

# Induction of Fine-grained Part-of-speech Taggers via Classifier Combination and Crosslingual Projection

**Elliott Franco Drábek**

Department of Computer Science  
Johns Hopkins University  
Baltimore, MD 21218  
edrabek@cs.jhu.edu

**David Yarowsky**

Department of Computer Science  
Johns Hopkins University  
Baltimore, MD 21218  
yarowsky@cs.jhu.edu

## Abstract

This paper presents an original approach to part-of-speech tagging of fine-grained features (such as case, aspect, and adjective person/number) in languages such as English where these properties are generally not morphologically marked.

The goals of such rich lexical tagging in English are to provide additional features for word alignment models in bilingual corpora (for statistical machine translation), and to provide an information source for part-of-speech tagger induction in new languages via tag projection across bilingual corpora.

First, we present a classifier-combination approach to tagging English bitext with very fine-grained part-of-speech tags necessary for annotating morphologically richer languages such as Czech and French, combining the extracted features of three major English parsers, and achieve fine-grained-tag-level syntactic analysis accuracy higher than any individual parser.

Second, we present experimental results for the cross-language projection of part-of-speech taggers in Czech and French via word-aligned bitext, achieving successful fine-grained part-of-speech tagging of these languages without any Czech or French training data of any kind.

## 1 Introduction

Most prior research in part-of-speech (POS) tagging has focused on supervised learning over a tagset such as the Penn Treebank tagset for English, which is restricted to features that are morphologically distinguished in the focus language. Thus the only verb person/number distinction made in the Brown Corpus/Penn Treebank tagset is VBZ (3rd-person-singular-present), with no corresponding person/number distinction in other tenses. Similarly, adjectives in English POS tagsets typically have no distinctions for person, number or case because such properties have no morphological surface distinction, although they do for many other languages.

This essential limitation of the Brown/Penn POS subtag inventory to morphologically realized distinctions in English dramatically simplifies the problem by reducing the tag entropy per surface form (the adjective *tall* has only one POS tag (JJ) rather than numerous singular, plural, nominative, accusative, etc. variants), increasing both the stand-alone effectiveness of lexical prior models and word-suffix models for part-of-speech tagging.

However, for many multilingual applications, including feature-based word alignment in bilingual corpora and machine translation into morphologically richer languages, it is helpful to extract finer-grained lexical analyses on the English side that more closely parallel the morphologically realized tagset of the second (source or target) language.

In particular, prior work on translingual part-of-speech tagger projection via parallel bilingual corpora (e.g. Yarowsky et al., 2001) has been limited to inducing part-of-speech taggers in second languages (such as French or Czech) that only assign tags at the granularity of their source language (i.e.

the Penn Treebank-granularity distinctions from English). The much richer English tagsets achieved here can allow these tagger projection techniques to transfer richer tag distinctions (such as case and verb person/number) that are important to the full analysis of these languages, using only bilingual corpora with the morphologically impoverished English.

For quickly retargetable machine translation, the primary focus of effort is overcoming the extreme scarcity of resources for the low density source language. Sparsity of conditioning events for a translation model can be greatly reduced by the availability of automatic source-language analysis. In this research we attempt to induce models for the automatic analysis of morphological features such as case, tense, number, and polarity in both the source and target languages with this end in mind.

## 2 Prior Work

### 2.1 Fine-grained part-of-speech tagging

Most prior work in fine-grained part-of-speech tagging has been limited to languages such as Czech (e.g. Hajič and Hladká, 1998) or French (e.g. Foster etc.) where finer-grained tagset distinctions are morphologically marked and hence natural for the language. In support of supervised tagger learning of these languages, fine-trained tagset inventories have been developed by the teams above at Charles University (Czech) and Université de Montréal (French). The tagset developed by Hajič forms the basis of the distinctions used in this paper.

The other major approach to fine-grained tagging involves using tree-based tags that capture grammatical structure. Bangalore and Joshi (1999) have utilized “supertags” based on tree-structures of various complexity in the tree-adjoining grammar model. Using such tags, Brants (2000) has achieved the automated tagging of a syntactic-structure-based set of grammatical function tags including phrase-chunk and syntactic-role modifiers trained in supervised mode from a treebank of German.

### 2.2 Classifier combination for part-of-speech tagging

There has been broad work in classifier combination at the tag-level for supervised POS tagging models. For example, Márquez and Rodríguez (1998) have performed voting over an ensemble of decision tree and HMM-based taggers for supervised En-

glish tagging. Murata et al. (2001) have combined neural networks, support vector machines, decision lists and transformation-based-learning approaches for Thai part-of-speech tagging. In each of these cases, annotated corpora containing the full tagset granularity are required for supervision.

Henderson and Brill (1999) have approached parsing through classifier combination, using bagging and boosting for the performance-weighted voting over the parse-trees from three anonymous statistical phrase-structure-based parsers. However, as their switching and voting models assumed equivalent phrase-structure conventions for merger compatibility, it is not clear how a dependency parsing model or other divergent syntactic models could be integrated into this framework. In contrast, the approach presented below can readily combine syntactic analyses from highly diverse parse structure models by first projecting out all syntactic analyses onto a common fine-grained lexical tag inventory.

### 2.3 Projection-based Bootstrapping

Yarowsky et al. (2001) performed early work in the cross-lingual projection of part-of-speech tag annotations from English to French and Czech, by way of word-aligned parallel bilingual corpora. They also used noise-robust supervised training techniques to train stand-alone French and Czech POS taggers based on these projected tags. Their projected tagsets, however, were limited to those distinctions captured in the English Penn treebank inventory, and hence failed to make many of the finer grained distinctions traditionally assumed for French and Czech POS tagging, such as verb person, number, and polarity and noun/adjective case.

Probst (2003) pursued a similar methodology for the purposes of tag projection, using a somewhat expanded tagset inventory (e.g. including adjective number but not case), and focusing on target-language monolingual modeling using morpheme analysis. Cucerzan and Yarowsky (2003) addressed the problem of grammatical gender projection via the use of small seed sets based on natural gender. Another distinct body of work addresses the problem of parser bootstrapping based on syntactic dependency projection (e.g. Hwa et al. 2002), often using approaches based in synchronous parsing (e.g. Smith and Smith, 2004).

Word	Core POS	Prsn	Num.	Case	Tns/ Asp.	Pol.	Voi.
The	DT	3	PL.	NOM.			
books	NN	3	PL.	NOM.			
were	VB	3	PL.		PAST	+	ACT.
provoking	VB	3	PL.		PAST- PROG.	+	ACT.
laughter	NN	3	S.	ACC.			
with	IN						
their	DT	3	PL.	'WITH'			
curious	JJ	3	PL.	'WITH'			
titles	NN	3	PL.	'WITH'			

Figure 1: Example of fine-grained English POS tags

Word	Core POS	Prsn	Num.	Case	Tns/ Asp.	Pol.	Voice
Les	DT	3	PL.	NOM.			
livres	NN	3	PL.	NOM.			
provoquaient	VB	3	PL.		PAST- PROGR.	+	ACT.
des	DT	3	PL.	ACC.			
rires	NN	3	PL.	ACC.			
avec	IN						
ses	DT	3	PL.	'WITH'			
titres	NN	3	PL.	'WITH'			
curieux	JJ	3	PL.	'WITH'			

Figure 2: Example of fined-grained POS tags projected onto a French translation

### 3 Tagsets

We use Penn treebank-style part-of-speech tags as a substrate for further enrichment (for all of the experiments described here, text was first tagged using the fnTBL part-of-speech tagger (Ngai and Florian, 2001)). Each Penn tag is mapped to a core part-of-speech tag, which determines the set of fine-grained tags further applicable to each word. The fine-grained tags applicable to nouns, verbs, and adjective are shown in Table 1. This paper concentrates on these most important core parts-of-speech.

The example English sentence in Figure 1 illustrates several key points about our tagset. Some of the information we are interested in is already expressed by the Penn-style tags – the NN *titles* is plural; the VBD *were* is in the past tense. For these, our goal is simply to make these facts explicit.

On the other hand, *curious* could also be meaningfully said to be semantically plural, and most importantly for us, the corresponding word in a translation of this sentence into many other languages would be morphologically plural. Similarly, the head verb *provoking* is also semantically in the past tense, and is likely to be translated to a past-tense form in many languages, even though in this example the actual tense marking is on *were*. We expect the ‘pastness’ of the action to be much more stable cross-linguistically, than the particular division of labor between the head word and the auxiliary. By prop-

	VB	JJ	NN	Range
<b>Person</b>	•	•	•	1 / 2 / 3
<b>Number</b>	•	•	•	SINGULAR PLURAL
<b>Case</b>		•	•	NOMINATIVE ACCUSATIVE GENITIVE PREPOSITION-‘IN’ PREPOSITION-‘OF’ ...
<b>Degree</b>		•		POSITIVE COMPARATIVE SUPERLATIVE
<b>Tense</b>	•			PAST PRESENT FUTURE
<b>Perfectivity</b>	•			+ / -
<b>Progressivity</b>	•			+ / -
<b>Polarity</b>	•			+ / -
<b>Voice</b>	•			ACTIVE / PASSIVE

Table 1: The fine-grained POS inventory used for English

agating these features from where they are explicit to where they are not, we hope to make information more directly available for projection. Another important class of information we would like to make available concerns syntactic relations, which many languages mark with morphological case. This is an issue that involves deep, complex, and ambiguous mappings, which we are not yet prepared to treat in their fullness. For now, we observe that *curious* and *titles* are both dominated by *with*.

Because of intent to mark whatever information is recoverable, some of our tags require some interpretation. For example, English has little or no morphological realization of syntactic case, but the essential information of case, relationship of a noun with its governor, is recoverable from contextual information, so we defined it in these terms. To avoid loss of information, we chose to remain agnostic about deeper analyses, such as the identification of theta roles or predicate-argument relationships, and restricted ourselves to a direct representation of surface relationships. We identified subjects, direct and indirect objects, non-heads of noun compounds, possessives, and temporal adjuncts, and created a distinct tag for the objects of each distinct preposition.

Our ideal would be to have as expansive and detailed a tagset as possible, a ‘quasi-universal’ tagset which could cover whatever set of distinctions might be relevant for any language onto which we might

Feature	Antecedent → CONSEQUENT
Noun Number	NN → SINGULAR NNS → PLURAL
Verb Tense	VBD → PAST (will shall) RB* VB → FUTURE

Figure 3: Examples of locally recoverable features

project our analysis. A completely universal tagset would require that the morphological distinctions made by the world’s languages come from a limited pool of possibilities, based on non-arbitrary semantic distinctions, and further would require that the relevant semantic information be recoverable from English text. The tagset we are using now is shaped in part by exceptions to these conditions. For example, we have put off implementing tagging of gender given the notoriously arbitrary and inconsistent assignment of grammatical gender across languages (although Cucerzan and Yarowsky (2003) were able to show success on projection-based analysis of grammatical gender as well).

In the end, we have settled on a set of distinctions very similar to those realized by the morphologically richer of the European languages, with the noticeable absence of gender. Table 1 describes the features we chose on this basis (definiteness and mood features were developed for English but not projected to French or Czech, and are not treated in this paper).

#### 4 Methods – English Tagging

The features we tagged vary widely in their degree of morphological versus syntactic marking, and the difficulty of their monolingual English detection. For some, tagging is simply a matter of explicitly separating information contained in the Penn part-of-speech tags, while others can be tagged to a high degree of accuracy with simple heuristics based on local word and part-of-speech tag patterns. These include number for nouns and adjectives, person (trivially) for nouns, degree for adjectives, polarity, voice, and aspect (perfectivity and progressivity) for verbs, as well as tense for some verbs. Figure 3 shows example rules for some of these easier cases.

The more difficult features are those whose detection requires some degree of syntactic analysis. These include case, which summarizes the relation of each noun with its governor, and the agreement-based features: we define person, number, and case

for attributive adjectives by agreement with their head nouns, number and person for verbs and predicate adjectives by agreement with their subjects, and tense for some verbs by agreement with their inflected auxiliaries.

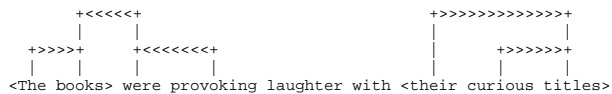
We investigated four individual approaches for the syntax-features – a regular-expression-based quasi-parser, a system based on Dekang Lin’s MiniPar (Lin, 1993), a system based on the Collins parser (Collins, 1999), and one based on the CMU Link Grammar Parser (Sleator and Temperley, 1993), as well as a family of voting-based combination schemes.

#### 4.1 Regular-expression Quasi-parser

The regular-expression ‘quasi-parser’ takes a direct approach, using several dozen heuristics based on regular-expression-like patterns over words, Penn part-of-speech tags, and the output of the fnTBL noun chunker. Use of the noun chunker facilitates identification of noun/dependent relationships within chunks, and extends the range of patterns identifying noun/governor relationships across chunks.

The output of the quasi-parser consists of two parts: a case tag for each noun in a sentence, and a set of agreement links across which other features are then spread. We call this a direct approach because the links are defined operationally, directly indicating the spreading action, rather than representing any deeper syntactic analysis.

In the diagram of the example sentence below, an arrow from one word to another indicates that the former takes features from the latter. The example also shows the context patterns by which the nouns in the sentence receive case.



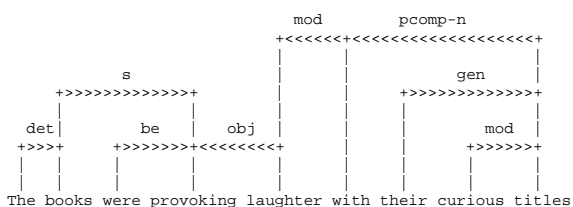
Word	Context Pattern → CASE TAG
<i>laughter</i>	VB (genitive-NP)* • → ACCUSATIVE
<i>titles</i>	with (genitive-NP)* • → PREP-WITH
<i>books</i>	default → NOMINATIVE

#### 4.2 MiniPar and the CMU Link Grammar Parser

For MiniPar, the Collins parser, and the CMU link grammar parser, we developed for each a set of minimal-complexity heuristics to transform the parser output into the specific conceptions of dependency and case we had developed for the first pass.

MiniPar produces a labeled dependency graph, which yields a straightforward extraction of the information needed for this task. Case tagging is a simple matter of mapping the set of dependency labels to our case inventory. Our agreement links are almost a subset of MiniPar’s dependencies (with some special treatment of subject/auxiliary/main-verb triads, as shown in the example sentence).

The figure below presents MiniPar’s raw output for the example sentence, along with some example dependency-label/case-tag rules. The agreement links extracted from the dependency graph are identical (in this case) to those produced by the regular-expression quasi-parser.



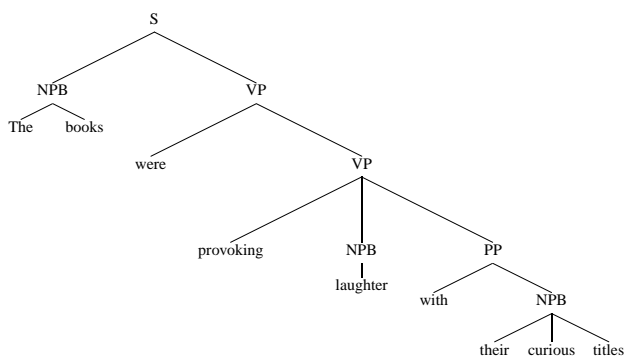
Word	Dependency Label → CASE TAG
<i>books</i>	s → NOMINATIVE
<i>laughter</i>	obj → ACCUSATIVE
<i>titles</i>	pcomp-n:with → PREP-WITH

The output of the CMU link grammar parser has properties similar to MiniPar, and thus tag extraction was handled in a similar fashion.

### 4.3 Collins Parser

The Collins Parser produces a Penn-Treebank-style constituency tree, with head labels. Although we could have used the head-labels to operate on the dependency graph as with MiniPar, we chose to concentrate on addressing the weakest point of our previous systems, the identification of case. Our algorithm traces the path from each noun to the root of the tree, stopping at the first node which we judged to reliably indicate case.

We did not directly extract any further information from the Collins parser output. Instead, the remainder of the system is identical to the regular-expression quasi-parser. However, because the system uses nominative case to identify verb subjects, we did expect to see some improvements in agreement-based features as well.



Word	Path to Root → CASE TAG
<i>books</i>	NPB:S → NOMINATIVE
<i>laughter</i>	NPB:VP:VP:VP:S → ACCUSATIVE
<i>titles</i>	NPB:PP(with):VP:VP:S → PREP-WITH

### 4.4 Parser Combination

The fine-grained taggers based on the four participating parsers exhibited significant differences in their strengths and weaknesses, suggesting potential benefit from combining them. Lacking tag-level numerical scores and development data for weight-training, we restricted ourselves to simple voting mechanisms. We chose to do all of the combinations at the end of the process, voting separately on tags for specific features of specific words. Without tag-level probabilities from the one-best parser outputs, we were still able to use the combination protocols to achieve a coarse-grained confidence measure.

We compared a series of seven combination protocols of increasing leniency to investigate precision/recall tradeoffs. The strictest, ‘4:0’, produces an output only when there are four votes for the favored tag, and no votes for any other. Analogously, protocols ‘3:0’, ‘2:0’ and ‘1:0’ also allow no dissent, but allow progressively more abstentions. Continuing the sequence, protocol ‘2:1’ proposes a tag as long as there is a clear majority, ‘2:2’ as long as supporters are not outnumbered by dissenters, and ‘1:3’ whenever possible. To break ties in the latter two protocols, we favored first the CMU Link Parser, then Collins, then MiniPar, then Regexp. (Lacking sufficient labeled data for fine-tuning, we ordered them arbitrarily.)

## 5 Evaluation of English POS Tagging

Before we began the development of our taggers, we created standard tagging guidelines, and hand annotated a 3013-word segment of the English side of the Canadian Hansards, to be used for evaluation.

Core POS	Feature	MiniPar	Regexp	Collins	CMU Link	1:3
JJ	num	86.8	87.7	87.7	87.9	<b>88.4</b>
	case	65.1	74.5	76.4	79.2	<b>80.6</b>
	deg	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
	'French'	86.8	87.7	87.7	87.9	<b>88.4</b>
NN	'Czech'	57.9	64.3	67.1	68.1	<b>70.5</b>
	num	<b>99.7</b>	<b>99.7</b>	<b>99.7</b>	<b>99.7</b>	<b>99.7</b>
	case	65.9	74.8	77.8	77.3	<b>80.0</b>
	'French'	<b>99.7</b>	<b>99.7</b>	<b>99.7</b>	<b>99.7</b>	<b>99.7</b>
VB	'Czech'	65.0	74.8	77.8	77.2	<b>79.9</b>
	num	77.2	64.8	65.5	66.8	<b>78.1</b>
	tns	<b>77.2</b>	66.8	67.1	67.1	76.3
	prsn	<b>88.0</b>	75.0	74.3	73.4	86.5
	pol	96.3	<b>96.6</b>	<b>96.6</b>	<b>96.6</b>	<b>96.6</b>
	voice	<b>88.0</b>	<b>88.0</b>	<b>88.0</b>	<b>88.0</b>	<b>88.0</b>
	'French'	61.8	61.3	61.0	61.3	<b>67.5</b>
overall	'Czech'	61.3	61.1	60.8	61.1	<b>67.1</b>
	'French'	82.6	82.5	82.4	83.2	<b>85.2</b>
	'Czech'	62.5	67.8	69.4	70.5	<b>73.3</b>

Table 2: English tagging forced-choice accuracy

Core POS	Feature	Mini Par	Regexp	Collins	CMU Link	2:0	1:0	1:2
JJ	num	79.1	81.3	81.3	82.2	81.2	83.8	<b>83.9</b>
	case	72.1	79.2	83.0	78.9	78.1	79.1	<b>84.2</b>
	deg	100	100	100	100	100	100	<b>100</b>
	'Czech'	67.6	72.2	76.0	74.3	70.4	73.4	<b>77.9</b>
NN	num	99.7	99.7	99.7	99.7	99.7	99.7	<b>99.7</b>
	case	68.5	75.5	<b>78.6</b>	77.9	72.6	72.5	78.1
	'Czech'	68.1	75.2	<b>78.3</b>	77.7	72.2	72.1	77.8
	tns	78.0	68.5	68.7	68.0	68.7	78.3	<b>78.3</b>
VB	num	72.7	61.3	61.2	61.3	61.1	76.1	<b>77.1</b>
	prsn	77.2	66.5	65.4	63.9	64.0	78.3	<b>79.0</b>
	pol	96.3	96.6	96.6	96.5	96.5	96.5	<b>96.6</b>
	voice	88.0	88.0	88.0	88.0	88.0	88.0	<b>88.0</b>
	'French'	61.7	50.7	50.2	50.1	50.6	64.8	<b>65.6</b>
	'Czech'	61.1	50.5	49.9	49.8	50.4	64.5	<b>65.2</b>
	all	'French'	81.9	78.7	78.5	78.5	83.6	78.9
	'Czech'	65.4	66.0	67.8	69.3	68.9	63.5	<b>72.9</b>

Table 3: English tagging F-measure

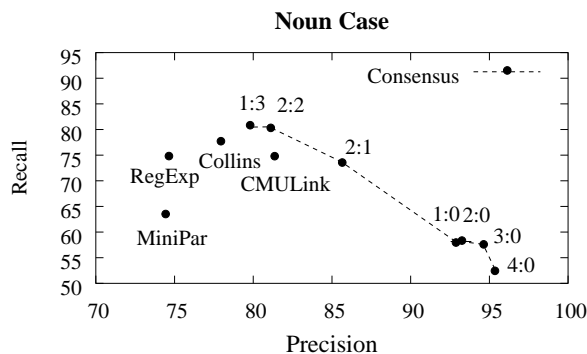


Figure 4: Precision versus Recall – Noun case

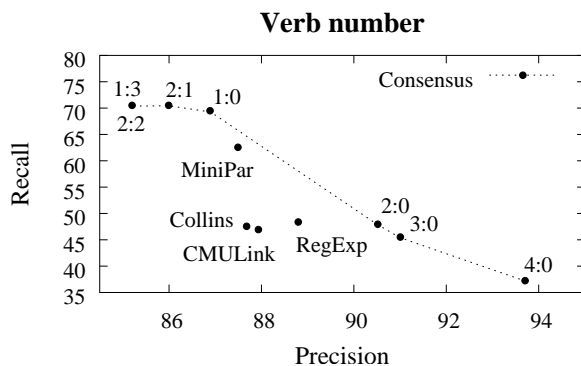


Figure 5: Precision versus Recall – Verb number

Table 2 shows system accuracy on test data in a forced-choice evaluation, where abstentions were replaced by the most common tag for the each situation (the combination system is that one biased most heavily towards recall.)

In addition to the individual features, we also list ‘pseudo-French’ and ‘pseudo-Czech’. These represent exact-match accuracies for composite features comprising those features typically realized in French or Czech POS taggers. For example, pseudo-Czech verb accuracy of 67.1% indicates that for 67.1% of verb instances, the Czech-realized features of number, tense, perfectivity, progressivity, polarity, and voice were *all* correct. These give an indication of the quality of the starting point for crosslingual bootstrapping to the respective languages.

Besides the forced-choice scenario, we were also interested in the effect of allowing abstentions for low-confidence cases. Table 3 shows the F-measure of precision and recall for the individual systems, as well as a range of combination systems. Figures 4 and 5 show (for two example features) the clear precision/recall tradeoff. Performance of the consensus systems is higher than the individual parser-based taggers at all levels of tag precision or recall.

Unfortunately, because MiniPar does its own integrated tokenization and part-of-speech tagging, we found that a significant portion of the errors seemed to stem from discrepancies where MiniPar disagreed on the segmentation or the core part-of-speech of the words in question.

## 6 Cross-lingual POS Tag Projection and Bootstrapping

Our cross-lingual POS tag projection process is similar to Yarowsky et al. (2001). It begins by performing a statistical sentence and word alignment of the bilingual corpora (described below), and then transfers both the coarse- and fine-grained tags achieved from classifier combination on the English side via the higher confidence word alignments (based on the intersection of the 1-best word alignments induced from French to English and English to French). The projected tags then serve as noisy monolingual training data in the source language.

There are several notable differences and extensions: The first major difference is that the projected fine-grained tag set is much more detailed, including such additional properties as noun case, adjective

case and number, and verb person, number, voice, and polarity. Because these span the subtag features normally assumed for Czech and French part-of-speech taggers, the projection work presented here for the first time shows the translingual projection and induction of full-granularity Czech and French taggers, rather than the much less complete and coarser-grained prior projection work.

The other major differences are in the method of target-language monolingual tagger generalization from the projected tags. We pursue a combination of trie-based lexical prior models and local-agreement-based context models. The lexical prior trie model, as illustrated in Figure 6 for noun number, shows how the hierarchically-smoothed lexical prior conditioned on variable length suffixes can assign noun number probabilities to both previously seen words (with full-word-length suffixes stored) and to new words in test data, based on backoff to partially matching suffixes.

The context models are based on exploiting agreement phenomena of the fine-grained tag features in local context.  $P(\text{subtag}|\text{context})$  for each word token is a distance-weighted linear interpolation of the posterior tag distributions assigned to its neighbors by the trie-based lexical-prior model. Finally  $P(\text{subtag}|\text{word})$  is an equally-weighted linear interpolation of the  $P(\text{subtag}|\text{affix})$  trie model probability and  $P(\text{subtag}|\text{context})$  context-agreement probability. Table 4 contrasts the performance of these two models in isolation and combination.

All of these models condition their probabilities first on the core part-of-speech of a word. We used the methods of Yarowsky et al. (2001) to develop a core part-of-speech tagger for French, based only on the projected core tags, and used this as a basis for fine-grained tags. We also ran experiments isolating the question of fine-grained tagging, assuming as input externally supplied core tags from the gold-standard data. Table 4 shows results under both of these assumptions.

For French, the training data was 15 million words from the Canadian Hansards. Word alignments were produced using GIZA++ (Och and Ney, 2000) set to produce a maximum of one English word link for each French word (i.e., a French-to-English model). The test data was 111,000 words of text from the Laboratoire de Recherche Appliquée en Linguistique Informatique at the Université de Montréal, annotated with person, number, and tense.

Suffix	Pr(PLURAL suffix)	Pr(SINGULAR suffix)
<i>none</i>	32.5	67.5
-s	66.5	33.5
-is	35.3	64.7
-ais	16.2	83.8

Figure 6: Example smoothed suffix trie probabilities for French noun number

Several factors contributed to a fairly successful set of results. The quality of the alignments is subjectively very good; the morphological system of French is relatively simple, and is a good match for our suffix tries; Perhaps most importantly, the mappings between the English and the French tagsets were for the most part simple and consistent. The most prominent exception is verb tense.

For Czech, the training and testing data were from the Reader’s Digest corpus. We used the first 63,000 words for testing, and the remaining 551,000 for training, ignoring the translations of the test data and the gold-standard tags on the training data.

It should be noted that the *baseline* (most likely tag) performance is actually a supervised model using the target language monolingual goldstandard data frequencies. The other results based on translingual projection have no knowledge of the true most likely tag, and hence occasionally underperform this supervised “baseline”. Finally, one of the major reasons for lower Czech performance is the currently very poor quality of the bilingual word alignments. However, using these diverse POS subtags as features offers the potential for substantially improved word alignment for morphologically rich languages, one of the central downstream benefits of this research.

## 7 Conclusion

We have demonstrated the feasibility of automatically annotating English text with morphosyntactic information at a much finer POS tag granularity than in the standard Brown/Penn tagset, but at a POS detail appropriate for tagging morphologically richer language such as Czech or French. This is accomplished by using a classifier combination strategy to integrate the analyses of four independent parsers, achieving a consensus tagging with higher accuracy than the best component parser.

Furthermore, we have demonstrated that the resulting fine-grained POS tags can be successfully

Feature	Engl. Comb.	Baseline	Trie	Vic.	Comb.
<b>French (using correct core POS)</b>					
JJ-num	1:0	67.0	97.6	98.0	<b>98.2</b>
	2:0	67.0	97.6	98.0	<b>98.2</b>
NN-num	1:0	71.2	94.3	<b>94.7</b>	94.6
	2:0	71.2	94.3	<b>94.7</b>	94.6
VB-num	1:0	53.4	<b>91.9</b>	73.2	90.2
	2:0	53.4	73.1	72.7	<b>73.2</b>
VB-prsn	1:0	<b>88.0</b>	76.9	78.7	77.7
	2:0	88.0	92.9	93.0	<b>93.4</b>
VB-tns	1:0	47.6	<b>86.2</b>	71.7	73.9
	2:0	47.6	<b>54.7</b>	51.9	53.8
VB-exact	1:0	26.8	<b>48.1</b>	43.4	47.1
	2:0	26.8	<b>50.0</b>	46.9	49.2
overall-exact	1:0	56.2	<b>79.7</b>	78.5	79.6
	2:0	56.2	80.3	79.6	<b>80.3</b>
<b>French (induced core POS)</b>					
JJ-num	1:0	65.1	87.1	<b>89.0</b>	88.3
	2:0	65.1	87.1	<b>89.1</b>	88.5
NN-num	1:0	66.6	87.5	87.8	<b>87.9</b>
	2:0	66.6	<b>87.5</b>	87.8	<b>87.9</b>
VB-num	1:0	53.0	<b>86.4</b>	79.5	84.9
	2:0	53.0	71.2	70.6	<b>71.4</b>
VB-prsn	1:0	<b>75.1</b>	67.4	69.7	68.4
	2:0	75.1	80.4	80.8	<b>81.1</b>
VB-tns	1:0	43.3	<b>65.1</b>	62.0	64.2
	2:0	43.3	<b>49.0</b>	46.3	48.2
VB-exact	1:0	24.1	<b>43.9</b>	40.2	43.0
	2:0	24.1	<b>45.3</b>	42.2	44.6
overall-exact	1:0	52.6	73.3	72.5	<b>73.4</b>
	2:0	52.6	73.7	73.1	<b>73.9</b>
<b>Czech (using correct core POS)</b>					
JJ-num	1:0	28.0	<b>46.4</b>	44.5	45.1
	2:0	28.0	<b>47.0</b>	44.6	46.0
JJ-case	1:0	7.1	40.2	<b>42.0</b>	40.9
	2:0	7.1	37.9	<b>41.4</b>	40.2
JJ-deg	1:0	89.2	85.6	<b>86.8</b>	86.6
	2:0	89.2	85.6	<b>86.8</b>	86.6
JJ-exact	1:0	6.9	<b>20.6</b>	19.1	19.4
	2:0	6.9	<b>20.9</b>	20.0	20.5
NN-num	1:0	52.2	<b>71.1</b>	69.6	70.7
	2:0	52.2	<b>71.1</b>	69.4	70.8
NN-case	1:0	<b>53.5</b>	39.5	39.2	39.6
	2:0	<b>53.5</b>	39.2	38.6	39.1
NN-exact	1:0	23.7	<b>29.5</b>	28.7	29.4
	2:0	23.7	<b>29.7</b>	28.6	29.4
VB-num	1:0	57.0	<b>71.6</b>	69.1	70.7
	2:0	57.0	71.2	69.7	<b>71.4</b>
VB-prsn	1:0	55.1	<b>65.9</b>	64.9	65.4
	2:0	55.1	<b>65.3</b>	64.3	64.9
VB-voice	1:0	<b>97.3</b>	93.2	93.9	93.4
	2:0	<b>97.3</b>	93.2	93.9	93.4
VB-pol	1:0	91.1	<b>93.8</b>	89.9	92.1
	2:0	91.1	<b>93.8</b>	89.9	92.1
VB-exact	1:0	9.9	<b>15.2</b>	14.6	14.8
	2:0	9.9	14.5	14.3	<b>14.7</b>
overall-exact	1:0	15.7	<b>22.6</b>	21.8	22.2
	2:0	15.7	<b>22.5</b>	21.7	22.3

Table 4: Accuracy of induced fine-grained taggers, by core part-of-speech, feature, underlying english tagger combination (*eng-comb.*), and french tagging method (most likely tag – *baseline*, suffix trie (prefix trie for Czech verb polarity) – *trie*, vicinity voting – *vic.*, or trie/vicinity combination – *comb.*)

projected to additional languages such as French and Czech, generating stand-alone taggers capturing the salient fine-grained POS subtag distinctions appropriate for these languages, including features such as adjective number and case that are not morphologically marked in the original English.

## References

- S. Bangalore and A. K. Joshi. 1999. Supertagging: An Approach to Almost Parsing. *Computational Linguistics*, 25(2): 237–265.
- T. Brants. 2000. TnT – a Statistical Part-of-Speech Tagger. In *Proceedings of ANLP-2000*, pp. 224–231.
- M. Collins. 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Ph.D. Dissertation, University of Pennsylvania, 1999.
- S. Cucerzan and D. Yarowsky. 2003. *Minimally Supervised Induction of Grammatical Gender*. In *Proceedings of HLT/NAACL-2003*, pp. 40–47.
- J. Hajič and B. Hladk’ a. 1998. Tagging inflective languages: prediction of morphological categories for a rich, structured tagset. In *Proceedings of COLING-ACL Conference*, pp. 483–490.
- R. Hwa, P. Resnik, and A. Weinberg. 2002. Breaking the Resource Bottleneck for Multilingual Parsing. In *Proceedings of LREC-2002*.
- D. Lin. 1993. Principle-based parsing without overgeneration. In *Proceedings of ACL-93*, pp. 112–120.
- L. Màrquez and H. Rodr’iguez. 1998. Part-of-speech tagging using decision trees. In *Proceedings of the European Conference on Machine Learning*.
- M. Murata, Q. Ma, and H. Isahara. 2001. Part of Speech Tagging in Thai Language Using Support Vector Machine. In *Proceedings of NLPNN-2001*, pp. 24–30.
- G. Ngai and R. Florian. 2001. Transformation-based Learning in the Fast Lane. In *Proceedings of NAACL-2001*, pp. 40–47.
- F. J. Och and H. Ney. 2000. Improved statistical alignment models. In *Proceedings of ACL-2000*, pp. 440–447.
- K. Probst. 2003. Using ‘smart’ bilingual projection to feature-tag a monolingual dictionary. In *Proceedings of CoNLL-2003*, pp. 103–110.
- D. Sleator and D. Temperley. 1993. Parsing English with a Link Grammar. In *Proceedings, Third International Workshop on Parsing Technologies*, pp. 277–292.
- D. Smith and N. Smith. 2004. Bilingual Parsing with Factored Estimation: Using English to Parse Korean. In *Proceedings of EMNLP-2004*, pp. 49–56.
- D. Yarowsky, G. Ngai, and R. Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of HLT-2001, First International Conference on Human Language Technology Research*, pp. 161–168.