Do POS Tags Help to Learn Better Morphological Segmentations?

Kairit Sirts
Department of Computing
Macquarie University
Australia
kairit.sirts@gmail.com

Mark Johnson
Department of Computing
Macquarie University
Australia
mark.johnson@mq.edu.au

Abstract

The utility of using morphological features in part-of-speech (POS) tagging is well established in the literature. However, the usefulness of exploiting information about POS tags for morphological segmentation is less clear. In this paper we study the POS-dependent morphological segmentation in the Adaptor Grammars framework. We experiment with three different scenarios: without POS tags, with gold-standard tags and with automatically induced tags, and show that the segmentation F1-score improves when the tags are used. We show that the gold-standard tags lead to the biggest improvement as expected. However, using automatically induced tags also brings some improvement over the tag-independent baseline.

1 Introduction

Linguistially, part-of-speech (POS) tagging and morphology are closely related and this relation has been heavily exploited in both supervised and unsupervised POS tagging. For instance, the supervised Stanford tagger (Toutanova et al., 2003) as well as some unsupervised POS taggers (Berg-Kirkpatrick et al., 2010; Lee et al., 2010) use character prefix and/or suffix features, while the model by Christodoulopoulos et al. (2011) makes use of suffixes learned with an unsupervised morphological segmentation model.

There have been some attempts to exploit the relation in the opposite direction to learn the segmentations dependent on POS tags. For instance, the segmentation procedures described by Freitag (2005) and Can and Manandhar (2009) find the syntactic clusters of words and then perform morphology learning using those clusters. Few works have included a small number of syntactic classes directly into the segmentation model (Goldwater et al., 2006; Lee et al., 2011). However, Goldwater et al. (2006) only trains the model on verbs, which means that the classes model different verb paradigms rather than POS tags. Secondly, the model is never evaluated in a single class configuration and thus it is not known whether incorporating those classes gives any actual improvement. The results of Lee et al. (2011) show small improvements when the POS-word component (a bigram HMM) is incorporated into the model. However, the number of syntactic categories they learn is only 5, which is smaller than the number of main POS categories in most annotated corpora. Moreover, the main gain in the segmentation F-score is obtained by modeling the agreements between adjacent words, rather than exploiting the relation to syntactic classes.

Another line of previous work has attempted to model the POS tags and morphological segmentations jointly in an unsupervised model (Can, 2011; Sirts and Alumäe, 2012; Frank et al., 2013). However, the results presented in those papers fail to demonstrate clearly the utility of using the tag information in segmentation learning over the scenario where the tags are missing.

The goal of this paper is to explore the relation between POS tags and morphological segmentations and in particular, to study if and how much the POS tags help to learn better segmentations. We start with experiments learning segmentations without POS tags as has been standard in previous literature (Goldsmith, 2001; Creutz and Lagus, 2007; Sirts and Goldwater, 2013) and then add the POS information. We first add the information about gold-standard tags, which provides a kind of upper bound of how much the segmentation accuracy can gain from POS information. Secondly, we also experiment with automatically induced tags. We expect to see that gold-standard POS tags improve the segmentation accuracy and that induced tags

Kairit Sirts and Mark Johnson. 2015. Do POS Tags Help to Learn Better Morphological Segmentations? . In *Proceedings of Australasian Language Technology Association Workshop*, pages 91−100.

are also helpful. The results of these experiments can be informative to whether directing effort into developing joint unsupervised models for POS tagging and segmentation is justified, or whether the efforts of exploiting synergies in morphology learning should be focused elsewhere.

We define the segmentation model in the Adaptor Grammars framework (Johnson et al., 2007) that has been previously successfully applied to learning morphological segmentations (Johnson, 2008; Sirts and Goldwater, 2013). In fact, we will use some of the grammars defined by Sirts and Goldwater (2013) but enrich the grammar rules with information about POS tags. Our POS-dependent grammars are inspired by the grammars used to learn topic models (Johnson, 2010), which have separate rules for each topic. In a similar fashion we will have a separate set of rules for each POS tag.

We conduct experiments both in English and Estonian—a morphologically rich inflective and agglutinative language—and show that the grammars exploiting information about the gold-standard POS tags indeed learn better morphological segmentations in terms of F1-score. The gain in scores when compared to the tag-independent segmentations is up to 14%, depending on the language and the grammar. When the model uses automatically induced tags, the learned segmentations in English are still better than the tag-independent baseline, but the differences in scores are smaller, reaching up to 11% absolute improvement. Although the scores show improvements in Estonian as well, the closer inspection of segmentations of different POS category words reveals that in most cases there are no major differences between segmentations learned with and without tags.

The rest of the paper is organized as follows. In section 2 we briefly introduce the Adaptor Grammars framework, section 3 describes the tag-dependent grammars used in experiments. Section 4 lists the experimental scenarios. In section 5 we describe the experimental setup. Section 6 presents the results, followed by the discussion in section 7, section 8 concludes the paper.

2 Adaptor Grammars

Adaptor Grammars (AG) (Johnson et al., 2007) is a non-parametric Bayesian framework for learning latent structures over sequences of strings. In the current context, the sequence of strings is a sequence of characters making up a word, and the latent structures of interest are the morphemes.

An AG consists of two components: a probabilistic context-free grammar (PCFG) that can generate all possible latent structures for the given inputs, and a Pitman-Yor process (PYP) adaptor function that transforms the probabilities of the parse trees in such a way that the probabilities of the frequently occurring subtrees are much higher than they would be under the PCFG model.

A simple morphological grammar for the AG model could be (Sirts and Goldwater, 2013):

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

where each word consists of one or more morphemes and each morpheme is a sequence of characters. The grammar here uses an abbreviated notation for denoting the recursive rules and thus the first rule is a short-hand writing for:

$$\begin{aligned} \text{Word} &\rightarrow \text{Morphs} \\ \text{Morphs} &\rightarrow \text{Morph} \\ \text{Morphs} &\rightarrow \text{Morph Morphs} \end{aligned}$$

The underline denotes the adapted non-terminals, i.e. the sub-trees rooted in those non-terminals are cached by the model and their probabilities are computed according to the PYP. In the given example the Morph non-terminal is adapted, which means that the model prefers to re-generate the same subtrees denoting the morphemes repeatedly.

We use in our experiments an existing AG implementation[1], the technical details of this implementation are described in (Johnson and Goldwater, 2009).

3 POS-dependent Grammars

The POS-dependent grammars are inspired by the grammars that have been used to learn topic models (Johnson, 2010). Whereas the topic modeling grammars have one rule for every latent topic, the POS-dependent grammars have one rule for each possible tag, which enables the model to cache the subtrees corresponding to morphemes in words with specific syntactic category.

[1]available from http://web.science.mq.edu.au/~mjohnson/Software.htm

Consider for instance a tagset that contains three tags: verb, noun and adjective. Then, in order to make the simple morpheme sequences generating grammar shown in the previous section to be POS-dependent, the rules for each POS tag have to be replicated:

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

Each rule rooted in Word now first generates a non-terminal that corresponds to a particular POS tag and a sequence of POS-specific morphemes. In order to make the grammar complete, we also add rules that generate the terminal symbols corresponding to specific POS tags. We add an underscore to the terminal symbols corresponding to tags to distinguish them from other terminal symbols that are used to generate the words themselves.

$$
\begin{aligned}
\text{Noun} &\rightarrow \text{N\_} \\
\text{Verb} &\rightarrow \text{V\_} \\
\text{Adj} &\rightarrow \text{A\_}
\end{aligned}
$$

We experiment with three different grammars that generate POS-dependent morphological segmentations. The first two of them, **MorphSeq** and **SubMorph** are essentially the same as the ones used for morphological segmentation in (Sirts and Goldwater, 2013). The third one, **CollocMorph**, adds another layer of latent structure on top of morphemes to model morpheme collocations. All three grammars are made tag-dependent by replicating the relevant rules by using tag-specific non-terminals as explained above.

The **MorphSeq**, which was also given as an example in Section 2, is the simplest grammar that just generates each word as a sequence of morphemes. It is essentially a unigram morphology model. The tag-dependent version we used is the following:

$$
\begin{aligned}
\text{Word} &\rightarrow \text{Tag Morph}_{\text{tag}}^{+} && \text{for } \forall \text{ tag} \in T \\
\underline{\text{Morph}}_{\text{tag}} &\rightarrow \text{Morph} && \text{for } \forall \text{ tag} \in T \\
\text{Tag} &\rightarrow \tau && \text{for } \forall \tau \in \mathcal{T} \\
\underline{\text{Morph}} &\rightarrow \text{Char}^{+}
\end{aligned}
$$

Here, T is the set of non-terminal symbols denoting different tags. For instance, this set could be $\{N, V, A\}$ denoting nouns, verbs and adjectives. $\mathcal{T}$ is the corresponding set of tag terminal symbols. Each tag-specific Morph non-terminal also generates a general back-off Morph non-terminal which is shared between all tags. This is desirable because words with different syntactic categories may share the same set of stems. Also, some suffixes are reused across different syntactic categories, either due to agreement or polysemy.

The **SubMorph** grammar adds an additional level of latent structure below the morphemes by generating each morpheme as a sequence of sub-morphemes. In (Sirts and Goldwater, 2013), this was shown to improve the segmentation results considerably. We define the morphemes as tag-specific and specify that sub-morphemes are shared across all tags. In preliminary experiments we also tried to make sub-morphemes tag-specific but this grammar did not produce good results.

$$
\begin{aligned}
\text{Word} &\rightarrow \text{Tag Morph}_{\text{tag}}^{+} && \text{for } \forall \text{ tag} \in T \\
\underline{\text{Morph}}_{\text{tag}} &\rightarrow \text{Morph} && \text{for } \forall \text{ tag} \in T \\
\text{Tag} &\rightarrow \tau && \text{for } \forall \tau \in \mathcal{T} \\
\underline{\text{Morph}} &\rightarrow \text{SubMorph}^{+} \\
\underline{\text{SubMorph}} &\rightarrow \text{Char}^{+}
\end{aligned}
$$

The third grammar, **CollocMorph**, extends the SubMorph grammar and adds another layer of morpheme collocations on top of Morphs. In this grammar both morpheme collocations and morphemes are tag-specific while sub-morphemes are again general:

$$
\begin{aligned}
\text{Word} &\rightarrow \text{Tag Colloc}_{\text{tag}}^{+} && \text{for } \forall \text{ tag} \in T \\
\underline{\text{Colloc}}_{\text{tag}} &\rightarrow \text{Morph}_{\text{tag}}^{+} && \text{for } \forall \text{ tag} \in T \\
\underline{\text{Morph}}_{\text{tag}} &\rightarrow \text{Morph} && \text{for } \forall \text{ tag} \in T \\
\text{Tag} &\rightarrow \tau && \text{for } \forall \tau \in \mathcal{T} \\
\underline{\text{Morph}} &\rightarrow \text{SubMorph}^{+} \\
\underline{\text{SubMorph}} &\rightarrow \text{Char}^{+}
\end{aligned}
$$

4 Experimental Scenarios

In order to assess how much the syntactic tags affect the accuracy of the morphological segmentatations, we conducted experiments using four different scenarios:

1. POS-independent morphological segmentation;
2. POS-dependent morphological segmentation using gold-standard tags;
3. POS-dependent segmentation using syntactic clustering learned with an unsupervised model;
4. POS-dependent segmentation using randomly generated tags.

The first scenario does not use any tags at all and is thus the standard setting used in previous work for conducting unsupervised morphological segmentation. This is the baseline we expect the other, tag-dependent scenarios to exceed.

The second scenario, which uses gold-standard POS tags, is an oracle setting that gives an upper bound of how much the tags can help to improve the segmentation accuracy when using a particular segmentation model. Hypothetically, there could exist tagging configurations, which improve the segmentations more than the oracle tags but in our experiments this was not the case.

The third scenario uses the tags learned with an unsupervised POS induction model. Our expectation here is that the segmentations learned with this scenario are better than the baseline without any tags and worse than using gold-standard tags. The experimental results presented later confirm that this is indeed the case.

The final scenario is the second baseline using tags generated uniformly at random. By evaluating this scenario we hope to show that not just *any* tagging configuration improves the segmentation results but the tags must really correspond at least to some extent to real syntactic tags.

5 Experimental Setup

We conduct experiments in both English and Estonian—a morphologically complex language belonging to Fenno-Ugric language group, using all four scenarios explained above and all three described grammars. AG is a stochastic model and thus it may produce slightly different results on different runs. Therefore, we run the AG in each setting consisting of the language-scenario-grammar triple for 10 times with different random initialisations. We run the sampler for 1000 iterations, after which we collect a single sample and aggregate the samples from all runs by using maximum marginal decoding (Johnson and Goldwater, 2009; Stallard et al., 2012). We use batch initialisation, table label resampling is turned on and all hyperparameters are inferred.

	English	Estonian
MTE types	8438	15132
Eval types	7659	15132
Eval nouns	3831	8162
Eval verbs	2691	4004
Eval adjectives	1629	3111

Table 1: The number of open class words (nouns, verbs and adjectives) used for training and evaluation.

5.1 Data

All experiments were conducted on English and Estonian parts of the Multext-East (MTE) corpus (Erjavec, 2004) that contains G. Orwell's novel "1984". The MTE corpora are morpho-syntactically annotated and the label of each word also contains the POS tag, which we can use in the oracle experiments that make use of gold-standard tags. However, the annotations do not include morphological segmentations. For Estonian, this text is also part of the morphologically disambiguated corpus,[2] which has been manually annotated and also contains segmentations. We use Celex (Baayen et al., 1995) as the source for English gold-standard segmentations, which have been extracted with the Hutmegs package (Creutz and Lindén, 2004). Although not all the words from the MTE English part are annotated in Celex, most of them do, which provides a reasonable basis for our evaluations.

We conduct experiments only on a subset of word types from the MTE corpora, in particular on nouns, verbs and adjectives only. These POS categories constitute open class words and thus are expected to contain the most morphological richness. The statistics about the number of word types in the training and evaluation sets as well as the number of words belonging to different POS categories for both English and Estonian are given in Table 1. The counts of nouns, verbs and adjectives

[2] `http://www.cl.ut.ee/korpused/morfkorpus/index.php?lang=en`

	English				Estonian			
	No POS	**Gold**	**Learned**	**Rand**	**No POS**	**Gold**	**Learned**	**Rand**
MorphSeq	51.4	54.3	55.7	52.5	48.1	53.2	52.5	49.1
SubMorph	63.3	69.6	68.1	64.3	66.5	66.5	64.3	65.5
CollocMorph	56.8	71.0	68.0	66.6	65.4	68.5	66.5	68.4

Table 2: F1-scores of all experiments in English and Estonian using different grammars and settings. **MorphSeq** generates sequences of morphemes, **SubMorph** adds the sub-morphemes, and **CollocMorph** adds the morpheme collocations. **No POS** are the models trained without tags, **Gold** uses goldstandard POS tags, **Learned** uses tags learned by an unsupervised POS induction model, and **Rand** uses randomly generated tags.

do not add up to the total number of evaluated word types because some of the words in the corpus are ambiguous and occur in different syntactic roles.

The automatically induced syntactic tags were learned with an unsupervised POS induction model (Sirts and Alumäe, 2012).[3] The main reason for choosing this model was the fact that it has been evaluated on the same MTE corpus we use for learning on both English and Estonian and has shown to produce reasonably good tagging results.

5.2 Input Format

For POS-independent segmentation we just train on the plain list of words. For tag-dependent experiments we have to reformat the input so that each word is preceded by its tag, which will be parsed by the left branch of the first rule in each grammar. For instance, the input for the tag-independent AG model for a noun `table` is just a sequence of characters separated by spaces:

```
t a b l e
```

However, for the tag-dependent model it has to be reformatted as:

```
N_ t a b l e,
```

where `N_` is the terminal symbol denoting the noun POS.

The tag assignments of the unsupervised POS induction model are just integer numbers and thus for instance, if the model has assigned a tag 3 to the noun `table` then the input has to be reformatted as:

```
3_ t a b l e,
```

where `3_` is the terminal symbol denoting the induced tag cluster 3.

[3]The results were obtained from the authors.

The number of different tags in automatically learned tagset is larger than three, although the training still contains only nouns, verbs and adjectives. This is because even the best unsupervised POS taggers usually learn quite noisy clusters, where one POS category may be split into several different clusters and each cluster may contain a set of words belonging to a mix of different POS categories.

For the random tag baseline we just generate for each word a tag uniformly at random from the set of three tags: $\{0, 1, 2\}$, and reformat the input in a similar way as explained above about the automatically induced tags.

5.3 Evaluation

We evaluate the segmentations using the F1-score of the learned boundaries. The evaluation is type-based (as is also our training), meaning that the segmentation of each word type is calculated into the score only once. This is the simplest evaluation method for morphological segmentation and has been widely used in previous work (Virpioja et al., 2011).

6 Results

We present two sets of results. First we give the F-scores of all evaluated words in each language and then we split the evaluation set into three and evaluate the results for all three POS classes separately.

6.1 General results

The segmentation results are given in Table 2. The first thing to notice is that the models trained with gold-standard POS tags always perform the best. Intuitively this was expected, however, the differences between segmentation F1-scores are in most

	English				Estonian			
	No POS	**Gold**	**Learned**	**Rand**	**No POS**	**Gold**	**Learned**	**Rand**
MorphSeq N	49.5	50.7	52.5	50.0	51.6	56.3	55.0	52.5
MorphSeq V	54.4	59.9	60.7	56.4	46.3	55.9	53.7	47.0
MorphSeq A	50.2	54.6	55.1	51.7	41.1	42.5	44.6	42.7
SubMorph N	61.1	66.9	65.7	61.3	64.6	65.8	64.1	63.8
SubMorph V	67.8	75.4	73.7	70.9	78.9	80.8	75.3	77.8
SubMorph A	61.2	67.0	64.7	60.8	56.6	51.3	51.6	55.5
CollocMorph N	55.0	68.5	65.9	64.3	66.2	67.9	66.8	67.4
CollocMorph V	60.5	75.9	73.4	72.1	68.7	76.0	75.2	79.6
CollocMorph A	54.4	69.1	64.6	62.8	60.0	61.7	55.7	57.6

Table 3: F1-scores of segmentations for different POS classes in English and Estonian using different grammars and settings. **MorphSeq** generates sequences of morphemes, **SubMorph** adds the sub-morphemes, and **CollocMorph** adds the morpheme collocations. **N** denotes nouns, **V** stands for verbs and **A** are adjectives. **No POS** are the models trained without tags, **Gold** uses goldstandard POS tags, **Learned** uses tags learned by an unsupervised POS induction model, and **Rand** uses randomly generated tags.

cases only few percentage points. The only notable exception is English trained with the CollocMorph grammar where the difference with the tag-independent baseline is 14%. However, the baseline score for the CollocMorph grammar in English is much lower than the baseline with the SubMorph grammar, which has a simpler structure. In order to understand why this was the case, we looked at the precision and recall of the CollocMorph grammar results. We found that for the baseline model, the precision is considerably lower than the recall, which means that the results are oversegmented. We always extracted the segmentations from the middle latent level of the CollocMorph grammar and in most cases this gave the best results. However, for the English baseline model, extracting the segmentations from the morpheme collocation level would have given more balanced precision and recall and also a higher F1-score, 60.9%, which would have reduced the difference with the segmentations learned with gold-standard POS tags to 10%.

When the gold-standard POS tags are substituted with the automatically learned tags, the segmentation scores drop as expected. However, in most cases the segmentations are still better than those learned without any tags, although the differences again fall in the range of only few percentage points. In one occasion, namely with the SubMorph grammar in Estonian, the score actually drops by 2% points and with CollocMorph grammar in Estonian the improvement is only about 1%. English segmentations learned with CollocMorph grammar again improve the most over the baseline without tags, gaining over 11% improvement in F1-score.

The last setting we tried used random POS tags. Here we can see that in most cases using random tags helps while in one case—again Estonian SubMorphs—it degrades the segmentation results, leading to lower scores than the baseline without tags. In English, the randomly generated tags always improve the segmentation results over the tag-independent baseline but the results are worse than the segmentations learned with the automatically induced tags. In Estonian, however, for the two more complex grammars, SubMorph and CollocMorph, the randomly generated tags lead to slightly better segmentations than the automatically induced tags. This is a curious result because it suggests that some kind of partitioning of words is helpful for learning better segmentations but that in some cases the resemblance to true POS clustering does not seem that relevant. It could also be that the partitioning of words into nouns, verbs and adjectives only was too coarse for Estonian, which realises many fine-grained morpho-syntactic functions inside each of those POS classes with different suffixes.

6.2 Results of different POS classes

In order to gain more insight into the presented results we also computed F1-scores separately for each of the three POS classes. Those results are given in Table 3. From this table we can see that the segmentation scores are quite different for words with different POS tags. For English, the scores for nouns and adjectives are similar, while the verbs are segmented much more accurately. This is reasonable because usually verbs are much simpler in structure, consisting usually of a stem and a single inflectional suffix, while nouns and adjectives can contain several stems and both derivational and inflectional suffixes. In all cases, segmentations learned with either gold or induced tags are better than segmentations learned with random or no tags at all. CollocMorph is the only grammar where the segmentations learned with random tags improve significantly over the tag-independent baseline. The gap is so large because, as explained above, the precision and recall of the CollocMorph grammar without tags evaluated on the middle grammar level are heavily biased towards recall and the results are oversegmented, while the grammar using randomly generated tags manages to learn segmentations with more balanced precision and recall.

In Estonian, the results are more mixed. For nouns, the only grammar where the POS tags seems to help is the simplest MorphSeq, while with other grammars even specifying gold standard POS tags only leads to minor improvements. Verbs, on the other hand gain quite heavily from tags when using MorphSeq or CollocMorph grammar, while with SubMorph grammar the tag-independent baseline is already very high. Closer inspection revealed that evaluating the CollocMorph grammar on the morpheme collocation level would have given more balanced precision and recall and a tag-independent F-score of 83.2%, which is even higher than the SubMorph 78.9%. Also, evaluating segmentations learned with gold standard tags on that level would have improved the F-score even more up to 89.4%. At the same time, the scores of segmentations learned with both random and induced tags would have dropped. Finally, the scores of the Estonian adjectives are in general the lowest and with both SubMorph and CollocMorph grammar adding the tags in most cases does not give any improvements.

7 Discussion

The main goal of this study was to assess, whether and how much do POS tags help to learn better morphological segmentations. The basis for this question was the intuition that because POS tags and morphological segmentations are linguistically related they should be able to exploit synergies during joint learning. However, the previous work in joint POS induction and morphological segmentation has failed to show the clear gains. Therefore we designed an oracle experiment that uses gold-standard POS tags to measure the upper bound of the gains the POS tags can provide in learning morphological segmentations.

On English, using gold-standard POS tags helps to gain 3-14% of F1-score depending on the grammar, while in Estonian the gains remain between 0-5%. The accuracy gained from tags varies for different POS classes. Both in Estonian and English verbs seem to benefit the most, which can be explained by the fact that in both languages verbs have the simple structure consisting mostly of a stem and an optional inflectional suffix which informs the POS class. At the same time, nouns and adjectives can also contain different derivational morphemes which can be shared by both POS classes. Also, in Estonian the adjectives must agree with nouns in case and number but the sets of suffixes both word classes use are not completely overlapping, which makes the relations between POS tags and segmentations more complex. Another reason for the difference between gains in English and Estonian can be that Estonian as morphologically more complex language may be able to exploit the capacity of the generative AG model more effectively even without tags. At the same time the morphologically simpler English gains more from adding additional information in the form of POS tags.

In general, the effect of POS tags on the segmentation accuracy is not huge, even when the linguistically correct gold-standard tags are used. One reason here can be that we provided the system with very coarse-grained syntactic tags while morphological suffixes are more closely related to the more fine-grained morpho-syntactic functions. This is especially true in English where for instance different verbal suffixes are almost in one-to-one relation with different morpho-syntactic functions. The situation is probably more complex with morphologically rich languages such as Estonian where there are different inflectional classes, which all express the same morpho-syntactic function with different allomorphic suffixes.

Correct	No POS	Gold	Learned	Rand
condemn_ed	cond_em_n_ed	condemn_ed	condem_n_ed	con_demn_ed
grovell_ing	grovel_ling	gro_vell_ing	gro_vell_ing	gro_velling
catalogue	cat_a_logue	cata_logue	cata_logue	cata_logue
propp_ed	pro_p_p_ed	prop_ped	prop_p_ed	propped
match_es	m_atch_e_s	match_es	match_es	matches
suuna_ga (N)	suu_na_ga	suuna_ga	suuna_ga	suunaga
sammal_t (N)	samm_al_t	samm_alt	samm_alt	samm_alt
pääse_ks (V)	pääs_e_ks	pääse_ks	pääse_ks	pääse_ks
pikkuse_d (A)	pikku_sed	pikku_se_d	pikkuse_d	pikku_sed
kükita_sid (V)	kü_ki_ta_sid	küki_ta_sid	küki_tas_id	kükita_sid

Table 4: Examples of both English and Estonian mostly incorrectly segmented words learned with CollocMorph grammar.

Although using automatically induced tags almost always improves the segmentation results, the gains are in most cases quite small. We assume that the induced tags cannot improve the segmentations more than the gold-standard tags. However, it is not clear whether the accuracy of the induced POS tags themselves affects the segmentations accuracy much. The experiments with the random baseline showed that the POS tags should not be completely random but how large differences in tagging accuracies start affecting the segmentations' quality remains to be studied in future works.

Some examples of segmented words for both English and Estonian are given in Table 4. For those examples, the POS-independent grammar has learned incorrect segmentations. The various POS-dependent grammars are in some cases able to learn correct segmentations, in some cases learn more correct segmentations, but in some cases also learn equally false segmentations. For instance for English, all POS-dependent grammars are able to improve the segmenation of the word *condemned*, but only the grammar informed by gold POS tags gets it exactly right. The word *matches* is segmented correctly by grammars using both gold and induced tags, while the grammar with random tags undersegments. In Estonian for instance, only the grammar using random tags gets the word *kükitasid* right, while all the other grammars oversegment it. On the other hand, the adjective *pikkused* is correctly segmented only by the grammar using automatically learned tags and all the other grammars either oversegment or place the segment boundary in an incorrect location.

8 Conclusion

Morphology is a complex language phenomenon which is related to many different phonological, orthographic, morpho-syntactic and morphotactic aspects. This complexity has the potential to create synergies in a generative model where several aspects of the morphology are learned jointly. However, setting up a joint model that correctly captures the desired regularities is difficult and thus it may be useful to study the synergistic potentials of different components in a more isolated setting.

The experiments in this paper focused on the relations between syntactic tags and concatenative morphological segmentations. We showed that both gold-standard POS tags as well as automatically induced tags can help to improve the morphological segmentations. However, the gains are on average not large—5.3% with gold-standard tags and 3.9% with induced tags. Moreover, deeper analysis by evaluating the segmentations of words from different POS classes separately reveals that in Estonian even the goldstandard POS tags do not affect the segmentations much.

These results suggest that perhaps other relations should be studied of how to use various aspects of morphology to create synergies. For instance, POS tags are clearly related to paradigmatic relations. Also, clustering words according to morpho-syntactic function could benefit from using methods developed for learning distributional representations. Finally, it could be helpful to learn morphological structures jointly on both orthographic and phonological level.

References

R. H. Baayen, R. Piepenbrock, and L. Gulikers. 1995. The CELEX lexical database (release 2).

Taylor Berg-Kirkpatrick, Alexandre B. Côté, John DeNero, and Dan Klein. 2010. Painless unsupervised learning with features. In *Proceedings of Human Language Technologies: The Conference of the North American Chapter of the Association for Computational Linguistics*, pages 582–590.

Burcu Can and Suresh Manandhar. 2009. Unsupervised learning of morphology by using syntactic categories. In Francesca Borri, Alessandro Nardi, and Carol Peters, editors, *Working Notes for the CLEF 2009 Workshop*.

Burcu Can. 2011. *Statistical Models for Unsupervised Learning of Morphology and POS Tagging*. Ph.D. thesis, The University of York.

Christos Christodoulopoulos, Sharon Goldwater, and Mark Steedman. 2011. A Bayesian mixture model for part-of-speech induction using multiple features. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 638–647.

Mathias Creutz and Krista Lagus. 2007. Unsupervised models for morpheme segmentation and morphology learning. *ACM Transactions on Speech and Language Processing*, 4(1):3:1–3:34.

Mathias Creutz and Krister Lindén. 2004. Morpheme segmentation gold standards for finnish and english. Publications in Computer and Information Science Report A77, Helsinki University of Technology.

Tomaž Erjavec. 2004. MULTEXT-East version 3: Multilingual morphosyntactic specifications, lexicons and corpora. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, pages 1535–1538.

Stella Frank, Frank Keller, and Sharon Goldwater. 2013. Exploring the utility of joint morphological and syntactic learning from child-directed speech. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*.

Dayne Freitag. 2005. Morphology induction from term clusters. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CONLL '05)*.

John Goldsmith. 2001. Unsupervised learning of the morphology of a natural language. *Computational Linguistics*, 27:153–198.

Sharon Goldwater, Thomas L. Griffiths, and Mark Johnson. 2006. Interpolating between types and tokens by estimating power-law generators. In *Advances in Neural Information Processing Systems 18*, pages 459–466.

Mark Johnson and Sharon Goldwater. 2009. Improving nonparametric Bayesian inference: Experiments on unsupervised word segmentation with Adaptor Grammars. In *Proceedings of Human Language Technologies: The Conference of the North American Chapter of the Association for Computational Linguistics*, pages 317–325.

Mark Johnson, Thomas L. Griffiths, and Sharon Goldwater. 2007. Adaptor Grammars: a framework for specifying compositional nonparametric Bayesian models. In *Advances in Neural Information Processing Systems 19*, pages 641–648.

Mark Johnson. 2008. Unsupervised word segmentation for Sesotho using Adaptor Grammars. In *Proceedings of the Tenth Meeting of ACL Special Interest Group on Computational Morphology and Phonology*.

Mark Johnson. 2010. PCFGs, topic models, Adaptor Grammars and learning topical collocations and the structure of proper names. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1148–1157.

Yoong Keok Lee, Aria Haghighi, and Regina Barzilay. 2010. Simple type-level unsupervised POS tagging. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 853–861.

Yoong Keok Lee, Aria Haghighi, and Regina Barzilay. 2011. Modeling syntactic context improves morphological segmentation. In *Proceedings of the Fifteenth Conference on Natural Language Learning*, pages 1–9.

Kairit Sirts and Tanel Alumäe. 2012. A hierarchical Dirichlet process model for joint part-of-speech and morphology induction. In *Proceedings of Human Language Technologies: The Conference of the North American Chapter of the Association for Computational Linguistics*, pages 407–416.

Kairit Sirts and Sharon Goldwater. 2013. Minimally-supervised morphological segmentation using adaptor grammars. *Transactions of the Association for Computational Linguistics*, 1(May):231–242.

David Stallard, Jacob Devlin, Michael Kayser, Yoong Keok Lee, and Regina Barzilay. 2012. Unsupervised morphology rivals supervised morphology for arabic mt. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2*, pages 322–327.

Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 173–180.

Sami Virpioja, Ville T. Turunen, Sebastian Spiegler, Oskar Kohonen, and Mikko Kurimo. 2011. Empirical comparison of evaluation methods for unsupervised learning of morphology. *Traitement Automatique des Langues*, 52(2):45–90.