

Beyond the synset: Swesaurus – a fuzzy Swedish wordnet

Lars Borin and Markus Forsberg
Språkbanken, University of Gothenburg, Sweden
lars.borin@svenska.gu.se, markus.forsberg@gu.se

Swesaurus is a free Swedish wordnet currently under construction in our research unit. Swesaurus is made up in part by a combination of a number of pre-existing freely available lexical resources. Two central resources for the purposes of this presentation are SALDO and Synlex.

SALDO is a full-scale Swedish lexical-semantic resource with non-classical, associative relations among word and multiword senses. The senses in SALDO are identified by carefully designed persistent formal identifiers, and for this reason, SALDO has become the “pivot” resource of all our computational lexicon activities, including Swesaurus. We will say something about the relationship of SALDO’s associative relations to classical lexical-semantic relations and how we intend to incorporate the SALDO relations in Swesaurus.

Synlex is a graded Swedish synonym list created by asking members of the public – users of an online Swedish-English dictionary – to judge the degree of synonymy of a random, automatically generated synonym pair candidate.

We will describe our experiments turning the graded synonymy relations of Synlex into *fuzzy synsets* in Swesaurus. The introduction of fuzziness into a wordnet raises many intricate methodological and theoretical questions, e.g., if w_1 is a graded synonym of w_2 , and w_1 is a hyponym of w_3 , what is the relation between w_2 and w_3 . Similarly, if w_a is a synonym of degree 0.75 of w_b and w_b is a synonym of degree 0.9 of w_c , what – if any – is the degree of synonymy between w_a and w_c ?

If synonymy is seen as an all-or-none transitive and symmetric relation, the construction of synsets from synonym pairs is arguably straightforward: We can simply compute the transitive closure of the synonymy relation.

When graded synonymy enters the picture, which method to use for collecting synonym pairs into synsets becomes much less obvious, and especially how to assign a degree of synonymy to “derived” pairs, i.e., pairs not in the original list. We have experimented with two different approaches for turning Synlex synonym pairs into fuzzy synsets in Swesaurus, transitive closure and clique formation. We will present the outcomes of these experiments and discuss the merits and disadvantages of the two methods.