Hungarian information structure: A comparison of Lexical-Functional Grammar and Cartography

Lindsay Hracs
University of Calgary

Abstract

A comparison of a Lexical-Functional Grammar analysis and a Cartography-based analysis of information structure suggests that Lexical-Functional Grammar can better account for phenomena associated with information structure. Also, Lexical-Functional Grammar seems to better capture the fact that the subject and topic roles do not have to coincide. Ultimately, Cartography lacks a method of expressing the interactions with respect to syntax, semantics, and prosody in a systematic way. This is because Cartography deals with information structure as a phenomenon of the peripheries.

Principles and Parameters based theories such as Cartography do not exhibit an interface between PF and LF, which is important for Hungarian. In Hungarian, prosody can affect scope-based interpretation. Lexical-Functional Grammar on the other hand, fully integrates c-structure, f-structure, i-structure, and p-structure in the Correspondence Architecture allowing for an interface between all of these components. Thus, Lexical-Functional Grammar is better suited to deal with the phenomena associated with information structure in Hungarian.
1. Introduction
This paper presents a comparison of a Lexical-Functional Grammar (LFG) analysis and a transformational generative Cartography-based analysis of information structure (IS) in Hungarian. The concentration of this paper is focus structure.

Hungarian word order is governed by notions of IS, rather than the need to identify the subject or the object of the sentence (Erteschik-Shir 2007: 81). Erteschik-Shir (2007: 55) states that multi-component theories such as LFG (Kaplan and Bresnan 1982; Asudeh 2006; Asudeh and Toivonen 2009; etc.) allow IS to be naturally integrated into levels of structure. However, Principles and Parameters models such as Cartography (Cinque and Rizzi 2008; Rizzi 1997, 2004; etc.) have some difficulties integrating prosody into the architecture, and this is a problem for Hungarian since prosody is a relevant component for IS and scope in Hungarian. I intend to show that LFG can better account for the phenomenon of IS in Hungarian. I will present some background information and assumptions regarding IS. Next, I will present a section illustrating the data that are relevant for the current discussion. Section 3 will provide an overview of LFG as well as an LFG analysis of the Hungarian data. Section 4 will provide an overview of Cartography as well as a cartographic analysis of the Hungarian data. Section 5 provides data showing that prosody must be considered in an accurate description of IS in Hungarian, and that LFG is better suited to this task. Finally, I will offer a conclusion to sum up the main points presented in this paper.

2. Hungarian focus structure
IS is treated as a grammatical component in that it is governed by rules and conventions of sentence grammar (Lambrecht 1994: xiii). The linear order of constituents is to some extent determined by concepts having to do with what is and what is not contextually known (Erteschik-Shir 2007: 1). However, how do we define what is contextually known and what is not contextually known? The following section provides some terminology that is relevant for the discussion of IS.

2.1 Terminology
There are three different notions that are relevant for the discussion of the Hungarian data below: topic, focus, and comment.

Erteschik-Shir (2007) states that a topic must be “old” information. More specifically, a topic must appear in a preceding context. In a sense, the topic of a particular discourse is identified by the context in which it is used. Consider the example in (1).

(1)   Q: What did John do?  
      A:  [He TOP] washed the dishes.  

      (Erteschik-Shir 2007: 1)
As we see in (1), the topic is an item that is referential to something already mentioned in the previous context. In this case he refers to John. We can say that the topic refers to the individual that will be asserted about in the sentence (Kiss 2002: 2).

Erteschik-Shir (2007) suggests that foci must be “new” in the discourse. In other words, focus is the non-presupposed information in the sentence. Consider the example in (2).

(2)   Q: What did John do?  
      A:  He [WASHED THE DISHES FOC].  

      (Erteschik-Shir 2007: 29)
In the example in (2), the context is not given from the question that is being asked, thus *washed the dishes* can be considered new information.

Erteschik-Shir (2007) defines the part of the sentence that is not the topic as the comment. As cited by Erteschik-Shir (2007), the example in (3) represents Rizzi’s (1997: 286) structural mapping of the topic-comment division. I will return to the structure represented in (3) for the discussion of Cartography.

(3)

```
             TopP
              |
             XP     Top'
              |
            Top    YP
```

XP = topic
YP = comment

(Rizzi 1997: 286)

It is widely agreed upon that all languages have IS. However, some languages tend to grammaticize IS. This can be seen in the structure in (3). In Cartography, as in other linguistic theories, topic and focus are instantiated as phrases and heads of phrases. The structure in (3) shows the topic and comment clearly instantiated as constituents, however, this is not the case for all frameworks and all languages.

The structure represented in (3) is relevant in describing the interaction of topic, focus, and comment. Given the structure in (3), we can see that the focus of the sentence is contained within the comment. Consider the examples in (4a) and (4b).

(4)

a. Q: What did John do?
   A: [He TOP] [washed THE DISHES FOC, COMMENT]

b. Q: What did John wash?
   A: [He TOP] [washed [THE DISHES FOC] COMMENT]

(Erteschik-Shir 2007: 28)

In example (4a), the comment and the focus are contained within the same constituent. However, with respect to the answer represented in example (4b), the focus is contained within a unit that itself is contained within the comment. Thus, we see that there is more than one way for the topic, focus, and comment to interact.

2.2 Information structure in Hungarian

It is often suggested that Hungarian has variable word order (Kiss 2002: 2). In the case of variable word order, the subject and object in a sentence are not linked to a fixed structural position, thus a verb and its two arguments can form sentences in any of the possible combinations (SVO, SOV, OVS, OSV VSO, VOS). This can be seen in the examples in (5).
(5) szereti ‘loves’, János ‘John’, Marit ‘Mary-ACC’

a. János szereti Marit. ‘It is John who loves Mary’
b. János Marit szereti. ‘As for John and Mary, he loves her’
c. Marit szereti János. ‘It is Mary whom loves John’
d. Marit János szereti. ‘As for Mary and John, he loves her’
e. Szereti János Marit. ‘John loves Mary’
f. Szereti Marit János. ‘John loves Mary’

(Kiss 1981: 187)

However, as Kiss (2002) suggests, the ordering of major sentence constituents in Hungarian is actually constrained in that the function that is associated with a specific structural position is a logical or IS function (i.e., topic or focus), rather than a grammatical function (i.e., subject or object) like in English.

The Hungarian sentence is made up of a topic and a comment (Varga 1980: 68). The notion of topic is independent of the grammatical function of subject (Kiss 2002: 2). Consider example (6).

(6)

a. [János TOP] [fel hívta Marit COMMENT]  
   John up called Mary-ACC  
   ‘John called up Mary’

b. [Marit TOP] [fel hívta János COMMENT]  
   Mary-ACC up called John  
   ‘Mary was called up by John’

(Kiss 2002: 2)

With respect to ACTION verbs, the agent is associated with the topic function, as in (6a), and the patient is associated with the topic function, as in (6b). This is very different from English, where the topic and subject (or agent roles) usually coincide. In the Hungarian sentence, the comment begins with a focused constituent (Varga 1980: 68). Consider the example in (7).

(7)

a. [János TOP] [MARIT kérte fel COMMENT]  
   John-NOM Mary-ACC asked VM  
   ‘As for John, it was Mary that he asked for a dance’

b. [Marit TOP] [JÁNOS kérte fel COMMENT]  
   Mary-ACC John-NOM asked VM  
   ‘As for Mary, it was John that asked her for a dance’

(Kiss 2002: 3)

In example (7a), Marit ‘Mary-ACC’ occupies the preverbal focus position, and in (7b), János ‘John-NOM’ occupies the preverbal focus position. Given that they can occupy different positions, the grammatical functions of subject and object are identified by case-marking. That is, the subject will be marked with Nominative case and the object will be marked with Accusative case. An important phonological consideration to point out is that the lexical item or items in the focused position receive primary stress (Varga 1980).

---

1 Although (6b) is translated as a passive sentence, in Hungarian there is only a difference of word order (Kiss 2002). As de Groot (1994) points out, there is no morphological marking of a voice difference.

2 The verb in Hungarian usually coexists with a verbal modifier (VM) (Kiss 2002: 3). The verbal modifier is said to be a non-referential complement that acts as an aspectual operator (Kiss 2002: 3).
3. Lexical-Functional Grammar and Hungarian information structure

LFG is a theory of generative grammar that conforms to lexicalist, non-transformational principles (Asudeh and Toivonen 2009: 1). LFG was first explored in the 1970’s by Joan Bresnan and Ron Kaplan and discussed in more detail in Kaplan and Bresnan (1982). A main assumption of LFG is that lexical entries are specified for functional and semantic information (i.e. the information specified at f-structure, discussed in section 3.1) (Asudeh and Toivonen 2009: 1). Another assumption is that different types of linguistic information have distinct formal representations simultaneously occurring in grammatical modules, and each grammatical module is subject to separate principles, constraints, and formal descriptions (Asudeh and Toivonen 2009: 1). In addition, the architecture of LFG postulates simple data structures with mappings that define the relationships between these structures. The syntax is non-derivational; transformations are not used for the manipulation of syntactic structure, for the expression of IS, or to create optional syntactic variants (Asudeh and Toivonen 2009: 1).

3.1 Representations of syntactic structure

Kaplan and Bresnan (1982) state that LFG assumes two distinct, but linked, syntactic structures for every sentence in every language: constituent structure (c-structure) and functional structure (f-structure).

C-structures are characterized by phrase structure trees that represent word order, dominance, constituency, and syntactic categories (Asudeh and Toivonen 2009: 2). The f-structure is the representation that captures abstract syntactic relations such as agreement, binding, and other various dependencies (Asudeh and Toivonen 2009: 2). F-structures are a finite set of grammatical functions that represent morphosyntactic information (i.e. case, agreement features, tense, and aspect). Essentially, an f-structure is made up of attribute-value pairs that form an atomic attribute value matrix for a particular sentence (Asudeh and Toivonen 2009: 2). An attribute is a symbol (e.g., PRED, SUBJ, OBJ, TENSE) and its value is either: a) a symbol (e.g. SINGULAR, PRESENT), b) a semantic form (a potentially complex symbol presented in single quotes, e.g., ‘house’), or c) an f-structure (Asudeh and Toivonen 2009: 2).

3.2 I-structure

In addition to c-structure and f-structure, other levels of representation have been proposed to account for various phenomena such as phonological structure (p-structure), morphological structure (m-structure), argument structure (a-structure), semantic structure (s-structure), and information structure (i-structure) (Asudeh and Toivonen 2009: 20). An important level of representation for the discussion of Hungarian is the i-structure. The i-structure level encodes the discourse functions of the sentence in the given context (Butt and King 1997: 9).

Specifiers instantiate either traditional grammatical functions such as subject and object or prominent discourse functions such as topic and focus (Butt and King 1997: 9). Butt and King (1997: 9) posit a lexocentric category S which captures the non-configurational part of the phrase structure. In addition, while syntactic positions are associated with discourse functions, a separate level of representation, i-structure, must be posited as these syntactic positions do not serve as the representation of the discourse structure of a given utterance (Butt and King 1997).

---

3 Asudeh and Toivonen (2009: 9) state that lexocentric phrase structure rules are instantiated in languages where grammatical functions are encoded morphologically rather than configurationally.
Thus, the f-structure and i-structure can be associated with the c-structure. This can be seen in the Urdu sentence represented in (8)\textsuperscript{4}.

(8) a. [naadyyaa \text{TOP}] (to) [abhii] [Tofii] [bazaar=se \text{FOC}] xariid
Nadya.F.NOM indeed just now toffee.F.NOM market.M=from buy
rah-ii \text{\textsuperscript{h}}-ii
Stat-FSG be-PST.FSG
‘Nadya was just buying toffee at the market’

b. Constituent structure:

c. Functional structure:

\textsuperscript{4} Arrows indexed with ‘\textit{i}’ represent IS.
The example in (8) illustrates that the subject and topic coincide as the same lexical item. However, this need not be the case. Thus, having both f-structure and i-structure captures the fact that the functions topic and focus do not always coincide specifically with the subject or object.

3.3 An LFG account of Hungarian information structure

It is important to note that unlike the structure seen in (8) as presented by Butt and King (1997), I will not be positing a lexocentric S in the structure. Butt and King (1997) proposed S to deal with the non-configurational aspects of the Urdu data. However, earlier in this paper it was mentioned that Hungarian is a configurational language since word order is restricted by topic and focus marking. That is, in Hungarian, topic and focus have designated structural positions.

(9) a. \[ \text{Marit}_{\text{TOP}} [\text{JÁNOS} \text{ kérte} \text{ fel}_{\text{COMMENT}}] \]
Mary-ACC John-NOM asked VM
‘As for Mary, it was John that asked her for a dance’

b. Constituent structure:

\[
\begin{align*}
\text{IP}_1 & \\
& \text{f}_2i \in (f_1i \text{ TOP}) \\
& (f_1 \text{ OBJ}) = f_2 \\
& \text{NP}_2 \\
& \text{Marit} \\
& \text{f}_4 \in (f_4i \text{ FOC}) \\
& (f_4 \text{ SUBJ}) = f_5 \\
& \text{NP}_5 \\
& \text{JÁNOS} \\
& \text{f}_6 = \text{f}_7 \\
& \text{V}_7 \\
& \text{VM}_8 \\
& \text{kérte} \\
& \text{fel}
\end{align*}
\]
c. Functional structure

```
PRED 'kérte <SUBJ,OBJ>'

SUBJ $f_1$  $f_4, f_5$  [PRED 'János'
NUMBER SINGULAR
PERSON 3]

OBJ $f_2$  [PRED 'Marit'
NUMBER SINGULAR
PERSON 3]

VM $f_6, f_8$  PRED 'fel'

TENSE PAST
```

Information structure:

```
$\begin{align*}
&f_1 \quad \text{TOP} \quad f_2 \quad \text{PRED} \quad 'Marit' \\
&f_3 \quad \text{FOC} \quad f_4, f_5 \quad \text{PRED} \quad 'János' \\
&f_7 \quad \text{COM} \quad f_6, f_8 \quad \text{PRED} \quad 'kérte'
&\quad \text{PRED} \quad 'fel'
\end{align*}
```

The structures represented in (9) are able to accurately characterize the relationships between both grammatical functions (i.e., subject and object) and IS functions (i.e. topic, focus, and comment). That is, although the c-structure accurately expresses the word order and IS functions of the sentence, the attributive value matrices represented in the f-structure and i-structure are required to express topic and focus, and subject and object relationships. More importantly example (9) shows that the topic and subject do not necessarily have to coincide.

4. Cartography and Hungarian information structure

Cartography is a research program developed within the Principles and Parameters framework (Shlonsky 2010: 417). Cartography is an attempt to draw detailed and precise maps of syntactic configurations and answer the question: What are the correct structural maps for natural language syntax (Cinque and Rizzi 2008)?

4.1 Fundamentals of Cartography

According to Cinque and Rizzi (2008) one of the heuristic principles guiding much of the work in Cartography is “one (morphosyntactic) property – one feature – one head”. It is important to note that this does not exclude featurally complex heads that may arise. That is, these featurally
complex heads cannot be “atoms” of syntactic computations (as in LFG), rather, they arise through derivational processes, namely, head movement (Cinque and Rizzi 2008). Cinque and Rizzi (2008: 50) state that a single surface position may be a complex collection of features such as tense, mood, and subject agreement, but it is important to note that this must be a consequence of the movement of a head to a different structural position.

A fundamental intuition which cartographic approaches attempt to validate is that natural language design tends to opt for local simplicity, that is, a syntactic head has a simple featural specification and can enter into only a few simple relations with its associates (Cinque and Rizzi 2008: 50). This local simplicity is achieved through movement in natural language syntax (Cinque and Rizzi 2008: 50). Consider an example where an expression must be interpreted as both the thematic argument of the verb and as the topic of the structure.

(10) This book, I will read _______ tomorrow.

(Cinque and Rizzi 2008: 50)

Cartographic approaches suggest that natural languages express such structures as seen in (10), by having the element occur twice. One occurrence is in the thematic position and the other occurrence is in the left-peripheral position (Cinque and Rizzi 2008). The thematic argument must be in that position in order for the verb to assign it a certain thematic role (i.e. UTAH (Baker 1988) – it receives its thematic role in D-structure), but how is a scope-discourse property like topicality accounted for (Cinque and Rizzi 2008: 50)? The simple answer to this question is that it is accounted for by dedicated structural positions.

4.2 Information structure from a cartographic perspective
Scope-discourse properties can be accounted for by assuming that there are dedicated structural positions for the elements in question. Cinque and Rizzi (2008: 51) suggest that there is a designated head, Top, that normally occurs in the left periphery of the clause activating the interpretation ‘the specifier should be interpreted as the topic, and the complement as the comment’. Thus, the sentence in (10) would be represented as follows:

(11) [This book]\text{\textsc{top}} [Top [I will read <this book> tomorrow]\text{\textsc{comment}}]

(Cinque and Rizzi 2008: 51)

The specifier of the Top head is interpreted as the Topic and is contextually familiar, and its complement is interpreted as the Comment, a complex predicate predicated of the Topic (Rizzi 2004: 7). The specifier of the Foc head is interpreted as the Focus and its complement is the presupposition. The structures are represented as follows:

(12)

\[
\begin{array}{c}
\text{TopP} \\
\text{XP = Topic} \\
\text{YP = Comment} \\
\text{FocP} \\
\text{XP = Focus} \\
\text{YP = Presupposition} \\
\end{array}
\]

(Rizzi 2004: 7)

In this view of IS, dedicated positions instantiate certain discourse functions (Rizzi 2004: 7).

4.3 A Cartography-based account of Hungarian information structure
In this section, I will apply the principles of Cartography to a portion of the same data seen in the Section 3. Consider the example in (13).
The verb phrase is assumed to be verb initial, with all of the other arguments in an arbitrary order (Kiss 2002: 27). However, there are a few tendencies with respect to postverbal word order, although, it is important to note that these are indeed just tendencies and order in the VP does not affect grammaticality (Kiss 2002: 27). First, specific constituents [+specific] tend to precede non-specific constituents [-specific], and human constituents [+human] tend to precede non-human constituents [-human]. Second, if the arguments share the same [± specific] and [±human] features, θ-role tends to affect the argument order in the postverbal space (Kiss 2002: 27). More specifically, the agent is more likely to precede the patient than the other way around (Kiss 2002: 27). Consider the structure in (14).

Given the underlying structure in (14), the tree structure in (15) can be proposed.

Although there is some variability with respect to the order of the postverbal arguments in the underlying structure, the structure in (15) adequately accounts for surface word order.
5. Prosody, position, and scope in Hungarian
In Hungarian, prosody plays an important role with respect to word order and scope. Not only is word order governed by IS, the grammaticality and interpretations of word order are governed by stress placement. In Hungarian, logical scope assignment is based on word order and prosody (Hunyadi 1996: 140).

(16)  a. 'János köszöntött 'mindenkit.
     John-NOM greeted everyone-ACC
     ‘John greeted everyone’

     b. "János köszöntött mindenkit.
     ‘It was John who greeted everyone (others greeted fewer people)’

     c. "János köszöntött 'mindenkit.
     ‘It was John who greeted everyone (everyone was greeted and each of
     them by John)’

     d. János TOP 'mindenkit köszöntött.
     ‘As for John, he greeted everyone (not just some people)’

     (Hunyadi 1996: 140)

5.1 Prosody in LFG and Cartography
Given the data presented in this paper, it is apparent that the notions of topic and focus interact with phonology (prosody), word order (position), and interpretation (scope) (Erteschik-Shir 2007: 55). As Erteschik-Shir (2007: 4) points out, within Principles and Parameters theories such as Cartography, IS is mainly a phenomenon dealt with in the peripheries, making it difficult to express the interactions with respect to syntax, semantics, and prosody in a systematic way. Rather, multi-component theories such as LFG are better suited to incorporate IS due to their parallel correspondence architectures where syntactic, semantic, phonological, and pragmatic representations are computed in parallel (Erteschik-Shir 2007: 4). A figure showing the parallel correspondence architectures can be seen below.5

![Figure 1: The Correspondence Architecture (Asudeh 2006)](image)

Represented in Figure 1, along with the already mentioned c-structure, f-structure, and i-structure, is phonological structure, or p-structure (Butt and King 1998, O’Connor 2006). P-structure models phrasal phonology and prosody (Asudeh and Toivonen 2009). As seen in the previous section, prosody plays and important role in IS in Hungarian.

On the other hand, the architectures within Principles and Parameters models pose a

---

5 The exact mappings with respect to p-structure will not be presented in this paper. The relevant point is that p-structure exists and that it can account for prosody related effects on IS and scope.
major problem with integrating IS (Erteschik-Shir 2007: 55). That is, syntactic computation interfaces with both phonology and semantics, yet there is no interface between the phonological component (PF) and the semantic component (LF). This can be seen in Figure 2.

![Figure 2: Principles and Parameters Architecture](image)

Ultimately, IS is relevant for both prosody and interpretation, thus, the relevant IS features must be available to both PF and LF, and therefore must already be present in the syntax.

6. Conclusion
Hungarian word order is governed by IS, rather than the need to identify the subject or the object of the sentence (Erteschik-Shir 2007: 81). Although Hungarian appears to have variable word order (refer to example (5)), Hungarian word order is restricted by topic and focus marking, and prosody (Erteschik-Shir 2007, Hunyadi 1996).

The comparison of an LFG analysis and a Cartography-based analysis suggests that LFG can better account for IS. Also, LFG seems to better capture the fact that the subject and topic roles do not have to coincide. Ultimately, Cartography lacks an ability to express the interactions with respect to syntax, semantics, and prosody in a systematic way since it deals with IS as a phenomenon of the peripheries. Principles and Parameters based theories do not exhibit an interface between PF and LF, which is important for Hungarian as prosody can affect scope-based interpretation. LFG on the other hand does this with the Correspondence Architecture. LFG fully integrates c-structure, f-structure, i-structure, and p-structure in the Correspondence Architecture allowing for an interface between all of these components. Thus, LFG is better suited to deal with the phenomenon of information structure in Hungarian.
References:


Contact Information:

Lindsay Hracs
l_kirkpatrick@shaw.ca

Department of Linguistics
The University of Calgary
SS 823, 2500 University Dr. N.W.
Calgary, AB, T2N 1N4
Canada