Inflectional periphrasis as collocation

Olivier Bonami[1]   Gert Webelhuth[2]

[1]Université Paris Sorbonne &
Institut Universitaire de France &
Laboratoire de Linguistique Formelle

[2]Goethe Universität Frankfurt

HPSG Workshop
Frankfurt, May 2012

Introduction

- Inflectional periphrases: multi-word constructions integrated in an inflectional paradigm.
- Inflectional periphrases are not the result of free syntactic combinations, but rather realize cells in the inflectional paradigms of lexemes.
  - ☞ Vincent and Börjars 1996, Börjars, Vincent, and Chapman 1997, Ackerman and Webelhuth 1998, Spencer 2001, Blevins 2001, 2007 (ms), Stump 2002, Sadler and Spencer 2001, Spencer 2003, Ackerman and Stump 2004, Stump 2006, Bonami and Samvelian 2009, Bonami and Webelhuth (in press)
- Our claim: None of theses proposals is compatible with all the desirable design properties of a theory of periphrasis as inflection.
- New proposal to solve this problem: periphrastic predicates as collocations.

Design properties of a theory of periphrasis

True periphrases integrate inflectional paradigms

- Not all multiword expressions with inflection-like content are inflectional periphrases.
- Persian has three ways of expressing progressivity:
    - Implicitly, by using an imperfective form

        (1) Maryam madrase mi-raft.
            Maryam school   IPFV-go.PST[3SG]

            'Maryam was going to school.'/'Maryam used to go to school.'

    - Combining a finite form of <u>dâštan</u> 'have' and a finite form of the main verb.

        (2) Maryam dâšt   madrase mi-raft.
            Maryam have.PST[3SG] school   IPFV-go.PST[3SG]

            'Maryam was going to school.'

    - Using the predicative adjective <u>mašqul</u> 'occupied' and an infinitive main verb

        (3) Maryam mašqul-e   madrase raft-an ast.
            Maryam occupied-EZ school   go-INF COP.PRS.3SG

            'Maryam was going to school.'

True periphrases integrate inflectional paradigms

- Despite identical semantics, only the second strategy is undisputably an inflectional periphrase:
    - No subjunctive progressive

(4) * Fekr   mi-kon-am   ke dâr-ad   be-dav-ad.
    thought IPFV-do-.PRS-1SG that have-PRS-3SG SBJV-run.PRS-3SG

    (intended) 'I think that he is running.'

(5) Fekr   mi-kon-am   ke mašqul-e   davidan bâš-ad.
    thought IPFV-do-1.SG that occupied-EZ run-INF be.SBJV-3.SG

    - No negative progressive

(6) a. * Maryam na-dâr-ad   (ne-)mi-dav-ad.
        Maryam NEG-have.PRS-3SG NEG-IPFV-run.PRS-3SG

    (intended) 'Maryam is not running.'

   b. Maryam mašqul-e   davidan nist.
      Maryam occupied-EZ run-INF NEG.COP.PRS.3SG

☞ Periphrases fill cells in a paradigm whose geometry is partly arbitrary

Auxiliaries have normal paradigms

- Many attemps to treat periphrases as ordinary syntax.
- Usually leads to systematic overgeneration.

		PERFECTIVE	IMPERFECTIVE	PERFECT
PRESENT		***	mi-xar-ad	xarid-e-ast
PAST	DIR.	xarid	mi-xarid	xarid-e bud
	IND.	xarid-e-ast	mi-xarid-e-ast	xarid-e bud-e-ast
SUBJUNCTIVE		be-xar-ad		xarid-e bâš-ad

Distribution of the Persian perfect periphrase (Bonami & Samvelian, 2009)

- Only way out:
    1. Either assume some kind of competition between morphology and syntax (e.g. Poser 1992, Bresnan 2001, Kiparsky 2005)
        ☞ Technically and conceptually problematic
        or assume that auxiliaries are by chance defective where morphology is available
        ☞ Strongly implausible
    2. Assume that all auxiliaries are deponent (here: [PRF −] forms expressing [PRF +])
        ☞ Strongly implausible

Arbitration between synthesis and periphrasis follows the logic of Pāṇini's principle

- Tundra Nenets nouns: declension is synthetic in general, periphrastic for local cases in the dual.

	SG	DU	PL
NOM	ti	tex°h	tiq
ACC	tim	tex°h	tí
GEN	tih	tex°h	tíq
DAT	ten°h	tex°h nyah	tex°q
LOC	tex°na	tex°h nyana	tex°qna
ABL	texød°	tex°h nyad°	texøt°
PROS	tew°na	tex°h nyamna	teqm°na

Absolute subparadigm of the Tundra Nenets noun TI 'male reindeer'
(Salminen 1997)

☞ This is despite the existence of a perfectly well-formed candidate synthetic form tex°h

Arbitration between synthesis and periphrasis follows the logic of Pāṇini's principle

- Czech verbs: Past conjugation is periphrastic except in the 3rd person.

	SG			PL		
	M	F	N	M	F	N
1	pekl jsem	pekla jsem	peklo jsem	pekli jsme	pekly jsme	pekla jsme
2	pekl jsi	pekla jsi	peklo jsi	pekli jste	pekly jste	pekla jste
3	pekl	pekla	peklo	pekli	pekly	pekla

Past of the verb PÉCT 'to bake'

- This is despite the existence of a perfectly well-formed (and otherwise obligatory) 3rd person copula: 3SG <u>je</u>, 3PL <u>jsou</u>
- ☞ Favors a view where arbitration between synthesis and periphrasis happens within the inflectional system.

Periphrasis is independent of phrase structure

- The parts of a periphrase can stand in varying phrase-structural configurations (Bonami & Webelhuth, in press):

(7) a. dass das Buch jemand [VC gekauft hat]
CPZR the book nobody buy.PST.PCPL have.PRS.3SG
'that nobody bought the book' (German)

b. Paul [VP a lu ce livre].
Paul have.PRS[3SG] read.PST.PCPL that book
'Paul read that book.' (French)

c. Paul [VP has [VP read that book]]. (English)

d. [S Maryam dâšt [S madrase mi-raft]].
Maryam have.PST[3SG] school IPFV-go.PST[3SG]
'Maryam was going to school.' (Persian)

e. [S Toj njama [CP da e v kâštata]].
he not-have THAT be.PRS[3SG] v house.DEF
'He will not be in the house.' (Bulgarian)

Grammatical relations between parts of a periphrase

- The parts of a periphrase stand in a syntactic head-argument or head-modifier relationship.

(8) a. John has [$_{VP}$ left the room].
     b. [$_{AP}$ more important]

- Syntactic operations can affect parts of a periphrase, as long as they do not disrupt the grammatical relations.

(9) Subject-auxiliary inversion
     a. Has John [$_{VP}$ left the room]?
     b. May John [$_{VP}$ leave the room]?

(10) Topicalization
     a. [$_{VP}$ Left the room] [$_{S}$ I believe [$_{S}$ he has __]].
     b. [$_{VP}$ Leave the room] [$_{S}$ I believe [$_{S}$ he may __]].

Word and Paradigm morphology, phrase-structure based syntax

- Inflectional systems are best described in word-and-paradigm approaches.
  - ☞ See among many others Hockett 1954, Robins 1959, Matthews 1972, Anderson 1992, Zwicky 1992, Aronoff 1994, Stump 2001, Blevins 2006
- Syntactic systems are best described in phrase-structural terms, as incrementally built combinations of signs.
  - ☞ See among many others Harman 1963, Bresnan 1978, Gazdar, Klein, Pullum & Sag 1985, Pollard & Sag 1987, 1994, Steedman 1996
- ☞ In Stump's (2001) terms, inflection is inferential-realizational, syntax is lexical-incremental.
- An adequate theory of periphrasis should be compatible with such a position.

Current approaches to periphrasis

- None of the existing proposals is satisfactory.
  - Ackerman and Webelhuth 1998: syntax of periphrasis is too inflexible to handle extraction or modification.
  - Sadler and Spencer 2001, Ackerman and Stump 2004: opposite problem: syntax too unconstrained or details not worked out.
  - Bonami and Samvelian 2009: the morphological component fails to be completely realizational.
  - Bonami and Webelhuth (in press): Panini's Principle does not apply within the morphology, can't deal with periphrases that rest on the modifier-head relation.

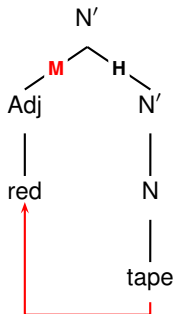An alternative: periphrases as collocations

The intuition

- The main element of a periphrase requires the presence of a selector in the same local environment.
- This is reminiscent of the mutual cooccurrence requirements we find in collocations.
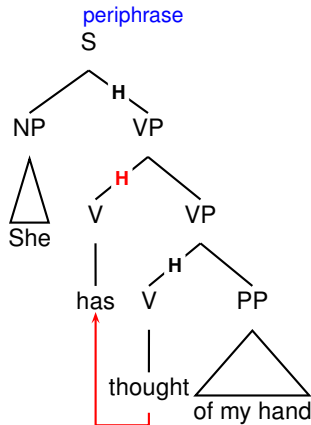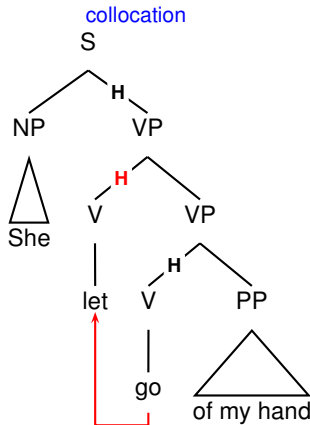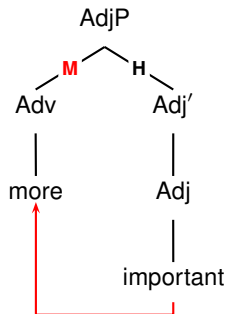
The intuition

- The same goes for modification structures.

A limited form of collocation

- Prior work on collocations in HPSG: Sailer 2000, Soehn & Sailer 2003, Soehn 2006, Richter & Sailer 2009.
- We adopt an analysis in the spirit of Soehn & Sailer 2003, but with important modifications.
    - Words can carry a REVerse-SELection requirement.
    - This amounts to asking for a selector to be present.
    - The distance between selector and selectee can be as long as the grammar allows independently for that kind of selection relation.
- Inflection rules may produce REV-SEL requirements.
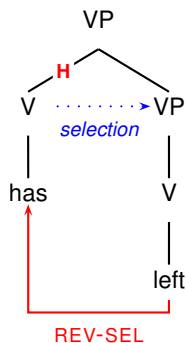
Consequences of reverse selection

- If a word carries a REV-SEL requirement, then it (or one of its projections) should be selected by a word whose morphological description unifies with that requirement.
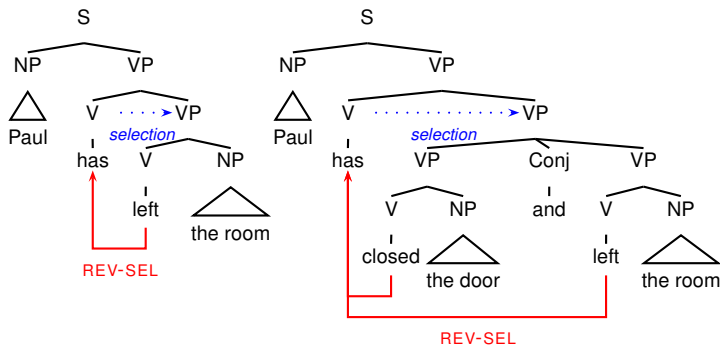
# Consequences of reverse selection (continued)

- Since we rely on the HPSG theory of selection, extraction of parts of periphrases is predicted to be possible without any further stipulations.
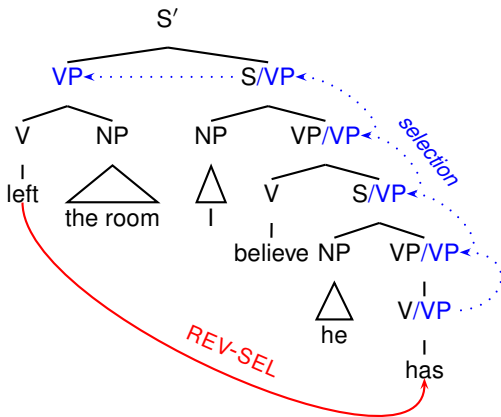
Periphrasis as collocation: the details

INFL and HEAD

- Introduction of the feature INFL on words:
  - INFL is what morphology realizes, HEAD is what syntax and semantics examine.
  - INFL features often have corresponding HEAD features, but there can be mismatches.
  - For ordinary words, INFL relates to HEAD lexically.
  - In periphrases, this relation is mediated by syntax.
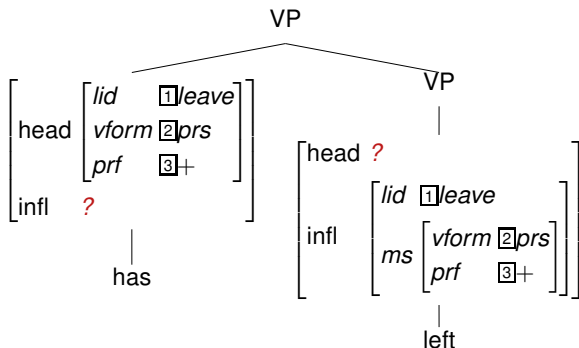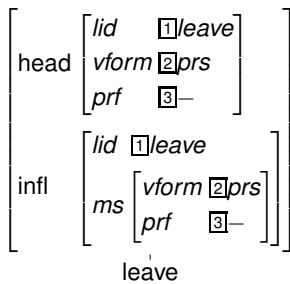
INFL and HEAD on periphrase-inducing main verbs

- The periphrase-inducing main verb needs to be a nonfinite form in terms of HEAD, because it heads a nonfinite VP
  - ☞ Can carry constituent negation: He hasn't [not left]
- But it needs to have a head value distinct from that of an ordinary present participle, so that the auxiliary can select specifically for it.
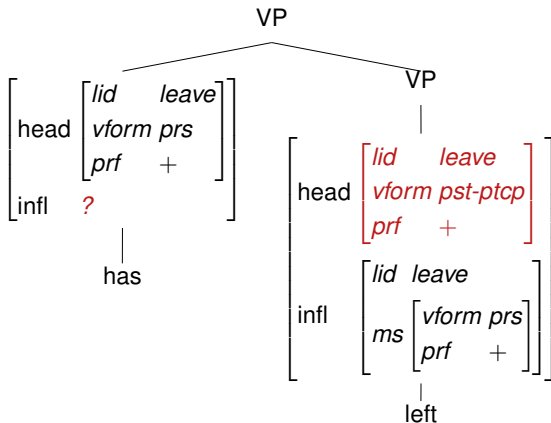
INFL and HEAD on auxiliaries

- The auxiliary needs to have the right INFL features to ensure its correct inflection.
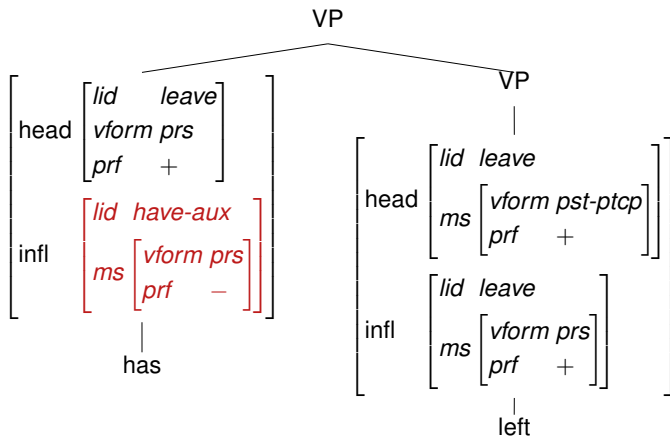
Information flow in periphrases

- We now need to ensure the right information flow:
    - The blue path:
        - Constraint on periphrastic words
        - Inheritance constraint on auxiliary lexemes
    - The red path:
        - REV-SEL requirement induced by inflection rule, cashed out on the auxiliary
        - Constraint on nonperiphrastic words, applied to auxiliary
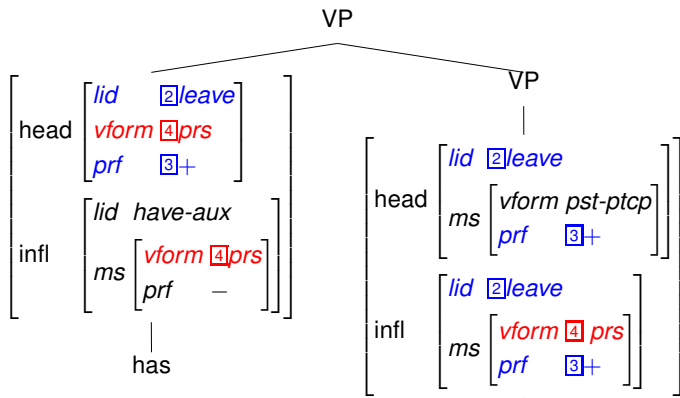
The Reverse Selection Principle

- We formulate the Reverse Selection Principle in terms of selection for INFL from a set-valued word level feature REV-SEL:

(11) Projection
   a. Every sign is a projection of itself
   b. A phrase is a projection of its head
   c. A coordination is a projection of each of its daughters

(12) Selection
   a. A sign selects all signs whose *synsem* occur on its ARG-ST
   b. A sign selects any sign whose *synsem* occurs on its MOD

(13) Reverse selection principle
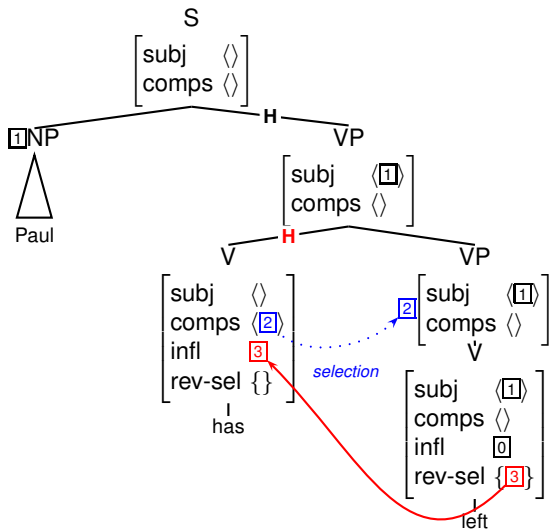   If a word *w* carries a reverse selection requirement *s* in its REV-SEL, then *s* must be token-identical to the INFL value of a word *w'* selecting for a projection of *w*.

The Reverse Selection Principle illustrated

Producing reverse selection requirements

- We embed a version of Paradigm Function Morphology (Stump, 2001) as a morphological component of our HPSG grammar.
- Realization rules may:
  - modify the phonological representation of their input
  - add reverse selectional requirements on the syntactic context.

$$\begin{bmatrix} \text{PHON} & X \\ \text{REV-SEL} & \{\} \end{bmatrix}, \sigma : \begin{bmatrix} \text{VFORM} & pst\text{-}ptcp \\ \text{PRF} & - \end{bmatrix} \longrightarrow \begin{bmatrix} \text{PHON} & Xed \\ \text{REV-SEL} & \{\} \end{bmatrix}$$

$$\begin{bmatrix} \text{PHON} & X \\ \text{REV-SEL} & \{\} \end{bmatrix}, \sigma : \begin{bmatrix} \text{PRF} & + \end{bmatrix} \longrightarrow \begin{bmatrix} \text{PHON} & \textbf{refer}\left(X, \sigma! \begin{bmatrix} \text{VFORM} & pst\text{-}ptcp \\ \text{PRF} & - \end{bmatrix} \right) \\ \\ \text{REV-SEL} & \left\{ \begin{bmatrix} \text{LID} & have\text{-}aux \\ \text{MORSYN} & \sigma! \begin{bmatrix} \text{PRF} & - \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix}$$

☞ left in has left is not a past participle, but a present perfect whose phonology is referred to that of a past participle.

Illustration

Periphrastic vs. nonperiphrastic words

- Some features project from INFL to HEAD if they are not used for periphrastic expression of some other feature.

$$(14) \quad word \rightarrow \left(\begin{bmatrix} \text{REV-SEL} & \{\} \end{bmatrix} \leftrightarrow \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{VFORM} & \boxed{1} \end{bmatrix} \\ \text{INFL} & \begin{bmatrix} \text{MS} & \begin{bmatrix} \text{VFORM} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right)$$
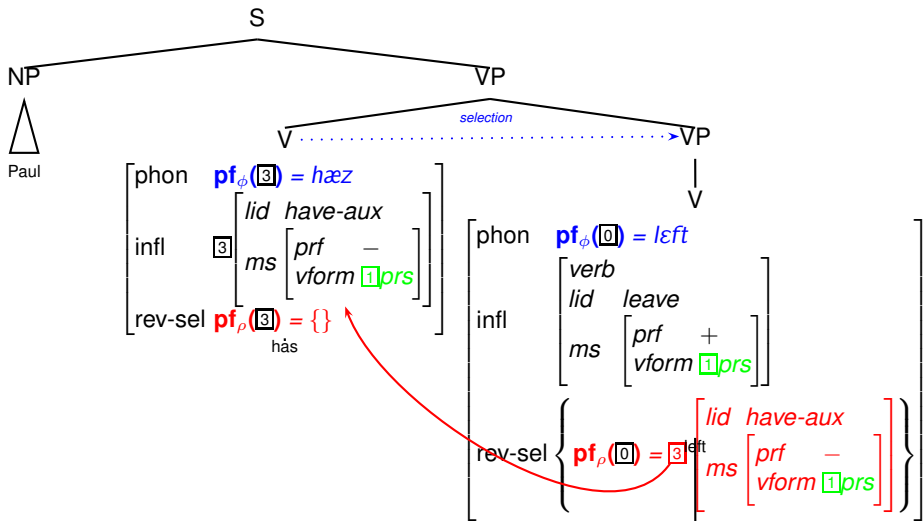
Plain vs. auxiliary lexemes

- Some features project from INFL to HEAD only for nonauxiliary lexemes

(15) *nonaux-lxm* →
$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{LID} & \boxed{1} \\ \text{PRF} & \boxed{2} \end{bmatrix} \\ \text{INFL} & \begin{bmatrix} \text{LID} & \boxed{1} \\ \text{MS} & \begin{bmatrix} \text{PRF} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$
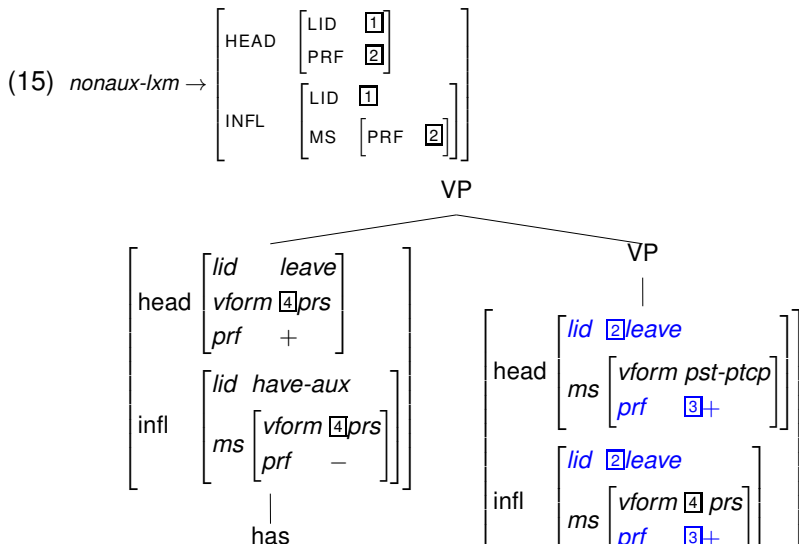
Lexical entries for auxiliaries

$$
\begin{bmatrix}
\textit{lexeme} \\[4pt]
\text{HEAD} \quad \begin{bmatrix} \text{LID} & \boxed{1} \\ \text{PRF} & \boxed{3} \end{bmatrix} \\[14pt]
\text{INFL} \quad \begin{bmatrix} \text{LID} & \textit{have-aux} \\ \text{MS} & \begin{bmatrix} \text{PRF} & - \end{bmatrix} \end{bmatrix} \\[14pt]
\text{CONT} \quad \boxed{4} \\[10pt]
\text{ARG-ST} \quad \left\langle \boxed{5}, \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{LID} & \boxed{1} \\ \text{VFORM} & \textit{pst-ptcp} \\ \text{PRF} & \boxed{3}+ \end{bmatrix} \\ \text{SUBJ} & \langle \boxed{5} \rangle \\ \text{CONT} & \boxed{4} \end{bmatrix} \right\rangle
\end{bmatrix}
$$
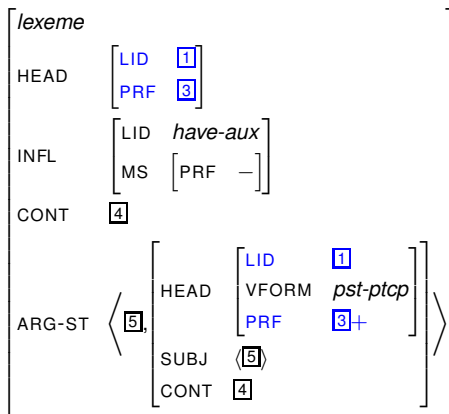
Auxiliary lexical entries: illustration
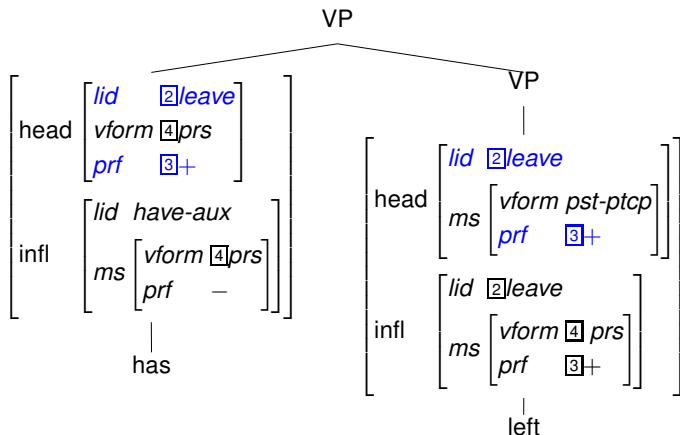
☞ No spurious ambiguity: because it selects for a [PRF +] complement, the auxiliary cannot combine with an ordinary ([PRF −]) participle.

Stacking periphrases

Stacking periphrases: the details

- The rule for perfect should not be applicable to progressive forms, so that we can prevent *is having left.

$$
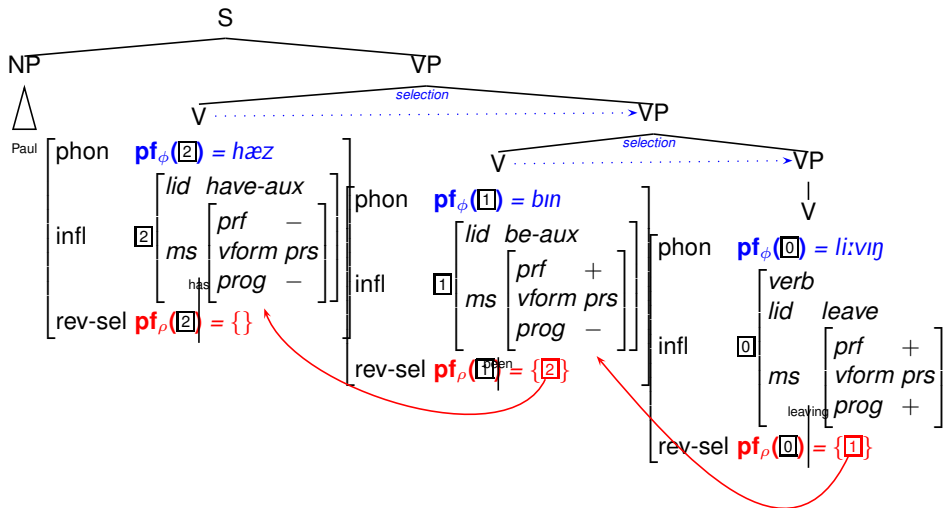\begin{bmatrix} \text{PHON} & X \\ \text{REV-SEL} & \{\} \end{bmatrix}, \sigma : \begin{bmatrix} \text{PRF} & + \\ \text{PROG} & - \end{bmatrix} \longrightarrow \begin{bmatrix} \text{PHON} & \textbf{refer}\left(X, \sigma! \begin{bmatrix} \text{VFORM} & pst\text{-}ptcp \\ \text{PRF} & - \end{bmatrix} \right) \\ \\ \text{REV-SEL} & \left\{ \begin{bmatrix} \text{LID} & have\text{-}aux \\ \text{MORSYN} & \sigma! \begin{bmatrix} \text{PRF} & - \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix}
$$

- Compare:

$$
\begin{bmatrix} \text{PHON} & X \\ \text{REV-SEL} & \{\} \end{bmatrix}, \sigma : \begin{bmatrix} \text{PROG} & + \end{bmatrix} \longrightarrow \begin{bmatrix} \text{PHON} & \textbf{refer}\left(X, \sigma! \begin{bmatrix} \text{VFORM} & prs\text{-}ptcp \\ \text{PROG} & - \end{bmatrix} \right) \\ \\ \text{REV-SEL} & \left\{ \begin{bmatrix} \text{LID} & be\text{-}aux \\ \text{MORSYN} & \sigma! \begin{bmatrix} \text{PROG} & - \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix}
$$

Periphrasis by modification

$$\begin{bmatrix} \text{PHON} & X \\ \text{REV-SEL} & \{\} \\ \text{LID} & bad \end{bmatrix}, \sigma : \begin{bmatrix} \text{GRADE} & comp \end{bmatrix} \longrightarrow \begin{bmatrix} \text{PHON} & worse \\ \text{REV-SEL} & \{\} \end{bmatrix}$$

$$\begin{bmatrix} \text{PHON} & X \\ \text{REV-SEL} & \{\} \\ \text{LID} & class\text{-}A \end{bmatrix}, \sigma : \begin{bmatrix} \text{GRADE} & comp \end{bmatrix} \longrightarrow \begin{bmatrix} \text{PHON} & X\,er \\ \text{REV-SEL} & \{\} \end{bmatrix}$$

$$\begin{bmatrix} \text{PHON} & X \\ \text{REV-SEL} & \{\} \end{bmatrix}, \sigma : \begin{bmatrix} \text{GRADE} & comp \end{bmatrix} \longrightarrow \begin{bmatrix} \text{PHON} & X \\ \text{REV-SEL} & \left\{ \begin{bmatrix} \text{LID} & more \end{bmatrix} \right\} \end{bmatrix}$$

AdjP

Adv

$$\begin{bmatrix} \text{phon} & \mathbf{pf}_\phi(\boxed{1}) = m\textit{ɔːɹ} \\ \text{infl} & \boxed{1} \begin{bmatrix} adv \\ lid \quad more \end{bmatrix} \\ \text{rev-sel} & \mathbf{pf}_\rho(\boxed{1}) = \{\} \end{bmatrix}$$

Adj′

Adj

$$\begin{bmatrix} \text{phon} & \mathbf{pf}_\phi(\boxed{0}) = imp\textit{ɔːɹtnt} \\ \text{infl} & \boxed{0} \begin{bmatrix} adj \\ lid \quad important \\ ms \begin{bmatrix} grade \quad comp \end{bmatrix} \end{bmatrix} \\ \text{rev-sel} & \mathbf{pf}_\rho(\boxed{0}) = \boxed{1} \end{bmatrix}$$
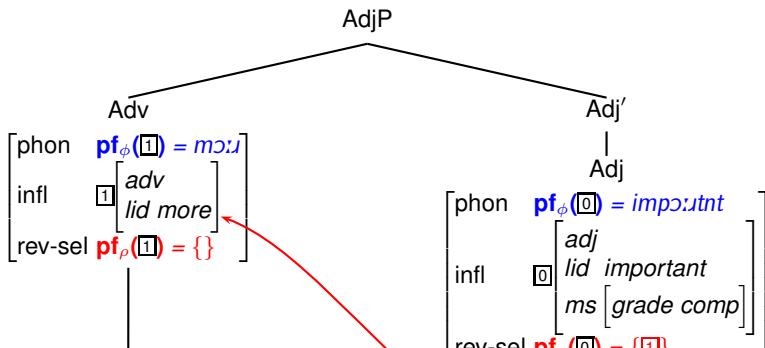
Avoiding unwarranted generalizations

- It is tempting to elevate constraints on feature percolation to the level of principles.
- This is unwarranted: much cross-linguistic and language-internal variation.
    - In familiar situations, paradigmatic opposition between the form of the main verb and the form of the auxiliary. In the Persian progressive, tense and mood expressed jointly on the main verb and auxiliary.

$$
\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{VFORM} & \boxed{1} \\ \text{PROG} & + \end{bmatrix} \\ \text{INFL} & \begin{bmatrix} \text{VFORM} & \boxed{1} \end{bmatrix} \end{bmatrix}
\qquad
\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{VFORM} & \boxed{1} \end{bmatrix} \\ \text{INFL} & \begin{bmatrix} \text{VFORM} & \boxed{1} \\ \text{PROG} & + \end{bmatrix} \end{bmatrix}
$$

    - In familiar situations, periphrastic expression of some feature relies on the exponents for a different feature set on the auxiliary. In Tundra nenets, local case is expressed by the congruent local case of the auxiliary postposition.

$$
\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{CASE} & gen \end{bmatrix} \\ \text{INFL} & \begin{bmatrix} \text{CASE} & \boxed{1} \end{bmatrix} \end{bmatrix}
\qquad
\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{CASE} & \boxed{1} \end{bmatrix} \\ \text{INFL} & \begin{bmatrix} \text{CASE} & \boxed{1}\,local \end{bmatrix} \end{bmatrix}
$$

Conclusions

Conclusion

Inflectional periphrases

- are not the result of free syntactic combinations.
- Rather: they realize cells in the inflectional paradigms of lexemes.
- None of the previous proposals is compatible with all the desirable design properties of a theory of periphrasis as inflection.
- New proposal to solve this problem: periphrastic predicates as collocations.

Conclusions

- **Syntax:**
  - Words can carry a REVerse-SELection requirement.
  - This amounts to asking for a specific selector to be present in the local environment of the word.
  - The distance between selector and selectee can be as long as the grammar allows independently for the selection relation involved.

- **Morphology:**
  - The paradigm function may produce
    - phonological effects, and in addition
    - REV-SEL requirements.

The theory captures the major desiderata for a theory of periphrasis as inflection:

- The two or more exponents can stand in various syntactic relationships.
- The degree of locality of these relationships is independent of periphrasis.
- Gaps in the paradigms of auxiliaries follow from paradigm structure.
- The morphological component is realizational rather than incremental.
- Arbitration between synthesis and periphrasis is decided within the morphological component.