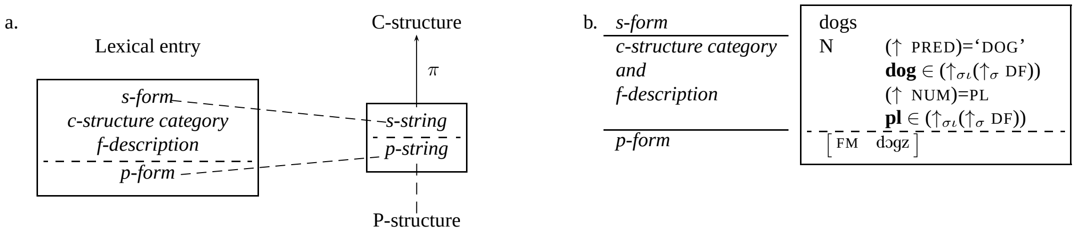
Morphology in the LFG architecture

Mary Dalrymple, University of Oxford

LFG treats linguistic structure as composed of a set of related but independent modules, each with its own internal structure and its own set of universal and language-specific constraints. LFG work on the morphology-syntax interface (notably Sadler & Spencer 2001; Kaplan & Butt 2002) adopts this view of morphology and its relation to other grammatical structures: the morphological component has its own internal structure and obeys universal and language-particular constraints on word formation that need not be shared by other levels.

Following Dalrymple & Mycock (2011) and Mycock & Lowe (2013), the structures in (1a) illustrate the role of the lexical entry in mediating between the s(yntatic)-string and the p(honological)-string. A sample lexical entry (simplified) is given in (1b). The f-description includes two meaning constructors for *dogs*, abbreviated as **dog** and **pl**; for the DF semantic feature, see Dalrymple & Nikolaeva (2011):

(1)


Following Kaplan & Butt (2002),[1] we propose that lexical entries are defined by a morphology relation $\mathcal{M}$ relating the different components of the lexical entry in (1): s(yntactic)-forms, p(honological)-forms, possibly complex c-structure categories, and f-descriptions.

(2) Elements of $\mathcal{M}$: $<$*s-form, p-form, category, f-description*$>$

Note that there can be different c-structure categories and f-descriptions associated with the same s-form and p-form. For example, *swim* has the same s-form and p-form whether it is a noun or a verb. Rewriting the lexical entry in (1b) in this format, the $\mathcal{M}$ relation for the word *dogs* is:

(3) $<$dogs, [dɔgz], N, {(↑ PRED)='DOG', **dog** $\in (\uparrow_{\sigma\iota}(\uparrow_\sigma$ DF)), (↑ NUM)=PL, **pl** $\in (\uparrow_{\sigma\iota}(\uparrow_\sigma$ DF))}$> \in \mathcal{M}$

$\mathcal{M}$ defines the set of lexical entries for a language; the task of defining a morphological component for LFG, then, consists in defining $\mathcal{M}$.

We adopt the distinction introduced by Sadler & Spencer (2001) and Spencer (2006) between purely morphological features or *m-features* (e.g., inflectional class) and syntactically relevant morphologically encoded features or *s-features* (e.g., number). This presupposes a *realizational* theory of morphology as proposed by, among others, Stump (2001), though (like Sadler & Spencer 2001, Kaplan & Butt 2002, and Spencer 2006) our proposal is compatible not only with Stump's theory, but with any realizational theory which relates words to feature sets encoding their morphological properties and structure, including finite state theories of morphology (Kaplan & Kay 1994; Beesley & Karttunen 2003).

We assume that each lexemic root is associated with a unique identifier, its LEXID (Stump 2001). The LEXID is relevant internal to the morphology, in defining the lexical entry, but is not represented as a part of the lexical entry. It cannot be equated with the semantic form, at least if we assume that the semantic form defines subcategorization requirements, since inflectionally related forms with the same LEXID (e.g., voice alternations: Spencer 2013) may have different subcategorization frames.

Building on and adapting proposals by Kaplan & Butt (2002), we decompose the relation $\mathcal{M}$ in terms of a morphological realization relation R and a description function D. The lexical entry encoded in the $\mathcal{M}$ relation is assembled from the s-form, p-form, and s-features contributed by the realization relation R and the c-structure category and f-description obtained via the description function D.

(4) $\mathcal{M}$ = {$<$*s-form, p-form, category, f-descr*$>$ |
$<$LEXID, *s-form, p-form, s-features*$> \in R \wedge$
D(LEXID, *s-features, category, f-descr*)}

The morphological realization relation R is a 4-tuple relating a lexical identifier LEXID, an s-form, and a p-form to a set of s-features. We follow the standard convention of prefixing morphological features, including s-features, with "m", so that the s-feature encoding morphological plural number is written as M-NUM:PL.

(5) $<$LEXID, s-form, p-form, s-features$> \in R$
Example: $<$DOG, dogs, [dɔgz], {M-CAT:N, M-NUM:PL}$> \in R$

[1] We follow the presentation in the unpublished Kaplan & Butt (2002) draft paper, which differs slightly from the version in the handout for the talk presented at LFG02.

R contains all of the 4-tuples for all of the words in the language, whether or not they are formed by a regular, productive morphological process. R for each language is defined entirely by the morphological realization component, and is compatible with any realizational theory of morphology, or any theory that associates s-features with p-forms and s-forms relative to a lexemic root. Purely morphological m-features are represented internal to R, and are not visible to $\mathcal{M}$. S-features are visible within $\mathcal{M}$, but do not appear in the lexical entries defined by $\mathcal{M}$. This maintains a clean separation between purely morphological m-features and syntactically relevant s-features, and between morphology and the other components of the grammar.

The interpretation function D relates a LEXID and a set of s-features to a possibly complex c-structure category and f-description:[2]

(6) D(LEXID, *s-features, category, f-description*)
D(DOG, {M-CAT:N, M-NUM:PL}, N, {(↑ PRED) = 'DOG', **dog** $\in$ ($\uparrow_{\sigma\iota}$($\uparrow_\sigma$ DF)), (↑ NUM) = PL, **pl** $\in$ ($\uparrow_{\sigma\iota}$($\uparrow_\sigma$ DF))})

D is defined in terms of S, a function from LEXIDs to the f-description which is shared among all word forms associated with the lexeme, and M, a function from s-features (and possibly the LEXID) to an f-description. The LEXID DOG has the following S entry:

(7) S(DOG, {(↑ PRED) = 'DOG', **dog** $\in$ ($\uparrow_{\sigma\iota}$($\uparrow_\sigma$ DF))})

The c-structure category and the rest of the f-description are determined by the interpretation function M on the basis of the s-features that appear as the second argument of D, which encode the syntactically relevant differences among the different word forms for the same LEXID. There are several issues in the definition of M: (1) how M is involved in determining the (possibly complex) c-structure category for a word; (2) whether M can produce different results for the same s-features associated with different LEXIDs; and (3) whether M must take into account defaults and feature cooccurrences when determining the f-description corresponding to a set of s-features. With respect to (1), complex c-structure categories encode fine-grained morphosyntactic information about phrase structure categories: for example, Frank & Zaenen (2002) propose categories like V[main,part] for participial main verbs. This information is encoded in s-features, and taken into account by M in defining the c-structure category for a word.

With respect to (2), the simplest assumption is that M does not need to take the LEXID into account, and that the f-description for a word is just the union of the f-description provided by S for the LEXID and the f-description corresponding to the relevant set of s-features:

(8) Simple definition of D, to be evaluated:

D(LEXID, *s-features, category, f-descr*$_S$$\cup$ *f-descr*$_M$) iff
S(LEXID, *f-descr*$_S$) $\wedge$ M(*s-features, category, f-descr*$_M$)

Whether this simple assumption is correct depends on how much information R has about the syntactic behaviour of words. If R always produces s-features that straightforwardly govern syntactic behaviour, the definition of M can be simple, and need not take the LEXID into account; on the other hand, if the interpretation of a set of s-features varies according to the LEXID, the definition of M must be enriched to include LEXID. Patterns of casemarking syncretism in Chukchee (Spencer 2006) will be presented to illustrate these two possibilities.

With respect to (3): the simplest assumption is that the f-description corresponding to a set of s-features can be constructed by examining one s-feature at a time, mapping it to a partial f-description independent of the presence or absence of other s-features. This is, for example, how Andrews (2005) defines M for features not involved in case stacking. This simple definition of M assumes that there are no dependencies among s-features, and no defaults: it leaves no way to introduce an f-description in the absence of an s-feature, for example. As Sadler & Spencer (2001) observe, defaults can play a role in the determination of f-descriptions, and the D function must be able to produce the correct f-description in such cases as well. A definition of M will be presented which allows for the interpretation of an s-feature to be conditioned by the presence or absence of other s-features.

REFERENCES: • Andrews, Avery D., III. 2005. F-structural spellout in LFG morphology. URL `http://members.iinet.net.au/~ada/AveryAndrews/Papers/fspell.pdf`. Unpublished manuscript, Australian National University. • Beesley, Kenneth R. & Lauri Karttunen. 2003. *Finite-State Morphology*. Stanford: CSLI Publications. • Dalrymple, Mary & Louise Mycock. 2011. The prosody-syntax interface. In Miriam Butt & Tracy Holloway King (editors), *On-Line Proceedings of the LFG2011 Conference*. Stanford: CSLI Publications. URL `http://csli-publications.stanford.edu/LFG/16/`. • Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and Information Structure*. Cambridge, UK: Cambridge University Press. • Frank, Anette & Annie Zaenen. 2002. Tense in LFG: Syntax and morphology. In Hans Kamp & Uwe Reyle (editors), *How We Say WHEN It Happens: Contributions to the Theory of Temporal Reference in Natural Language*, pp. 17–51. Tübingen: Niemeyer. Reprinted in Sadler & Spencer (2004, 23–66). • Kaplan, Ronald M. & Miriam

[2] In a full treatment, the f-description produced by D will be a set of templates representing collections of syntactic, semantic, and other specifications, but for simplicity we leave templates aside here.

Butt. 2002. The morphology-syntax interface in LFG. URL http://cslipublications.stanford.edu/LFG/7/lfg02.html. Abstract in *On-line Proceedings of the LFG2002 Conference*, ed. Miriam Butt and Tracy Holloway King. • Kaplan, Ronald M. & Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics* 20(3), pp. 331–378. • Mycock, Louise & John J. Lowe. 2013. The prosodic marking of discourse functions. In Miriam Butt & Tracy Holloway King (editors), *On-Line Proceedings of the LFG2013 Conference*. Stanford: CSLI Publications. URL http://csli-publications.stanford.edu/LFG/18/. • Sadler, Louisa & Andrew Spencer. 2001. Syntax as an exponent of morphological features. In Geert Booij & Jaap van Marle (editors), *Yearbook of Morphology*, pp. 71–96. Dordrecht: Kluwer. • Sadler, Louisa & Andrew Spencer (editors). 2004. *Projecting Morphology*. Stanford: CSLI Publications. • Spencer, Andrew. 2006. Syntactic vs. morphological case: Implications for morphosyntax. In Leonid Kulikov, Andrej Malchukov, & Peter de Swart (editors), *Case, Valency and Transitivity*, pp. 3–22. Amsterdam: John Benjamins. • Spencer, Andrew. 2013. *Lexical Relatedness: A Paradigm-Based Model*. Oxford: Oxford University Press. • Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge, UK: Cambridge University Press.