

Formal Approaches to The Interface of Syntax and Information Structure

Part 1: Introduction

Detmar Meurers
Ohio State University
dm@ling.osu.edu

Kordula De Kuthy
Ohio State University
kdk@ling.osu.edu

LOT Winter School 2005, Groningen

General Remarks

- I'd like to know more **about you**; please hand in a piece of paper with:
 - name, email, affiliation
 - research interests and background (syntax, semantics, formal pragmatics, HPSG, CG, LFG, GB, formal/computational linguistics, . . .)
- I'll distribute **handouts** with 4up-copies of the slides.
If you're missing some of the **articles** I mentioned as readings in the course announcements, let me know and I can make an original available for copying.
The **course web page** has links to handouts and some papers:
<http://ling.osu.edu/~dm/05/lot/>
- Don't hesitate to ask **questions**:
 - At LOT, during the lectures, during the breaks, . . . anytime!
 - After LOT, feel free to send emails to continue the discussion.

2/35

Outline of the course

- Formal Approaches to the Interface of Syntax and Information Structure
 - Background on Information Structure
 - Approaches in three frameworks:
 - * Lexical-Functional Grammar (LFG)
 - * Head-Driven Phrase Structure Grammar (HPSG)
 - * Combinatory Categorical Grammar (CCG)
 - Comparison, Open Issues, Outlook
- How to search in corpora for linguistically relevant data

3/35

Background on Information Structure

- Introduction: What is information structure and basic notions
- Historical development of information structure approaches (largely based on von Stechow 1999, ch. 3)
 - The Beginnings of Information Structure
 - The Prague School
 - Halliday and the American structuralism
 - Information Packaging
 - Information Structure in Generative Grammar
- The Semantics of Information Structure
 - Structured Meaning
 - Alternative Semantics
- Intonation and Information Structure
- Word order and Information Structure

4/35

There is more than just syntax and semantics

A simple sentence (1) can be used in many different contexts (2–4), conveying different kinds of information.

- (1) Tim bought a new car.
- (2) a. There is a brand-new Mercedes outside. Did anybody buy a new car?
b. TIM bought a new car.
- (3) a. Tim looks so happy these days. What did he do?
b. Tim bought a new CAR.
- (4) a. What did Tim do after his old car broke down? Did he lease a new car?
b. No, Tim BOUGHT a new car.

What is information structure?

- Very generally speaking, the information structure encodes which part of an utterance is informative in which way, in relation to a particular context.
- A wide range of approaches exists with respect to the question what should be regarded as the primitives of the information structure, with diverse and often confusing terminology.

Two primitives of information structure

Many approaches include one or both of the following distinctions:

- **Givenness:** A distinction between
 - what is new information advancing the discourse (**focus**)
 - what is known, i.e., anchoring the sentence in existing (or presupposed) knowledge or discourse (**background**)
- **Aboutness:** A distinction between
 - what the utterance is about (**topic, theme**)
 - what the speaker has to say about it (**comment, rheme**).

Example: (5) a. What does John drink?

	<i>background</i>	<i>focus</i>
b.	John drinks	BEER.
	<i>topic</i>	<i>comment</i>

The Focus/Background distinction

- A sentence can be structured into two units according to their informativeness, i.e., which part is informative (new) with respect to the discourse, the **focus**; and which part is uninformative (known), the **background**.
- The typical test for the focus unit of a sentence is the constituent question:
 - (6) a. Q: Who did Sue introduce to Bill?
A: Sue introduced [JOHN]_F to Bill.
 - b. Q: Who did Sue introduce to Bill?
A: Sue introduced [the woman with the red SCARF]_F to Bill.
 - c. Q: What happened?
A: [Sue introduced John to BILL]_F
- The focus part of a sentence can be one word, a phrase, or the whole sentence.
- The background part of the sentence can be derived from the focus part, i.e., it is the part of the utterance that is not the focus.

The Focus/Background distinction (cont.)

- Linguistic means of marking such an information structuring are, for example, word order, morphology and prosody.
- English and German are so called intonation languages, i.e., they use pitch accents to highlight informational units of the utterance in a particular way.
- The intonationally highlighted part is associated with the most informative part, i.e., the focus, while the remainder of the sentence contains mainly background knowledge, i.e., information that is already available in the discourse.

The Topic/Comment distinction

- In the topic-comment structure, topic refers to what the utterance is about and comment what the speaker says about it.
- The topical element can be associated with the question: *What about X?*
- In English, topic is marked by a pitch accent, just like focus is, but of a different kind: The focus accent is a typical falling movement while the topic accent is realized as a fall-rise.

(7) Q: Well, what about FRED? What did HE eat?

A: FRED ate the BEANS.
 topic focus

(8) Q: Well, what about the BEANS? Who ate THEM?

A: FRED ate the BEANS.
 focus topic

Semantic effects

A sign in the London underground reads (Halliday 1967):

(9) Dogs must be carried.

This sentence can be read in two different ways:

- (10) a. Dogs must be CARRIED.
 b. DOGS must be carried.

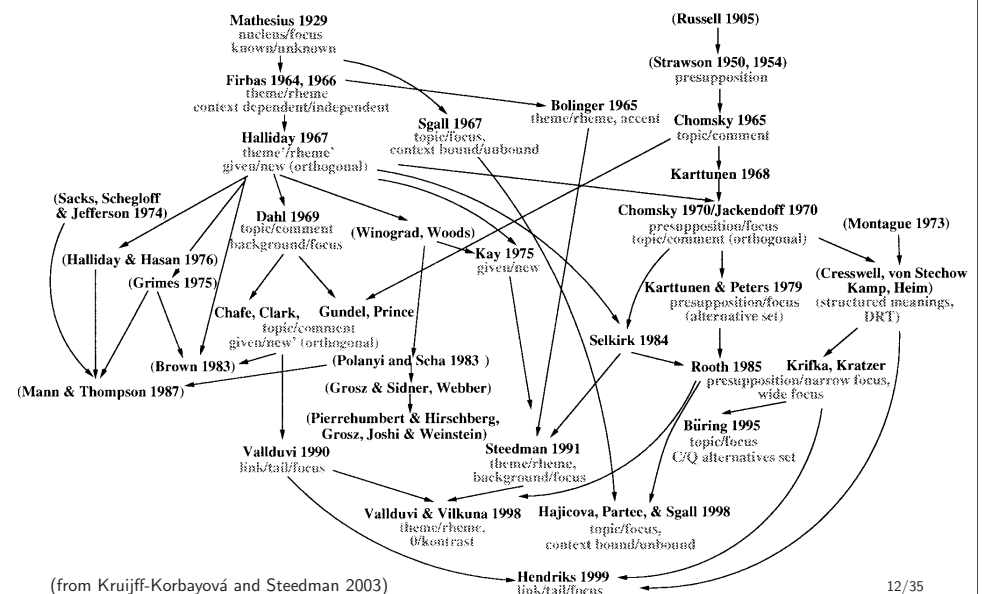
There is a difference in meaning:

- (11) a. If you have a dog, you must carry it.
 b. What you must do is carry a dog.

The second reading is odd here, but it is the preferred one for:

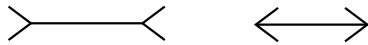
(12) Shoes must be worn.

Historical development of information structure approaches



The beginnings of information structure

- In the course of the 19th century it became obvious that the grammatical description of the sentence does not cover all aspects of sentence meaning. Differences in the presentation of the sentence content were attributed to an underlying psychological structure.
- One important movement in psychology, the so-called **Gestalt theory**, assumed that perception functions as a whole gestalt and not by constructing something out of small units. The gestalt perception includes two different parts: **figure** and **ground**.
- The figure is recognized only against the ground. This is the principle behind many optical illusions, as illustrated below, where one and the same stimulus (the line) is perceived differently depending on the ground.



Gestalt and language

- Related to the Gestalt theory in psychology, the idea of a dichotomy of the sentence organization was developed, which inherited the terms figure and ground.
- The figure represents the prominent or highlighted part, while the ground represents the given or less informative material of the sentence.

The communicative function of language

- At the beginning of the 20th century, the interest in the communicative function of language increased.
- In order to distinguish between the grammatical structure of the sentence (subject–predicate), the psychological structure of concepts or ideas, and the informational structure of a message in a communication, Ammann (1928) introduces a new pair of terms for the latter: **theme** and **rheme**.
- The Prague School integrated the distinction between theme and rheme into the grammatical system, with Mathesius (1929) employing it for describing word order phenomena in Czech and other Slavic languages, where word order typically reflects an ordering from the contextually given to the new.

The Prague School

- The most characteristic feature of the Prague structuralists, in contrast to other structuralist schools, is the functional perspective:
 - Language is understood as a tool for communication and the information structure is important for both the system of language and the process of communication.
- Firbas (1964) argues that information structure is not a dichotomy but rather a whole scale, or hierarchy, or what he calls **communicative dynamism**.
- Daneš (1970) extends the thematic relation of the sentence to one of a text.
- The newer Prague School (cf., e.g., Sgall et al. 1973, 1986) derive the **topic-focus articulation** from a notion of contextual-boundedness and make it part of the grammatical model of a sentence.

Halliday and the American structuralists

- Halliday (1967) introduced the Praguean distinction of theme and rheme into American structuralist linguistics.
- He is the first who uses the term **information structure** and establishes an independent concept of it. He assumes that an utterance is organized into “information units”, which do not correspond to constituent structure.
- Information structure is realized phonologically, i.e., the utterance is divided into different tone groups (intonational phrases). These phrases exhibit an internal structure.
- Analogously, Halliday assumes two structural aspects of information structure:
 - the informational partition of the utterance, the **thematic structure** (theme-rheme), organizes the linear ordering of the informational units.
 - the internal organization of each informational unit, the **givenness**, elements are marked with respect to their discourse anchoring.

Historical development of information structure approaches

17/35

Information Packaging

- Chafe (1976) is interested in the way discourse is structured, and he introduces the term **information packaging**.
- The idea of information packaging was further developed by Vallduví (1990), who assumes an information structure that merges the two most prominent aspects of information structure (focus/background and topic/comment):
 - focus (= focus & comment)
 - link (= background & topic)
 - tail (= background & comment),

(13) a. What about John? What does he drink?

b.

John	drinks	BEER.
<i>background</i>		<i>focus</i>
<i>topic</i>		<i>comment</i>
<i>link</i>	<i>tail</i>	<i>focus</i>

Historical development of information structure approaches

18/35

Information Structure in Generative Grammar

- Chomsky (1971) assumes a **focus/presupposition** distinction. The function of focus is to determine the relation of the utterance to responses, to utterances to which it is a possible response, and to other sentences in the discourse.
 - Focus is defined as the phrase containing the intonation center.
 - Presupposition is described as that part of the sentence that is conveyed independently of the speech act or a negation made in the sentence.
- E.g., both sentences in (14) presuppose that John writes poetry somewhere.

(14) a. Does John write poetry in his STUDY?
b. John doesn't write poetry in his STUDY.

Both can be followed by:

(15) No, John writes poetry in the [GARDEN]_{Focus}.

- On this basis, Jackendoff (1972) developed an approach which is the basis for a number of semantic theories of focus.

Historical development of information structure approaches

19/35

The semantics of focus

- Information structure plays an essential role in several aspects of meaning:
 - Information structure is essential for
 - * the construction and the coherence of a discourse,
 - * the choice of anaphoric elements, and
 - * the identification of the adequate speech act.
 - Information structure is necessary for the interpretation of sentences with focus-sensitive particles (e.g., *only*, *also*, *too*), or adverbs of quantification (e.g., *always*, *sometimes*).
- (16) a. John only introduced [BILL]_F to Sue.
b. John only introduced Bill to [SUE]_F.
c. John only introduced [BILL]_F to [SUE]_F.
d. John only [INTRODUCED]_F Bill to Sue.

The semantics of focus

20/35

The semantics of focus: Starting point

- In the wake of Jackendoff (1972), formal theories of the semantics of focus associate with each sentence a model-theoretic entity which directly reflects its focal structure. This entity is often called the **focus-induced interpretation**.
 - The value of the focus, i.e., the ordinary denotation of the focused expression, is part of the **set of alternatives**, the **p(resuppositional)-set**.
 - The rest of the sentence corresponds to a semantic structure that is called **p-skeleton**. It is formed by substituting the focused expressions with appropriate variables, for example:

(17) a. John introduced [BILL]_F to Sue.
p-skeleton: John introduced x to Sue.

b. John introduced Bill to [SUE]_F.
p-skeleton: John introduced Bill to y.

The semantics of focus: Starting point (cont.)

- The focus induces the presupposition that the set of values which verify the p-skeleton is under discussion. E.g., (18) is the focal presupposition of (17a):

(18) $\{d \in D : \text{John introduced } d \text{ to Sue}\}$ is under discussion.
- The sentence is taken to assert that the denotation of the focus is a member of this set. For example, (19) is the assertion of (17a):

(19) $\text{Bill} \in \{d \in D : \text{John introduced } d \text{ to Sue}\}$
- Based on the insights of Jackendoff (1972), in the 1980's two strands of semantic theories of focus were developed:
 - Structured meaning approach
 - Alternative semantics

The semantics of focus: Structured meaning

- The structured meaning theory of focus was developed by Stechow (1981), Stechow and Cresswell (1983), Jacobs (1983), and Krifka (1992).
- The focus-induced interpretation of a sentence is an ordered sequence, the **structured meaning**, whose members are
 - the property obtained by λ -abstracting on the focus (or foci), and
 - the ordinary semantic interpretation of the focus (or foci).

As an example, consider the structured meaning representation of the examples repeated in (17):

- (17) a. John introduced [BILL]_F to Sue.
 $\langle \lambda x[\text{introduce}(\text{john}', x, \text{sue}')], \text{bill}' \rangle$
- b. John introduced Bill to [SUE]_F.
 $\langle \lambda y[\text{introduce}(\text{john}', \text{bill}', y)], \text{sue}' \rangle$

The semantics of focus: Alternative semantics

- The alternative semantics theory of focus was proposed in Rooth (1985).
- Each sentence receives two distinct model-theoretic interpretations:
 - an ordinary semantic value (written as $\llbracket \]^o$), and
 - a separate focus-induced interpretation called the p-set or **focus-semantic value** (written as $\llbracket \]^f$), which is the set of all propositions obtainable by replacing each focus with an alternative of the same type.

The focus semantic value of (17a), i.e., $\llbracket \text{John introduced [BILL]_F to Sue.} \rrbracket^f$ is shown in (20a). In (20b), it is spelled out assuming that the only individuals in D are John, Bill, Sue, and Mary.

- (20) a. $\{\text{the proposition that John introduced } d \text{ to Sue} : d \in D\}$
- b. $\left\{ \begin{array}{l} \llbracket \text{John introduced John to Sue} \rrbracket^o, \llbracket \text{John introduced Bill to Sue} \rrbracket^o, \\ \llbracket \text{John introduced Sue to Sue} \rrbracket^o, \llbracket \text{John introduced Mary to Sue} \rrbracket^o \end{array} \right\}$

How information structure is manifested in a language

- Languages differ with respect to how the information structure of an utterance is represented.
- Linguistic means of marking information structure are, for example: word order, morphology and prosody.
- English and German are so-called **intonation languages**:
 - Information structuring is signaled by the intonation (contour) of an utterance, including pitch accents.
 - The absence or presence of an accent thus is an indicator of the discourse function of a particular constituent in a sentence.

25/35

Characterizing intonation

- Intonation patterns consist of intonation features or subsystems of various kinds and names.
- The terms **intonational contour (tune)**, **prominence (stress)**, **intonational phrasing**, and **pitch range** are used to refer to these features.
 - The **contour** indicates the movement of pitch. For example, the intonation pattern of an assertion has a distinct contour from that of a question.
 - **Intonational phrasing** divides the sequence of words into intonational units, the intonational (or prosodic) phrases. Phrase boundaries are marked by pauses, boundary tones and duration patterns.
 - **Pitch range** controls the limits in which the contours are realized.

Intonation and Information Structure

26/35

Autosegmental-metrical approaches to intonation

- Pierrehumbert (1980) proposes a description of intonation consisting of:
 - the grammar of phrasal tunes, consisting of L and H tones, which are instantiated as pitch accents, phrase accents and boundary tones.
 - the metrical representation of the text
 - rules for lining up the tune with the text
- Phonological tones
 - Each phrase requires at least one pitch accent (for English: H*,L*, or bitonal as H*+L, H+L*, L*+H, L+H* and H*+H)
 - Each phrase receives a phrase accent (H⁻, L⁻) at the end of the word that is associated with the last pitch accent
 - Each phrase ends with a boundary tone (H%, L%).
- The approach has been modified and further developed, e.g., in Beckman and Pierrehumbert (1986).

Intonation and Information Structure

27/35

Intonational meaning

- There are two main questions with respect to intonational meaning:
 - What are the meaningful units of intonation?
 - What kind of meanings are associated with these units?
- Domains of intonational patterns: tune, phrasing, and pitch accent
- Meaning types that are associated with each of the domains:
 - Tune is often correlated with speech acts.
 - Phrasing is mostly associated with information structure.
 - The pitch accent is linked with the notion of focus.

Intonation and Information Structure

28/35

Discourse function of pitch accents

- In languages like English and German, pitch accents can have different shapes allowing them to signal different functions in the discourse.
- For example, Beckman and Pierrehumbert (1986) assume seven different pitch accents for English: H*, L* and bitonals H*+L, H+L*, L*+H, L+H*, H*+H.
- The H* pitch accents are often assumed to signal focus, i.e., new material. A more precise mapping of each possible pitch accent to its information structure role is subject of debate.

Discourse function of pitch accents (cont.)

- For example for German, Féry (1993) describes the so called hat pattern, consisting of a L*+H accent and a H*+L accent, which signals a topic-comment structure.

(21) Q: Habt ihr gut geschlafen? / Did you sleep well?

A: GESCHLAFEN hat KEINER von uns (aber unterhalten haben wir uns).

L*+H H*+L
slept has no-one of us but talked have we refl.

'None of us slept (instead, we talked).'

Relating intonation and interpretation: focus projection

- The word marked by a pitch accent and the extension of the focus are related to each other by rules of **focus projection**.

(22) Mary bought a book about BATS.

(23) a. Q: What did Mary buy a book about?
A: Mary bought a book about [BATS.]_F

b. Q: What did Mary buy?
A: Mary bought [a book about BATS.]_F

c. Q: What did Mary do?
A: Mary [bought a book about BATS.]_F

d. Q: What happened?
A: [Mary bought a book about BATS.]_F

Focus projection (cont.)

- The focus projection rules determine the focus projection potential of a pitch accent dependent on the syntactic surface structure.

For example, the pitch accents in (24) and (25) cannot project focus to larger constituents, i.e. they are not felicitous answers to the questions in (26).

(24) Q: Who bought a book about bats?
A: [MARY]_F bought a book about bats.

(25) Q: What related to bats did Mary buy?
A: Mary bought a [BOOK]_F about bats.

(26) a. What did Mary buy?
b. What did Mary do?
c. What happened?

Focus projection and constituency

- A rarely noted fact is that the focus resulting from one pitch accent does not always correspond to a constituent, as shown by the German (27) (Höhle 1982).

(27) Q: Was hat das Kind erlebt? / What did the child experience?

A: [Karl]_F hat dem Kind [das BUCH geschenkt]_F .

Karl has the child the book given

'Karl gave the child the book as a present.'

Word order and information structure

- Languages such as Russian, Hungarian, Czech, Catalan, or Turkish mark information structure through word order.
- But even intonational languages like English and German combine intonation and word order to mark certain information structurings.

1. Topicalization in English

(28) Q: Who did you meet in Germany?

A: In Germany, I met a lot of old friends.

(29) Q: You look so happy, what happened?

A: # In Germany, I met a lot of old friends.

Topicalization in English is not possible when an all-focus answer is expected.

Word order and information structure (cont.)

2. Scrambling in German

(30) Where was the spy supposed to meet the courier yesterday?

a. Gestern sollte der Spion den Kurier [in FRANKfurt]_{FOCUS} treffen.
yesterday should the spy the courier in Frankfurt meet

b. # Gestern sollte [in FRANKfurt]_{FOCUS} der Spion den Kurier treffen.
yesterday should in Frankfurt the spy the courier meet

'Yesterday, the spy was supposed to meet the courier in Frankfurt.'

Who was the spy supposed to meet in Frankfurt?

b'. Gestern sollte in Frankfurt der Spion [den KuRIER]_{FOCUS} treffen.
yesterday should in Frankfurt the spy the courier meet

The focused constituent cannot scramble (30 b), even though the word order is in principle possible (30 b').

References

- Ammann, Hermann (1928). *Die menschliche Rede. Sprachphilosophische Untersuchungen. 2. Teil: Der Satz*. Darmstadt: Wissenschaftliche Buchgesellschaft. Reprinted 1962, Lahr im Schwarzwald: Moritz Schauenburg.
- Beckman, Mary and Janet Pierrehumbert (1986). Intonational Structure in Japanese and English. *Phonology Yearbook* pp. 255–309.
- Chafe, Wallace (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*. New York: Academic Press, pp. 27–55.
- Chomsky, Noam (1971). Deep Structure, Surface Structure, and Semantic Interpretation. In D. Steinberg and L. Jakobovits (eds.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*, Cambridge, UK: Cambridge University Press.
- Daneš, František (1970). One Instance of Prague School Methodology: Functional Analysis of Utterance and Text. In Paul L. Garvin (ed.), *Method and Theory in Linguistics*, The Hague: Mouton, pp. 132–146.
- Féry, Caroline (1993). *German Intonational Patterns*. No. 285 in *Linguistische Arbeiten*. Tübingen: Max Niemeyer Verlag.
- Firbas, Jan (1964). On Defining the Theme in Functional Sentence Analysis. *Travaux Linguistique de Prague* 1, 267–280.
- Halliday, Michael (1967). Notes on Transitivity and Theme in English. Part 1 and 2. *Journal of Linguistics* 3, 37–81, 199–244.
- Höhle, Tilman N. (1982). Explikationen für 'normale Betonung' und 'normale Wortstellung'. In Werner Abraham (ed.), *Satzglieder im Deutschen*, Tübingen: Gunter Narr Verlag, pp. 75–153.
- Jackendoff, Ray (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.
- Jacobs, Joachim (1983). *Fokus und Skalen. Zur Syntax und Semantik der Gradpartikel im Deutschen*. Tübingen: Max Niemeyer Verlag.
- Krifička, Manfred (1992). A Compositional Semantics for Multiple Focus Constructions. In Joachim Jacobs (ed.), *Informationsstruktur und Grammatik*, Opladen: Westdeutscher Verlag, pp. 17–54.
- Kruijff-Korbayová, Ivana and Mark Steedman (2003). Discourse and Information Structure. *Journal of Logic, Language and Information (Introduction to the Special Issue)*.
- Mathesius, Vilém (1929). Zur Satzperspektive im modernen Englisch. *Archiv für das Studium der modernen Sprachen und Literaturen* 155, 200–210.
- Pierrehumbert, Janet (1980). The Phonetics and Phonology of English Intonation. Ph.d., MIT.

- Rooth, Mats (1985). Association with Focus. Ph.D. thesis, University of Massachusetts, Amherst, MA.
- Sgall, Petr, Eva Hajičová and Eva Benešová (1973). *Topic, Focus and Generative Semantics*. Kronberg/Taunus: Scriptor.
- Sgall, Petr, Eva Hajičová and Jarmila Panevová (1986). *The meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Prague and Dordrecht: Academia and Reidel.
- Stechow, Arnim von (1981). Topic, Focus, and Local Relevance. In Wolfgang Klein and W. Levelt (eds.), *Crossing the Boundaries in Linguistics*, Dordrecht: Reidel, pp. 95–130.
- Stechow, Arnim von and M. J. Cresswell (1983). De Re Belief Generalized. *Linguistics & Philosophy* 5, 503–535.
- Vallduví, Enric (1990). The Informational Component. Ph.D. thesis, University of Pennsylvania, Philadelphia. Published in 1992 by New York: Garland.
- von Heusinger, Klaus (1999). Intonation and Information Structure. The Representation of Focus in Phonology and Semantics. Habilitationsschrift, Universität Konstanz, Konstanz, Germany.
<http://www.ilg.uni-stuttgart.de/vonHeusinger/publications/abstracts/habil.html>.