

# Information Structure as Parallel Tree Building

Jon Stevens

July 15, 2011

Information Structure (IS) can be seen as a separate component of grammar (see Vallduví 1990, Vallduví and Engdahl 1996, Eilam 2011, and others), which determines how information is assimilated into a hearer's knowledge store, and which mediates syntax and phonology. Such accounts take the focus-ground distinction to be the most basic distinction made by IS. In contrast, analyses such as that of Rooth (1992) treat focus as a syntactic feature that is interpreted by the semantic component of grammar, and which may have phonological consequences. Accounts like Rooth's give focus a privileged status, capturing the intuition that focused material uniquely serves to answer the question under discussion. In this paper we propose an account that, like Vallduví and others, takes focus and ground to be primitive elements of an informational component, but which gives focus the unique property of receiving a pragmatically determined interpretation.

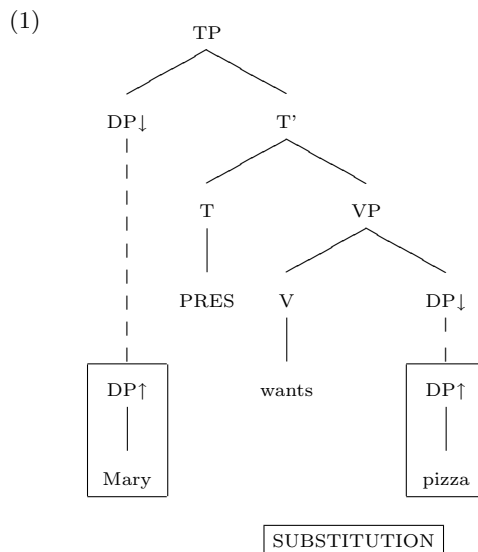
Our basic claim is that linguistic material in the ground must be primed by the context of utterance. We formalize this by building a system in which focus and ground correspond to two different syntactic structures, constructed in parallel. Structure in the ground is interpreted by the semantic component as it is built by selecting the most salient denotation for each constituent given the context. Focused material is interpreted pragmatically, using Gricean principles to select among possible denotations and taking into account the ground material that has already been interpreted. Although the main ideas behind this proposal are independent of formalism, we believe that a minimally extended Tree Adjoining Grammar is elegantly suited to the task.

There are at least two advantages to this approach. First, it successfully constrains unified models of semantic-pragmatic interpretation like those outlined in Parikh (2010) and Clark (2011), eliminating both redundancies and false predictions from these otherwise attractive frameworks. Second, it can account for data that is problematic under other analyses of IS.

After briefly reviewing some basics of the Tree Adjoining Grammar formalism, we outline the proposal and offer some linguistic data that we believe the proposal accounts for nicely.

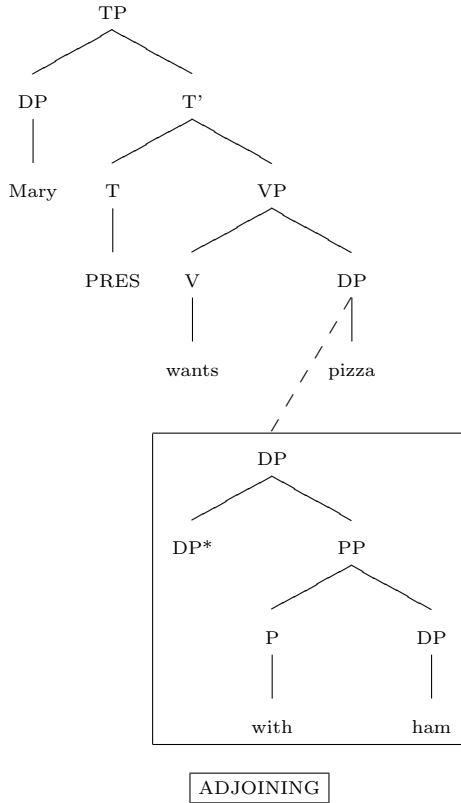
## 1 Tree Adjoining Grammar

Tree Adjoining Grammar (TAG) is a mildly context-sensitive (Joshi 1985) grammar formalism in which complex tree structures are built up from atomic units called Elementary Trees, via two operations: Substitution and Adjoining. In the illustration below we see the elementary tree for the present tense verb *wants* being supplied with two DP arguments by substituting the DPs *Mary* and *pizza* in for the empty DP argument nodes.



Simple sentences are built up this way, inserting argument constituents into lexically determined verbal structures. The Adjoining operation inserts structure into a tree by splitting a node and performing two substitutions. In the following example, the DP node dominating *pizza* is pulled apart from the main tree, at which point the structure  $[_{DP} DP [_{PP} \textit{with ham} ]]$  is substituted for the direct object DP node of *wants*. Then, the separated DP *pizza* is substituted in for the sister DP of *with ham*, creating the structure  $[_{DP} \textit{pizza} [_{PP} \textit{with ham} ]]$ . This transforms the sentence *Mary wants pizza* into *Mary wants pizza with ham*.

(2)



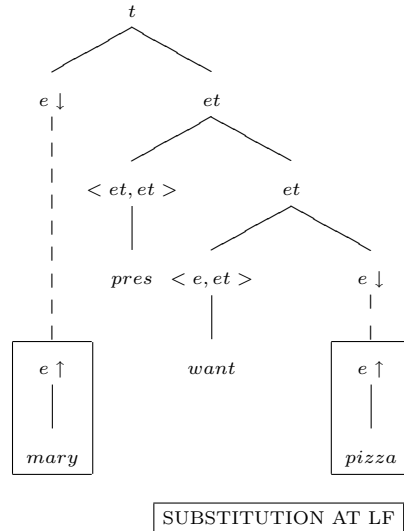
The mildly context-sensitive status of TAG gives it enough power to derive crossed dependencies (via Adjoining), but is more constrained than other context-sensitive systems (Joshi 1985). Also, the formalism has proven to have advantages in deriving certain locality phenomena that are found in natural language (Kroch and Joshi 1985, Frank 2002). As we see from the examples given above, the Elementary Trees of TAG are highly lexicalized. Proposed derivational operations such as movement are accounted for within a TAG framework by constraints on the inventory of Elementary Trees in a language. These meta-constraints may themselves be modeled with a grammar formalism, such as a Minimalist Grammar (Frank 2002).

Schabes and Schieber (1990) propose Synchronous TAG (STAG) to formalize the isomorphism between syntax and semantics. Simply put, a STAG formalism builds a logical form (LF) for a sentence as a separate tree with nodes that are “linked” to nodes in the syntactic tree. Every Substitution or Adjoining operation that affects a particular node on the syntactic tree must analogously affect its linked node on the LF tree. So, substituting *Mary* and *pizza* in for the DP arguments of *wants* is necessarily

<sup>1</sup>Note that secondary stress on *fisherman* and primary stress on *lives* is also a possible prosodic configuration for (5). It is this configuration that would correspond to the tensed verb being in focus along with the subject. Assuming a systematic alignment between IS and prosody in English, the existence of these options shows that it is felicitous, though not mandatory, to put the verb *lives* in the ground in this context.

accompanied by the substitution of those constituents’ denotations into the LF tree corresponding to *wants*, which is shown below.

(3)



In the proposal that follows, the focus and the ground of a sentence are built up separately, and only the ground material receives an interpretation within a STAG structure. The focused material remains uninterpreted until a special Unifying operation allows it to receive a denotation that is felicitous given the semantic material already present in the ground.

## 2 Parallel Tree Building

We take the ground of a sentence to be that material which is immediately interpretable within the context of the utterance of that sentence. By immediately interpretable, we mean that there is a single associated meaning that is primed, or easily retrievable from the common ground of the interlocutors within that context. It follows that if there is a particular question under discussion, linguistic material that refers back to that question will tend to be part of the ground. Thus, we have the famous question-answer pair diagnostic for focus and ground.

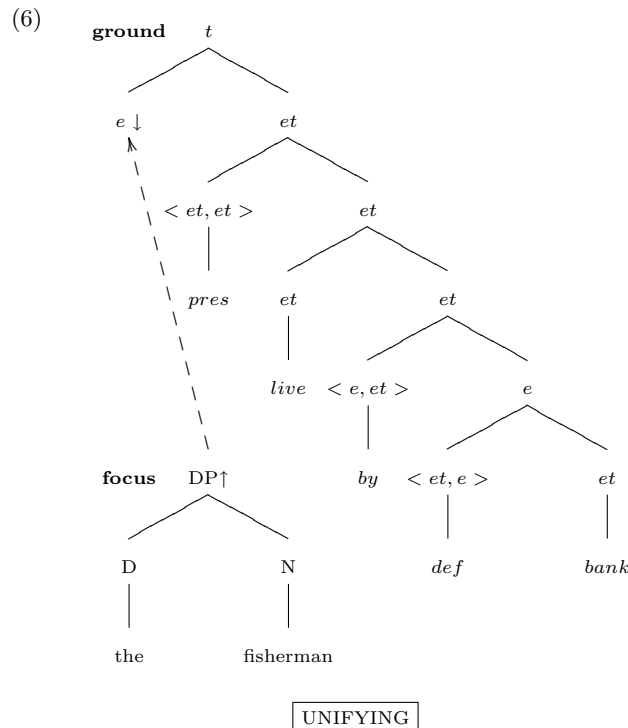
- (4) Q: Who lives by the bank?  
A: [<sub>F</sub> The FISHERMAN ] [<sub>G</sub> lives by the bank ] .

Note, however, that narrow focus does not need to align exactly with the question under discussion.<sup>1</sup>

- (5) Q: Why do you want to visit the bank?  
A: [<sub>F</sub> The FISHERMAN ] [<sub>G</sub> lives by the bank ] .

Here, the response is indirect, with the implicature ‘I want to visit the fisherman’, and yet this response can carry strong accent on *fisherman*, signaling narrow focus on the subject. Although the verb *lives* is not pulled directly from the question under discussion, it is allowed to be part of the ground because it is immediately interpretable given the context of wanting to visit somewhere. After all, having a friend who lives in a particular place is a common reason for wanting to visit that place. Note that if we replace *lives* with a contextually unusual verb like *dances*, accent on the verb becomes mandatory. We formalize this idea within a TAG framework by constructing a separate syntactic tree for the ground and linking it to an LF tree. This creates an isomorphism between LF construction in a STAG structure and speakers’ selection of immediately available interpretations for background material.

Because we take synchronous LF construction to be isomorphic to immediate interpretation, and because the focus, under our analysis, is not immediately interpretable, the focus tree in our parallel structure is not linked to an LF tree. Instead, the focus tree remains uninterpreted until the two trees are joined. Like rounding off a square peg, there must be an operation that takes the syntactic material in the focus and assigns it a denotation as it is incorporated into the ground. We call this operation Unifying.



Since this operation is distinct from the STAG operations that interpret the ground, it does not need to be isomorphic to immediate interpretation. Formally, this opera-

tion could assign any number of interpretations to the focused material, as long as there is no type mismatch. Crucially, however, speakers of natural language have a mechanism in place to determine the most likely interpretation given (a) the possible conventional meanings of the focused words, (b) the interpretation of the ground, and (c) the context in which they are situated, including their beliefs about the intentions of their interlocutors. In other words, the denotation of what is in focus is determined pragmatically. This is exactly the kind of interpretive model proposed by Parikh (2010), Clark (2011), and others who work within the growing framework of game theoretic pragmatics. In the following section we illustrate how this interpretive process works and why using IS to constrain it is advantageous, before looking at more examples.

### 3 Advantages

#### 3.1 Constrained Pragmatics

Consider the broad focus counterpart to the sentence we’ve been dealing with, *the fisherman lives by the bank*. Let’s first construct a context. On a recent riverboat cruise, an anonymous fisherman saved you from falling into the river after you slipped and fell over the railing on the deck. You wished to compensate him with money, but he vanished before you could get his contact information. Weeks later, you receive a voicemail from a friend claiming that she has something interesting to tell you. You call back, and the following exchange occurs.

- (7) Q: What did you want to tell me?  
A: [<sub>F</sub> The fisherman lives by the BANK ] .

You may take this statement to mean that the fisherman who saved you from falling lives close to a nearby section of the bank of the river you were on, with the implicature that you should spend some time there in order to locate and reward this mysterious fellow. However, you could also take it to mean that he lives close to the financial institution that both you and your friend have dealings with. Within a game theoretic pragmatics like that of Parikh (2010) or Clark (2011), disambiguation is determined by the hearer’s reasoning about the speaker’s intent given the context of utterance, which is modeled as a cooperative game between the speaker and the hearer. Essentially, the speaker wants to say as little as possible and still be understood. In order to play this game, the interlocutors must weigh probabilities for lexical meanings within the context.

In most contexts, the ‘financial institution’ reading for *bank* is considerably more common than the ‘riverbank’ reading. Hence, the speaker should specify *riverbank* for the latter in order to avoid misunderstanding. However, context can boost the probability of one meaning (like

if we're talking about rivers). Furthermore, surrounding linguistic material in a sentence can also alter the probability of a particular meaning. In the example above, the presence of *fisherman* will make the 'riverbank' interpretation more likely because the co-occurrence of fishermen and riverbanks is sufficiently high. Other contextual factors can boost the likelihood such that, in some situations, 'riverbank' will be the preferred reading for the word. Thus, the speaker is justified in using the shorter word, *bank*, rather than being more specific. This same sort of reasoning is the source of pragmatic enrichment: the hearer will reason that the speaker is referring to an easily reachable section of the riverbank because to do so results in a more useful message.

The crucial claim underlying a system like that of Parikh (2010) is that pragmatic principles are at play at each step of a semantic derivation. Referents are chosen rationally, denotations of individual constituents are enriched by context, and pragmatically determined material feeds the compositional derivation of an intended proposition, which in turn determines implicatures.

Our claim is that this is only half-true. These pragmatic principles apply *only* to the interpretation of what is in focus. The ground is pre-interpreted directly from context; none of the accompanying linguistic material can have an effect.<sup>2</sup> This makes for a more efficient system. In the broad focus declarative in (7), the presence of *fisherman* may make the use of *bank* in place of the unambiguous *riverbank* more reasonable. However, in the narrow focus counterpart seen in (4) and (5), the hearer does not need to consider the presence of *fisherman* when disambiguating *bank*. Rather, the disambiguation must already have taken place. Otherwise, the utterance is infelicitous. To put it in Gricean terms, placing narrow focus on the subject when the object is not easily disambiguated from the preceding context is a flagrant violation of the Maxim of Manner.

In addition to simplifying a theory of pragmatically enriched interpretation, our account makes better empirical predictions. Let's modify the dialog in (7) to make the two interpretations in question more or less equiprobable.

- (8) Q: What did you want to tell me?  
 A: Two things. First, I'll need to take you up on that loan offer. The bank wouldn't give me one.  
 Q: What else?  
 A: I found out something interesting. [<sub>F</sub> The fisherman lives by the BANK ] .

How is this last sentence to be interpreted? It seems unnatural, maybe even infelicitous, not to be more specific about the whereabouts of the life-saving fisherman, but the 'riverbank' meaning is not completely ruled out. Con-

trast this with the same dialogue, but with narrow focus on the subject in the last sentence ([<sub>F</sub> The FISHERMAN ] [<sub>G</sub> lives by the bank ]). This IS configuration *cannot* be used to convey the 'riverbank' meaning. By not addressing IS at all, Parikh's (2010) system does not predict any difference in interpretation between these two configurations. Our proposal, on the other hand, makes the correct prediction: in the case of narrow focus on *the fisherman*, the backgrounded word *bank* must be interpreted with respect to the preceding context, independent of the other linguistic material in the sentence. The result is that the most obvious interpretation given the preceding discourse, 'financial institution', is chosen, and the alternative ruled out.

### 3.2 Examples

Defining focus and ground solely in terms of the interpretive mechanism breaks with recent work on IS in that the focus/ground distinction is not explicitly tied to question-answer pairs or open propositions under discussion. This can be a valuable diagnostic, as in the following.

- (9) Q: Who shot J.R.?  
 A: [<sub>F</sub> KRISTIN ] shot J.R.

The question implies an open proposition *shot(x, JR)*, and the focused element of the answer provides the answer. As mentioned, however, there is a little bit more to the distinction than that. Consider the following.

- (10) Q: Why do you know so much about Miami University?  
 A: [<sub>F</sub> My MOTHER ] went to Miami.
- (11) (Out of the blue, fiddling with a cell phone)  
 I [<sub>F</sub> HATE ] this phone.

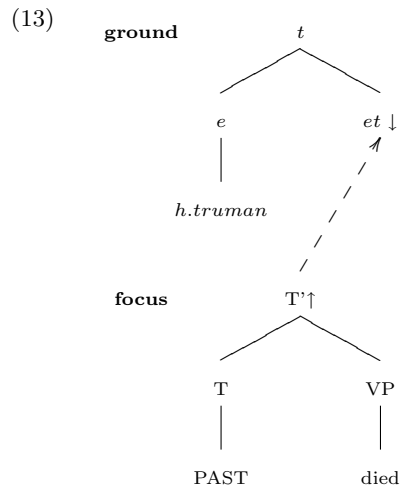
In both of these cases there is some semantic material (the denotations of *went* and *this phone*) that is not part of an open proposition or question that is explicitly known or under discussion, but is backgrounded nonetheless. Similarly, consider Schmerling's famous (1974) contrast between two answers to the question, "what happened?"

- (12) JOHNSON died.  
 Truman DIED.

The default intonation for a broad-focus sentence of the form *x died* has prominence on the subject. We see a non-default intonation in the utterance, Truman DIED. This, Schmerling argues, is a product of the context: Truman had been in the news at the time of the utterance due to his failing health, and so it was no surprise to hear bad news about him. This licensed the backgrounding, and with it the de-accenting, of *Truman*. Johnson's death, on

<sup>2</sup>This is not to say that speakers cannot repair infelicity by reconsidering an utterance without this restriction. Perhaps they can, but this would represent the post-hoc application of higher-level reasoning, rather than something internal to the system.

the other hand, came somewhat out of the blue, and so the broad focus was retained. In the case of the former utterance, the focused verb *died* receives a denotation as it is incorporated into the ground, as shown in (13). In the case of the latter, there is no ground, and thus the referent of *Johnson* is determined not by priming, but rather by calculating the most likely referent given the context of somebody dying



Of course, it would be erroneous to claim that any constituent whose intended denotation is primed by the context must be part of the ground. Rather, we are making the converse claim that inclusion in the ground must be licensed by contextual priming. Contextually salient material ends up in focus quite often; for example, discourse-old definite DPs can be focused, as in (9). Just because some linguistic expression could in theory be disambiguated from prior context alone does not mean that it must be part of the ground, only that it can be. To round out the analysis, then, it will be necessary to say why elements end up in focus which could have been backgrounded. One possibility is that such a configuration, by forcing the hearer to do unnecessary pragmatic calculations, conveys an implicature (though it may be conventionalized). The implicature in example (9) is that the narrowly focused subject is the most important contributor of information to the sentence. That is to say, the extra interpretive burden placed on the hearer signals that the material in question is particularly interesting to the hearer in some way. This is consistent with the fact that in the answer in (9), all but the focus can be elided. It is only the subject of the sentence that the hearer needs to process before having the answer to their question. Further work will determine whether this simple line of reasoning is sufficient to maintain the simplicity of this conception of focus and ground.

## 4 Summary

The claims of this paper can be summarized as follows.

- Information Structure is a separate component of grammar and consists of two primitives: focus and ground.
- Focus and ground represent two different syntactic structures, built in tandem, that are interpreted quite differently.
- Focused material, and only focused material, is interpreted pragmatically as it is unified with the ground, with conventional meanings feeding a method of reasoning about speaker intent.
- Such an analysis improves theories of compositional pragmatics, and accounts for a wide range of data.

Much remains to be seen regarding the empirical predictions of this proposal, and we hope that this work will spark further discussion and investigation of the seemingly dual nature of Information Structure.

## References

- [1] Clark, Robin. 2011. *Meaningful Games*. Cambridge, MA: MIT Press.
- [2] Frank, Robert. 2002. *Phrase Structure Composition and Syntactic Dependencies*. Cambridge, MA: MIT Press.
- [3] Eilam, Aviad. *Explorations in the Informational Component*. Ph.D Dissertation, University of Pennsylvania.
- [4] Frey, Werner. 2005. Pragmatic properties of certain German and English left peripheral constructions. *Linguistics* 43:1, 89-129.
- [5] Joshi, Aravind. 1985. How much context-sensitivity is necessary for characterizing structural descriptions. In D. Dowty, L. Karttunen, and A. Zwicky (eds.), *Natural Language Processing: Theoretical, Computational, and Psychological Perspectives*. New York: Cambridge University Press.
- [6] Kiss, Katalin. 2007. Topic and focus: Two structural positions associated with logical functions in the left periphery of the Hungarian sentence. In C. Féry, G. Fanselow, and M. Krifka (eds.), *Interdisciplinary Studies in Information Structure* 6, 69-81.
- [7] Kroch, Anthony and Aravind Joshi. 1985. *The Linguistic Relevance of Tree Adjoining Grammar*. Technical Report MS-CS-85-16, Department of Computer and Information Sciences, University of Pennsylvania.
- [8] Parikh, Prashant. 2010. *Language and Equilibrium*. Cambridge, MA: MIT Press.
- [9] Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1, 75-116.
- [10] Schabes, Yves and Stuart Schieber. 1990. Synchronous Tree Adjoining Grammars. In *Proceedings of the 13th International Conference on Computational Linguistics*, Vol. 2, 253-258.
- [11] Schmerling, Sue. 1974. A re-examination of normal stress. *Language* 50, 66-73.
- [12] Selkirk, Elizabeth. 2007. Contrastive focus, givenness and unmarked status of discourse new. In C. Féry, G. Fanselow, and M. Krifka (eds.), *Interdisciplinary Studies in Information Structure* 6, 125-145.
- [13] Tomioka, Satoshi. 2010. Contrastive topics operate on speech acts. In M. Zimmermann and C. Féry (eds.), *Information Structure: Theoretical, Typological and Experimental Perspectives*. Oxford: Oxford University Press.
- [14] Vallduví, Enric. 1990. *The Informational Component*. Ph.D thesis, University of Pennsylvania.
- [15] Vallduví, Enric and Elisabet Engdahl. 1996. The linguistic realization of information packaging. *Linguistics* 34, 459-519.