A Second-order Co-Occurrence Model for Selectional Preferences

Sabine Schulte im Walde

Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

schulte@ims.uni-stuttgart.de

Predicates impose selectional restrictions on the realisation of their complements, as first illustrated by Chomsky (1957) through his famous example *Colorless green ideas sleep furiously*. Though the sentence is syntactically well-formed, it is not semantically meaningful, unless interpreted metaphorically. Approaches in computational linguistics that model selectional restrictions commonly refer to them as *selectional preferences*. This illustrates the fact that selectional restrictions refer to a certain degree of acceptance, rather than a binary decision, and that the degree of acceptance is often modelled by probabilities. Selectional preferences are of great interest to computational linguists, both from a lexicographic and from an applied perspective.

Our approach to selectional preference induction is both intuitive and cheap. The underlying idea is that selectional preferences of a predicate's complement are defined by the properties of the complement realisations. For example, a typical direct object of the verb *drink* is usually fluid, might be hot or cold, can be bought, might be bottled, etc. We thus suggest a second-order co-occurrence model for selectional preferences: a predicate's restrictions to the semantic realisation of its complements are expressed through the properties of the complements. The crucial ingredients of our model are corpus-based joint frequencies $freq(p,r_1,n)$ of predicates p and nouns n with respect to some functional relationship r_1, and corpus-based joint frequencies $freq(n,r_2,prop)$ of nouns n and noun properties $prop$ with respect to some functional relationship r_2. In a first step, the selectional preference of r_1 is represented by the joint noun-property corpus frequencies across the nominal arguments, cf. Equation (1). I.e., the feature vector of the predicate is a union of the properties of the nouns. For example, if the predicate is the verb *drink*, the verb-noun relation is a direct object, and the property is adjectives that modify nouns: then, the verb's selectional restrictions are defined by an adjective feature vector, where the set of adjectives is the union of the adjectives modifying the nouns subcategorised by the verbs. The feature values *score(drink,dir-obj,adj)* rely on the frequencies of all nouns that appeared as direct objects of *drink*, *freq(drink,dir-obj,n)*, and on the frequencies of the adjectives those nouns appeared with (not necessarily in the same context with the verb), *freq(n,n-mod,adj)*. For example, if *coffee* appeared 50 times as direct object of *drink*, and *tea* appeared 5 times, and if *coffee* was modified by the adjective *hot* 100 times and by *fluid* 30 times, and if *tea* was modified by *hot* 60 times and by *fluid* 15 times, then *score(drink,dir-obj,hot) = 50*100+5*60 = 5,300*, and *score(drink,dir-obj,fluid) = 50*30+5*15 = 1,575*. The scoring provided in Equation (3) is the most simplest, using raw corpus frequencies. Alternative versions incorporate normalisations of the raw frequencies.

(1) $score(p,r_1,prop) = \Sigma_{n\epsilon(p,r1)} freq(p,r_1,n) * freq(n,r_2,prop)$

The resulting selectional preference descriptions are predicate vectors over complement properties. In order to determine the selectional preference for a specific (seen or unseen) noun, we calculate the vector-based similarity of the predicate's preference vector and the specific noun's vector. The measures to calculate the similarities and thus the natural fit of a specific noun to a selectional preference description can be varied. We experimented with four standard measures, the cosine of the vector's angle (a standard measure in linear algebra), the skew divergence, an information-theoretic measure and variant of the Kullback-Leibler divergence (Lee, 2001), Kendall's *tau*, a measure for rank correlation (Hatzivassiloglou and McKeown, 1993; Lapata, 2006), and the *jaccard index*, a binary distance measure (Manning and Schütze, 1999). Our method is similar to an approach by Erk (2007) who also used complements' corpus-based properties to describe selectional preferences. We addressed the task from a different direction, though, and the result is a simplified version of her approach. Furthermore, our goal is different from hers in that we are interested in the contributions of the various properties, in addition to determining the natural fit of nouns to selectional preferences.

The joint frequencies are based on approx. 650 million words from the German web corpus *DeWaC* (Baroni and Kilgarriff, 2006), after the corpus was preprocessed by the *Tree Tagger* (Schmid, 1994) and by a dependency parser (Schiehlen, 2003). We experimented with verbs and their subjects, direct objects and pp objects, and used modifying adjectives, subcategorising verbs and subcategorising prepositions as noun properties. The evaluation was carried out in accordance with Brockmann and Lapata (2003) who collected human judgements on selectional preference realisations for 30 subjects, 30 direct objects, and 30 pp objects (10 verbs in each three cases), and correlated the judgements with system scores by linear regression. The baseline of the experiments was determined by correlating the joint corpus-based predicate-noun frequencies of the subjects, direct objects and pp-objects with the human judgements (also by linear regression). We used two baselines, one based on the raw frequencies, and the other based on the frequencies transformed by *log10*. The upper bound of the approaches is the inter-subject agreement (isa) on the selectional preference judgements, as calculated by Brockmann and Lapata. The table below presents an extract of the correlation results, with the best results per relation in bold font.

The talk will describe our approach and related work in detail, present examples of the second-order co-occurrence vectors, analyse the contribution of the various noun properties, assess the results with respect to the model parameters, and discuss alternative parameter settings.

	SUBJECT			***DIRECT OBJECT***			***PP-OBJECT***		
	cos(f)	*cos(log(f))*	*tau*	*cos(f)*	*cos(log(f))*	*tau*	*cos(f)*	*cos(log(f))*	*tau*
adj (a)	.427	.447	.343	.357	.200	.205	.053	.185	.207
verb (v)	.420	.461	.320	.183	.142	.250	.106	.226	.303
prep (p)	.440	.344	.337	**.669**	.220	.289	.284	.403	.345
v+a	.419	.468	.358	.338	.205	.299	.150	.288	.334
v+a+p	.437	**.504**	.423	.652	.242	.312	.342	**.445**	.403
baseline	.298/.652			.315/.559			.319/.565		
isa	.790			.810			.820		

References

Marco Baroni and Adam Kilgarriff. Large Linguistically-processed Web Corpora for Multiple Languages. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics,* Trento, Italy, 2006.

Carsten Brockmann and Mirella Lapata. Evaluating and Combining Approaches to Selectional Preference Acquisition. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics,* Budapest, Hungary, 2003.

Noam Chomsky. *Syntactic Structures.* Mouton, The Hague, 1957.

Katrin Erk. A Simple, Similarity-based Model for Selectional Preferences. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics,* Prague, Czech Republic, 2007.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. Towards the Automatic Identification of Adjectival Scales: Clustering Adjectives According to Meaning. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics,* Columbus, OH, 1993.

Mirella Lapata. Automatic Evaluation of Information Ordering: Kendall's Tau. *Computational Linguistics,* 32(4), 2006.

Lillian Lee. On the Effectiveness of the Skew Divergence for Statistical Language Analysis. *Artificial Intelligence and Statistics,* 2001.

Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing.* MIT Press, Cambridge, MA, 1999.

Michael Schiehlen. A Cascaded Finite-State Parser for German. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics,* Budapest, Hungary, 2003.

Helmut Schmid. Probabilistic Part-of-Speech Tagging using Decision Trees. In *Proceedings of the 1st Conference on New Methods in Language Processing,* 1994.