Do POS Tags Help to Learn Better Morphological Segmentations?

Kairit Sirts and Mark Johnson

Macquarie University

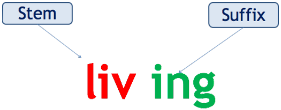
ALTA 2015 workshop

09.12.2015

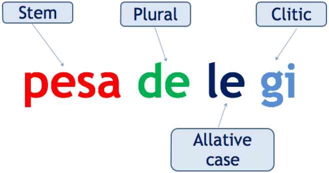
Morphological segmentation

- The task of splitting words into morphemes
- Morphemes are the smallest meaning-bearing units in language
- The current context is unsupervised segmentation

Simple English example



More complex Estonian example



Segmentation and POS tags

POS describe the syntactic function of words

DET	NOUN	VERB	ADJ	PUNCT
The	mouse	is	blue	.

POS and segmentation are related

- **sing_ing** → **VERB**
- **walk_ed** → **VERB**
- **home_less** → **ADJ**
- **walk_s** → **NOUN** or **VERB**

- **boring ADJ** → **boring**
- **singing VERB** → **sing_ing**
- **speed NOUN** → **speed**
- **walked VERB** → **walk_ed**

Previous work

POS dependence on segmentation

Several previous works have demonstrated the utility of using morphological features for unsupervised POS induction (Berg-Kirkpatrick et al., 2010; Lee et al., 2010; Christodoulopoulos et al., 2011).

Segmentation models utilising POS

- Clusters learned during preprocessing (Freitag, 2005; Can and Mandahar, 2009);
- Learn pseudo-syntactic clusters together with segmentations (Goldwater et al., 2006; Lee et al., 2011);
- Joint models of POS induction and morphological segmentation (Can, 2011; Sirts and Alumäe, 2012; Frank et al., 2013).
- **The benefits are mostly not clear.**

The goal of this work

Model the segmentation dependence on POS tags in order to learn:

- Whether the POS tags help to improve segmentations **as expected** (but not clearly demonstrated) in previous works;
- How much do the tags improve the segmentations (if at all)?
- Whether the segmentations are improved when using gold standard tags or do the automatically learned tags help as well.

The setting

- Use the Adaptor Grammars framework (Johnson et al., 2007) that have been previously demonstrated to perform state-of-the art morphological segmentation (Sirts and Goldwater, 2013);
- Adaptor Grammars enable easily to experiment with different morphological grammars;
- Use the same grammars with and without POS tags.

Adaptor Grammars

- Framework for learning non-parametric Bayesian models for parse trees over sequences of strings.
- Consists of a PCFG and a PYP adaptor function:
  - PCFG defines all possible parse tree structures;
  - PYP adaptor changes the probability of the parse trees such that frequently occurring subtrees are more probable.

A simple morphological grammar — MorphSeq grammar

$$\text{Word} \rightarrow \text{Morph}^{+}$$
$$\underline{\text{Morph}} \rightarrow \text{Char}^{+}$$

POS-dependent grammars

- Inspired by the grammars used to learn topic models (Johnson, 2010).
- Defines a separate set of rules for each POS tag.

POS-dependent MorphSeq grammar

$$
\begin{aligned}
\text{Word} &\rightarrow \text{Noun}\ \text{Morph}^{+}_{\text{Noun}} \\
\text{Word} &\rightarrow \text{Verb}\ \text{Morph}^{+}_{\text{Verb}} \\
\text{Word} &\rightarrow \text{Adj}\ \text{Morph}^{+}_{\text{Adj}} \\
\underline{\text{Morph}_{\text{Noun}}} &\rightarrow \text{Morph} \\
\underline{\text{Morph}_{\text{Verb}}} &\rightarrow \text{Morph} \\
\underline{\text{Morph}_{\text{Adj}}} &\rightarrow \text{Morph}
\end{aligned}
\qquad\qquad
\begin{aligned}
\text{Noun} &\rightarrow \text{N\_} \\
\text{Verb} &\rightarrow \text{V\_} \\
\text{Adj} &\rightarrow \text{A\_} \\
\underline{\text{Morph}} &\rightarrow \text{Char}^{+}
\end{aligned}
$$

Experimental grammars

- Tag-dependent **Morph** and **Colloc** rules
- General **Morph** and **SubMorph** rules

SubMorph grammar

$$\begin{aligned} \text{Word} &\rightarrow \text{Morph}^{+} \\ \underline{\text{Morph}} &\rightarrow \text{SubMorph}^{+} \\ \underline{\text{SubMorph}} &\rightarrow \text{Char}^{+} \end{aligned}$$

CollocMorph grammar

$$\begin{aligned} \text{Word} &\rightarrow \text{Colloc}^{+} \\ \underline{\text{Colloc}} &\rightarrow \text{Morph}^{+} \\ \underline{\text{Morph}} &\rightarrow \text{SubMorph}^{+} \\ \underline{\text{SubMorph}} &\rightarrow \text{Char}^{+} \end{aligned}$$

Experimental scenarios

1. Baseline without tags;
2. Oracle setting using gold standard POS tags;
3. Using tags learned with an unsupervised model (Sirts and Alumäe, 2012);
4. POS-dependent segmentation baseline using random tags.

Data

- English and Estonian nouns, verbs and adjectives;
- Word types and gold-standard tags from Multext-East corpus (G. Orwell "1984");
- English gold-standard segmentations from Celex;
- Estonian gold-standard segmentations from Estonian morphologically disambiguated corpus;
- Evaluated using segment boundary F1-score.

	English	**Estonian**
MTE types	8438	15132
Eval types	7659	15132

Results on English

	No POS	Gold	Learned	Rand
MorphSeq	51.4	54.3	**55.7**	52.5
SubMorph	63.3	**69.6**	68.1	64.3
CollocMorph	56.8	**71.0**	68.0	66.6

Results on Estonian

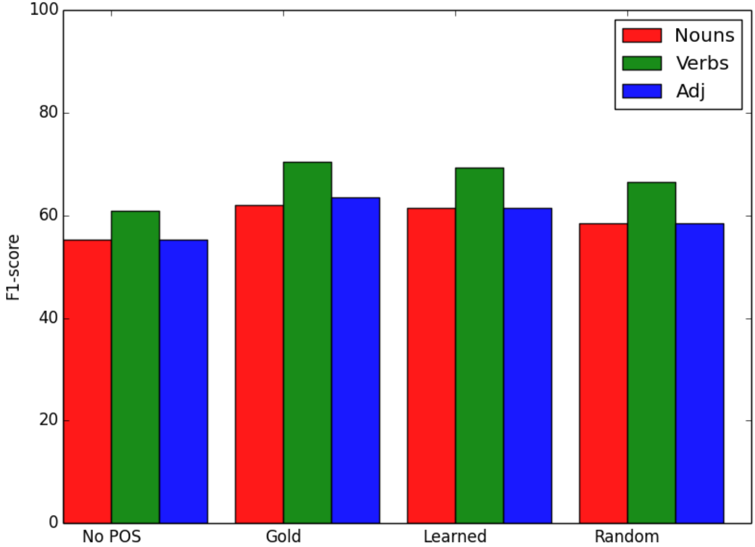
	No POS	Gold	Learned	Rand
MorphSeq	48.1	**53.2**	52.5	49.1
SubMorph	**66.5**	**66.5**	64.3	65.5
CollocMorph	65.4	**68.5**	66.5	68.4

Evaluate the results for each POS tag separately

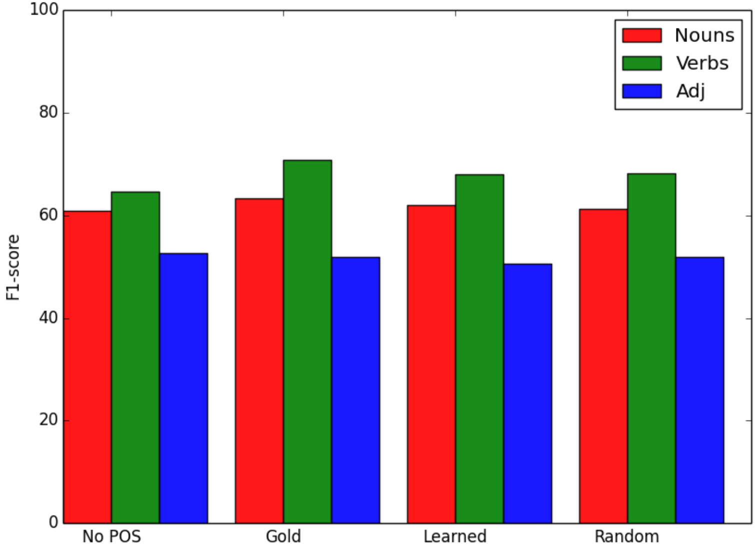
- The differences between scenarios are in several cases small.
- Evaluate the segmentation results for each POS category words separately.

	English	**Estonian**
Nouns	3831	8162
Verbs	2691	4004
Adjectives	1629	3111

Results for different POS tags in English



Results for different POS tags in Estonian



What did we learn?

- Grammars without tags give the lowest performance **as expected**.
- Gold-standard tags give the largest improvement **as expected**.
- The POS tags make more difference in English than in Estonian (**which is not what we expected**).
- In English, the induced tags perform better than random tags, but random tags are better than no tags.
- In Estonian there doesn't seem to be much difference between induced and random tags.
- The results of words with different POS are different, the F-score of verbs is much higher than nouns and adjectives.
- The absolute differences between different scenarios are not great.

Future work

- In morphologically complex languages experiment with more fine-grained morpho-syntactic tags;
- Develop even more complex grammars, such that precision and recall would be better balanced;
- Study the usefulness of POS tags in semi-supervised segmentation setting;
- Experiment with tags learned with a supervised tagger.

Conclusions

- Experiments to assess whether and how much do POS tags help to learn better morphological segmentations.
- Used Adaptor Grammars framework to define grammars of different complexity and utilizing different POS tags (no tags, gold tags, learned tags and random tags).
- Results in English showed that using tags helps to improve segmentations results: gold tags help the most but induced tags help as well.
- In Estonian the differences between different settings are small and thus the results are not convincing, the reasons of which should be explored in future work.