

Surface Case – Abstract Case (Project notes, summary)

Mátyás Naszódi: Morphologic 1118 Budapest, Késmárki utca 8.
 naszodim@morphologic.hu

In the Hungarian language, there are 17-29 case suffixes depending on the linguistic model. These surface cases do not determine their semantic properties exactly. On the other hand, the nominal postpositions have the same role in the sentence syntax that case suffixes do. If you add the number of postpositions to that of case suffixes, the number of surface cases exceeds a hundred.

From the point of view of parsing and interpreting sentences, the surface forms of a case are rarely relevant. Rather, the functionality of a nominal phrase is important. Usually the noun phrases are categorized by their adverbial functionality in the phrase. That is what I call abstract case. Several surface cases stand for more than one abstract case.

This large amount of cases makes the syntax formalism complicated, and makes the computational interpretation hard.

In the paper I attempt to describe the abstract cases by their independent features. The features appear to be semantic ones. There are five classes of the features.

1. Type: Base, Time, Location, Mood, Quantity
2. Direction: Source, Target, Fix
3. Covering: Point, Surface, Interval, Inside, Beside, Through...
4. Preciseness: Precise, Uncertain, About...
5. Negation Positive, Negative

Some of the combination of these features have no surface interpretations. Each class has less than 10 possible values that make their usage manageable.

Examples: <i>-t</i> suffix(<i>accusative</i>)	[Base, Target, Point, Precise, Positive]
<i>-val</i> suffix(<i>by/with/together</i>)	[Base/Mood, Fix, Beside, any, Positive]
<i>nélkül</i> postposition(<i>without</i>)	[Base/Mood, Fix, Beside, any, Negative]

The categories are language dependent. Mainly the Covering class was subjectively determined. The classification was extended to all of the adverbial words and expressions, question words, etc.

For example: <i>valamikor</i> (<i>sometime</i>):	[Time, Fix, any, Uncertain, Positive]
<i>sehova</i> (<i>to nowhere</i>):	[Location, Fix, any, Uncertain,
Negative]	

<i>ott</i> (<i>there</i>):	[Location, Fix, any, Precise, Positive]
------------------------------	---

In some cases a surface case belongs to a single abstract case, like *kor* is, while the suffixes *-ig*, *-ban*... exactly determine their directions, covering preciseness features, but they can mark location, time, or mood as well. The ambiguous feature depends on the semantics of the word they follow. For example the words expressing time are recognizable. Most of them are measure units, while others are recognized by the sequence of the derivative affixes.

This classification was written in the HUMORESKE formalism, and it can be applied in any feature (affix) grammar. It has appeared useful in application of parsing Hungarian sentences, and analyses of psychological reports.