Information Structure as Information-Based Partition[*]

Satoshi Tomioka

University of Delaware

While the *Information Structure (IS)* is most naturally interpreted as 'structure of information', some may argue that it is structure of something else, and others may object to the use of the word 'structure'. This paper focuses on the question of whether the informational component can have structural properties such that it can be called 'structure'. The preliminary conclusion is that, although there are some vague indications of structurehood in it, it is perhaps better understood to be a representation that encodes a finite set of information-based partitions, rather than structure.

Keywords: Partition, Topic, Recursivity, Second Occurrence Focus

1 Introduction

Let me begin this paper with this candid admission: The term *Information Structure (IS)*, which goes back to Halliday (1967), is perhaps a little confusing. Without any theoretical biases or inclinations, one would most naturally interpret the term as 'structure of information'. I suspect, however, that this way of interpreting it may invite objections from those working on IS and related issues.

Some would argue that IS refers to a representation of linguistic objects that has structural properties. The information itself is not a linguistic object, so it does not make sense to say that IS is the structure of information. The first

---

half of the compound 'information' should, therefore, be interpreted as a modifier of some sort, meaning 'informational', 'information-based', 'information-related' or something along those lines. One of the most prominent advocates of this view is Erteschik-Shir (1997), whose f(ocus)-structure is an annotated S-structure that encodes foreground/background information. This f-structure is meant to replace LF as the input representation to the semantic translation, and there is a finite set of mapping algorithms that connect f-structure and the file-card semantic system fashioned after Heim (1982, Chapter 3). Although similar approaches are found in Vallduví (1992) and Lambrecht (1994), I think that Erteschik-Shir's approach is more 'structural' than its alternatives.[1] As a replacement for LF, f-structure is where scope is computed, and it is determined in structural terms at that level.[2] Erteschik-Shir also provides f-structural (re)analyses of many syntactic phenomena, ranging from extraction out of islands and crossover phenomena to anaphora binding. The existence of subordinate f-structure gives additional hierarchical flavor to f-structure.

While a sizable contingency of IS researchers assume the thesis of 'IS as a linguistic representation', it seems that there are still many others, myself among them, who use the term 'Information Structure' without commitment or belief that it entails the existence of an independent linguistic representation, like Erteschik-Shir's f-structure. For us, IS tells us the state of affairs of information or it says something about how information is organized. Although it is a representation of non-linguistic objects, IS is still considered linguistically

---

[1] Erteschik-Shir compares her approach to the two alternatives mentioned above in her book (Erteschik-Shir 1997: 1.8, p. 55–56).

[2] Unlike in the 'Transparent LF' (a term borrowed from von Stechow), the position of a quantifier at f-structure does not play a deterministic role in scope. For instance, Q1 sits lower than Q2 but takes scope over Q2 if Q1 is co-indexed with a topic. See Erteschik-Shir (1997: 5.3) for more discussion.

relevant because the way information is organized has a significant impact on linguistic structures of different modules. Within this view, one can still object to the idea of 'Information Structure as structure of information', but the objection is not about 'of information' but to the choice of the word 'structure'.

2   Structure vs. Partition

The key notions often cited in connection to IS are essentially bipartite: Theme vs. Rheme, Topic vs. Comment/Focus, Ground vs. Focus, Given vs. New, and perhaps a few others like these. The question is whether this 'informational partition' should be described as structure.

The notion of 'structure' has played a central role in shaping modern linguistic theories. We are indeed quite used to using structure at so many different levels and have come to expect certain properties from it. For instance, (1) is often associated with a linguistic representation that has structural properties.

(1)   The existence of hierarchy, and/or the fixed hierarchical ordering of
         primitives.

There are numerous instances that exemplify (1): the X-bar Schema in Syntax (X^0 – X' – XP) or the Prosodic Hierarchy in Phonology (segment – mora – syllable – foot…). It is also worth pointing out that (1) has led to the emergence of relational notions that are defined in structural terms. C-command in syntax is perhaps among the most recognizable of these. Almost all notions of locality in syntax are structure-sensitive as well.

How about information structure? Does it have any attributes that can be described as hierarchical? Are there any structural notions that are relevant to this level of representation? I suspect that I am not the only one who is inclined

to say 'no' to these questions. The various ways of partitioning in the information component, like those I listed above, do not seem to encode any hierarchy in any obvious way. Nor are there any relational notions for IS that are based on hierarchical structure. One potential exception to this generalization is Vallduví's (1992) version of *Information Packaging,* which partitions information into Focus and Ground, the latter of which is further divided into Link and Tail.

(2)

```
              Focus        Ground
                          Link      Tail
```
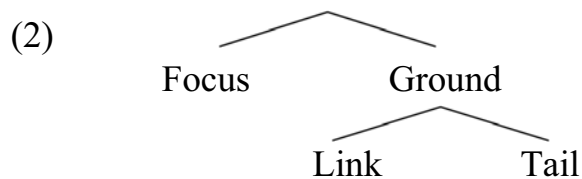
This classification does look hierarchical, but despite its appearance, the representation does not make any use of the hierarchy, nor does it have any isomorphic mapping relations to other representations. For instance, Link can be regarded as structurally lower than Focus in this representation, but such a structural asymmetry does not correlate with semantic scope. The general scope tendency often noted in the literature (e.g., Erteschik-Shir 1997, Krifka 2001) is that the element that corresponds to Link (often equated to a topic) takes the widest scope.

Although informational partitioning itself is not enriched enough to encode structural ordering, IS can still show hierarchical properties if it has another well-known structural property in linguistics, namely (3).

(3)   Recursivity or embeddability of a part of structure within a larger part.

IS seems to fare better with this criterion. One empirical phenomenon that may call for recursivity in information partitioning is the nested foci or the second occurrence focus phenomenon, which has attracted a lot of attention lately (cf.

Krifka 1991, Beaver et al. 2004, Féry and Ishihara 2005, Rooth to appear, Büring 2006).

(4)   A: Bill only eats VEgetables.
        B: FRED also only eats vegetables.

In this example, B's utterance has focus on the subject NP *Fred*, which is associated with the focus-sensitive adverb *also*. Although the VP *only eats vegetables* should be regarded as the background (or whatever one assumes to be the opposite to focus), the semantics of *only* requires the presence of focus-marking on its associate, namely the object NP *vegetables*. This situation can be interpreted as the recursive Focus-Background partition within the matrix background portion, as in (5).

(5)

Focus   Background

Focus   Background
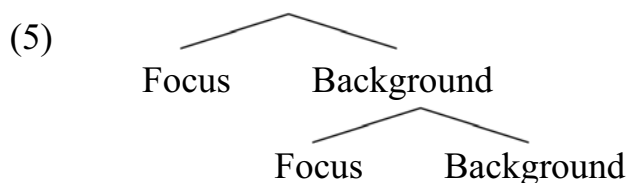
I have been ignoring one important detail here, however. On the one hand, the issue of the second occurrence focus is discussed most frequently in connection with 'association with focus', a popular phenomenon among formal semanticists in which focus affects truth conditions (in the case of *only* or *always*) or presuppositions (in the case of *also* or *even*). The information-based partitions, on the other hand, come from more pragmatic or discourse-oriented frameworks. In other words, the second occurrence focus, which is motivated by the semantics of a focus-sensitive adverb, may or may not be integrated comfortably in the information-based partition in the way that it renders support for the idea of recursivity in IS. Even if we clear up this issue, there is another pressing

question: How many levels of embedding are possible? We can certainly add another layer to the example in (4)..

(6)   A: Bill only eats VEgetables.
        B: FRED also only eats vegetables.
        A: Well, you THINK Fred also only eats vegetables, but actually he quit being a vegetarian.

While there can be more than two layers of partitioning, it is still unclear whether we need to make a distinction between the second occurrence and the third occurrence foci. Should the previous embedding structure be preserved under additional embedding in a way that mimics syntactic embedding

All in all, it can be speculated that the second occurrence focus calls for structural IS, but it presents more questions than answers.

3 Embedded Topics in Japanese

More promising evidence for recursivity or embeddability in the information component is the recursive topic-marking in Japanese and Korean, which have morphological marking for topicality.[3] As the following Japanese example shows, topic-marking with the topic particle *wa* can be reiterated under syntactic embedding.

---

[3] The view that *wa* is the marker of a topic was popularized by Susumu Kuno (e.g., Kuno 1973) but has recently been challenged by Kuroda (2005). One of Kuroda's main arguments is that a *wa* phrase can be used as an answer to a Wh-question, which is often regarded as a sign of being focused. Kuroda is very careful in making his point by eliminating non-exhaustive, partial answers, which he correctly identifies as contrastive uses of *wa*. However, his crucial examples ((11) and (12) on p. 9), judged acceptable by Kuroda, do not get universal approval. All speakers that I consulted (and myself) still find them odd. The common feeling among us is that the sentences themselves are fine, but they are not really answering the question. At this point, I cannot offer anything more substantial and leave this issue for future research.
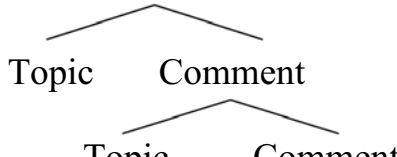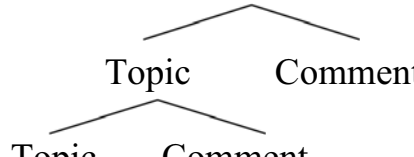
(7)   A: What did Ken say?
        B: Ken-**wa**   [Erika-**wa** baka-da]-to   itta
            Ken-**TOP** Erika-**TOP** fool-COP-COMP   said
            'Ken said Erika is a fool.'

Note to the editor: COP is 'copula'.

The embedded CP in B's sentence is considered focus since it corresponds to the Wh-phrase in A's question. Within the embedded clause, we find another topic-marked phrase. Thus, this will be an instance of Topic-Comment (or Focus) partition within Comment/Focus. It is also possible to have a topic-marked phrase within another topic, as shown in (8).

(8)   [[[Erika-**wa**   baka-da]-to   itta]-no]-**wa**   Ken-da
         Erika-**TOP** fool-COP-COMP   said-one-**TOP**   Ken-be
        'The one who said Erika is a fool is Ken.'

(9a,b) illustrate how these Japanese facts translate into the embedding of informational partitioning.

(9)   a.   Topic   Comment   b.   Topic   Comment
                        Topic   Comment   Topic   Comment

Unlike 'association with focus' cases, the embedded topic-marking in Japanese is not made a victim of the tension between formal semantics and pragmatics. The notion of topicality is firmly grounded in pragmatic/discourse theories that make use of information-based partitions. [4] Topics can be embedded more than once, and there does not seem to be any limit on the levels of embedding. It closely

---

[4]   As a matter of fact, taking 'topic' as a part of information-based partitions is not so popular in formal semantics. Rather, a topic is often interpreted as a question-under-discussion (QUD), as in von Fintel (1994), Roberts (1996), and Büring (2003), among others.

mirrors the syntax of embedding, so unlike the second/third occurrence foci, we can easily talk about some topic being more embedded than another topic.

Before congratulating ourselves that we have found evidence for the 'structurehood' of IS, however, I would like to make some cautionary notes. First, it is not the case that a topic-marking can be found in any embedded clause. Kuroda (2005) and Portner (2004) independently note that *wa*-marking in an embedded clause is possible when there is the presence of an agent of a cognitive act, such as believing, thinking or doubting, or of a speech act, such as saying or reporting, in the embedded clause. Thus, embedding topics are found most typically in complement clauses of verbs of attitude reports. On the other hand, they cannot appear in relative clauses or in certain adjunct clauses (e.g., *when*, *if*, etc.). I am not too optimistic about the prospect that this restriction is derivable entirely from informational properties: Given/New, Topic/Comment, Theme/Rheme partitions cannot be easily used to explain the subtle distinctions among various embedding structures. Second, embedded topics do not share all the characteristics that matrix topics have. Kuroda (1965) observed, for instance, that with an individual-level predicate, the nominative subject necessarily induces the exhaustive interpretation while the topic subject gets the neutral interpretation.

(10) a. John-ga   zurugasikoi
         John-NOM sly
         Exhaustive reading: 'Of all the relevant people, it is John who is sly.'

     b. John-wa   zurugasikoi
         John-TOP   sly
         Neutral reading: 'Speaking of John, he is sly.'

Heycock (1994) accounts for the contrast by making an appeal to the concept of competition between a topic and a nominative. A nominative subject gets a

focalized interpretation when it could have been marked with *wa* but isn't. Interestingly, the obligatory exhaustive reading is not applicable to embedded nominative subjects.

(11)   Erika-wa [CP John-ga   zurugasikoi to]   omo-ttei-ru
        Erika-TOP   John-NOM sly   COMP think-PROG-PRES
        'Erika thinks that John is smart' (Neutral reading possible)

For some reason, the notion of competition between *wa* and *ga* does not arise in embedded contexts despite the fact that embedded topics are possible. This needs to be accounted for, and as far as I know, there has not been any satisfactory explanation proposed.

4   Summary

There are a few signs of structural attributes in Information Structure. All in all, we have to admit, however, that structure in IS is rather rudimentary, and that we have to look hard for any linguistic relevance that such structural characteristics may bring about. The question remains of whether we should continue to allow ourselves to use the term without believing that it is structure. My inclination is that, as long as we share the understanding that IS is a representation, not necessarily structural, where a finite set of bipartite distinctions apply, there is not much harm in using it.

References

Beaver, David, Brady Clark, Edward Flemming, Florian Jäger, and Maria Wolters. 2004. When semantics meets phonetics: Acoustical studies of second occurrence focus. Manuscript, Stanford University.

Büring, Daniel. 2003. On D-trees, beans, and B-accents. Linguistics and Philosophy 26: 511-545.

Büring, Daniel. 2006 Been there, marked that – A tentative theory of second occurrence focus. Manuscript, UCLA.

Erteschik-Shir, Nomi. 1997. *The Dynamics of Focus Structure*. Cambridge: Cambridge University Press.

Féry, Caroline, and Shinichiro Ishihara. 2005. Interpreting second occurrence focus. Manuscript, Universität Potsdam.

Fintel, Kai, von. 1994. *Restrictions on Quantifier Domains*. Doctoral Dissertation, University of Massachusetts at Amherst.

Halliday, Michael Alexander Kirkwood. 1967. "Notes on Transitivity and Theme in English, Part II." *Journal of Linguistics* 3, 199–244.

Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Doctoral Dissertation, University of Massachusetts at Amherst.

Heycock, Caroline. 1994. Focus projection in Japanese. In M. Gonzalez (ed.) *Proceedings of NELS* 24. GLSA, University of Massachusetts at Amherst, 157–171.

Krifka, Manfred. 1991. A compositional semantics for multiple focus constructions. In Joachim Jacobs (ed.) *Informationsstruktur und Grammatik*, Sonderheft der *Linguistischen Berichte*, 17–53.

Krifka, Manfred. 2001. Quantifying into Question Acts, *Natural Language Semantics*, 9.1, 1–40.

Kuno, Susumu. 1973. *The Structure of the Japanese Language*. Cambridge: MIT Press.

Kuroda, Shige-yuki. 1965. *Generative Grammatical Studies in the Japanese Language*. PhD dissertation, MIT.

Kuroda, Shige-yuki. 2005. Focusing on the matter of topic: A study of *wa* and *ga* in Japanese. *Journal of East Asian Linguistics*, 14, 1–58.

Lambrecht, Knut. 1994. *Information Structure and Sentence Form. A theory of topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Portner, Paul (2004) 'The Semantics of Imperatives within a Theory of Clause Types,' in K. Watanabe and R. B. Young (eds.) Proceedings of Semantics and Linguistic Theory 14. Cornell University: CLC Publications.Roberts,

Craige. 1996. Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. In J. H. Yoon and Andreas Kathol (eds.) *OSU Working Papers in Linguistics* 49: *Papers in Semantics*, 91–136.

Rooth, Mats. to appear. Second Occurrence Focus and Relativized Stress F. In Zimmermann, M. and C. Féry (Eds.) Information Structure in different perspectives. OUP.

Vallduvi, Enric. 1992. The Information Component. Doctoral dissertation, University of Pennsylvania

Satoshi Tomioka
University of Delaware
Department of Linguistics
46 E. Delaware Ave.,
Newark, DE 19716
U.S.A.
stomioka@udel.edu
http://www.ling.udel.edu/tomioka