Periphrasis and Morphology in LFG
Mary Dalrymple, University of Oxford
Workshop on Morphology, Waseda University, 21 July 2015

Matthews (1991), Börjars et al. (1997), Sadler & Spencer (2001): some of the cells of the Latin verbal paradigm are filled by a multi-word periphrase.

(1) Latin *laudo* 'I praise':

IMPERFECTIVE	Active	Passive
Present	laudat	laudatur
Past	laudabat	laudabatur
Future	laudabit	laudabitur
PERFECTIVE	Active	Passive
Present	laudavit	laudatus/a/um est
Past	laudaverat	laudatus/a/um erat
Future	laudaverit	laudatus/a/um erit

1 Definitions of periphrasis

Spencer (2012):

- A periphrasis must consist of more than one word.
- A periphrasis must realize a grammatical property.

Ackerman & Stump (2004):

- A lexeme may be realized synthetically (as a single syntactic atom) or periphrastically (by two or more syntactic atoms co-occurring in a c-structure).
- The contentive information associated with a periphrase is not determined by the contentive information associated with its individual, syntactically independent parts through the mediation of unification principles defined on syntactic structures; rather, the contentive information associated with a periphrase is specified morpholexically. That is, syntactic principles of constituency and linearity determine the distribution of a periphrase's individual parts, but not the functional information which that periphrase expresses.

2 Identifying periphrases

Ackerman & Stump (2004):

- Criterion I [feature intersection]: If an analytic combination C has a featurally intersective distribution, then C is a periphrase. [Latin]
- Criterion II [non-compositionality]: If the morphosyntactic property set associated with an analytic combination C is not the composition of the property sets associated with its parts, then C is a periphrase.
- Criterion III [distributed exponence]: If the morphosyntactic property set associated with an analytic combination C has its exponents distributed among C's parts, then C is a periphrase. [but see Spencer 2012]

Criterion II: English auxiliary system (Spencer, 2012):

(2) a. Perfect: He has eaten the cake.
HAVE PAST.PARTICIPLE

b. Passive: The cake was eaten.
BE PAST.PARTICIPLE

c. Progressive: He is eating the cake.
BE PRESENT.PARTICIPLE

Spencer (2001): French past tense formed from present tense auxiliary and a tenseless participle:

(3) Jean a lu ce livre.
have.3SG.PRS read.PASTPART this book
'Jean read this book.'

3 Otoguro (2012): Periphrasis and Japanese verb inflection

The negative polite past form of a verb is periphrastic (see also Spencer 2008):

(4) *taberu* 'eat':

	present positive	past positive	present negative	past negative
plain	taberu	tabeta	tabenai	tabenakatta
polite	tabemasu	tabemasita	tabemasen	tabemasen desita

(5) Otoguro (2012), example 15:

Taroo wa ringo o tabemasen desita.
TOP apple ACC eat.NEG.POLITE COPULA.POLITE.PAST

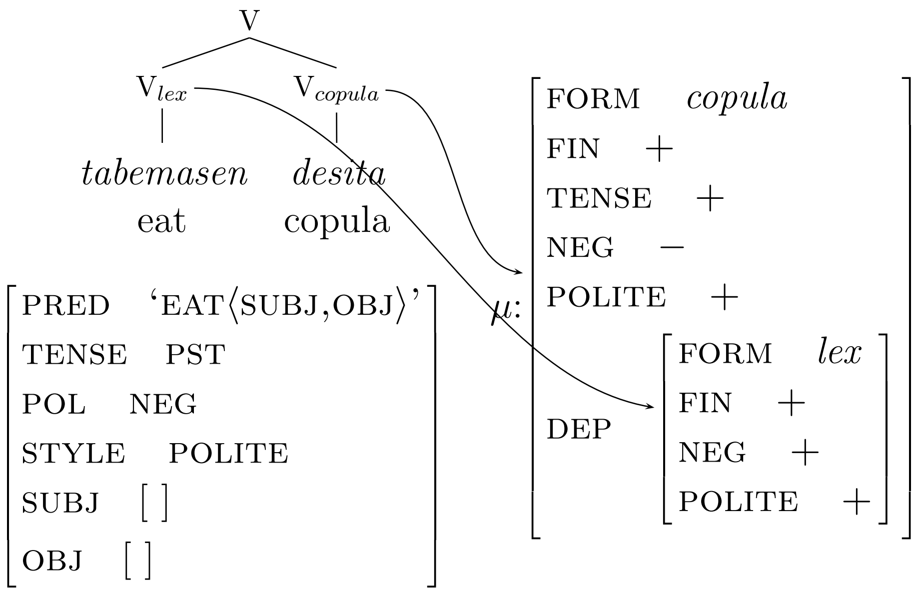
'Taroo didn't eat an apple.'

Analysis (Otoguro, 2012, (18), (19)):

(6) a. *desita* V_{copula}
$(\hat{*}_\mu \text{ FIN}) = +$
$(\hat{*}_\mu \text{ TENSE}) = +$
$(\hat{*}_\mu \text{ NEG}) = -$
$(\hat{*}_\mu \text{ POLITE}) = +$
$(\hat{*}_\mu \text{ LINK}) = +$
$(\hat{*}_\mu \text{ DEP NEG}) =_c +$
$(\hat{*}_\mu \text{ DEP POLITE}) =_c +$
$(\uparrow \text{TENSE}) = \text{PST}$
$(\uparrow \text{STYLE}) = \text{POLITE}$

b. *tabemasen* V_{lex}
$(\hat{*}_\mu \text{ FIN}) = +$
$(\hat{*}_\mu \text{ NEG}) = +$
$(\hat{*}_\mu \text{ POLITE}) = +$
$(\uparrow \text{POL}) = \text{NEG}$
$(\uparrow \text{STYLE}) = \text{POLITE}$
$(\uparrow \text{PRED}) = \text{'eat}\langle\text{SUBJ,OBJ}\rangle\text{'}$

(7)

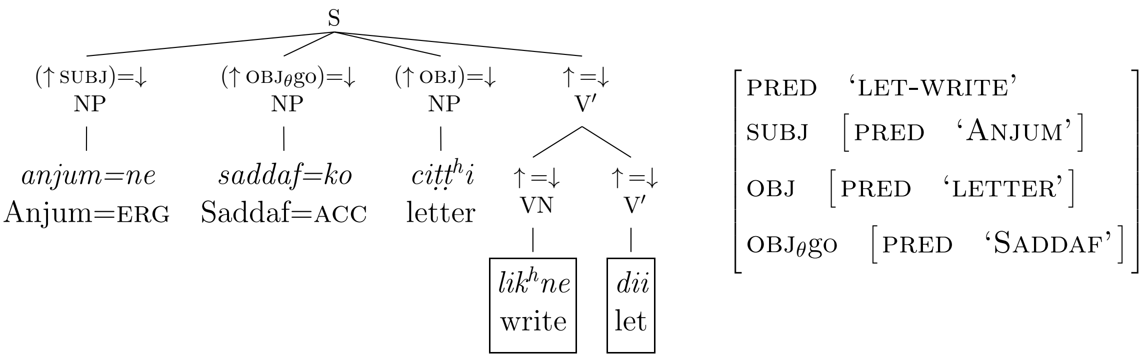


Successfully captures monoclausality of *tabemasen desita*.

4 Complex predicates

But what about complex predicates?

(8) Urdu, Butt (1995): more than one verb, monoclausal f-structure



- Complex predicates: more than one word, monoclausal, **formed in syntax**
- Verbal periphrases: more than one word, monoclausal, **formed in morphology** (cf. Ackerman & Stump 2004).

5 Periphrases in HPSG

Bonami & Samvelian (2015) show that 'perfect' forms in Persian are periphrastic, realized as an inflected form of *budan* 'be' and a perfect participle:

(9) Bonami & Samvelian (2015), example (2a):

Maryam	in	tâblo=râ	foruxte	bud.
	this	painting=RA	sell.PRF.PTCP	be.PST.3SG

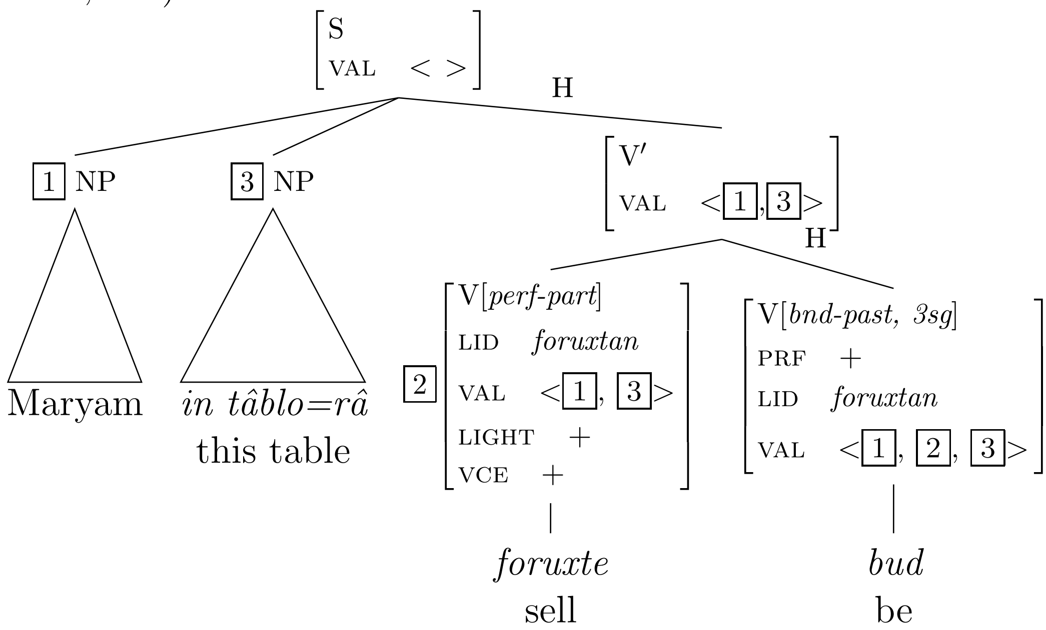
'Maryam had sold this painting.'

(10) Lexical description for *bud* in combination with *foruxtan* (Bonami & Samvelian, 2015, 358):

$$\begin{bmatrix} \textsc{phon} & \textit{bud} \\ \textsc{head} & \begin{bmatrix} \textsc{lid} & \textit{foruxtan} \\ \textsc{morsyn} & \begin{bmatrix} \textsc{tns} & \textit{past} \\ \textsc{prf} & + \\ \textsc{agr} & \textit{3sg} \\ \textsc{pol} & + \end{bmatrix} \end{bmatrix} \\ \textsc{arg-st} & \left\langle \text{NP}, \begin{bmatrix} \textsc{light} & + \\ \textsc{vce} & + \\ \textsc{lid} & \textit{foruxtan} \\ \textsc{morsyn} & \begin{bmatrix} \textsc{form} & \textit{part} \\ \textsc{prf} & + \\ \textsc{pol} & + \end{bmatrix} \end{bmatrix}, \text{NP} \right\rangle \end{bmatrix}$$

Tight relation between two verbs encoded by VCE + specification.

(11) Analysis of 'Maryam had sold this painting' (Bonami & Samvelian, 2015, 359):



Issues:

- A special form of *bud* is produced for every verb in the past perfect.
- The structure in (11) appears to involve complementation, and does not resemble the structure we would expect for a nonperiphrastic, single-word cell in the same paradigm.

6 Desiderata for an LFG treatment

- The f-structure of a periphrasis is indistinguishable from the f-structure of a nonperiphrastic (single-word) form in the same paradigm, except for the different inflectional features.
- The relation between the word forms in the periphrasis is established in the morphology component, specifically in the realization component.
- The analysis does not blow up the lexicon with word forms specific to a particular periphrastic cell in a paradigm.

7 The morphology relation $\mathcal{M}$

The full lexical entry for the plural noun *dogs* (Dalrymple & Mycock, 2011; Mycock & Lowe, 2013; Dalrymple et al., 2015):

(12) Full lexical entry for *dogs*:

s-form	$(\bullet\ \text{FM}) =$	dogs
c-structure category	$\pi(\bullet) =$	N
f-description	$(\uparrow \text{PRED}) =$	'dog'
	$(\uparrow \text{NUM}) =$	PL
	dog $\in$	$(\uparrow_{\sigma\iota}\ (\uparrow_{\sigma}\ \text{DF}))$
	pl $\in$	$(\uparrow_{\sigma\iota}\ (\uparrow_{\sigma}\ \text{DF}))$
p-form	/dɔgz/	

The morphology relation $\mathcal{M}$:

(13) $\mathcal{M}$ = {<s-form, p-form, category, f-description$_L$∪ f-description$_M$> |
LE <root (and idiosyncratic stem forms), f-description$_L$, LexemicIndex> ∧
R <LexemicIndex, s-form, p-form, m-features> ∧
D <LexemicIndex, m-features, category, f-description$_M$ >}

7.1 The Lexemic Entry *LI*

A **lexemic entry** is a three-place relation *LE* involving (1) the form of the root and any non-predictable stem forms; (2) an f-description$_L$ associated with the lexeme; and (3) the Lexemic Index.

(14) General form of lexemic entry:

LE <root & idiosyncratic stem forms, syntax & semantics, Lexemic Index>

(15) Lexemic entry for the lexeme DOG1:

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

(16) Lexemic entry for CHILD1:

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

(This simple f-description is a standin for the fully complete lexemic entry, which encodes syntactic, semantic, information-structural, and other information by means of *templates* (Dalrymple et al., 2004) enabling the statement of generalizations about classes and subclasses of lexemic entries.)

7.2 The Realization Relation R

The morphological realization relation R encodes a relation between a word form and its associated morphological features.

Definition: R is a set of four-place relations which we will call *m-entries*, associating a Lexemic Index, an s-form, and a p-form with a set of m-features.

(17) General form of m-entry:

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

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

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

- The m-entries for each language are defined entirely by the morphological realization component R.
- We make no assumptions about the nature of R; it is compatible with any means of associating m-features with p-forms and s-forms relative to a lexemic root.

7.3 The Functional Description Function D

The functional description function D maps a set of m-features to the appropriate c-structure category and f-description$_M$, given a Lexemic Index (cf. Kaplan & Butt 2002).

(19) General form of the description function D:

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

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

7.4 $\mathcal{M}$ Defined in Terms of D, LE, and R

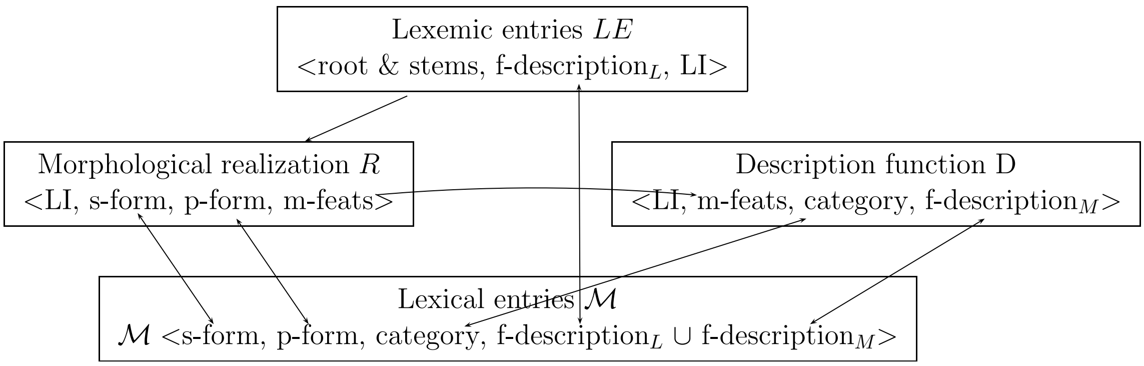
$\mathcal{M}$ is the set of all lexical entries

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

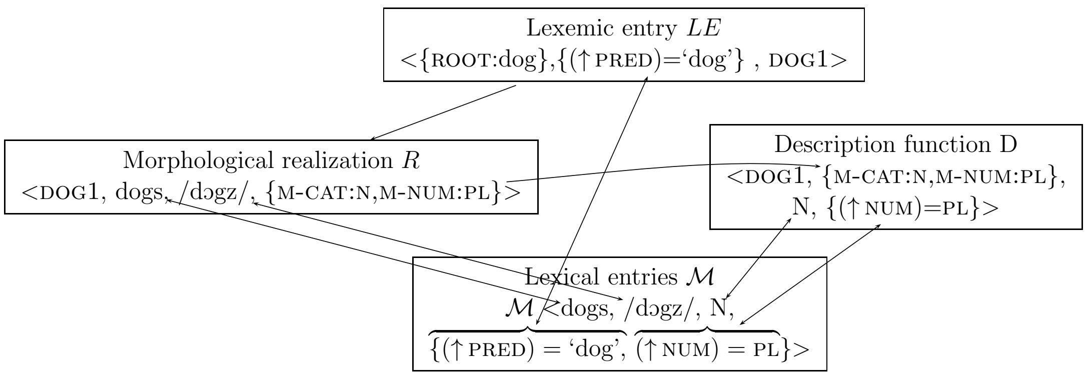
that meet the conditions imposed by LE, R, and D:

(21) $\mathcal{M}$ = {<s-form, p-form, category, f-description$_L$∪ f-description$_M$> |
LE <root (and idiosyncratic stem forms), f-description$_L$, LexemicIndex> ∧
R <LexemicIndex, s-form, p-form, m-features> ∧
D <LexemicIndex, m-features, category, f-description$_M$ >}

(22)



(23)



8 Periphrastic realizations

The definition of $\mathcal{M}$ in the previous section assumes a single-word m-entry in the realization relation R:

(24) R for nonperiphrastic m-entries:

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

Let us now assume that R can license more than one m-entry for a paradigm cell. We assume that there is a *primary* or *main* m-entry, which is the one that will be designated as carrying the f-description$_L$ associated with the lexemic entry, and any number of *secondary m-entries*, annotated with primes in (25).

(25) Revised R allowing for periphrastic or nonperiphrastic realizations of a paradigm cell:

R <m-entry, {m-entry′, m-entry″, ...}>

(26) Nonperiphrastic realizations have no secondary m-entries:

R <m-entry, ∅ >

Example: Nonperiphrastic positive past tense polite form of TABE1 (‘ate’):

(27) R <<TABE1, tabemasita, /tabemaɕita/,
{M-CAT:V, M-VTYPE:LEX, M-TENSE:PAST, M-STYLE:POLITE, M-POL:POS}>, ∅ >

Example: Nonperiphrastic negative present tense polite form of TABE1 (‘does not eat’):

(28) R <<TABE1, tabemasen, /tabemaseɴ/,
{M-CAT:V, M-VTYPE:LEX, M-TENSE:PRES, M-STYLE:POLITE, M-POL:NEG}>, ∅ >

Example: Periphrastic negative past tense polite form of TABE1 (‘did not eat’), schematically:

(29) <m-entry$_{tabemasen}$, {m-entry$_{desita}$}>

Assuming the following definitions:

(30) *tabemasen-m-features*:
{M-CAT:V, M-VTYPE:LEX, M-POL:NEG, M-STYLE:POLITE, M-FIN:+}

desita-m-features:
{M-CAT:V, M-VTYPE:COP, M-TENSE:PAST, M-NEG:−, M-STYLE:POLITE, M-DEP-NEG:+, M-DEP-POLITE:+}

we have the following periphrastic morphological realization:

(31) R <<TABE1, tabemasen, /tabemaseɴ/, *tabemasen-m-features*>,
{<TABE1, desita, /deɕita/, *desita-m-features*>} >

Definitions of D for Japanese:

(32) D_{cat} <LI, m-features, V_{lex}> if and only if {M-CAT:V, M-VTYPE:LEX} $\subseteq$ m-features.

D_{cat} <LI, m-features, V_{copula}> if and only if {M-CAT:V, M-VTYPE:COP} $\subseteq$ m-features.

(33) D_{feats} <LI, M-TENSE:_TNS, m-features, {($\hat{*}_\mu$FIN)=+, ($\hat{*}_\mu$TENSE)+, ($\uparrow$ TENSE)=_TNS}>.

(34) D_{feats} <LI, M-STYLE:POLITE, m-features, {($\hat{*}_\mu$POLITE)=+, ($\uparrow$ STYLE)=POLITE}>.

(35) D_{feats} <LI, M-POL:POS, m-features, {($\hat{*}_\mu$NEG)=−, ($\uparrow$ POL)=POS}>.
D_{feats} <LI, M-POL:NEG, m-features, {($\hat{*}_\mu$NEG)=+, ($\uparrow$ POL)=NEG}>.

(36) D_{feats} <LI, M-FIN:+, m-features, {($\hat{*}_\mu$FIN)=+}>.

(37) D_{feats} <LI, M-NEG:_NEG, m-features, {($\hat{*}_\mu$NEG)=_NEG}>.

(38) D_{feats} <LI, M-DEP-NEG:_NEG, m-features, {($\hat{*}_\mu$LINK)=+, ($\hat{*}_\mu$DEP NEG)=_NEG}>.

(39) D_{feats} <LI, M-DEP-POLITE:_POL, m-features, {($\hat{*}_\mu$DEP POLITE)=_POL}>.

(40) D_{feats} <LI, M-VTYPE, m-features, $\emptyset$>.

We must now adjust D to allow for the possibility of periphrastic morphological realizations, given this new definition of R.

(41) $\mathcal{M} = \cup\{\{$<s-form$_1$, p-form$_1$, category$_1$, f-description$_L \cup$ f-description$_{M1}$>,
<s-form$_2$, p-form$_2$, category$_2$, f-description$_{M2}$>,
⋮
<s-form$_n$, p-form$_n$, category$_n$, f-description$_{Mn}$>} |
LE <{root, stems}, f-description$_L$, LI> ∧
R <<LI, s-form$_1$, p-form$_1$, m-features$_1$ >,
{<LI, s-form$_2$, p-form$_2$, m-features$_2$ >,
⋮
<LI, s-form$_n$, p-form$_n$, m-features$_n$ >} > ∧
D <LI, m-features$_1$, category$_1$, f-description$_{M1}$ > ∧
D <LI, m-features$_2$, category$_2$, f-description$_{M2}$ > ∧
⋮
D <LI, m-features$_n$, category$_n$, f-description$_{Mn}$ >}

- The *LE* component is unchanged, since we are still retrieving information about a single lexeme.
- We now use the revised definition of *R*, which allows a paradigm cell to be realized periphrastically, by more than one m-entry.
- For each m-entry in *R*, we must map the m-features to the appropriate f-description, so we apply *D* to each m-entry.

For periphrastic *tabemasen desita* 'did not eat', as desired:

(42) $\mathcal{M} \supseteq$ {<tabemasen, /tabemaseɴ/, V$_{lex}$,
{(↑ PRED)='EAT⟨SUBJ, OBJ⟩' ∪ {($\hat{*}_\mu$FIN)=+, etc. (as in 6b)}>,
<desita, /deɕita/, V$_{copula}$, {($\hat{*}_\mu$FIN)=+, etc. (as in 6a)}>}
given the definitions of *D* above, and:

LE <{root:tabe}, {(↑ PRED)='EAT⟨SUBJ, OBJ⟩'}, TABE1> ∧

R <<TABE1, tabemasen, /tabemaseɴ/, *tabemasen-m-features*>,
{<TABE1, desita, /deɕita/, *desita-m-features*>} > ∧

D <TABE1, *tabemasen-m-features*, V$_{lex}$, {($\hat{*}_\mu$FIN)=+, etc. (as in 6b)}> ∧

D <TABE1, *desita-m-features*, V$_{copula}$, {($\hat{*}_\mu$FIN)=+, etc. (as in 6a)}>}

9 Observations and conclusion

- Following Otoguro's analysis, the f-structure for periphrastic *tabemasen desita* is monoclausal, and resembles the f-structures for nonperiphrastic realizations involving the same lexeme.
- The relation between the components of a periphrasis is established by the morphological realization component R.
- Our new definition of $\mathcal{M}$ adds a PRED-less lexical entry for the secondary m-entries (here, the copula) to the full set of lexical entries for the language. The form added in this way will generally be identical to the form required by periphrastic entries with the same features for other lexemes, so we do not blow up the lexicon with multiple versions of the same lexical entry for each lexeme.

References

Ackerman, Farrell & Gregory Stump. 2004. Paradigms and periphrastic expression: A study in realization-based lexicalism. In Louisa Sadler & Andrew Spencer (editors), *Projecting Morphology*, pp. 111–157. Stanford: CSLI Publications.

Bonami, Olivier & Pollet Samvelian. 2015. The diversity of inflectional periphrasis in Persian. *Journal of Linguistics* 51(2), pp. 327–382.

Börjars, Kersti, Nigel Vincent, & Carol Chapman. 1997. Paradigms, periphrases and pronominal inflection: A feature-based account. In Geert Booij & Jaap van Marle (editors), *Yearbook of Morphology*, pp. 155–180. Dordrecht: Kluwer.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Dissertations in Linguistics. Stanford: CSLI Publications. Revised and corrected version of 1993 Stanford University dissertation.

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

Dalrymple, Mary, Ronald M. Kaplan, & Tracy Holloway King. 2004. Linguistic generalizations over descriptions. In Miriam Butt & Tracy Holloway King (editors), *On-Line Proceedings of the LFG2004 Conference*. Stanford: CSLI Publications. URL `http://csli-publications.stanford.edu/LFG/9/lfg04.html`.

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

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/`.

Kaplan, Ronald M. & Miriam 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.

Matthews, Peter H. 1991. *Morphology*. Cambridge, UK: Cambridge University Press.

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/`.

Otoguro, Ryo. 2012. Constructional paradigm in constraint-based morphosyntax: A case of Japanese verb inflection. In Oana David, Kayla Carpenter, Florian Lionnet, Christine Sheil, Tammy Stark, & Vivian Wauters (editors), *Proceedings of the 38th Annual Meeting of the Berkeley Linguistics Society*, volume 38. Berkeley: Berkeley Linguistics Society.

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.

Spencer, Andrew. 2001. The paradigm-based model of morphosyntax. *Transactions of the Philological Society* 99(2), pp. 279–314.

Spencer, Andrew. 2008. Negation in Japanese: A case of morphosyntactic mismatch. *Lingua* 118(7), pp. 997–1017.

Spencer, Andrew. 2012. Sentence negation and periphrasis. In *Periphrasis: The Role of Syntax and Morphology in Paradigms*. Oxford: Oxford University Press/British Academy.