# Discourse Structure and Sentential Information Structure

*An Initial Proposal*

Livia Polanyi (`polanyi@fxpal.com`)
*FX Palo Alto Laboratory*

Martin van den Berg (`vdberg@fxpal.com`)
*FX Palo Alto Laboratory*

David Ahn (`davidahn@cs.rochester.edu`)[*]
*University of Rochester*

**Abstract.** I this article we argue that discourse structure constrains the set of possible constituents in a discourse that can provide the relevant context for structuring information in a target sentence, while information structure critically constrains discourse structure ambiguity. For the speaker, the discourse structure provides a set of possible contexts for continuation while information structure assignment is independent of discourse structure. For the hearer, the information structure of a sentence together with discourse structure instructs dynamic semantics how rhematic information should be used to update the meaning representation of the discourse (Polanyi and van den Berg, 1996).

**Keywords:** discourse structure, theme/rheme, focus/background, information structure, prosody, Linguistic Discourse Model (LDM)

## 1. Introduction

Although it is generally accepted that a sentence's information structure is determined in large part by its relationship to previous text, the question of how to establish the appropriate discourse context for information structure assignment has never been addressed. Analyses of information structure normally assume that the relevant context is a question and the target sentence is the answer to that question (Vallduvi, 1994). The assumption is that this prior context immediately precedes the target sentence. While this assumption may be a convenient convention for investigating the subtler aspects of intra-sentential information structure, serious problems arise in analyzing actual texts when the critical context for assigning information structure is not part of the immediately preceding sentence. For these cases, it is necessary to have a theory of how discourse structure constrains the choice of possible contexts for determining sentential information structure. This paper addresses how discourse structure and information structure

are related and sketches an integrated approach to the phenomena of discourse continuity.

In brief, our claim is that discourse structure constrains the set of possible constituents in a discourse that can provide the relevant context for structuring information in a target sentence, while information structure critically constrains discourse structure ambiguity. To develop an argument to support this claim, we rely upon the Linguistic Discourse Model (LDM) as a theory of discourse structure (Polanyi and Scha, 1984), and follow Steedman (1991, 2000) in assuming that sentence-level information structure consists of a *theme* and *rheme*, each further subdivided into *background* and *focus*.

## 2.   Discourse Structure

Under the LDM, a discourse consists of a sequence of basic discourse units (BDUs), which are parsed incrementally to form a discourse tree with BDUs as leaves. A sentence may be made up of one or more BDUs. Incremental tree construction is carried out according to a set of discourse construction rules. Which particular rule applies is determined by using information in the surface structure of the incoming unit, possible genre constraints and inference over the meaning of the utterances. More precisely, each (possibly incomplete) sentence of the discourse is parsed individually, and the different units that constitute the sentence are determined. Once the sentence segmentation has taken place, an appropriate discourse rule applies to attach the sentence as a whole somewhere along the right edge of the discourse tree. Once attached as a unit to the tree, the sentence loses its identity as a sentence. The units contribute as units to the discourse tree as a whole.

That an incoming utterance can only be attached to the right-edge of the tree, often referred to as the right-edge constraint, is by now uncontroversial. Only nodes on the right edge have a span that includes the end of the discourse so far. Any other attachment would result in a tree in which the span of the nodes does not correspond to the linear order of the utterances in the text. Through the right-edge constraint, the tree encodes a natural notion of accessibility: the nodes along the right edge are the only nodes accessible for continuation, and antecedents for the resolution of anaphors and other informational linking are also only available if they are visible along the right-edge[1].

---

[1] This constraint does not apply within the sentence or to deictic anaphors. In particular, deictically used definite descriptions can refer to an entity in an otherwise inaccessible position if they uniquely determine an entity in the speaker and hearer's model of the world.

Because the meaning of the preceding material is not available on the right edge, an anaphor cannot refer back to an antecedent in a preceding subordinated constituent.

In the LDM, all nodes are first class objects with a common structure[2]. The leaf nodes contain semantic and structural information, which is calculated based on traditional syntactic and semantic theories, while non-terminal nodes contain semantic and structural information calculated in terms of their children and the particular discourse rule that created them[3].

The main discourse rules are *coordination* and *subordination*. In the case of a coordination, all children contribute equally to the structure of the constructed node. In the case of a subordination, only the subordinating node contributes to the structure, and the structure of the subordinated child does not play a role. Besides these two constructions, all languages have a number of ad hoc constructions, collectively referred to as *n-aries* including question-answer pairs (1a), greetings (1b) and preposed context specifiers (underlined in 1c).

1a  *Who proved completeness? John did.*

1b  *Hi John, how are you? Very well, thanks.*

1c  *Depending on the season, noodles might be served in a hot soup or cold like a salad.*

All n-ary constructions have well-defined ways of calculating the constructed node. In this paper, we will restrict our attention to discussing the general cases of coordination and subordination[4].

## 3. Information Structure

In this paper, we do not want to argue for a particular theory of information structure. We follow Steedman in assuming that information structure is assigned as a pair of a theme and rheme, each further

---

[2] In other places we have referred to BDUs and non-terminal nodes together as discourse constituent units (DCUs).

[3] Information structure information such as which elements are thematic or rhematic may also be represented at non-terminals in some cases.

[4] Traditionally, non-terminals are labeled with the discourse relation (coordination, subordination, or n-ary) that was used to construct them. However, it is important to note that the labels have no intrinsic meaning. They are only a convenient shorthand to be able to draw manageable trees. In particular, construction rules cannot depend on the label of the node that an incoming unit is attaching to.

subdivided into a background, expressing old or given information, and focus, expressing information that is new[5] or at least expressed in a new way (Steedman 1991, 2000). We also follow Steedman in assuming that themes and rhemes can cross traditional syntactic boundaries, although they do conform to the more relaxed notion of constituency inherent in Combinatory Categorial Grammar. We depart from Steedman and from many other researchers, however, in allowing information structure to be assigned to units smaller than the sentence. As we will see in our example, sub-sentential BDUs are sometimes, but not always, the appropriate unit for information structure assignment. How the appropriate unit for information structure assignment is determined is still an open question and beyond the scope of this paper.

In the context of the discussion in this paper, an utterance is divided into four possibly discontinuous or empty parts:

$$(\text{theme-background THEME-FOCUS})_\theta \text{ rheme-background}$$
$$\text{RHEME-FOCUS}$$

SMALL CAPS denote the focused elements. We surround the theme with parentheses indexed with $\theta$. We refer to themes as $\theta$ and rhemes as $\rho$.

## 4.  Discourse Structure and Sentence Topic

From a discourse perspective, the information structure of an incoming unit divides it into a theme, which must be linked back to the preceding discourse, and a rheme, which need not be. Establishing a link between the theme of the main clause of a new sentence and information available at an accessible node in the tree determines the sentence's attachment point. The type of attachment, coordination, subordination, or n-ary, reflects the theme's relation to the information structure of the BDU represented at the attachment node.

Like other types of structure that individual units contribute to the discourse tree, information structure is inherited by the non-terminal

---

[5] In this paper, we assume that *new* means new with respect to what is accessible in the discourse and applies to discourse referents and information about them, and not to entities. A discourse referent can be new even if the entity it refers to has been mentioned before. In particular, a new discourse referent to an entity has to be introduced if earlier discourse referents referring to that same entity are inaccessible. Also, a new discourse referent may be introduced if this entity denotes a subset or an element of an earlier introduced discourse referent or constitutes a different perspective on it.

nodes. A C-node inherits the generalization of the themes of its constituent nodes and of their rhemes. An S-node inherits the information structure of its subordinating daughter directly.

We assume that the following holds:

1. When linking a new unit into the discourse, the theme of the unit is used to decide where and how the new constituent should be integrated into the discourse tree. The decision for an attachment point includes comparing the theme of the incoming unit with the themes and the rhemes of accessible nodes along the right edge of the tree. The rheme of the new unit is necessary to resolve some remaining ambiguities, in particular the choice of discourse construction that obtains.

2. The focused elements in both the theme and the rheme introduce new discourse referents or new information about the theme. In the theme, which links up with earlier material, focused elements are often derived from an antecedent whereas in the rheme, the focused elements are typically either truly new to the discourse or require re-introduction because earlier mentions have been rendered inaccessible.

As an initial proposal, we suggest that there are five ways the theme of a new node can relate to the information structure of preceding text. In the following, we will denote the theme and rheme of the discourse node we are considering attachment to as $\theta_d$ and $\rho_d$ respectively, and the theme and rheme of the utterance to be attached as $\theta_u$ and $\rho_u$.

**theme-theme chaining**     In theme-theme chaining the discourse node is constructed on the basis of a relationship between the themes.

1. If $\theta_u$ essentially repeats $\theta_d$, and $\rho_u$ is derived from $\rho_d$, the node is attached as a subordination.

   *(John)$_\theta$ is a nice looking guy. (He)$_\theta$ has blond hair.*

2. If $\theta_u$ essentially repeats $\theta_d$, and $\rho_u$ is not derived from $\rho_d$, the node is attached as a subordination. the construction is a coordination.

   *(John)$_\theta$ is a nice looking guy. (He)$_\theta$ works for a bank.*
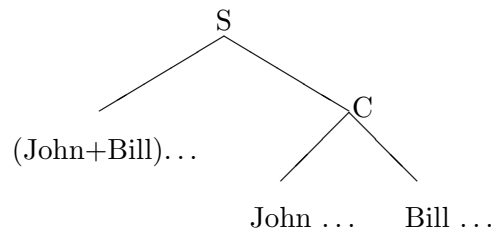
3. If $\theta_u$ does not essentially repeat but is derived from $\theta_d$, the discourse relation is a subordination.

   *(John and Bill)$_\theta$ are nice looking guys. (John)$_\theta$ works for a bank.*

4. If $\theta_d$ and $\theta_u$ are not directly related to each other, but both are related in a similar way to a dominating subordinating node, then the construction is a coordination, as illustrated by the last two sentences of the following example:

*(John and Bill)$_\theta$ are nice looking guys. (John)$_\theta$ works for a bank. (Bill)$_\theta$ plays tennis.*

These sentences are coordinated because they bear a similar relationship to the dominating sentence: *John and Bill are nice looking guys.*

```
              S
            /   \
           /      C
          /      / \
   (John+Bill)...      
              John ...   Bill ...
```

**theme-rheme chaining**     In theme-rheme chaining, the discourse node is constructed based on a relationship between the new theme and a preceding rheme.

5. If $\theta_u$ is derived from $\rho_d$, the discourse relation is a subordination. *(John and Bill)$_\theta$ work for a bank in Palo Alto. (Palo Alto)$_\theta$ is a town in Silicon Valley.*

Although the notions of focus and background are not explicitly mentioned in these rules, they play a role indirectly through the conditions of the form *B essentially repeats A* and *B is derived from A*. We will assume a model of focus as outlined in (Steedman 1991, 2000), where the background forms an open proposition, and the focused elements fill positions in that proposition. Ignoring theme-rheme structure for the moment, the following example gives a simple illustration of this:

2a  MATHEMATICIANS *proved* THEOREMS

2b  $\lambda y \lambda x(\texttt{proved}(x,y))(\texttt{mathematicians}, \texttt{theorems})$

3a  MAURICE *proved* COMPACTNESS

3b  $\lambda y \lambda x(\texttt{proved}(x,y))(\texttt{Maurice}, \texttt{compactness})$

We say that **B essentially repeats A** if, after all the anaphoric elements are filled in, B has the same meaning as A.

We say that **B is derived from A** if, after all the anaphoric elements are filled in: (1) the background of A and the background of B are equivalent, and (2) the focused elements of B are more specific or more general than the corresponding elements of A[6]. For example, (3) is derived from (2), because Maurice is a mathematician (more specific), and compactness is a particular theorem (more specific).

In analyzing a discourse, discourse syntax assigns each incoming sentence its place in the emerging discourse tree. In current approaches, lexical information, syntactic and semantic structure, tense and aspect, and world knowledge are used to infer the attachment point and relation (Lascarides and Asher, 1993). However, after exploiting these resources, attachment ambiguities often still remain. Given that normal language users seldom experience discourse attachment ambiguities, additional sources of information must be used in attachment decisions. We believe that the information structure of both the incoming sentence and accessible DCUs provides information critical for disambiguation.

The relationship between discourse structure and information structure is not a one-way street, however. The surface structure of a sentence, relying on prosody and syntactic constructions, encodes its information structure only partially. It is often necessary to consider the context to determine information structure. Discourse structure provides a constraint on the choice of the appropriate context—as in anaphora resolution, the context for information structure assignment must be along the right edge of the tree and therefore accessible (Polanyi and van den Berg, 1999). Even more strongly, if an incoming unit can be placed into the tree on the basis of factors other than information structure, the node to which it is attached provides the context for assigning information structure. In cases in which there are both attachment and assignment ambiguities, the choices for each kind of structure provide mutual constraints which narrow down the possibilities for attachments. This is similar to how anaphora resolution and discourse structure have been shown to provide mutual constraints in Polanyi and van den Berg (1999).

---

[6] It seems safe to assume all elements in B should either be more specific or equal, or all elements in B should be more general or equal. It seems unlikely that in one case some are more specific and some others more general.

## 5.   Analysis of an example text

We have chosen as an example text a short exposition on Japanese noodle eating[7]

> *Japanese people often eat noodles. Noodles are usually eaten for lunch or a light snack. Depending on the season, noodles might be served in a hot soup or cold like a salad. When noodles are served in a hot soup, vegetables, tofu, and meat are also used for the soup. There are several types of noodles that people eat in Japan. Udon are thick, white noodles made fresh from wheat flour and are usually served with a hot soup. Soba are thin, buckwheat noodles which are firmer than udon. They can be served in a soup like udon, but are usually served as a cool dish in the summer. Ramen are very thin, curly wheat noodles served as a quickmeal or a late night snack. In Japan, noodles lend variety to daily meals.*

In the example, themes are marked with a $\theta$; rhemes are unmarked. Words receiving stress are in SMALL CAPS.

1. Japanese people often eat NOODLES.

2. (Noodles are USUALLY eaten)$_\theta$ for LUNCH or a LIGHT SNACK.

3. (Depending on the SEASON)$_\theta$, (noodles might be served)$_\theta$ in a HOT SOUP or COLD like a salad.

4. (When noodles are served in a hot SOUP)$_\theta$, VEGETABLES, TOFU, and MEAT are ALSO used for the soup.

5. (There are)$_\theta$ several TYPES of noodles (that people eat in Japan)$_\theta$.

6. (UDON)$_\theta$ are THICK, WHITE noodles made fresh from wheat flour and are USUALLY served with a hot SOUP.

7. (SOBA)$_\theta$ are THIN, BUCKWHEAT noodles which are FIRMER than udon.

8. (They can be served)$_\theta$ in a SOUP like UDON, (but are USUALLY served)$_\theta$ as a COOL dish in the SUMMER.

9. (RAMEN)$_\theta$ are very thin, CURLY wheat noodles served as a QUICK meal or a LATE night SNACK.

---

[7] Example text adapted from a slightly longer text that was published on the Web as a student webpage produced in the Fall 2000 class, "Introduction to Japanese Culture" given by Professor Shigeru Miyagawa at MIT. The most significant change made in the text was the insertion of sentence (5) which was introduced in order to illustrate some aspects of the LDM that could not be shown in the original text.

10. (In Japan)$_\theta$, (noodles)$_\theta$ lend VARIETY to daily meals.

From a discourse point of view, the first sentence of any text is unproblematic. The first sentence, composed of one or more BDUs, is attached as a unit to the initial null-tree. This is what we have done, then, with sentence (1). As the introduction to this discourse, sentence (1) is entirely rhematic; since there is no preceding discourse to link to, nothing in sentence (1) could serve as a link. Much of sentence (1) does, however, serve as background. The fact that Japanese people often eat noodles of a specific type while new to the discourse, is not informative, in the sense that we from general world knowledge know that humans all over the world eat one or another type of food. Thus, the nontraditional constituent[8] *Japanese people often eat* is the background of the rheme of sentence (1), and the NP NOODLES is the focus of the rhe

nme.d next sentence, the constituent *Noodles are* USUALLY *eaten* is thematic. This theme is clearly derived from the rheme of sentence (1), and so by the theme-rheme chaining rule 5[9], sentence (2) is subordinated to sentence (1). Given that the theme of (2) and the rheme of (1) are both about noodle-eating, it is clear that (2) elaborates on the rhematic noodle-eating introduced in (1) by specifying that on those occasions when the Japanese eat noodles, they USUALLY are eaten for LUNCH or a LIGHT SNACK.

In sentence (3), further elaborating information is given about how the focused NOODLES in (1), are served: in a HOT SOUP or COLD. Sentence (3) consists of two BDUs: (3a), the preposed temporal adverbial *Depending on the* SEASON; and (3b) the main clause, *noodles might be served in a* HOT SOUP *or* COLD *like a salad*. (3a) specifies the temporal context which determines the choice of preparation style given by (3b). The analysis of (3), however, is not altogether straightforward because the sentence gives rise to a discourse ambiguity. We can interpret this sentence as elaborating on (2) or interpret (3) and (2) as independent elaborations on (1). Note that this attachment ambiguity corresponds to a difference in meaning. If (3) elaborates on (2), what is eaten for lunch or a light snack depends on the seasons. If (3) elaborates on (1), on the other hand, what is eaten depends on the seasons even if the noodles are eaten on an occasion other than lunch or a cold snack.

---

[8] Many of the themes and rhemes and backgrounds and foci we consider in this example consist of strings that would be analyzed as non-constituents in traditional syntactic theories but as possible constituents in a syntactic formalism such as CCG. For the sake of convenience, we will simply refer to these constituents as *constituents*. For further discussion of the relationship between CCG and information structure, see Steedman (1991,2000).

[9] See Section 4, for definitions of the rules.

In either case, (3) attaches to (2), which therefore provides the context for information structure assignment. Relative to that context, what is thematic in sentence (3) is the serving of noodles and its relation to time. Thus, both the BDU *Depending on the* SEASON and the constituent *noodles might be served* are thematic. The class of time intervals referred to in (3)—*seasons*—while related to the class referred to in (2)—*mealtimes*—is nonetheless distinct from it, which is why SEASON is focal. The remainder of the sentence, then, is rhematic. For the purpose of this paper, we have chosen here to interpret (3) as an elaboration on (1), so (3)'s rheme is independent of (2)'s rheme. Thus, since the themes of (2) and (3)—*Japanese noodle-eating*—are essentially identical and the rhemes are independent, (2) and (3) are linked with each other, via rule 2, by a coordination relation at a new node dominating both sentences. This new node is labeled with the common information and is, itself, immediately dominated by (1), as shown in in Figure 1a.

Like sentence (3), sentence (4) consists of 2 BDUs: a preposed adverbial, (4a), *When noodles are served in a hot* SOUP; and the main clause, (4b). Unlike sentence (3), all of sentence (4), rather than any of its embedded BDUs, serves as the unit of information structure assignment—the theme-rheme division corresponds exactly to the discourse segmentation. (4a) is entirely thematic and simply elaborates on one of the choices—hot soup—offered in the rheme of sentence (3). Because the theme of (4) is entirely derived from the rheme of (3), by rule 5, (4) is subordinated to (3) on the discourse tree (Figure 1b).
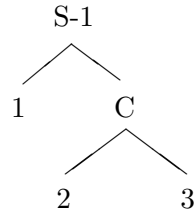
```
        S-1                              S-1
       /   \                            /   \
      1     C                          1     C
           / \                              / \
          2   3                            2   S-3
                                              / \
                                             3   4
```
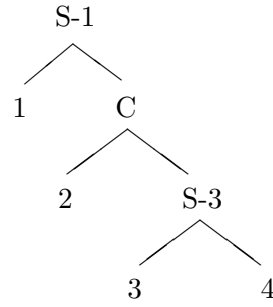
*Figure 1a.* Sentences (1) to (3)        *Figure 1b.* Sentences (1) to (4)

```
              S-1
             /   \
            1     C
                 /|\
                2 S-3 5
                 / \
                3   4
```
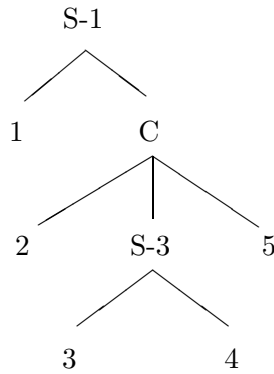
Figure 1c. Sentences (1) to (5)

Sentence (5) is an existential sentence, and following (Milsark, 1977), we take it to have two separate VP arguments: the weak NP *several* TYPES *of noodles* and the predicative coda constituent *that people eat in Japan.* Weak NPs usually correspond to unfamiliar or newly introduced entities; thus, the NP argument is rhematic. The embedded PP *of noodles*, which is not new information, serves as background to the focused TYPES. The predicative coda argument, which is a separate BDU, consists entirely of old information and is thus thematic. We have marked the expletive *There are* as thematic.

The theme of sentence (5) is not related to the elaborating discussion of soup begun in (4), nor to the discussion of styles of preparation begun in (3) and still accessible along the right edge of the discourse tree. Instead, the theme of (5) is derived from the rheme of sentence (1). (5) specifies in its rheme that there are several TYPES of the category of noodles eaten in Japan, a category which is introduced in the rheme of (1) and repeated in the theme of (5). (1), of course, is still available along the right edge of the tree. Because (5) bears a similar

relation to (1) as (2) and (3), (5) is attached as a third item in the coordination of BDUs (2) and (3) (by rule 4). By attaching BDU (5) under the node dominating BDUs (2) and (3), BDUs (2), (3) and (4) are rendered inaccessible (Figure 1c).

Sentence (6) consists of 3 BDUs: (6a) UDON *are* THICK WHITE *noodles*; (6b) [UDON] *are made fresh from wheat flour*; and (6c) [UDON] *are* USUALLY *served with a* HOT SOUP. From a discourse perspective, the relation of BDU (6b) to (6a) is one of elaboration, because (6b) gives more information about the udon introduced in (6a). The NP UDON introduces one element of the set of types of noodles introduced in sentence (5) and is thus thematic. Because the particular type is unfamiliar, however, it is also focused. The remainder of (6a) and all of (6b) and (6c) serve as the rheme of the sentence. Because its theme is derived from the rheme of (5), (6) is subordinated to (5) by rule 5. Note that having one type of noodle being introduced in focus raises the expectation that this is the first element of a list of different types of noodles (Figure 2a).
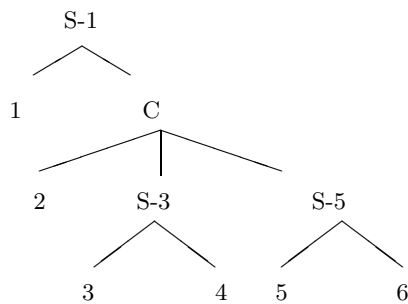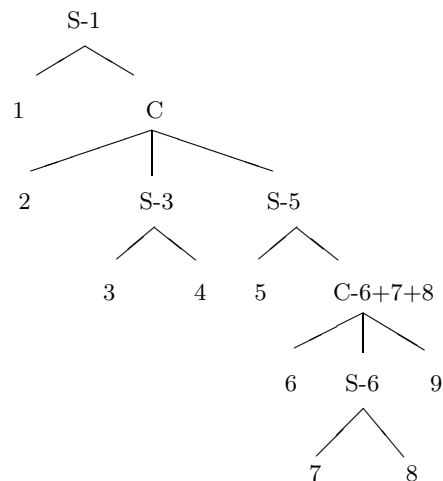


*Figure 2a.* Sentences (1) to (6)          *Figure 2b.* Sentences (1) to (9)

In line with this expectation, the theme of (7) is another type of noodle, soba, and because it is also unfamiliar, it is focused. The rheme of (7) consists of the remainder of the sentence, both the predicate *are* THIN, BUCKWHEAT *noodles* and the subordinate clause *which are* FIRMER *than udon*, which is a separate BDU that elaborates on the embedding BDU. The themes of (6) and (7) relate to sentence (5) in

the same way but are not directly related to each other. They are coordinated (by rule 4).

(8) elaborates on (7) because the information in (8) concerns serving styles for soba. (8) consists of two BDUs: (8a) *They can be served in a* SOUP *like* UDON and (8b) *but are* USUALLY *served as a* COOL *dish in the* SUMMER. This provides us with our first example of a sentence which does not serve as the unit of information structure assignment; instead, each BDU receives its own information structure assignment. *They can be served* is the theme in (8a), while *in a* SOUP *like* UDON is rhematic. The theme of the contrastive clause (8b), *but are* USUALLY *served* is derived from the theme of (8a), with the addition of the contrastive focus USUALLY. The rheme of (8b) also contains contrastive foci: COOL and SUMMER. (8a) and (8b) are joined by an n-ary relation triggered by the connective *but*, and because the shared theme of that n-ary—serving styles for soba—is derived from the theme of (7), (8) as a whole is subordinated to (7) by rule 3.

Sentence (9) is also a complex sentence composed of multiple BDUs. Information structure, though, is assigned to the sentence as a whole. The thematic information in (9) consists of the focused term RAMEN which contrasts with UDON in (6) and SOBA in (7). The remainder of the sentence, which provides new information about ramen, constitutes its rheme. Because (9)'s theme is related to the types of noodles introduced in (5) in the same way as the themes of (6) and (7), it is coordinated to those two by rule 4. At this point, the discourse tree has the structure shown in Figure 2b.

As competent readers familiar with the type of undergraduate paragraph writing of this sample text, we treat the final sentence in this text as a concluding remark. (10) sums up the information in the entire noodle description.

While we may all accept this analysis as reasonable, we must still justify why we believe this. There are other possible analyses—ones which would be less final, closing off fewer of the nodes in the existing tree for continuation. Faced with the problem of possibly ambiguous discourse attachment, how can we make a strong case for choosing one attachment point for (10) over another?

To address this question, let us briefly adopt the conventional question/answer analysis methodology of current theoretical information structure research. Let us begin by suggesting that the rhematic information in (10), *to lend* VARIETY *to daily meals*, provides an answer to the question *Why are* NOODLES *often eaten in* JAPAN*?* Note that this analytic question has precisely the form of the initial sentence in this text prefaced with the interrogative *why*. (10) therefore, elaborates on (1) because it gives more information about the focused thematic

information in (1) . But there are already accessible nodes in the tree which also represent structures subordinated to (1): namely, the sub-trees (2-9) and (5-9). The question, then, is whether (10) should be attached directly under (1) at a new node above the sub-tree (2-9) or whether it should be integrated as a right daughter under either the node dominating (2-9) or the node dominating (5-9). To answer this question, we must consider the relationship of the rhematic information in (10) to the information available at the nodes dominating (2-9) and (5-9).

To analyze the information at (2-9) and (5-9), we will also turn to the conventional question/answer methodology. Currently, (2), (3), and (5) are dominated by a C-node which is labeled with the information common to its three daughters. All have noodle-eating as their theme: (2) specifies *when* noodles are eaten; (3) specifies *how* noodles are eaten; and (5) specifies *what* types of noodles are eaten. Since when-, how- and what-type responses have nothing in common beyond the fact that they all provide types of specific information, all that is available at the C-node dominating (2-9) is specific aspects of noodle eating. This very high level predicate subsumes the question of why noodles are eaten as well.

In general, however, it is preferable to attach as low as possible. Low attachment closes off the smallest span of text. In this case, the only lower open nodes available are (5) and (9). (10) obviously has nothing to do with (9), which involves specifics of ramen, which leaves only (5). At node (5), the full information from (5) is available, including the only focused information: the rhematic TYPES. The focused rhematic information in (10), *lending* VARIETY, is not a further specification of TYPES, so (10) it is not an elaboration on (5). For similar reasons, neither does (10) dominate (5). Since there is no logical or rhetorical connective that would license an n-ary construction, the only possible relation (10) could bear to (5) would be a coordination relation. But noodles lending variety and there being several types of noodles do not have anything relevant in common, therefore (5) and (10) are not coordinated. The only possibility then appears to be our original pro-posal: (10) should be attached as the right sister of (5) in the C-node immediately dominated by (1).

There is one last possibility however, and this comes not only from the semantics of the constituents, but takes into account genre con-straints as well. If we remember our intuition that (10) sums up the text, then (10) should be attached as a sister node to (1) under a newly constructed dominating node that captures the rhetorical and logical relationship between the focused *noodles* in (1) and the focused *variety*

in (10). *Why do the Japanese often eat noodles? Because eating noodles lends variety to daily meals.*
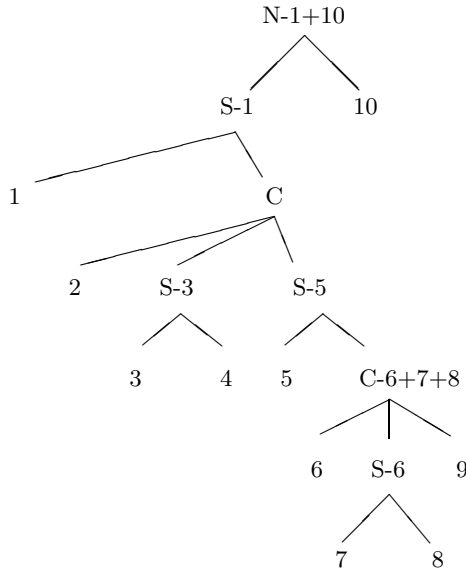


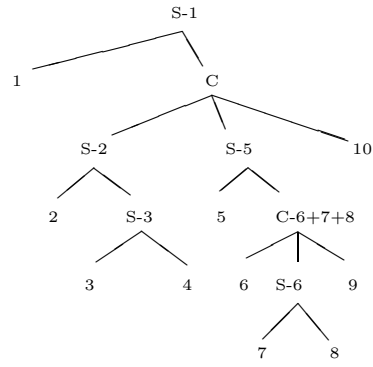*Figure 3a.* Sentences (1) to (10)



*Figure 3b.* Alternative attachments of (3) and (10)

## 6.  Conclusion

Although the assignment of information structure to a sentence depends on the discourse structure, and the construction of the discourse structure may depend on the information structure of the units involved, the dependency between information structure and discourse structure is complementary and not circular. For the speaker, the discourse structure provides a set of possible contexts for continuation while information structure assignment is independent of discourse structure. For the hearer, the information structure of a sentence together with discourse structure instructs dynamic semantics how rhematic information should be used to update the meaning representation of the discourse (Polanyi and van den Berg, 1996). Thus, the relationship between discourse structure and information structure reflects the different but deeply related tasks of speaker and hearer in a communicative situation.

# References

Kruijff-Korbayová, I., 1997, "Czech noun phrases in File Change Semantics", pp. 107–118 in *Proceedings of the Second ESSLLI Student Session.* A. Drewery, G. M. Kruijff & R. Zuber (eds.), Aix en Provence, France.

Lascarides, A. and N. Asher, 1993, "Temporal interpretation, discourse relations and Commonsense Entailment", Linguistics and Philosophy, **16**, 437–494.

Milsark, G., 1977, "Towards an explanation of certain peculiarities of the existential construction in English", Linguistic Analysis, **3**, 1–29.

Polanyi, L. and M. van den Berg, 1999, "Logical structure and discourse anaphora resolution", in *Proceedings of the Workshop on The relation of Discourse/Dialogue Structure and Reference*, D. Cristea, N. Ide, D. Marcu (eds.), 37th Annual Meeting of the Association of Computational Linguistics, College Park, Maryland.

Polanyi, L. and M. van den Berg, 1996, "Discourse structure and discourse interpretation", in *Proceedings of the Tenth Amsterdam Colloquium*, P. Dekker and M. Stokhof, ILLC/Department of Philosophy, University of Amsterdam.

Polanyi, L. and R. Scha (1984) "A syntactic approach to discourse semantics", pp 413–419, *Proceedings COLING10*, Stanford, CA.

Steedman, M., 1991, *Structure and intonation*, Language **68**, 260–296.

Steedman, M., 2000, "Information structure and the syntax-phonology interface," Linguistic Inquiry, **31:4**, 649–689.

Vallduví, E., 1994, "Information Packaging: A Survey, Word Order, Prosody and Information Structure Project Report'," University of Edinburgh.