Weakly supervised learning of allomorphy

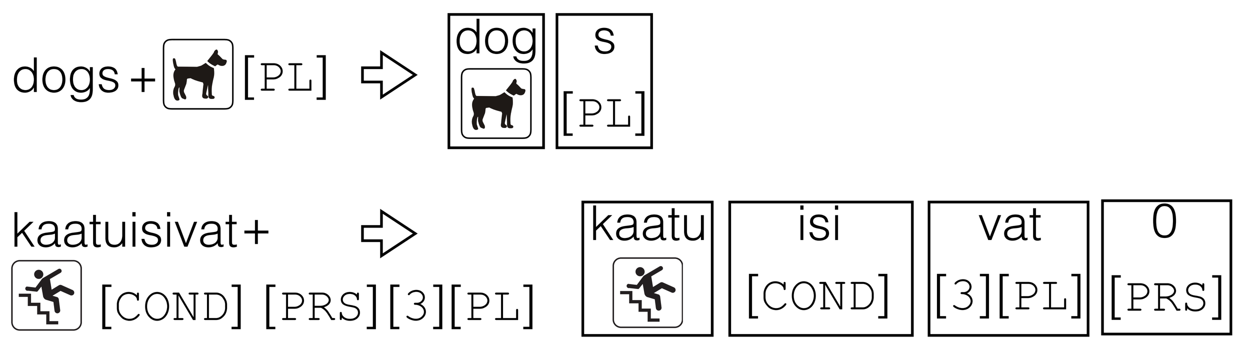
University of Colorado **Boulder**

Miikka Silfverberg and Mans Hulden
University of Colorado Boulder, Department of Linguistics
firstname.lastname@colorado.edu

Overview

We present weakly supervised learning of allomorphy – **a new learning problem in the field of computational morphology** and evaluate several methods for solving it. We present experiments on English, Finnish, Spanish and Turkish.

Our training data consists of plain unsegmented word forms (for example **kaatuisivat** "they would fall over") and morphological feature sets (for example V, COND, PRS, 3, PL). We call this a **weak labeling**.



The result is a morphologically segmented corpus where each morphological segment is associated with at least one morphological feature.

The crucial constraint provided by the weak labeling is that only a subset of all labels is present in a each word form.

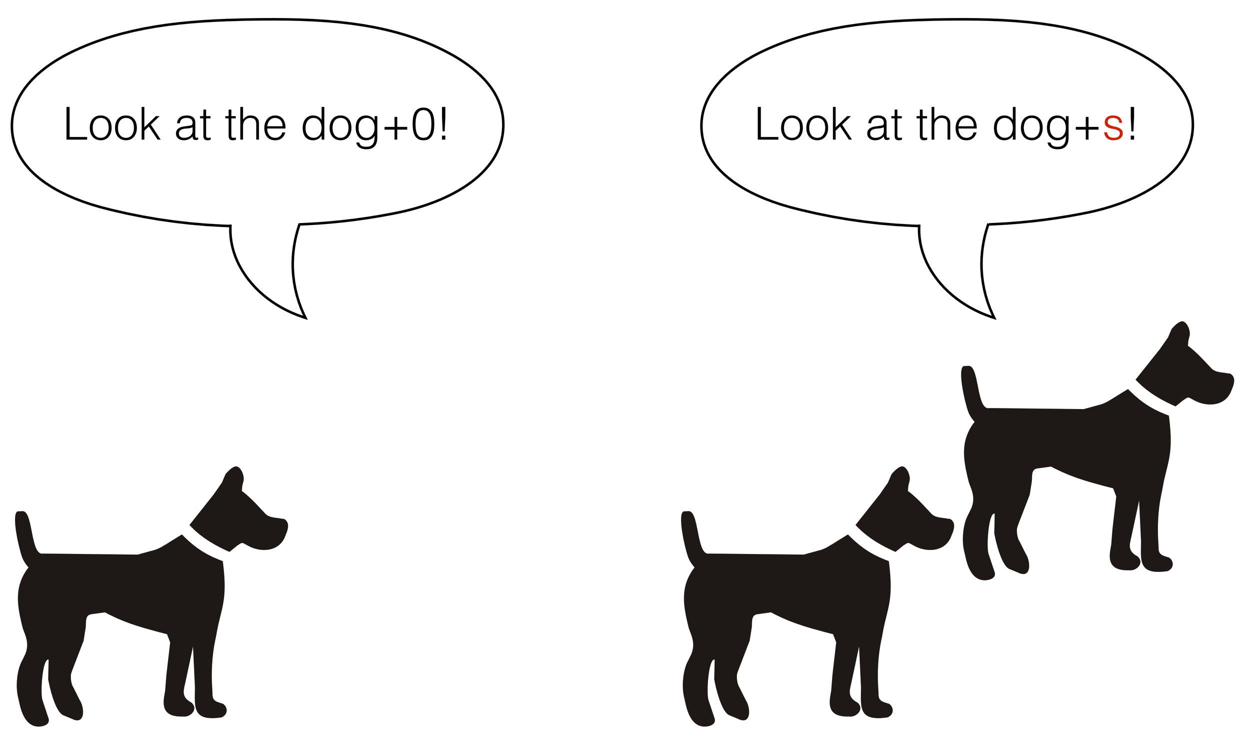
$$\Theta(\text{“isi”}, \texttt{COND})$$

We formalize a generic objective function Θ that scores the goodness of segmentations and labeling globally in a corpus and treat the problem of joint segmentation and feature assignment as a search problem in the space of all possible segmentations and labelings.

Motivation

Most treebanks include some morphological annotation at the word level. Typically, the annotation is not aligned with the relevant substrings in the word forms themselves.

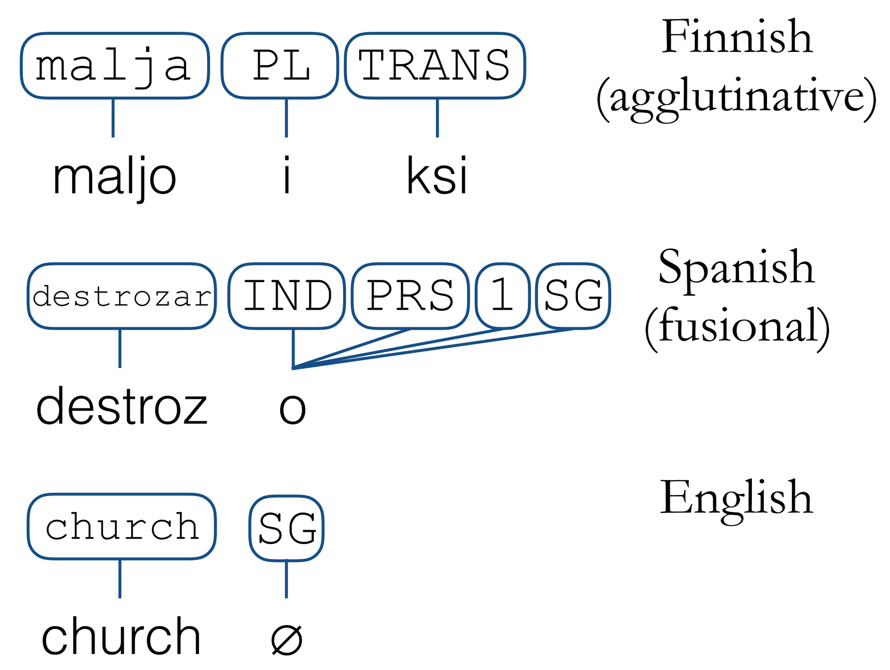
We want to **improve the usability of treebanks** by inducing annotating at the morpheme level without additional supervision.



Moreover, this type of input and the related inference problem is consistent with the assumptions of relevant inputs witnessed in **L1 acquisition**—a combination of stems and other affixes where the learner knows from the environment some semantic signal from the immediate discourse, e.g. plurality, tense, etc.

Scope

We explore assignments where each segment is aligned with at least one morphological feature. One segment may, however, be aligned with several features. This is required when a morpheme encodes for several morphological features.



A typologically interesting case not modeled in our approach is templatic morphology. The objective functions we develop may be adapted to this case, however, at the cost of enlarging the search space.

Objective Function

We accomplish segmentation and feature assignment by learning an objective function $\Theta : \Sigma^* \times Y \rightarrow \mathbb{R}$, which scores combinations of segments and morphological features. Here Σ^* is the set possible segments and Y is the set of morphological features. For each word form, we then seek the segmentation and label assignment which maximizes the total score over all segments and features.

We experiment with different objective functions:

1. We use the log of the symmetric conditional probability of substrings and morphological features as the score.

$$\mathrm{SCP}(x, y) = p(x|y)p(y|x) = \frac{p(x, y)^2}{p(x)p(y)}$$

2. We train a perceptron model to predict morphological features based on substrings and use its parameters as scores.
3. Substrings and morphological features define distributions over co-occurring morphological features. We use the negative Kullback-Leibler Divergence of the distributions as a scoring function.

$$\mathrm{KL}(p||q) = \sum_{y \in Y} p(y) \log \frac{p(y)}{q(y)}$$

4. We train a Rescorla-Wagner model to predict morphological features based on substrings and use its parameters as scores.
5. As baseline, we train a Morfessor model on the plain word forms and use a very simple model for assigning morphological features.

Results

(a)

	Eng	Fin	Swe	Tur
Kullback-Leibler Divergence				
R	93.91	82.74	73.81	81.25
P	87.15	80.36	65.89	76.60
F_1	90.41	81.54	69.63	78.86
Perceptron				
R	98.81	80.67	86.15	86.68
P	95.06	88.05	76.54	90.54
F_1	96.90	84.20	81.06	**88.57**
Rescorla-Wagner				
R	98.93	83.74	82.58	82.88
P	97.87	86.81	82.22	91.96
F_1	**98.40**	85.23	82.40	87.18
Symmetric Conditional Probability				
R	89.02	81.56	77.38	70.65
P	95.15	91.58	94.33	90.28
F_1	91.99	**86.28**	**85.02**	79.27
Morfessor baseline				
R	80.79	67.36	76.19	77.26
P	61.10	65.67	72.35	93.22
F_1	69.58	66.50	74.22	84.50

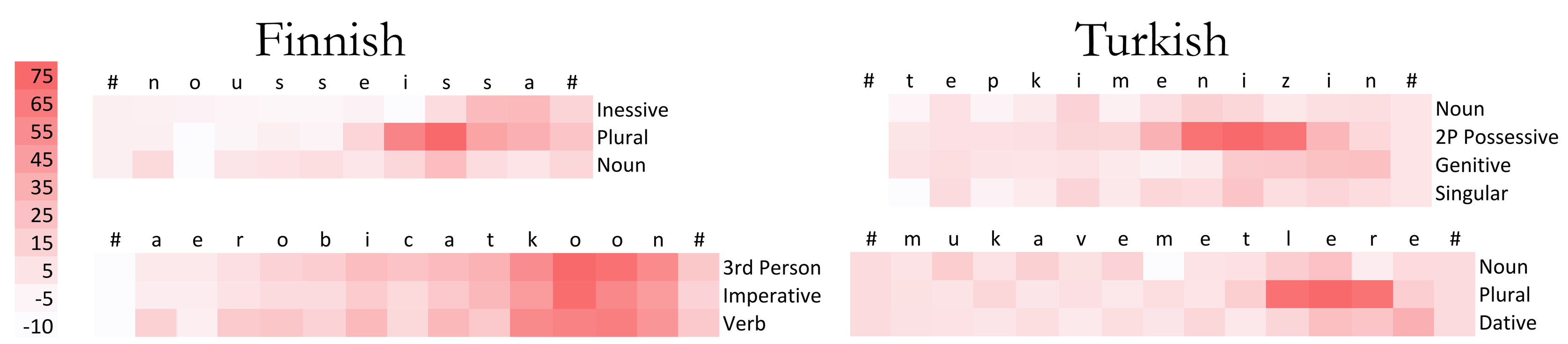
(b)

	Eng	Fin	Swe	Tur
Kullback-Leibler Divergence				
R	69.52	45.66	15.71	44.28
P	62.02	43.82	13.35	40.97
F_1	65.56	44.72	14.43	42.56
Perceptron				
R	90.15	47.62	44.55	61.32
P	84.94	54.05	37.57	65.13
F_1	87.47	50.63	40.76	**63.16**
Rescorla-Wagner				
R	94.98	50.84	43.91	56.84
P	93.42	53.53	43.63	65.76
F_1	**94.19**	52.16	**43.77**	60.97
Symmetric Conditional Probability				
R	64.31	50.14	37.82	27.99
P	71.49	59.37	51.53	39.89
F_1	67.71	**54.37**	43.62	32.89
Morfessor baseline				
R	21.19	9.94	21.79	39.93
P	14.14	9.59	20.27	52.20
F_1	16.96	9.77	21.00	45.24

(c)

	Eng	Fin	Swe	Tur
Kullback-Leibler Divergence				
R	74.18	39.70	8.88	31.36
P	74.11	37.07	7.64	27.09
F_1	74.15	38.34	8.22	29.07
Perceptron				
R	90.06	34.77	30.30	59.91
P	90.06	33.59	27.30	54.70
F_1	**90.06**	34.17	28.72	**57.19**
Rescorla-Wagner				
R	43.96	29.98	24.31	43.14
P	43.96	29.98	24.96	43.58
F_1	43.96	29.98	24.63	43.36
Symmetric Conditional Probability				
R	66.37	56.45	62.88	52.81
P	66.37	53.90	64.07	53.35
F_1	66.37	**55.15**	**63.47**	53.08
Morfessor baseline				
R	1.77	9.66	20.11	8.74
P	1.76	10.97	25.09	11.21
F_1	1.77	10.27	22.32	9.82

Results for (a) morpheme boundaries; (b) unlabeled morphemes; (c) labeled morphemes.



Activations of the Rescorla-Wagner model for morphological features and the characters in Finnish and Turkish word forms.

Conclusions

- We have presented a new learning problem for natural language processing, namely weakly supervised learning of allomorphy.
- The problem is important from a practical point of view because there are many morphologically annotated corpora where the annotation is not extended to the morpheme level.
- It is also relevant from a theoretical point of view because it is related to L1 morphology learning.
- We explored four different learning methods and compared these to a baseline consisting of unsupervised morphological segmentation augmented by a straightforward labeling mechanism.
- Our results show that weak supervision delivers sizable improvements when evaluated with regard to F_1-score on labeled and unlabeled segmentation. Rescorla-Wagner delivers the best results.
- Evaluation for the task is challenging because even linguists do not always agree upon the best way to segment words into morphemes.