MORPHOLOGY IN THE LFG ARCHITECTURE

Mary Dalrymple
University of Oxford

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

In line with the overall modular approach of LFG, we assume that 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 of structure. Following Sadler and Spencer (2001), Kaplan and Butt (2002), Spencer (2006, 2013), and many others, we assume that the morphological component of the grammar associates a word form with a set of morphological features representing the structure and contribution of the word, often analyzed as identifying a slot in a paradigm. This view presupposes a *realizational* theory of morphology as proposed by, among others, Stump (2001, 2006, 2012); it is, however, compatible not only with explicitly paradigm-based models, but with any realizational theory which relates words to feature sets encoding their grammatical properties and structure, including finite state theories of morphology (Kaplan and Kay, 1994; Beesley and Karttunen, 2003). Here, we show how lexical entries for word forms are produced on the basis of input from a realizational morphological component.

1 The Morphology-Syntax Interface in LFG

In common with much LFG work, we assume that the morphological component of the grammar associates syntactic, semantic, and other information with word forms, producing lexical entries for word forms.

A note about terminology is in order. Morphologists often use the term 'lexical entry' to refer to information associated with a **lexeme** rather than a word form. Here, we use the term 'lexical entry' to refer to a **word form** (for example, the plural noun *dogs*) and its associated syntactic, semantic, and phonological information. We will use the term *lexemic entry* to refer to the pairing between a lexeme and the f-description encoding grammatical information that all word forms of the lexeme have in common (what Ackerman and Stump (2004) call the 'lexemicon').

1.1 The Lexical Entry: Grammatical Information Associated with Word Forms

We assume that the full lexical entry for the plural noun *dogs* contains at least the following information[1] (Dalrymple and Mycock 2011; Mycock and Lowe 2013; Dalrymple et al. 2015; see Bögel 2015 for a related proposal):

[†]Thanks to Miriam Butt, Ron Kaplan, John Lowe, Louise Mycock, and Andy Spencer for detailed comments on drafts, and to the audience at LFG15 for helpful discussion.

[1]Depending on theoretical assumptions, additional features such as PERS or CASE may also be required to be present.

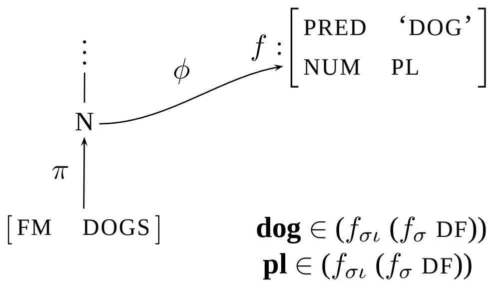
(1) Full lexical entry for *dogs*:

s-form	$(\bullet\ \textsc{fm}) =$	dogs
c-structure category	$\lambda(\pi(\bullet)) =$	N
f-description	$(\uparrow \textsc{pred}) =$	'dog'
	$(\uparrow \textsc{num}) =$	PL
	dog $\in$	$(\uparrow_{\sigma\iota}\ (\uparrow_{\sigma}\ \textsc{df}))$
	pl $\in$	$(\uparrow_{\sigma\iota}\ (\uparrow_{\sigma}\ \textsc{df}))$
p-form	/dɔgz/	

We follow Dalrymple and Mycock (2011) and Mycock and Lowe (2013) in distinguishing two aspects of the string, the *s-string* and the *p-string*. The p-string for a word form is divided into prosodic units, each of which is related by rules of phonology and prosody to the p-form. In this lexical entry, the p-form is /dɔgz/. For more on prosodic structure and its representation, see Mycock and Lowe (2013), Dalrymple et al. (2015), and Bögel (2015).

The s-string is composed of s-string units. The s-string unit for a word form is represented in its lexical entry by the symbol •, meaning the current s-string unit: its use is similar to the $*$ symbol standing for the current node of the phrase structure tree. Each s-string unit is an attribute-value structure containing the attribute FM whose value is a string representing the form of the word, as well as additional attributes and values which we will not discuss here. S-string units are related to terminal nodes of the c-structure tree via the projection function π, as shown in (2). In (1), the s-string unit contributed by the word form *dogs* is related to the c-structure node labelled with the category 'N', as specified in the second line of the entry: π is the function from s-string units to terminal nodes of the c-structure tree, and λ is the labelling function for c-structure nodes (Kaplan, 1995). The rest of the lexical entry contains f-structural information (specification of the PRED and NUM of *dogs*) and two meaning contributions, the meaning of the lexeme DOG and the semantic contribution of the plural morphology.

(2)



Our purpose in the following is to show how the s-form, p-form, c-structure category, and f-description for a word form are determined on the basis of the morphological structure of the word form, given a theory of morphological realization for the language. For simplicity, in the following exposition we will omit meaning constructors and other nonsyntactic constraints in the f-description, working with a simplified lexical entry such as (3):

(3) Simplified lexical entry assumed here:

s-form	$(\bullet$ FM$)$ =	dogs
c-structure category	$\lambda(\pi(\bullet))$ =	N
f-description	$(\uparrow$ PRED$)$ =	'dog'
	$(\uparrow$ NUM$)$ =	PL
p-form	/dɔgz/	

It is important to keep in mind that this simple f-description is a stand-in for the fully complete lexical entry, which encodes syntactic, semantic, information-structural, and other information by means of *templates* (Dalrymple et al., 2004) enabling the statement of generalizations across lexical entries, lexemic entries, and rule annotations within and across languages.

1.2 Lexical entries $\mathcal{L}$

Building on proposals by Kaplan and Butt (2002), we recast the different components of the lexical entry in (3) as a relation $\mathcal{L}$ involving an s-form, a p-form, a possibly complex c-structure category, and an f-description.

(4) $\mathcal{L}<$s-form, p-form, category, f-description$>$

The lexical entry given in (5) represents exactly the same information as in (3), but in a more convenient format for the definitions that we will provide.

(5) $\mathcal{L}<$dogs,
/dɔgz/,
N,
$\{(\uparrow$ PRED$)$='dog', $(\uparrow$ NUM$)$=PL$\}>$

1.3 Types of Morphological Features

We assume a realizational morphological component in which two types of morphological features are relevant. *Morphomic features* are relevant only for morphological realization, and play no role in other components of the grammar. The proper treatment of morphomic features and their role in morphological realization is an important issue in morphological theory, but since we do not depend on a specific theory of morphological realization, we will have nothing to say about morphomic features. A standard example of a morphomic feature is inflectional or declensional class.

M-features are any morphological features that have relevance for other components of the grammar: that is, any morphological features other than morphomic features (Sadler and Spencer, 2001; Spencer, 2006). We follow Sadler and Spencer

(2001) in prefixing morphological features with M-: for instance, writing the m-feature for morphologically encoded tense as M-TENSE, and the m-feature for morphologically encoded past tense as M-TENSE:PAST (Sadler and Spencer, 2001).

1.4 Lexemic Entries and the Lexemic Index

Spencer (2013) proposes that lexemic entries (which he calls 'lexical representations') have the following four components:

(6) Lexemic entries: Spencer (2013)

FORM:	the form of the root and any non-predictable stem forms
SYN:	syntactic information and requirements
SEM:	a representation of the meaning of the lexeme
LI:	a Lexemic Index, an arbitrary label identifying the lexeme

SYN and SEM constitute the f-description associated with the lexeme. We follow Spencer (2013) in assuming that each lexemic root is associated with a unique identifier, its Lexemic Index (similar to the LexID proposed by Stump 2001).

Building on Spencer (2013), we define a **lexemic entry** as a three-place relation LE involving (1) the form of the root and any non-predictable stem forms; (2) an f-description$_L$ that encodes syntactic, semantic, and other information associated with the lexeme, filling the role of Spencer's SYN and SEM; and (3) the Lexemic Index.

(7) General form of lexemic entry:

LE <root & idiosyncratic stem forms, f-description, Lexemic Index>

Lexemic entries for the lexemes with Lexemic Index DOG1 and CHILD1 are as follows:

(8) Lexemic entry for the lexeme DOG1:

LE <{ROOT:dog}, {(↑ PRED)='dog'}, DOG1>

(9) Lexemic entry for CHILD1:

LE <{ROOT: child; STEM1: children}, {(↑ PRED)='child'}, CHILD1>

The full f-description for a word form is obtained by combining the f-description$_L$ for the lexeme and the f-description$_M$ representing morphologically encoded grammatical information, as we will soon see.

1.5 The Realization Relation R

We assume that the morphological component specifies a morphological realization relation R, a set of four-place relations which we will call *m-entries*: R associates a Lexemic Index, an s-form, and a p-form with a set of m-features.

(10) General form of m-entry:

$R <$LexemicIndex, s-form, p-form, m-features$>$

For the word form *dogs*, we have the following m-entry:

(11) M-entry for the word form *dogs*:

$R <$DOG1, dogs, /dɔgz/, $\{$M-CAT:NOUN, M-NUM:PL$\}>$

We assume that the m-entries for each language are defined entirely by the morphological realization component R. The realization relation R for a language accounts for all aspects of the realization of word forms in the language, encompassing a theory of derivational and inflectional morphology, and encoding generalizations about affix ordering, stress placement, and other morphological patterns for the language. Our modular theory of the interface between the morphological component and the rest of the grammar makes no assumptions about the precise nature of R or the internal details of the morphological component; in the current context, R is simply a means of associating m-features with p-forms and s-forms relative to a lexemic root, and is compatible with any realizational theory of morphology.

1.6 The Description Function D

Finally, we require a means of interpreting the m-features for a word form as they are relevant to the rest of the grammar. We follow Kaplan and Butt (2002) in positing a description function D, which maps a set of m-features to the appropriate c-structure category and f-description$_M$, given a Lexemic Index (LI). D corresponds to what Andrews (2005) calls $\mathcal{F}$, and to what Sadler and Nordlinger (2004) call a "lexical transducer" relating m-features to grammatical specifications.

(12) General form of the description function D:

$D <$LI, m-features, category, f-description$_M>$

For the word form *dogs*, D maps the m-features $\{$M-CAT:NOUN, M-NUM:PL$\}$ to the c-structure category N and the simplified f-description $\{(\uparrow$ NUM)=PL$\}$:

(13) $D <$DOG1,
$\{$M-CAT:NOUN, M-NUM:PL$\}$,
N,
$\{(\uparrow$ NUM)=PL$\}>$

1.7 $\mathcal{L}$ Defined in Terms of D, LE, and R

We can now define $\mathcal{L}$ in terms of the set of lexemic entries LE, the morphological realization relation R, and the description function D which interprets the m-features to produce a c-structure category and f-description. $\mathcal{L}$ is the set of all lexical entries of the following form:

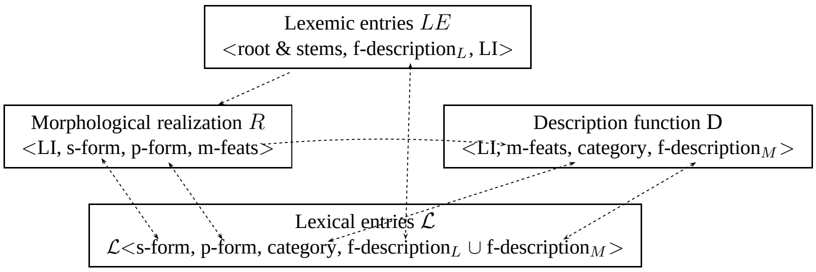
<s-form, p-form, category, f-description$_L \cup$ f-description$_M$>

which meet the conditions imposed by LE, R, and D:

(14) $\mathcal{L}$ = {<s-form, p-form, category, f-description$_L \cup$ f-description$_M$> :
LE <root (and idiosyncratic stem forms), f-description$_L$, LI> $\wedge$
R <LI, s-form, p-form, m-features> $\wedge$
D <LI, m-features, category, f-description$_M$ >}

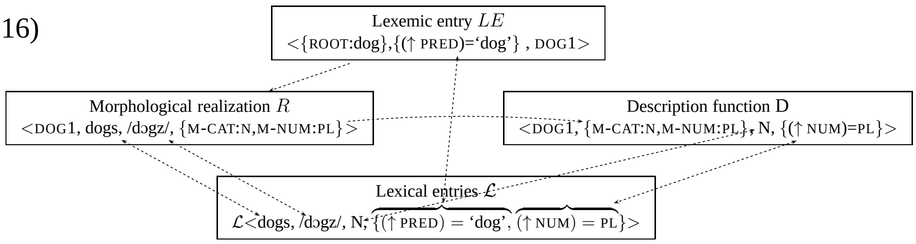
The diagram in (15) shows the flow of information in determining $\mathcal{L}$:

(15)



The diagram in (16) shows how the lexical entry for *dogs* is defined, given the other components:

(16)



Different features and structures are relevant and visible to different components of the grammar. Morphomic features are represented internal to R, and are not visible to $\mathcal{L}$. The Lexemic Index and m-features that are relevant for other components of the grammar are interpreted within $\mathcal{L}$, but do not appear in the lexical entries defined by $\mathcal{L}$ (Kaplan and Butt, 2002). This maintains a clean separation between morphomic features and other m-features, and between morphology and the other components of the grammar. In this way, the current proposal aligns itself with the Principle of Morphology-Free Syntax (Pullum and Zwicky, 1988; Zwicky, 1992), and contrasts with proposals that reject the Lexicalist Hypothesis, including Distributed Morphology (Embick and Noyer, 2007) and the Exo-Skeletal Model (Borer, 2013).

2 Morphological Features and Morphological Classes

In this section, we present some simple and informal examples of the D-mapping. The intention is to illustrate the range and types of D-mappings that may be required, given various alternative assumptions about the best way of treating a particular grammatical construction or the grammatical consequences of a particular morphological alternation. In line with the overall architecture of LFG, our approach is modular in the sense that any particular theory of the realization relation R is generally compatible with a range of different possibilities for syntactic and semantic analysis. An example of this is presented in Section 2.1, where two alternative syntactic analyses of the English "affix hopping" pattern are considered in the context of the same theory of realization R for English. Our theory of the D-mapping as the interface between the morphological component and the rest of the grammar must be flexible enough to allow expression of alternative grammatical analyses on the basis of the same morphological realization relation R, and also to encompass alternative morphological assumptions about the nature of R, given a particular body of assumptions about the proper grammatical analysis of a syntactic construction.

We use $\stackrel{D}{\Rightarrow}$ informally for the D-mapping in this section. The full formal definition of D, to be provided in Section 3, covers all of the types of D-mappings to be examined in the rest of this section.

2.1 C-Structurally Relevant M-Features

English auxiliaries must appear in a particular linear sequence; this is the well-known "affix hopping" pattern (Chomsky, 1957):

(17) a. They are swimming.

b. They have swum.

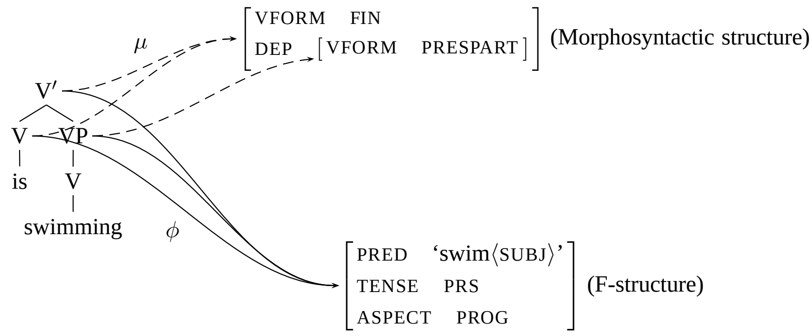
c. They have been swimming.

d. *They have are swum./*They have swimming.

Here we discuss two different LFG analyses of affix hopping. We do not take a position on which analysis is correct; rather, we use the two alternative analyses as illustrations of how the D-mapping works, and in particular to show how different definitions of the D-mapping from the same m-features give rise to different c-structure categories and f-descriptions depending on the syntactic analysis that is assumed.

Butt et al. (1996a,b) introduce a separate projection, *morphosyntactic structure*, reachable via the μ function from the c-structure. The role of morphosyntactic structure is to keep track of morphosyntactic dependencies such as affix hopping: on their analysis, embedding relations in morphosyntactic structure mirror embedding relations at c-structure, and the f-structure is monoclausal.

(18) *is swimming*, Butt et al. (1996a,b):



We assume that the present participle form *swimming* is associated via the realization relation R with the m-features M-CAT:VERB and M-VFORM:PRESPART:

(19) M-entry for the word form *swimming*:

R <SWIM1, swimming, /swɪmɪŋ/, {M-CAT:VERB, M-VFORM:PRESPART}>

On the analysis of Butt et al. (1996a,b), the m-feature M-VFORM:PRESPART corresponds to the feature VFORM with value PRESPART at morphosyntactic structure, and the feature ASPECT with value PROG at f-structure. For this analysis, the required D-mapping is given in (20):

(20) D mapping, Butt et al. (1996a,b):

$$\text{M-VFORM:PRESPART} \overset{D}{\Rightarrow} \{(\hat{*}_\mu\ \text{VFORM})=\text{PRESPART}, (\uparrow \text{ASPECT})=\text{PROG}\}$$

On this analysis, then, a single m-feature (here, the m-feature M-VFORM:PRESPART) can map to an f-description consisting of more than one equation.

Frank and Zaenen (2002) present an alternative analysis of affix hopping in which morphosyntactic structure is projected from f-structure rather than from c-structure. Their analysis also appeals to complex c-structure categories such as V[fin,be], encoding fine-grained specifications over subtypes of standard categories like VP, V′, or V.[2] The parameters of a complex category are written in square brackets after the category label: a complex category like V[fin,be] is appropriate for a word form that is a verb (V) and can be categorized as finite (fin) and a form of the verb *be*. Parameter matching in c-structure rules allows for featural information to be passed through the c-structure, and for the c-structure position of words with particular parameters to be constrained. For example, if we assume that the first parameter of the V′ category can be instantiated to one of the three values ‘fin’, ‘inf’, and ‘part’, and furthermore that the V′’s parameter must match

[2] Miriam Butt [p.c.] notes that a similar complex-category-based analysis is implemented in the English Pargram grammar, and can be inspected through the INESS interface at `http://clarino.uib.no/iness/xle-web`.

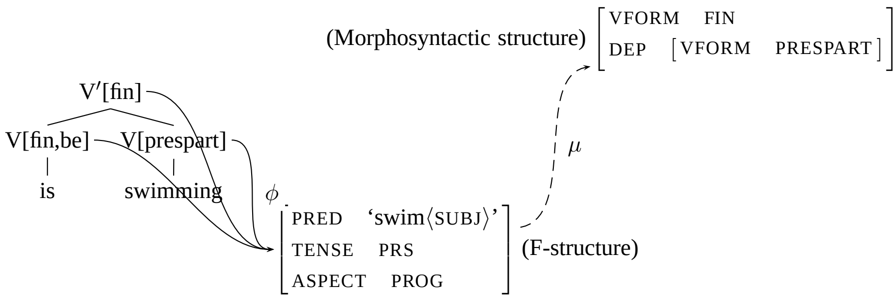
the first parameter of its V daughter, the rule in (21a) is a concise abbreviation of the three rules in (21b). All three of the rules in (21b) require a form of the verb *be* to be followed by a present participle form, and as such license a phrase such as *is swimming*.

(21) a. V′[_ftness] ⟶ V[_ftness,be] V[prespart]

b. V′[fin] ⟶ V[fin,be] V[prespart]
V′[inf] ⟶ V[inf,be] V[prespart]
V′[part] ⟶ V[part,be] V[prespart]

Under the assumptions of Frank and Zaenen (2002), the c-structure, f-structure, and morphosyntactic structure for *is swimming* is as in (22):

(22) *is swimming*, Frank and Zaenen (2002):



We assume that the finite form *is* has the following m-entry:

(23) M-entry for the word form *is*:

R <BE1, is, /ɪz/, {M-CAT:VERB, M-VTYPE:BE, M-VFORM:FIN}>

On Frank and Zaenen’s analysis, the D-mapping maps from m-features to complex c-structure categories as well as determining the f-description. A combination of three m-features determines the complex c-structure category V[fin,be] for the word form *is*:

(24) D-mapping to complex c-structure category, Frank and Zaenen (2002):

{M-CAT:VERB, M-VTYPE:BE, M-VFORM:FIN} $\overset{D}{\Rightarrow}$ V[fin,be]

The M-VFORM:FIN feature is also involved in specifying the value of the VFORM feature at morphosyntactic structure:

(25) D-mapping to f-description, Frank and Zaenen (2002):

M-VFORM:FIN $\overset{D}{\Rightarrow}$ {($\uparrow_\mu$VFORM)=FIN}

Note that in the Frank and Zaenen (2002) analysis, more than one m-feature is involved in a single D-mapping relation, as in (24); furthermore, two different D-mapping relations may depend on the same m-feature, here, the M-VFORM feature, which appears in both (24) and (25). Our formal definition of D must therefore allow for these possibilities.

2.2 F-description$_M$ Dependent on Inflectional Class

It may sometimes be necessary for the D-mapping to depend on the simultaneous presence of more than one m-feature, and our formal definition of D must therefore be formulated to allow the mapping for each m-feature to be constrained by reference to the other m-features associated with a word form. For example, the past tense form of an English verb may be the same as its past participial form, or the two forms may differ. In what we will call Class 1, the past participle form is the same as the past tense form; in what we will call Class 2, there are two separate forms.

(26)

	ROOT	PAST TENSE	PAST PARTICIPLE
CLASS 1:	meet	met	
	talk	talked	
	make	made	
CLASS 2:	give	gave	given
	take	took	taken

One possible analysis of these patterns is to assume that inflectional class, represented as M-CLASS, is an m-feature which can be relevant in D-mapping. According to this analysis, the following m-entries are required:

(27) M-entries for English verb forms:

R <MEET1, met, /mɛt/, {M-CAT:VERB, M-ICLASS:1, M-TENSE:PAST}>

R <GIVE1, gave, /geɪv/, {M-CAT:VERB, M-ICLASS:2, M-TENSE:PAST}>

R <GIVE1, given, /gɪvən/, {M-CAT:VERB, M-ICLASS:2, M-FORM:PPART}>

Notice that on this analysis, there is only one m-entry for the Class 1 verb form *met*, associated with the m-feature M-TENSE:PAST. The D-mapping rules given in (28) produce the correct f-descriptions for these word forms: a disjunction specifying either past tense or the past participial form for Class 1 verbs like *meet*, and separate rules for past tense and past participial forms for Class 2 verbs like *give*.

(28) a. Past participle homophonous with past tense form in Class 1 verbs:

M-TENSE:PAST $\stackrel{D}{\Rightarrow}$ {{(↑ TENSE)=PAST | (↑ VFORM)=PPART}} in the presence of the m-feature M-ICLASS:1.

b. Separate past participle and past tense form in Class 2 verbs:

M-TENSE:PAST $\stackrel{D}{\Rightarrow}$ {(↑ TENSE)=PAST} in the presence of the m-feature M-ICLASS:2.

M-FORM:PPART $\stackrel{D}{\Rightarrow}$ {(↑ VFORM)=PPART}

If we adopt this analysis, M-ICLASS is not a morphomic feature relevant only for morphological realization, but is crucial in the definition of D in determining the full f-description for a verb form.

2.3 Lexical Exceptions: F-description$_M$ Dependent on Lexemic Index

We include the Lexemic Index as a component of the description function D in order to allow for the possibility that the interpretation of a set of m-features varies for different lexemes, and is thus dependent on the Lexemic Index. This is another kind of context-dependence that may be relevant in the definition of the D-mapping.

For example, Acquaviva (2008, page 19) proposes that the noun *measles* is exceptional in being morphologically plural (involving suffixation of plural *-s* to a base, and so carrying the m-feature M-NUM:PL) but syntactically singular, and so bearing the f-structure equation (↑ NUM)=SG:

(29) Measles is/*are a terrible disease.

Support for the view that *measles* is morphologically plural, consisting of the root *measle* followed by plural *-s*, is provided by attested examples of the uninflected form *measle* as the first member of a compound:

(30) a. New needle-free **measle vaccine** 'could save thousands of children's lives' (headline in *The Telegraph*, 17 Aug 2009)

b. ...it is reasonable because though we have never found a **measle germ** associated with **measle-symptoms** we have in cases with like symptoms found, not indeed **measle germs**, but things of the same sort... (Wisdom, 1968)

If we adopt this generalization, the D-mapping for a set of m-features may differ depending on the Lexemic Index LI. In the case at hand, we have one mapping involving the M-NUM feature for *measles*, and another mapping for all other nouns:

(31) *Measles* as a lexical exception:

M-NUM:PL $\stackrel{D}{\Rightarrow}$ {(↑ NUM)=SG} if LI = MEASLES1,
otherwise M-NUM:PL $\stackrel{D}{\Rightarrow}$ {(↑ NUM)=PL}.

This analytic possibility may or may not arise, depending on the particular theory of the morphology-syntax-semantics interface that is adopted. An alternative, more restrictive hypothesis is that D never takes the LI into account, and always relates a given set of m-features to the same f-description; this would require an alternative account of the difference between a syntactically singular noun like *measles* and a syntactically plural noun like *dogs*. If such an account is shown to be available and preferable for all word forms, we need not include the LI as a component of the description function D. Future work will show whether the more restrictive hypothesis is viable, or if the LI must be taken into account in at least some cases of the D-mapping.

2.4 Defaults in F-description$_M$

Some morphological theories assume defaults: that is, the absence of a feature is interpreted as indicating the presence of some grammatical property. Hence, the D-mapping must be formulated so as to allow for the possibility of introducing a particular f-description if no m-feature of a certain type is present.[3]

For example, we might propose that nouns are assumed to be syntactically singular if the plural m-feature does not appear, so that plural nouns are associated with the m-feature M-NUM:PL, and singular nouns lack a M-NUM feature.

(32) Singular number as a morphological default for nouns:

Introduce the f-description {(↑ NUM)=SG} if there is an m-feature M-CAT:N but no M-NUM m-feature in the m-description.

3 The Description Function D

3.1 Previous Definitions of D

In their analysis of the morphology-syntax interface, Kaplan and Butt (2002) assume that the D-mapping is defined very simply: the f-description corresponding to a set of m-features is constructed by examining one m-feature at a time, mapping it to a partial f-description independent of the presence or absence of other m-features. To illustrate their approach, they propose an analysis of the German noun *Kätzchen* which assumes that it is ambiguous, and that the correlation between number and case is captured by positing two different R-relations producing two different lexical entries, one with singular number and indeterminate NOM/DAT/ACC case, and the other with plural number and indeterminate NOM/GEN/DAT/ACC case:

(33) Kaplan and Butt (2002), *Kätzchen*:

a. R <KÄTZCHEN1, Kätzchen, /kɛts.çən/, {M-ROOT:KATZE, M-CATEGORY:NOUN, M-DIMIN, M-GEND:NEUT, M-NUM:SG, M-CASE:NOM/DAT/ACC}>

b. R <KÄTZCHEN1, Kätzchen, /kɛts.çən/, {M-ROOT:KATZE, M-CATEGORY:NOUN, M-DIMIN, M-GEND:NEUT, M-NUM:PL, M-CASE:NOM/GEN/DAT/ACC}>

Kaplan and Butt (2002) provide the following set of D-mappings from m-features to f-descriptions:

[3] Note that we neither advocate nor deplore the use of defaults in defining the D-mapping relation; whether or not defaults are needed in the D-mapping for a language depends on the characteristics of the morphological realization relation R and the syntactic rules and constraints for the language. Our aim is to formulate the D-mapping in a flexible enough way to allow for various alternative assumptions about the realization relation R and D-mapping for the language under analysis.

(34) Kaplan and Butt's (2002) description function D for each m-feature:

$$\begin{aligned}
D(\textsc{m-root:Katze}) &= \{(\uparrow \textsc{pred}) = \text{‘Katze’}\} \\
D(\textsc{m-category:noun}) &= \{\text{N}, (\uparrow \textsc{ntype}) = \textsc{count}\} \\
D(\textsc{m-dimin}) &= \{(\uparrow \textsc{dimin}) = +\} \\
D(\textsc{m-gend:neut}) &= \{(\uparrow \textsc{gend}) = \textsc{n}\} \\
D(\textsc{m-num:sg}) &= \{(\uparrow \textsc{num}) = \textsc{sg}\} \\
D(\textsc{m-num:pl}) &= \{(\uparrow \textsc{num}) = \textsc{pl}\} \\
D(\textsc{m-case:nom/dat/acc}) &= \{(\uparrow \textsc{case}) \in \{\textsc{nom,dat,acc}\}\} \\
D(\textsc{m-case:nom/gen/dat/acc}) &= \{(\uparrow \textsc{case}) \in \{\textsc{nom,gen,dat,acc}\}\}
\end{aligned}$$

On this simple view, there are no mappings from multiple m-features to a complex c-structure category (as we saw in 24), and the same m-feature cannot be involved in two different D-mapping relations (as we saw in 24 and 25). This view also does not allow for lexical exceptions taking into account the Lexemic Index (as we saw in 31), or for defaults in the D-mapping (as we saw in 32). Rather, the f-description corresponding to a set of m-features is just the union of all of the f-descriptions for each m-feature:

(35) Kaplan and Butt's (2002) description function D for a set of m-features:

$$D(\{d_1, d_2, ..., d_n\}) = D(\{d_1\}) \cup D(\{d_2\}) \cup ...\ D(\{d_n\})$$

This assumption is shared by Andrews (2005), who proposes a similar definition for his version of D, which he calls $\mathcal{F}$. This simple approach is adequate for many cases, but as we have seen, it is not adequate for all of the analytical possibilities that may arise.

3.2 Definition of D

We propose a definition of D that allows for the more complex cases examined in Section 2:

(36) D $<$LI, m-features, category, f-descr$_{default}$$\cup$ f-descr$_{feat}$ $>$ if and only if
D_{cat} $<$LI, m-features, category$>$ and
$D_{default}$ $<$LI, m-features, f-descr$_{default}$ $>$ and
D_{feats} $<$LI, m-features, f-descr$_{feats}$ $>$.

The subsidiary definitions D_{cat}, $D_{default}$, and D_{feats} are specified on a language-by-language basis, though there is likely to be a great deal of commonality in their definitions across languages; this is an important topic of research in the interface between morphology and other components of the grammar.

The (possibly complex) c-structure category for a word form is specified by D_{cat} on the basis of the L(exical) I(ndex) and the m-features. The f-description for the word form is determined by combining two subsidiary f-descriptions: D_{feats} contributes the f-description that is specified by the m-features associated with the word form, and $D_{default}$ contributes the default f-description that appears in the absence of certain marked m-features.

3.2.1 Defining the C-Structure Category

If complex categories are not assumed, the definition of D_{cat} is very simple, appealing to a straightforward specification of the c-structure category by the m-feature M-CAT:

(37) D_{cat} <LI, m-features, N> if and only if M-CAT:N $\in$ m-features.
D_{cat} <LI, m-features, V> if and only if M-CAT:V $\in$ m-features.
D_{cat} <LI, m-features, Adj> if and only if M-CAT:ADJ $\in$ m-features.
$\vdots$

If complex categories are assumed, more than one m-feature might be involved in the full specification of a complex category. Here is a representative example of a D_{cat} rule for the complex category V[fin,be] in the analysis of affix hopping shown in example (22):

(38) D_{cat} <LI, m-features, V[fin,be]>
if and only if {M-CAT:VERB,M-VTYPE:BE,M-VFORM:FIN} $\subseteq$ m-features.

According to the definition of D in (36), D_{cat} is required to apply in order to determine the c-structure category of a word form. It does not do any "feature accounting", however; whether the definition of D_{cat} appeals to one m-feature or more than one, all of the m-features are passed on to $D_{default}$ to check for the application of rules involving privative m-features and defaults.

3.2.2 Privative M-features and Defaults

We assume that a set of default/privative D-mapping rules is defined for each language (including the possibility of no default mapping rules). Assuming that $D_{default}$ contains n default rules, the default f-description results from applying each of the n rules in turn:

(39) Default mappings $D_{default}$:

$D_{default}$ <LI, m-features, $d_1 \cup d_2 \cup ...d_n$ > if and only if
$D1_{default}$ <LI, m-features, d_1 > and
$D2_{default}$ <LI, m-features, d_2 > and
$\vdots$
$Dn_{default}$ <LI, m-features, d_n >.

The rule in (40) is a schematic rule illustrating the general form of default rules:

(40) Schematic default rule:

$D1_{default}$ <LI, m-features, f_1 > if $m_1 \notin$ m-features (and possibly other conditions as well),

otherwise $D1_{default}$ <LI, m-features, $\emptyset$ >.

For example, the following rule introduces a singular f-description for a noun that does not have a M-NUM:PL feature:

(41) Example: Default mapping to a singular f-description in the absence of a plural m-feature

$D1_{default}$ <LI, m-features, $\{(\uparrow$ NUM)=SG$\}$>
  if M-CAT:N $\in$ m-features and M-NUM:PL $\notin$ m-features,
otherwise $D1_{default}$ <LI, m-features, $\emptyset$ >.

As with the D_{cat} rules, the $D_{default}$ rules are not involved in "feature accounting": once the default rules have applied to a set of m-features, that set is passed unchanged to the D_{feats} rule.

3.2.3 F-descriptions corresponding to m-features

Finally, the D_{feats} rule applies. This rule keeps track of features, and each m-feature must be accounted for by a D_{feats} rule. However, the D-mapping for an m-feature might correspond to an empty f-description: for example, a feature that specifies only c-structure category information might correspond to the empty f-description.

(42) Mapping m-features to f-descriptions:

D_{feats} <LI, $\{m_1, m_2, ...m_n\}$, $d_1 \cup d_2 \cup ...d_n$ > if and only if
  D_{feats} <LI, m_1, $\{m_1, m_2, ...m_n\}$, d_1 > and
  D_{feats} <LI, m_2, $\{m_1, m_2, ...m_n\}$, d_2 > and
  $\vdots$
  D_{feats} <LI, m_n, $\{m_1, m_2, ...m_n\}$, d_n >.

Simple D-mappings from m-features to f-descriptions An example of a simple mapping from the M-CASE feature to syntactic case is shown in (43):

(43) Example: Simple mapping from M-CASE to syntactic case

D_{feats} <LI, M-CASE:NOM, m-features, $\{(\uparrow$ CASE)=NOM$\}$>.
D_{feats} <LI, M-CASE:ACC, m-features, $\{(\uparrow$ CASE)=ACC$\}$>.
D_{feats} <LI, M-CASE:DAT, m-features, $\{(\uparrow$ CASE)=DAT$\}$>.
D_{feats} <LI, M-CASE:GEN, m-features, $\{(\uparrow$ CASE)=GEN$\}$>.

In formulating the D_{feats} rules for a language, it may be useful to appeal to a notational convention that allows for reuse of an m-feature value as the value of the corresponding f-structure feature, borrowing the underscore notation for the argument of a parametrized template (Dalrymple et al., 2004) to indicate that morphological case always matches syntactic case. Here, the value of the M-CASE feature is represented as _CASE with a preceding underscore, and is reused as the value of the f-structure feature CASE.

(44) Notational convention: General mapping from any M-CASE to the corresponding f-structure CASE specification, abbreviating the rules in (43)

D_{feats} <LI, M-CASE:_CASE, m-features, $\{(\uparrow$ CASE)=_CASE$\}$>.

Context-sensitive D-mappings We have seen that a D-mapping rule for a particular m-feature may depend on the presence or absence of other m-features. For example, in Section 2.2 we saw that the f-description for a verb with m-feature M-VFORM:PAST can depend on the inflectional class of the verb. The following D_{feats} rule captures this dependency:

(45) Example: D-mapping dependent on the M-CLASS m-feature

D_{feats} <LI, M-VFORM:PAST, m-features, $\{\{(\uparrow$ TENSE)=PAST | $(\uparrow$ VFORM)=PPART$\}\}$>
if and only if M-CLASS:1 $\in$ m-features.

D_{feats} <LI, M-VFORM:PAST, m-features, $\{(\uparrow$ TENSE)=PAST$\}\}$>
if and only if M-CLASS:2 $\in$ m-features.

Vacuous D-mappings Since the D_{feats} mapping maps each individual m-feature to an f-description, we require a D_{feats} mapping rule for each m-feature, even those that do not correspond to an f-description. A schematic role for such inert m-features is the following:

(46) D-mapping to the empty f-description for an inert m-feature m_1

D_{feats} <LI, m_1, m-features, $\emptyset$ >.

For example, the M-CLASS feature may be important in controlling the mapping of other features (such as the M-VFORM feature, as shown in 45), but it may not itself correspond to any f-description; that is to say, it maps to the empty f-description $\emptyset$.

(47) Example: D-mapping to the empty f-description for the M-CLASS:1 and M-CLASS:2 m-features

D_{feats} <LI, M-CLASS:1, m-features, $\emptyset$ >.
D_{feats} <LI, M-CLASS:2, m-features, $\emptyset$ >.

For succinctness, we can introduce an additional notational convention: specification of an m-feature attribute like M-CLASS without specifying a value is interpreted as signifying all possible values for the M-CLASS feature.

(48) Example: D-mapping to the empty f-description for the M-CLASS m-feature with any value

D_{feats} <LI, M-CLASS, m-features, $\emptyset$ >.

D-mappings dependent on the LI Finally, the Lexemic Index is relevant for analyses involving lexically idiosyncratic D-mappings. Schematically, such analyses are of the following form:

(49) Schematic mapping for a lexical exception, dependent on the Lexemic Index:

D_{feats} <LI, m_1, m-features, f_1 > if LI = l_1,
otherwise D_{feats} <LI, m_1, m-features, f_2 >.

For example, we can treat the noun *measles* as a lexical exception, morphologically plural but syntactically singular:

(50) Example: Mapping to syntactically singular f-description for the morphologically plural noun *measles*, and to plural f-description for all other nouns

D_{feats} <LI, M-NUM:PL, m-features, $\{(\uparrow$ NUM)=SG$\}$> if LI = MEASLES1,
otherwise D_{feats} <LI, M-NUM:PL, m-features, $\{(\uparrow$ NUM)=PL$\}$>.

4 Conclusion

We have proposed a definition of $\mathcal{L}$ as the set of lexical entries for the word forms of a language. We rely on a set of lexemic entries LE and a morphological realizational component R which associates a set of m-features with a word form of a lexeme in the language. The description function D for the language maps from m-features to c-structure categories and f-descriptions. We hope that the proposals we have made will enable further exploration of the place of morphology in the architecture of LFG and the interface between morphology and the rest of the grammar.

References

Ackerman, Farrell and Stump, Gregory. 2004. Paradigms and Periphrastic Expression: A Study in Realization-Based Lexicalism. In Sadler and Spencer (2004), pages 111–157.

Acquaviva, Paolo. 2008. *Lexical Plurals: A Morphosemantic Approach*. Oxford: Oxford University Press.

Andrews, III, Avery D. 2005. F-Structural Spellout in LFG Morphology, unpublished manuscript, Australian National University.

Beesley, Kenneth R. and Karttunen, Lauri. 2003. *Finite-State Morphology*. Stanford: CSLI Publications.

Bögel, Tina. 2015. *The Syntax-Prosody Interface in Lexical Functional Grammar*. Ph. D.thesis, Universität Konstanz.

Borer, Hagit. 2013. *Structuring Sense, Volume III: Taking Form*. Oxford: Oxford University Press.

Butt, Miriam, Fortmann, Christian and Rohrer, Christian. 1996a. Syntactic Analyses for Parallel Grammars: Auxiliaries and Genitive NPs. In *Proceedings of the International Conference on Computational Linguistics (COLING96)*, pages 182–187, International Committee on Computational Linguistics, Copenhagen.

Butt, Miriam, Niño, María-Eugenia and Segond, Frédérique. 1996b. Multilingual Processing of Auxiliaries Within LFG. In Dafydd Gibbon (ed.), *KONVENS '96: Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference*, pages 111–122, Mouton de Gruyter, reprinted in Sadler and Spencer (2004, 11–22).

Chomsky, Noam. 1957. *Syntactic Structures*. Mouton de Gruyter.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*, volume 34 of *Syntax and Semantics*. New York: Academic Press.

Dalrymple, Mary, Kaplan, Ronald M. and King, Tracy Holloway. 2004. Linguistic Generalizations Over Descriptions. In Miriam Butt and Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2004 Conference*, Stanford: CSLI Publications.

Dalrymple, Mary, Lowe, John and Mycock, Louise. 2015. Lexical Functional Grammar. Second edition of Dalrymple (2001), in preparation.

Dalrymple, Mary and Mycock, Louise. 2011. The Prosody-Syntax Interface. In Miriam Butt and Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2011 Conference*, Stanford: CSLI Publications.

Embick, David and Noyer, Rolf. 2007. Distributed Morphology and the Syntax-Morphology Interface. In Gillian Ramchand and Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*, pages 289–324, Oxford: Oxford University Press.

Frank, Anette and Zaenen, Annie. 2002. Tense in LFG: Syntax and Morphology. In Hans Kamp and Uwe Reyle (eds.), *How We Say WHEN It Happens: Contributions to the Theory of Temporal Reference in Natural Language*, pages 17–51, Tübingen: Niemeyer, reprinted in Sadler and Spencer (2004, 23–66).

Kaplan, Ronald M. 1995. The Formal Architecture of Lexical-Functional Grammar. In Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell, III and Annie Zaenen (eds.), *Formal Issues in Lexical-Functional Grammar*, pages 7–27, Stanford: CSLI Publications, previously published in *Proceedings of ROCLING II*, ed. C.-R. Huang and K.-J. Chen, Tapei 1989, pp. 1–18, and in *Journal of Information Science and Engineering 5* (1989), pp. 305–322.

Kaplan, Ronald M. and Butt, Miriam. 2002. The Morphology-Syntax Interface in LFG, abstract in *On-line Proceedings of the LFG2002 Conference*, ed. Miriam Butt and Tracy Holloway King.

Kaplan, Ronald M. and Kay, Martin. 1994. Regular Models of Phonological Rule Systems. *Computational Linguistics* 20(3), 331–378.

Mycock, Louise and Lowe, John J. 2013. The Prosodic Marking of Discourse Functions. In Miriam Butt and Tracy Holloway King (eds.), *On-Line Proceedings of the LFG2013 Conference*, Stanford: CSLI Publications.

Pullum, Geoffrey K. and Zwicky, Arnold M. 1988. The Syntax-Phonology Interface. In Frederick J. Newmeyer (ed.), *Linguistics: The Cambridge Survey*, Cambridge, UK: Cambridge University Press.

Sadler, Louisa and Nordlinger, Rachel. 2004. Relating Morphology to Syntax. In Sadler and Spencer (2004), pages 159–186.

Sadler, Louisa and Spencer, Andrew. 2001. Syntax as an Exponent of Morphological Features. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology*, pages 71–96, Dordrecht: Kluwer.

Sadler, Louisa and Spencer, Andrew (eds.). 2004. *Projecting Morphology*. Stanford: CSLI Publications.

Spencer, Andrew. 2006. Syntactic vs. Morphological Case: Implications for Morphosyntax. In Leonid Kulikov, Andrej Malchukov and Peter de Swart (eds.), *Case, Valency and Transitivity*, pages 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.

Stump, Gregory T. 2006. Heteroclisis and Paradigm Linkage. *Language* 82(2), 279–322.

Stump, Gregory T. 2012. The Formal and Functional Architecture of Inflectional Morphology. In Angela Ralli, Geert Booij, Sergio Scalise and Athanasios Karasimos (eds.), *On-line Proceedings of the Eighth Mediterranean Morphology Meeting (MMM8) Cagliari, Italy, 14–17 September 2011*, pages 254–268, Patras: University of Patras.

Wisdom, John. 1968. *Other Minds*. Berkeley: University of California Press.

Zwicky, Arnold M. 1992. Some Choices in the Theory of Morphology. In Robert D. Levine (ed.), *Formal Grammar: Theory and Implementation*, pages 327–371, Oxford: Oxford University Press.