A Foundational Survey on Accuracy of Taggers

Masatsugu ONO*

Abstract

This paper shows that it will be possible to study grammar based on a corpus tagged by taggers although taggers sometimes assign incorrect tags. It has been found that there were some tendencies of incorrect tags which appear when tags are assigned by taggers. As long as tagger users recognize the mistagging tendencies, taggers will be very useful for English vocabulary and grammar studies.

1. Overview

In corpus-based studies of grammar, it is necessary to prepare special corpora which have grammatical information in texts. The grammatical information can usually be represented by part-of-speech indexes called *tags* which are assigned to each word in a text. In general, tags are assigned by software called a *tagger*.

However, taggers sometimes assign the wrong tags to words. Therefore, a survey on accuracy of taggers is required on before a precise corpus-based analyses of grammar can be done.

In this paper, I will explain how taggers can correctly assign part-of-speech tags and what kind of tags tend to be wrong, and I will also present what kinds of grammatical studies are possible based on the accuracy of taggers.

2. Tags and Taggers

Tags play important roles. One of the most important roles of tags is to identify rich and complex grammatical structures and functions explicitly, which are invisible in normal texts. In order to identify all of the structures and functions as accurately as possible, many kinds of tags exist in one tagset. Currently, there are four tagsets, as shown in Table 1.

Taggers are indispensable in corpus studies of grammar. It is possible to assign tags by hand, but this is not feasible when we are dealing with huge texts. Therefore, when tagging we often use software called *tagger*. However, tags assigned by taggers are not always correct because taggers assign tags based on a probabilistic algorithm. Therefore, it has been found that a corpus-based study of grammar is impossible in the absence of accurate taggers.

3. Procedures

The purpose of this survey is to show whether the grammatical study based on the corpus tagged by taggers is possible or not. I dealt with two taggers in this survey. One was Apple Pie Parser 5.9 (APP) and the other was Rule Based Tagger 1.14 (RBT).

To analyze the accuracy of these taggers, a correctly tagged corpus was required as a model. In this survey, I made use of the Brown Corpus Tagged Version as the correctly tagged corpus.

* Lecturer, Department of General Education

Table 1. Types of Tags

Types of Tags	
Types	Representations
Brown Corpus	ABL ABN ABX AP AT BE BED BEDZ BEG BEM BEN BER BEZ CC CD CS DO DOD DOZ DT DTI DTS DTX EX FW HL HV HVD HVG HVN HVZ IN JJ JJR JJS JJT MD NC NN NNS NN$ NNS$ NP NP$ NPS NPS$ NR NRS NR$ OD PN PN$ PP$ PP$$ PPL PPLS PPO PPS PPSS QL QLP RB RBR RBT RN RP TL TO UH VB VBD VBG VBN VBZ WDT WP$ WPO WPS WQL WRB , . ( ) * -¥ - :
LOB Corpus	ABL ABN ABX AP AP'' AP$ APS AT ATI BE BED BEDZ BEG BEM BEN BER BEZ CC CC'' CD CD-CD CD1 CD1$ CD1S CDS CS CS'' DO DOD DOZ DT DTI DTS DTX EX HV HVD HVG HVN HVZ IN IN'' JJ JJ'' JJB JJB'' JJR JJT JNP MD NC NN NN'' NN$ NNP NNP$ NNPS NNPS$ NNS NNS$ NNU NNU'' NNUS NP NP$ NPL NPL$ NPLS NPS
	NPS$ NPT NPT'' NPT$ NPTS NPTS$ NR NR$ NRS OD PN PN'' PN$ PP$ PP$$ PP1A PP1AS PP1O PP1OS PP2 PP3 PP3A PP3AS PPAO PP3OS PPL PPLS PPLS'' QL QLP RB RB'' RB$ RBR RBT RI RN RP TO TO'' UH VB VBD VBG VBN VBZ WDT WDTR WP WP$R WPA WPO WPOR WPR WRB XNOT ZZ &FO &FW , . ( ) ... : ; ? ! *' **' *-
British National Corpus	AJ0 AJC AJS AT0 AV0 AVP AVQ CJC CJS CJT CRD DPS DT0 DTQ EX0 ITJ NN0 NN1 NN2 NP0 ORD PNI PNP PNQ PNX POS PRF PRP PUL PUN PUQ PUR TO0 UNC VBB VBD VBG VBI VBN VBZ VDB VDD VDG VDI VDN VDZ VHB VHD VHG VHI VHN VHZ VM0 VVB VVD VVG VVI VVN VVZ XX0 ZZ0
Penn TreeBank	CC CD DT EX FW IN JJ JJR JJS LS MD NN NNS NNP NNPS PDT POS PRP PRP¥$ RB RBR RBS RP SYM TO UH VB VBD VBG VBN VBP VBZ WDT WP WP¥$ WRB -LRB- -RRB- ¥$. , '' :
Apple Pie Parser 5.9	NNPX NNX -LRB- -RRB- ¥$. , '' :
Rule Based Tagger 1.14	( ) . , :

The method of this survey is very simple. It is to compare tags assigned by the taggers with the tags in the Brown Corpus. However, it has been found that there were several steps required to merge the two tags.

First of all, it was necessary to change the tags used in the Brown Corpus into the tags of the Penn TreeBank Tagset because the former tags were different from the latter ones. The tags which APP and RBT assign follow the Penn TreeBank Tagset, but the tags in the Brown Corpus do not follow the Penn TreeBank Tagset because the Tagset of the Brown Corpus is original.

Secondly, it was indispensable to determine the levels to research for the accuracy of taggers. There are a lot of tags which represent the word forms in detail in one tagset. In this survey, I set up the two levels to analyze the accuracy of taggers: one was the word-form level and the other was the word-class level. These are illustrated in Table 2.

Table 2. Levels of Comparison

Levels of Comparison	
Level of Word Class	Level of Word Form
Adjectives	positive, comparative, superative
Adverbs	positive, comparative, superative, particles
Cardinal numbers	Cardinal numbers
Conjunctions and Propositions	Coordinatings, Subordinatings, Propositions
Determiners	Determiner
Existential *there*	Existential *there*
Foreign words	Foreign words
Interjections	Interjections
List item markers	List item markers
Modals	Modals
Nouns	singular and mass, plural, proper noun singular, proper noun plural, possessive nouns
Predeterminers	Predeterminers
Pronouns	personal, possessive
Symbols	Symbols
Infinitive *to*	Infinitive *to*
Verbs	base form, past tense, gerund or present participle, past participle, 3rd person singular present
Wh-words	wh-determiners, wh-pronouns, possessive wh-pronouns, wh-adverbs

Finally, all tags were divided into two groups with respect to agreement. I analyzed the accuracy of taggers based on a relative rate of the number of tags in each group.

4. Results and Discussion

The tags assigned by APP and RBT are in agreement with the tags in the Brown Corpus at a rate of 89.08% and 92.33% respectively on the word-form level. Similarly, the tags assigned by the taggers correspond to the tags in the Brown Corpus at a rate of 90.39% and 93.55% respectively on the word-class level. These rates can be regarded as indicative of the accuracy of the taggers.

Looking at the taggers individually, the accuracy of RBT is slightly better than that of APP. In addition, with respect to a pair of tags which is in disagreement, it has been found that there were some consistent tendencies about the wrong tags even though the taggers assigned them by mistake. The examples are as follows:

(1)

a. Most of the words starting with capital letters are tagged with NNP.

b. Normally, words which have an -*ing* form are tagged with VBG. However, sometimes they are tagged with JJ or NN.

c. Normally, words which have a past participle form are tagged with VBN. However, sometimes they are tagged with JJ or VBD.

d. Normally, numbers are tagged with CD. However, sometimes they are tagged with JJ and NN.

e. Complex words such as hyphenated words tend to be mistagged.

f. Function words tend to be mistagged. For example, "that", "what", "to" etc.

The ten most frequent tag-merging patterns and the ten most frequent words when tags were assigned by mistake are illustrated in Table 3. to 10. And the following tagged sentences show the rate at which a tag in APP and RBT was in agreement with a corresponding tag in the Brown Corpus in each genre.

Table 3. Frequency of merging patterns in mistagging between BRN and APP (Word Form)

Results	
Merging Patterns (BRN — APP)	Frequency
NN — NNP	15324
IN — TO	11163
VBN — VBD	6455
JJ — NNP	5183
IN — RB	4601
JJ — NN	3931
VBG — NN	3706
NN — JJ	3508
DT — RB	3084
VB — NN	3051

Table 4. Frequency of merging patterns in mistagging between BRN and RBT (Word Form)

Results	
Merging Patterns (BRN — RBT)	Frequency
NN — NNP	14738
IN — TO	11163
JJ — NNP	5008
RP — IN	4355
PDT — DT	2939
JJ — NN	2900
. — :	2808
NN — JJ	2730
NNP+POS — NNP	2538
) — SYM	2429

Table 5. Frequency of merging patterns in mistagging between BRN and APP (Word Class)

Results	
Merging Patterns (BRN – APP)	Frequency
IN – TO	11163
JJ – NN	9667
VB – NN	8640
IN – RB	4723
VB – JJ	4259
NN – JJ	3809
RB – JJ	3093
DT – RB	3084
JJ – RB	2989
NN – VB	2902

Table 6. Frequency of merging patterns in mistagging between BRN and RBT (Word Class)

Results	
Merging Patterns (BRN – RBT)	Frequency
IN – TO	11163
JJ – NN	7941
VB – NN	6206
RP – IN	4355
NN – JJ	3239
PDT – DT	2939
RB – JJ	2918
. – :	2808
NN – VB	2520
) – SYM	2429

Table 7. Frequency of words in mistagging between BRN and APP (Word Form)

Results		
Words	Merging Patterns (BRN – APP)	Frequency
to	IN – TO	11139
all	PDT – RB	2756
;	. – JJ	1988
what	WDT – WP	1900
that	IN – DT	1743
J	NN – NNP	1645
some	DT – RB	1594
any	DT – RB	1333
about	IN – RB	1238
that	WP – IN	1006

Table 8. Frequency of words in mistagging between BRN and RBT (Word Form)

Results		
Words	Merging Patterns (BRN - RBT)	Frequency
to	IN - TO	11139
;	. - :	2781
)	) - SYM	2429
all	PDT - DT	2200
what	WDT - WP	1900
J	NN - NNP	1642
up	RP - IN	1545
that	WP - WDT	1464
out	RP - IN	1090
more	RB - RBR	933

Table 9. Frequency of words in mistagging between BRN and APP (Word Class)

Results		
Words	Merging Patterns (BRN - APP)	Frequency
to	IN- TO	11139
all	PDT - RB	2756
;	. - JJ	1988
that	IN - DT	1743
some	DT - RB	1594
any	DT - RB	1333
about	IN - RB	1238
that	WP - IN	1006
first	JJ - RB	981
up	RP - RB	912

Table 10. Frequency of words in mistagging between BRN and RBT (Word Class)

Results		
Words	Merging Patterns (BRN - RBT)	Frequency
to	IN - TO	11139
;	. - :	2781
)	) - SYM	2429
all	PDT - DT	2200
up	RP - IN	1545
out	RP - IN	1090
]	. -)	798
both	PDT - DT	731
down	RP - RB	629
New	JJ - NN	530

(2)

Tagged Brown Corpus

a.   the _AT Fulton _NP County _NN Grand _JJ Jury _textbfNN said _VBD Friday _NR an _AT investigation _NN of _IN Atlanta's _NP$ recent _JJ primary _NN election _NN produced _VBD no _AT evidence _NN that _CS any _DTI irregularities _NNS took _VBD place _NN ._.

Brown Corpus Tagged by Apple Pie Parser

b.   the/DT Fulton/NNP County/NNP Grand/NNP Jury/NNPX said/VBD Friday/NNP an/DT investigation/NN of/IN Atlanta/NNP 's/POS recent/JJ primary/JJ election/NN produced/VBD no/DT evidence/NN that/IN any/RB irregularities/NNS took/VBD place/NN -PERIOD-/.

Brown Corpus Tagged by Rule Based Tagger

c.   the/DT Fulton/NNP County/NNP Grand/NNP Jury/NNP said/VBD Friday/NNP an/DT investigation/NN of/IN Atlanta's/NNP recent/JJ primary/JJ election/NN produced/VBD no/DT evidence/NN that/IN any/DT irregularities/NNS took/VBD place/NN ./.

The results above summarize the following. NNP means singular proper noun in the definition of the Penn TreeBank Tagset. If a noun is a proper noun, it is written with an initial capital. Therefore, a tagger will tag every word beginning with a capital letter with NNP. As a matter of fact, in the tagging algorithm, if unknown words of taggers appear, taggers tag words beginning with a capital letter with NNP and tag all other words with NN.

With respect to (1b) and (1c), it is difficult for us as well as for computers to distinguish VBG and VBN from JJ and NN. "Interesting" and "surprised" are good examples. That is why such mistaggings cannot be avoided.

(1d) may result from the fact that some words have many kinds of grammatical functions. For example, "that" has the grammatical function of a demonstrative pronoun, a conjunction and a relative pronoun. So it is difficult for taggers to determine the correct tag because they are based on probabilistic models.

As to (1e), there are some words which have a constant or nearly constant mistagging pattern. If users know the patterns of mistagging, those mistaggings are regarded as having been tagged correctly.

Based on the discussion above, it is clear that the studies using a corpus tagged by a tagger, in this case, APP and RBT, are limited. For example, quantitative studies is impossible. To analyze the frequency of "that" as a relative pronoun is impossible because the taggers cannot distinguish between "that" as a relative pronoun and "that" when it has other grammatical functions such as a demonstrative pronoun or a conjunction. However, qualitative studies, such as analyzing the collocation or occurance of "that" as a relative pronoun will be possible.

In conclusion, the results of the accuracy of APP and RBT were about 90% respectively. In addition, it has been found that there were some tendencies of incorrect tags which appear when tags are assigned by taggers. If the mistagging tendencies were clear, the incorrect tags based on the tendencies were regarded as the correct tags. Therefore, the accuracy of the taggers will be better. Furthermore, even if all of the mistagging tendencies are not demonstrated explicitly, it will be possible to study grammar based on a corpus tagged by the taggers above if the tagger users recognize the mistagging tendencies.

References

Aston, G., and L. Burnard (1998) *The BNC Handbook: Exploring the British National Corpus with SARA*, Edinburgh: Edinburgh University Press

Belmore, N. (1991) 'Tagging Brown with the LOB Tagging Suite', *ICAME Journal No. 15*, Bergen: Norwegian Computing Centre for the Humanities

Brill, E. (1993) *Some Advances in Transformation-Based Part of Speech Tagging*, Cambridge, Massachusetts: MIT

Francis, W. N., and H. Kucera (1989) *Manual of Information to accompany A Standard Corpus of Present-Day Edited American English, for use with Digital Computers, Providence*, Rhode Island: Brown University Press (Corrected and Revised)

Kucera, H. and W. N. Francis (1967) *Computational Analysis of Present-day American English*, Providence, Rhode Island: Brown University Press

Leech, G. (1997) *A Brief User's Guide to the Grammatical Tagging of the British National Corpus*, (Available: http://info.ox.ac.uk/bnc/what/gramtag.html [November 1st, 2003])

Ota, A. (1963) *Tense and Aspect of Present-Day American English*, Tokyo : Kenkyusha

Saito, T., J. Nakamura and I. Akano (Ed.) (1998) *Eigo Corpus Gengogaku*, Tokyo: Kenkyusha

Santorini, B. (1990) *Part-of-Speech Tagging Guidelines for the Penn Treebank Project* (3rd Rivision, 2nd printing), Tokyo : Kenkyusha

Sekine, S (1996) *Manual of Apple Pie Parser 5.7*, New York : New York University

Takaie, H., and H. Suga (1998) *Jissen Corpus Gengogaku*, Tokyo: Kirihara Uni

Tools

Brill, E. (1993) *Rule Based Tagger 1.14*, MIT and the University of Pennsylvania
Sekine, S (1996) *Apple Pie Parser 5.9*, New York University