

Annotating Information Structure: The Case of *Topic*

Philippa Cook & Felix Bildhauer
Freie Universität Berlin

Abstract

This paper deals with the annotation of Sentence Topics/Aboutness Topics in naturally occurring data. We report on a corpus study in which relatively poor inter-rater agreement was attained for the annotation of topics, although both coders were adhering to the same annotation instructions. Tokens that were particularly difficult to assess are identified, systematized, and discussed in some detail. In sum, the cases that are most likely to lead to non-matching annotations are those that either require a decision between “thetic” or “topic-comment”, or involve an overlap between Focus and Topic. The findings raise a number of issues that may contribute to the discussion in theoretical linguistics, and they also may alert other researchers planning a similar enterprise to some pitfalls they may encounter.

1 Introduction

Research on information structure may serve a twofold purpose: first, information structure constitutes an intriguing area of investigation in its own right, where numerous concepts and their interrelations are still in need of further refinement. Second, insights in this field may lead to promising (re-)analyses of linguistic phenomena on the basis of information structure, that is, using information structural constraints in describing phenomena previously accounted for in terms of syntax (e.g. De Kuthy, 2002; Cook and Payne, 2006; Ambridge and Goldberg, 2008; Cook and Ørsnes, 2010). In this respect, corpora annotated for information structure are particularly valuable, as they put one in a position to test linguistic analyses that are based on notions such as “topic”, “focus” and “givenness”. However, not only are these notions used in different ways across different currents of research, but they also cause considerable problems when applied to naturally occurring data by researchers who otherwise agree largely on the definitions of these concepts and who even adhere to the same set of annotation guidelines.

In the present paper, we will take a closer look at the annotation of Sentence Topics/Aboutness Topics in naturally occurring data. The data we will discuss were extracted from the DeReKo¹ corpus and coded for information structure as part of a study on preferential topic realization in German newspaper texts, that is, a corpus study not initially related to the present work. Section 2 outlines the criteria used by the annotators for identifying Aboutness Topics and relates them to alternative approaches to topic-hood. Section 3 reports the relevant details of the corpus study, including measurements of inter-rater agreement for the annotation of Aboutness Topics. In Section 4, we identify the type of data that turns out to be particularly difficult to assess

¹<http://www.ids-mannheim.de/kl/projekte/korpora/>

and seek to establish exactly which features are involved in these cases and why they give rise to diverging annotations.

2 “Topic” in theory and in annotation guidelines

The notion of Topic we are dealing with in the study being reported on here is that of Aboutness Topic. Since there are considerable differences in the way in which the notion of Topic has been used, and since the actual operationalizability of this notion is the crux of the current contribution, we will lay out here some of the basic assumptions taken by researchers working with the notion of Aboutness Topic. Krifka (2007), in his concise overview of the basic notions of information structure, points out that the use of the terms Topic and Comment reflect what von der Gabelentz (1869) called ‘psychological subject’ and ‘psychological predicate’ respectively, that is “the entity that a speaker identifies about which then information, the comment, is given ” (Krifka, 2007, 40). This approach to Topic was further elaborated by Reinhart (1981), who adopts Stalnaker’s (1978) notion of “context set” (a set of propositions which interlocutors accept to be true; that is, a “Common Ground”). In addition, Reinhart assumes that the Common Ground is structured in a such a way that information is stored in terms of a pairing of an entity and a proposition (or set of propositions) about that entity. New information is added to the Common Ground in the form of structured propositions, where the Sentence Topic designates an entity and the remainder of the sentence contributes the information to be associated with that entity (just like information in a file-card system is stored on a certain file card bearing a heading).² Building on Reinhart’s approach, Krifka (2007) formulates the following definition:

- (1) The topic constituent identifies the entity or set of entities under which the information expressed in the comment constituent should be stored in the CG content.

The notions of Topic/Comment have sometimes been mixed up with the notions of Background/Focus such that, for instance, Focus is believed to be the complement of Topic. The reason for this mixing of dimensions is presumably due to the fact that Topics are in practice prototypically given whereas, in contrast, foci are canonically new. Thus there seems to be a simple dichotomy in which newness and givenness align independently with Focus and Topic respectively. Such a merging of the dimensions is, however, problematic because there are cases which deviate from the canonical alignment in that (i) there are topics which contain a Focus, viz. e. g. (2) below. Such examples involving so-called contrastive topics can, but do not have to, involve an aboutness topic. Rather, the unifying feature of so-called contrastive topics is their function in discourse, where they are assumed to indicate a discourse strategy (Roberts, 1996; Büring, 2003; Krifka, 2007). The next problematic case is (ii) that there do appear to be new (i. e. non-given) topics as in (3) where an entity is introduced as new into the discourse

²The file-card metaphor has since been used by a number of authors. For a critical evaluation, see e. g. Hendriks and Dekker (1996).

(*a good friend of mine*) but still serves as aboutness topic (although see the discussion in Section 4.2 below since this possibility is not wholly uncontroversial). Thirdly, there is the possibility (iii) that the Comment is not always identical to the Focus, i. e. the Focus could be just a sub-part of the Comment as shown in (4) below. Finally, there is also the possibility of non-new foci as in (5) (viz. the discussion of second-occurrence foci; Partee, 1999).

- (2) a. What do siblings do?
 b. [My [SISTER]_{FOC}]_{TOP} [studies MEDicine]_{FOC}
- (3) [A good friend of mine]_{TOP} [married Britney Spears last year]_{COMMENT}
- (4) a. When did [Aristotle Onassis]_{TOP} marry Jacqueline Kennedy?
 b. [He]_{TOP} [married her [in 1968]_{FOC}]_{COMMENT} (Krifka, 2007, 42–44)
- (5) a. Everyone already knew that Mary only eats [vegetables]_{FOC}
 b. If even [Paul]_{FOC} knew that Mary only eats [vegetables]_{SOF}, then he should have suggested a different restaurant.
 (Partee, 1999, 216)

Thus, the possibility of such non-canonical alignments (e. g. non-given topics, non-new foci) must be accommodated in a model of information structure. We have chosen to follow the multi-partitioning approach espoused by Krifka which assumes both a Topic/Comment and an orthogonal Focus/Background partition in order to be able to do justice to the non-canonical as well as canonical pairings.

The characterization of Aboutness Topic that we adopt is also distinct from Valluví's (1992) "Link", which is defined positionally as the sentence-initial topic. Further, since the Focus-Background partition is independent of the Topic-Comment partition, an Aboutness Topic can in principle be identical to a focus of an utterance (though it is unclear whether or not cases other than those in (2) exist; we will return to this point below). Generally speaking, under this approach, a sentence has only one Aboutness Topic. Sentences which lack a Topic – or perhaps more precisely, a Topic-Comment articulation – are classed as thetic (cf. Krifka, 2007, 43). We will have more to say about thetic utterances in general and about the distinction between thetic vs. topic-comment utterances in particular in Section 4 below.

The guidelines for the annotation of information structure (Götze et al., 2007), which were produced by the collaborative research cluster (SFB) 632, and which closely mirror the proposals of Krifka (2007), provide instructions for the annotation of Information Status (or 'givenness'), Topic, and Focus. Under the notion of Topic, both Aboutness Topic and Frame-setting Topic are identified. It is the former that concerns us here (see Krifka, 2007, for a more detailed discussion of frame-setting). Götze et al. (2007, 165) offer the following tests for determining the Aboutness Topic of an utterance:

- (6) An NP X is the Aboutness Topic of a sentence S containing X if
 a. S would be a natural continuation to the announcement
Let me tell you something about X

- b. S would be a good answer to the question
What about X?
- c. S could be naturally transformed into the sentence
Concerning X, S'
where S' differs from S only insofar as X has been replaced by a suitable pronoun.

Applying these diagnostics to naturally-occurring data is not without problems, as will become clear in the following sections.

3 A corpus study

The initial purpose of the corpus study was to test a hypothesis about which verbal dependents (argument or adjunct) are most frequently realized as an Aboutness Topic (AT). On the basis of a prior study, occurrences of four verbs were sampled from the DeReKo corpus. The data was filtered such that only one argument-frame per verb was considered, all occurrences instantiating a different argument frame were discarded (see Table 1; a subject-XP is taken for granted in each case and therefore not listed explicitly). After also discarding occurrences in questions, relative clauses, conditionals, titles and in the first sentence of quotations, between 135 and 167 sentence tokens per verb were included in the study.

Argument frame	Verb	Example
PP _{mit} XP _{LOC}	geraten 'to get (caught in)'	Er gerät [mit seiner Hose] [in die Kette]. 'He got his trousers caught in the chain.'
XP PP _{auf}	reagieren 'to react'	Sie reagiert [überrascht] [auf den Vorschlag]. 'She reacted surprisedly to this suggestion.'
PP _{von}	profitieren 'to profit'	Sie profitieren [von den Steuersenkungen]. 'They profit from the tax reductions.'
	herrschen 'to reign'	Dort herrscht Ruhe. 'There reigns peace.'

Table 1: Verbs and argument frames

Two independent coders (the authors of this paper) annotated a total of 587 sentence tokens, using the annotation schema proposed by the SFB 632 (Götze et al., 2007). The annotation task consisted in selecting the AT from among the NPs (and deictic expressions such as *hier* 'here', *dann* 'then' etc.) contained in a sentence, or alternatively stating that a sentence has no AT. Cohen's κ was used as a measure of inter-rater agreement, which was calculated separately for the annotation of each verb, as shown in Table 2.³

³Cohen's kappa is the proportion of agreement that remains after chance agreement has been factored out (cf. Cohen, 1960). Inter-rater agreement calculated for whether or not a sentence was judged to contain an aboutness topic at all is also highly variable across the four verbs: *profitieren*: $\kappa=.01$, *herrschen*: $\kappa=.51$, *geraten*: $\kappa=.33$, *reagieren*: $\kappa=.25$; the most dramatic change is observed in *profitieren*, where kappa indicates concordance is (slightly) below chance level. This fact strongly suggests that the annotation guidelines can be interpreted in substantially different ways.

Verb	N	Coinciding Annotations	κ
<i>profitieren</i>	135	109	.57
<i>herrschen</i>	138	102	.55
<i>geraten</i>	147	99	.33
<i>reagieren</i>	167	106	.22

Table 2: Inter-rater agreement

Inter-rater agreement is highly variable across the four verbs, but it never exceeds $\kappa = .57$, which in our view is much less than could be expected in a case where both annotators base their judgements on the same guidelines. On inspecting more closely the tokens on which the annotators did not agree, we could identify data that proves particularly difficult to assess. Most of the controversial cases can be grouped into one of the following categories:

- Problems in deciding whether the sentence has an Aboutness Topic at all, including cases where the status of potential topic expressions is unclear because the interaction between topic and focus (especially their overlapping) is not covered exhaustively in the guidelines (nor in the literature).
- The annotators’ different interpretation of “Aboutness”; most commonly, deciding “what the sentence is about” when there is more than one expression that could plausibly serve as the Aboutness Topic: in many cases, the diagnostics sketched in (6) do not yield an answer, or their application is not straightforward.

In what follows, Section 4.1 will briefly illustrate a number of cases where the annotators did in fact agree, and Section 4.2 will address examples from the two problematic categories listed above.

4 Discussion

4.1 Agreement

Examples (7)–(8) are typical of the cases in which the annotators agreed on the AT of the sentence. (In addition to the critical data (b), we also provide some of the immediately preceding context in (a).) In both examples, a non-subject was chosen as the AT. This is probably in part due to the fact that the subject, being a non-specific indefinite, is not suitable as an AT (see Endriss, 2009; Götze et al., 2007). In addition, in terms of givenness, the referent of the non-subject is either “active” (as in (7)), or “accessible” (as in (8)), which are prototypical properties of Aboutness Topics (see Section 2) above.

- (7) a. Ein besonderer Fall ist der sogenannte „Promilleweg_i“, der von Rothenbach Richtung Brandscheid führt.
‘The so-called “promille-road”, leading from Rothenbach to Brandscheid, is a special case.’
- b. *Auf [dem idyllisch gelegenen Wirtschaftsweg_i]_{TOP} herrscht nämlich emsiger Autobetrieb.*
on the picturesquely situated farm road reigns actually active through-traffic
‘The picturesque farm road is actually busy with through-traffic.’
- (8) a. „Es ist schön und lustig, aber die Produktion eines solchen Spiels ist teuer, lohnen sich denn überhaupt die Kosten?“
‘“It’s beautiful and funny, but producing a game like this is expensive, is the cost really worth it?”’
- b. *Auf [diese Frage]_{TOP} würde wohl mancher Nicht-Betriebswirt mit „Typisch BWLer“ reagieren.*
on this question would probably many non-economists with typically economist react
‘Many non-economists would probably react to this question by saying “this is typical of economists”.’

Example (9) illustrates a class of cases where annotators agreed that there is no Aboutness Topic. (9b) is the first sentence of a newspaper article, with no prior context related to it except for the heading, given in (9a). However, cases similar to this one also gave rise to non-matching annotations in our study, as example (14) in the next section shows.

- (9) a. Gegen Leitschiene
‘Against the guardrail’
- b. *Mit ihrem Pkw geriet auf der A 14 in Höhe Ortsgebiet Koblach eine Frau (18) aus Mellau ins Schleudern.*
with her car got on the A 14 in height municipal.area Koblach a woman (18) from Mellau into.the skid
‘A woman (18 yrs.) from Mellau got into a skid on the A 14 near the municipal area of Koblach.’

4.2 Disagreement

The examples presented in this section are representative of the numerous cases that caused difficulties. Example (10b) is representative of a large number of cases that involve two expressions, each of which could justifiably be analysed as the AT of the sentence.

- (10) a. *Dazugelernt habe ich besonders im Bereich der Öffentlichkeitsarbeit. Ich merkte, welche Handlung welche Reaktion auslöst und wie man gewisse Ereignisse richtig kommuniziert.*
‘I learned more in the area of public relations work in particular. I noticed what sort of reaction was caused by which actions and how to communicate certain events correctly.’
- b. *Von [dieser Erfahrung]_{TOP}? kann [ich]_{TOP}? am neuen Ort selbstverständlich profitieren*
from this experience can I at.the new place evidently profit
‘I will clearly be able to profit from this experience at the new place.’

In many cases, one of these candidates is a non-subject that is realized in initial position. However, we do not adopt Vallduví's (1992) approach of identifying the aboutness topic positionally, as it is well known that the latter can occupy positions other than the initial position in German. The difficulty lies in deciding whether the prominent position of the PP should have priority over the fact that (i) the subject is commonly considered the default topic of a sentence and (ii) the topic of preceding sentences (in (10a), arguably the subject) is likely to be the topic of the current sentence as well ("topic chain"; see Givón, 1983).

Example (11b) is similar to (10b) in that it, too, contains two candidate expressions, but it also differs from (10) because one of these expressions (namely the subject NP) should probably bear a focus accent.

- (11) a. *Diese Busspur ermöglicht die neue Buslinie, die ab 1. Juni eingerichtet wird: (...) Damit erhalten zum Beispiel die Bretzenheimer einen flotteren Anschluß nach Hechtsheim (...) Auch in die Altstadt geht's schneller.*
 'This bus lane made possible the new bus route, which will operate as of June 1st: (...) The residents of Bretzenheimer will thus have a better connection to Hechtsheim (...) It will be even quicker to get into the old town-centre too.'
- b. *Außerdem profitiert [der ORN-Bus aus Nieder-Olm]_{TOP?} von [der Spur]_{TOP?}*
 furthermore profits the ORN-Bus from Nieder-Olm from the lane
 'The ORN-Bus from Nieder-Olm will also benefit from the lane.'

Note that the two possible choices of AT in (11) correspond to different discourse strategies: analysing *Spur* as the AT yields topic continuity (cf. Givón, 1983) as *Spur* is arguably the topic of (many of) the preceding sentences. On the other hand, choosing the subject-NP as the AT entails a topic switch.⁴

Turning now to examples (13b) and (14b), the annotators disagreed here on whether they were dealing with a topic-comment structure or rather with a topicless/thetic sentence. At the heart of the disagreement about these examples lies the question of precisely how the two orthogonal IS-partitions assumed here (Topic-Comment vs. Focus-Background) interact with one another, and in particular, how topic and focus may overlap. Various authors (e.g. Krifka, 1992; Steedman, 2000) suggest that both the topic (theme) and the comment (rheme) section of an utterance have their own focus-background structure. To our knowledge, the only cases discussed in which topic and focus overlap are cases of so-called contrastive topic; that is, they involve a semantic focus (marked by a rise) within the initial phrase that induces alternatives in addition to a focus later in the clause which also induces alternatives. The overall function is to indicate a discourse strategy whereby only a question that is subordinate to the (possibly implicit) question under discussion is answered. Independently of such discourse configurations, the question of the possible overlap of Topic and Focus has been less explicitly spelt out. For one annotator, there is no intrinsic problem with a complete

⁴In the terminology of the Prague School (Daneš, 1974), these strategies correspond to a 'thematic progression' with a continuous theme and a thematic progression with derived themes, respectively. The latter describes a configuration where there is one 'hypertheme' (i.e., a discourse topic; the bus lane, in our example), on which individual sentences elaborate. Each one of these sentence presents a theme of its own that is 'derived' in some way from the hypertheme.

overlap of (new-information) focus and AT as sketched in (12), but the other annotator tends to rule this out:

- (12) Q. Who ate the apple?
 A. Kim ate the apple.
 []_{FOC} []_{BACKGROUND}
 []_{TOP} []_{COMMENT}

If one disallows a total overlap of Topic and Focus as in (12), then the question is how the utterance should instead be analysed. One possibility which we will discuss below is that examples such as (13b) and probably also (14b) are topicless sentences. This, however, raises questions about the possible complexity ofthetic utterances.

- (13) a. *In Wil wird das seit Anfang Oktober gültige Rauchverbot nicht überall umgesetzt, und in gewissen Lokalen wird noch immer geraucht. Häufig wird der Gast darauf aufmerksam gemacht, dass es in seiner Verantwortung liegt, zu rauchen.*
 ‘The smoking ban that has been in place since the beginning of October is not put into practice everywhere in Wil and people still smoke in certain bars. Frequently the customer is told that they’re smoking at their own risk.’
- b. *Eine andere Stimmung herrscht im [Kirchberger Restaurant
 a different atmosphere reigns in.the Kirchbergian Restaurant
 Eintracht]_{TOP?}, wo das Rauchverbot strikt eingehalten wird.
 Eintracht where the smoking.ban strictly kept is*
 ‘It’s a different situation at Kirchberg’s Eintracht Restaurant, where the smoking ban is strictly adhered to.’
- (14) a. *Großes Bedauern über Becks Rücktritt*
 ‘Deep regret over Beck’s Resignation’
- b. *Mit großem Bedauern und totaler Überraschung reagierte gestern [die
 with big regret and total surprise reacted yesterday the
 Ludwigshafener SPD-Prominenz]_{TOP?} auf den Rücktritt des
 Ludwigshafen SPD-dignitaries on the resignation of.the
 Bundesvorsitzenden Kurt Beck.
 federal party leader Kurt Beck*
 ‘The SPD-dignitaries in Ludwigshafen reacted with deep regret and utter shock at the resignation of the party leader Kurt Beck.’

Both annotators agree that the example in (13) can be analyzed as introducing a new referent in the main clause, about which the relative clause makes a further predication. The actual information structure within the main clause itself is, however, not so evident. One annotator selected *Kirchberger Restaurant Eintracht* as the AT of the main clause, irrespective of the fact that the same phrase appears to coincide with the final focus of the main clause. The other annotator elected that there was no AT in the main clause (i. e. the introduced referent does not function as Topic until later, in the relative clause). Note that the only other potential Topic candidate, the subject NP, as a non-specific indefinite cannot normally function as an Aboutness Topic.⁵ Under the latter

⁵It is worth noticing here that (13b) might be a case of i-topicalisation (Jacobs, 1997): both *eine andere Stimmung* and *Kirchberger Restaurant Eintracht* are contrasted against elements that have been previously mentioned or can be inferred from the preceding text. However, identifying (13b) as i-topicalisation does not help in deciding whether or not the sentence has an Aboutness Topic, for it has been shown that i-topics behave differently and crucially do not necessarily involve Aboutness. See in particular Jacobs (2001); Büring (2003); Krifka (2007).

view, the main clause does not constitute a Topic-Comment utterance at all. Lacking a Topic-Comment partition is one of the defining features ofthetic utterances (e.g. Lambrecht, 1994; Krifka, 2007), but classifying (13) asthetic is not without difficulties either. It is customarily said ofthetic utterances that the focus spreads across the whole utterance (e.g. Lambrecht, 1994; Rosengren, 1997), and thatthetic sentences in German bear a single accent on the subject (thus, the subject phonologically integrates with the predicate) (e.g. Krifka, 1984; Sasse, 1987). However, our intuition is that (13b) requires two prosodic peaks. Furthermore, a description of this utterance as event-reporting, a further characteristic ofthetics, (cf. Götze et al., 2007, 163) does not seem quite correct either since, as mentioned above, the function of sentences like (13) is to introduce or present a new entity which may then later function as Aboutness Topic in the next discourse chunk.

As for example (14), a similar situation holds. One annotator chose the subject NP as topic and the other elected that the sentence had no AT. However, this example differs from (13) in that there is no contrastive element in initial position. Further, while it was clear in (13) that the main accent falls on the PP, here it could be either on the subject NP or on the final PP. If one assumes it to fall on the subject NP, and if one assumes this to be the AT (as one annotator did), then a similar configuration to that in (13) holds. For the other annotator, who opted for a topicless analysis, the fact that the subject-NP follows the adverb *gestern* guided the decision that it is not an AT, as (14b) does not seem to be a felicitous answer to a question like “What about the Ludwigshafen SPD-dignitaries?”, at least for that annotator. The sentence is discourse-initial, preceded only by a headline, and unless the sentence final NP is to be analysed as AT (an option neither annotator took), the only remaining possibility is to classify it as lacking an AT. However, as was the case with example (13b), in its natural context, sentence (14b) requires more than a single prosodic peak, thus it does not conform to the description usually given ofthetic sentences.

Thus, analyzing examples like (13b) and (14b) asthetic gives rise to difficulties unless one is willing to adopt a definition oftheticity which allows for a type ofthetic utterance that introduces or presents an entity (rather than a situation or event). Such a definition fits in with the approach totheticity found in Lambrecht (1994, 2000) who terms this type ofthetic ‘presentational’ (vs. ‘event-reporting’) and Sasse’s (1987) type referred to as ‘entity-central’ (vs. ‘event central’). Given this bifurcation of the notion oftheticity, and bearing in mind the two orthogonal dimensions of IS along which sentences are analysed in the model we are assuming, one may classifythetics in general as ‘all-comment’ but not necessarily as ‘all-focus’. The difference between entity-centralthetics and event-centralthetics can then be captured by recourse to their differing focus structures. Only event-centralthetics involve focus spreading across the whole utterance whilst with entity-centralthetics it is merely the phrase that denotes

the introduced referent that is focused (illustrated in Figure 1). A distinction between event-reporting and entity-presenting thetics was not part of the annotation guidelines at the time of the study and has now been proposed as a modification.

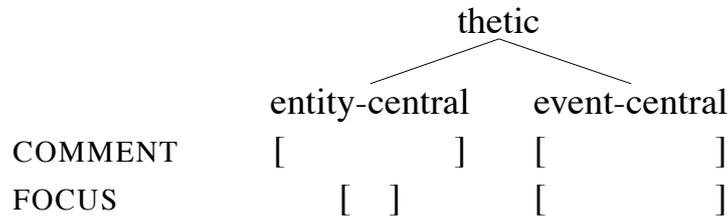


Figure 1: Analysis of different types of thetics in terms of FOCUS and COMMENT

Summing up, then, the two options for example (14) can be sketched thus (assuming that the main stress is on the subject-NP):

- (15) a. [Mit großem Bedauern und totaler Überraschung reagierte gestern]_{COMMENT} [[die Ludwigshafener SPD-Prominenz]_{FOC}]_{TOP} [auf den Rücktritt des Bundesvorsitzenden Kurt Beck]_{COMMENT}
- b. [Mit großem Bedauern und totaler Überraschung reagierte gestern [die Ludwigshafener SPD-Prominenz]_{FOC} auf den Rücktritt des Bundesvorsitzenden Kurt Beck]_{COMMENT}

Data of the type exemplified in (13) and (14) came up frequently and the problem is thus clearly one that should be clarified in other such annotation tasks in the future. Moreover, these data show that it is necessary for annotators to state (in rough terms) the accent pattern they assumed when annotating a sentence token, as different accentuations are sometimes possible and may be indicative of different information structural partitionings.

5 Conclusion

In the present contribution, we reported on difficulties that arose from an annotation task in which we sought to operationalize the notion of Aboutness Topic. As a starting point, the annotators took the guidelines produced by a team of researchers affiliated to a collaborative research centre focusing on information structure (Götze et al., 2007). These guidelines (which, incidentally, are currently undergoing a revision phase) are undoubtedly a valuable resource and a good starting point in bringing terminological clarity to a domain of study (information structure) which is notorious for involving many conflicting definitions on the one hand but also uses of the same terminology in different senses on the other (see, e. g., Kruijff-Korbayová and Steedman, 2003). Nevertheless, once the domain of study shifts to naturally-occurring data, the concept of Aboutness Topic presents various difficulties, as thematised here.

Summing up, we hope to have alerted other researchers planning a similar enterprise to some pitfalls they may encounter and hope we can contribute to the discussion concerning issues which also have a resonance for theoretical linguistics such as (i) the

potential overlap of Aboutness Topic and focus and (ii) the correct delineation of thetic utterances and the role that presentation may play there. We remain optimistic that a careful discussion of many of the areas of contention that arose whilst conducting this study will lead to a fine-tuning of the notion of Aboutness Topic which renders it usable in future studies.

Acknowledgements

This work was funded by Deutsche Forschungsgemeinschaft within Sonderforschungsbereich 632 “Informationsstruktur” (Project A6 “Theorie und Implementation einer Analyse der Informationsstruktur im Deutschen unter besonderer Berücksichtigung der linken Satzperipherie”).

References

- Ben Ambridge and Adele E. Goldberg. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics*, 19(3):357–389, 2008.
- Daniel Büring. On D-Trees, Beans, and B-Accents. *Linguistics & Philosophy*, 26(5):511–545, 2003.
- Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.
- Philippa Cook and Bjarne Ørsnes. Coherence with adjectives in German. In Stefan Müller, editor, *The Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar*, Stanford, 2010. CSLI Publications.
- Philippa Cook and John Payne. Information structure and scope in German. In Miriam Butt and Tracy Holloway King, editors, *Proceedings of LFG06, University of Constance*, pages 124–144, Stanford, 2006. CSLI.
- František Daneš. Functional sentence perspective and the organization of the text. In *Papers on Functional Sentence Perspective*, pages 106–208. Mouton, The Hague and Paris, 1974.
- Kordula De Kuthy. *Discontinuous NPs in German*. CSLI Publications, Stanford, 2002.
- Cornelia Endriss. *Quantificational Topics. A Scopal Treatment of Exceptional Wide Scope Phenomena*. Springer, Dordrecht, 2009.
- Talmy Givón. Introduction. In Talmy Givón, editor, *Topic continuity in discourse. A quantitative cross-language study*, pages 1–41. Benjamins, Amsterdam, 1983.
- Michael Götze, Thomas Weskott, Cornelia Endriss, Ines Fiedler, Stefan Hinterwimmer, Svetlana Petrova, Anne Schwarz, Stavros Skopeteas, and Ruben Stoel. Information structure. In Stefanie Dipper, Michael Götze, and Stavros Skopeteas, editors, *Interdisciplinary Studies on Information Structure*, volume 7 of *Working Papers of the SFB 632*, pages 147–187. Universitätsverlag, Potsdam, 2007.
- Herman Hendriks and Paul Dekker. Links without locations. Information packaging and non-monotone anaphora. In Paul Dekker and Martin Stokhof, editors, *Proceedings of the Tenth Amsterdam Colloquium*, pages 339–358, University of Amsterdam, 1996. ILLC-Department of Philosophy.
- Joachim Jacobs. I-Topikalisierung. *Linguistische Berichte*, 168:91–134, 1997.
- Joachim Jacobs. The dimensions of topic-comment. *Linguistics*, 39:641–681, 2001.
- Manfred Krifka. Fokus, Topik, syntaktische Struktur und semantische Interpretation. Unpublished manuscript, 1984. URL <http://amor.rz.hu-berlin.de/~h2816i3x/Publications/Krifka%201984%20Fokus.PDF>.
- Manfred Krifka. A compositional semantics for multiple focus constructions. In Joachim Jacobs, editor, *Informationsstruktur und Grammatik*, pages 17–54. Westdeutscher Verlag, Opladen, 1992.

- Manfred Krifka. Basic notions of information structure. In Caroline Féry, Gisbert Fanselow, and Manfred Krifka, editors, *Interdisciplinary Studies on Information Structure*, number 6 in Working Papers of the SFB 632, pages 13–56. Universitätsverlag, Potsdam, 2007.
- Ivana Kruijff-Korbayová and Mark Steedman. Discourse on information structure. *Journal of Logic, Language and Information: Special Issue on Discourse and Information Structure*, 12(3):249–259, 2003.
- Knud Lambrecht. *Information Structure and Sentence Form. Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press, Cambridge, 1994.
- Knud Lambrecht. When subjects behave like objects. an analysis of the merging of S and O in sentence-focus constructions across languages. *Studies in Language*, 24:611–682, 2000.
- Barbara H. Partee. Focus, quantification, and semantics-pragmatics issues. In Peter Bosch and Rob van der Sandt, editors, *Focus. Linguistic, cognitive, and computational perspectives*, pages 213–231. Cambridge University Press, Cambridge, 1999.
- Tanya Reinhart. Pragmatics and linguistics. An analysis of sentence topics. *Philosophica*, 27:53–94, 1981.
- Craige Roberts. Information structure in discourse. towards an integrated formal theory of pragmatics. In Jae Hak Yoon and Andreas Kathol, editors, *Papers in Semantics*, volume 49 of *OSU Working Papers in Linguistics*, pages 91–136. The Ohio State University Department of Linguistics, Ohio, 1996.
- Inger Rosengren. The thematic/categorical distinction revisited. *Linguistics*, 35:439–479, 1997.
- Hans-Jürgen Sasse. The thematic-categorical distinction revisited. *Linguistics*, 25:511–580, 1987.
- Robert Stalnaker. Assertion. In Peter Cole, editor, *Pragmatics*, number 9 in *Syntax and Semantics*, pages 315–332. Academic Press, New York, 1978.
- Mark Steedman. Information structure and the syntax-phonology interface. *Linguistic Inquiry*, 31: 649 – 689, 2000.
- Enric Vallduví. *The informational component*. Garland, New York, 1992.
- Georg von der Gabelentz. Ideen zu einer vergleichenden Syntax. *Zeitschrift für Völkerpsychologie und Sprachwissenschaft*, 6:376–384, 1869.