An evaluation of three POS taggers for the tagging of the Tswana Learner English Corpus

Bertus van Rooy & Lande Schäfer
Potchefstroom University
Private Bag X6001, Potchefstroom
2520, South Africa

Phone: +27-18-299 1487/1504 Fax: +27-18-299 1562
E-mail: ntlajvr@puknet.puk.ac.za (Van Rooy) ntlyvs@puknet.puk.ac.za (Schäfer)

1. Introduction

Before starting with part of speech (POS) tagging on our corpus of learner English we decided to evaluate three POS taggers to see which one gives the best results when tagging written second language English. We evaluated the taggers' performance to determine which tagger would be most suitable for linguistic analyses on a POS-tagged corpus that had not been tag-edited. Once the accuracy of the taggers had been determined, we investigated the factors that contributed to inaccuracy with a view to establish time and cost effective ways of increasing tagger accuracy without necessarily tag-editing the corpus from beginning to end. The aim of this research was to explore the possibility of selective tag editing based upon specific tokens or tags frequently associated with tagging errors.

2. The Tswana Learner English Corpus project and the POS taggers

At the end of 2000 we started compiling the Tswana Learner English Corpus (TLEC). Setswana is spoken as a mother tongue in the North West and Northern Cape Provinces of South Africa and in Botswana. Most of the South African speakers of Setswana learn English as a second language in school. When it is complete, this 200 000-word corpus will form part of the International Corpus of Learner English (ICLE).

Currently there is significant movement towards the fostering of a local human language technology industry in South Africa. Our aim is to collect and POS tag more corpora of South African varieties of English and of the other ten indigenous languages. As TLEC will be the first POS tagged corpus of a South African variety of English, we wanted to establish procedures for time and cost-effective POS tag-editing.

Last year we started with POS tagging. As TLEC will form part of ICLE, it needs to be tagged by TOSCA-ICLE. TOSCA-ICLE has a very large tagset and undertakes complex subcategorisation of the main wordclasses. We were interested to see how TOSCA-ICLE would perform on our data and how its performance would compare to the performance of taggers with less complex subcategorisation withint its tagset. We decided to also evaluate the Brill-tagger (Brill 1999) and CLAWS7 (see Garside & Smith 1997 for discussion of an earlier version of CLAWS).

3. Method of evaluation of taggers

We based our evaluation of the taggers on the four considerations for the selection of a tagger that are identified by Van Halteren (1999b:95-98):

- the tagset,
- documentation,
- the tagging process, and
- performance.

The Brill-tagger makes use of the Penn Treebank Tagset, which is quite a small tagset consisting of only 36 tags. The CLAWS7 tagset has 137 tags, excluding the punctuation tags. TOSCA-ICLE has a possible 220 tags that it can assign, due to the fact that it allows for subcategorisation on various levels. Thus, the tagset as criterion concerns the types of contrasts that the tagset identifies. We have taken the limitations of each tagset into account in our judgment of tag errors. Documentation refers to whether all the relevant documentation about the tagset, the tagging process and the types of assumptions underlying the various operations are available, to facilitate correct interpretation of results. We easily obtained documentation regarding all three taggers via the Internet, and they all contained the relevant information. Considerations regarding the tagging process includes the time it takes, the user interface it provides and the platform on which it operates. Performance refers to how accurate a tagger is at assigning POS-tags to corpus data.

We focused our evaluation on measuring the taggers' performance. Van Halteren (1999a) proposes a number of measures for this. The most basic measure is overall tagger accuracy, which is calculated by dividing the number of correct tags into the total number of tags. Precision measures how many tokens that received a tag Y received that tag correctly, thus p*recision* EQUALS number of tokens tagged Y correctly DIVIDED BY total number of tokens tagged X. Recall measures how many tokens that should be tagged Y are indeed tagged Y. Thus, *recall* EQUALS number of tokens tagged Y correctly DIVIDED BY total number of tokens that must have been tagged Y.

We randomly selected 13 essays (each approximately 400 words) from the corpus which provided us with a 5000 word sample (a total of 5618 tokens, including punctuation). The sample was tagged with all three taggers. The Brill and TOSCA-ICLE tagging were done in the language technology laboratory in Potchefstroom on Linux and DOS respectively. The CLAWS7 tagging was done by UCREL.

We followed a comparative method in tag-editing. We entered the data into a Microsoft Excel file. Each tagger's data was represented in three columns. The first column contained the tokens, the second the tag assigned to that token, and the third the tag correction if there was a correction. We placed the data of the three taggers next to each other and aligned the data according to the tokens. The authors of this paper undertook the editing together in order to validate each other's interpretation of the learner's sentences as well as each other's judgement of the correctness of the tags. Being able to see the tags assigned by all three taggers sometimes helped us in deciding whether we would accept a certain POS-tag as correct or not.

After editing the tags we calculated the accuracy of the taggers to determine whether it would be possible to undertake linguistic analyses and the accompanying POS-based searches on the unedited tagged corpus. We wanted to establish which tagger would give us the most peace of mind with searches prior to tag-editing, and whether some searches could take place without tag-editing having taken place.

4. Results of evaluation of taggers

The overall accuracy for all three taggers are presented in Table 1. In terms of performance, CLAWS7 fared the best, while the TOSCA-ICLE tagger fares substantially poorer than the 05% accuracy reported by De Haan (2000) for the same tagger on other varieties of learner English. It would be possible to undertake POS-based searches for specific grammatical constructions on a corpus that has been tagged with CLAWS7 but not tag-edited. However, the ideal is to work with a corpus with POS tags that are as accurate as possible, and the next step was to find ways to improve tagger accuracy by way of tag-editing, as explained in section 5.

Table 1 Overall accuracy

	Accuracy
Brill	86.34%
CLAWS7	96.26%
TOSCA-ICLE	88.04%

We have made the decision to tag TLEC with both TOSCA-ICLE and CLAWS, and we therefore set out to examine ways of improving their accuracy. Because we don't rule out the possibility of using the Brill-tagger on data in future, we also looked at ways in which its accuracy could be improved. As we have already tagged most of our corpus with TOSCA-ICLE, we will begin by editing the TOSCA-ICLE tags first, and therefore tagging procedures for this tagger will be discussed in more detail in section 6.

5. Method of improving tagger accuracy

The time and effort that it took to tag-edit the three tagged versions of our 5000-word sample indicated to us that it would not be viable to tag-edit the whole corpus (and also other corpora) in this manner. We decided to carry out selective editing, by establishing which tags were frequently assigned incorrectly and which tokens were frequently tagged incorrectly. One would then be able to search for problematic tags or tokens to check whether they need correcting or not.

After editing the tags, the precision and recall rates for each tag was calculated. From this it was possible to identify tags which had a particularly low precision or recall. Table 2 represents the three tags with the lowest precision for each tagger. For the sake of interest the recall rates for these tags are also given. The data in Table 2 indicates that it is not enough to simply consider the precision rates for the individual tags. The frequency with which the tag occurs in the corpus should also be taken into account. It would, for instance, be worthwhile to do a search for all CD tags in the Brill-tagged sample and to systematically correct all of them, because they constitute 5.59% of all tags, whereas it would not be worth the effort to search for all PRON(cleft) tags to ensure that they are correct, since they only constitute 0.02% of all the tags.

Table 2: Tags with the lowest precision for each tagger

		Frequency	Precision	Recall
Brill-tagger	CD	5.59%	13.31%	95.35%
	RBS	0.16%	44.40%	100%
	NNPS	0.09%	60.00%	100%
CLAWS7	RGR	0.08%	50.00%	100%
	RRR	0.24%	50.00%	100%
	DDQ	0.61%	58.06%	100%
TOSCA-ICLE	PRON(cleft)	0.02%	0.00%	n/a
	PROFM(so,clause)	0.02%	0.00%	n/a
	VB(aux,modal,infin)	0.04%	0.00%	n/a

The question then is not "What are the tags with the lowest accuracy?", but "What are the tags that contribute most significantly to the overall error rate (the opposite of the overall accuracy)?". We calculated each tag's contribution to the overall error rate of the tagger by dividing the number of times a specific tag was assigned incorrectly by the total number of tag errors. Tables 3, 4 and 5 shows which tags contributed most significantly to the tagging errors made by the various taggers.

Table 3 Contribution to errors (Brill-tagger)

Brill tag	Percentage of total tag errors
VB	19.55%
NN	16.46%
IN	14.81%
VBP	11.11%

Table 4 Contribution to errors (CLAWS7)

CLAWS tag	Percentage of total tag errors
NN1	24.87%
JJ	10.52%
VV0	10.52%
ND1	7.94%

Table 5 Contribution to errors (TOSCA-ICLE)

TOSCA-ICLE tag	Percentage of total tag errors
ADJ(ge,pos)	12.23%
VB(lex,montr,infin)	8.10%
ADV(ge,pos)	6.27%
CONJUNC(subord)	5.50%
N(sing)	5.50%
PREP(ge)	5.20%
N(plu)	3.82%
VB(lex,montr,imp)	3.82%

Editing only the tags contained in Table 3 would eliminate more than 60% of the tagging errors made by the Brill-tagger. Similarly, editing the tags listed in Table 4 would eliminate more than half of the tagging errors made by CLAWS7. Editing only the tags in Table 5 would also result in the elimination of more than half of the errors in the sample tagged by TOSCA-ICLE.

Once we had established which tags should be submitted to selective manual editing by considering each tag's contribution to the overall error rate, we were faced with the question of developing procedures for selective editing based on the data contained in the 5000-word sample. One very important factor that we had to take into account was the availability of people who could do manual tag editing. Our team consists of two people who have a sound knowledge of parts of speech and who are familiar with the full tagsets of the taggers we used (henceforth referred to as full editors), and four people who have only a rudimentary knowledge of parts of speech and no knowledge of the tagsets (henceforth referred to as mini editors). Selective tag-editing meant that we would only need to provide selective training which is time-effective. In devising a work plan, we had to decide which tags could be successfully dealt with by mini editors, and which tags should be left to the full editors. This entailed a closer look at tagging results for each tagger.

6. Selective tag editing: procedures for TOSCA-ICLE

The aim was to provide each mini editor with tag-specific training. For instance a mini editor could be informed about the ADJ(ge,pos) class, which is the tag that made the greatest contribution to the tagging errors made by TOSCA-ICLE (see Table 5). The editor would also be informed that the ADJ(ge,pos) tag is often assigned to words that should have been tagged N(sing). Table 6 shows that in our data more than one fifth of the ADJ(ge,pos) tags, should have been N(sing) tags. The mini editor would be instructed about the identification of ADJ(ge,pos) and N(sing). This would also entail general instruction about the attributes of nouns and adjectives.

Table 6 Assignment of the ADJ(ge,pos) tag by TOSCA-ICLE

Correct	Number of times ADJ(ge,pos) was assigned
ADJ(ge,pos)	198
N(sing)	60
ADV(ge,pos)	7
PREP(ge)	4
NADJ(ge,pos)	1
PRON(quant)	1
PRON(pers,plu)	1
PRON(pers,sing)	1
VB(lex,*,*)	2

The N(sing) tag, in turn, contributes to 5.5% of the errors (see Table 5). However, most of the N(sing) tag errors were made on lexical verbs (see Table 7). One could instruct the mini-editor working with N(sing) and ADJ(ge,pos) tags to check all the N(sing) tags and simply flag instances where the tag was not assigned correctly, except of course where the confusion is clearly with ADJ(ge,pos), since this mini editor already knows the identification criteria for ADJ(ge,pos) and can therefore correct the tag.

Table 7 Assignment of the N(sing) tag by TOSCA-ICLE

Correct	Number of times N(sing) was assigned
N(sing)	605
ADJ(ge,pos)	6
ADJ(ge,pos,ingp)	1
ADV(wh)	1
VB(aux,semip,pres)	1
VB(lex,ditr,pres)	2
VB(lex,intr,infin)	1
VB(lex,intr,ingp)	4
VB(lex,intr,pres)	3
VB(lex,montr,infin)	1
VB(lex,montr,ingp)	4
VB(lex,montr,pres)	7

The same mini-editor could also edit the N(plu) tag, which contributed to 5.2% of the tagging errors as s/he would be familiar with the properties of nouns. Table 8 shows that N(plu) was mostly assigned incorrectly to N(sing). In adopting this procedure, one mini-editor can correct a substantial portion of the tag errors while only having to be trained on three tags. This should hopefully ensure fair accuracy in his/her work, since there is little room for mistakes to be made. At the same time, potentially difficult cases are flagged. The full editors can retrieve these cases much more easily, as they do not need to work through all instances of these three tags, and make the necessary corrections.

Table 8 Assignment of the N(plu) tag by TOSCA-ICLE

Correct	Number of times N(plu) was assigned
N(plu)	369
N(sing)	27
VB(lex,montr,pres)	6

Errors pertaining to lexical verbs contributed significantly to the overall error rate (see Table 9). Therefore it would be viable to train one mini editor on the category of lexical verbs. The incorrect assignment of lexical verbs can sometimes be ascribed to confusion with another main word class, such as nouns, but most of the VB(lex,…) errors are due to confusion pertaining to transitivity (see table 10) and confusion pertaining to the form of the verb (see Table 11).

Table 9 Contribution to error rate by VB(lex,…) tags

TOSCA-ICLE lexical verb tag	Contribution to error rate
VB(lex,montr,infin)	8.10%
VB(lex,montr,imp)	3.52%
VB(lex,intr,infin)	3.37%
VB(lex,montr,edp)	2.91%
VB(lex,montr,pres)	7.99%
VB(lex,cop,pres)	1.68%
VB(lex,dirt,infin)	1.53%
VB(lex,intr,imp)	1.53%
VB(lex,intr,subjun)	1.53%
VB(lex,(intr,pres)	1.38%
VB(lex,montr,past)	1.07%

Table 10 Confusion matrix for transitivity features of the lexical verb

	Number of times tag was assigned				
Correct ↓	VB(lex, cxtr,…)	VB(lex, dimontr,…)	VB(lex, ditr,…)	VB(lex, intr,…)	VB(lex, montr,…)
VB(lex,cxtr,…)	10	0	0	0	2
VB(lex,dimontr,…)	0	0	1	0	0
VB(lex,ditr,…)	0	0	24	1	11
VB(lex, intr, …)	1	0	0	166	14
VB(lex,montr,…)	3	0	8	7	370

A closer look at the data in table 10 indicates that the two more frequent categories, monotransitive and intransitive verb, are the ones causing most of the problems although their precision is more than 90% in both cases, while the precision of the tag for ditransitives is less than 70%. The confusion caused by the ditransitive tag is mostly with the monotransitive tag. If one starts the mini editor on the tag for ditransitives, and also train him/her on monotransitives, then proceed to editing for monostransitives, which should include training for intransitives, before finally dealing with the intransitives themselves, a fair result should be obtained.

Table 11 Confusion matrix for forms of the lexical verb

	Number of times tag was assigned						
Correct ↓	edp	imp	infin	ingp	past	pres	subjun
edp	84	0	0	0	0	0	0
imp	0	3	0	0	0	0	0
infin	0	12	263	0	0	3	0
ingp	0	0	0	79	0	0	0
past	3	0	0	0	31	0	1
pres	0	14	41	0	0	217	2
subjun	0	0	0	0	0	0	0

We believe that the formal features of the verb are fairly easy to understand for our mini editors, and therefore propose to provide training on these when the first training on the transitivity features is given. No separate editing for the formal features will be done, rather, these should be inspected and corrected at the same time as the transitivity features. What should be clear from table 11 is that the tag for imperatives is assigned with a precision of about 10% and the tag for infinitives with a precision of about 87%. Very often, the tokens should rather have been tagged as present tense forms of the verb, although it also happens frequently that the tag imperative is assigned to infinitives. One should therefore concentrate on disambiguation criteria for these three tags in particular during training of the mini editor.

As is clear from table 5, the tag responsible for the third most errors is the tag for general adverbs. What makes this tag more difficult to handle, is that there is no single category with which it is often confused. All the possible correct tags of tokens tagged as ADV(ge,pos) are presented in table 12.

Table 12 Assignment of the ADV(ge,pos) tag in TOSCA-ICLE

	ADV(ge,pos)
ADV(ge,pos)	161
ADJ(ge,pos)	2
ADV(connec)	4
ADV(phras)	3
ART(def)	1
ART(indef)	1
CONJUNC(subord)	3
EXTHERE	2
N(sing)	4
PREP(ge)	13
PRON(dem,sing)	2
PRON(quant)	3
PRON(univ)	1
VB(lex,montr,pres)	1

The only tag that is confused with ADV(ge,pos) in a substantial number of cases is the tag PREP(ge), general prepositions. Since there is a limited number of lexical items, such as 'out' and 'behind' that are potentially ambiguous between PREP(ge) and ADV(ge,pos), it is feasible to train mini editors to distinguish between these two. We believe that it is similarly possible to train mini editors to resolve confusion between ADV(ge,pos) and EXTHERE, the tag for existential 'there', since it is limited to a single lexical item. For the remainder, we propose to focus the training of the mini editor on good identification criteria for the tag ADV(ge,pos) itself, and instruct him/her to flag all cases that are not clearly correct or clear confusions with PREP(ge) or EXTHERE. If one happens to employ a mini editor who has already worked on another tag, and therefore knows how to identify that, he/she can be asked to also resolve confusions with that tag, e.g. the mini editor who has already handled N(sing).

The fifth biggest contributor to tag errors is the tag for subordinating conjunctions CONJUNC(subord). As is clear from table 13, there is no single tag that is consistently confused with this tag, similar to the situation with ADV(ge,pos).

Table 13 Assignment of the CONJUNC(subord) tag in TOSCA-ICLE

	CONJUNC(subord)
CONJUNC(subord)	153
ADV(connec)	1
ADV(ge,pos)	2
ADV(wh)	5
PREP(ge)	5
PRON(dem,sing)	10
PRON(inter)	1
PRON(rel)	6

However, there are a couple of lexical items that are subject to some confusion. It might therefore be worthwhile to rather concentrate on these lexical items. These are the tokens ‘so’ and ’that’. Let us look at confusion tables for individual tokens to come to an understanding of the type of confusion to see how we can use this information to improve the precision of the tag ADV(ge,pos).

Table 14 Confusion table for the token ‘so’

		Proposed tag				
		ADV(connec)	CONJUNC (subord)	CONJUNC (subord):1/2	ADV(ge,pos)	PTCL(to):1/3
Correct tag	ADV(connec)	3	0	0	0	0
	CONJUNC(subord)	0	5	0	3	0
	CONJUNC(subord):1/2	0	0	11	0	0
	ADV(ge,pos)	0	2	0	5	0
	PTCL(to):1/3	0	0	0	0	0

The only confusion that is evident from this confusion table is the confusion between the tags CONJUNC(subord) and ADV(ge,pos). If one reconsiders tables 12 and 13, it is clear that the confusion involving these two tags is related to the token ‘so’ in all cases. It would therefore be worthwhile to instruct a mini editor to deal with the tagging of the token ‘so’ before turning to either CONJUNC(subord) or ADV(ge,pos).

Table 15 Confusion table for the token ‘that’

		Proposed tag			
		CONJUNC (subord)	CONJUNC (subord):2/2	PRON(dem,sing)	PRON(rel)
Correct tag	CONJUNC(subord)	39	0	0	0
	CONJUNC(subord):2/2	0	11	0	0
	PRON(dem,sing)	10	0	7	2
	PRON(rel)	6	0	3	16

From this table, it is clear that all instances involving confusion between the tags CONJUNC(subord) and PRON(dem,sing) in table 13 involve the token ‘that’. Additionally, this token is also responsible for all confusions with the tag PRON(rel) in table 13. Thus, instructing a mini editor to inspect and correct all instances of the token ‘that’ will contribute significantly to the overall precision of the tag CONJUNC(subord) and a number of other tags too. Correcting the tags for the tokens ‘so’ and ‘that’ will lead to the correction of 60% of tag errors involving the tag CONJUNC(subord), involving much less trouble than dealing with all tokens tagged as CONJUNC(subord).

As a last example, we turn our attention to the seventh tag on the list in table 5, the tag for general prepositions, PREP(ge). The results in table 16 indicate that there is one tag responsible for almost half of the errors involving this tag, which is ADV(connec).

Table 16 Assignment of the PREP(ge) tag in TOSCA-ICLE

Correct tag	Number of times tagged PREP(ge)
PREP(ge)	104
ADV(connec)	14
ADV(ge,pos)	2
ADV(phras)	2
CONJUNC(coord)	1
CONJUNC(subord)	4
N(sing)	1
PREP(phras)	1
PRTCL(to)	2
VB(lex,montr,infin)	1
VB(lex,montr,pres)	3

It would therefore be possible to instruct a mini editor to examine all occurrences of the tag PREP(ge) and inspect them for confusion with ADV(connec). However, to understand the meaning of the tag ADV(connec), it is also necessary to understand the meaning of the related tags for subordinating and coordinating conjunctions. However, close inspection of the confusion table for the token 'like' indicates that this token is responsible for 13 of the 14 errors involving the confusion between PREP(ge) and ADV(connec). It therefore seems far more efficient to instruct a mini editor to deal with the token 'like'. It has the further gain that all instances of confusion between PREP(ge) and PREP(phras), as well as VB(lex,montr,pres) also involve 'like'. Thus 17 of the 31 errors involving the tag PREP(ge) can be corrected by only inspecting and correcting the errors involving the token 'like'.

The preceding discussion in section 6 has hopefully demonstrated how significant improvement to overall tagger accuracy can be achieved by training mini editors on a selected number of tags and in some cases on the possible tags for a selected number of tokens. More than half of all tag errors (50,9%) can be removed following the procedure above, while examining far less than half of all the tokens in the corpus.

Once this has been done, the full editors can work through the corpus on the basis of a different list, the one presented in table 2. Working through the tags in the order of least precision means that more substantial improvement in tag accuracy is effected per number of tags inspected. For instances, a first step could be to inspect and correct all tags with an accuracy of less than 80%. These tags are listed in table 17.

Table 17 Tags not yet corrected with precision of less than 80%

POS Tag	Frequency	Correct	Precision
PROFM(so,clause)	1	0	0.0000%
PRON(cleft)	1	0	0.0000%
VB(aux,modal,infin)	2	0	0.0000%
VB(aux,modal,pres,encl)	3	0	0.0000%
VB(aux,pass,past,neg)	1	0	0.0000%
VB(aux,semi,pres)	5	0	0.0000%
VB(lex,cxtr,past)	1	0	0.0000%
ADJ(ge,pos,ingp)	10	1	10.0000%
VB(aux,perf,pres)	21	4	19.0476%
ADV(ge,sup)	4	1	25.0000%
PRON(antit)	3	1	33.3333%
VB(aux,perf,infin)	9	3	33.3333%
VB(aux,semi,infin)	3	1	33.3333%
VB(lex,cxtr,edp)	3	1	33.3333%
VB(lex,cxtr,infin)	10	4	40.0000%
VB(lex,montr,past)	14	7	50.0000%
ADV(ge,comp)	11	6	54.5455%
PRON(inter)	33	18	54.5455%
VB(aux,pass,pres)	29	19	65.5172%
NUM(ord,sing)	3	2	66.6667%
ADJ(ge,pos,edp)	19	13	68.4211%
VB(aux,do,pres)	28	21	75.0000%
VB(lex,cop,ingp)	8	6	75.0000%
VB(aux,pass,past)	9	7	77.7778%

If a full editor were to inspect and correct the 231 tokens in table 17, he/she will encounter 115 errors. Thus a further 17,6% of the errors can be removed if a full editor is tasked with correcting the tags in table 17. Combining the two procedures leads to the removal of 68,5% of all tag errors, which is, we believe, a very good yield for relatively little input. If this remains valid for the entire corpus, we can achieve a tagging accuracy of 96,2%. Additioally, we will have an accurate record of which tags have been corrected and which haven't. For most of our purposes, this is sufficient. If a linguist or applied linguist were interested in a feature whose tags have not been corrected, assuming that that person will understand the grammar of that particular feature thoroughly, we can simply ask him/her to correct the relevant tags before commencing the study, as part of the licensing conditions for using our corpus of learner English.

9. Conclusions

This paper has presented results on the evaluation of three taggers, Brill, CLAWS7 and TOSCA-ICLE, using a non-native variety of English as testing data. Overall, CLAWS7 performs substantially better than the other two.

Since our concern is with the accuracy of the TOSCA-ICLE tagger in particular, we have proposed a method for selective tag error editing to improve the overall accuracy of the tags in the tagged corpus. This procedure involves identifying the tags that contribute the biggest proportions to the total number of errors, together with a clear indication of the tags with which a particular tag is confused. On this basis, mini editors are trained to deal with one tag or a set of related tags at a time. Following this, full editors are employed to correct all remaining tags with a precision of less than 80%. In this way, which we believe is time efficient, we manage to remove more than two thirds of all tag errors, without having to manually edit two thirds of the corpus.

10. References

Brill E 1999 Corpus-based rules. In Van Halteren H (ed) *Syntactic wordclass tagging*. Dordrecht, Kluwer, pp 247-262.

De Haan P 2000 Tagging non-native English with the TOSCA-ICLE tagger. In Mair C, Hundt M (eds) *Corpus Linguistics and Linguistic Theory.* Amsterdam, Rodopi, pp 69-79.

Garside R, Smith N 1997 A hybrid- grammatical tagger: CLAWS4. In Garside R, Leech G, McEnery A (eds) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. Essex, Addison Wesley Longman, pp 102-121.

Van Halteren H 1999a Performance of taggers. In Van Halteren H (ed) *Syntactic wordclass tagging.* Dordrecht, Kluwer, pp 81-94.

Van Halteren H 1999b Selection and operation of taggers. In Van Halteren H. (ed) *Syntactic wordclass tagging*. Dordrecht, Kluwer, pp 95-104.