
5. the quitting event took place before the person referred to as Mary was informed of it.

The assertion made in the sentence in (1) can be expressed as follows: "propositions 1–4 are taken for granted; the person referred to as Mary has learned, on the day before the speech event, that her brother has quit smoking".

One thing which remains open is whether the proposition expressed by the complement clause *that her brother has quit smoking* is presupposed or asserted. The traditional analysis is that it is presupposed and not asserted. However, it may also be argued that whether the proposition is presupposed or asserted may depend on the discourse context. I will return to this point in section 3.1.

While presupposition and assertion relate to propositions and e.g. their givenness and acceptability, identifiability and activation relate to referents and their discourse properties. As Lambrecht points out (1994: 77–78), the distinction between identifiable and unidentifiable referents is similar to that between pragmatically presupposed and pragmatically asserted propositions. The speaker and the hearer have shared knowledge of a presupposed proposition prior to the utterance, while only the speaker has knowledge of an asserted proposition before the utterance. Correspondingly, the speaker and the hearer have a shared representation of an identifiable referent at the time of the utterance,

while only the speaker has a representation of an unidentifiable referent at the time of the utterance.[3]

Thus, a central question behind identifiability is whether or not the speaker expects the hearer to understand, at the time of the utterance, which object or proposition the speaker has in mind when they refer to that object or proposition in the sentence. Activation, in turn, concerns the question as to whether or not the referent—object or proposition—is present in the universe of discourse:[4] is it present in the speech situation, has it been already mentioned in the discourse, is it "reachable" through some other already active referent, etc.

Overall, as e.g. Kuningas (2007: 26) points out, we need information structure as the "third component" of grammar in order to explain *why,* in a given context, the speaker chooses to use a given syntactic structure or expression type (say, a cleft sentence or passive voice) rather than some other one (e.g. the canonical sentence structure or active voice). Typically, if not always, there are several possible ways of expressing the same proposition—labeled *allosentences* by Lambrecht (1994)—but their suitability to any given context varies and is dependent, crucially, of such factors as what the interlocutors expect each other to know beforehand, what has and what has not been already mentioned in the same discourse, and the like.

Several construction grammarians—notably, Lambrecht himself included—have incorporated information structure notions as parts of their frameworks. Categories like those presented above have proven useful in the descriptions of various constructions and clause-level phenomena. As an early example, Lambrecht (1990) discusses what he calls the "Mad Magazine sentence", illustrated by the sentence *What, me worry?* and shows convincingly that a proper understanding of the functions of this sentence type requires an understanding of its information structure properties. Briefly, the sentence (or

---

[3] This rather black-and-white distinction is an overgeneralization. For a broader point of view concerning the identifiability of a referent, see e.g. Dubois (1980), Chafe (1994: 93–107), as well as the discussion in Lambrecht (1994: 77–92) and the references cited in these works.

[4] By the *universe of discourse,* Lambrecht refers to an abstraction of the participants, objects, propositions, etc. which "exist" for the purposes of the discourse (cf. Lambrecht 1994: 36–43). It consists of the interlocutors (or "speech participants"), the speech setting, and the "text-internal world" (which, in turn, consists of linguistic expressions and their meanings). The universe of discourse regulates e.g. what is and what isn't "present" in terms of speech.

"conventionalized sentence fragment", as Lambrecht calls it) presents topical information, and is often followed by a "follow-up proposition" which is focal.

Lambrecht, like several others, also spells out information structure requirements or constraints in feature matrix descriptions of constructions. For instance, Lambrecht (2004) presents an analysis of the French *comme*-N construction (e.g. *C'est pas marrant, comme histoire.* 'That's not a funny story.') in which the description of the first nominal element (in the previous example, the pronominal *ce,* truncated as *c'*) includes the feature [ACT active]. In other words, this sentence type requires that the first nominal element of any actual sentence instantiating this sentence type be active in the present discourse. What this means in actual practice will be discussed in section 3.2 below.

Information structure phenomena are often associated with something akin to sentence types, or expression types. The English "Mad Magazine sentence" and the French *comme*-N construction are cases in point, and also such phenomena as cleft sentences, dislocations and extraposition structures have often been studied with an information structural point of view. Construction Grammar approaches grammar in a manner which makes it relatively easy to describe the properties of e.g. single NPs as dependent on their position in a larger conventionalized structure, it seems like no surprise that information structure has attracted a lot of interest within Construction Grammar—and, conversely, that Construction Grammar seems like an attractive framework for information structure researchers.

2.   Information structure and grammar

Research on information structure focuses on issues like the givenness and accessibility of information, the organization of information in sentences and in discourse, and various related topics. What, then, is the place of information structure in the language system, or grammar? How does it relate to syntax or semantics? Are information structure phenomena semantic by nature, or are they primarily connected to the structure and organization of sentences, and thereby to syntax?

Lambrecht (1994: 6–13) presents information structure as the third part of a tripartite language system, on a par with syntax and semantics. He cites several linguists

who have proposed fundamentally similar architectures of grammar, comprising three parts or levels. According to this line of thought, grammar consists of form (or formal structure), or (morpho)syntax; meaning (and its structure), or semantics; and a third part which doesn't have an equally conventional name, though in all cases cited by Lambrecht, this "third part" of grammar has a strong pragmatic flavor to it.

While a steadily increasing number of linguists consider information structure an integral part of the language system, it is quite common that in actual practice, compared to (morpho)syntax and semantics, information structure is the "odd one out", and is often left out of grammars and descriptions of grammatical phenomena. Östman and Virtanen (1999: 92–93) point this out as follows:

> "It is true that, in very general terms, linguists do take Information Structuring, or (following Case and Construction Grammar terminology) 'rhetorical relations' into account on a par with—and as being in principle equal in importance to—investigations of the situational frame-roles (like 'buyer' and 'seller' in a commercial transaction), semantic, or theta roles (like Agent, Patient, and Experiencer), and grammatical relations or functions (like Subject and Object). But the specific nature of these rhetorical roles is often left vague. In many theories—including most versions of Construction Grammar (cf. Fillmore & Kay, forthcoming; but see Lambrecht 1994)—they seem so far to have been relegated to the status of 'extras' that might have to be referred to at some point, in order to understand and be able to describe some constructions, but most of the time, for most languages, linguists seem to be quite happy not referring to them.
>
> And yet, Information Structure in a very obvious sense forms the bridge between grammar and use, an interface with a 'missing link' status for understanding language."

Information structure is, thus, an element of sentence grammar, supplementary to morphosyntax and semantics. It is concerned with the manner in which the message is conveyed, or with the question as to *why* the speaker makes the particular syntactic and semantic choices and uses the particular expression types rather than some other ones (cf. e.g. Kuningas 2007: 26), or, more broadly, "why there are so many kinds of sentence structures" (Lambrecht 1994: 9). This last point may be illustrated with a parallel pointed out by Lambrecht (ibid.):

> "If ecology focuses on the interaction of organisms and their environments, the study of information structure focuses on the interaction of sentences and their contexts."

In other words, while the ecologist asks why there are so many species of living organisms, the (functional) linguist asks why there are so many competing grammatical structures. We may push this parallel one step further by thinking of different communication situations as the "living conditions" of linguistic expressions and expression types. Biology and ecology speak of *ecological niches.* Those species survive which have a "suitable slot" in the ecosystem: sufficient nutrition, not too many enemies, etc. Similarly, linguistics could (and sometimes does) speak of *grammatical niches.* Those expression types survive—grammaticalize, or remain parts of the expressive apparatus of the language—which have a slot in the language system: sufficient expressive need, not too many competing expression types, etc.

3. Central concepts of information structure

In order to discuss information structure in any detail, and to illustrate the topic with some concrete examples, some basic concepts need to be introduced. This discussion cannot possibly do justice to the wide spectrum of approaches and plethora of literature on these and related concepts. Due to a vast oversupply of competing theories of and approaches to information structure, and due to limitations of space, I shall focus mainly on the ideas and analytical tools presented in Knud Lambrecht's work, notably in his 1994 book *Information Structure and Sentence Form,* which is by far the single most influential work on information structure written in the spirit of Construction Grammar.

3.1 Presupposition and assertion

The notions *pragmatic presupposition* and *pragmatic assertion,* as defined by Lambrecht (1994), were discussed briefly in section 1. It was pointed out that the sentence *Mary heard yesterday that her brother has finally quit smoking* presupposes a number of things (see the discussion above) and asserts, essentially, that 'the person referred to as Mary has now learned that her brother has quit smoking'. However, it was left open as to whether the proposition expressed by the complement clause *that her brother has quit smoking* is presupposed or asserted. The answer to this question will ultimately depend on how exactly we understand the notions *presupposition* and *assertion.*

Lambrecht speaks, specifically, of *pragmatic* presupposition and assertion. He defines these notions as follows (1994: 52):

- Pragmatic presupposition: The set of propositions lexicogrammatically evoked in a sentence which the speaker assumes the hearer already knows or is ready to take for granted at the time the sentence is uttered.
- Pragmatic assertion: The proposition expressed by a sentence which the hearer is expected to know or take for granted as a result of hearing the sentence uttered.

A traditional analysis would state that the complement-taking verb *hear* is factive (in the sense of Kiparsky and Kiparsky 1971) and, therefore, the proposition expressed by the complement clause is, by definition, presupposed. However, it is not obvious that the traditional analysis is coherent with Lambrecht's definitions of pragmatic presupposition and assertion.

First of all, the fact that Mary's brother has quit smoking may or may not have been discussed earlier during the same discussion. If it has, then the speaker may rightfully assume that "the hearer already knows or is ready to take for granted" this proposition "at the time the sentence is uttered"—hence, the proposition would be pragmatically presupposed. If it has not, the speaker may expect the hearer "to know or take for granted" this proposition "as a result of hearing the sentence uttered"—and hence it would be pragmatically asserted, despite the fact that the proposition comes up in a complement clause.

Second, however, it is perhaps problematic to say that the speaker expects the hearer "to know or take for granted" the proposition expressed by a complement clause "as a result of hearing the sentence uttered". Traditionally,[5] asserting has been taken to be a property of main clauses, and the fact that Lambrecht's definition of pragmatic presupposition does not, strictly speaking, fit together with our example does not entail that his definition of pragmatic assertion does. Furthermore, Lambrecht defines pragmatic

---

[5] It must be pointed out, however, that speaking of "traditional" analyses of presuppositions and assertions is a gross oversimplification. As already Levinson (1983: 167) points out, "there is more literature on presupposition than on almost any other topic in pragmatics", and the amount of relevant literature has multiplied since he made this observation.

presupposition as a *set of propositions,* but pragmatic assertion as *the proposition* expressed by a sentence. If this is to be taken literally, then pragmatic assertion would apparently only include the proposition expressed by the main clause.

Instead of pushing this discussion too far, it may be more useful to think of the propositions expressed by complement clauses as discourse referents. This makes it more relevant to speak of their *activation status* and their role as the *topic* or *focus* of the sentence rather than their presupposedness or assertedness.

3.2   Activation

Another central concept in information structure besides presupposition and assertion is the *activation state* of a referent. A relatively widespread account of activation states, particularly within CxG, is that formulated by Chafe (1987) and subsequently used by e.g. Lambrecht (1994) as well as a number of others. The notion of activation state refers to the presence or absence of the referent in the speaker's and the hearer's consciousness. The three states proposed by Chafe are *active, semi-active* (also often called *accessible*) and *inactive.* This three-step division is closely related to the classical division of *given* vs. *new,* supplemented by the third step, previously proposed e.g. by Prince (1981) by the name *evoked.* The three states can be characterized as follows, in the spirit of Chafe (1987: 25) and Lambrecht (1994: 94):

– An *active* referent is conceptually present in the person's focus of consciousness. It can be said to be in the limelight (or "currently lit up", as Chafe puts it).[6]
– A *semi-active* (or *accessible*) referent is in the person's consciousness only peripherally but is available to attention despite the fact that it is not in focus.
– An *inactive* referent is not even peripherally in the person's consciousness at the moment but, rather, stored in the long-term memory.

---

[6] Chafe speaks of activation states as properties of *concepts.* I prefer to follow Lambrecht in interpreting activation statuses as properties of referents (or their mental representations).

A referent can be semi-active for two general reasons. First, it may have lost part of its previous activeness ("typically by having been active at an earlier point in the discourse" according to Chafe). Secondly, a referent may be active because it is a part of a frame evoked by the discourse.[7] Thus, speaking of a right triangle (to use a classic example used e.g. in Langacker 1986: 7 and in Goldberg 1995: 26) evokes the concept of hypotenuse: it becomes accessible, or semi-active. However, merely speaking of a right triangle does not bring the hypotenuse to the focus of attention to the extent that it would become a fully active referent.

The activation status of a referent may be a relevant feature with regard to the pragmatically adequate use of a given sentence structure. The French *comme*-N construction mentioned in section 1 is a case in point, and others will be discussed in section 4. In a limited set of instances, activation status may be set even for a single lexeme, a pronoun or a determiner in particular (e.g. Gundel and Fretheim point out that "the proximal demonstrative determiner *this/these* codes the cognitive status 'activated'"), but it is much more common that activation status is determined by a clause-level construction, word order, or some other non-lexical factor. Therefore, looking at constructions in a holistic manner is necessary for understanding the mechanisms in which the activation status of a given referent is coded in the linguistic expression associated with that referent.

The recognizability, accessibility, and activation of referents relates first and foremost to the image which the interlocutors have of one another's consciousness, and to how speakers structure their utterances to fit what the hearers do and do not know. As Tomlin et al. put it (1997: 80), "If the speaker assumes, prior to uttering an intonation unit, that a concept is already active in the listener's mind, he will verbalize that concept in an attenuated manner, most probably proniminalizing it. If he assumes that a concept is not presently activated in the listener's consciousness, he will verbalize that concept in a less attenuated manner, most probably nominalizing it."

---

[7] Both Lambrecht and Chafe speak of a *schema* in this context. However, as Lambrecht himself points out (1994: 99), "Chafe's notion of a schema and its associated expectations, which he takes from cognitive psychology, is closely related to the Fillmorean notion of a semantic frame".

In order for the speaker to be able to accommodate his utterance to the speaker's consciousness, he naturally needs to have knowledge or at least assumptions about what the hearer knows beforehand and what is and what is not currently active in the hearer's mind. To put it simply, the basic idea is that both the speaker and the hearer have a *mind* with *thoughts* and *concepts* in it. Among these thoughts and concepts are assumptions about what kinds of thoughts and concepts exist in the other person's mind; we might call these *projections* of the other person's mind. Thus, the speaker has in his mind a projection of the hearer's mind, and vice versa:
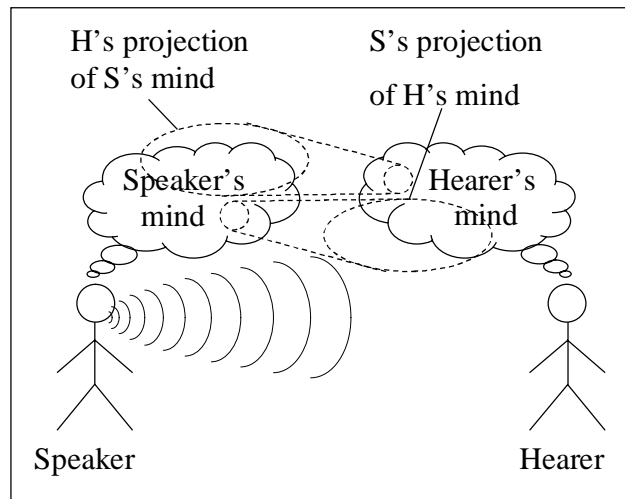

Figure 1: The Speaker – Hearer setting.

Different linguistic units may contain information about what these projections are like. For instance, the use of indefinite and definite articles has to do with whether or not the speaker expects the hearer to recognize the referent. Similarly, what Chafe, Lambrecht, and others have written concerning the activation statuses of concepts and/or referents also relates to this setting: inactive status applies to referents which the speaker doesn't assume to "exist" in the hearer's mind, i.e. which are not included in the speaker's projection of the hearer's mind. The difference between active and semi-active referents, in turn, depends on the status of the referent in the speakers projection of the hearer's mind and whether the speaker assumes the hearer to be able to access the referent.

This overall setting is simple to the point of being somewhat naïve. Yet, there is plenty of evidence to support it. One piece of evidence is the *theory of mind* approach to autism (originally Baron-Cohen et al. 1985, in more detail and illustration e.g. Happé 1994: 38–44). The main idea behind this line of research is, roughly, that autistic individuals lack (to some extent) the capability to conceive of the fact that other individuals have knowledge and beliefs which differ from their own knowledge and beliefs. This may be reformulated by saying that autistic individuals either lack the capability or have a different way of forming projections of other people's minds.

Happé (ibid.) illustrates the theory-of-mind-related deficit with a simple test setting known as the *Smarties task* (Perner et al. 1989). In the Smarties task, the subject is first shown a closed candy box which contains a pencil, and asked what is in the box. Typically, the subject—whether autistic or not—replies that the box contains candy. Next, the subject is shown that the box in fact contains a pencil. Then, the subject is told something like the following: "In a minute, Billy will come in. I will show the box to him like I showed it to you, and ask what is in it. What will he say?" A four-year-old non-autistic child is able to reply this question taking into account the fact that Billy doesn't know the contents of the box. An autistic child, in contrast, assumes Billy to reply that the box contains a pencil.

I pointed out in section 1 that, in Lambrecht's view, an identifiable referent is one for which there is a shared representation in the minds of the speaker and the hearer, whereas for an unidentifiable referent, there only is a representation in the speaker's mind. The whole concept of identifiability is thus dependent on the setting shown in figure 1, i.e. on how the interlocutors form projections or ideas of one another's mind and consciousness.

3.3 Topic and focus

Lambrecht builds most of his information structure theory on the notions of *topic* and *focus.* He devotes most of his 1994 book to a discussion of these two concepts and analyses which illustrate different kinds of topic and focus structures. He defines topic and focus as follows:

> **Topic:** that which the sentence is about. (Lambrecht 1994: 118)
>
> **Focus:** The semantic component of a pragmatically structured proposition whereby the assertion differs from the presupposition. (id. 213)

As vague as this definition of *topic* may sound, it is fairly widely used (for discussion on the notion and its definition, see e.g. Lambrecht 1994: 117–131, Erteschik-Shir 2007: 7–27).

A range of phenomena typically used for illustrating the concept of topic is *topicalization,* or *left dislocation* (though the two are not completely synonymous). Erteschik-Shir gives the following Danish examples (2007: 7–8):

(2) a.   Hun hilste på Ole. <u>Ham</u> havde hun ikke mødt før...
           She   greeted Ole. Him had   she not met before

     b.   Hun hilste på Ole. Hun havde ikke mødt <u>ham</u> før...
           She   greeted Ole. She had   not met him before

The examples (2a) and (b) differ with regard to the topic. In (2a), the object pronoun *ham* is topicalized and occurs in the sentence initial position. Correspondingly, the sentence is interpreted as saying something about the referent of the pronoun, i.e. about Ole. In (2b), the object pronoun is in its unmarked position after the verb, and the sentence is construed (most probably) as being about the person referred to as *she*.

Erteschik-Shir (2007: 27) interprets Lambrecht's definition of *focus* as "the non-presupposed information in the sentence". She illustrates the concept with the following example (id. p. 28) which, in her analysis, shows the presupposition in *italics* and the focus in CAPITALS:

(3) a.   Q: What did John do?
           A: *He* WASHED THE DISHES.

b. Q: What did John wash?

A: *He washed* THE DISHES.

c. Q: Who washed the dishes?

A: JOHN *washed them.*

d. Q: What happened to the dishes?

A: JOHN WASHED *them.*

e. Q: What happened?

A: JOHN WASHED THE DISHES.

f. Q: What did John do with the dishes?

A: *He* WASHED *them.*

In (3a), the subject pronoun *he* is the only element in the sentence which refers to a presupposed entity. The rest of the sentence, i.e. the whole VP, is thus the focus. In (3b), not only the subject but also the event type is presupposed: in other words, it is established in the discourse that John washed something. The non-presupposed part, and hence the focus, is what John washed, i.e. the dishes. In (3c–f), the presupposed parts of the sentence vary, and the focus varies accordingly, one extreme being (3e) which has no presupposed elements. Hence, it is a *sentence focus* structure in Lambrecht's terms (1994: 233–235).[8]

4. Information structure and sentence form

With regard to Construction Grammar, the most important aspect of information structure phenomena is, perhaps, the influence which they have on the formal structure of linguistic expressions. To the extent that information structure phenomena are

---

[8] For a discussion of this term and its relationship to the perhaps more widespread notion of *thetic sentence* (as opposed to *categorical sentence*), see Lambrecht (1987).

conventionalized in a language, they must be coded as properties of constructions (at least if we adopt the view that the grammar of a language consists solely of constructions). Therefore, a proper and comprehensive account of the grammar of any given language will include information structure, and not as a separate addendum or a pragmatic component but, rather, as a set of features and properties which any of the grammatical constructions in that language may have.

To this end, together with Laura Michaelis, Lambrecht has proceeded to implement his information structure theory in the feature matrix notation used in Construction Grammar, notably in the branch of CxG referred to as "Berkeley Construction Grammar" by Fillmore (this volume) and "Unification Construction Grammar" by Goldberg (2006: 213–217). They have incorporated into the CxG feature matrices information concerning e.g. the activation statuses of each part of a sentence, and the topic and focus of a sentence (see Michaelis and Lambrecht 1996, Lambrecht and Michaelis 1998).

Central to the work of Michaelis and Lambrecht is the fact that information structure properties, including activation status, topic vs. focus role, and presupposedness, tend to be conventionalized—or, indeed, grammaticalized—as parts of sentence-sized constructions. This, and more generally integrating information structure features in sentence grammar, may be illustrated with their account of the English nominal extraposition construction—or, more accurately, the sentence (or "construct") *It's amazing, the difference* which instantiates the nominal extraposition construction (Michaelis and Lambrecht 1996: 227):

```
┌─────────────────────────────────────────────────────────────────────────────┐
│  syn  [ cat V, max +, hsbj + ]                                                │
│  sem  #3                                    Inherit Subject-Predicate          │
│  ┌──────────────────────────────────────────────────────────────────────┐    │
│  │  syn  [ cat V, hsbj +, max + ]                                         │    │
│  │  sem  #3                                                               │    │
│  │  ┌──────────────────┐  ┌──────────────────────────┐  ┌──────────────┐ │    │
│  │  │ syn ⎡cat N      ⎤ │  │ syn [ cat V, max +, hsbj+]│ │ syn [cat N, max+]│   │
│  │  │     ⎣max +, pro+⎦ │  │ sem  #3                   │ │ sem  #1        │ │    │
│  │  │ sem  #1           │  │ prag focus                │ │ role ⎡grfn null⎤│    │
│  │  │ role ⎡grfn subj⎤  │  │ val  ⎡syn N, pro +    ⎤   │ │      │theta #2 ││   │
│  │  │      │theta #2 │  │  │      │sem  #1         │   │ │      ⎣prag topic⎦│   │
│  │  │      ⎣prag topic⎦ │  │      │role ⎡grfn subj⎤│   │ │                 │   │
│  │  │ act    active +   │  │      ⎣     ⎣theta #2⎦⎦   │ │ act    active + │ │   │
│  │  │ lfm    it         │  │ lfm      is AMAZING       │ │ lfm  the difference│  │
│  │  └──────────────────┘  └──────────────────────────┘  └──────────────┘ │    │
│  └──────────────────────────────────────────────────────────────────────┘    │
└─────────────────────────────────────────────────────────────────────────────┘
```
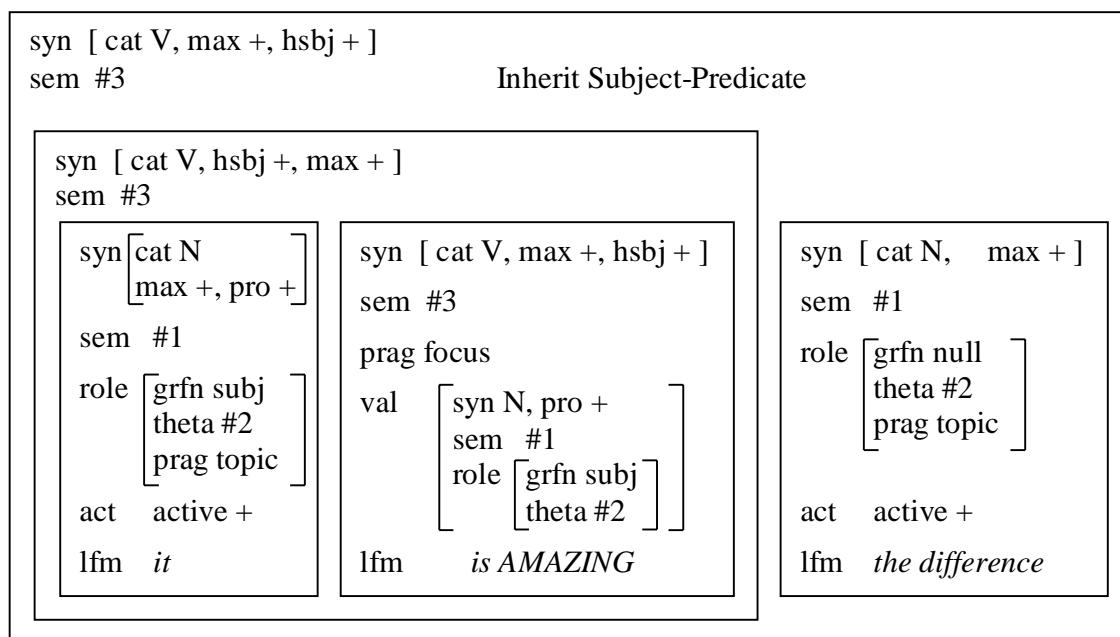
Figure 2 (taken from Michaelis and Lambrecht 1996: 227).

As the figure shows, the subject position of the sentence is occupied by the pronoun *it*.
Yet, the sentence also includes the coreferential NP *the difference*. The construction
instantiated by this sentence includes the information that these two represent the same
theta role (#2) and that their common referent is active in the discourse. The same
referent is also the topic of the sentence. In other words, the speaker assumes the referent
of the NP *the difference* to be active in the hearer's consciousness at the time of the
utterance, and he expresses, by uttering this sentence, a proposition which concerns that
particular referent (i.e. the idea that the referent is amazing).

One of the central motivating forces behind information structure research is the
observation that any given language makes it possible to express the very same
propositional meaning often with quite a number of formally different sentences. In this
context, Lambrecht uses the term *allosentence* (presented in Daneš 1964). The term has
its roots in the structuralist tradition: morphology speaks of morphs, morphemes, and
allomorphs, and phonology speaks of phones, phonemes, and allophones. Similarly,
sentence grammar may speak of allosentences. The idea is that in the sentence grammar

context, *propositions* correspond to the "emic" units. Thus, different sentences expressing the same proposition are considered allosentences of that proposition.[9]

As phonology describes the contexts of occurrence for each of the allophones of a given phoneme, a grammar which takes information structure seriously describes the contexts of occurrence for different allosentences of a given proposition. This idea may be illustrated by a set of sentences exemplifying word order variations in Finnish. E.g. the sentences in ((4a–d) all express the same proposition and are its allosentences:[10]

(4)　a.　Kalle　　　antoi　　　　omenan　Villelle.
　　　　name-NOM give-PST-3SG apple-ACC name-ALL
　　　　'Kalle gave the/an apple to Ville.'

　　　b.　Villelle　Kalle　　　omenan　　antoi.
　　　　name-ALL name-NOM apple-ACC give-PST-3SG
　　　　'It was to Ville that Kalle gave the/an apple.'

　　　c.　Antoi　　　　Kalle　　Villelle　omenan.
　　　　give-PST-3SG name-NOM name-ALL apple-ACC
　　　　'Kalle did give the/an apple to Ville.'

　　　d.　Omenan　Ville　　　Kallelle　antoi.
　　　　apple-ACC name-ALL name-NOM give-PST-3SG
　　　　'It was the/an apple that Kalle gave to Ville.'

Each one of these sentences is natural in a different set of contexts, and their naturalness and suitability to context is regulated first and foremost by information structure: what the speaker and the hearer know beforehand, what has and has not been discussed previously in the same discourse, what is it that is being talked about, etc.

As a side note, Lambrecht specifically warns of the fallacy of a "pragmatically neutral" sentence. There can be no pragmatically neutral allosentence, since every allosentence has its pragmatic usage conditions:

---

[9] As Mukherjee (2001: 89) points out, Lambrecht uses this term more broadly than Daneš, who only uses the term for different word order variations of the same "sentence pattern" (cf. Daneš 1964: 233).
[10] NOM = nominative case, PST = pas tense, 3SG = 3rd person singular, ACC = accusative case, ALL = allative case.

"Just as there are no sentences without morphosyntactic and phonological structure, there are no sentences without information structure. Saying that some syntactic or prosodic structures 'have a special pragmatic function' while others do not is somewhat like saying that some mechanical tools have a special function while others are functionally neutral. According to this logic, a screwdriver for example would be said to have a 'special function' because the objects manipulated with it (i.e. screws) must have a special shape, while a hammer would be said to be functionally neutral because it may be used to drive in various kinds of objects including nails, fence poles, and if need be even screws." (1994: 16–17)

Thus, e.g. the examples ((4a–d) all have their specific information structures and possible discourse functions, which may be partially overlapping and partially separate. One might claim that the sentence ((4a) is unmarked as compared to ((4b–d) in the sense that it apparently has a broader set of possible contexts.[11] However, it is certainly not without a pragmatic function, constraints on possible usage contexts, and information structure.

The term *allosentence* was first introduced by Daneš to capture word order variations like these, and word order variation is often the starting point of information structure accounts (e.g. Erteschik-Shir 2007: 1). However, in the Construction Grammar framework, word order has received surprisingly little attention. Some (mostly preliminary) suggestions have been made (e.g. Fried and Östman 2004: 69–71; Leino and Kuningas 2005; Kuningas and Leino 2006), but one can hardly speak of a tradition, let alone convention, of construction-based word order research.

While the tools of information structure analysis presented above are essential to describing word order variations, and while it hardly makes sense to discuss word order variations without making reference to information structure, one cannot equate the two. For the purposes of constructional analysis, word order phenomena and information structure features (i.e. notions like topic vs. focus, presupposition vs. assertion, activation, etc.) are separate features or properties of constructions. While there are obvious correlations between word order and information structure, it is equally obvious that word order has other functions beside marking information structure (for a brief and informative overview, see e.g. Fried 2003). Furthermore, word order is by no means the only formal way of coding information structure features.

---

[11] Lambrecht (1994: 17) presents the pragmatic markedness of grammatical structures as a relative notion: given two allosentences, one is unmarked if it serves two discourse functions and the other one only serves one of the two.

Rather than thinking of word order and information structure as inherently interdependent, it is often more fruitful to postulate constructions which combine one or both of the two with other linguistic features and phenomena. Michaelis's and Lambrecht's account of the English nominal extraposition construction, shown in figure 2 above, is a case in point: it specifies the word order and the topic and the focus of the sentence, but it also includes quite a lot of other information, e.g. concerning the subject and the predicate of the sentence. The construction also specifies that the sentence-initial pronoun and the sentence-final NP are coreferential, etc.

Also the work of Mettouchi (e.g. 2003, 2008) and Kuningas (2007) on Kabyle (Berber/Afro-Asiatic, spoken in Algeria) sentence types and information structure illustrates the interconnectedness of information structure on the one hand and word order phenomena on the other with such issues as grammatical functions, case marking, sentence meaning, and the like. Consider the following examples, taken from Kuningas (2007: 79–80, glossed in English by J.L.):

(5) a. teggent   tlawin tiɣrifin
         they.make women pancakes
         'The women make pancakes.'

     b. tilawin, teggent   tiɣrifin
         women they.make pancakes
         'The women, they make pancakes.'

     c. d   tiɣrifin i   teggent   tlawin
         AUX pancakes REL they.make women
         'It's pancakes   that the women make.'

The sentence (5a) exemplifies what Kuningas calls the *canonical sentence* ("phrase canonique"), by and large corresponding to the English unmarked SVO sentence. (5b) exemplifies the *left dislocation sentence* ("phrase disloquée à gauche") which, for the purposes of this illustration, corresponds to the English left dislocation construction

discussed above. (5c) exemplifies the *cleft sentence* ("phrase clivée"), which more or less corresponds to the English cleft sentence used here as its translation.

As the English translations suggest, these sentence types differ with regard to their information structure, notably with regard to their choice of topic and focus, and with regard to what is taken to be presupposed and/or active. Formally, (5a) and (5b) differ from each other with regard to word order, but also with regard to case marking: the word for 'women' is in the annexed case ("état d'annexion") in (5a), i.e. *tlawin.* In (5b), it is in the free case ("état libre"), i.e. *tilawin.* The cleft sentence (5c) resembles (5a) with regard to case marking but differs from both (5a) and (5b) with regard to word order and in other respects as well: taken literally, it is a predicative clause which essentially corresponds to its English counterpart.[12]

This rather quick look at a small set of examples is sufficient to show that information structure phenomena typically associate, and grammaticalize together, with sentence types: conventional pairings of morphosyntactic and semantic features. Thus, overall, word order and information structure do relate to each other very closely, but they are also intrinsically intertwined with other aspects and phenomena of grammar.

5.   Conclusion

Information structure is a splendidly wide topic, and it is impossible to give a comprehensive picture of such a range of different, though interrelated, phenomena within the limits of a brief book chapter. The aim of this chapter has been to provide the reader with a selection of references and a rough map for navigating towards those and other related sources. For an in-depth introduction to information structure, especially in the CxG tradition, the interested reader is referred first and foremost to the work of Knud Lambrecht, in particular to his 1994 monograph.

---

[12] The *d* before the word *tiɣrifin* in (5c) is an auxiliary of predication which marks the sentence as a predicative structure—like roughly half of the world's languages, Kabyle does not use a copula here—and the *i* marks the rest as a relative clause.

For a full CxG description of these sentence types, the interested reader is referred to Kuningas (2007). A comparison of the left dislocation construction exemplified in (5b) with similar structures in (spoken) French and Finnish, see Leino and Kuningas (2005).

As for the place of information structure in (construction) grammar, we might conclude that information structure depends on grammar and grammar depends on information structure. It may sometimes be useful to temporarily separate information structure, morphosyntax, and semantics from one another for the sake of simplicity and to focus on one over the others. More often than not, however, giving a really useful and complete description of a construction, or sentence type, or a single utterance, requires taking all three aspects of grammar into account and integrating them into the same constructional description. Here, as in so many other cases, the exceptional strength of CxG is its capability of integrating seemingly different but intrinsically interrelated phenomena into a coherent descriptive apparatus and into credible and comprehensive analyses.

References

Baron-Cohen, Simon and Alan M. Leslie and Uta Frith (1985). 'Does the autistic child have a "theory of mind"?' In *Cognition* 21, 37–46.

Chafe, Wallace (1976). 'Givenness, contrastiveness, definiteness, subjects, and topics', in Li, Charles N. (ed.): *Subject and Topic.* London: Academic Press. 27–55.

——— (1987). 'Cognitive constraints on information flow', in Tomlin, Russell (ed.): *Coherence and Grounding in Discourse.* Typological Studies in Language, vol. XI. Amsterdam: John Benjamins. 21–52.

——— (1994). *Discourse, Consciousness, and Time.* Chicago: University of Chicago Press.

Daneš, František (1964). 'A three level approach to syntax', in Daneš, František (ed.): *Travaux de linguistique de Prague,* vol. 1. University of Alabama Press. 225–240.

Du Bois, John (1980). 'Beyond definiteness: the trace of identity in discourse', in Chafe, Wallace L. (ed.): *The pear stories: Cognitive, cultural and linguistic aspects of narrative production.* Norwood: Ablex. 230–274.

Erteschik-Shir, Nomi (2007). *Information Structure: The Syntax–Discourse Interface.* Oxford: Oxford University Press.

Fillmore, Charles J. and Paul Kay (forthcoming). *Construction Grammar.* Stanford: Center for the Study of Language and Information. [Referred to in Östman and Virtanen 1999.]

Firbas, Jan (1962). 'Notes on the function of the sentence in the act of communication', in *Journal of Studies of the Philosophical Faculty of Brno* Series A, 10, 133–148.

————— (1964). 'On defining the theme in functional sentence analysis', in Daneš, František (ed.): *Travaux de linguistique de Prague,* vol. 1. University of Alabama Press. 267–280.

Fried, Mirjam (2003). 'Word Order', in Östman, Jan-Ola and Jan Blommaert (eds.): *Handbook of Pragmatics.* Amsterdam: John Benjamins.

Fried, Mirjam and Jan-Ola Östman (2004). 'Construction Grammar: A Thumbnail Sketch', in Fried, Mirjam and Jan-Ola Östman (eds.): *Construction Grammar in a Cross-Language Perspective.* Constructional Approaches to Language 2. John Benjamins, Amsterdam. 11–86.

Gundel, Jeanette and Thorstein Fretheim (2003). 'Information Structure", in Östman, Jan-Ola and Jan Blommaert (eds.): *Handbook of Pragmatics.* Amsterdam: John Benjamins.

Halliday, M. A. K. (1967). 'Notes on transitivity and theme in English: part 2', in *Journal of Linguistics* 3, 199–244.

Happé, Francesca (1994). *Autism: an introduction to psychological theory.* London: UCL Press.

Kiparsky, Paul and Carol Kiparsky (1971). 'Fact', in Steinberg, Danny D. and Leon A. Jakobovits (eds.): *Semantics. An Interdisciplinary Reader in Philosophy, Linguistics and Psychology.* Cambridge: Cambridge University Press. 345–369.

Kuningas, Johanna (2007). *Structure informationelle et constructions du kabyle: Etude de trois types de phrase dans le cadre de la grammaire constructionelle.*

Publications, vol. 42. Helsinki: University of Helsinki Department of General Linguistics.

Kuningas, Johanna and Jaakko Leino (2006). 'Word Orders and Construction Grammars', in Suominen, Mickael et al. (eds.): *A Man of Measure. Festschrift in Honour of Fred Karlsson on his 60th Birthday.* A special supplement to SKY Journal of Linguistics, 19:2006. 301–309.

Lambrecht, Knud (1981). *Topic, Antitopic and Verb Agreement in Non-standard French.* [Pragmatics and Beyond II 6.] Amsterdam: John Benjamins.

——— (1984). 'Formulaicity, frame semantics, and pragmatics in German binominal expressions', in *Language* 60:3, 753–796.

——— (1987). 'Sentence focus, information structure, and the thetic–categorial distinction', in *Proceedings of the 13th Annual Meeting of the Berkeley Linguistics Society.* 366–382.

——— (1990). ''What, me worry?' Mad Magazine sentences revisited.' In *Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society.* 215–228.

——— (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents.* Cambridge: Cambridge University Press.

——— (2000). 'Prédication seconde et structure informationnelle. La relative de perception comme construction présentative.' In *Langue Française* 127, 49–66.

——— (2001). 'Dislocation', in Haspelmath, Martin and Ekkehard König and Wulf Oesterreicher and Wolfgang Raible (eds.): *Language Typology and Language Universals: An International Handbook, vol. 2.* Berlin: Walter de Gruyter. 1056–1078.

——— (2004). 'On the interaction of information structure and formal structure in constructions: the case of French right-detached *comme*-N', in Fried, Mirjam and Jan-Ola Östman(Ed.): *Construction Grammar in a Cross-Language Perspective.* [Constructional Approaches to Language 2.] Amsterdam: John Benjamins. 157–199.

Lambrecht, Knud and Laura A. Michaelis (1998). 'Sentence Accent in Information Questions', in Koenig, Jean-Pierre (Ed.): *Discourse and Cognition: Bridging the Gap.* CSLI Publications, Stanford.

Langacker, Ronald W. (1986). 'An Introduction to Cognitive Grammar', in *Cognitive Science* 10, 1–40.

Leino, Jaakko and Johanna Kuningas (2005). 'Structure informationnelle et l'ordre des mots avec les exemples du français, kabyle et du finnois', in Fernandez-Vest, Jocelyne (ed.): *Les langues ouraliennes aujourd'hui – The Uralic languages today.* Bibliothèque de l'Ecole des Hautes Etudes. Paris: Editions Honoré Champion. 587–598.

Levinson, Stephen C. (1983). *Pragmatics.* Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Mettouchi, Amina (2003) 'Contrastive Focalization on clefts in Taqbaylit Berber', in *Proceedings of IP2003 Interfaces Prosodiques, Nantes 27-29 mars 2003.* 143–148.

——— (2008) 'Case-marking, Syntactic domains and Information structure in Kabyle (Berber)', in Frajzyngier, Zygmund and Erin Shay (eds.): *Interaction of Morphology and Syntax: Case Studies in Afroasiatic.* Amsterdam: John Benjamins. 7–40.

Michaelis, Laura – Lambrecht, Knud (1994). 'On Nominal Extraposition: A Constructional Analysis', in *Proceedings of the 20th Annual Meeting of the Berkeley Linguistics Society.* 262–273.

——— (1996). 'Towards a Construction-Based Theory of Language Function: the Case of Nominal Extraposition', in *Language* 72, 215–247.

Mukherjee, Joybrato (2001). *Form and Function of Parasyntactic Presentation Structures: A Corpus-based Study of Talk Units in Spoken English.* Amsterdam: Rodopi.

Östman, Jan-Ola and Tuija Virtanen (1999). 'Theme, comment, and newness as figures in information structuring', in van Hoek, Karen and Andrej A. Kibrik and Leo Noordman (eds.): *Discourse Studies in Cognitive Linguistics.* Current Issues in Linguistic Theory, vol. 176. Amsterdam: John Benjamins. 91–110.

Perner Josef and Uta Frith and Alan M. Leslie and Susan R. Leekam (1989). 'Exploration of the autistic child's theory of mind: Knowledge, belief, and communication', in *Child Development* 60, 688–700.

Prince, Ellen (1981). Toward a taxonomy of given–new information. In Cole, Peter (ed.): *Radical pragmatics.* New York: Academic Press. 223–255.

Tomlin, Russell S. and Linda Forrest and Ming Ming Pu and Myung Hee Kim (1997). 'Discourse Semantics', in Van Dijk, Teun A. (ed.): *Discourse as Structure and Process.* London: Sage. 63–111.