

Identifying “aboutness topics”: two annotation experiments*

Philippa Cook

PHILIPPA.COOK@FU-BERLIN.DE

*Institut für Deutsche und Niederländische Philologie
Freie Universität Berlin
14195 Berlin, Germany*

Felix Bildhauer

FELIX.BILDHAUER@FU-BERLIN.DE

*Institut für Deutsche und Niederländische Philologie
Freie Universität Berlin
14195 Berlin, Germany*

Editor: Stefanie Dipper, Heike Zinsmeister, Bonnie Webber

Abstract

This paper deals with the annotation of “aboutness topic” (also known as “sentence topic”) in naturally occurring data. We report on two annotation experiments involving German newspaper texts: In experiment 1, based on the annotation guidelines by Götze et al. (2007), two expert annotators had to select the aboutness topic from among a small number of pre-defined choices, for a total of 588 sentence tokens. Although the results are disappointing in terms of inter-annotator agreement (with Fleiss’ κ between .19 and .57, depending on sentence category), they allowed us to identify and analyze cases that resist easy classification. Next, the original guidelines were tentatively modified, capturing insights from experiment 1. In a second experiment, based on the revised guidelines and following several training sessions, four raters annotated a total of 56 sentences, deciding for each NP whether or not it is an aboutness topic. Overall, the results are still disappointing from an inter-rater agreement perspective, but again, a closer look at the problematic cases is instructive as to the weaknesses of the current operationalizations of information structural notions and, perhaps most importantly, it reveals a number of basic theoretical questions that still await satisfactory answers.

Keywords: information structure, topic, focus, thematic, annotation, inter-rater agreement

1. Introduction

Research on information structure (IS) may serve a twofold purpose: first, information structure constitutes an intriguing area of investigation in its own right, although numerous concepts and their interrelations are nevertheless 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 Ørnes, 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

*. We would like to thank Julia Ritz for providing the text material used in experiment #2 and for helping us with the annotator training. We are also grateful for the constructive comments by the anonymous reviewers. All remaining shortcomings are of course ours. The research reported in this paper was funded by the *Deutsche Forschungsgemeinschaft*, SFB 632 “Information Structure”, Project A6.

“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 largely agree 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 “aboutness topics” (also known as “sentence topics”) in naturally occurring data. The data used in the annotation experiments were extracted from the German reference corpus DeReKo (Institut für Deutsche Sprache, 2009) (first experiment), as well as from the *Tübinger Baumbank des Deutschen / Zeitungskorpus (TüBa-D/Z)*¹ (second experiment). Section 2 outlines the criteria used by the annotators of the first experiment for identifying aboutness topics and relates them to alternative approaches to topic-hood. Section 3 reports the first experiment, the results of which are discussed in Section 3.5, where we identify the type of data that turns out to be particularly difficult to assess and seek to establish exactly which features are involved in these cases and why they give rise to diverging annotations. On the basis of these results, we finally modify the annotation guidelines used in experiment #1. Section 4 reports on a second experiment that was carried out using the revised version of the guidelines and discusses its outcome. In Section 5, we raise several questions relating to both the theoretical underpinnings of the guidelines and to the more practical issue of how the guidelines may be further improved by making them more specific about particular phenomena.

2. “Topic” in theory and in the annotation guidelines

The notion of “topic” we are dealing with in the studies being reported on here is that of “aboutness topic” (AT). 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 topicality 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 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, 41) 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 C[common] G[round] content.

1. <http://www.sfs.uni-tuebingen.de/tuebadz.shtml>

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

The notions of “topic” and “comment” have sometimes been mixed up with the notions of “background” and “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 discourse-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. (2b) 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 (*a good friend of mine*) but still serves as aboutness topic (although see the discussion in Section 3.5.2 below since this possibility is not wholly uncontroversial). Finally, there is the possibility that the comment is not always identical to the focus, i. e. the focus may only be a sub-part of the comment as shown in (4b) below. Finally, there is also the possibility of non-new foci, as in (5b) (viz. the discussion of second-occurrence foci; Partee, 1999).

- (2) a. What do your siblings do?
 b. [My [SISter]_{FOC}]_{TOP} [studies MEDicine]_{FOC} (Krifka, 2007, 44)
- (3) [A good friend of mine]_{TOP} [married Britney Spears last year]_{COMMENT} (Krifka, 2007, 42)
- (4) a. When did [Aristotle Onassis]_{TOP} marry Jacqueline Kennedy?
 b. [He]_{TOP} [married her [in 1968]_{FOC}]_{COMMENT} (Krifka, 2007, 42)
- (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 Vallduví’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 (2b) should be allowed by the theory; see Section 3.5.2). Under Krifka’s approach that we adopt here, a sentence typically has only one aboutness topic (while Reinhart, 1981, p. 86 fn. 17, excludes altogether the possibility that a sentence has more than one aboutness topic). Sentences which lack a topic – or perhaps more precisely, a topic/comment articulation – are classed as topicless/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 3.5 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 “framesetting 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 next section where we report on an annotation experiment based on the aforementioned criteria. Motivated by the outcome of the first experiment, these criteria were then modified and the effects were assessed in a second annotation experiment, reported in Section 4.

3. Annotation experiment #1

3.1 Method

In the first annotation experiment, the annotators had to decide for a number of target sentences which constituent, if any, is the aboutness topic, basing their decision on the annotation guidelines of Götze et al. (2007). For each sentence, the annotators made a choice from a small number of pre-selected topic-candidates (such as the subject-XP, object-XPs and deictic adverbial expressions such as *hier* ‘here’, *dann* ‘then’).

3.2 Materials

All target sentences were sampled, along with some preceding and following context, from the DeReKo corpus (Institut für Deutsche Sprache, 2009) (the vast majority of the material in DeReKo consists of newspaper texts). They were selected such that every sentence contained one of the four verbs shown in Table 1, instantiating a specific argument frame (a subject-XP is taken for granted in each case and therefore not listed explicitly).³ After discarding occurrences in questions, relative clauses, conditionals, titles and in the first sentence of quotations, a total of 588 sentence tokens (between 136 and 167 per verb) were included in the study.

3.3 Subjects

Each target sentence was assessed by two independent coders (the authors of this paper). The annotators were familiar with concepts of information structure beyond the explanations in the annotation guidelines that were used (this is important because differences in annotation behaviour

3. The choice of these four verbs was driven by the requirements of a pilot study in which we investigated the preferred realization of the sentence topic in connection with non-agentive (or not prototypically agentive) subjects.

| 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

| | | aboutness topic | | | thetic vs. topic-comment | | |
|--------------------|-----|-----------------|------------------|--------|--------------------------|------------------|--------|
| Verb | N | % agr. | Fleiss' κ | p | % agr. | Fleiss' κ | p |
| <i>profitieren</i> | 136 | 80.7 | .57 | < .001 | 95.6 | -.02 | .79 |
| <i>herrschen</i> | 138 | 73.9 | .54 | < .001 | 76.8 | .51 | < .001 |
| <i>geraten</i> | 147 | 67.3 | .25 | < .01 | 68.0 | .26 | < .01 |
| <i>reagieren</i> | 167 | 63.5 | .19 | < .001 | 88.6 | .23 | < .01 |
| all sentences | 588 | | | | 82.3 | .44 | <.001 |

Table 2: Inter-rater agreement in experiment #1

could in principle be due to annotators having different underlying concepts of “topic”, even if they are using the same guidelines).

3.4 Results

For each one of the four verbs, we calculated Fleiss' κ (Fleiss, 1971) as a measure of inter-rater agreement on the choice of the aboutness topic, as shown in Table 2.⁴ In addition, we calculated a κ -value for the more coarse-grained distinction between thetic (topicless) and topic-comment sentences, both per-verb and for all four verbs taken together.

All κ values for the choice of an aboutness topic indicate a statistically significant degree of agreement. That said, inter-rater agreement is highly variable across the four verbs, never exceeding $\kappa = .57$. In the case of *reagieren*, agreement does not even reach a “fair” level, following Landis and Koch's (1977) interpretation of κ -values. Turning to the distinction between thetic and topic-comment sentences, κ -values are even lower here although most of them are still statistically significant. The fact that agreement is basically at chance level in the case of *profitieren* raises the question whether the annotation guidelines can be interpreted in substantially different ways by different annotators (we will turn to this question in the discussion of the second experiment).

3.5 Discussion

In our view, the low level of inter-rater agreement observed in experiment #1 is quite unacceptable given that both annotators based their judgements on the same guidelines. On the one hand, it

4. We are using Fleiss' κ for m raters instead of Cohen's (1960) κ here because we want to be able to compare the figures to the outcome of the second annotation experiment, in which m = 4 raters participated.

could be the case that the concept of “aboutness topic” simply cannot be operationalized in a way that allows annotators to consistently categorize naturally-occurring data. On the other hand, it is possible that the annotation guidelines and/or the underlying theory are not specific enough and thus leave too much room for interpretation. A good starting point to decide between these two alternatives then is to inspect more closely the tokens on which the annotators did not agree. In this regard, we could identify data that proved particularly difficult to assess. Most of the controversial cases fall 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 that we are aware of)
- 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 3.5.1 will briefly illustrate a number of cases where the annotators did in fact agree, and Section 3.5.2 will address examples from the two problematic categories listed above.

3.5.1 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 (7b)), or “accessible” (as in (8b)), 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
 on the picturesquely situated farm road reigns actually active
 Autobetrieb.
 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
on this question would probably many non-economists with typically
BWLer” reagieren.
economist react
‘Many non-economists would probably react to this question by saying “that’s 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)
with her car got on the A 14 in height municipal.area Koblach a woman (18)
aus Mellau ins Schleudern.
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.’

3.5.2 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 analyzed 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 kommu-
niziert.
‘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
from this experience can I at.the new place evidently
profitieren
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, as mentioned in Section 2 above, we do not adopt Vallduví’s (1992) approach of identifying the aboutness topic positionally, as it is well known that the aboutness topic can occupy positions other than the initial position in German. The difficulty lies in weighing up 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 (10b) because one of these expressions (namely the subject NP) is arguably focussed.

- (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 (11b) correspond to different discourse strategies: analyzing *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.⁵

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 each 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. Independent 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 overlap of (new-information) focus and AT as sketched in (12b) in which *Kim* bears both topic and focus but the other annotator rules this out. This issue is in fact far from resolved in the literature on information structure. An anonymous reviewer of this paper was surprised that we should even consider that this complete overlap of topic and focus should be ruled out but, on the other hand, one finds occasional references in the linguistics literature which do indeed seem to rule out this possibility (see e. g. Féry, 2007, 168, whose definition of aboutness topic includes the restriction that it be “crucially followed by a focus constituent”).

- (12) a. Who ate the apple?

5. In 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.

- b. Kim ate the apple.
 []_{FOC} []_{BACKGROUND}
 []_{TOP} []_{COMMENT}

If one disallows a total overlap of topic and focus, i.e. one explicitly disallows a partitioning such as that in (12b), then the question is, of course, how the utterance should instead be analyzed. One possibility which we will discuss now is that in such cases we actually have a topicless sentence. Examples such as (13b) and probably also (14b) could thus perhaps be topicless. This, however, raises questions about the possible complexity of topicless/thetic 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 Eintracht]_{TOP?},
 a different atmosphere reigns in.the Kirchbergian Restaurant Eintracht
 wo das Rauchverbot strikt eingehalten wird.
 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. 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 Bundesvorsitzenden
 Ludwigshafen SPD-dignitaries on the resignation of.the federal party leader
 Kurt Beck.
 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 (13b) 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 aboutness 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.⁶ Under the latter view, the

6. It is worth noticing here that (13b) might be a contrastive topic: both *eine andere Stimmung* and *Kirchberger Restaurant Eintracht* are contrasted against elements that have been previously mentioned or can be inferred from 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 (13b) 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 (the main clause of) (13b) requires two prosodic peaks. Furthermore, a description of this utterance as event-reporting, a further characteristic ofthetic, (cf. Götze et al., 2007, 163) does not seem quite correct either since, as mentioned above, the function of sentences like (13b) is to introduce or present a new entity (not an event) which may then later function as aboutness topic in the next discourse chunk.

Concerning example (14b) now, 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 (13b) in that there is no contrastive element in initial position (thus we are not dealing with a contrastive topic strategy). Further, while it was clear in (13b) that the main accent would fall 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 (13b) holds, i.e. we have an overlap of focus and topic (which, recall, some researchers find problematic while others do not). 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?”. The sentence is discourse-initial, preceded only by a headline, and unless the sentence-final NP is to be analyzed as AT (an option neither annotator took), the only remaining possibility is to classify it as lacking an AT if a complete overlap of topic and focus is not an option. However, as was the case with example (13b), in its natural context sentence (14b) requires more than a single prosodic peak and thus 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, 2001) who terms this type ofthetic ‘presentational’ (vs. ‘event-reporting’) as well as Sasse (1987) where this type is 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 analyzed in the model we are assuming, one may classifythetic in general as “all-comment” but not necessarily as “all-focus”. The difference between entity-centralthetic and event-centralthetic can then be captured by recourse to their differing focus structures. Only event-centralthetic involve focus spreading across the whole utterance whilst with entity-centralthetic it is merely the phrase that denotes the introduced referent that is focused (illustrated in Figure 1). A distinction between event-reporting and entity-presentingthetic was not part of the annotation guidelines at the time of the first annotation experiment. It has been included in the revised version of the guidelines, which will serve as the basis for the second experiment (see Section 4).

preceding text. However, identifying (13b) as a contrastive topic does not help in deciding whether or not the sentence has an aboutness topic, for it has been shown that contrastive topics behave differently and crucially do not necessarily involve “aboutness”. See in particular Jacobs (1997, 2001); Büring (2003); Krifka (2007).

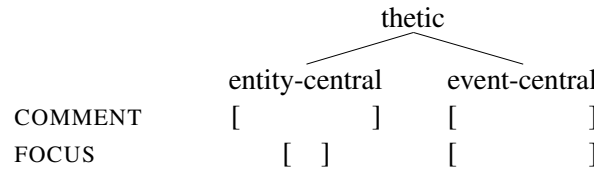


Figure 1: Analysis of different types of thetics in terms of FOCUS and COMMENT

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

- (15) a. TOPIC-COMMENT-STRUCTURE:
 [Mit großem Bedauern und totaler Überraschung reagierte gestern]_{COMMENT} [[die Ludwigshafener SPD-Prominenz]_{FOC}]_{TOP} [auf den Rücktritt des Bundesvorsitzenden Kurt Beck]_{COMMENT}
- b. ENTITY-CENTRAL THETIC:
 [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 (13b) and (14b) 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. In sum, then, the annotation guideline that were used in this experiment constitute a 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, that have to be addressed both theoretically and in terms of application (i. e., via modifications to the annotation guidelines). The next section presents some such changes to the guidelines and reports on a second annotation experiment in which these modifications were taken into account.

4. Annotation experiment #2

4.1 Revised annotation guidelines

In order to evaluate the revised version of the annotation guidelines that had been used in experiment #1, a second annotation experiment was conducted. The relevant modifications of the guidelines are as follows:

- From among the tests for topic-hood listed in (6) above, (6a) was dropped (because it was unclear how deaccentuation of the given material should be handled):

- (6a) S would be a natural continuation to the announcement
Let me tell you something about X

- The distinction between entity-central and event-central thetics that we presented above was incorporated into the guidelines. The revised version gives an example of each of these types, along with an explanation:

Topicless sentences come in two variants: The main function of an “entity-central” topicless sentence (as in (16a)) is to introduce a new entity (which, in many cases, serves as the topic in the subsequent sentence). In an “event-central” topicless sentence (as in (17)), the emphasis is not so much on introducing an entity, but rather on presenting a situation as a whole.

- (16) a. (First sentence of a newspaper article)
 [Detectives investigating phone hacking at News International have arrested [a 31-year-old woman]_{FOC}]_{THETIC}.
- b. (Continuation)
 [The woman]_{TOP} is believed to be. . .
- (17) [[It is raining]_{FOC}]_{THETIC}.

4.2 Method

In the second annotation experiment, annotators were presented with several short texts and had to decide for every NP or PP whether or not it is an aboutness topic (or, alternatively, whether they were dealing with a thetic sentence as defined above). This differs from the approach taken in the first experiment, where annotators chose the aboutness topic from a small, pre-selected set of options (like *subject*-NP, *object*-NP etc.), and it reflects more closely the typical context of use of the annotation guidelines. The annotation task was done electronically, using the open-source tool MMAX2.⁷ Prior to the actual annotation task, all annotators participated in a training phase consisting of (a) answering clarification questions about the annotation guidelines, (b) performing a test annotation jointly with the experimenters, (c) annotators performing a test annotation on their own and (d) discussing the results of the individual test annotations. After the initial training phase, the annotators carried out the actual annotation task within three weeks, at home.

4.3 Materials

Seven short texts (5 – 11 sentences each; 8 on average) from the German newspaper *die tageszeitung* (extracted from the *TüBa-D/Z* corpus) were used in the experiment, thus matching the text type used in experiment #1. The markables (NP and PP constituents) were indicated, thus the annotators did not have to identify them on their own.

4.4 Subjects

A total of four (2 m, 2 f; age 21–42 years, 29.5 on average) students with a background in linguistics (0.5–4.5 years of study) participated in the experiment. All of them are native speakers of German and were largely unfamiliar with notions of information structure prior to the experiment.

7. <http://mmax2.net>

4.5 Results

The distinction between thetic and topic-comment sentences yields 56 data points per annotator (one for each sentence), with 39.3% raw agreement between the annotators. In 21 cases, all annotators agree that a sentence has a topic-comment structure, while there is only a single case in which all four agree that they are dealing with a thetic sentence (though they do not agree on the subtype of thetic as defined in Section 4.1). With respect to the identification of aboutness topics, we examined a total of 516 data points for each annotator. In 17 cases, all annotators agree in their choice of the aboutness topic (and in 392 cases they agree that a constituent is not an aboutness topic), thus raw agreement is relatively high (79.3%). In order to factor out chance agreement, Fleiss’s kappa for $m = 4$ raters was calculated, both for the annotation of aboutness-topics and for the more coarse-grained distinction between thetic and topic-comment sentences, as shown in Table 3. Although the results reach statistical significance, the amount of agreement beyond chance is only “moderate” for the choice of the aboutness topic, while there is merely “fair” agreement on the thetic/topic-comment distinction.⁸

| | N items | % raw agreement | $\kappa(\text{Fleiss})$ | p |
|----------------------|-----------|-----------------|-------------------------|--------|
| aboutness topic | 516 | 79.3 | .447 | < .001 |
| thetic/topic-comment | 56 | 39.3 | .225 | < .001 |

Table 3: Overall inter-rater agreement in experiment #2

4.6 Discussion

It might look odd at first glance that there is considerably less agreement on the question of whether or not raters are dealing with a non-thetic (i. e., topic-comment) sentence at all, than on the actual choice of the aboutness topic. This is probably an artifact of the experiment design: a decision (topic or not?) had to be made for a very large number of markables, including sub-parts of NPs to which raters are not likely to attribute information structural properties in a principled way, at least not at the level of training of our participants. A large amount of “not a topic” annotations thus makes up for the major part of the observed agreement. This is illustrated by the fact that, if only those items are taken into account which at least one of the raters annotated as aboutness topic ($N = 143$), there is only slight agreement ($\kappa = .067$) and statistical significance is marginal ($p = .051$). There are 5 cases, though, in which different annotators did class different sub-parts of an NP as the topic. Since such cases are not explicitly discussed in the annotation guidelines, this kind of embedded markable might have caused confusion that ultimately lowered overall inter-rater agreement. We “repaired” these annotations by attributing topic status to the embedding NP. Even with this correction applied, inter rater agreement does not improve substantially ($\kappa = .476$).

Inter-rater agreement on the choice of a specific subtype of “thetic” cannot be evaluated in the same way since there is only a single case in which all annotators agree that a sentence is thetic in the first place.

Another interesting question to ask is whether or not the annotators agree on a specific topic once they all have decided that they are dealing with a topic-comment sentence (and not with a thetic one). We recalculated inter-rater agreement on the topic constituent for the subset of sentences that

8. As before, according to Landis and Koch’s (1977) interpretation of κ -values.

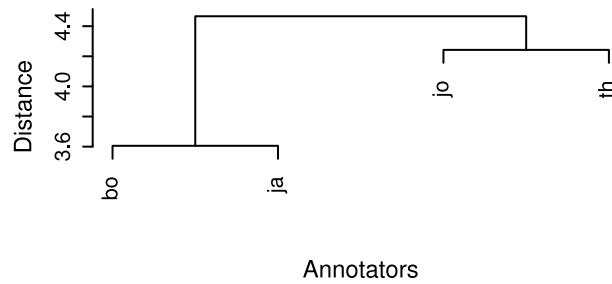


Figure 2: Similarity in distinction “thetic/topic-comment” (euclidean distance, clustering based on average distance): annotators fall into two groups.

all annotators had rated as a topic-comment-structure. The resulting, higher coefficient ($\kappa = .563$, $p < 0$, $N = 21$) indicates that in those cases which are clear instances of topic-comment sentences, annotators tend to agree more on the aboutness topic, though agreement is far from being perfect in these cases, too. Example (18b) illustrates one such sentence for which all four annotators agreed on the choice of the aboutness topic.

- (18) a. Friedensbewegung, christliche und linke Gruppen sowie Gewerkschafter aus der ganzen Republik haben dagegen zu einer Demonstration gegen den Krieg in Jugoslawien aufgerufen.
 ‘The peace movement, christian and leftist groups as well as trade unionists from all over Germany have announced a demonstration against the war in Yugoslavia.’
- b. [Die Protestveranstaltung] steht unter dem Motto „Stoppt den Krieg! Helfen
 the protest.event stands under the motto stop the war help
 statt Bomben!“. instead.of bombs
 ‘The protests will take place under the motto “Stop the war! Help them, don’t bomb them!”

In the discussion of experiment #1, we briefly pointed to the possibility that the annotation guidelines leave too much room for interpretation and thus can be understood in different ways. In order to investigate this question further, we computed a distance matrix for the annotations of our four participants that would allow us to group them according to the similarity of their annotations. With respect to the distinction between “thetic” and “topic-comment”, there is in fact evidence that the guidelines are ambiguous, as the cluster tree in Figure 2 illustrates: the annotators fall into two groups that differ in their rating behaviour. The picture is not so clear, though, for the choice of the aboutness topic. Here, two of our annotators pattern closely together, while the third and the fourth annotator differ both from each other and from the first two annotators (see Figure 3).

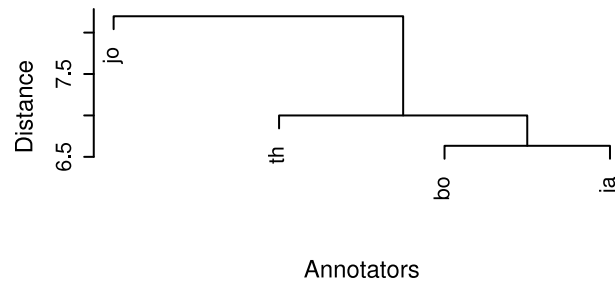


Figure 3: Similarity in choice of aboutness topic (euclidean distance, clustering based on average distance): no clear pattern.

5. General Discussion

Topic: selecting among various candidates

A high degree of agreement on topic choice was often observed when the phrase in question refers to an entity that can be analyzed as the topic of the preceding utterance(s), and is realized in a syntactically prominent position (e. g. as the subject), as is the case in example (18b) given in Section 4. On the other hand, we observe that a typical configuration where annotators disagree involves cases in which the topic of the preceding utterance(s) is realized in a syntactically or functionally less salient position, while another expression that was not the topic in the preceding utterance(s) occupies a more salient position in the utterance in question. This pattern, illustrated in (19), is reminiscent of the “retain” transition type of Centering Theory (Grosz et al., 1983, 1995), which typically announces a shift of the “backward looking center” (marked as C_b in example (19)). In (19b), the (referent of the) NP *Hamburg* counts as the backward-looking center because it is coreferential with the most salient referent of the preceding utterance. It is arguably also the most salient referent in (19b). Now, in (19c), it is referred to again, thus (the referent of) *Hamburg* is also the backward-looking center of that utterance. However, it is not realized in the most salient position (the subject position) in (19c); instead the subject position is occupied by the NP *Kanal4*, which is probably why some annotators analyzed *Kanal4* as the aboutness topic. While the concept of “backward looking center” is not fully coextensive with our understanding of “aboutness topic” (but see e. g. Beaver, 2004, who does equate these two concepts), the two notions are close enough to make this observation noteworthy, if only because it allows one to describe in precise terms one of the contexts in which confusion is likely to arise: the manifest conflict between structural saliency on the one hand and referential continuity on the other hand is surely an issue that should be addressed in any future version of the annotation guidelines.

- (19) a. Hamburg - Eine Stadt verabschiedet sich vom Rechtsstaat, RTLplus , So. , 0.40 Uhr
 ‘Hamburg - a city bids farewell to constitutional law, RTLplus, Sunday, 12:40 a. m.’
 b. Ist von [Hamburg] C_b die Rede, denken viele an eine weltoffene und liberale Stadt.
 is from Hamburg the speech, think many of a world.open and liberal city
 ‘If they hear „Hamburg”, many people imagine an open-minded and liberal city.’

- c. [Kanal4] zeigt am Sonntag zu später Stunde bundesweit auf RTLplus eine andere
 Kanal4 shows on Sunday at late hour nationwide on RTLplus a other
 Seite [der Hansestadt]_{C_b}.
 side of.the hanseatic.city
 ‘Kanal 4 will be showing a different side of the hanseatic city late on Sunday in a
 national broadcast on RTLplus.’

Topic-Comment vs. Thetic

One of the main current weaknesses leading to poor inter-rater agreement, we believe, lies in the fact that in both annotation tasks the premise was adopted that every sentence either has an aboutness topic or else it is necessarily a thetic (i. e. topicless) utterance. The annotation guidelines (both the original and the revised version) simply do not allow for any sentence type that does not fit into either of these categories. This is perhaps an artifact of current research on information structure in theoretical linguistics. Typically, studies on information structure in that particular research tradition have not thus far focused on naturally-occurring data. When one examines such data, however, as was the case in the annotation experiments, it can be observed that utterances which do not straightforwardly fit into the postulated available utterance types (topic-comment vs. thetic) actually seem quite frequent.

From the results reported in the preceding section, it is obvious that the modification of the annotation guidelines, i. e. slightly greater explicitness in characterizing what should count as a thetic sentence, along with the distinction between two types of thetics and examples for each one of these, is not helpful when it comes to deciding whether or not a given sentence is to be classed as thetic. For example, while in the second experiment, there is a number of sentences (37.5%) that seem to qualify as clear instances of a “topic-comment” structure (those which the annotators unanimously classed as such), the same does not hold for the category “thetic”, for which there is only one case where all annotators agree. Given the fact that not even the revised definition of “thetic”, as adopted in the revised guidelines, seems to accommodate naturally-occurring data, although it already covers cases that are structurally much more complex than the kind of sentences traditionally called “thetic”, there seems to be an urgent need to clarify the notion of “thetic” from a theoretical perspective. In particular, it is quite unclear how complex a thetic sentence can be and how one can even tell with any certainty whether a sentence is thetic or not, as there are no clear criteria offered in the literature that we are aware of. For instance, some authors restrict theticity to intransitive verbs while others clearly do not, yet this contrast does not appear to be openly addressed or commented on in the theoretical literature. In this connection, the question arises whether the clear-cut distinction between thetic and non-thetic is in fact an idealization that is not necessarily reflected in naturally-occurring data: alternatively, it is conceivable that there is a gradual transition from sentences with a clear topic-comment structure via sentences with a less salient topic-comment structure to sentences that do not display a topic-comment structure at all (thetics).

Appropriateness of thetic vs. topic-comment

There is also the possibility that the distinction between “thetic” vs. “topic-comment” is not equally pertinent for different kinds of utterances. It is plausible, and indeed sometimes necessary, to assume the existence of other utterance types which do not serve to present an entity nor an event (i. e. are not thetic) but which also do not have the function of a topic-comment utterance, namely advancing

the Common Ground by predicating the information expressed in the “comment” of some entity under whose file card this information is to be stored. While the exact characterization of such utterances and a detailed description of constraints on their occurrence is currently not available, we will briefly discuss a few examples which deserve further study in this respect, viz. sentences that do not express an assertion, and subordinate clauses.

In the case of sentences that do not serve to make an assertion, such as interrogatives, imperatives and exclamatives, the simple dichotomy of topic-comment vs.thetic does not seem sufficient. Our intuition is that since these encode a different type of speech act, it seems very likely that they have a different function in terms of discourse structuring to that normally attributed to topic-comment configurations. This intuition is of course in need of closer scrutiny and empirical testing. Currently, the annotation guidelines do not state whether or not such sentences should be considered to host an aboutness topic and so, in the absence of any specific instruction to the contrary, the annotators sought to annotate these utterance types as either containing an aboutness topic or as instantiating a thetic utterance. Consider, for instance, the following example which is a verb-first conditional:

- (20) Ist von Hamburg die Rede, denken viele an eine weltoffene und liberale Stadt
 is of Hamburg the speech think many of a world.open and liberal city
 ‘If they hear “Hamburg”, many people imagine an open-minded and liberal city.’

Two annotators in experiment #2 selected the NP *Hamburg* as the aboutness topic. One of these annotators also selected a further aboutness topic in the consequent clause of the conditional (namely the NP *eine weltoffene und liberale Stadt*). The other two annotators selected no aboutness topic at all in the whole utterance (neither in the antecedent nor the consequent clause). Thus, with the current dichotomy offered in the annotation guidelines (if no topic, then thetic), two annotators classed (or were indirectly forced to class) this utterance as thetic. As mentioned above, it is not implausible to assume that the antecedent of a conditional necessarily does not involve a topic-comment partition at all. This, of course, should not entail that the sentence must therefore automatically be considered thetic either. Rather, the conditional clause and the consequent clause taken together appear to provide the correct domain for attributing discourse functions such as topic and comment, with the conditional clause establishing the topic, and the consequent clause predicating of it. However, the antecedent clause, considered in isolation, seems to simply have a very different discourse structuring function and as such the question of whether it is thetic or not or of which phrase instantiates the aboutness topic is just not a relevant question. A similar situation holds for exclamative utterances such as (21) which probably have a different function in discourse than either that of predication (about a topic) or theticity. One annotator in experiment #2 selected the expletive object *es* (which is part of a fixed idiomatic expression) as aboutness topic. The others didn’t mark any element as aboutness topic, which again illustrates the difficulties that annotators are facing when confronted with this type of sentence.

- (21) Besser kann man es gar nicht machen!
 better can one it not make
 ‘There is no better way to do it.’

For future studies, it would probably be advisable to give detailed instructions as to how these sentence types should be handled, since in the absence of explicit instructions, annotators are bound to become confused.

Turning now to subordinate clauses, we observe that the task of identifying their information structural partitioning poses a major problem for the annotators of experiment #2, giving rise to numerous conflicting judgements. In some cases, the subordinate clause was not independently assessed and in some it was (recall the discussion about annotating the main- and the relative clause structure of (13b) in the first experiment). It is not evident whether all subordinate clauses should be considered to constitute an independent IS domain of their own, independent of that of their matrix clause, or whether matrix and subordinate clause together should be considered to constitute one broad IS Domain. For the case in point, the question boils down to whether or not a subordinate clause should be considered to potentially embed a topic-comment partition separate from the topic-comment partition of the matrix. In principle, there seems to be no ban on the embedding of thetic utterances per se.

At present, there is no detailed discussion of the treatment of subordinate clauses in the annotation guidelines and annotators were simply instructed to annotate all finite subordinate clauses except restrictive relative clauses. Again, we believe there is a need for careful empirical testing of whether or not this is the correct procedure. It seems intuitively to make sense to distinguish finite and non-finite embedded clauses since the latter typically do not overtly express the subject of the embedded predicate and this therefore is simply not available as a contender for topic status. Often this leaves no argument other than the object role available as a topic. Annotators were thus instructed in the guidelines not to consider non-finite clauses as separate IS domains. Further it also seems likely that the choice of the subordinating conjunction plays a role in the answer as to whether or not the embedded clause involves an independent IS domain. It has been observed that different (classes of) conjunctions instantiate different discourse relations and impose different constraints on the “common ground”. It is thus not implausible that the embedded propositions these classes introduce differ with respect to their internal IS. This point perhaps also applies to different coordinating conjunctions (cf. *und* ‘and’ vs. *aber* ‘but’). Further, the semantics of the embedding predicate may well affect whether or not a subordinated proposition constitutes an independent IS domain. It has been suggested, for instance, that presupposed complements (of e. g. factive verbs) are not asserted and may therefore lack a topic-comment articulation altogether (cf. Cook and Ørsnes, 2010; Ebert et al., 2009; Kuroda, 2005). In this respect, example (22c) is typical of a number of cases that caused problems. The matrix verb *aufdecken* ‘reveal’ is factive and is followed by a subordinate clause in the function of direct object which is introduced by the subordinating conjunction *wie* ‘how’. The annotators did not agree here on whether or not *Hamburg* (or *der Stadtstaat Hamburg*) should count as an aboutness topic.

- (22) a. Ist von Hamburg die Rede, denken viele an eine weltoffene und liberale Stadt. Kanal4 zeigt am Sonntag zu später Stunde bundesweit auf RTLplus eine andere Seite der Hansestadt.

‘If they hear “Hamburg”, many people imagine an open-minded and liberal city.’

- b. Kanal4 zeigt am Sonntag zu später Stunde bundesweit auf RTLplus eine andere Seite der Hansestadt.

‘Kanal 4 will be showing a different side of the hanseatic city late on Sunday in a national broadcast on RTLplus.’

- c. Die beiden Fernsehjournalisten Ernst Matthiesen und Oliver Neß decken auf, wie
 the two tv.journalists E. M. and O. N. reveal up how
 sich der Stadtstaat Hamburg immer mehr in Richtung eines „Polizeistaats“
 itself the sovereign city of Hamburg ever more in direction of a police.state
 bewegt.
 moves
 ‘The two TV journalists E. M. and O. N. reveal how the sovereign city of Hamburg
 begins to turn into a police state.’

Similarly, the different possible syntactic realizations of embedded clauses could be of significance. German permits embedded verb-last and verb-second clauses (both with and without subordinating conjunction) and it is plausible that this too correlates with different IS possibilities. Moreover, it is not clear whether the distinction between argument clauses and adverbial clauses also relates to IS differences; and indeed whether different sub-types within both argument and non-argument clauses need to be distinguished. Finally, the treatment of complements of non-verbal categories such as nouns and adjectives must also be considered.

Discourse relations

Finally, another issue which seems worthy of consideration in this connection concerns the discourse relations between utterances (in the sense of e. g. Asher and Lascarides, 2003; Mann and Thompson, 1988) and the role of the notion of aboutness topic in such discourse relations. While certain discourse relations do appear to serve to advance the “common ground” in the way described by Krifka (2007) such as, for instance, “restatement”, “elaboration” and “narration”, it is perhaps worth considering whether a particular utterance may or may not instantiate a topic-comment partition as a function of the discourse relation that connects this utterance to the remaining discourse. As mentioned by Krifka (2007) at the end of his IS overview, this interaction between discourse relations and IS has not been extensively examined and there is a need for further research here. In this vein, an anonymous reviewer also points out to us that frameworks such as e. g. Asher and Lascarides (2003) and related frameworks provide a tool for identifying the arguments of certain discourse relations as contributing static (rather than eventive) information. This raises the (attractive) possibility that a corpus annotated with this level of description (which is independently useful anyway) could turn out to be helpful in clarifying (or even perhaps obviating the need for) the distinction betweenthetic and topic-comment utterances which were, recall, a particular source of problems in both of the annotation exercises. In general, these comments are suggestive to us that an optimally annotated corpus should include annotations of numerous different levels (focus-background, topic-comment, given-new, discourse relations etc.) in combination with one another so that cross-cutting information can be drawn on in order to enhance the precision of annotation. This is an important and promising direction for future research, we believe.

Explicitness of the annotation guidelines

As a consequence of what has been discussed so far, it is obvious that the annotation guidelines must be much more specific as to how particular constructions are to be analyzed. The guidelines should, for instance, provide concrete instructions as to how different types of subordinated and non-assertive utterances should be treated and how coordinations are to be handled. While it is

desirable to use the same basic notions of information structure as a fundament for annotating texts in a variety of different languages, we doubt that the necessary degree of explicitness in the guidelines can be achieved without language-specific instructions on how particular phenomena are to be treated. Taking the results reported in this paper as a baseline, we believe that introducing a number of language-specific rough-and-ready distinctions would help raise the overall figure of inter-rater agreement, even though it might not do justice to a putatively small number of cases that feature a somehow “uncanonical” link between syntactic form and information structure.

Broadening the concept of Topicality

As we have mentioned at various points above, there are many different properties that are associated with topicality. The notion of aboutness topic we selected to work with and which was outlined in Section 1 explicitly allows topics to be new in the discourse, although it has long been acknowledged (and we would also agree) that notions such as familiarity and givenness prototypically align with topicality (cf. Chafe, 1976 but also Reinhart, 1981). The definition of aboutness topic is also not tied to any syntactic position but the idea that topics tend to be realized in syntactically prominent positions or undergo a kind of separation (cf. Rizzi, 1997 but also e. g. Jacobs, 2001) is also well-known. The concept of invoking lists of referents or of partiality or of addressing sub-questions under discussion also appears to tendentially align with topical status (cf. Roberts, 1996; Büring, 2003 and also the notion of a partially-ordered set relation (poset) of Birner and Ward, 1998). Again, the definition of aboutness topic used here is formulated independently of such effects. The question arises (given the clear difficulty annotators had with the annotation of aboutness topic as defined here in naturally-occurring data) whether an optimal annotation of topic can better succeed if a “bundle” approach is adopted as advocated by e. g. Smith (2003, 198–199) who lists a set of “topic cues”. Annotators could plausibly be instructed to differentiate between canonical or highly prototypical topics and less prototypical topics. Examples of the former type, i. e. in which a bundle of prototypical topic properties coincide on one candidate expression, were those which achieved a higher degree of inter-rater agreement. On a slightly different note, it is plausible that there is some cross-linguistic variation concerning which properties play a greater role in defining topicality (see McNally, 1998, for a similar suggestion) and that language-specific “bundles” of topic features should be identified and spelt out in annotation guidelines.

6. Summary

In the present contribution, we reported on the results of two annotation experiments involving naturally occurring data, in the course of which raters categorized sentences as either “thetic” or “topic-comment” and, in the latter case, also identified the aboutness topic of the sentence. In experiment #1, two expert raters sought to follow the annotation guidelines of Götze et al. (2007). This resulted in a relatively low degree of inter-rater agreement as a consequence of which we identified a number of typical cases that gave rise to difficulties. As the distinction between “thetic” and “topic-comment” turned out to be particularly difficult to assess, a modification to the original guidelines was proposed that elaborated on the notion of “thetic” and broadened the range of phenomena that it can be applied to.

The revised version of the annotation guidelines served as the basis of experiment #2, in which four non-expert raters (i. e. students with a background in linguistics but no expertise in information structure) annotated several short newspaper texts. The modifications to the the original annotation

guidelines did not improve inter-rater agreement (Fleiss κ does not exceed .23 for thetic/topic-comment distinction and .48 for the choice of the aboutness topic). We conclude that neither the original nor the revised version of the guidelines provide a particularly reliable set of rules when it comes to annotation of aboutness topics. To help improve this situation, we highlighted a number of issues that should be taken into account in any future revision of the guidelines (or in any other set of annotation guidelines for information structure).

In particular, we suggest that the notion of “thetic” (and its relation to “topic-comment”) should be subjected to further scrutiny because its applicability is less than clear once the domain of study shifts from idealized, constructed examples to naturally-occurring data. In this connection, we also raised the question of whether or not the thetic/topic-comment distinction is a pertinent one for all sentences types alike. Among other things, conditionals and various other kinds of non-assertive utterances do not necessarily lend themselves to an analysis in terms of “thetic” vs. “topic-comment”. Moreover, the discussion here makes clear that the treatment of subordination is complex and probably requires consideration of numerous different parameters. Finally, we pointed to the fact that the concept of “aboutness topic” is still lacking a level of operationalization that would allow annotators to reliably identify this information structural construct when scanning naturally-occurring data. Borrowing some of the descriptive machinery of Centering Theory, we described one of the typical configurations in which annotators disagree on their choice of a sentence’s aboutness topic as a conflict between referential continuity and structural salience. In addition to these points, we also expressed serious doubts that any annotation guidelines for information structure can be stated in a way that is neutral with respect to the target language. While the guidelines used in the two experiments no doubt constitute a valuable resource and starting point, we suggest supplementing them with instructions specific to particular languages, addressing a concrete set of phenomena and explaining how these should be handled. This probably also implies introducing a number of possibly controversial decisions into the guidelines, that is, the guidelines would not necessarily reflect a relatively consensual stance on information structure anymore. However, this would seem to be preferable to a set of theoretically consensual, but vastly underspecified, annotation rules.

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. At the moment, we must concede that a careful and thorough discussion of the points raised above is required before the notion of aboutness topic, as defined here, can be employed with reasonable confidence in annotation tasks.

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.
- Nicholas Asher and Alex Lascarides. *Logics of conversation*. Cambridge University Press, Cambridge, 2003.
- David I. Beaver. The optimization of discourse anaphora. *Linguistics and Philosophy*, 27(1):3–56, 2004.
- Betty J. Birner and Gregory Ward. *Information status and noncanonical word order in English*. Benjamins, Amsterdam, 1998.

- Daniel Büring. On d-trees, beans, and B-accent. *Linguistics & Philosophy*, 26(5):511–545, 2003.
- Wallace L. Chafe. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles N. Li, editor, *Subject and topic*, pages 25–55. Academic Press, New York, 1976.
- 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*, pages 122–142, 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 Publications.
- 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.
- Christian Ebert, Cornelia Endriss, and Stefan Hinterwimmer. Embedding topic-comment structures results in intermediate scope readings. In Muhammad Abdurrahman, Anisa Schardl, and Martin Walkow, editors, *Proceedings of NELS 38*, Amherst, 2009. GLSA.
- Cornelia Endriss. *Quantificational topics. A scopal treatment of exceptional wide scope phenomena*. Springer, Dordrecht, 2009.
- Caroline Féry. Information structural notions and the fallacy of invariant correlates. In Caroline Féry, Gisbert Fanselow, and Manfred Krifka, editors, *Interdisciplinary studies on information structure*, volume 6 of *Working papers of the SFB 632*, pages 161–184. Universitätsverlag, Potsdam, 2007.
- Joseph L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378—382, 1971.
- Georg von der Gabelentz. Ideen zu einer vergleichenden Syntax. *Zeitschrift für Völkerpsychologie und Sprachwissenschaft*, 6:376–384, 1869.
- 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.

- Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, pages 44–50, Cambridge, Massachusetts, USA, 1983. Association for Computational Linguistics.
- Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–226, 1995.
- 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.
- Institut für Deutsche Sprache. Deutsches Referenzkorpus. Archiv der Korpora geschriebener Gegenwartssprache 2009-I. (Release vom 28.02.2009), 2009. <http://www.ids-mannheim.de/DeReKo>.
- Joachim Jacobs. I-Topikalisierung. *Linguistische Berichte*, 168:91–134, 1997.
- Joachim Jacobs. The dimensions of topic-comment. *Linguistics*, 39(4):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*, volume 6 of *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.
- Shige-Yuki Kuroda. Focusing on the matter of topic: a study of *wa* and *ga* in Japanese. *Journal of East Asian Linguistics*, 14:1–58, 2005.
- Knud Lambrecht. *Information structure and sentence form. Ttopic, 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(3):611–682, 2001.
- J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.
- William C. Mann and Sandra A. Thompson. Rhetorical structure theory. Toward a functional theory of text organization. *Text*, 8(3):243–281, 1988.

- Louise McNally. Towards a theory of the linguistic coding of information packaging instructions. In Peter Culicover and Louise McNally, editors, *The Limits of Syntax*, volume 29 of *Syntax and Semantics*, pages 161–183. Academic Press, New York, 1998.
- 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.
- Luigi Rizzi. The fine structure of the left periphery. In Liliane Haegeman, editor, *Elements of Grammar*, pages 281–337. Kluwer, Dordrecht, 1997.
- Craig 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, Ohio, 1996.
- Inger Rosengren. The thematic/categorical distinction revisited once more. *Linguistics*, 35(3):439–479, 1997.
- Hans-Jürgen Sasse. The thematic-categorical distinction revisited. *Linguistics*, 25(3):511–580, 1987.
- Carlota S. Smith. *Modes of discourse. The local structure of texts*. Cambridge University Press, Cambridge, 2003.
- Robert Stalnaker. Assertion. In Peter Cole, editor, *Pragmatics*, volume 9 of *Syntax and Semantics*, pages 315–332. Academic Press, New York, 1978.
- Mark Steedman. Information structure and the syntax-phonology interface. *Linguistic Inquiry*, 31(4):649 – 689, 2000.
- Enric Vallduví. *The informational component*. Garland, New York, 1992.