German Perception Verbs:
Automatic Classification of Prototypical and Multiple Non-Literal Meanings

Benjamin **David**, Sylvia **Springorum** & Sabine **Schulte im Walde**

IMS, **University of Stuttgart**, 2014

Abstract

This project presents a **token-based automatic classification of German perception verbs** into literal vs. multiple non-literal senses. Based on a corpus-based dataset of German perception verbs and their systematic meaning shifts (following Ibarretxe-Antuñano, 1999), we identify one verb of each of the four perception classes optical, acoustic, olfactory and haptic, and used Decision Trees relying on syntactic and semantic corpus-based features to classify the verb uses into 3-4 senses each. **Our classifier reaches accuracies between 45.5% and 69.4%**, in comparison to baselines between 27.5% and 39.0%. In three out of four cases analysed, our classifier's accuracy is significantly higher than the according baseline.

Gathered Verbs

gustatory	acoustic	haptic	olfactory	optic
1 verb	16 verbs	13 verbs	11 verbs	61 verbs

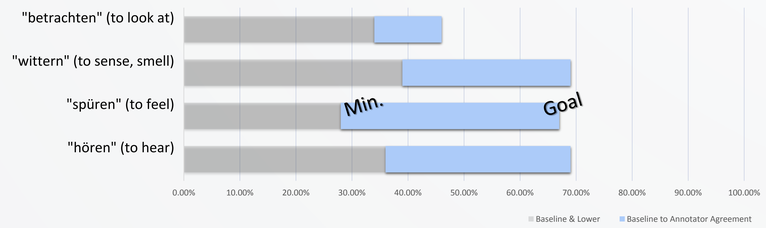
Selected Verbs & Meanings

Filtering by the following criteria:
Verbs of both **active and passive** perception
with significant **corpus frequency**
which convey **several non-literal meanings**

acoustic	haptic	olfactory	optic
hören *to hear*	**spüren** *to feel*	**wittern** *to sense (smell)*	**betrachten** *to look at*
to hear (prototypical)	to feel (prototypical)	to smell (prototypical)	to look at (prototypical)
to (dis)like	to realize	to advance towards	to define
to obey	to feel (emotions)	to predict	to analyse (objective)
to be informed	to suspect	-	to judge (subjective)

Sample Data

3 linguists independently annotated **over 750 sample sentences** for the selected perception verbs, with **~70% agreement** rate. This agreement rate is the goal in terms of classifier accuracy. The sample sentences are chosen at random from the **SdeWaC-Corpus[1].**



Perception Verbs

By starting an **exhaustive search** through thesauri and online thesauri, we create a database of German perception verbs. The sole gustatory verb found ("schmecken", to taste) conveys only one very rare non-prototypical meaning and is not analysed.

Classification

The classification is done utilizing **WEKA**. Most features used were extracted by the SubCatExtractor[2]. It provides **subcategorization frames** for sentences in pre-parsed corpora. Further, **hypernymy data** extracted from the SdeWaC[1] corpus with help of GermaNet data as well as **sentiment data** from GermanPolarityClues[3] is consulted.

Features

Syntactic Features		Verb-Modifying Features		Semantic Features	
Sentence Rule	These rules state which one out of 12 methods was used by the SubCat Extractor to extract the subcategorisation frame.	**Verb Form**	The tag given for the verb by the TreeTagger with STTS Tagset.	**Subject Hypernym**	Subject's hypernym extracted from GermaNet.
Sentence Form	Describes the dependency relations of the verb complex according to the TIGER corpus annotation format.	**Adverb**	Presence of an adverb represented by a Boolean value.	**Accusative Hypernym**	Accusative object's hypernym extracted from GermaNet.
Adjective, Accusative Object & Negation	Presence of the accusative object represented (Boolean value).	**Adverbial & Prepositional Object**	For each preposition introducing a prepositional or adverbial object one Boolean feature is introduced.	**Adverb & Adjective Sentiment**	The sentiment data extracted from GermanPolarityClues[3]. (positive/neutral/negative/none)
Subjunction & Modal Verb	Either 'None' or the lemma of the subjunction/modal verb if found.				

Observations

Our classifier significantly outperforms the respective baselines, with the exception of the olfactory perception verb classification. Despite the hypernym data being insufficient in many cases, the semantic sub-vector returns the best prediction accuracies, with an average of **47% accuracy**. The choice of meanings is intentionally rough: Some are dropped or merged to allow easier analysis of the results. The biggest deviations can be accounted to ambiguity between only 2 of the meanings and the prototypical meaning acting as a residual class.

Results

Nearest Centroid
Classification with WEKA
Baseline / **Accuracy** / Goal

Perception Type (Verb)		**Syntactic Features Acc.**	**Verb-Modifying Features Acc.**	**Semantic Features Acc.**	Baseline / **Combining all Features** / Goal
acoustic	(hören)	**46%**	**56%**	**53%**	36% / **57%** / 69%
haptic	(spüren)	**43%**	**41%**	**36%**	28% / **42%** / 67%
olfactory	(wittern)	**44%**	**40%**	**38%**	39% / **32%** / 69%
optic	(betrachten)	**41%**	**53%**	**56%**	34% / **46%** / 46%

[1] Faaß and Eckart (2013) [2] Scheible (2012) [3] Waltinger (2013)