Lexical Relations

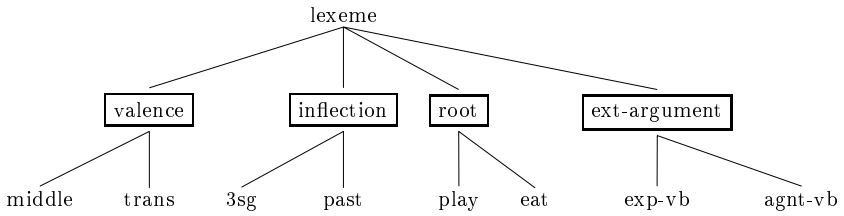
1 Introduction

The notes below give an overview of an approach to morphology presented in Jean Pierre Koenig, 1999, Lexical Relations. CSLI Publications.

Koenig is particularly concerned with productivity and partial idiosyncrasies. He argues for a model which will allow speakers to interpret word forms they have never come across before. He argues that there should be structural slots in the definition of words. Since this is not currently the case in HPSG, lexical productivity cannot result from various ways of filling structural positions. Koenig also modifies the hierarchical lexicon allowing for type underspecification and on-line category construction.

2 On-Line Type Construction

There is a multidimensional hierarchy of lexemes of the kind below:



From the stored lexical hierarchy we can only retrieve underspecified lexical entries like the following entry for *play*:

$$\begin{bmatrix} play \\ \text{STEM-PHON} \ \langle \text{ play} \rangle \\ \text{CONT} \begin{bmatrix} \text{NUCLEUS} \begin{bmatrix} play\text{-}rel \\ \text{ACT} \ \boxed{1} \\ \text{UND} \ \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{CAT} \begin{bmatrix} \text{ARG-ST} \ \langle \text{NP}\boxed{1}, \text{NP}\boxed{2} \rangle \end{bmatrix} \end{bmatrix}$$

Any well-formed word must belong to a single class from each of the four dimensions involved in the hierarchy above. Intuitively this means choosing one type in each dimension and combining their information structure via unification.

Consider the sentences below:

(1) Claudio played the sonata.

(2) This sonata plays well on an organ.

(3) Claudio plays the sonata well on her harpischord.

The sentences above illustrate different uses of the verb *play*, i.e. different choices along the dimensions of the hierarchy.

Below you can see the category of agentive verbs:

$$\begin{bmatrix} \textit{agnt-vb} \\ \text{CONT} \begin{bmatrix} \text{NUCLEUS} \begin{bmatrix} \textit{act-rel} \\ \text{ACT } \boxed{1} \end{bmatrix} \\ \text{DESIG-ARG } \boxed{1} \end{bmatrix} \end{bmatrix}$$

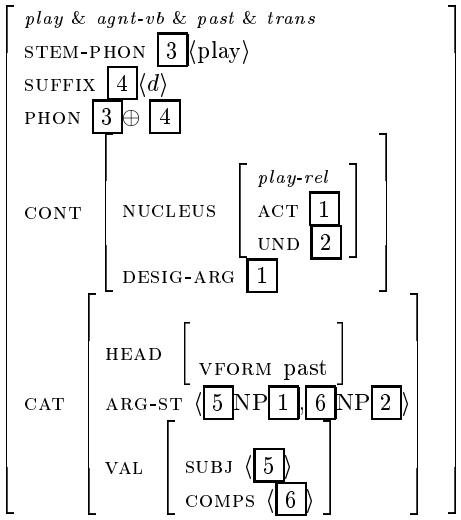
The result of combining the *agnt-vb* and *play* categories will be as follows:

$$\begin{bmatrix} \textit{play \& agnt-vb} \\ \text{STEM-PHON } \langle \text{play} \rangle \\ \text{CONT} \begin{bmatrix} \text{NUCLEUS} \begin{bmatrix} \textit{play-rel} \\ \text{ACT } \boxed{1} \\ \text{UND } \boxed{2} \end{bmatrix} \\ \text{DESIG-ARG } \boxed{1} \end{bmatrix} \\ \text{CAT} \begin{bmatrix} \text{ARG-ST } \langle \text{NP}\boxed{1}, \text{NP}\boxed{2} \rangle \end{bmatrix} \end{bmatrix}$$

The category of *past verbs* will look like this:

$$\begin{bmatrix} \textit{past} \\ \text{STEM-PHON } \boxed{3} \\ \text{SUFFIX } \boxed{4}\langle d \rangle \\ \text{PHON } \boxed{3} \oplus \boxed{4} \\ \text{CAT} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{VFORM past} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The intersection of the *play*, *agnt-vb*, *past* and *trans* categories will give the following result:



To infer the full type of a lexeme, the parser/interpreter must

- successfully choose a (maximally specific) class from each dimension;
- successfully combine (via unification) the information shared by all members of the class;
- disregard all combinations that fail.

Where a traditional hierarchical lexicon describes morphological relations through rules which apply to fully specified categories, the type-underspecified hierarchical lexicon describes them through a shared, underspecified type.

One way to think of this shared type, or "abstract" lexeme category, is that it represents what is common to all morphological and morphosyntactic categories of the particular word one finds in English sentences.

3 Exceptions

Problem:

Many processes are productive and regular. But they do not just apply to an intensionally defined subcategory, they also must exceptionally apply to some list of words.

Example:

(4) Je croyais qu'il avait de la classe.
I believe-PST that he have-PST of the class

'I thought he had some class'.

(5) Je lui croyais de la classe.
I to.he/she believe-PST of the class

'I thought he/she had some class'.

(6) *On prête qu'il a de bons sentiments.
People say.PR that he have.PR INDEF good feelings

'People say he has finer feelings'.

(7) On lui prête de bon sentiments.
People to.he/she say.PR INDEF good feelings

'People say that he/she has finer feelings'.

Solution:

We need to define an open-ended subcategory which accounts for the productivity of the process: any lexeme that satisfies its internal characterization can combine with it.

We also need to define a list of exceptions. The subcategory whose characterization includes a list of members accounts for the positive exceptions from the process. These lexemes are registered in the mental lexicon as members of the category.

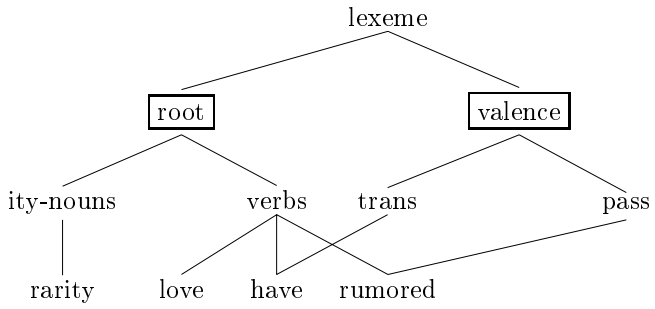
Problem:

The English *-ity* noun class gives an example of a pattern, where all words share a set of properties, but no new members can be added.

Solution:

The entire category must be both intensionally defined to capture its members' common properties and extensionally defined by enumerating the set of lexemes to which it applies.

Illustration:



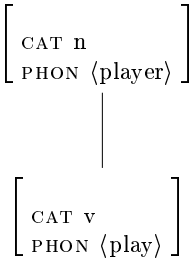
4 A Typed Constituent Structure-Based Morphology

Because category intersection is an additive process it cannot account for instances in which the information associated with the input of a putative rule is incompatible with that of the output.

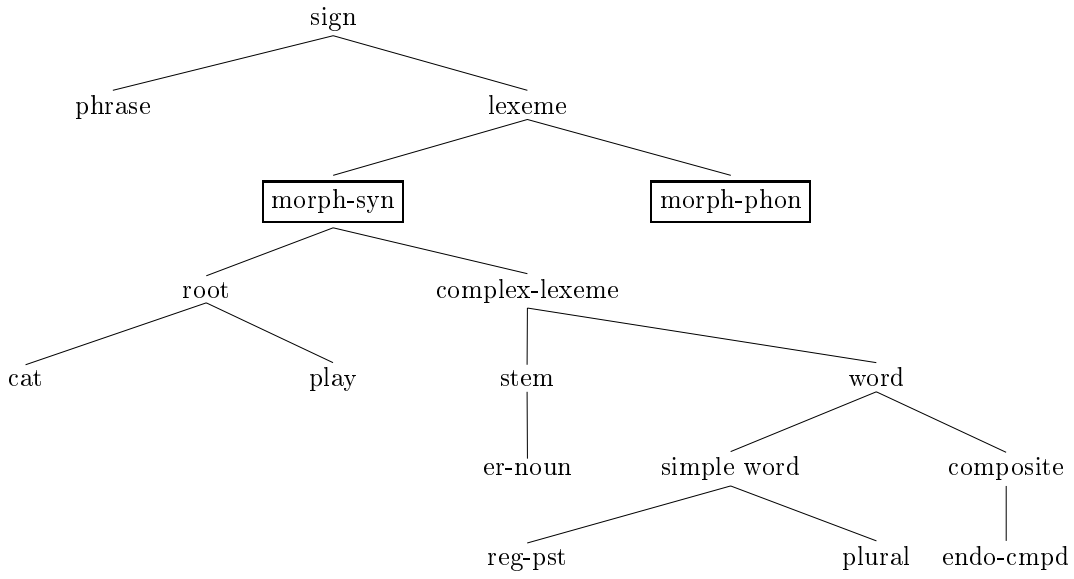
Therefore we need to assume morphological structure. The structure assumed in *Lexical Relations* is degenerate: a typical local tree configuration consists of a mother and a single daughter.

The affix is not a daughter, it is simply added to the phonology of the mother.

The approach to morphological structure is construction based, rather than morpheme bases. I.e. affixes are not terminals in the morphological trees and any node may carry phonological information.



4.1 The Morphosyntactic Side of Morphological Constructions



Lexemes are all linguistic signs smaller than or equal to word: roots, stems, and words. *Complex lexemes* have internal constituent structure. *Roots* do not. They are minimal grammatical units and do not have internal constituency. They are the atoms of morphology. *Words* are syntactic atoms selected by phrase structural schemata. *Stems* are defined negatively, as neither roots, nor words. *Simple words* are inflected words such as 'dog', 'dogs', 'played', etc. *Composite words* are both the ones preceded by particles (e.g. 'withstand') and compounds (e.g. 'mountain-climber').

Morphological structure is expressed through the feature μ-STRUC, defined on *complex-lexeme*.

$$\left[\begin{array}{l} \textit{complex-lexeme} \\ \mu\text{-STRUC} \left[\begin{array}{l} \textit{μ-struc} \\ \text{DGHTR lexeme} \end{array}\right] \end{array}\right]$$

The following AVM shows how the singular noun "cat" is derived from the root "cat", given the modifications above.

$$\left[\begin{array}{l} \textit{sg-noun} \\ \text{PHON} \left[\text{AFF} \left[\text{SUFF } \langle\text{e-string}\rangle\right]\right] \\ \text{CAT} \left[\text{HEAD} \left[\mu\text{-FEAT} \left[\text{SGR|NUM sg}\right]\right]\right] \\ \mu\text{-STRUC} \left[\text{DGHTR cat}\right] \end{array}\right]$$

The past form of "play" now looks as follows:

$$\left[\begin{array}{l} \textit{past} \\ \text{PHON} \left[\text{AFF} \left[\text{SUFF } \langle\text{d}\rangle\right]\right] \\ \text{CAT} \left[\text{HEAD} \left[\text{VFORM past}\right]\right] \\ \mu\text{-STRUC} \left[\text{DGHTR play}\right] \end{array}\right]$$

Note: affixes are not morphological daughters. They are only phonological objects and do not exist independently of the constituent structure construction itself.

Affixes are never the heads of morphological constructions.