

Information- and discourse-structure analysis with questions under discussion

Session 1: Introduction to information structure

Kordula De Kuthy, Eberhard-Karls-Universität Tübingen

Arndt Riester, Universität Bielefeld

September 2022

Summer School *Corpus Annotation and Data Analysis (CANDA)*

Georg-August-Universität Göttingen

Introduction

Information structure

A sentence like (1) can be uttered in different contexts, giving rise to different informational structuring.

(1) I eat meat.

(2) A: What do you eat?

B: [I]_T eat [MEAT]_F.

(3) A: Who eats meat?

B: [I]_F eat [meat]_T.

- **Focus (F)**: Answer to a question. What drives the discourse. In English, marked by a pitch accent (sentence stress).
- **Sentence topic (T)**: Salient entity currently being talked about. Connection to previous discourse.

Information structure in Japanese

In Japanese, a topic is marked syntactically (early position in sentence) and by use of the grammatical particle は 'wa' – or omitted.

(4) (私 は) 肉 を 食べる。

watashi wa niku o taberu

I TOP meat ACC eat

[I]_T eat [meat]_F.

(5) 肉 は 食べない。

niku wa tabe-nai

meat TOP eat-NEG

[I]_{T?} [don't eat]_F [meat]_T. (As for meat, I don't eat it.)

The non-wa-marked parts – object in (4), verb in (5) – are (part of) the focus, since there must be an informative contribution in every sentence.

Asking questions

While (4) and (5) are the standard ways to express 'I eat meat' / 'I don't eat meat', they typically occur in two different situations, characterised by two different **questions**:

Q₄: {What are **you eating?**}

(4) (私 は) 肉 を 食べる。
watashi wa niku o taberu
I TOP meat ACC eat
[I]_T eat [meat]_F.

Q₅: {What is your attitude towards **meat?**}

(5) 肉 は 食べない。
niku wa tabe-nai
meat TOP eat-NEG
[I]_{T?} [don't eat]_F [meat]_T. (As for meat, I don't eat it.)

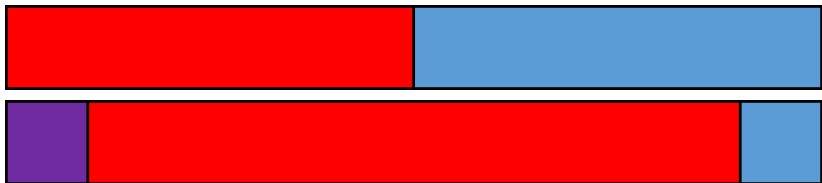
Topical elements occur in the question, focal elements answer it.

Goals of this class

Goals of this class

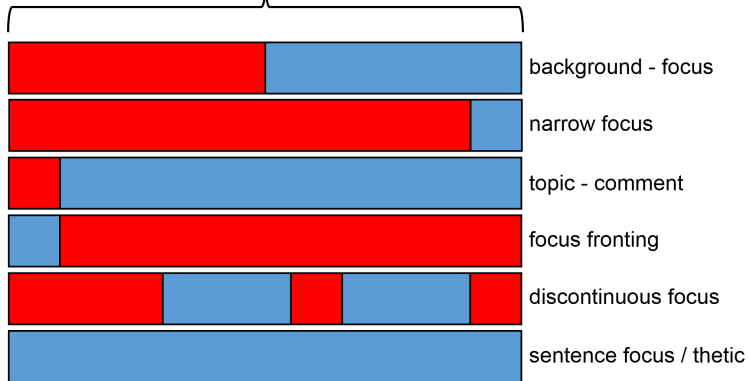
- Introduction to basic information structural concepts
- Present a view on discourse as being structured by questions – so-called **questions under discussion (QUDs)**
- Provide a QUD based method for the practical analysis and annotation of natural discourse (text, dialogue etc.)

Two basic types of discourse units

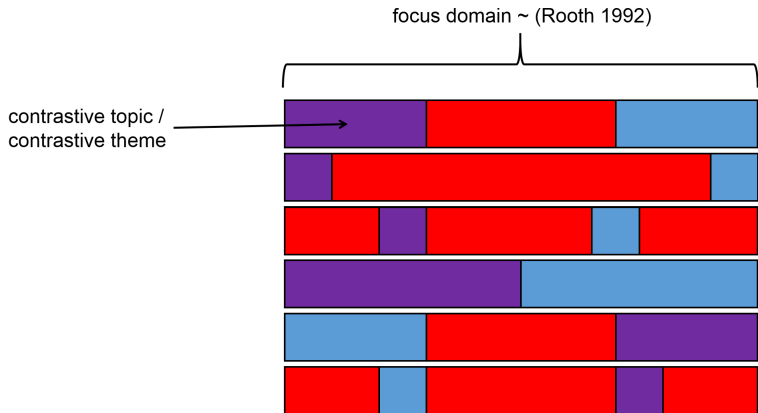


Bipartite structure

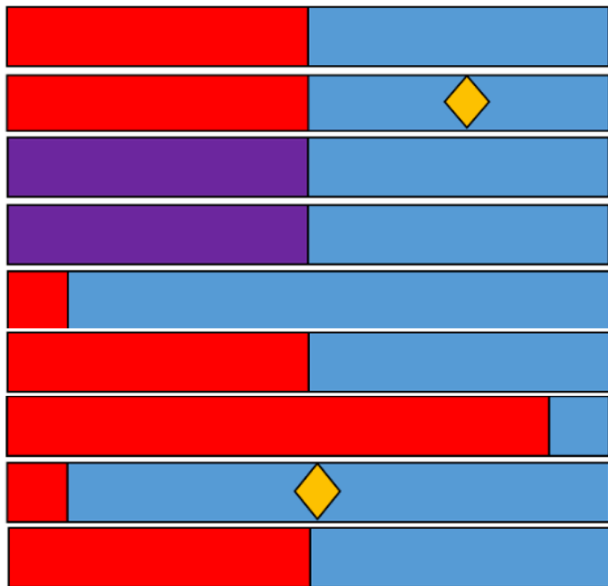
focus domain ~ (Rooth 1992)



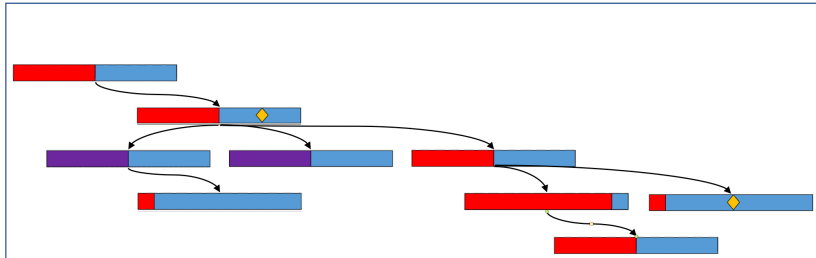
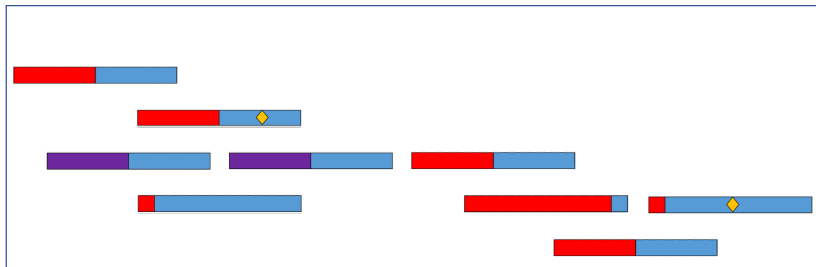
Tripartite structure



A simple picture of discourse



Discourse structure



Focus, information status
(given/new)

History of information structure

- 19th century (von der Gabelentz 1867): *psychological subject/predicate* vs. grammatical subject/predicate
- Prague Linguistic Circle (early 20th century): **Vilém Mathesius** (1882-1945), father of modern information structure theory (Mathesius 1929, 1975), coins the notions
 - *Theme*: what the sentence is about
 - *Rheme*: what is being said about the theme
- Firbas 1964: *Communicative dynamism*: expressions which push communication forward to a varying extent
- Analytical philosophy (Strawson 1964): *Topic* as aboutness and link to known information
- **Michael A.K. Halliday** (1925-2018): Term *information structure* used for the first time (Halliday 1967).
 - Text is structured into *information units* (\neq sentences).
 - Each unit carries an *information focus* (tonal prominence).
 - *Contrastive focus*



Halliday 1967: p. 204 on *information focus* and *given/new*:

*“What is focal is ‘new’ information; not in the sense that it cannot have been previously mentioned, although it is often the case that it has not been, but in the sense that the speaker presents it as **not being recoverable from the preceding discourse.**”*

*“If we use the – admittedly rather inappropriate – term ‘given’ to label what is not ‘new’, we can say that the system of information focus assigns to the information unit **a structure in terms of the two functions ‘given’ and ‘new’.**”*

How to define *given/new*?

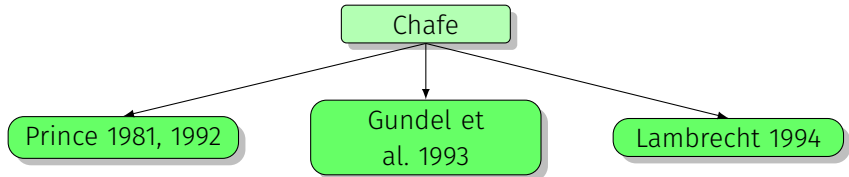
Wallace Chafe
(1927-2019)



Cognitive interpretation (Chafe 1976, 1994):

<i>Given / old:</i>	already <i>active</i> (in consciousness) at this point in conversation
<i>Accessible:</i>	<i>semi-active</i> until mention, i.e. related to some active entity
<i>New:</i>	<i>inactive</i> until mention

- Activation can happen in different ways (e.g. by means of communication, visual presence, or other).
- Inactive entities can be known or unknown.



- *Information status* notion
- Assumed familiarity
- *evoked / inferrable / new*
- *discourse-new vs. hearer-new*

- *Givenness hierarchy*
- Mapping “*cognitive status*” (information status) onto different types of referring expressions

- Subclassification of *inactive* information: *identifiable* (\approx definite) vs. *unidentifiable*

Identifiability

Lambrecht 1994: 77ff. Distinctions among *inactive (new)* referents

Identifiable for both speaker and addressee

- | | |
|------------------------------|------------------------------------|
| · known / familiar | the sun, mom, Joe Biden |
| · description | the tree house in my garden |
| · (activating) demonstrative | those ugly pictures |
| · generic definite | She is studying the whale . |
| · generic indefinite | A bike is useful. |
-
-

Identifiable only for the speaker

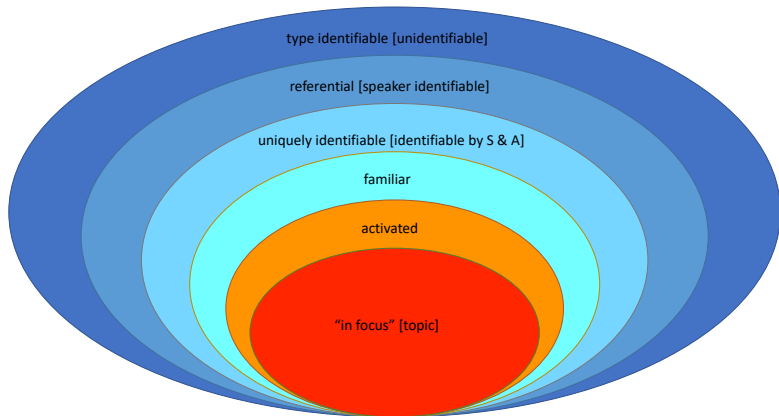
- | | |
|-----------------------------|---|
| · “indefinite <i>this</i> ” | I met this guy on the train. |
| · combination with numeral | I was introduced to one John Smith . |
| · specific indefinite | I got a book for my birthday. |
-
-

Unidentifiable

- | | |
|---------------------------|---|
| · non-specific indefinite | I am looking for a good book . (any) |
|---------------------------|---|

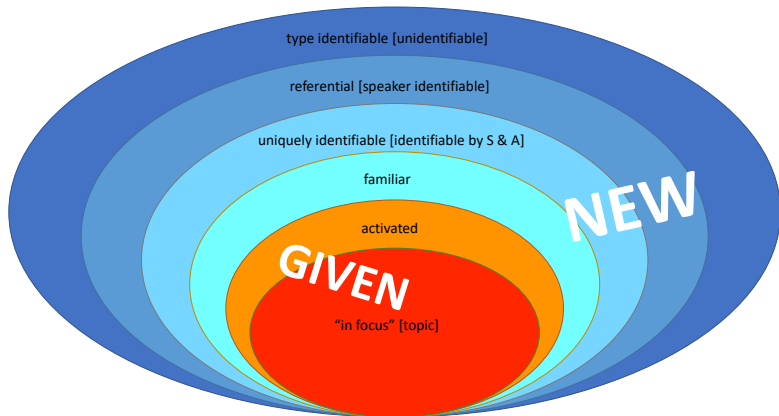
Givenness hierarchy

Properties of “referential” expressions according to Gundel et al. 1993
[+ translations into Lambrecht’s terminology]



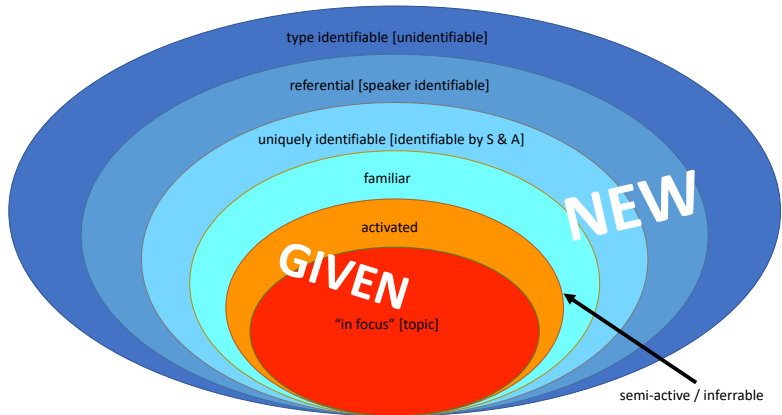
Givenness hierarchy

Properties of “referential” expressions according to Gundel et al. 1993
[+ translations into Lambrecht’s terminology]



Givenness hierarchy


Properties of “referential” expressions according to Gundel et al. 1993
[+ translations into Lambrecht’s terminology]



Defining focus

- In the default case, *focus* represents the **new** information of a sentence.
- *New*: understood as *inactive*, in particular *unmentioned* information

- (6) a. [There was once a princess.]_F
b. She_T [lived on a giant pancake]_F.

- Instead of *focus*, some scholars prefer the *rheme* notion e.g. Daneš 1974, Vallduví and Vilkuna 1998, Steedman 2000.
- The complement of **focus/rheme** is called **background, theme** or **topic**, but these notions are not used uniformly, and with varying meanings. 

- Recall: definiteness \approx identifiability for speaker and addressee (Lambrecht 1994)
- Coincides with the classical presuppositional account of definite descriptions and proper nouns / names since Frege 1892:
 1. The meaningful use of a definite expression presupposes the (at least fictional) **existence** of an entity to which the expression refers.
 2. This entity is **unique**, i.e. there is exactly one object that fulfills the description.

Identifiability, uniqueness, definiteness (cont.)

Definite expressions identifiable because they refer to a unique entity...

domain	referential information status
... in the current discourse	given
... in the context of another, previously mentioned entity (frame)	bridging (associative)
... in the world	unused

(cf. Baumann and Riester 2012, Riester and Baumann 2017)



Coreference, uniqueness in previous discourse

- (7) I met a friend yesterday.
- a. [He] told me a story. (pronoun)
 - b. [The friend] came from Hamburg. (repetition)
 - c. [The funny guy] was in a good mood. (epithet)
 - d. I hadn't seen [Albert] for months. (name)
- (8) The West is suspecting Iran of building nuclear arms. But negotiations with [Tehran] continue.
(metonymy / pars pro toto / synecdoche)
- (9) [Paul [sings under the shower]_k]_i
- a. Mary finds [that]_i weird.
 - b. John does [it]_k, too. (abstract anaphor)

[Barack Obama, 2004 DNC speech]

My father was a foreign student, born and raised in a small village in Kenya. He grew up herding goats, went to school in a tin-roof shack.

His father, my grandfather, was a cook, a domestic servant to the British. But my grandfather had larger dreams for his son. Through hard work and perseverance my father got a scholarship to study in a magical place; America, that shone as a beacon of freedom and opportunity to so many who had come before. While studying here, my father met my mother.

(cf. coreference annotation in OntoNotes, Weischedel et al. 2012, Zeldes 2022)

Note that nominal predicates are not coreferential / given, but new information!

[Barack Obama, 2004 DNC speech]

My father was a foreign student, born and raised in a small village in Kenya. He grew up herding goats, went to school in a tin-roof shack.

His father, my grandfather, was a cook, a domestic servant to the British. But my grandfather had larger dreams for his son. Through hard work and perseverance my father got a scholarship to study in a magical place; America, that shone as a beacon of freedom and opportunity to so many who had come before. While studying here, my father met my mother.

(cf. coreference annotation in OntoNotes, Weischedel et al. 2012, Zeldes 2022)

Activation / givenness via situational presence (symbolic deixis)

On behalf of the great state of Illinois, crossroads of a nation, land of Lincoln, let me express my deepest gratitude for the privilege of addressing this convention . Tonight is a particular honor for me because, let's face it, my presence on this stage is pretty unlikely.

Other (rare) cases of givenness/activation/salience

[Game of Thrones, Season 1, Episode 5, *King Robert Baratheon and Queen Cersei*]

Robert: Ah, so here we sit, seventeen years later, holding it all together. Don't you get tired?

Cersei: Every day.

Robert: How long can hate hold a thing together?

Cersei: Well, seventeen years is quite a long time.

Robert: Yes, it is.

Cersei: Yes, it is. What was **she** like?

Robert: You've never asked about her, not once.



(**she** = Lyanna Stark, Robert's true love, who died long ago)

- Discourse-new, but dependent on (anaphoric to) previous discourse (H. Clark 1977, Asher and Lascarides 1998, Poesio and Vieira 1998, Löbner 1998, Rösiger et al. 2018)
- Identifiability within a frame (Fillmore et al. 1976, Fillmore 2008)
- Relational noun with implicit (given) argument
- It is the implicit argument that provides the link to the previous discourse.

(10) When they entered the house, they kicked [the door] in.

(11) I saw Santa Claus. [His nose] was red.

(12) Our correspondent in Egypt is reporting that [the opposition] is holding a rally against [the constitutional referendum].

- (Globally) unique entity during first mention
- Speaker expects the addressee to know the referent



- (13) [The Pope] stood on [St. Peter's Square].
- (14) [Space probe Voyager 1] passed [planet Jupiter] in [1979].
- (15) [Igor Stravinsky] died in [New York] and was buried in [Venice].
- Note that whether a referent is actually known to the addressee (audience) or not is not a linguistic question!
 - The speaker can be mistaken.
 - Fame is group-specific and fades over time.

- (Globally) unique entity during first mention
- Unfamiliar to (but descriptively identifiable for) the addressee

(16) [The woman Max went out with last night] is an astrophysicist.

(17) [Carl, my neighbour,] never gets up before 11 o'clock.

r-unused

On behalf of the great state of Illinois, crossroads of a nation, land of Lincoln, let me express my deepest gratitude for the privilege of addressing this convention . [...] The day after Pearl Harbor my grandfather signed up for duty, joined Patton's army, marched across Europe .

- Label for indefinite expressions (unidentifiable for addressee)



- (18) [A stormtrooper] threatened me.
- (19) [A military spokesman] confirmed [explosions] and the death of [at least two soldiers].
- (20) He is married to [a computer scientist].



- Recall that, from an information-structural perspective, *new (inactive)* covers **r-new**, **r-unused** and, potentially, **r-bridging (semi-active)**.

- Nesting of referring expressions
- Each referent should receive its own label.

(21) [All operations [at the nearby airport]] were suspended.

- It is easy to overlook personal, reflexive and possessive pronouns.

(22) [They] met in [[his] office].

Topic, focus, questions

Defining topic

- A **topic** is usually defined as **what the sentence is about** (e.g. Strawson 1964, Reinhart 1981, Lambrecht 1994, Jacobs 2001, Krifka 2008).
- By default, an (ordinary) topic is **THE most salient/given referent** in an utterance.
- Note that all of these attributes have been questioned:
 - *Maybe there can there be several topics in one utterance.*
 - *Maybe there are new topics.*
 - *Maybe topics do not have to be referential.*
- Despite these concerns, we can say that **a topic (background) always defines an (implicit) question under discussion (QUD) about that topic:**

- (23) Q: {What about the princess?}
> A: She_T [lived on a giant pancake]_F.



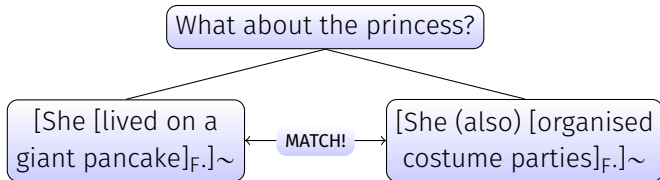
- A focus is then – generally – defined as the answer to the QUD (Roberts 2012 [1996], Velleman and Beaver 2016).
- Preliminary procedure: For each sentence, identify...
Given information → topic / background → QUD → focus
- Warning: This only works in default cases!

Alternative Semantics (Rooth 1992, Beaver and B. Clark 2008, Wagner 2012, Büring 2016)

- The focus of an utterance is interpreted as a **variable** of the corresponding semantic type.
- The set of all possible instantiations of the variable is called the *focus semantic value* $\llbracket \cdot \rrbracket^f$.
 - **Ordinary semantic value:**
 $\llbracket \textit{She [lived on a giant pancake]}_F \rrbracket^o = \textit{live_on_pancake}(pr)$
 - **Focus semantic value:**
 $\llbracket \textit{She [lived on a giant pancake]}_F \rrbracket^f = \{P(pr) \mid P \in \langle e, t \rangle\}$
- Note that the “alternatives” in the focus semantic value are purely abstract!

Focus interpretation operator

- A focus-interpretation operator (\sim) delimits the **information unit** (the *focus domain*, Rooth 1992).
- It defines a **template** for matching the utterance against an actual local alternative in the discourse.
- It also ensures **question-answer congruence**.



Information focus

- The default focus
- Background-focus divide based on the given-new distinction
- Alternatives remain abstract or are irrelevant
- Default syntax, sentence accent on new information

Contrastive focus

- Possible deviation from the default
- Non-standard syntax possible (language-dependent): object fronting, clefting, dislocation, scrambling
- Alternatives are identifiable in discourse or implicated.
- Contrast may imply the exclusion of other alternatives (É. Kiss 1998, Umbach 2004, Repp 2010, Cruschina 2021).
- **Given information can become focal.**

The exact delimitation between information focus and contrastive focus is unclear. (Scale? cf. Umbach (2004) and Repp (2010))

The limits of givenness

Information focus (new):

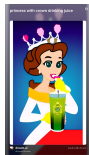
- (24) Q: {What did the princess drink?}
> A: [She drank [raspberry juice]_F.]~

Note that the background ("topic"?) covers more than just the (referential) subject!

Focus on new information, using overt alternatives:

- (25) Q: {What did the princess drink?}
> A': [She drank [raspberry juice]_F.]~
> A'': [She also drank [ginger lemonade]_F.]~

(list)



- (26) Q: {What did the princess drink?}
> A': [She drank [raspberry juice]_F.]~
> A''': No, [She drank [pomegranate smoothie]_F!]~ (correction)

Focus on given information:

- (27) > A₁: One day, the wizard and the princess went shopping.
> Q₂: Who bought sunglasses? (overt question)
>> A₂: [[The wizard]_F (bought sunglasses).]~

- Overt questions can introduce new information.
- Background information can/should be elided.



The limits of givenness (cont.)

> A₁: One day, the wizard and the princess went shopping.

> Q_{3a}: {What did the princess do with the wizard?}

> > A_{3a}: [The princess [advised]_F the wizard.]~

> Q_{3b}: {Who advised whom?}

> > A_{3b}: [[The princess]_F advised [the wizard]_F.]~

Focus on new information
(Default interpretation)

Focus on given information
(Contrastive interpretation)

New information backgrounded

> > A_{3b''}: No, [[The wizard]_F advised [the princess]_F!]~

Annotation of an example text

- Annotate all referring expressions in the text about Göttingen (available from course webpage)
- Use the following labels:

given

bridging / associative







unused







new









predicate







How to reconstruct implicit questions in discourse and identify focus







References







-  Asher, N. and A. Lascarides. 1998. Bridging. *Journal of Semantics* 15. 83–113.
-  Baumann, S. and A. Riester. 2012. Referential and Lexical Givenness: Semantic, prosodic and cognitive aspects. In G. Elordieta and P. Prieto (eds.), *Prosody and meaning* (Interface Explorations 25), 119–161. Berlin: De Gruyter Mouton.
-  Beaver, D. and B. Clark. 2008. *Sense and sensitivity. how focus determines meaning*. Oxford: Wiley-Blackwell.
-  Büring, D. 2016. *Intonation and meaning*. Oxford University Press.
-  Chafe, W. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics and Point of View. In C. N. Li (ed.), *Subject and topic*, 25–55. New York: Academic Press.
-  Chafe, W. 1994. *Discourse, consciousness, and time*. University of Chicago Press.



-  Clark, H. 1977. Bridging. In P. Johnson-Laird and P. Wason (eds.), *Thinking: readings in cognitive science*, 169–174. Cambridge University Press.
-  Cruschina, S. 2021. The greater the contrast, the greater the potential: On the effects of focus in syntax. *Glossa* 6.
<https://doi.org/10.5334/gjgl.1100>.
-  Daneš, F. 1974. Functional Sentence Perspective and the organization of text. In F. Daneš (ed.), *Papers on Functional Sentence Perspective*, 2017–2222. Prague & The Hague: Academia, Mouton.
-  É. Kiss, E. 1998. Identificational Focus versus Information Focus. *Language* 74(2). 245–273.
-  Fillmore, C. et al. 1976. Frame semantics and the nature of language. In *Annals of the new york academy of sciences: conference on the origin and development of language and speech*, vol. 280, 20–32.
-  Fillmore, C. 2008. Frame Semantics. In D. Geeraerts (ed.), *Cognitive linguistics: basic readings*, 373–40. Berlin: Mouton de Gruyter.

-  Firbas, J. 1964. On defining the theme in functional sentence perspective. *Travaux Linguistiques de Prague* 1. 267–280.
-  Frege, G. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100. 25–50.
-  Gundel, J., N. Hedberg and R. Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69(2). 274–307.
-  Halliday, M. 1967. Notes on transitivity and theme in English. Part 2. *Journal of Linguistics* 3. 199–244.
-  Jacobs, J. 2001. The dimension of topic-comment. *Linguistics* 39. 641–681.
-  Krifka, M. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3-4). 243–276.
-  Lambrecht, K. 1994. *Information structure and sentence form*. Cambridge University Press.
-  Löbner, S. 1998. Definite Associative Anaphora. In *Proceedings of the discourse anaphora and resolution colloquium (daarc)*. Lancaster.

-  Mathesius, V. 1929. Functional linguistics. In J. Vachek (ed.), *Praguiana: Some basic and less well-known aspects of the Prague Linguistics School*, 121–142. publ. 1983. Amsterdam: Benjamins.
-  Mathesius, V. 1975. *A functional analysis of present day English on a general linguistic basis*. Translated by Libuše Dušková. The Hague: Mouton.
-  Poesio, M. and R. Vieira. 1998. A Corpus-Based Investigation of Definite Description Use. *Computational Linguistics* 24(2). 183–216.
-  Prince, E. 1981. Toward a taxonomy of given-new information. In P. Cole (ed.), *Radical pragmatics*, 233–255. New York: Academic Press.
-  Prince, E. 1992. The ZPG letter: Subjects, definiteness and information status. In W. Mann and S. Thompson (eds.), *Discourse description: Diverse linguistic analyses of a fund-raising text*, 295–325. Amsterdam: Benjamins.
-  Reinhart, T. 1981. Pragmatics and linguistics: An analysis of sentence topics in pragmatics and philosophy I. *Philosophica* 27(1). 53–94.

-  Repp, S. 2010. Defining ‘Contrast’ as an Information-Structural Notion in Grammar. *Lingua* 120. 1333–1345.
-  Riester, A. and S. Baumann. 2017. *The RefLex Scheme – Annotation Guidelines*. Vol. 14 (SinSpeC. Working Papers of the SFB 732). University of Stuttgart. <http://dx.doi.org/10.18419/opus-9011>.
-  Roberts, C. 2012 [1996]. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6). Earlier version in OSU Working Papers in Linguistics, Vol. 49, 1–69.
-  Rooth, M. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116.
-  Rösiger, I., A. Riester and J. Kuhn. 2018. Bridging resolution: Task definition, corpus resources and rule-based experiments. In *Proceedings of the 27th international conference on computational linguistics*, 3516–3528. Santa Fe, NM, USA.
-  Steedman, M. 2000. Information structure and the syntax-phonology interface. *Linguistic Inquiry* 31(4). 649–689.

-  Strawson, P. 1964. Identifying reference and truth-values. *Theoria* 30. 96–118.
-  Umbach, C. 2004. On the Notion of Contrast in Information Structure and Discourse Structure. *Journal of Semantics* 21(2).
-  Vallduví, E. and M. Vilkuna. 1998. On Rheme and Kontrast. In P. Culicover and L. McNally (eds.), *The Limits of Syntax*, vol. 29 (Syntax & Semantics). New York: Academic Press.
-  Velleman, L. and D. Beaver. 2016. Question-based models of information structure. In C. Féry and S. Ishihara (eds.), *The oxford handbook of information structure*, 86–107. Oxford University Press.
-  von der Gabelentz, G. 1867. Ideen zu einer vergleichenden Syntax. *Zeitschrift für Völkerpsychologie und Sprachwissenschaft* 6. 376–384.
-  Wagner, M. 2012. Focus and givenness: a unified approach. In I. Kučerová and A. Neeleman (eds.), *Contrasts and positions in information structure*, 102–148. Cambridge University Press.

-  Weischedel, R. et al. 2012. *OntoNotes Release 5.0*. Tech. rep. Technical report, Linguistic Data Consortium. Philadelphia.
-  Zeldes, A. 2022. Opinion piece: can we fix the scope for coreference? problems and solutions for benchmarks beyond ontonotes. *Dialogue and Discourse* 13(1). <https://doi.org/10.5210/dad.2022.102>.