S-structure Features for Information Structure Analysis

Louise Mycock & John J. Lowe, University of Oxford
louise.mycock@ling-phil.ox.ac.uk, john.lowe@ling-phil.ox.ac.uk

Dalrymple & Nikolaeva (2011) present an approach to information structure (henceforth I.S.) within LFG according to which s-structures contain a feature DF whose value (TOPIC, FOCUS, BACKGROUND or COMPLETIVE) specifies the discourse function of a referent, which in turn determines the role that the relevant meaning constructor will bear at the level of i-structure. As formulated, this approach treats I.S. categories as atomic. While this is unproblematic when analysing a language in which some or all of those discourse functions are associated with a particular syntactic position (e.g. Hindi/Urdu; Butt & King 1996), such an approach obscures generalizations about all the ways in which syntax can encode I.S. For example, in Nandi (Creider & Creider 1989), the principle determining word order is best characterized as (Switch Topic–) Comment(–Continuing Topic), with any clause-initial topic accompanied by the particle *kò*; thus in (1) we see syntactic positions associated with discourse functions not included in Dalrymple & Nikolaeva's inventory.

(1) a. ki:-rì:p a:rtê:t [la:kwe:t]$_{\textbf{Continuing Topic}}$ NANDI
PAST-watch lamb.OBL child.NOM
b. [kò la:kwé:t kò]$_{\textbf{Switch Topic}}$ ki:-rì:p a:rtê:t
PRT child.OBL PRT PAST-watch lamb.OBL
'The child watched the lamb.' (Creider & Creider 1983: 3, 8)

Simply extending the list of atomic DF values to include SWITCH TOPIC and CONTINUING TOPIC would obscure what they have in common: as a result it would be necessary to complicate the description of, e.g., Hungarian by stating that clause-initial Topic position in this language hosts SWITCH TOPICs and CONTINUING TOPICs. A list of atomic DFs also fails to capture important traditional distinctions such as Topic/Comment and Background/Focus. In reality, the values of DF represent labels for more complex statements about I.S. status. We propose to address this important issue by replacing DF with a set of s-structure attributes that more accurately define discourse functions on the basis of principled decompositions of the notions of relative salience and newness, concepts which are often vaguely defined in the wider literature on I.S. but which are understood to be fundamental to its analysis.[1] We propose the five s-structure attributes defined below: (2)–(4) cover the notion of pragmatic prominence/salience; (5) and (6), based on Prince (1992), account for how a hearer establishes and tracks relations between discourse entities on the basis of informational links that relate utterance and context. The latter capture 'newness' and the notion of accessibility, a prerequisite for topichood; see Erteschik-Shir (2007).

(2) **ABOUTNESS (ABOUT±)** The proposition expressed is about this meaning, in the sense that it represents a matter of current concern and is the pivot for truth value assessment.
(3) **INFORMATIVENESS (INFORM±)** A relation is established between this meaning and other elements of the relevant proposition, conveying new information and changing the addressee's representation of the world as a result.
(4) **UPDATE (UPDATE±)** This meaning provides an information update that develops the communication due to its novel information structure status relative to the current discourse context.
(5) **DISCOURSE NEWNESS (DISC_NEW±)** A meaning explicitly evoked in the discourse.
(6) **HEARER NEWNESS (HEAR_NEW±)** A meaning that cannot reasonably be assumed to be already known to the addressee.

With these five binary-valued attributes in place, we can combine Dalrymple & Nikolaeva's (2011) approach to s-structure and i-structure with a feature-based view of discourse functions similar to that in Butt & King (1996) and Choi (1999). Possible feature combinations enable us to define four discourse functions which are essential for cross-linguistic analysis (Table 1), while at the same time capturing other classic distinctions in the status of information (Table 2):[2] **Topic$_{\textbf{E}}$** (Topic Establisher) is a semantic unit which is a topic for the first time, a category including

[1] We set aside contrastive discourse functions as the claim that contrast should be regarded as a separate information structure category is controversial; compare, e.g., Lambrecht (1994: 290) and Dretske (1972).

[2] [DISC_NEW−, HEAR_NEW+] does not exist because something already mentioned in the discourse cannot be new to the hearer (Prince 1992); [ABOUT+] is incompatible with [HEAR_NEW+] because one cannot be informative about an entity which the hearer cannot reasonably be expected to identify.

Switch Topics; a referent has the status of **Topic$_C$** (Continuing Topic) if it is a current (ongoing), constant or repeated topic; **Focus$_{NI}$** (Focus New Information) provides information which is not assumed to be shared by the interlocutors; **Background Information** is not of primary importance to the discourse. (Note that Butt & King's (1996) category Completive Information is a subtype of Background Information under our approach, having a positive value for DISC_NEW.)

Table 1: Discourse Functions defined in terms of s-structure attributes

	prominence/salience			newness/accessibility	
Discourse Function	**ABOUT**	**INFORM**	**UPDATE**	**DISC_NEW**	**HEAR_NEW**
Topic$_E$	+	−	+	±	−
Focus$_{NI}$	−	+	+	+	±
Topic$_C$	+	−	−	−	−
Backg. Info.	−	−	−	±	±

Table 2: Classic I.S. distinctions defined

Topic	ABOUT+
Comment	ABOUT−
Focus	INFORM+
(Back)Ground	INFORM−
'Newsworthy'	UPDATE+

This represents an improvement on previous approaches to non-contrastive discourse functions: Butt & King's (1996) system does not treat Switch Topics as distinct from Continuing Topics; in Choi's (1999) system, Continuing Topics are not distinct from Background Information; and Cook & Payne's (2006) system does not have the capacity to define either Switch Topic or Completive Information as separate I.S. categories. Under our approach, the crucial salience features for 'child' in (1a) and (1b) are represented at s-structure as (7a) and (7b), respectively. In Nandi, the s-structure attribute-value pairs that distinguish between Topic$_C$ and Topic$_E$ will be associated with distinct c-structure positions, just as the f-structure attributes FOCUS and TOPIC were in earlier LFG work, and as DF values are under Dalrymple & Nikolaeva's (2011) approach.

(7) a. **Topic$_C$** $\textbf{child}: \begin{bmatrix} \text{ABOUT} & + \\ \text{INFORM} & - \\ \text{UPDATE} & - \end{bmatrix}$ b. **Topic$_E$** $\textbf{child}: \begin{bmatrix} \text{ABOUT} & + \\ \text{INFORM} & - \\ \text{UPDATE} & + \end{bmatrix}$

Furthermore, the features we propose enable us to capture important generalizations about the relationship between I.S. and syntax cross-linguistically. Herring (1990) presents a survey of 36 languages representing five major word order types (excluding OSV for lack relevant data) and 'free' word order languages, as summarized in Table 3.

Table 3: Capturing generalizations about the syntax–I.S. interface using s-structure attributes

Word order (GFs)	**Tendencies in relation to I.S.**	**Generalization (s-structure attributes)**
SVO	topics in initial position	initial: ABOUT+; ABOUT+ > ABOUT−
SOV	topics are initial; focus is immediately preverbal; backgrounded, predictable information including Topic$_C$ may appear postverbally	initial: ABOUT+; ABOUT+ > ABOUT− immediately preverbal: INFORM+ final: UPDATE−, INFORM−, DISC_NEW−, HEAR_NEW−
VSO, VOS, OVS, 'free'	focus–topic order, but any Switch Topic is initial	initial: UPDATE+, ABOUT+ otherwise: INFORM+ > INFORM−

Under our LFG approach, s-structure attributes are available to associate a specific c-structure position not only with a feature 'bundle' that defines one of the discourse functions shown in Table 1 (as in a discourse-configurational language with an identifiable Topic position such as Hungarian), but also with specific attribute-value pairs to capture other equally important facts about the syntax–I.S. relationship of the type outlined in Tables 2 and 3. This more refined analysis improves the descriptive adequacy of LFG's model of I.S., permitting it to account for a wider range of data and generalizations over the different I.S. categories attested cross-linguistically.

Selected References: Butt, M. & T. H. King (1996). 'Structural Topic and Focus Without Movement'. *Proceedings of LFG96*. ◆ Choi, H.-W. (1997). *Optimizing Structure in Context: scrambling and information structure*. CSLI. ◆ Cook, P. & J. Payne (2006). 'Information Structure and Scope in German'. *Proceedings of LFG06*. ◆ Creider, C. A. & J.T. Creider (1983). 'Topic–Comment Relations in a Verb-Initial Language'. *Journal of African Linguistics* 5: 1–15. ◆ Dalrymple, M. & I. Nikolaeva (2011). *Objects and Information Structure*. CUP. ◆ Herring, S. C. (1990). 'Information Structure as a Consequence of Word Order Type'. *BLS 16*: 163–74. ◆