A Modular Account of Information Structure in Extensible Dependency Grammar

Ralph Debusmann[1], Oana Postolache[2], and Maarika Traat[3]

[1] Programming Systems Lab, Saarland University, Saarbrücken, Germany
rade@ps.uni-sb.de
[2] Computational Linguistics, Saarland University, Saarbrücken, Germany
Al. I. Cuza University, Iaşi, Romania
oana@coli.uni-sb.de
[3] Informatics, University of Edinburgh, Scotland
M.Traat@sms.ed.ac.uk

Abstract. We introduce a modular, dependency-based formalization of Information Structure (IS) based on Steedman's prosodic account [1, 2]. We state it in terms of Extensible Dependency Grammar (XDG) [3], introducing two new dimensions modeling 1) prosodic structure, and 2) theme/rheme and focus/background partitionings. The approach goes without a non-standard syntactic notion of constituency and can be straightforwardly extended to model interactions between IS and other dimensions such as word order.

1 Introduction

Information Structure (IS) is the way in which people organize their utterances. Usually, in an utterance there is a part that links the content to the context, and another that advances the discourse by adding or modifying some information. IS is an important factor in determining the felicity of an utterance in a given context. Among the many applications where IS is of crucial importance are content-to-speech systems (CTS), where IS helps to improve the quality of the speech output [4], and machine translation (MT), where IS improves target word order, especially that of free word order languages [5].

In this paper we present a modular, dependency-based account of IS based on Steedman's prosodic account of IS for Combinatory Categorial Grammar (CCG) [1, 2]. Similarly to Steedman, we establish a bi-directional correspondence between IS and prosodic structure, i.e. when the IS is known, we can determine the prosodic structure (e.g. in CTS systems), and when we have the prosodic information, we can extract the IS (e.g. to augment dialog transcripts).

We state our approach in terms of Extensible Dependency Grammar (XDG) [3], which allows us to take a modular perspective on IS. We distinguish three notions of constituency: syntactic, prosodic, and information structural, which are related, but not identical. Thus, differently from Steedman, we can decouple

A. Gelbukh (Ed.): CICLing 2005, LNCS 3406, pp. 25–36, 2005.

syntax from information structure, and do not assume non-standard syntactic constituents. By this, we can monotonically add IS to existing XDG grammars. Moreover, our technique is prepared to straightforwardly state constraints on the interplay of IS, prosody and word order, as required for free word order languages such as Czech. This would bring XDG closer to Functional Generative Description (FGD) [6], Kruijff's Dependency Grammar Logic (DGL) [7], Kruijff's and Baldridge's generalized CCG approach [8], and Kruijff's and Duchier's approach using Topological Dependency Grammar (TDG) [9]. The latter account, although stated in a similar framework, is quite different from ours: it concentrates less on modularity, and more on the interaction of different aspects (prosody, word order etc.) in the realization of IS.

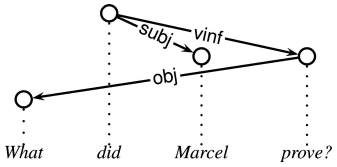
The paper is organized into five sections. Section 2 gives an overview of the XDG grammar formalism. Section 3 is a short introduction to the existing IS approaches; we concentrate on Steedman's prosodic account and his two levels of IS: theme/rheme and focus/background. In section 4, we integrate IS into XDG, introducing two new dimensions: one to model the prosodic structure of the sentence and one to describe the theme/rheme and focus/background distinctions. In section 5, we conclude and outline avenues for future research.

2 Extensible Dependency Grammar

In this section we introduce Extensible Dependency Grammar (XDG) [3]. XDG is a grammar formalism based on dependency grammar [10] and a generalization of Topological Dependency Grammar (TDG) [11]. XDG is all about *modularity*, striving to transplant ideas from software engineering into the context of grammar formalisms. Modularity ensures both *re-usability* and *compositionality*: XDG grammars are consequently composed from layers of simple, re-usable modules. This yields new possibilities for grammar engineering and cross-linguistic modeling.

2.1 Dependency Grammar

Dependency grammar models the syntax of a natural language in terms of relations between words, which correspond 1:1 to nodes. Dependency relations involve *heads* and *dependents*. For example, in the dependency analysis displayed below, in (1), the finite verb *did* is the head of the dependent *Marcel* and *prove* is the head of *what*. Dependency relations are often further specified: in (1), *Marcel* is a subj-dependent of *did*, i.e. the subject, and *what* is the object of *prove*.



(1)

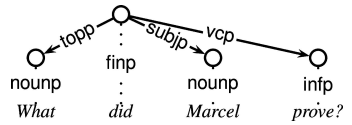
2.2 Multiple Dimensions

Dependency grammar was originally concerned with surface syntax only [10]. However, nothing stops the general concept of dependency grammar—stating relations between words—from being transferred to other linguistic areas, including morphology, deep syntax and semantics. In this generalized sense, pioneered by the Prague and Russian Schools [6, 12], a dependency analysis consists of *multiple* dependency graphs, one for each linguistic dimension. XDG also adopts this idea.

The components of a multi-dimensional dependency analysis are not independent. For instance, semantic arguments can only be realized by appropriate syntactic functions (e.g. agents by subjects). [6, 12] use functional mappings between architecturally adjacent dimensions (e.g. surface and deep syntax). XDG goes beyond that: each dimension can be made to interact with *any* other by bi-directional, relational constraints.

2.3 Word Order

XDG allow splitting up surface syntax into the dimensions *Immediate Dominance* (ID) and *Linear Precedence* (LP). This is essential for the successful treatment of complex word order phenomena in [11]. The ID dimension is solely devoted to syntactic function: with word order factored out, an ID analysis is an *unordered* tree as in (1) above. Word order is taken care of in the LP dimension. LP analyses are *ordered* trees, flatter than the corresponding ID trees. We display an example LP analysis in (2) below:



(2)

Here the finite verb *did* is the head of three dependents: *what*, *Marcel* and *prove*. *What* is a topp-dependent, i.e. it is in topicalized position. Similarly, *Marcel* is in subject position, and *prove* in verbal complement position. In the LP analysis each node carries a node label. This is used to order heads with respect to their dependents. In (2), *did* has node label finp. A well-formed LP analysis must be correctly ordered according to a global order on the set of labels, e.g.:

$$\mathsf{topp} \prec \mathsf{finp} \prec \mathsf{subjp} \prec \mathsf{vcp} \tag{3}$$

Here, we state that topicalized words must precede finite verbs, subjects and verbal complements.

2.4 Semantics

XDG allows us to go far beyond surface syntax. In [3], the authors introduce the dimensions of *Predicate-Argument structure* and *Scope structure* to represent semantics. Because of the relational nature of XDG, this syntax-semantics

interface is automatically bi-directional: syntax can disambiguate semantics and vice-versa, e.g. semantic attachment preferences can resolve modifier attachments in syntax. However, as semantics does not concern us in this paper, we omit further mention of it.

2.5 Principles

The well-formedness conditions of an XDG analysis are specified by *principles* from an extensible *principle library*. Each principle has a declarative semantics and can be parametrized. The *tree principle*, for example, constrains an analysis to be a tree and is parametrized by dimension. Thus, the same principle can be used to constrain the analyses on the ID and LP dimensions to be trees. The *valency principle*, in addition, is *lexicalized*, and constrains the incoming and outgoing edges of each node. On ID, for instance, a finite verb such as *did* requires a subject (outgoing edges), and only nouns *can* be subjects (incoming edges).

The principles so far were *one-dimensional principles*, constraining only one dimension. To constrain the relation between multiple dimensions, XDG offers two means: 1) the lexicon, and 2) multi-dimensional principles. Firstly, the lexicon assigns to each word a set of lexical entries simultaneously constraining all dimensions. Secondly, the principle library includes *multi-dimensional principles*, parametrized by multiple dimensions, which directly constrain their relation.

2.6 Lexicon

XDG grammars rely heavily on the lexicon. To ease the creation of the lexicon and the statement of linguistic generalizations, XDG provides facilities in the spirit of Candito's metagrammar [13], extended in [14]. Basically, the XDG *metagrammar* is an abstract language to describe the lexicon, which is automatically compiled out to yield the lexicon itself.

2.7 Parsing and Generation

XDG parsing and generation is done by the constraint-based XDG solver. Given that XDG solving is NP-complete in the worst case, handcrafted grammars have yielded good performance in the average case. This makes XDG already interesting for the exploration of new linguistic theories such as the one presented here. A comprehensive grammar development toolkit including the XDG solver is freely available and easy to install and use [15]. So far, the XDG solver cannot yet parse induced grammars (e.g. from the Prague Dependency Treebank) competitively [16], but research is underway to improve its performance.

3 Information Structure

In this section, we introduce the concept of Information Structure (IS), illustrate it with examples, and briefly touch upon selected issues related to it. We devote

specific attention to Steedman's [1, 2] prosodic account of information structure, which we have chosen as the basis for our realization of IS in XDG.

3.1 Information Structure Basics

By Information Structure (IS) we mean the way people organize the content they want to communicate in an utterance. There are usually several ways for the same propositional content to be presented. An alternative name for the same concept is Information Packaging, introduced by Chafe [17]. He illustrated its meaning as follows:

> [The phenomena at issue here] have to do primarily with how the message is sent and only secondarily with the message itself, just as the packaging of toothpaste can affect sales in partial independence of the quality of the toothpaste inside.

IS is typically realized by a combination of various means, depending on the typology of the language. In languages with relatively fixed word order, such as English, prosody is often a prominent factor. Free word order languages are more likely to realize IS by word order variation, whereas other languages, such as Japanese, realize IS by morphology (e.g. the special topic marker *-wa*).

Different names have been used for the sub-divisions in IS: topic and focus, theme and rheme, ground and focus, relatum and attributum, to name just a few. What all these divisions have in common, with minor differences, is that they distinguish a part of an utterance that links it to the previous discourse, and another part that is a novel contribution. For a more extensive overview of different approaches to IS, see [18] and [7]. We use the terms *theme* and *rheme* as introduced by the Prague circle of linguists (note that our use of these terms differs from the use by Halliday [19]). Theme is the part that relates the utterance to the previous discourse, while rheme adds or modifies some information about the theme.

As hinted at in Chafe's definition, IS does not affect the propositional content of an utterance. What it does affect, is the contextual felicity of the utterance. For example, while (4)a is a suitable answer for the question in (4), (4)b is not acceptable in the given context:

What did Marcel prove?
a. $[\text{Marcel proved}]_{th}$ $[\text{COMPLETENESS.}]_{rh}$ (4)
b. $*$ $[\text{MARCEL}]_{rh}$ $[\text{proved completeness.}]_{th}$

The words in small capitals in (4) carry the main prosodic accent of the sentence. Assuming that this accent marks the rheme, we can see why only (4)a, but not (4)b is an appropriate answer to the question: *completeness* is the new information asked for, not *Marcel*.

3.2 Prosodic Account of Information Structure

Steedman [1, 2] divides IS into theme and rheme. In his approach, the IS division follows prosodic phrasing. Both theme and rheme can be further divided into

background and focus. The focused material in the theme and rheme are the words that carry pitch accents, while the unaccented words are the background. The most common kind of theme is the so-called "un-marked" theme, where no words carry a pitch accent. Marked themes, as in (5), are used when one item stands in explicit contrast with another from the previous discourse.

I know that Marcel likes the man who wrote the musical.
But who does he ADMIRE?

Marcel	ADMIRES	the woman who	DIRECTED	the musical.
background	*focus*	*background*	*focus*	*background*
theme		*rheme*		

(5)

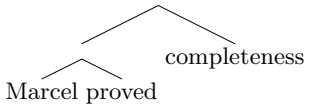
Steedman claims that there is a specific set of pitch accents in English that can accompany the theme, and another that can accompany the rheme, the most common theme pitch accent being L+H* and the most common rheme pitch accent being H*.[1]

Boundary tones delimit prosodic phrases. There are various boundary tones, the most frequently occurring being a low boundary — LL% — and a rising boundary — LH%. There is a tendency for LH% to occur at the end of an intonational phrase containing the theme pitch accent L+H*, and for LL% to occur after the rheme pitch accent H*.

According to the prosodic phrasing, Combinatory Categorial Grammar (CCG) [1] provides different parses for the same string of words, giving rise to different interpretations with respect to the information structure:

I know what Mark proved, but what did Marcel prove?

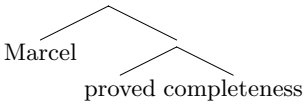
MARCEL proved		COMPLETENESS.	
L+H*	LH%	H*	LL%



(6)

I know what Mark did, but what did Marcel do?

MARCEL		proved	COMPLETENESS.
L+H*	LH%	H*	LL%



(7)

While pitch accents are seen as properties of the words that carry them, boundary tones are seen as individual lexical entries and independent phrasal constituents.

[1] The intonational notation used is due to Pierrehumbert [20]. According to her, intonational phrases are made up of the following components: pitch accent(s), phrasal tone and boundary tone. In Steedman's [1], [2] representation the last two have been joined together under the name '*boundary tone*'. L stands for low pitch, and H for high pitch.

4 Adding Information Structure to XDG

In this section, we present a way of modeling information structure within the XDG formalism. We follow Steedman's approach [1, 2], sketched in section 3.2, which we adapt to XDG by introducing two new dimensions: *Prosodic Structure* (PS) and *Information Structure* (IS). While Steedman views only pitch accents as properties of words, and treats boundary tones as separate lexical items, we treat both pitch accents and boundary tones as properties of words.

4.1 PS Dimension

An analysis on the PS dimension is a tree whose shape is determined by edges representing boundary tones and pitch accents. The root of the tree corresponds to the punctuation mark at the end of the sentence. The daughters of the root are the words carrying boundary tones. Thus, the outgoing edges of the root may be labeled with LL% (low boundary tone), LH% (high boundary tone), H*_LL% (falling pitch accent and low boundary tone) and L+H*_LH% (rising pitch accent and high boundary tone). For simplicity, we consider only the two most frequent types of boundary tones (LL% and LH%), the two most frequent types of pitch accents (H* and L+H*) and their combinations.

Boundary tones delimit non-overlapping, contiguous prosodic constituents. Each word that has a boundary tone attached to it is the head of a prosodic constituent and has the node label b (for *boundary*). To its left it can have accented (may be labeled with H* or L+H*) or non-accented daughters (na), both having the node label nb (for *non-boundary*).

We constrain the PS dimension by the following one-dimensional principles: 1) tree, 2) valency, and 3) order. The tree principle constrains PS analyses to be trees, and the valency principle lexically restricts the incoming and outgoing edges of each node. The order principle serves three purposes: a) it restricts the node labels of each node, b) it requires PS constituents to be projective, i.e. non-overlapping, and c) the order of the daughters of each node must be compatible with the following global order, stating that boundary tones follow everything else:[2]

$$\{\mathsf{L{+}H^*},\ \mathsf{H^*},\ \mathsf{LH\%},\ \mathsf{LL\%},\ \mathsf{L{+}H^*\_LH\%},\ \mathsf{H^*\_LL\%},\ \mathsf{na},\ \mathsf{nb}\} \prec \{\mathsf{b}\} \tag{8}$$

The restrictions on the incoming and outgoing edges (valency: in and out features) and on the node labels (order: on feature) of each node are stipulated in the lexicon. As an example, we show the lexical class *pa_bo* for words carrying both an H* pitch accent and an LL% boundary tone:

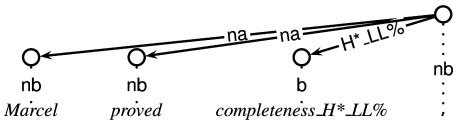
$$pa\_bo ::= \left[PS : \left[\begin{array}{r} \mathsf{in} : \{\mathsf{H^*\_LL\%?}\} \\ \mathsf{out} : \{\mathsf{H^*}*, \mathsf{na}*\} \\ \mathsf{on} : \{\mathsf{b}\} \end{array} \right] \right] \tag{9}$$

[2] This global order actually orders sets of labels instead of just labels, contrary to the total order given in (3) above. The order of labels within these sets is unrestricted.

Words inheriting from this class can only have an incoming edge H*_LL%, and can have an arbitrary number of outgoing edges to accented words, labeled H*, or non-accented ones, labeled na. Their node label is b.

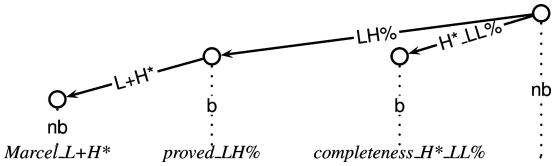
For illustration, we display some example PS trees below, corresponding respectively to (4), (6) and (7) from section 3:

What did Marcel prove?
Marcel proved COMPLETENESS.
H* LL%



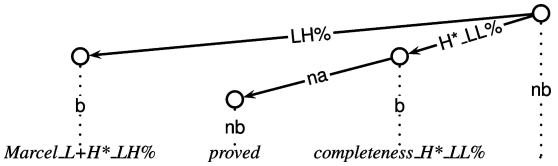
(10)

I know what Mark proved, but what did Marcel prove?
MARCEL proved COMPLETENESS.
L+H* LH% H* LL%



(11)

I know what Mark did, but what did Marcel do?
MARCEL proved COMPLETENESS.
L+H* LH% H* LL%



(12)

4.2 IS Dimension

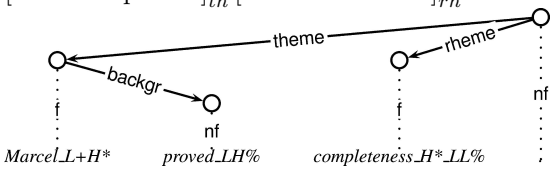
Using XDG's modular methodology, we can simplify our account by first specifying the IS dimension independently from prosodic structure (PS dimension). Only later we will regulate the interplay of the two using the lexicon, and a multi-dimensional principle.

An IS analysis is again a tree whose root corresponds to the punctuation mark. The daughters of the root are the words carrying a pitch accent (instead of those carrying a boundary tone as in the PS), which we call *foci* following Steedman. Their incoming edge label is either theme or rheme, and their node label is f (for *focus*). We require each sentence to have at least one rheme (but cf. *all theme utterances* in [1]), while the theme is optional.

Foci are the heads of non-overlapping, contiguous information structural constituents (i.e. themes or rhemes). Their daughters are the words constituting the

background (edge label backgr). These have node label nf (for *non-focus*). Contrary to boundary tones on the PS, which have to be positioned rightmost in PS constituents, the position of foci within IS constituents is unconstrained. Here is an example IS analysis (cf. (11) in section 4.1):

I know what Mark proved, but what did Marcel prove?
[MARCEL proved]$_{th}$ [COMPLETENESS.]$_{rh}$



(13)

We constrain the IS dimension by re-using the following one-dimensional principles: 1) tree, 2) valency, and 3) order. In the IS dimension, the purpose of the order principle is twofold: a) it restricts the node labels of each node, and b) it requires IS constituents to be non-overlapping. It does not, however, prescribe an order on the set of labels.

As an example, we show the lexical class *rf* for the foci of rhemes:

$$rf ::= \left[IS: \left[\begin{array}{r} \mathsf{in}: \{\mathsf{rheme?}\} \\ \mathsf{out}: \{\mathsf{backgr}*\} \\ \mathsf{on}: \{\mathsf{f}\} \end{array} \right] \right] \tag{14}$$

Words which inherit from this class have the node label f. They can only have an incoming edge labeled rheme, whilst the number of outgoing edges, which are labeled backgr, is arbitrary.

The dimensions of IS and PS are certainly not independent. We constrain their relationship by two means: 1) the lexicon, and 2) a multi-dimensional principle. Firstly, we constrain the lexicon such that nodes with incoming edges L+H* and L+H*_LH% in the PS must have the incoming edge theme in the IS, and those with incoming edges H* and H*_LL% in the PS must have the incoming edge rheme in the IS. Secondly, we use a multi-dimensional principle called the *island principle*, which states that IS constituents must always either coincide with a corresponding PS constituent, or be subparts of it. In other words, IS constituents cannot cross the prosodic constituent boundaries. This principle generalizes over the two cases of marked themes (where the IS constituents coincide with the PS constituents), and unmarked themes (where the IS constituents are subparts of the PS constituents).

As an example for the lexicon constraining the relation between PS and IS, we show the lexical class *rheme_pa_bo*, resulting from the combination of the classes *pa_bo* ((9) above) and *rf* ((14) above):

$$rheme\_pa\_bo = \left[\begin{array}{ll} PS: & \left[\begin{array}{r} \mathsf{in}: \{\mathsf{H*\_LL\%?}\} \\ \mathsf{out}: \{\mathsf{H}*, \mathsf{na}*\} \\ \mathsf{on}: \{\mathsf{b}\} \end{array} \right] \\ IS: & \left[\begin{array}{r} \mathsf{in}: \{\mathsf{rheme?}\} \\ \mathsf{out}: \{\mathsf{backgr}*\} \\ \mathsf{on}: \{\mathsf{f}\} \end{array} \right] \end{array} \right] \tag{15}$$

Words inheriting from this class have the pitch accent H* and the boundary tone LL% (incoming edge label H*_LL%) on the PS. Since these tones accompany only rhemes, they must consequently have incoming edge label rheme in the IS.

So far, we have not dealt with the issue of unmarked themes, which contain no pitch accents and consequently no foci. Here, the IS can be ambiguous, while the PS is unambiguous. Consider (16) which could be an answer to any of the questions (17), (18) and (19):

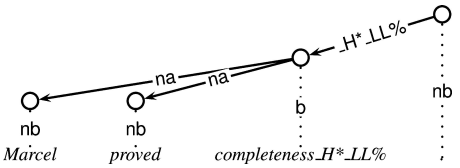
Marcel proved completeness_H*_LL%. (16)

What did Marcel prove? (17)

What did Marcel do? (18)

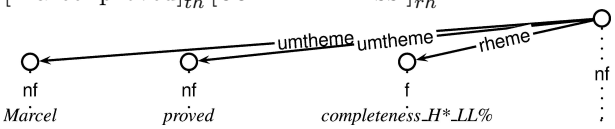
What's new? (19)

The PS of (16), displayed in (20), is unambiguous. Given this PS, however, the IS is ambiguous. The three alternative analyses (21), (22) and (23) correspond respectively to questions (17), (18) and (19). In these analyses, we make each word in the unmarked theme form a singleton IS constituent (including only itself), and having the incoming edge label umtheme (for *un-marked theme*).



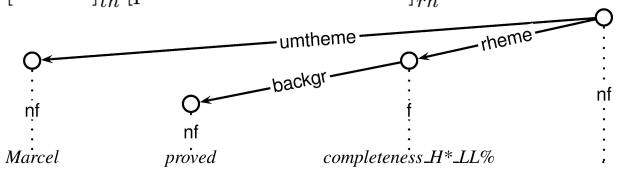
(20)

What did Marcel prove?
[Marcel proved]$_{th}$ [COMPLETENESS.]$_{rh}$



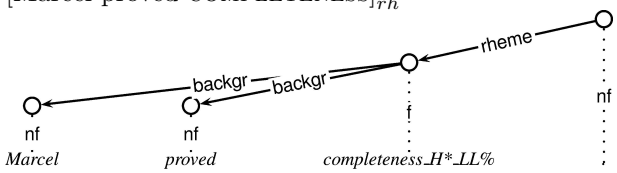
(21)

What did Marcel do?
[Marcel]$_{th}$ [proved COMPLETENESS.]$_{rh}$



(22)

What's new?
[Marcel proved COMPLETENESS]$_{rh}$



(23)

5 Conclusions

We presented a new, modular and dependency-based formalization of IS couched in the framework of Extensible Dependency Grammar (XDG) [3]. As a starting point, we chose Steedman's prosodic account of IS [1, 2]. Our reformulation of his ideas in XDG resulted in a different perspective on the interplay of IS and syntax, decoupling the two to a much higher degree. Thus, we did not introduce non-standard syntactic constituents and could add IS monotonically to any existing XDG grammar.

The approach presented here is not just theoretical: we have already implemented an English grammar using the XDG Development Kit (XDK) [15], which reflects precisely the account given in this paper.

The most interesting avenue for future research is the interplay of IS with other linguistic areas such as word order. In English, IS is mostly realized by prosody, but the picture changes for free word order languages such as Czech, where word order is another prominent factor [7]. Our XDG-based account is perfectly prepared to accommodate such interactions. It allows for the straightforward statement of constraints that relate the IS dimension to the dimension of word order, for example, that topicalized material must be the theme (e.g. in certain dialects of English).

Acknowledgements

We thank Ciprian Gerstenberger and Stefan Thater for their valuable contribution during the IGK summer school project in which this new approach has been developed. We also thank Timothy Jones for his helpful advice when proofreading the pre-final versions of the paper.

References

1. Steedman, M.: The Syntactic Process. MIT Press (2000)
2. Steedman, M.: Information Structure and the Syntax-Phonology Interface. Linguistic Inquiry **31** (2000) 649–689
3. Debusmann, R., Duchier, D., Koller, A., Kuhlmann, M., Smolka, G., Thater, S.: A Relational Syntax-Semantics Interface Based on Dependency Grammar. In: Proceedings of COLING 2004, Geneva/SUI (2004)
4. Prevost, S., Steedman, M.: Information Based Intonation Synthesis. In: Proceedings of the ARPA Workshop on Human Language Technology, Princeton/USA (1994)
5. Stys, M., Zemke, S.: Incorporating Discourse Aspects in English-Polish MT: Towards Robust Implementation. In: Recent Advances in NLP, Velingrad/BUL (1995)
6. Sgall, P., Hajicova, E., Panevova, J.: The Meaning of the Sentence in its Semantic and Pragmatic Aspects. D. Reidel, Dordrecht/NL (1986)
7. Kruijff, G.J.M.: A Categorial-Modal Architecture of Informativity. PhD thesis, Charles University, Prague/CZ (2001)

8. Kruijff, G.J.M., Baldridge, J.: Generalizing Dimensionality in Combinatory Categorial Grammar. In: Proceedings of COLING 2004, Geneva/SUI (2004)
9. Kruijff, G.J.M., Duchier, D.: Information Structure in Topological Dependency Grammar. In: Proceedings of EACL 2003, Budapest/HUN (2003)
10. Tesnière, L.: Eléments de Syntaxe Structurale. Klincksiek, Paris/FRA (1959)
11. Duchier, D., Debusmann, R.: Topological Dependency Trees: A Constraint-based Account of Linear Precedence. In: Proceedings of ACL 2001, Toulouse/FRA (2001)
12. Mel'čuk, I.: Dependency Syntax: Theory and Practice. State Univ. Press of New York, Albany/USA (1988)
13. Candito, M.H.: A Principle-based Hierarchical Representation of LTAG. In: Proceedings of COLING 1996, Kopenhagen/DK (1996)
14. Crabbé, B., Duchier, D.: Metagrammar Redux. In: Proceedings of the International Workshop on Constraint Solving and Language Processing, Roskilde/DK (2004)
15. Debusmann, R., Duchier, D.: XDG Development Kit (2004) http://www.mozart-oz.org/mogul/info/debusmann/xdk.html.
16. Bojar, O.: Problems of Inducing Large Coverage Constraint-Based Dependency Grammar. In: Proceedings of the International Workshop on Constraint Solving and Language Processing, Roskilde/DK (2004)
17. Chafe, W.L.: Givenness, Contrastiveness, Definiteness, Subjects, Topic, and Point of View. In Li, C., ed.: Subject and Topic. Academic Press, New York/USA (1976) 25–55
18. Vallduví, E., Engdahl, E.: The Linguistic Realisation of Information Packaging. Journal of Linguistics **34** (1996) 459–519
19. Halliday, M.A.K.: Notes on Transitivity and Theme in English, Part II. Journal of Linguistics **3** (1967) 199–244
20. Pierrehumbert, J.: The Phonetics and Phonology of English Intonation. PhD thesis, Massachusetts Institute of Technology, Bloomington/USA (1980)