Uninflecting and uninflectable lexemes: implications for paradigm structure

Andrew Spencer
University of Essex

1 Uninflecting and uninflectable lexemes

In inflecting languages some (classes of) lexemes are **uninflecting**: *of, the, almost,* …, but some members of otherwise inflecting classes may (unexpectedly) fail to inflect, i.e. they show **uninflectability**. Russian nouns distinguish 6 x 2 = 12 case/number forms but about 3000 are indeclinable (uninflectable): all the forms of their paradigm are identical to the root, e.g. *kenguru* 'kangaroo'. We distinguish such uninflectable lexemes from uninflecting lexemes on the one hand and defective lexemes (e.g. Russian *mečta* 'dream', lacking gen.pl) on the other. Uninflectable lexemes can occur in all the contexts open to inflecting lexemes (unlike defective lexemes), but the form is invariable, e.g. Russian (where UNINFL indicates and uninflecting or uninflectable form):

(1) a. odin vombat/kenguru
one.NOM.M.SG wombat[M].NOM.SG/kangaroo[UNINFL]
'one wombat/kangaroo'

b. s ètimi vombat-**ami**/kenguru
with these.INSTR.PL wombat[M].INSTR.PL/ kangaroo[UNINFL]
'with these wombats/kangaroos'

Lexemes can also be partially (un)inflectable: Polish *muzeum* inflects in the plural but is indeclinable in the singular; verbs of the English *hit* class have only *-s* and *-ing* inflections; Macedonian adjectives such as *kasmetlija* 'lucky, sg', *kasmetlii* 'lucky, pl' fail to inflect for [GENDER:{m,f,n}] in the singular.

2 Typology of uninflectedness

Russian KENGURU is uninflectable in all its occurrences — lexical uninflectability. However, English KANGAROO is inflectable except in compounds: *kangaroo(*s) tails* . This is *constructional uninflectability*, in which an otherwise inflecting lexeme has to be realized as a non-inflecting word form, or even as an uninflected bound stem, as in many cases of noun incorporation. A further case of constructional uninflectability is the predicative adjective in German:

(2) a. Ich bin ein **kleines** Känguru
I am a little.N.SG.NOM/ACC kangaroo
'I am a little kangaroo'

b. das Känguru ist **klein**(*-e/*-es/…)
the kangaroo is little[UNINFL]
'The kangaroo is little'

A particularly interesting case is that of indeclinable foreign names in Russian (though not necessarily in other Slav languages). Many borrowed or cited foreign words, and especially names,

have a phonological shape which is difficult to accommodate to the Russian morphological system, like *kenguru* (no native nouns have a stem ending in *-u*) or *Si (Czin′pin)* 'Xí (Jìnpíng)'. Other foreign names, however, can easily be inflected as though they were Russian: *Bil* 'Bill', *Bil-a* 'genitive singular' (cf *Kiril*, *Kiril-a*), or *Klinton* 'Clinton', *Klinton-a* 'gen. sg.' (cf *Solženicyn*, *Solženicyn-a*). However, female referent names do not have phonological forms such as *Klinton*, and native given names do not end in *-i* for either sex, so that *Xilari Klinton* 'Hillary Clinton' is indeclinable, like *Si* 'Xí':

(3) a. reč′ Xilari Klinton
speech Hillary[UNINFL] Clinton[UNINFL]
'Hillary Clinton's speech'

b. reč′ Bil-a Klinton-a
speech Bill[M]-GEN.SG Clinton[M]-GEN.SG
'Bill Clinton's speech'

c. reč′ Si Czin′pin-a
speech Xí[M] Jìnpíng[M]-GEN.SG
'Xí Jìnpíng's speech'

where *Czin′pin*, but not *Si*, inflects. Cases such as these seem to be intermediate between lexical and constructional uninflectability.

3 Theoretical questions

I address two of the several questions that a theoretical model of morphology has to answer in order to accommodate uninflectability, and link these to Question 3.

1. How can an uninflected form of a lexeme be treated by the morphosyntax as though it were fully inflected (and what prevents **all** lexemes from behaving in this way)?

2. In cases of constructional uninflectability with inflectable lexemes, such as the German predicative adjective construction, what is the morphosyntactic description of the uninflected lexical form? How is that form accessed by the grammar?

3. In the PFM2 class of models, what is the interface form between the entry for an uninflecting lexeme and the syntactic terminal it occupies, i.e. what 'word form' of a lexeme such as ALMOST undergoes 'lexical insertion'?

4 A PFM2-class approach

With paradigm-driven approaches, discussion of the morphology-syntax interface seldom asks how a completely uninflecting lexeme, such as an English prepositions, is represented in the syntax. Specifically, how do we interpret the notion 'word form of lexeme $\mathscr{L}$' where $\mathscr{L}$ belongs to a class which has no inflectional paradigm, in a model in which the morphology-syntax interface is supposed to be mediated through the inflectional paradigm, as in PFM2? Sag (2012, 119) is unusual in recognizing and addressing this problem. In SBCG, uninflecting lexemes are shifted to the type *word (form)* by a Zero Inflection Construction. This effectively treats such words (even particles, *oh, y'know,* …) as trivially inflected forms of lexemes with one-celled paradigms. While Sag's proposal permits uninflecting words to enter the syntax it does not directly address the problem of uninflectable lexemes, whether lexically uninflectable or constructionally.

PFM2 distinguishes crucially between CONTENT and FORM paradigms (Π_C, Π_F). Π_C defines all the syntactically accessible inflectional contrasts a lexeme is obliged to make, Π_F defines the morphophonological forms expressing those contrasts, a ***Correspondence*** function ***Corr*** specifies the mapping $\Pi_C \mapsto \Pi_F$, in part defined by the function ***pm*** defined over the feature sets, Σ, T of Π_C, Π_F. In addition, we make explicit the obvious assumption that every lexeme has a lexical (lexemic, dictionary) entry specifying a basic morphophonological form, or 'root' (i.e. a default lexical stem, STEM_0). For a completely uninflecting lexeme (class) we take $\Sigma = \text{T} (= \Pi_C = \Pi_F) = \emptyset$. This applies to, say, English prepositions or to uninflecting coverbs in languages with extensive LVCs. For uninflectable, non-defective members of an otherwise inflecting lexical class we take $\Sigma (= \Pi_C) \neq \emptyset$, T $(= \Pi_F) = \emptyset$, e.g. *kenguru* lexemes.

We now extend the definition of the ***Corr*** function minimally. Consider the application of ***Corr*** to a lexeme with lexemic index/LID λ, that is associated with no CONTENT paradigm features, Σ, at all (i.e. an uninflecting lexeme, say, English ALMOST). Stump's definition of ***Corr*** maps the Π_C cells to corresponding Π_F cells via the Paradigm Function, $\text{PF}(\langle \text{Z}, \tau \rangle)$, which maps the root of the lexeme, Z, and the form feature set, τ, to a cell in Π_F, (4).

(4) Given λ the lexical index of any lexeme/lexemic entry, Λ = the complete set of lexemic entries (= the lexicon).
Then $\forall \sigma \in \Sigma$ (possibly null), $\forall \lambda \in \Lambda$, $\textbf{\textit{Corr}}(\lambda, \sigma) = \text{PF}(\langle \text{Z}, \textbf{\textit{pm}}(\sigma) \rangle)$

However, this mapping will be undefined as things stand for cases where $\Sigma = \text{T} = \emptyset$. We therefore assume mapping (5), the Default Exponence Principle, as the default morphosyntactic expression of all lexemes defined over null feature sets.

(5) Default Exponence Principle (DEP)
$\text{PF}(\langle \text{Z}, \emptyset \rangle) = \text{Z} = \text{STEM}_0(\lambda)$

Together, (4, 5) state that the default realization of all lexemes is STEM_0 ($\equiv$ Z, the lexical root). For inflecting lexemes the DEP (5) is overridden by the more specific ***Corr(espondence)*** function. However, for an uninflecting lexeme DEP has the same effect as Sag's Zero Inflection Construction. This means that we do not require uninflectable lexemes to have a non-null inflectional paradigm. The uninflected lexemes whose form is defined by (4, 5) have just a root form but no paradigm, CONTENT or FORM.

For Russian uninflecting KENGURU, KLINTON[F], SI etc, Π_C defines 12 cells, as expected. However, we assume a special application of ***pm*** triggered by a lexical class feature under which the Π_F is undefined (for partially uninflectable lexemes the relevant part of Π_F is undefined). This will mean that the ***Corr*** function will deliver the value $\text{PF}(\langle \text{Z}, \emptyset \rangle)$. By (5) this means that the uninflectable lexeme reverts to its root form for all or part of its CONTENT paradigm.

Note that this account correctly characterizes *kenguru*-words as (highly) irregular: the otherwise default mapping to FORM paradigm is overridden and undefined. This has consequences for entropy-based perspectives on paradigm structure. Entropy can only be defined over cells in FORM paradigms (CONTENT paradigms are maximally regular by definition). But for uninflectable lexemes we seem to minimize entropy, giving the false impression that such lexemes are highly regular. On my approach, entropy measures are simply undefined for uninflecting/uninflectable lexemes.

We have accounted for non-inflecting lexemes with exactly the same machinery, and shown how ordinary inflecting words are the result of a specific application of the ***Corr*** function overriding the DEP, answering Question (1) above. At the same time we have begun to address Question (2): the uninflected forms of lexemes that appear in compounds or German predicative adjective constructions are also root forms. The most direct way to handle this is to assume

that such a construction includes an interface specification overriding the normal morphology-syntax interface under which lexical insertion is defined over cells in the Π_C. Instead, such a construction has to be so formulated that it specifies the Π_C of any lexeme corresponding to the appropriate syntactic terminal as 'locally undefined' with respect to that construction. How this is to be implemented formally depends on precisely how the morphology-syntax interface is formalized.[1]

The approach makes the prediction that there can be no defective lexemes such as Russian MEČTA. A defective lexeme is one which unexpectedly has an unfillable, undefined Π_C cell, and a fortiori a corresponding gap in Π_F. However, given the DEP, such cells should be treated like (the unique form of) an uninflecting lexeme, and so should revert to the root form, contrary to fact. This means that additional machinery will have to be developed to actively prevent defective forms from getting realized. Given the very peculiar ontological status of defective lexemes/cells, this seems to be the correct conclusion.

References

Sag, Ivan A. 2012. Sign-Based Construction Grammar: an informal synopsis. In Hans C. Boas & Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, 69–202. Stanford, CA: CSLI Publications.

[1] Polish *muzeum*-words have to be given special treatment: their uninflectable singular form, *muzeum*, is not presumably the STEM_0 form. Such cases appear to be rather rare.