Deep cases in the `4lang` concept lexicon

Márton Makrai

Hungarian Academy of Sciences, Institute for Computer Science and Control
e-mail: makrai@sztaki.mta.hu

`4lang` is a multilingual lexicon for general human language understanding containing formal representations of word meaning in the monosemic approach to lexical semantics, which means that items are language independent concepts covering different uses of the same word, uses in different sentence patters and even in different parts of speech with the same meaning representation.[1] Multilinguality and abstractness of items have the effect that a simple deep case (or thematic) frame captures uses with different arity (i.e. transitive and intransitive). Deep cases denote the nodes in the graph representing the meaning of a predicate where the representation of the argument (single word, entity or phrase) has to be inserted.

`4lang` makes no clear cut between complements and adjuncts. Basically an argument is represented by a deep case whenever its needed for building of the representation of the verb. As uses of the same verb with different arities are handled in the same item, deep cases are used consequently in different verb patterns, and all possible arguments are included in the representation. However, as verbs can be defined as special cases of other verbs (biting is cutting with teeth), arguements are inherited, so not every argument is listed directly in the definition of some verb. An other source of inplicite arguments are constructions providing verbs with outer arguments e.g. *pain a picture **for somebody***.

Most frequent verbal deep cases are agents (denoted by `AGT`), patients (`PAT`), and datives (`DAT`). Patient plays the role of the neutral case it seems to play in many systems (Somers 1987)[2]. Following the unaccusative hypothesis, arguments of intransitive verbs split to agents and patients. The label "dative" is taken from Fillmore (1968), but our understanding is narrower as we mainly restrict dative to recipients in ditransitives (verbs of communication (e.g. *tell*) and transfer (e.g. *give*)). There are three locative cases in `4lang` (`TO`, `FROM`, and `AT`), the latter being used for the abstract goal of relational nouns such as *occasion* and *need* as well. A greater group of relational nouns require the possessive (`POSS`) such as *absence* and *duty*. Quirky cases can be marked in a language dependent module.

Deep cases in `4lang` are not restricted to verbs. Some grammatical features such as plural contribute to meaning, so morphemes expressing them have deep cases. Representations of productive derivational suffixes and adpositions also refer to the conceptual element they attach to with deep cases (`REL`).

---

[1] The lexicon, automatically collected word forms in 50 languages, a verctor space language model (embedding) computed from `4lang`, and articles can be found at `http://hlt.sztaki.hu/resources/4lang/`

[2] References can be found in the full version of the article that is in Hungarian.