Information Prosody Correspondence in HPSG

Mohammad Haji-Abdolhosseini
University of Toronto

This paper provides a constraint-based account of information-prosody correspondence within the HPSG framework. The starting point of the paper is Klein’s (2000) account of prosodic constituency in HPSG. However, it departs from the standard syntactocentric architecture of grammar, and adopts a grammar design in which syntax, phonology, and information structure are generated in parallel, with all three applying to a common list of domain objects. It is shown that this theoretical architecture elegantly captures many of the various constraints that have been shown to hold in classical views of grammar.

1 Introduction

This paper lays down the groundwork for a unification-based model of prosody that is sensitive to the syntax and information structure of the sentence. The approach adopted is a more modular one. The theory developed here derives syntactic and prosodic structures at different layers interacting at interfaces only. The model of prosodic constituency laid out here is no-longer syntax-driven. Prosodic structure is defined in parallel with syntactic structure over a list of lexical items commonly accessed from syntax, phonology, and information structure. The architecture of this information-based and modular model of prosody is depicted in Figure 1. According to this model, the syntactic/semantic, prosodic and information structures are all constructed from a unique list of lexical items, $W$. The arrows pointing from $W$ to various structures represent well-formedness constraints on those structures. The arrows that point back to $W$ represent constraints on the features of the members of $W$ imposed by those structures. Structural constraints are basically those found in standard HPSG literature such as the rule schemata and the like. Informational constraints define well-formed information structures. We do not discuss these in this paper. ISPC, ITAC and

Copyright ©2003 Mohammad Haji-Abdolhosseini
Figure 1: Architecture of the information-based model of prosodic constituency

mkMtr are discussed in detail in section 3 where the formal account of the data is presented.

Further motivation for adopting the architecture presented in Figure 1 comes from the myriad mismatches observed between syntactic and prosodic structures. As Zwicky (1982) puts it, “[t]he divergence between the syntactic and phonological organizations of the same material has long been recognized as a problem in analysis and a challenge to theorizing, finding recognition in works as diverse as Kahane and Beym (1948); Pulgram (1970); Bing (1970); Cooper and Paccia-Cooper (1980) and the writing of the ‘metrical phonologists’, in particular Selkirk (1981).” Basically, the mainstream literature assumes that the prosodic structure mirrors syntactic structure unless otherwise specified in order to satisfy certain phonological constraints. These constraints, however, render virtually every prosodic structure different from the syntactic structure of the same sentence. For example, invariably in every Det, Adj, N sequence, the Adj gets “promoted” to the sister of Det giving rise to the following prosodic structure [[Det Adj] N] which is different from the syntactic structure [Det [Adj N]]. The modular model proposed in this paper accounts for the phenomena that Butt and King (1998) call “prosodic promotion”, and “prosodic flattening” straightforwardly without having to manipulate syntactic structures. In addition, information structure-prosody correspondence is handled elegantly in a modular fashion without recourse to unnecessary and ad hoc operations and/or levels of representation. This approach allows for the extension of the model to straightforwardly account for word-order variations as well.

2 Data

The following example is frequently mentioned by Steedman (e.g. Steedman, 2000b, 94) as one that needs to be accounted for by any theory that deals with syntax-phonology mismatches.

(1)  * [[I want to begin to][try to write a play]]
In this example a pause has been placed between a leaner and the prosodic word that it leans on. Clearly, a pause should not be allowed to intervene within leaner groups and we should make provisions in our theory to reject such ill-formed structures.

Klein’s account incorrectly marks (2) ungrammatical as *I*, being a personal pronoun is considered a leaner in the model.

(2)  [I] [want to begin to try to write a play].

The sentences in (2) and (3) appear in Steedman (2000b, 93). He suggests a model of syntax whose surface structures correspond directly to intonational contours. Thus, in these examples, all of the observed intonational contours correspond to alternate surface structures for the sentence in a CCG framework.

(3)  a. [I want][to begin to try to write a play].
    b. [I want to begin][to try to write a play].
    c. [I want to begin to try][to write a play].
    d. [I want to begin to try to write][a play].

In our framework, we would like to develop a model that not only is able to account for these alternate intonational contours and their corresponding semantics, but also maintains the modularity of its component theories. Another example that Steedman (2000b), *inter alia*, discusses is (4).

(4)  *[Three mathematicians] [in ten prefer margarine].

Selkirk (1984) attributes the ungrammaticality of (4) to the violation of the Sense Unit Condition, meaning that the prepositional phrase in ten and the verb phrase prefer margarine fail to form a sense unit as neither is a complement or modifier of the other. Steedman’s CCG model accounts for this. Again, approaching the problem from our standpoint, we would like a multi-partite account for this fact.

Another type of data that we want to account for here is:

(5)  a. [Jane gave the book to Mary]
    b. [Jane] [gave the book to Mary]
    c. [Jane gave the book] [to Mary]
    d. [Jane gave] [the book] [to Mary]
    e. * [Jane] [gave] [the book to Mary]
    f. * [Jane gave] [the book to Mary]
These data have been discussed in Selkirk (1984), and similar examples have been talked about in Steedman (2000a). Selkirk (1984) also attributes the ungrammaticality of (5e, f) to the violation of the Sense Unit Condition: The phrases the book and to Mary do not form a sense unit because neither is a complement or modifier of the other.

3 Analysis

3.1 Information Status and Intonation

Like Steedman, who adopts a Hallidayan tradition, we use the term theme to refer to given information and rheme to new information. Steedman (2000b, 101), following Pierrehumbert (1980), attributes L+H* LH% intonation contour to theme and H*LL% to rheme. L+H* LH% and H*LL% are in Pierrehumbert’s notation (Pierrehumbert, 1980), and respectively correspond to rise-fall-rise and fall intonation in British style (Ladd, 1996, 82). Going back to our example about writing a play (extended here as (6)), we can discuss some of the interaction between information structure and prosody. Hereafter, θ stands for theme and ρ for rheme.

(6) a. [I]θ [want [(to begin) [(to try) [(to write) (a play)]]]]ρ L+H* LH% H*LL%
   b. [(I want)]θ [(to begin) [(to try) [(to write) (a play)]]]ρ L+H* LH% H*LL%
   c. [(I want) (to begin)]θ [(to try) [(to write) (a play)]]ρ L+H* LH% H*LL%
   d. [(I want) (to begin) (to try)]θ [(to write) (a play)]ρ L+H* LH% H*LL%
   e. [(I want) (to begin) (to try) (to write)]θ [(a play)]ρ L+H* LH% H*LL%
   f. [(I want) [(to begin) [(to try) [(to write) (a play)]]]]

Other terms used in the partitioning of information include (back)ground/focus, and topic/comment among others. For the purposes of this paper, we assume that all of these correspond to given/new information. Steedman (2000b) makes a distinction between background/focus and theme/rheme. For him, theme or rheme can be partitioned into background and focus. In this account, the DTE can be thought of Steedman’s focus and whatever that is not a DTE can be considered as background. For a survey of literature on information packaging, see Vallduví and Engdahl (1996).
In (6a–e), each sentence is marked with respect to its information structure; whereas (6f) is un-marked. Assuming that the correlation between information structure and intonation holds and ignoring the possibility of foregrounding items other than the last in an intonational phrase, we conclude that in (6a–e) the last prosodic word (i.e. the default DTE) in theme bears a L+H* LH% (rise-fall-rise) intonation and the last prosodic word in rheme bears a H*LL% (fall) intonation.

3.2 The Type Hierarchy and Constraints

The present model does not have provisions for relating the information status of the constituents in the sentence to prosody. It is clear, however, that in order for it to be able to return the correct intonational phrasing, such a correspondence is necessary. We need to make sure that themes and rhemes (when marked) bear the right intonation and do not occupy the same intonation phrase. Sensitivity to contextual information by the prosodic component entails modification in the feature appropriateness conditions in the prosodic type hierarchy as well as having new constraints introduced on them. Pollard and Sag (1994) assume the presence of a CONTEXT feature for SIGN|SYNSEM|LOCAL. It only seems natural to place information structure within context. However as Engdahl and Vallduví (1994) propose, placing information structure in local objects is problematic for a trace-based account of unbounded dependencies. It is exactly for this reason that De Kuthy (2002), in her theory of information structure, assumes that information structure is a feature appropriate to sign in par with PHON, and SYNSEM. This is another step towards a tripartite architecture of grammar and we are going to adopt it in this work as well. But unlike De Kuthy, we are not going to assume that the scope of information status is represented as a symbolic language with a model-theoretic interpretation. There are two reasons for this: Firstly, taking De Kuthy’s approach requires adherence to one particular semantic theory. In this work, we would like to remain theory-neutral as much as possible when it comes to the internal structures of phonology and semantics. Secondly, linking semantics directly to information structure and in turn phonology adds to the syntactocentrism of the theory. In addition to Jackendoﬃ (2002), a considerable body of work suggests that semantics, syntax, and phonology should be allowed to work separately while making sure that they constrain one another. For more information see Penn (1999a,b); Penn and Haji-Abdolhosseini (2003). What is assumed here is that phonology, syntax and information structure all operate as independently as possible while working on one common list of domain objects that we assume to be lexical items here for convenience. Thus, sign will have (at least) the following feature appropriateness constraint defined over it.
Type \textit{info} has two subtypes: \textit{marked-info} and \textit{unmarked-info}. The type \textit{marked-info} itself subsumes \textit{theme} and \textit{rheme}.

In the prosody partition, we need a place to record the tonal information. Therefore, we add the feature \textit{TONE} to \text{mtr}(\tau). Feature \textit{TONE} takes as its value a list of \textit{tone} objects, which have the following subtypes: \textit{marked-tone} and \textit{unmarked-tone}. The type \textit{marked-tone} (at least) subsumes \textit{rfr}, which stands for rise-fall-rise (L+H* LH%) intonation, and \textit{fall}, which stands for falling (H*LL%) intonation (see (9)). Our revised prosodic type hierarchy takes the form shown in Figure 2.

Another point to discuss here is Klein’s type hierarchy of phrases. What that hierarchy assumes is that all syntactic phrases match some prosodic phrase in their yield. While this is a logical starting point since syntactic trees and prosodic trees often look very similar, even isomorphic in some cases, they clearly are not the same as we observe in the data above and in the literature. Sometimes prosodic phrases do not correspond to any syntactic constituent and vice versa. In our move towards a tripartite architecture, we should therefore treat these two types of constituency differently. Klein’s approach is heavily syntax-driven and involves making prosodic trees by manipulating syntactic trees in a semi-derivational fashion. What we need to do instead is to modify \text{mkMtr} such that it declaratively defines prosodic trees without the need to refer to syntax. This will also greatly simplify \text{mkMtr} as we shall see shortly. Prosodic structure is defined over the list of domain objects as opposed to a list of partial prosodic structures. Figure 3.2 presents the type hierarchy of phrases that we assume in this paper.

A constraint is now required to associate the tones introduced in (9) with the information that they convey. This constraint has to be declared for any object of type \textit{word}. This can be regarded as
an interface point between conceptual structure and phonological structure in Jackendoff’s terms. The constraint, which is called the Information-Tone Association Constraint (ITAC), is formulated in Figure (4). The first disjunct in (4) relates theme with the rise-fall-rise (L+H* LH%) intonation. The second disjunct relates rheme with falling (H*LL%) intonation, and the third one is the default situation where lexical items are left unmarked with regard to their information status and tone. The last disjunct states that some word objects are prosodically leaners.

3.3 Klein’s mkMtr Function Revised

We now need to revise the mkMtr function to handle the new formalism. Before we do that, however, let us go over the type of change that needs to be made. Take the examples in (10).

(10) a. [Jane [drank milk]]
    b. [[Jane drank] milk]

In (10a), Jane is the theme and drank milk the rheme; whereas, in (10b), Jane drank is the theme and milk the rheme. (10a) is compatible with Prosodic Isomorphism Hypthesis (PIH), but (10b) is not. Jane and drank form their own prosodic constituent because they both correspond to the theme of the sentence and milk belongs to a different prosodic constituent because its informational status is different. Therefore, what we want mkMtr to do is to relate prosodic structure and information structure. What this amounts to theoretically is that a weak form of PIH in this model holds for prosody and information structure as opposed to syntactic structure.

(11) The mkMtr Function (Revised)

a. \( \text{mkMtr} : \text{list(pros)} \mapsto \text{mtr(pros)} \)

\[
\text{mkMtr}[\text{\text{[a]}}] = \text{mkMtr}^{\text{full}}(\text{mkAllLnrs}[\text{\text{[a]}}])
\]
Figure 4: Information-Tone Association Constraint (ITAC)

b. \( mkMtr^{\tau<_{pros}} : list(pros) \mapsto mtr(\tau) \)
   
   \( mkMtr^{\tau} \left( \left[ PHON \ \text{pros} \right] \right) = \left[ \right] \)

c. \( mkMtr^{lnr} : list(pros) \mapsto mtr(pros) \)
   
   \( mkMtr^{lnr} \left( \left[ \text{lnr, \ldots, lnr, wrd} \right] \right) = \left[ mtr(lnr) \right] \)

\[
\begin{align*}
\text{DOM} & \left( \text{\ldots, lnr, wrd, p-wrd} \right) \\
\text{TONE} & \left( \text{\ldots, lnr, wrd, p-wrd} \right) \\
\end{align*}
\]

\[
\begin{align*}
\text{DOM} & \left( \text{\ldots, lnr, wrd, p-wrd} \right) \\
\text{TONE} & \left( \text{\ldots, lnr, wrd, p-wrd} \right) \\
\end{align*}
\]

d. \( mkMtr^{full} : list(pros) \mapsto mtr(full) \)

i. \( mkMtr^{full} \left( \left[ \text{wrd, TONE} \right] \right) = \left[ \text{TONE} \right] \)
The new mkMtr function is used in a constraint on sign objects as formalised in (13). The function collect-phon that is defined below in (12) and used in (13) takes a list of domain objects and returns a list of the PHON values of those objects. Theoretically, relations like collect-phon not only ensure the correct input type to other relations or modules of the grammar, they are also ideal in restricting access. In this case, collect-phon allows phonology to only see the phonological data inside DOM. Except for the interface constraints (such as ITAC, and ISPC), nothing from phonology can access the data in the syntactic/semantic, or information-structural modules.

Note that we no longer make use of base-pr and ext-pr; rather, we let what has been described as prosodic flattening and prosodic promotion follow naturally from general constraints on prosody and information structure.

(12) \textit{collect-phon}: \text{list(dom-obj)} \rightarrow \text{list(pros)}
    
    a. \text{collect-phon(\{}\text{\} )=} \text{\{\} }
    
    b. \text{collect-phon}(\text{[1 | 2]}\text{)} = \text{[PHON [1 ] | collect-phon([2]) ]}

(13) \textit{sign} \Rightarrow \begin{bmatrix}
    \text{PHON} & \text{mkMtr} \left( \text{collect-phon([1])} \right) \\
    \text{DOM} & \text{[1]}
\end{bmatrix}
(14) \( mkAllLnrs : \text{list}(\text{pros}) \rightarrow \text{list}(\text{pros}) \)

a. \( mkAllLnrs(1 \oplus 2 \oplus 3) = mkAllLnrs(1 \oplus (\text{mkMtr}^{\text{lnr}}(2)) \oplus 3) \)

b. \( mkAllLnrs(3) = 3 \)

(11a) is the top-level function called by \textit{sign} objects. It uses the \textit{mkAllLnrs} function defined in (14) to generate all the possible leaner groups in the list of domain objects, and passes the resulting mixed list of leaner groups and prosodic words to \textit{mkMtr}^{\text{full}} to generate a complete prosodic structure for the original list of domain objects.

(11b) simply returns a singleton argument intact because a metrical tree requires at least two daughters. (11c) defines metrical trees as consisting of a group of leaners attached to a final prosodic word with the latter being the DTE. The leaner group has the value of its TONE feature structure-shared with that of the prosodic word of the leaner group. (11d-i) is the first of the two definitions for \textit{mkMtr}^{\text{full}}. It requires that all the members of its argument list share the same tone value, which means they should all belong to the same intonational phrase (IP). In that case, it makes a metrical tree in the usual manner and structure-shares its tone value with that of the daughters. (11d-ii) places metrical objects in the same prosodic constituent just in case those objects bear the same tone specification. Then it makes a metrical tree out of the result with the remainder of the list of prosodic objects passed to it.

3.4 Scope of Theme/Rheme Status

The issue of the scope of \textit{theme} and \textit{rheme}, also known as “the projection problem” is approached in this subsection. We define this concept in the form of the \textit{Information Status Projection Constraint (ISPC)} as a type constraint on \textit{hd-cx}. ISPC is formalised in Figure 5.

According to ISPC the arguments of the head daughter in a headed construction by default inherit the information status of that predicate through structure sharing. When an argument is overtly marked for \textit{theme} or \textit{rheme}, it will not inherit the information status (and tone) of the head. Thus in (6c), repeated here as (15), for example, \textit{begin} inherits theme status from \textit{want}, and \textit{write} and \textit{play} inherit \textit{rheme} from \textit{try}.

(15) \([(I \text{ want}) (\text{to begin})]_\theta [(\text{to try}) [(\text{to write}) (\text{a play})]]]_\rho
\[ \text{L+H* LH}\% \quad \text{H*LL}\% \]

Multiple theme and rheme markings are also possible and they can be distinguished by the fact that multiple themes/rhemes are listed separately in the \textit{INFO} feature. We do not consider the projection problem in non-head constructions in this work. Since we assume that the rule schemata allow for the union of the domain objects of their daughters as well as the lists of informational objects, we always have access to the information status of any given prosodic word.
3.5 Accounting for the Data

Let us now go over the derivation of the examples in (10). These derivations are straightforward. In the following two derivations, we use the AVM notation for better exposition. Subsequent examples are represented in Klein’s more succinct notation.

\[
\begin{align*}
\text{(16)} \quad & \quad \text{mkMtr} \left( \begin{bmatrix}
\text{Jane} \\
\text{milk}
\end{bmatrix} \right) \\
\text{mkMtr}^{\text{full}} \left( \text{mkAllLnrs} \left( \begin{bmatrix}
\text{Jane} \\
\text{milk}
\end{bmatrix} \right) \right)
\end{align*}
\]

The application of \( \text{mkMtr} \) to the list of domain objects shown in (16) is represented in (17). The second example, (10b) is derived analogously.
We can again consider the play writing examples, which are shown in (18). Let us assume that these sentences roughly correspond to the semantic specifications represented in Figure 6. In fact, we present the semantic specifications that correspond to (18c). The difference between Figure 6 and the semantic specifications of (18a, b, d) is merely in the scope of theme/rheme (see section 3.4). (18e) is not marked for theme/rheme and gets the default prosodic constituency. (18c), therefore, receives the prosodic structure shown in (19). The cases of (18b, d) are similar.

(18) a. [I want]₀[to begin to try to write a play]₀.
    b. [I want to begin]₀[to try to write a play]₀.
    c. [I want to begin to try]₀[to write a play]₀.
    d. [I want to begin to try to write]₀[a play]₀.
    e. [I want to begin to try to write a play].

(19) \[
\begin{align*}
\text{mkMtr}^\text{full}(\langle \text{I, want, to, begin, to, try, to, write, a, play} \rangle) &= \\
\text{mkMtr}^\text{full}(\text{mkAllLnrs}([\square])) &= \\
\text{mkMtr}^\text{full}(\langle \langle \text{I want}, \text{to begin}, \text{to try}, \text{to write}, \text{a play} \rangle \rangle) &= \\
\left[(\text{I want})\text{to begin}\right]^{\text{refr}}\left[(\text{to try})\text{to write}(\text{a play})\right]^{\text{full}}
\end{align*}
\]

(20) \[
\left[(\text{I want})\text{to begin}(\text{to try})(\text{to write})(\text{a play})\right]
\]

Notice that because the lexical items are unmarked in (18e) with respect to their information status, the prosodic structure that emerges is flat as shown in (20). This is an example where we see that
Figure 6: Basic semantics and information structure of (18c)

what is generally known as prosodic flattening follows naturally from this account and no special theoretical devices are required to derive that structure from a highly structured syntactic tree.

The case of (6a) is somewhat different from the others. In this example, the pronoun I, a leaner, forms its own prosodic phrase bearing the L+H* LH% intonation that corresponds to theme. According to our model, however, the feature TONE is not appropriate to lnr because leaners by definition need a prosodic word to attach to. This can be solved by introducing a lexical rule that type-shifts leaners when their INFO feature is marked. This is formulated as (21) below.

(21) **lnr Type-Shifting Rule**

\[
\text{PHON } \text{lnr} \quad \text{INFO } \text{marked-info} \Rightarrow \text{PHON } \text{p-wrd}
\]

The above example brings us to our next set of data presented earlier in (5) repeated below as (22).

(22) a. [Jane gave the book to Mary]
    b. [Jane] [gave the book to Mary]
    c. [Jane gave the book] [to Mary]
    d. [Jane gave] [the book] [to Mary]
e. * [Jane] [gave] [the book to Mary]
f. * [Jane gave] [the book to Mary]
g. [Jane] [gave the book] [to Mary]
h. [Jane] [gave] [the book] [to Mary]

According to our analysis, (22a) is considered the unmarked case. In (22b), Jane has been marked as theme and gave as rheme, which passes down this status to its arguments book and Mary. Furthermore, in (22c), gave has been marked as theme and Mary as rheme. As mentioned earlier, Selkirk (1984) attributes the ungrammaticality of (22e, f) to the violation of the Sense Unit Condition since the book and to Mary do not form a sense unit. We achieve the same effect in this approach by ISPC and assuming that no more than one information unit (i.e. theme/rheme) can be present in one IP. In other words, each intonation phrase corresponds to only one information unit. This is in line with our version of PIH. Such an analysis entails that in (22d, g, h), there are multiple themes or rhemes and those multiple themes or rhemes are reflected as separate IPs in phonology. (22e, f) are ungrammatical because the book and to Mary have different informational markings, i.e. theme/rheme, rheme1/rheme2 or the like. This condition also prevents (4) because the only way that in ten can be separated from three mathematicians is to have a different informational marking, which by ISPC could not be structure-shared with the informational marking of prefer margarine. Not only ISPC ensures that each information unit reflects the right intonation in phonology; together with the mkMtr function, they also provide an implementations of Selkirk’s (1984) Sense Unit Condition without resorting to another level of representation and unnecessary complication of the theory.

As an example, let us look at the sentences in (22) again. (22d, g, h) have multiple themes or rhemes. The indexed info and its corresponding tone value ensure that multiple themes or rhemes are not mistakenly grouped together. (22c) receives the following prosodic and information structure if we assume that give and book are marked as multiple themes.

(23) [[Jane gave]_{θ1}^{rfr1} (the book)_{θ2}^{rfr2} (to Mary)_{ρ1}^{fall1}]

Examples (22e, f) are automatically rejected because the two arguments of give are sisters of one another; therefore, they cannot bear the same information status by ISPC, and thus, cannot be in the same IP.

Another interesting consequence of the information-based account of prosody in a tripartite grammar architecture is the fact that an ill-formed prosodic structure like (24) never arises because of the way mkMtr has been defined and this relieves us from positing the Lexical Head Association Constraint (see Klein, 2000), which according to Klein is a partial implementation of Selkirk’s end-based mapping.
4 Concluding Remarks

This paper started off with Klein’s (2000) analysis of prosodic constituency in HPSG and extended it to account for some prosodic variation phenomena that are dependent upon the information structure of the sentence. Because a constraint-based approach to prosodic phenomena is employed here, we can capture some interesting linguistic generalities without recourse to *ad hoc* operational rules. In addition, the modular design of the theory allows for better readability and maintainability. The departure from a syntactocentric theory towards a tripartite one in terms of Jackendoff (2002) proved to be a promising approach as it captured a lot of the phenomena previously discussed in the literature in much simpler terms.

The most natural course of action to take from this point is to map all the other intonation forms with information structure in this approach and see what effects they have on the grammar overall. We should also try to find more constraints that syntax, semantics, or pragmatics impose on prosodic structure and even word order. For example, an account of heavy-NP shift and other similar phenomena in this model seems promising.

This model should also be supplemented with an implementation of the model in a CTS system, which then can be used in a study to ascertain how much more or less natural a system based on this model sounds as opposed to a simple TTS system and human voice.

References


