

Towards a Dynamic Model of Topic-Marking*

Tim Florian Jaeger and David Y. Oshima
Stanford University

ABSTRACT – This paper addresses the gap between the research on D[iscourse] S[tructure] T[opics] (cf. Chafe 2001, Fraser 1988) and I[nformation] S[tructure] T[opics] (cf. Jacobs 2001, Büring 1999, Lambrecht 1994, Vallduvi 1992). This paper takes a uniform approach to topics where ISTs and DSTs are treated as ontologically identical. We propose a hierarchical model of T(topic)-trees, which are constructed dynamically in the course of discourse. The nodes of a T-tree consist of sets of semantic objects (e.g. entities, propositions, predicates, etc.; contrasting with Dik's 1989 D-topics). The higher a T-tree node is in the tree, the bigger is the linguistic form unit that it corresponds to (e.g. the whole text, a paragraph, a sentence). The highest topic in a T-tree is always the global DST, the lowest topics are the ISTs.

Based on T-trees, we show how the (topic-)coherence of texts could be modelled. The T-trees are motivated by two case studies, one on the 'as for' construction in English and one on multiple topic constructions in Japanese. The proposed model allows us to formalize the constraints which those constructions put on the context in which they appear. The remainder of the paper, discusses the interaction of the proposed model with the notion of focus and contrast (cf. Vallduví and Vilkuña 1998) making reference to such various constructions as left-dislocation, it-clefts, and topicalization in English, as well as, Multiple Nominative Constructions in Japanese.

1 Introduction

In recent years several attempts have been made to formalize the discourse-related notions such as topic and focus (Jacobs 2001, Büring *to appear*, Portner and Yabushita 1998, Roberts 1996, Engdahl and Vallduví 1996, Vallduví 1992). In this paper, we will show that, drawing on data from English and Japanese, a proper model of topic-marking should be able to capture its global and dynamic nature, and we propose a tentative model of topic marking that satisfies this desideratum. The organization of the paper is as follows. In section 2, we will set up basic terminology and ontology. In section 3, we will sketch out a hierarchical/dynamic model of topic-marking, which we will call “T(topic)-tree model”, and demonstrate empirical data that motivates it. In section 4, we discuss bearing of our model on the theory of focus and information packaging in general; in particular, we argue that the concept of topic developed in this paper subsumes some subtypes of (what is called) contrastive focus (the amount of text does not stand in relation to the other sections). Finally, section 5 contains our conclusion.

Parts of our discussion will remain at a highly conceptual level, partly due to the shortness of this paper. We will however, provide data taken from corpora to provide a foundational basis for our most important claims (cf. section 3.2 and 3.3).

2 Where we come from: Terminological Clarifications

Many of the terms used in this paper such as, among others, *topic*, *focus*, *information structure*, *discourse structure* and *information packaging* have received a plethora of different definitions in the literature. We therefore think that it is necessary to make clear where we come from and what we subscribe to. We use the term *information packaging* to refer to the group of phenomena that figure in the structuring of information in the course of communication in a very general sense. By information we mean the propositions expressed in a text, as well as the conversational implicatures, and presuppositions that go along with those propositions. Our use of the term information packaging implies that we do not restrict the study of such phenomena to the sentence or clause level (cf. Vallduvi

* The authors would like to thank Elizabeth Traugott for valuable comments and discussion. Also, we thank Lev Blumenfeld and Andrew Koontz-Garboden for being guinea pigs for some parts of the argumentation in this paper. We are solely responsible for the remaining errors.

1992 and in some sense Lambrecht 1994) but also include structuring of information above that level also known as discourse structure.

In the last 40 years there has been great variety of approaches to the notion of what has been called *topic*, *link*, or *subject* from different schools of linguistics. Most current theories of information structure go back to the foundational works in the 1970s and 1980s, such as the Prague School of the Functional Sentence Perspective (cf. Firbas 1971, 1974; Sgall 1969, 1975, 1978; Benesova et al. 1973; for an introductory summary to the work of the Prague Schools, see Sgall 1993), the influential articles in Li's (1976) "Subject and Topic" (cf. Li & Thompson 1976, Chafe 1976, Givón 1976) or the first generative approaches (cf. Chomsky 1971, Jackendoff 1972). Most of those first writings have in common that they were primarily concerned with the linguistically overt effects of information structural concepts. Further, all of those approaches share that the object of their study was the sentence or the clause. Other early approaches provided the foundation of what has become known as the study of discourse structure (cf. Halliday 1970, 1974).

In section 2.1, we discuss the development of the research on topics which focused in the sentences level (i.e. IS-Topics; henceforth IST). In section 2.2, we provide some background on the research on discourse structure topics, by which we mean the study of topics on the level above the sentence (i.e. DS-Topics; henceforth DST). Throughout this paper, we will use the term *topic expression* when we mean an expression which refers to a topic, which itself is a semantic element. We slightly modify Lambrecht's (1994: 131) definition of topic expression to capture expressions which refer to ISTs as well as expressions which refer to DSTs.

Definition of topic expression (general)

- (1) A constituent is a topic expression iff the message¹ conveyed by the linguistic form unit (e.g. a sentence or a paragraph) with which it is associated is pragmatically construed as being about the denotation of (a part of) this constituent.

2.1 Information Structure Topics

Broadly speaking, varieties of the definitions of the term (sentence-level) topic can be classified into two groups, namely, those based on givenness (familiarity) and those based on *aboutness*. Below we will briefly review some leading approaches, to clarify our terminology and standpoint.

The givenness-based approach defines the topic in terms of the givenness and its varieties [+/- mentioned] (originally [+/- vorerwähnt], cf. Fries 1971), [+/-new] (cf. Choi 1999), [+/- contextually bound] (cf. Sgall 1975), [+/- on-stage] (cf. Chafe 1976 who ultimately means [+/- activated] in a psychological sense), [+/- recoverable] (cf. Halliday 1967/8, 1970), and the grade of communicative dynamism (cf. Firbas 1974). In her well-known article "Toward a Taxonomy of Given-New", Prince (1981) shows that the subvarieties of givenness can be summarized into three types: i.e. Givenness_p, which corresponds to *predictability* and *recoverability* (cf. Prince 1981: 226), Givenness_s, which corresponds to *saliency* (ibid.: 228), and Givenness_k, which corresponds to the *shared knowledge* of the speaker and the listeners (ibid.: 230). Those three types of Givenness are shown to be not mutually independent. Since, each of the three types of Givenness turns out to be insufficient alone, Prince then defines a taxonomy of *Assumed Familiarity*, which is supposed to cover those information structural phenomena which turn out to be relevant for linguistic form (ibid: 233).

Sgall (1975: 303), Dik (1978, 1980: 42, 1989: 266), Vallduví (1992), Lambrecht (1994) and É. Kiss (1995), on the other hand, all consider topic to be "what the sentence is about".² An elaboration of

¹ By *message* we mean a communicative function of utterances. Note that we use the term message to refer to a level of representation which is richer than the purely truth-conditional form (i.e. propositions).

² Note that Dik (1989) does not limit "aboutness" to the sentence level (cf. section 3.1). In some sense, other framework that distinguish between, for example, *topic* with *comment* (cf. Gundel 1974, Dahl 1974, Kuno

the notion of *aboutness* is given in Vallduví (1992). Utilizing the idea of File Change Semantics (Heim 1992), he characterizes the topic (link, in his terminology) as the phrase that denotes the discourse entity on whose file card the sentence's information is entered (see also Engdahl and Vallduví 1996, Portner and Yabushita 1998).

Another line that divides existing approaches to topic is the one between relational and non-relational views. In the relational view, taken by Gundel, Halliday and Vallduví among others, the topic is defined only in contrast to other notions, such as "comment" and "rheme". Chafe and Dik, in contrast, adopts an extrinsic definition of topic. In this paper, we take the aboutness-based/non-relational view³, although we do not limit our interest in the phenomenon at sentential levels, nor consider topic as solely sentence/clause-internal phenomenon. We adopt the following, tentative definition of what we refer to as Information Structure Topic (IST) expression based on Lambrecht (1994: 118):

Definition of Information Structure Topic

- (2) The referent of a constituent X is an Information Structure Topic iff X denotes the thing the message conveyed by the sentence or clause is ABOUT.

2.2 Discourse Structure Topics

As mentioned above, there has been another research tradition which arose from the early seminal works on "topics", namely the study of discourse structure (cf. Schiffrin 1994: 23-40). Here, we are not interested in the whole area of discourse analysis (e.g. Hopper 1979, Labov 1997), but in the study of topics above the sentence level (cf. Chafe 2001, Brown and Yule 1983, Fraser 1988). A definition of discourse topic is given by Chafe (2001), as the following:

- (3) A coherent aggregate of thoughts introduced by some participant in a conversation, developed either by that participant or by several participants jointly, and then either explicitly closed or allowed to peter out. (ibid: 674)

Since we will arrive at a hierarchical model of DST (cf. section 3), we need a more refined definition of DSTs here. We use the term DST to refer to any kind of topic which is defined above the sentence level (e.g. *discourse topics*, i.e. paragraph level topics, or *supertopics*, as mentioned in Chafe 2001: 674). Parallel to our definition of ISTs, we use the *aboutness* relation to provide a descriptive definition of DSTs :

Discourse Structure Topic definition

- (4) The topic of a discourse chunk is the thing which the messages conveyed by that discourse chunk are ABOUT.

A marker of any kind of DST will be referred to as *discourse structure topic marker* (henceforth DST-marker). In the following section we propose a hierarchical model of topics, which we motivate in section 3.2 and 3.3.

1980, Reinhart 1982) and *theme* vs. *rheme* (Firbas 1966, 1971 Halliday 1967/8) also used aboutness to define the term topic.

³ In this respect, our stance is the closest to authors such as Sgall (1975), Dik (1978, 1989), Lambrecht (1994).

3 Hierarchical Model of Discourse Structure Topics

In many previous approaches mentioned in the last section, authors explicitly limit their attention to the level above/below sentences, implicitly assuming that the notions of IST/DST could not be treated in a homogeneous way (Chafe 2001:673-5, Lambrecht 1994). Of course, if one takes all information packaging phenomena with syntactic correlates to be ISTs (cf. Lambrecht 1994: 118), the distinction between ISTs and DSTs is self-fulfilling. However, to the best of our knowledge, there has been no attempt to argue for the ontological/interpretative distinction of those two types of topics.

The position against this division, which we call the “uniform view of topic”, is taken by Dik (1989). He tries to base the notion of topic⁴ on D(iscourse)-topics, which he defines to be the entities which a discourse⁵ (i.e. a coherent text, cf. *ibid*: 266-7) is about. Here, we follow the uniform view in that we, unless proven otherwise, assume that all topics are ontologically homogeneous. In the next section, we sketch out a model which describes hierarchically organized sets of topics (both DSTs and ISTs). While our model shares certain conceptual properties with Dik's (1989) and Buring's (*to appear*) work, it differs in important aspects, which we clarify along the way.

3.1 T-trees

A discourse may have different DSTs, and DSTs may be hierarchically and/or sequentially organized. Thus, the notion of topic should be interpreted relative to the relevant stretch of discourse (e.g. book, chapter, section, paragraph, and ultimately the individual clause; cf. Dik 1989, Chafe 2001). We propose that the topics are better understood as an ordered set, which typically has a structure of (algebraic) tree.

⁴ For Dik (1989) topic is one of the clause-internal pragmatic functions along with focus. Pragmatic functions “concern the informational status of constituents of the clause in relation to the wider communicative setting”, where “‘communicative setting’ can be understood in terms of the Speaker’s estimate of the Addressee’s pragmatic information at the moment of speaking” (*ibid*: 265).

⁵ The term *discourse* is used to refer to at least three different things, namely roughly (a) language above the sentence, (b) language in use, and (c) utterances (cf. Schiffrin 1994: 23-40). Also, there is a fourth use of discourse in the term DISCOURSE FUNCTION used in literature on LFG (Lexical Functional Grammar, cf. Bresnan 2001), which refers to a part of the F(unctional)-structure and, at the end, is quasi equivalent to an information structural role. Here, we use the term *discourse structure* to refer to the structure of language above the level of the sentence.

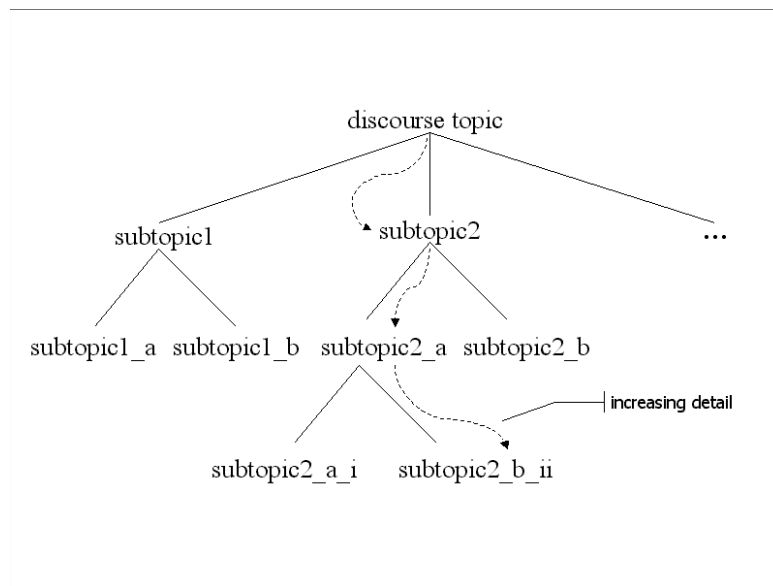


Figure 1 – Schema of a T-tree

Let us call this conceptual entity, T(topic)-tree.⁶ The root of the T-tree corresponds to the global discourse topic (*supertopic* in Chafe 2001), whereas “leaves” (at the moment of each utterance), which are introduced or picked up at the sentence level, correspond to ISTs (cf. Dik 1989: 266-77). We depart from Dik (1989) in assuming that the members of the T-tree are limited to entities i.e. objects of type *e*. or group of entities. In the following examples, the sentence level topic can be of higher type like $\langle e, t \rangle$ (property, proposition abstract) and $\langle t, t \rangle$.⁷

- (5) Taro-ga tabeta-no-**wa** mame da.
Taro-NOM ate-COMP-TOP beans be
What Taro ate was beans.
- (6) Ookii-koto-**wa** ii koto da.
big-COMP-TOP good fact be
Being big is good. (lit.) / the size matters.
- (7) Heya-no naka-de-**wa** Taro-ga hon-o yondeiru.
room-GEN inside-LOC-TOP Taro-NOM book-ACC be reading
In the room, Taro is reading a book.

⁶ Although our topic-tree might look reminiscent of “question-based” models of discourse structure, such as Roberts’ (1996) “push-down store” model, and Büring’s D(iscourse)-tree, our model differs from them in essential respects. In question-based analysis, the members of the tree are either a question or proposition, and in McNally (1998), Büring (1997), for example, the notion of topic is derived from that of focus. In our model, topics are treated as primitives that form the T-tree, and their semantic type is typically (but not necessarily) an entity or group of entities. See Portner and Yabushita (1998), who argue for “topic-as-entity” view over “topic-as-question” view. Moreover, what McNally (1997) and Büring (1997) term as topic is a rather different from our notion of topic, but is what is commonly called “contrastive topic” (cf. Büring *to appear*, Lee 1999; Oshima 2002 argues that contrastive topic is not a topic at all, but a type of focus associated with presupposition about alternative propositions).

⁷ In this respect, our position resembles Chafe (2001) who assumes that discourse topics can be any “aggregate of thought”.

The Topic-tree expands (or grows “down”) in course of the discourse, as new entities, introduced as topics, get dynamically “adjoined” to it. In other words each utterance in the discourse could either maintain the topic of the previous utterance or “shift” it to another topic in either the same or a lower level of the T-tree. Figure 2 illustrates the different choices to expand a T-tree at a given moment in discourse.

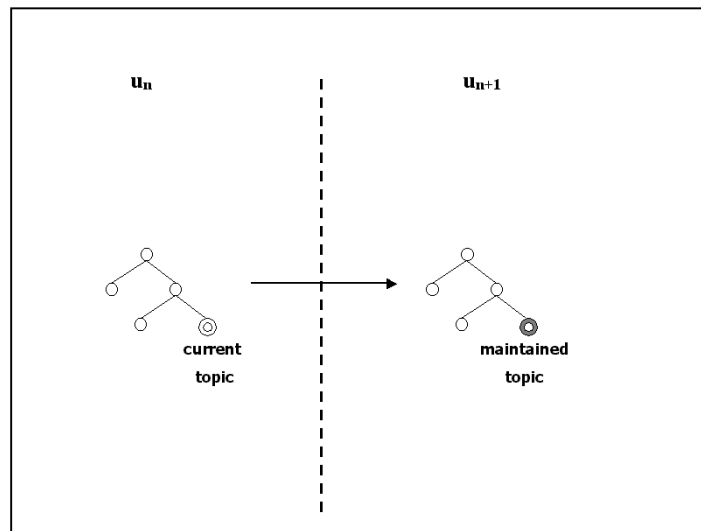


Figure 2 – Representation of Topic maintenance

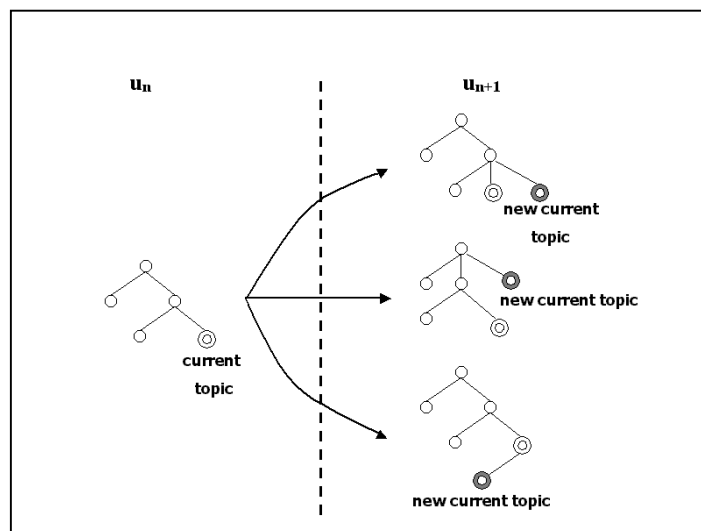


Figure 3 – Representation of Topic shifts

At any given point of discourse (i.e. a utterance u_n) we define the discourse status of a T-tree to be the set of the most current ISTs:

Discourse Status

- (8) $\langle T, n, \{t_1 \dots t_m\} \rangle$ is a discourse status when T is a T-tree and $\{t_1 \dots t_m\}$ are a subset of T -tree whose members are ISTs of the n^{th} utterance in a discourse (i.e. u_n).

The adoption of the uniform view of topic is not only a matter of conceptual parsimony or ontological aestheticism. In the following, we demonstrate two phenomena that suggest syntactic, morphological or lexical devices, which are commonly believed to be IST markers, should be able to “look up” to the global topic structure i.e. T-tree.

3.2 Argument from English 'as for' construction

The 'as for' construction in English (henceforth AFC) has been assumed to mark an IST (e.g. Lambrecht 1994: 149, Gundel 1974, Givon 1992: 308), as in (9), or contrastiveness (e.g. Chafe 1976: 49, 1994), as in (10). The AFC therefore seems to be commonly considered to be a grammatic correlate on the sentence level (cf. Lambrecht 1994: 117).

- (9) As for Rosa, John didn't really love her (cf. Lambrecht 1994: 149)
(10) As for the PLAY, he saw it YESTERDAY. (cf. Chafe 1976:49)

However, the pragmatic AFC's constraints cannot be fully captured at the level of individual sentences. One of the authors (Jaeger 2002, *in preparation*) has argued that what the AFC marks is a shift of a discourse structure subtopic (henceforth DSST), which has to be compatible with a higher DST in the T-tree. DSST roughly corresponds to a DST below the paragraph level in terms of the model proposed above. The following examples illustrate the AFC's function. The paragraph level DST (cf. the term "discourse topic" as used in Chafe 2001: 674, referring to Brown and Yule 1983: 71) in (11) is (something like) |catering a large number of people|. The DSST immediately preceding the AFC is |choice of food|. The DSST introduced by the AFC is |quantities|. Even though there are no directly coreferential phrases in the following sentence, the introduced DSST is picked up by the numbers in "six canapés [..]", "eight to ten [..]" in the immediately following sentence. Furthermore the DSST |quantities| is maintained in the following sentence and picked up by the expression "six glasses [..]", "25 measures [..]", "four glasses", and "six mixers [..]".⁹

- (11) It's not easy catering for a large number of people, as the chances are they won't fit round your dining table. A fun solution is to treat them to finger food. Provide a wide selection, making sure there are lots of different colours, flavours and textures. As for quantities, allow four to six canapés and nibbles per person as a pre-dinner appetiser, and eight to 10 for a longer party. Alcohol and soft drinks are much harder to estimate. As a guide, there are roughly six glasses to a bottle of wine, 25 measures to a 70cl bottle of spirits and four glasses, or enough for six mixers in a 500ml/18fl oz bottle of tonic.

As already mentioned, the new DSST introduced by an AFC has to be a subtopic (which is, typically, part, member or aspect) of an already established "higher" DST. In (12), the speaker/writer talks/writes about his house, which is located in the middle of nowhere. One problem with this house is related to water. This |water problem| is the paragraph level DST. The DSST immediately preceding the AFC in (12) is |laundry|. The DSST introduces a new DSST, namely |toilet facilities|. Intuitively, both DSSTs are compatible (i.e. coherent) with the paragraph level DST |water problem|.

- (12) But it's a fair distance to walk there and it is, I suppose, a lot handier to pop down to the reservoir. I can have the washing done and

⁸ We use the notation |*expression*| when we mean the denotation of a linguistic expression.

⁹ All of the following examples in this section have been collected in the British National Corpus.

back home again by the time I get to Richard's place. To be honest I never get much washing done during the winter months. Usually it is done during the summer, certainly for my better clothes. As for toilet facilities, I have an earth closet so the lack of water does not affect that side of things.

Below, we will discuss the two common claims about the AFC mentioned above: (i) it indicates contrastiveness and (ii) it marks an IST.

While the AFC *can* correlate with contrastiveness if the AFC's argument NP receives an emphatic accent, e.g. 'PLAY' in (10), contrastiveness is not required. Consider (13), where it is by no means clear what the referent of 'Dennis' (i.e. |Dennis|) is contrasted with. To be precise, there are (at least) two different possible readings of (13). In the first, "HIM" receives contrastive accent. In the second (group of) readings, "KILLing" or, considering the context, more likely "KINDness" receives focus accent (i.e. new information accent) and "him" does not get assigned any special stress. The first reading contrasts |Dennis| with other *candidates for* "killing X is a kindness". Here, we are especially interested in the second reading.¹⁰

- (13) Never happy unless she has something in her mouth," he confided to me. `You wouldn't think it to look at her, would you?" `Oh, I don't know." Karen giggled hysterically, spluttering water all over the counter. I knew then that we were bound to go all the way, wherever it might lead, whether we wanted to or not. As for Dennis, well, after that killing him would have been a kindness, wouldn't it?

We do not deny that the referent of the AFC's argument is part of a *partially ordered set* (henceforth "poset", cf. Hirschberg 1985) in both readings and therefore in some sense contrastive. However, the members of the posets which go with the two above-mentioned readings are of a very different type.

For the first reading of (13), |Dennis| is a member of a poset that contains the possible instances of the variable x in the proposition [BE(killing(x), a-kindness)]. Note that this meaning is available for the last sentence of (13) even if it is not preceded by an AFC, as long as "HIM" is realized with contrastive stress. Similar kinds of readings are available for the case that other constituents in the sentence receive contrastive accent. We call this prosody-driven contrastiveness $\text{KontraST}_{P[\text{proposition}]}$. Now consider (13) with no contrastive stress on either the argument expression of the AFC (i.e. "Dennis") or the coreferential expression in the sentence (i.e. "him"). In this case, |Dennis| is still involved in some kind of contrast.¹¹ This contrast exists between the following two formulas $\text{DSST}=|\text{Dennis}|$ and $\text{DSST}=\alpha \ \& \ \alpha \neq |\text{Dennis}|$. In other words that the new DSST is |Dennis| could be said to contrast with possible other DSSTs at the point of discourse before the AFC. We call this contrastiveness between possible DSST (or more generally different kinds of topics) $\text{KontraST}_{T[\text{topic}]}$. KontraST_T is found in all examples in our corpus containing an AFC, whereas KontraST_P is independent of (though highly compatible with) the AFC and, in English, always indicated by prosody. Although it is a subtle difference, we ask the reader to pause a second and convince herself of the qualitative

¹⁰ We used the following mark-up conventions: the AFC as well the sentence which is left-detached to are underlined. In addition, the AFC is given in *italic* type-setting.

¹¹ Some informants (including linguists) have problems seeing any contrastiveness related to |Dennis| in (13). This can be understood if we take Schwenter (1999) into consideration. He argued that contrastiveness and additiveness are two extremes of the same scale. Although this may be surprising at the first glance, the AFC illustrates Schwenter's point. Below, we show that the AFC adds a new DSST to the T-tree. Also, the AFC, in some sense, contrasts the introduced topic with other referents which would have been valid DSSTs at that point of discourse.

difference between the two types of contrast just discussed.¹² Another example, which shows that the AFC does not mark Kontrast_p is given in (11) above.

Now that we have discussed the hypothesis that the AFC marks contrastiveness, we turn to discuss the second common claim about the AFC that can be found in the literature. According to Lambrecht (1994:149) and Gundel (1974), among others, the AFC marks ISTs. This is in some sense true. However, there are a couple of points to be mentioned about this claim.

First, we have shown above that the AFC also marks a shift of DSSTs. Lambrecht (1994: 188) points to this when he mentions that the AFC is rather a *topic announcing* expression than a topic expression. This is just another way to capture the intuition which can already be addressed under such different labels as DSST shift, introduction of a new DSST or contrast between possible other DSST.

Second, for many examples the intuitions what "the sentence is about" diverge crucially. Consider the following example, where – probably primarily depending on one's own intuitions could argue that either or all of "justice and peace", "the Zuwaya" and "judges" are ISTs. For now it remains an open question whether the topic marked by the AFC has different properties than the topic within the sentence.¹³ In any case, the AFC introduces a new topic and is subject to certain constraints, which will be spelled-out below.

- (14) But that was generally the practice of settled people: Zuwaya were scattered over a vast territory, and although they owned land they were not permanently resident on it. Most shrines were the property of people who could visit them easily, and who might find it expedient to assemble there to show their strength if they had land disputes with their neighbours. As for justice and peace, the Zuwaya had no judges.

To sum up, we have shown that the AFC adds a new DSST to the list of DSTs, which often is maintained for a couple of sentences. In that sense, the AFC is what we informally can refer to as *forward-sensitive*. Furthermore, the new DST has to be different as the immediately preceding DSST. More carefully formulated, the AFC cannot introduce a topic which is identical to an IST of the immediately preceding sentence. Consider (14), where the AFC's argument would be "shrines", or, (11), where the AFC's argument would be "washing". All informants we asked agree that those examples sound really awkward. In this sense, the AFC is *backward-sensitive*. Finally, the new DSST has to be compatible with the current paragraph level DST (maybe a general well-formedness condition in order to guarantee cohesion). We will refer to this property of the AFC as being *hierarchy-sensitive*. Note that the claimed hierarchy-sensitiveness correctly predicts that the AFC cannot occur text-initial, since there is no already established "higher" topic available at this point.¹⁴ This contrasts with, for example, the Japanese thematic *wa*-construction (cf. Portner and Yabushita 1998, and section 3.3).

The model proposed in section 3.1, enables us to formalize the AFC's backward-, forward- and hierarchy-sensitivity. Formal descriptions of the AFC's constraints are given below, where a discourse status is defined as in (8). The constraint in (15) captures the AFC's backward-sensitivity; the constraint

¹² To be precise we are not talking about two kinds of contrast but about the same kind of contrast (relation) over ontologically different kinds of formal objects.

¹³ Preliminary tests were not conclusive. Tested hypotheses included "The topic marked by the AFC is [+FRAME-SETTING] in Jacob's (2001: 654)" and "The topic marked by the AFC is less concrete in a semantic type hierarchy" – both so far with unclear results.

¹⁴ There are some apparent exceptions to this generalization. Imagine a class room situation, where the teacher comes in and starts the discourse with "As for last week's homework, you can find it in front of my office". While we think that this example sounds really awkward, it could be explained by the fact that the discourse provides such a salient "higher" topic (in this case a *supertopic*, cf. Chafe 2001: 675; also *global DST*) that it can be assumed to be accessible to everyone in the audience.

in (16) captures the hierarchy-sensitivity. Finally, (17) is a first attempt to describe the forward-sensitive *preference* of the AFC to maintain the newly introduced DSST for a couple of utterances.

Constraints on the AFC

- (15) Given a discourse status $\langle T, n-1, \{t_1 \dots t_m\} \rangle$, u_n of the form $[_s [as\ for\ X], \dots]$ is felicitous only if the referent of X is not identical to any of $\{t_1 \dots t_m\}$ and nor to a member of T-tree that dominates $\{t_1 \dots t_m\}$.
- (16) In a T-tree T, u_n of $[_s [as\ for\ X], \dots]$ is felicitous only if the referent of X is not identical to ROOT(T).
- (17) In a T-tree T, u_n of $[_s [as\ for\ X], \dots]$ *prefers* the referent of X to be an IST in u_{n+1}, \dots, u_{n+k} .

3.3 Argument from Multiple Topic Sentences

Sentences involving “multiple topics” pose another problem to the non-uniform view of topic. Some people have implicitly or explicitly assumed that a sentence has at most one topic (e.g. Kuno 1973). Consider the following examples from Japanese (taken from Kuno 1973: 48) and English.

(18) *As for John, as for Mary, he married her.

(19) Watakushi-wa tabako-wa suimasu.
I-TOP cigarette smoke
'Speaking of myself, I DO smoke.'

The second *wa* in (19) should be construed as contrastive, rather than thematic (cf. Kuno 1973 48; see also Oshima 2002). However, there are exceptions to this rule. Consider the following examples taken from a Japanese corpus:¹⁵

[CONTEXT: the preceding paragraph talks about the atoms-for-peace program of the United States in the 1950's, and its connection with the United Nations Second International Conference on the Peaceful Uses of Atomic Energy held in 1958.]

(20) Nihon-mo, sono yokunen-ni "gensiryoku-kihon-hoo"-o jyootee simasita.

Nihon-wa gensiryoku-ni-kansite-wa wariiai hayai
Japan-TOP atomic-energy-TOP relatively progressive
Japan put the "Atomic Energy Act" on the agenda in the next year, too. About atomic power, Japan has been relatively pioneering.

[CONTEXT: none (the beginning sentence of an essay)]

(21) Zinrui hazimete-no tuki-e-no choosenkeekaku-ni-kansite-wa
mankind first-GEN moon-TOWARD attempt-project-TOP

sekaijyuu-no SF-sakka-ni-wa soozikohihan
entire-world-GEN sci-fi-authors-DAT-TOP entire-self-criticism

sinakutewa-naranai ten-ga aru.
do-must point-NOM exist
Speaking of the first attempt to land the moon in the human history, the sci-fi authors in the world have to criticize themselves in one point

¹⁵ The example sentences in (13) and (14) are taken from text #1226 and #512 in Komatsu Sakyo Corpus, which can be accessed online via <http://castelj.soken.ac.jp/groups/komatsu>

The above examples have two topics, one marked by the particle *-wa* and one by an adpositional expression *ni-kansite-wa*, which roughly translates as 'as for' or 'regarding'.¹⁶ There are even cases, where more than one element is *wa*-marked and neither occurrence of the *wa*-marked elements has to be interpreted as contrastive. Those examples directly contradict Kuno's generalization:

- (22) Kono kuni-de-**wa** kare-**wa** itumo kimyoo-na taiken-o suru.
 this country-LOC-TOP he-TOP always bizarre experience do
In this country, he always has strange experiences.
- (23) Sono situmon-ni-**wa** watasi-**wa** kotaeraremasen.
 that question-DAT-TOP I-TOP answer-can-not
*I cannot answer that question.*¹⁷

What controls the possibility of multiple topics within a single sentence? The generalization seems to be that one of the topics must be "higher" than the other one in the T-tree. In (20), for example, "atomic energy" and "Japan" stand in the relation of a supertopic/subtopic. In this concrete case, "atomic energy" corresponds to the paragraph-level DST and "Japan" to the section-level DST). The sentence in (21), appearing at the beginning of discourse, gives a hint at the author's information packaging strategy, which could roughly be described as "I'm going to talk about moon landing, quite a big issue which everyone is talking about these days. On this occasion, I particularly want to talk about its impact on the sci-fi authors".¹⁸

In the model proposed here, the constraint on sentences with multiple topics in Japanese (and perhaps in other languages, too¹⁹) can be captured as follows:

Constraint on Multiple Topic Construction(s)

- (24) Given a T-tree T, the utterance of [S [TOP X] ... [TOP Y] ...] is felicitous only if the referent of X dominates the referent of Y, or the referent of Y dominates the referent of X in T.

This constraint seems to hold for both types of multiple topic constructions mentioned above (i.e. the construction where both topics are marked by *-wa* and the construction where one topic is marked by *ni-kansite-wa*).

To sum up, we have shown how the model sketched out in section 3.1, is able to describe the discussed data from English and Japanese. The proposed model provides a platform for formalizations of topic-related phenomena. Applying the model to the AFC and multiple topic sentences, we have shown in section 3.2 and 3.3 how our formalizations (15), (16) and (24) make attestable predictions.

4 Kontrast, C-construability and Topic

In this section, we briefly examine some bearing of our "T-tree" model on the theory of focus: namely, we will argue there is possible overlapping between those entities which can be a topic (a member of T-tree) and those which can be focalized.

It has been widely argued that there exist (at least) two kinds of foci, which are called contrastive (or identificational) focus and presentational (or completive, information etc.) focus. Although the terminology varies among scholars (e.g. CONTRASTIVE vs. COMPLETIVE focus in Choi

¹⁶ Although less apparent (due to the lack of morphological marking) the same seems to hold in English. Consider, for example, (14) where one could argue that both "Peace and justice" and "the Zuwaya" are topics.

¹⁷ Compare example (23) with Noda (1996: 213): 'Watasi-wa sono situmon-ni-wa kotaeraremasen.' which corresponds to 'As for that question, I cannot answer.'

¹⁸ The author of this essay, incidentally, is a sci-fi author.

¹⁹ Bulgarian multiple topic constructions may be one exceptions to the constraint stated in (24).

1999, Dik et al. 1981; CONTRASTIVE vs. PRESENTATIONAL in Herring 1990, Rochemont 1986; IDENTIFICATIONAL vs. INFORMATIONAL in É.Kiss 1998), the basic ideas and insights are shared. Characteristics that distinguish the contrastive focus from the completive are as follows: (i) contrastive focus is implicitly or explicitly opposed to some other piece of information, while completive focus simply fills in a gap in the pragmatic information of the addressee (Dik et al. 1981); contrastive focus induces a set of alternatives (possible candidates) (Chafe, 1976: 36), (ii) contrastive focus (unlike presentational focus), can be c(ontext)-construable i.e. be ‘under discussion’ or ‘have a semantic or discourse antecedent in the discourse’ (Rochemont 1986: 47); contrastively focused arguments are already explicitly or implicitly present in the discourse context (Herring 1990:164), and (iii) the contrastive focus implicates ‘exhaustiveness’, while the completive focus does not (É. Kiss 1998, Vallduví and Vilkuna 1998).²⁰ Although in English the contrastive and completive foci are not always formally differentiated (though the former has been mentioned to be often accompanied by an extra heavy accent), there are morpho-syntactic means to single out the contrastive focus e.g. *it*-cleft construction:²¹

- (25) It’s potatoes I like(, not tomatoes).
 (26) It’s Mary I like(, not John).

Vallduví and Vilkuna (1998) distinguish the notion of “kontrast” from “rhematicity”, demonstrating syntactic evidence from languages like Finnish and Hungarian. The concept of rhematicity belongs to the domain of information packaging, and indicates the new information (update potential) of the sentence. Kontrast, on the other hand, is an “operator-like” element, that generates a membership set *M* (which corresponds to *C*-sets in Rooth 1992) which serves as some sort of quantificational domain for semantic computation (involving, for example, operators like *only* and *wh*-words) It, thus, captures the semantic import of “focus” in alternative semantics.

Within Vallduví and Vilkuna’s model, what has been called contrastive focus, identificational focus, exhaustive focus etc. refers to elements which are both rhemes and kontrastive (i.e. [K:+, Rh:+]). Utterances like (27) are interpreted to be identificational kontrast, as defined in (28).

- (27) John introduced BILL to Sue.

Identificational Kontrast

- (28) If $M = \{a, b, c\}$ and $P(x \in M)$, then $P(a)$ where $P = \lambda x$. John introduced *x* to Sue.

Property (ii) of contrastive foci (i.e. the ‘not entirely new’ property) seems to partially overlap with the notion of topichood.²² As mentioned in Section 2.1, topics often have been argued to stand in an correlation with Assumed Familiarity (cf. Prince 1981). We assume that the T-tree of our model largely overlaps or is subsumed by the set of *c*-construable objects (cf. Rochemont 1986: 47). Furthermore, some subtypes of contrastive foci seem to be topics as well. The “major subject” (the first nominative NP in the sentence) of Japanese multiple nominative construction (henceforth MNC) is one example:²³

²⁰ Whereas É. Kiss (1998) claims that the exhaustiveness induced by identificational focus is semantic (truth-conditional), Vallduví and Vilkuna (1998) assume that it is a conversational implicature of the identificational effect.

²¹ Example (25) and (26) are taken from Givón (1992: 309).

²² In this connection, Choi (1999) proposes to postulate the discourse feature [+prominent] to group topic and contrastive focus as a natural class. One flaw of her feature setting is that it cannot differentiate subtypes of contrastive focus (see below).

²³ For discussion about syntactic aspects of the MNC, see Heycock (1993), Takahashi (1994) among others.

- (29) Kono michi-**ga** jidoosha-**ga** jiko-o okosiyasui.
 this street car accident-ACC cause-likely
It is on this street that car accidents often break out.
 (cf. Jidoosha-**ga** kono michi-**de** jiko-o okoshiyasui.)
 car this street-at accident-ACC cause-be-likely
Car accidents often break out on this street.
- (30) Zo-**ga** hana-**ga** nagai.
 elephant trunk long
It is the elephant whose trunk is long.
 (cf. Zo-**no** hana-**ga** nagai.)
 elephant trunk long
The trunk of elephant is long.

The major subject of MNC has all the characteristics of contrastive focus given above (cf. Mihara 1994, Gunji 1987, Sugimoto 1995). In addition, it has been argued that, there are additional semantic/pragmatic conditions required to license the MNC construction: Mihara (1994: 110), for example, proposes that the relation between the major subject and the following sentence (that denotes some property) is licensed by the “aboutness” relation. Masuoka (1982) and Amano (1990) make similar remarks using the terms “characterization” and “seesitu jyojyutu”(property description) respectively. Given this, it seems reasonable to assume that the MNC promotes an element marked {[+topic], [+focus], [+kontrast]} to the major subject.²⁴ The topic marked by the thematic “wa”, in contrast, cannot be [+focus] (cf. Kuno 1973, Noda 1996, Oshima 2002). The discourse function associated with the major subject of MNC and *wa*-marked constituent in thematic-*wa* construction can be summarized as follows:

- (31) MNC: {[+topic], [+focus], [+kontrast]}
 (32) Thematic -*wa*: {[+topic], [-focus]}

Next, consider the two types of English "topicalization" as displayed in (33) and (34).

- (33) John she CALLED.
 (34) JOHN she called.

We propose that their shared characteristics, too, could be adequately captured utilizing the notion of kontrast. It is well known that the preposed ("topicalized") element can be either focalized or not (Gundel 1974: 143-50, Prince 1981, 1999, Choi 1999: 87-8). Although we have neither clear intuition nor evidence to explicate the exact informational import of the preposed NPs in (33) and (34), previous studies argue that (i) the preposed element, when focalized, is interpreted as a contrastive focus (Choi 1999), and (ii) the preposed element, even having a tonic accent on it (i.e. being focalized) has the status of a topic (i.e. it is what the sentence is about; cf. Givón 1992: 309, Gregory and Michaelis 2001) (iii) when the pitch accent does not fall on the preposed element, an inference is induced that the entity represented by the preposed element stands in a salient partially ordered set ('poset') relation to some entity or entities already evoked in the discourse (cf. Prince 1999: 7ff.). This suggests that the topicalization construction in English is licensed by the features {[+topic], [+kontrast]}, and is underspecified for the [+/-focus] feature.

²⁴ Note that this feature pattern is impossible in the model of Vallduví and Vilkkuna (1998), where topicality is defined as [-Rheme]. In this regard (too), our position is more similar to Dik (1989: Ch. 13) who assumes that the dimensions of topicality and focality may overlap (i.e. a "new topic" is possible).

(35) English Topicalization: {[+topic], [+kontrast]}

It is not the case, however, that all occurrences of contrastive focus are topics. For instance, we have no particular reason to believe that the focus of *it*-cleft construction should be a topic (but see Givón 1992: 308-9).

(36) *it*-cleft: {[+focus], [+kontrast]}

In summary, the T-tree and a given membership set M (which is generated by kontrastive operators, contrastive focus etc.), may and may not overlap, and languages have phonological / morpho-syntactic means to specify (a subset of) possible combinations of the orthogonal features [+topic], [+kontrast], and [+focus].

5 Conclusion and Outlook

In this paper, we argued that the global structure of DSTs may affect the topic-marking at the sentence level, which lead us to adopt the “uniform” approach to topic. In this approach discourse structure topics (DSTs) and information structure topics (ISTs) are situated on a single plane, which we formalized by means of T-trees. The proposed model provides a tool for first steps towards the formalization of both DST and IST-related phenomena such as the 'as for' construction in English or multiple topics in Japanese. Among other things T-trees allow for the formalization of hierarchy-sensitive constraints.

It should in general be possible to examine the well-formedness conditions for T-tree growth (i.e. the general constraints on the coherence of DST and IST shifts/maintenance). By providing a notion of T-trees on which possible tree growth can be formalized, we move one step closer to a dynamic model of the construction of discourse (topics). At this point, it would be interesting to investigate the compatibility of the model proposed here with other discourse models such as Centering Theory (cf. Grosz, Joshi and Weinstein 1995).

We also proposed that our notion of topic participates in the inventory of information packaging strategy along with the focality/rhematicity and the “kontrast” discussed by Vallduví and Vilkuna (1998), and that it subsumes some subtypes of so-called contrastive focus. We leave elaboration of the proposed (and informally presented) model to future research. Preliminary application of our model in interaction with focus and kontrast allowed us to sketch an analysis of a variety of discourse sensitive constructions, such as the English topicalization construction(s), *it*-clefts, and Japanese multiple nominative constructions.

6 References

- Amano, Midori (1990) “Hukushuakubunkô - hukushukakubun no imi to, sêritsu ni kakawaru imitekisêyaku” (essay on multiple nominative sentence – meaning and semantic constraints on the formation of multiple nominative sentence), *Nihongogaku* 9:5, 27-42.
- Büiring, Daniel (to appear) “On D-Trees, Beans, and B-Accents”, ms. UCLA, to appear in *Linguistics & Philosophy*.
- Chafe, Wallace, L. (2001) “The Analysis of Discourse Flow; in Deborah Schiffrin, Deborah Tannen, and Heidi E. Hamilton (eds.) *The Handbook of Discourse Analysis*, Cambridge: Blackwell Publishers, 673-687.
- Dik, Simon C. (1978) *Functional Grammar*, Amsterdam: Elsevier.
- Dik, Simon C. (1989) *The Theory of Functional Grammar Part I*, Dordrecht: Foris Publications.

- Dik, Simon C., Maria E. Hoffmann, Jan R. de Long, Sie Ing Djiang, Harry Stroomer and Lourens Devries (1981) "On the Typology of Focus Phenomena" in Teun Hoekstra, Harry van der Hulst and Michael. Moortgat (eds) *Perspectives on Functional Grammar*, Dordrecht: Foris Publications, 41-74.
- É. Kiss, Katalin (1995) "Introduction" in Katalin É. Kiss (ed.) *Discourse Configurational Languages*, Oxford: Oxford University Press, 3-27.
- É. Kiss, Katalin (1998) "Identificational Focus versus Information Focus", *Language* 74(2), 245-273.
- Engdahl, Elizabet, and Enric Vallduví (1996) "Information Packaging in HPSG" in Claire Grover and Enric Vallduví (eds.) *Edinburgh Working Papers in Cognitive Science*, Vol 12 *Studies in HPSG*, 1-31.
- Firbas, J. (1971): "On the concept of Communicative Dynamism in the Theory of Functional Sentence Perspective", in *Papers on Functional sentence perspective*, Prague: Academie.
- Fraser, Bruce (1988) "Types of English discourse markers" in *Acta Linguistica Hungarica* 38, 19-33.
- Givón, Talmy (1976) "Topic, Pronoun, and Grammatical Agreement" in Charles N. Li (ed.) *Subject and Topic*, New York: Academic Press, 149-188.
- Givón, Talmy (1992) "On Interpreting Text-Distributional Correlations" in Doris L. Payne (ed.) *Pragmatics of Word Order Flexibility*, Amsterdam: John Benjamins
- Gregory, Michelle L. and Laura A. Michaelis (2001) "Topicalization and Left Dislocation: A Functional Opposition Revisited", *Journal of Pragmatics* 33: 1665-1706)
- Grosz, Barbara J.; Joshi, Aravind K. and Scott Weinstein (1995) "Centering: A Framework for Modelling the Local Coherence of Discourse", in *Computational Linguistics* 21(2), 203-225.
- Gundel, Jeanette K. (1974) *The Role of Topic and Comment in Linguistic Theory*. Ph.D. Dissertation, University of Texas.
- Gundel, Jeanette K. (1985) "'Shared Knowledge' and Topicality" *Journal of Pragmatics* 9.1, 83-107.
- Gunji, Takao (1987) *Japanese Phrase Structure Grammar*, Dordrecht: Reidel.
- Halliday, Michael A. K. (1967/8) "Notes in transitivity and theme in English: Parts 1, 2 & 3", *Journal of Linguistics* no.3.1, no. 3.2, no. 4.2.
- Halliday, Michael A. K. (1970) "Language Structure and Language Function" in John Lyons (ed.) *New Horizons in Linguistics*, Middlesex: Pelican Books.
- Heim, Irene (1982) *The Semantics of Definite and Indefinite Noun Phrases*, Ph.D. Dissertation, University of Massachusetts.
- Heycock, Caroline (1993). "Syntactic Predication in Japanese", *Journal of East Asian Linguistics* 2:2.
- Hirschberg, J. (1985) *A theory of scalar implicature*, University of Pennsylvania, Ph.D. dissertation.
- Hopper, Paul J. (1979) "Aspect and Foregrounding in Discourse", in Talmy Givón (ed.) *Syntax and Semantics 12: Discourse and Syntax*, 213-219, New York: Academic Press.
- Jacobs, Joachim (2001) "The dimensions of topic-comment", *Linguistics* 39-4, 641-682.
- Jaeger, T. Florian (2002) "The Discourse functions of 'as for': An Indicator of Discourse Subtopic Shifts", Handout and Poster presented at the NWAV 2002, Stanford.
- Labov, William (1997) "Some further steps in narrative analysis", in *Journal of Narrative Life History* 7, 395-415.
- Lambrecht, Knud (1994) *Information structure and sentence form: Topic, Focus, and the mental representations of discourse referents*, Cambridge: Cambridge University Press.
- Lee, Chungmin (1999) "Contrastive Topic: A Locus of the Interface" in K. Turner (ed.) *The Semantics/Pragmatics Interface from Different Points of View*, Amsterdam: Elsevier.
- Li, Charles N. and Sandra A. Thompson (1976) "Subject and Topic: A New Typology of Language" in Charles N. Li (ed.) *Subject and Topic*, New York: Academic Press, 457-490.
- Masuoka, Takashi (1982) "Some Thoughts on the Functions of Subjectiviation in Japanese," *Kansai Linguistics Society* 2, 52-62, Kansai Linguistics Society.

- McNally, Louise (1998) "On Recent Formal Analysis of Topic" in Jonathan Ginzburg (ed.) *The Tbilisi Symposium on Language, Logic, and Computation: Selected Papers*, CSLI/Cambridge University Press 149-162
- Mihara, Kenichi (1994) *Nihongo no tōgokōkozō: sēsēbunporiron to sono ōyō* (syntactic structure of Japanese: generative grammar and its applications), Tokyo: Shohakusha.
- Noda, Hisashi (1996) *Wa to Ga* (Wa and Ga), Tokyo: Kuroshio Publisher.
- Oshima, David Y. (2002) "Contrastive Topic as Paradigmatic Operator" a paper to be presented at Workshop on Information Structure in Context, Stuttgart University; draft available from <http://www.stanford.edu/~davidyo/>
- Portner, Paul and Katsuhiko Yabushita (1998) "The Semantics and Pragmatics of Topic Phrases", *Linguistics & Philosophy* vol.21 no.2, 117-157.
- Prince, Ellen (1981) "Toward a Taxonomy of Given-New Information" in Peter Cole (ed.) *Radical Pragmatics*, New York: Academic Press, pp. 223-255.
- Prince, Ellen F. (1999) "How not to Mark Topics: 'Topicalization' in English and Yiddish" Ch. 8 of *Texas Linguistic Forum*, University of Texas, Austin.
- Rochemont, Michael S. (1986) *Focus in Generative Grammar*. Amsterdam: John Benjamins.
- Schiffirin, Deborah (1994) *Approaches to discourse*, Cambridge: Blackwell.
- Schiffirin, Deborah (1987) *Discourse markers*, Cambridge: Cambridge University Press.
- Schwenter, Scott (1999): *Pragmatics of conditional marking: implicature, scalarity, and exclusivity*; New York: Garland Publications.
- Sgall, Peter (1993) "The Czech Tradition" in Joachim Jacobs, A. von Stechow, W. Sternefeld and T. Vennemann (eds.) *Syntax: Ein internationales Handbuch zeitgenössischer Forschung - An international Handbook of Contemporary Research*, Berlin: de Gruyter, 349-368.
- Sgall, Peter (1975) "Conditions of the Use of Sentence and a Semantic Representation of Topic and Focus" in E.L. Keenan (ed.) *Formal Semantics of Natural Language*, 297-312, Cambridge: Cambridge University Press.
- Sugimoto (1995) "Daishugokōbun to Sōki no kaishaku" (major subject construction and exhaustive interpretation) in Takashi Masuoka, Hisashi Noda, and Yoshiko Numata (eds.) *Nihongo no shudai to toritate* (Topic and Association in Japanese) Tokyo Kuroshio Publisher.
- Takahashi, Chioko (1994) "Case, Agreement, and Multiple Subjects: Subjectivization in Syntax and LF", in Noriko Akatsuka (ed.) *Japanese/Korean Linguistics* vol. 4 . Stanford: CSLI.
- Vallduví, Enric (1992) *The Informational Component*, New York: Garland.
- Vallduví, Enric and Maria Villkuna (1998) "On Rheme and Kontrast" in Peter W. Culicover, Louise McNally (eds.) *Syntax and Semantics 29 the Limits of Syntax*, New York: Academic Press, 241-279.