Annotation of Information Structure: An Evaluation Across Different Types of Texts

Julia Ritz*, **Stefanie Dipper**†, **Michael Götze***

*Institut für Linguistik, Universität Potsdam
Karl-Liebknecht-Str. 24-25
14476 Potsdam
{julia, goetze}@ling.uni-potsdam.de
†Sprachwissenschaftliches Institut, Ruhr-Universität Bochum
44780 Bochum
dipper@linguistics.rub.de

Abstract

We report on the evaluation of information structural annotation according to the Linguistic Information Structure Annotation Guidelines (LISA, (Dipper et al., 2007)). The annotation scheme differentiates between the categories of information status, topic, and focus. It aims at being language-independent and has been applied to highly heterogeneous data: written and spoken evidence from typologically diverse languages. For the evaluation presented here, we focused on German texts of different types, both written texts and transcriptions of spoken language, and analyzed the annotation quantitatively and qualitatively.

1. Introduction

Information structure (IS) deals with properties of utterances that relate to information transfer between interlocutors, e.g., properties that refer to concepts such as the information states of speaker and hearer, their attentional states, beliefs, intentions, etc. Languages differ widely with regard to the linguistic means they use to express these concepts. Such means are, for example: (de)accentuation, word order, use of particles; typically, these means do not occur in isolation but simultaneously, and they seem to be interdependent to a certain degree.

To single out the impact of the individual factors that are involved, one possibility is to collect and annotate data with elementary, preferably theory-neutral, features such as "givenness" or "contrastiveness". Having annotated a set of data, the interplay and role of the features can be studied in combination, both qualitatively (e.g. by using search tools that allow cross-level queries), and quantitatively with statistical methods.

This approach was chosen by the Collaborative Research Center "SFB 632: Information Structure – the linguistic means for structuring utterances, sentences and texts"[1] (henceforth SFB). At the SFB, corpora for IS-related research have been created, containing transcribed speech data from more than twenty typologically different languages (elicited with the typological Questionnaire on Information Structure 'QUIS' (Skopeteas et al., 2006)[2]) and digitalized historical manuscripts. For the creation of this resource, guidelines for several linguistic layers (including morphology, syntax, IS etc.) have been defined (Dipper et al., 2007). To secure comparability of annotation, we have evaluated the guidelines on data elicited under controlled conditions and – exploratively – on unrestricted text.

Apart from the typological and historical interest in IS, annotation at this level is also valuable in disourse structure analysis (cf. (Polanyi et al., 2003)). Additionally, a number of NLP applications can profit from it, e.g. anaphora resolution (cf. (Strube, 1998)), text to speech, summarization and machine translation systems.

The paper is structured as follows: Section 2. gives a brief overview of the Linguistic Information Structure Annotation Guidelines (LISA, (Dipper et al., 2007)). Section 3. describes the annotation setup and reports on the evaluation results. In Section 4., we compare our evaluation to related studies in the field, and in Section 5. we draw conclusions and delineate directions for future research.

2. Annotation Guidelines for Information Structure

Major objectives in the design of the LISA guidelines were (i) reliability of annotation, (ii) language independence, and (iii) openness towards different theories. Whereas the first objective is a standard one for many guidelines, the second follows from the diversity of language data to be annotated within the SFB. The third objective results from the wish to be rather independent from specific theories, which is of course a difficult enterprise. Another important criterion for the guidelines is applicability to different modalities and text types.

These objectives resulted in a number of design decisions in the guidelines. For instance, we use decision trees and hierarchical annotation schemes for facilitating a reliable annotation. Annotation instructions rely mainly on functional tests, rather than tests involving linguistic form, enabling the application to data of different languages. Furthermore, possibly different dimensions of information structure are annotated independently from each other, postulating no relation between these different features (as one could do, e.g., for Topic and Focus).

[1] http://www.sfb632.uni-potsdam.de/
[2] http://www.sfb632.uni-potsdam.de/~d2/

2.1. The LISA Tagset

The guidelines cover three dimensions of IS: *information status*, *topic*, and *focus*. The choice was driven by the prominence of these dimensions in linguistic theories about IS, and by their usage across different theoretical frameworks and within the SFB. The single dimensions distinguish further subcategories, e.g., `aboutness topic` vs. `frame-setting topic` within *topic*, or `new-information focus` vs. `contrastive focus` within *focus*.

Table 1 shows the core tagset of all three dimensions of IS. For *information status* and *focus*, both a core and an extended tagset (i.e. the option for finer-grained distinctions, see Table 2) are available.

IS feature	Values	Description
Information Status	giv	given
	acc	accessible
	new	new
Topic	ab	aboutness topic
	fs	frame-setting topic
Focus	nf	new information focus
	cf	contrastive focus

Table 1: Core tagset of LISA Guidelines

2.1.1. Information Status

A rather well-studied dimension of IS is the *information status* of discourse referents (cf. Prince's (1981) *givenness*, Gundel et al.'s (1993) *cognitive status*). This portion of our guidelines is closely related to Nissim et al.'s (2004) annotation scheme for information structure in dialogue (henceforth AID). Both AID and LISA are structured hierarchically, with comparable labels[3], and decision trees to direct the annotator. LISA differs from AID with respect to granularity (AID specifies more subclasses) and the treatment of expressions referring to the dialogue participants and of generic pronouns. For anaphoric expressions, LISA also takes into account activation (cf. Gundel et al.'s (1993) *in focus*). A comparison between (Prince, 1981), AID and LISA, as well Riester et al.'s (2008) more recently developed scheme (which also takes into account underspecification), is provided in (Riester, 2008).

2.1.2. Topic and Focus

Another aspect of IS is the *topic/focus* distinction. Previous approaches (e.g. (Hajičová et al., 2000), (Paggio, 2006)) have defined *topic* and *focus* as mutually exclusive categories of the same level. LISA distinguishes between the levels of *topic/comment* and *focus/background*. They are annotated independently from each other, postulating no prior relation between them. Also, no relation to a deep syntactic annotation layer is presupposed (in contrast to, e.g. (Hajičová et al., 2000)).

[3] `old` (AID) corresponds to `given` in LISA, `inferable` to `accessible`, and `new` to `new`.

IS feature	Values	Description
Information Status	giv	given (underspecified)
	giv-active	active
	giv-inactive	inactive
	acc	accessible (underspecified)
	acc-sit	situationally accessible
	acc-aggr	aggregation
	acc-inf	inferable
	acc-gen	general
	new	new
Focus	nf	new-information focus
	nf-sol	solicited
	nf-unsol	unsolicited
	cf	contrastive focus
	cf-repl	replacement
	cf-sel	selection
	cf-part	partiality
	cf-impl	implication
	cf-ver	truth value (verum)

Table 2: Extended tagset of LISA Guidelines

2.2. Illustrative Examples

For better illustration, we will present a few examples annotated according to LISA:[4]

Information status (to be annotated to every referential noun phrase (NP)): Discourse referents are `given` when coreferent with previously mentioned material (e.g. *he* referring to *Peter* in (1)). A `given` element is `inactive` if there is at least one sentence between the mentions.

(1) [Peter]$_{new}$ went into [the garden]$_{new}$. [It]$_{giv-active}$ was blooming. [He]$_{giv-inactive}$ was happy.

Referents are `accessible` if they can be inferred (a) from the situative context of the discourse (examples (2) and (3)), (b) as an aggregation of previously mentioned referents (4), (c) through relational information to mentioned referents, i.e. via bridging (relations like entity-attribute, part-whole, sub-/superset, member of the same set, etc., e.g. *its entrance* as part of *the garden* in (5)), or (d) from the assumed world knowledge of the hearer (6).

(2) (In a dialogue during breakfast) Could [you]$_{acc-sit}$ pass [the sugar]$_{acc-sit}$, please?

(3) (pointing at pictures in a book) [The kid]$_{acc-sit}$ hits [the cow]$_{acc-sit}$.

(4) [Peter]$_{new}$ went shopping with [Maria]$_{new}$. [They]$_{acc-aggr}$ bought [flowers]$_{new}$.

(5) [The garden]$_{new}$ is beautiful. [Its entrance]$_{acc-inf}$ is just across [this river]$_{acc-sit}$.

(6) [The sun]$_{acc-gen}$ set. [Pele]$_{acc-gen}$ scored [his second goal]$_{new}$.

Other referents are `new` (7).

(7) [Peter]$_{new}$ went into [the garden]$_{new}$. [Another man]$_{new}$ appeared.

[4] Annotation boundaries are represented by brackets, labels by subscripts. As to abbreviations, see Table 2. Null context is assumed unless specified.

Nonreferentials are left unannotated ((8) and (9)).

(8) [It] always rains on [Sundays]$_{acc-gen}$.

(9) [Peter]$_{new}$ kicked [the bucket].

Topic: The `aboutness topic` is the entity about which the sentence under discussion makes a predication. The `framesetting topic` specifies the frame under which this predication holds (see examples (10) and (11)). Topics may be NPs, but also prepositional phrases, adverbs etc. All-new sentences (i.e. sentences uttered as an answer to the question *What's happening?/What happended?*) do not have topics.

(10) [Peter]$_{ab}$ was wearing red socks.

(11) [Physically]$_{fs}$, [Peter]$_{ab}$ is doing very well.

Focus: The part of the utterance serving to develop the discourse is annotated as the `new information focus` (either `solicited` by a question as in (12) or `unsolicited` (13)). As for `contrastive foci`, both contrasting elements are marked. In sentences with more than one contrastive pair, co-indexing is used to link pairs (14).

(12) [Who]$_{nf}$ is reading a book? [Mary]$_{nf-sol}$ is reading a book.

(13) [Once upon a time, there was a wizard]$_{nf-unsol}$. He [lived in a beautiful castle]$_{nf-unsol}$.

(14) My [older]$_{cf_1}$ sister [works as a secretary]$_{cf_2}$, but my [younger]$_{cf_1}$ sister [is still going to school]$_{cf_2}$.

3. Evaluation

In a first evaluation of the LISA guidelines, we focused on the criterion of reliability of annotation and used the inter-annotator agreement as a widely used and accepted indicator. Additionally, we considered data of different text types and modalities for the evaluation.

3.1. Evaluation setup

For the evaluation corpus, we chose language samples reflecting the heterogeneity of the data in the SFB. It consisted partly of data elicited with QUIS, 42 question/answer pairs and 2 map task dialogues (comparable to the HCRC map task (Anderson et al., 1991)), and partly of newspaper commentaries from the Potsdam Commentary Corpus (PCC, (Stede, 2004)). Data elicited with QUIS is strongly controlled, with the majority of discourse referents denoting concrete objects. The map task dialogues also contain locative and directive NPs, references to the dialogue participants and their beliefs (each of them assuming her map as *common ground* of all discourse participants) and fragmentary utterances. Data from the PCC is syntactically more complex and semantically more vague; these texts make demands upon the reader in that they only comment on events that have been introduced elsewhere in the newspaper. The size of the data sample wrt. numbers of texts, tokens, nominal phrases and sentences is shown in Table 3.

Corp.	Text Type	Txt.	Tok.	NPs	Sent.
QuAn	question/answer	42	573	196	85
Dial	map task dialogue	2	478	99	71
PCC	commentary	5	889	220	115
total		49	1,940	515	271

Table 3: Data used in the evaluation

The annotators, two undergraduate students of linguistics (both native speakers of German), took part in a three-day test annotation. The students started with an intensive half-day training for annotation of both syntax and IS. In the actual test annotation, they first annotated syntactic constituent structure (constituents and their categorial labels). Then, the students annotated IS, based on a corrected gold standard of the syntax annotation. As an annotation tool, the EXMARaLDA Partitur Editor (Schmidt, 2004)[5] was used.

3.2. Method and Results

On the annotated data, we calculated κ (Cohen, 1960), (a) based on NPs for annotation layers on which predefined NPs are labeled (like information status and, to a large extent, topic) and (b) based on tokens for tasks that include defining extensions (like focus and adverbial framesetting topics). The results are shown in Table 4.

		core scheme		ext.
Feature	**TextType**	κ_{tok}	κ_{NP}	κ_{NP}
Information Status	QuAn	.77	.80	.73
	Dial	.80	.66	.61
	PCC	.68	.60	.55
Topic	QuAn	.75	.91*	-
	Dial	.51	.50*	-
	PCC	.44	.46*	-
Focus	QuAn	.51	.62*	-
	Dial	.44	.48*	-
	PCC	.19	.41*	-

Table 4: Kappa values based on tokens vs. NPs, for core and extended scheme; asterisked numbers represent a partial evaluation only, since topic and focus spans may include other material than NPs.

The results mirror that, in general, agreement is declining with increasing complexity of the annotation task. Information status is the layer with highest agreement across all text types (values for the extended tagset being only slightly, but constantly, lower – just as expected). Topic and focus layers result in a wider range of values, with high agreement[6] only on topics in question/answer data.

A qualitative data inspection shows that for information status there was some disagreement on the referentiality

[5] `http://www.exmaralda.org`

[6] Commonly, quality of annotation is considered high when κ (or similar measures) $> .8$ and 'allowing for tentative conclusions' when $.67 < \kappa < .8$ (Carletta, 1996). The interpretation of κ and similar measures, however, is a matter under continuing discussion, see e.g. (Carletta, 1996; Artstein and Poesio, 2005).

of elements (i.e. the decision whether to assign a label at all), especially of relative and reflexive pronouns (Dial and PCC), and (in PCC) expressions in metaphors and collocations that are either compositionally interpretable (see (15) and (16))[7] or object to anaphoric links (though not referential in the strict sense, as in example (17[8])).

(15) Dienstleister erwartet eventuell
service providers.ACC awaits possibly
[das Aus].
[the.NOM off].

'Service providers should anticipate the possibility of [demise].'

(16) Wer zu lange wartet, verliert bei der
Who too long waits loses at the
Euroumstellung [den Überblick].
Euro conversion [the overview].

'He who waits too long, will lose [track] of the Euro conversion.'

(17) Sie haben gelernt, [die "heißen Eisen"]
They have learnt [the hot irons]
möglichst gleich anzupacken. Olympia war
preferably immediately to tackle. Olympia was
[ein "heißes Eisen"], [das andere]...
one hot iron, the other one...

'They learnt to strike while [the irons] are hot. [One of the 'hot irons'] was the Olympic games, [another one] is...'

Concerning topic annotation, disagreement was found in Dial data mainly in clarification requests that involve presupposition cancellation (18). In PCC data, there were sentences with ambiguous topic analyses (two aboutness topic candidates in (19); in (20) either *the upgrade* is the aboutness topic, or *Radewege* is the aboutness topic and *the upgrade* is a framesetting topic).

(18) Vom Fahrrad gehen wir zum
From.DEF bike go we to.DEF
Schmetterling... - Moment mal, da ist [kein
butterfly... - wait a minute, there is [no bike]
Fahrrad] drauf.
on it.

'From the bike, we go to the butterfly... Wait a minute, there is [no bike] there.'

(19) [Bundesagrarministerin Renate Künast]
Federal Minister of Agriculture Renate Künast
will [das Halten der Tiere in engen
wants to the rearing the.GEN animals in narrow
Legebatterien] bereits vom Jahr 2006 an
battery cages already from.DEF year 2006
verbieten.
ban.

'[Renate Künast, Federal Minister of Agriculture,] wants to ban [the rearing of these animals in tiny battery cages] from as early as 2006.'

(20) Doch mit [der Nachrüstung] tut sich
But with [the upgrade] does REFL
[Radewege] schwer.
[Radewege] difficult.

'[The town of Radewege] has its difficulties with [an upgrade], though.'

On a closer look at the focus annotation, agreement is in fact substantial (the main difference being one annotator's tendency to define focus extensions to phrasal heads rather than whole phrases). Taking partial matches fully into account, we obtain f-scores of 67.42% (QuAn), 65.22% (Dial) and 80.49% (PCC)[9].
Generally, some disagreement emerged due to misinterpretation of the guidelines, which may be a result of the shortness of the training period.

4. Discussion

Despite the fair amount of work in the field, only a few studies are actually comparable to ours. Some focus on subtasks, e.g. assignment of information status to definite descriptions (Poesio and Vieira, 1998; Spenader, 2003), definite and demonstrative descriptions ((Salmon-Alt and Vieira, 2002) for French and Portuguese), pronouns (Navaretta, 2004; Hedberg et al., 2007), or named entities of type person (Nenkova et al., 2005). An overview of more closely related scheme evaluations is given in Table 5.
For *information status*, Nissim et al. (2004) report kappa values of .845 for a four category classification (κ=.788 for the finer-grained version of their scheme) of dialogue data, which indicates high quality of annotation. For one thing, annotators were provided with a very thorough training[10], probably more profound than ours. For another thing, they exclude (a) disfluencies, (b) locative and directional NPs (for obvious reasons not an option in our map task dialogues) and adverbial NPs, and (c) NPs that were tagged *not understood* by either annotator. Hempelmann et al. (2005) report κ=.72 for Prince's (1981) seven category distinction and κ=.74 for six categories (collapsing categories E and E_s after the annotation), which indicates a reasonable amount of agreement. From the results of these and our studies we conclude that reliable annotation of *information status* is feasible, with a few restrictions: the gradual character of referentiality (e.g. in metaphors, a problem in PCC data) and an inherently vague definition of accessibility relations (a general problem).
With *topic/focus*, the picture is more diverse. Not only do definitions vary (and languages under discussion), so do reported results: Komagata (2001) reports κ=.38 to .44 for a binary distinction of matrix subjects in Japanese translations, a result well above chance level, but doubtful with respect to implications. However, bearing in mind

[7] Abbreviations in glosses: ACC - accusative, DEF - definite article, GEN - genitive, NOM - nominative, REFL - reflexive pronoun.

[8] Quotes in the original. Brackets mark entities under discussion.

[9] In comparison, for exact matches only, we obtained 33.33% (QuAn), 23.19% (Dial) and 15.85% (PCC).

[10] The role of training is emphasized by the fact that after an intermediate discussion phase, κ rises by about .05.

	familiarity/information status related schemes		theme/topic related schemes		
Publication	Nissim et al. (2004)	Hempelmann et al. (2005)	Komagata (2001)	Veselá et al. (2004)	Paggio (2006)
IS category {**value set**}	*Information Status* {old,med,new,n.a.} excluded: disfluencies, locative, directional, and adverbial NPs, entities tagged *not-understood* by either annotator	*Givenness/Newness* (Prince, 1981) {E,E_S,U,I,I_C,BN_A,BN} $E_{(S)}$: (situationally) evoked, U: unused, $I_{(C)}$: (containing) inferable, $BN_{(A)}$: brand new (anchored)	*Theme/Rheme* {wa, ga} (particles) matrix subjects within theme: to be *wa*-marked in Japanese target sentence, within rheme: *ga*-marked	*Topic/Focus* (Sgall et al., 1986) {topic,contrast, focus} topic, contrast: context-bound, focus: non-bound	*Topic/Focus* {topic, focus}
Corpus	Switchboard dialogues	narratives from 4^{th} grade textbooks	*The Physician and Sportsmedicine* medical case reports	Prague Treebank	DanPASS
Language	English (US)	English (US)	Japanese (translations from English)	Czech	Danish
Evaluation setup	2 annotators 1,502 NPs	2 annotators 478 NPs	4 annotators 109 subjects 6 annotators appr. 60 subjects	3 annotators 6,402 nodes	2 annotators 4,402 + 8,562 tokens (2 sections)
Results	κ=.845, κ=.788 for finer-grained version of hierarchical scheme	82%;κ=.72, 88%;κ=.74 (E and E_s conflated a posteriori)	κ=.38 to .44	82.24%*; 86.42%* (conflating topic and contrast a posteriori) *average	κ=.7 to .8

Table 5: Evaluations of Annotation Schemes for Information Structure

that translation involves some interpretation already, this is not astonishing. Veselá et al. (2004) only report percent agreement (on nodes): 82.24% for a three way distinction, 86.42% for a two way distinction (average between 3 annotators) on Czech data. Paggio (2006) found κ=.7 to .8 for a binary distinction on Danish data.

5. Conclusions and Outlook

The results obtained by our test annotation are highly varied. Generally, agreement appears to decline with increasing complexity of the annotation task, reflecting related work for other languages. Across different types of text, QuAn data (question/answer pairs) were annotated more consistently than Dial (map task dialogues) and PCC (newspaper commentaries). Especially topic annotation varied considerably depending on the text type.

Regarding different dimensions of IS, results for *information status* were acceptable, agreement of *topic* and *focus* annotation was low. In accordance to (Paggio, 2006), we found that the definition of where foci start will need improvement.

Means of improvement are training (Nissim et al., 2004; Veselá et al., 2004) – in our three-day evaluation the annotators certainly did not have much time for absorbing and discussing the guidelines –, which is limited by cost, and further specification of the guidelines, which is limited by LISA's objectives of language independence and openness towards different theories.

In the future development of the guidelines we will focus on i) enriching the guidelines with text-type-specific instructions, ii) the explicit encoding of subjective knowledge, in particular for inferable entities and entities accessible via world knowledge, and iii) the encoding of subjective interpretations. The latter is proposed, e.g., by (Reitter and Stede, 2003) for the annotation of discourse structure, which —like sentences— can often be assigned more than one interpretation. In this vein, an annotation encodes one possible interpretation, and strategies have to be developed for classifying and dealing with competing annotations: disagreement e.g. in IS annotation might thus result either from annotation errors or from differences in interpretation.

LISA has been applied to data from diverse languages by experts of linguistics with a native knowledge of the respective language. Thus, we will further validate LISA across languages.

6. References

Anne H. Anderson, Miles Bader, Ellen G. Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Carrod, Stephen Isard, Jacqlene Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry Thompson, and Regina Weinert. 1991. The HCRC Map Task Corpus. *Language and Speech*, 34:351–366.

Ron Artstein and Massimo Poesio. 2005. Kappa^3 = Alpha (or Beta). CSM 437, Department of Computer Science, University of Essex, UK.

Jean Carletta. 1996. Assessing Agreement on Classification Tasks: The Kappa Statistic. *Computational Linguistics*, 22(2):249–254.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.

Stefanie Dipper, Michael Götze, and Stavros Skopeteas, editors. 2007. *Information Structure in Cross-Linguistic Corpora: Annotation Guidelines for Phonology, Morphology, Syntax, Semantics, and Information Structure*, volume 7 of *Interdisciplinary Studies on Information Structure (ISIS)*. Universitätsverlag Potsdam, Potsdam, Germany.

Jeanette K. Gundel, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):279–307.

Eva Hajičová, Jarmila Panevová, Petr Sgall, Alena Böhmová, Markéta Ceplová, and Veronika Řezníčková. 2000. A Manual for Tectogrammatical Tagging of the Prague Dependency Treebank. TR 2000-09, ÚFAL/CKL, Prague, Czech Republic.

Nancy Hedberg, Jeanette K. Gundel, and Ron Zacharski. 2007. Directly and Indirectly Anaphoric Demonstrative

and Personal Pronouns in Newspaper Articles. In *Proceedings of DAARC 2007 (the Sixth Discourse Anaphora and Anaphora Resolution Colloquium)*, pages 31–36, Lagos, Portugal.

Christian F. Hempelmann, David Dufty, Philip M. McCarthy, Arthur C. Graesser, Zhiqiang Cai, and Danielle S. McNamara. 2005. Using LSA to automatically identify givenness and newness of noun-phrases in written discourse. In B. Bara, editor, *Proceedings of the 27th Annual Meetings of the Cognitive Science Society*, pages 941–949, Stresa, Italy. Mahwah, NJ: Erlbaum.

Nobo Komagata, 2001. *An Evaluation Method for Identifying Information Structure*. Department of Computer and Information Science, University of Pennsylvania. Manuscript.

Costanza Navarretta. 2004. An algorithm for resolving individual and abstract anaphora in danish texts and dialogues. In *Proceedings of the Workshop on Reference Resolution and its Applications: ACL 2004*, pages 95–102, Barcelona, Spain. Association for Computational Linguistics.

Ani Nenkova, Advaith Siddharthan, and Kathleen McKeown. 2005. Automatically learning cognitive status for multi-document summarization of newswire. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 241–248, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Malvina Nissim, Shipra Dingare, Jean Carletta, and Mark Steedman. 2004. An Annotation Scheme for Information Status in Dialogue. In *Proceedings of the 4th Language Resources and Evaluation Conference (LREC 2004)*, pages 1023–1026, Lisbon, Portugal.

Patrizia Paggio. 2006. Annotating Information Structure in a Corpus of Spoken Danish. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, pages 1606–1609, Genova, Italy.

Massimo Poesio and Renata Vieira. 1998. A Corpus-based Investigation of Definite Description Use. *Computational Linguistics*, 24(2):183–216.

Livia Polanyi, Martin van den Berg, and David Ahn. 2003. Discourse Structure and Sentential Information Structure. An Initial Proposal. *Journal of Logic, Language and Information*, 12(3):337–350.

Ellen F. Prince. 1981. Toward a Taxonomy of Given-New Information. *Radical Pragmatics*, pages 223–55.

David Reitter and Manfred Stede. 2003. Step by step: underspecified markup in incremental rhetorical analysis. In *Proc. of the 4th International Workshop on Linguistically Interpreted Corpora (LINC-03)*, pages 77–84, Budapest, Hungary.

Arndt Riester, Lorena Killmann, David Lorenz, and Melanie Portz, 2008. *Richtlinien zur Annotation von Gegebenheit und Kontrast in Projekt A1 SFB 732 (draft version as of March 2008)*. University of Stuttgart, Germany. Manuscript.

Arndt Riester. 2008. A semantic explication of *Information Status* and the underspecification of the recipients' knowledge. In A. Gronn, editor, *Proceedings of Sinn und Bedeutung 12*, Oslo.

Susanne Salmon-Alt and Renata Vieira. 2002. Nominal Expressions in Multilingual Corpora. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, pages 1627–1634, Las Palmas, Canary Islands, Spain.

Thomas Schmidt. 2004. EXMARaLDA - ein System zur computergestützten Diskurstranskription. In Alexander Mehler and Henning Lobin, editors, *Automatische Textanalyse - Systeme und Methoden zur Annotation und Analyse natürlichsprachlicher Texte*, pages 203–218. Wiesbaden: VS Verlag für Sozialwissenschaften, Germany.

Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Reidel, Dordrecht.

Stavros Skopeteas, Ines Fiedler, Sam Hellmuth, Anne Schwarz, Ruben Stoel, Gisbert Fanselow, and Manfred Krifka. 2006. Questionnaire on Information Structure: Reference Manual. Interdisciplinary Studies on Information Structure (ISIS), Vol. 4, Universitätsverlag Potsdam, Potsdam, Germany.

Jennifer Spenader. 2003. Between binding and accomodation. In Peter Kuehnlein, Hannes Rieser, and Henk Zeevat, editors, *Perspectives on Dialogue in the New Millennium*, pages 79–110. John Benjamins.

Manfred Stede. 2004. The Potsdam Commentary Corpus. In Bonnie Webber and Donna K. Byron, editors, *Association for Computational Linguistics (ACL) 2004 Workshop on Discourse Annotation*, pages 96–102, Barcelona, Spain.

Michael Strube. 1998. Never Look Back: An Alternative to Centering. In *Proceedings of COLING-ACL '98*, pages 1251–1257, Montréal, Québec, Canada.

Kateřina Veselá, Jirı Havelka, and Eva Hajičová. 2004. Annotators' agreement: The Case of Topic-Focus Articulation. In *Proceedings of the 4th Language Resources and Evaluation Conference (LREC 2004)*, pages 2191–2194, Lisbon, Portugal.