

A PROBABILISTIC APPROACH TO GRAMMATICAL ANALYSIS  
OF WRITTEN ENGLISH BY COMPUTER.

Andrew David Beale,  
Unit for Computer Research on the English Language,  
University of Lancaster, Bowland College,  
Bailrigg, Lancaster, England LA1 4YT.

ABSTRACT

Work at the Unit for Computer Research on the English Language at the University of Lancaster has been directed towards producing a grammatically annotated version of the Lancaster-Oslo/Bergen (LOB) Corpus of written British English texts as the preliminary stage in developing computer programs and data files for providing a grammatical analysis of unrestricted English text.

From 1981-83, a suite of PASCAL programs was devised to automatically produce a single level of grammatical description with one word tag representing the word class or part of speech of each word token in the corpus. Error analysis and subsequent modification to the system resulted in over 96 per cent of word tags being correctly assigned automatically. The remaining 3 to 4 per cent were corrected by human post-editors.

Work is now in progress to devise a suite of programs to provide a constituent analysis of the sentences in the corpus. So far, sample sentences have been automatically assigned phrase and clause tags using a probabilistic system similar to word tagging. It is hoped that the entire corpus will eventually be parsed.

THE LOB CORPUS

The LOB Corpus (Johansson, Leech and Goodluck, 1978) is a collection of 500 text samples, each containing about 2,000 word tokens of written British English published in a single year (1961). The 500 text samples fall into 15 different text categories representing a variety of styles such as press reporting, science fiction, scholarly and scientific writing, romantic fiction and religious writing. There are two main sections: informative prose and imaginative prose. The corpus contains just over 1 million word tokens in all.

Preparation of the LOB corpus in machine readable form began at the

Department of Linguistics and Modern English Language at the University of Lancaster in the early 1970s under the direction of G.N. Leech. Work was transferred, in 1977, to the Department of English at the University of Oslo, Norway and the Norwegian Computing Centre for the Humanities at Bergen. Assembly of the corpus was completed in 1978.

The LOB Corpus was designed to be a British English equivalent of the Standard Corpus of Present-Day Edited American English, for use with Digital Computers, otherwise known as the Brown Corpus (Kučera and Francis, 1964; Hauge and Hofland, 1978). The year of publication of all text samples (1961) and the division into 15 text categories is the same for both corpora for the purposes of a systematic comparison of British and American natural language and for collaboration between researchers at the various universities.

Word Tagging of the LOB Corpus.

The initial method devised for automatic word tagging of the LOB corpus can be represented by the following simplified schematic diagram:

WORD FORMS → POTENTIAL WORD TAG  
ASSIGNMENT (for each word in isolation)  
→ TAG SELECTION (of words in context)  
→ TAGGED WORD FORMS

Sample texts from the corpus are input to the tagging system which then performs essentially two main tasks: firstly, one or more potential tags and, where appropriate, probability markers, are assigned to each input word by a look up procedure that matches the input form against a list of full word forms, or, by default, against a list of one to five word final characters, known as the 'suffixlist'; subsequently, in cases where more than one potential tag has been assigned, the most probable tag is selected by using a matrix of one-step transition probabilities giving the likelihood of one word tag following another (Marshall, 1983: 141ff).

The tag selection procedure disambiguates the word class membership of many common English words (such as CONTACT, SHOW, TALK, TELEPHONE, WATCH and WHISPER). Moreover, the method is suitable for disambiguating strings of adjacent ambiguities by calculating the most likely path through a sequence of alternative one-step transition probabilities.

Error analysis of the method (Marshall, op. cit.: 143) showed that the system was over 93 per cent successful in assigning and selecting the appropriate tag in tests on the running text of the LOB corpus. But it became clear that this figure could be improved by retagging problematic sequences of words prior to word tag disambiguation and, in addition, by altering the probability weightings of a small set of sequences of three tags, known as 'tag triples' (Marshall, op. cit.: 147). In this way, the system makes use of a few heuristic procedures in addition to the one-step probability method to automatically annotate the input text.

We have recently devised an interactive version of the word tagging system so that users may type in test sentences at a terminal to obtain tagged sentences in response. Additionally, we are substantially extending and modifying the word tag set. The programs and data files used for automatic word tagging are being modified to reduce manual intervention and to provide more detailed subcategorizations.

#### Phrase and Clause Tagging.

The success of the probabilistic model for word tagging prompted us to devise a similar system for providing a constituent analysis. Input to the constituent analysis module of the system is at present taken to be LOB text with post-edited word tags, the output from the word tagging system. We envisage an interactive system for the future.

A separate set of phrase and clause tags, known as the hypertag set, has been devised for this purpose. A hypertag consists of a single capital letter indicating a general phrase or clause category, such as 'N' for noun phrase or 'F' for finite verb clause. This initial capital letter may be followed by one or more lower-case letters representing subcategories within the general hypertag class. For instance, 'Na' is a noun phrase with a subject pronoun head, 'Vzb' is a verb phrase with the first word in the phrase inflected as a third person singular form and the last word being a form of the verb BE.

Strict rules on the permissible combinations of subcategory symbols have been formulated in a Case Law Manual (Sampson, 1984) which provides the rules and symbols for checking the output of the automatic constituent analysis. The detailed distinctions made by the subcategory symbols are devised with the aim of providing helpful information for automatic constituent analysis and, for the time being, many subcategory symbols are not included in the output of the present system. (For the current set of hypertags and subcategory symbols, see Appendix A).

The procedures for parsing the corpus may be represented in the following simplified schematic diagram:

```
WORD TAGGED CORPUS -> T-TAG ASSIGNMENT  
(PARTIAL PARSE) -> BRACKET CLOSING AND  
T-TAG SELECTION -> CONSTITUENT ANALYSIS
```

Phrasal and clausal categories and boundaries are assigned on the basis of the likelihood of word tag pairs opening, closing or continuing phrasal and clausal constituencies. This first part of the parsing procedure is known as T-tag assignment. A table of word tag pairs (with, in some cases, default values) is used to assign a string of symbols, known as a T-tag, representing parts of the constituent structure of each sentence. The word tag pair input stage of parsing resembles the word- or suffixlist look up stage in the word tagging system. Subsequently, the most likely string of T-tags, representing the most probable parse, is selected by using statistical data giving the likelihood of the immediate dominance relations of constituents. Other procedures, which I will deal with later, are incorporated into the system, but, in very broad outline, the automatic constituent analysis system resembles word tagging in that potential categories (and boundaries) are first assigned and later disambiguated by calculating the most likely path through the alternative choices.

In the case of word tagging, the word tagged Brown corpus enabled us to derive word tag adjacency statistics for potential word tag disambiguation. But no parsed corpus exists yet for the purposes of deriving statistics for disambiguating parsing information. A sample databank of constituent structures has therefore been manually compiled for initial trials of T-tag assignment and disambiguation.

## The Tree Bank

When the original set of hypertags and rules was devised, G.R. Sampson began the task of drawing tree diagrams of the constituent analysis of sample sentences on computer print-outs of the word tagged version of the corpus. As tree drawing proceeded, amendments and extensions to the rules for tree drawing and the inventory of hypertags were proposed, on the basis of problems encountered by the linguist in providing a satisfactory grammatical analysis of the constructions in the corpus. The rationale for the original set of rules and symbols, and of subsequent modifications, is documented in a set of Tree Notes (Sampson, 1983 - ).

So far, about 1,500 complete sentences have been manually parsed according to the rules described in the Case Law Manual and these structures have been keyed into an ICL VME 2900 machine which represents them in bracketed notation as four fields of data on each record of a serial file. The fields or columns of data are:- (1) a reference number, (2) a word token of sample text, (3) the word tag for the word and (4) a field of hypertags and brackets showing the constituency-level status of each word token.

Any amendments to the rules and symbols for hypertagging necessitate corresponding amendments to the tree structures in the tree databank.

### The Case Law Manual.

The Case Law Manual (Sampson, 1984) is a document that summarizes the rules and symbols for tree drawing as they were originally decided and subsequently modified after problems encountered by the linguist in working through samples of the word tagged corpus. I will only give a brief sketch of the principles contained in the Case Law Manual in this paper.

Any sequence in the word tagged corpus marked as a sentence is given a root hypertag, 'S'. Between 'S' and the word tag level of analysis, all constituents perceived by the linguist to be consisting of more than one word and, in some cases, single word constituents, are labelled with the appropriate hypertag. Any clause or sentence tag must dominate at least one phrase tag but otherwise unary branching is generally avoided.

Form takes precedence over function so that, for instance, in fact is labelled as a prepositional phrase rather than as an adverbial phrase. No attempt is made to show any paraphrase relationships. Putative deleted or

transposed elements are, in general, not referred to in the Case Law Manual, the exceptions to this general principle being in the treatment of some coordinated constructions and in the analysis of constructions involving what transformational grammarians call unbounded movement rules (Sampson, 1984: 2).

The sentences in the LOB corpus present the linguist with the enormously rich variety of English syntactic constructions that occurs in newspapers, books and journals; and they also force issues - such as how to incorporate punctuation into the parsing scheme, how to deal with numbered lists and dates in brackets - issues which, although present and familiar in ordinary written language, are not generally, if at all, accounted for in current formalized grammars.

### T-TAG ASSIGNMENT

A T-tag is part of the constituent structure immediately dominating a word tag pair, together with any closures of constituents that have been opened, and left unclosed, by previous word tag pairs. Originally, it was decided to start the parsing process by using a table of all the possible combinations of word tag pairs, each with its own T-tag output. Rules of this sort may be exemplified as follows:-

```
      .
      .
(N)   CS - IN = YCP
(N+1) VBN - JJ = JIN : VJ[J : VJ[N
(N+2)   - RB = Y J : YJ[R
(N+3) VBG - RP = Y N : YJ[R
      .
      .
      .
```

A word tag pair, to the left of the equals sign, is accepted as the input to the rule which, by look-up, assigns a T-tag or string of T-tag options (separated by colons) as alternative possible analyses for the input tag pair. In example (N), a subordinating conjunction followed by a preposition indicates that a prepositional phrase is to be opened as daughter of the previous constituent (denoted by the 'wild card' hypertag 'Y'); in example (N+1), a past participle form of a verb followed by an adjective indicates three options:

- a. either close a previously opened adjective phrase and continue an already opened noun phrase or

- b. close a previously opened verb phrase and open an adjective phrase or
- c. close a previously opened verb phrase and open a noun phrase constituent.

In this way, the constituent analysis begins by an examination of the immediately local context and a considerable proportion of information about correct parsing structure is obtained by considering the sequence of adjacent word tag pairs in the input string. In some cases, surplus information is supplied about hypertag choices which later has to be discarded by T-tag selection; in other cases, word tag pairs do not provide sufficient clues for appropriate constituent boundary assignment. Word tag pair input should therefore be thought of as producing an incomplete tree structure with surplus alternative paths, the remaining task being to complete the parse by filling in the gaps and selecting the appropriate path where more than one has been assigned.

#### Cover Symbols.

For the purposes of T-tag look up, word tag categories have been conflated where it is considered unnecessary to match the input against distinct word tags; often, the initial part of a T-tag closes the previous constituent, whatever the identity of the constituent is, and specification of rules for every distinct pair of word tags is redundant. This prevents T-tag assignment requiring an unwieldy 133 \* 133 matrix.

The more general word tag categories are known as cover symbols. These usually contain part of a word tag string of characters with an asterisk replacing symbols denoting the redundant subclassifications. (See Appendix B for a list of cover symbols.)

#### Three stages of T-tag assignment.

T-tag assignment is now divided into three look-up procedures: (1) pairs of word tags (2) pairs of cover symbols (3) single word tags or cover symbols, preceded or followed by an unspecified tag. Each procedure operates in an order designed to deal with exceptional cases first and most general cases last. For instance, if no rules in (1) and (2) are invoked by an input pair of tags, where the second input tag denotes some form of verb, then the default rule - VB = Y]V is invoked such that any tag followed by any form of verb closes the constituent left open by a previous

T-tag look-up rule (where 'Y' is a symbol denoting any hypertag). Subsequently, a verb phrase is opened.

If the first tag of the input pair denotes a form of the verb BE, then the rule BE - VB = Y V in procedure (2) is invoked. Finally, if the first tag of the input pair is 'JJR', denoting a comparative adjective, and the second tag is 'VBN', denoting the past participle form of a verb, then the rule JJR - VBN = Y J in (1) is invoked.

The T-tag table was initially constructed by linguistic intuition and subsequently keyed into the ICL VME 2900 machine. Comparison of results with sections of samples from the tree bank enables a more empirical validation of the entries by checking the output of the T-tag look up procedure against samples of the corpus that have been manually parsed according to the rules contained in the Case Law Manual.

Where alternative T-tags are assigned for any word or cover tag pair, the options are entered in order of probability and unlikely options are marked with the token '@'. This information can be used for adjusting probability weightings downwards in comparison of alternative paths through potential parse trees.

#### Reducing T-tag options.

Some procedures are incorporated into T-tag assignment which serve to reduce the explosive combinatorial possibilities of a long partial parse with several T-tag options. Sometimes, T-tag options can be discarded immediately after T-tag assignment because adjacent T-tag information is incompatible; a T-tag that closes a constituency level that has not previously been opened is not a viable alternative. In cases where adjacent T-tags are compatible, the assignment program collapses common elements at either end of the options and the optional elements are enclosed within curly brackets, separated by one or more colons. Here is the representation in cover symbols and alternative constituent structures of the sentence, "Their offering last night differed little from their earlier act on this show a week or so ago." (LOB reference: CO\* 80 001 - 81 081). Cover symbols and word tags appear in angle brackets:

```
[S[N<DT*>N<N*>{:}N<AP*>N<N*>][V<VB*>]R<R*>
{:}P<IN>N<DT*>N<J*>N<N*>{:}]P<IN>[
N<DT*>N<N*>{:}]F:LR]N<DT*>N<N*>N
+<CC>N<P*>][R<R*>:J<R*>:R<R*>]S<. *>]
```

Gaps in the analysis.

Since the T-tag selection phase of the system does not insert constituents, it follows that any gaps in the analysis produced by T-tag look up must be filled before the T-tag selection stage. By intuition or by checking the output of T-tag assignment against the same samples contained in the tree bank, rules have been incorporated into T-tag assignment to insert additional T-tag data after look up but before probability analysis. When T-tag look up produces [PN] (open prepositional phrase, open and close noun phrase), a further rule is incorporated that closes the prepositional phrase immediately after the noun phrase. Similarly, a preposition tag followed by a wh-determiner (e.g. with whom, to which, by whatever, etc) indicates that a finite clause should be opened between the previous two word tags (whatever precedes the preposition and the preposition itself).

Rules of this sort, which we call "heuristic rules", could be dealt with by including extra entries in the T-tag look up table, but since the constituency status is more clearly indicated by sequences of more than two tags, it is considered appropriate, at this stage, to include a few rules to overwrite the output from T-tag look up, in the same way that heuristics such as 'tag triples' and a procedure for adjusting probability weightings were included in the word tagging system, prior to word tag selection, to deal with awkward cases there.

Long distance dependencies.

Genitive phrases and co-ordinated constructions are particularly problematic. For instance, in The Queen of England's Palace, T-tag look up is not, at present, able to establish that a potential genitive phrase has been encountered until the apostrophe is reached. We know that a genitive constituent might be closed according to whether the potential genitive constituent contains more than one word. Consequently a procedure must be built in to establish where the genitive constituent should be opened, if at all. Co-ordinated constructions present similar problems.

#### T-TAG SELECTION AND BRACKET CLOSING

It is the task of the final phase of the parser to fill in any remaining closing brackets in the appropriate places and calculate the most probable tree structure given the various T-tag options. The bracket closing procedure works

backwards through the T-tag string, selecting unclosed constituents, constructing possible subtrees and assigning each a probability, using immediate dominance probability statistics. Each of the possible closing structures is incorporated into the calculation for the next unclosed constituent; the bracket closing procedure works its way up and down constituency levels until the root node, 'S', has been reached and the most probable analysis calculated.

T-tag options are treated in a similar manner to bracket closing; probabilities are calculated for the alternative structures and the most likely one is selected.

Immediate dominance probabilities.

A program has been devised to record the distinct immediate dominance relationships in the tree bank for each hypertag; the number of permissible sequences of hypertags or word tags that any hypertag can dominate is stored in a statistics file. At initial trials, this was the databank used for selecting the most likely parse, but because the tree bank was not sufficiently large enough to provide the appropriate analysis for structures that, by chance, were not yet included in the tree bank, other methods for calculating probabilities were tried out.

At present, daughter sequences are split into consecutive pairs and the probability of a particular option is calculated by multiplying probabilities of pairs of daughter constituents for each subtree. This method prevents sequences not accounted for in the tree bank from being rejected. Sample sentences have been successfully parsed using this method, but we acknowledge that further work is required. One problem created by the method is that, because probabilities are multiplied, there is a bias against long strings. It is envisaged that normalization factors, which would take account of the depth of the tree, would counterbalance the distortion created by multiplication of probabilities.

#### CONCLUSION

We have found that the success rate for grammatically annotating the LOB corpus using probabilistic techniques for lexical disambiguation is surprisingly high and we have consequently endeavoured to apply similar techniques to provide a constituent analysis.

Corpus data provides us with the rich variety of extant English constructions that are the real test of the grammarian's and the computer programmer's skill in devising an automatic parsing system. The present method provides an analysis, albeit a fallible one, for any input sentence and therefore the success rate of the tagging scheme can be assessed and where appropriate, improved.

#### ACKNOWLEDGEMENTS

The author of this paper is one member of a team of staff and research associates working at the Unit for Computer Research on the English Language at the University of Lancaster. The reader should not assume that I have contributed any more than a small part of the total work described in the paper. Other members of the team are R. Garside, G. Sampson, G. Leech (joint directors); F.A. Leech, B. Booth, S. Blackwell.

The work described in this paper is currently supported by Science and Engineering Research Council Grant GR/C/47700.

#### REFERENCES

- Hauge, J. and Hofland, K. (1978). Microfiche version of the Brown University Corpus of Present-Day American English. Bergen: NAVF's EDB-Senter for Humanistisk Forskning.
- Johansson, S., Leech, G. and Goodluck, H. (1978). Manual of information to accompany the Lancaster-Oslo/Bergen corpus of British English, for use with digital computers. Unpublished document: Department of English, University of Oslo.
- Kučera, H. and Francis, W.N. (1964, revised 1971 and 1979). 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.
- Marshall, I. (1983). 'Choice of Grammatical Word-Class without Global Syntactic Analysis: Tagging Words in the LOB Corpus', *Computers and the Humanities*, Vol 17, No. 3, 139-150.
- Sampson, G.R. (1984). UCREL Symbols and Rules for Manual Tree-Drawing. Unpublished document: Unit for Computer Research on the English Language, University of Lancaster.
- (1983). Tree Notes I-XIV. Unpublished documents: Unit for Computer Research on the English Language, University of Lancaster.

#### APPENDIX A Hypertags and Subscripts.

The initial capital letter of each hypertag represents a general constituent class and subsequent lower case letters represent subcategories of the constituent class. The reader is warned that, in some cases, one lower case letter occurring after a capital letter has a different meaning to the same letter occurring after a different capital letter.

A	As-clause
D	Determiner phrase
Dq	beginning with a wh-word
Dqv	beginning with wh-ever word
E	Existential THERE
F	Finite-verb clause
Fa	Adverbial clause
Fc	Comparative clause
Ff	Antecedentless relative clause
Fn	Nominal clause
Fr	Relative clause
Fs	Semi-co-ordinating clause
G	Germanic genitive phrase
J	Adjective phrase
Jq	beginning with a wh-word
Jqv	beginning with a wh-ever word
Jr	Comparative adjective phrase
Jx	with a measured gradable
L	Verbless clause
M	Number phrase
Mf	Fractional number phrase
Mi	with ONE as head
N	Noun phrase
Na	with subject pronoun head
Nc	with count noun head
Ne	Emphatic reflexive pronoun
Nf	Foreign expression or formula
Ni	IT occurring with extraposition
Nj	with adjective head
Nm	with mass noun head
Nn	with proper name head
No	with object pronoun head
Np	Plural noun phrase
Nq	beginning with a wh-word
Nqv	beginning with a wh-ever word
Ns	Singular noun phrase
Nt	Title
Nu	with abbreviated unit noun head
Nx	premodified by a measure expression
P	Prepositional phrase
Po	beginning with OF
Pq	with wh-word nominal
Pqv	with wh-ever word nominal
Ps	Stranded preposition

R	Adverbial phrase		including pre-nominal possessive pronouns, <u>her</u> , <u>your</u> , <u>my</u> , <u>our</u> ...).
Rq	beginning with a wh-word		
Rqv	beginning with a wh-ever word		
Rr	Comparative adverb phrase	HV*	Grammatical forms of the verb HAVE, ( <u>have</u> , <u>had</u> (past tense), <u>having</u> , <u>had</u> (past participle), <u>has</u> ).
Rx	with a measured gradable		
S	Sentence		
Si	Interpolation	J*	Adjective (including attributive, comparative and superlative adjectives: <u>enormous</u> , <u>tantamount</u> , <u>worse</u> , <u>brightest</u> ...).
Sq	Direct quotation		
T	Non-finite-verb clause		
Tb	Bare non-finite-verb clause		
Tr	FOR-TO clause	N*	Noun (including formulae, foreign words, singular common nouns, with or without word initial capitals, abbreviated units of measurement, singular proper nouns, singular locative nouns with word initial capitals, singular titular nouns with word initial capitals, singular adverbial nouns and letters of the alphabet).
Tg	with <u>-ing</u> participle as head		
Ti	with an infinitive head		
Tn	with past participle head		
Tq	Infinitival indirect question		
U	Exclamation or Grammatical Isolate		
V	Verb phrase		
Vb	ending with a form of the verb BE	P*	Pronoun ( <u>none</u> , <u>anyone</u> , <u>everything</u> , <u>anybody</u> , <u>me</u> , <u>us</u> , <u>you</u> , <u>it</u> , <u>him</u> , <u>her</u> , <u>them</u> , <u>hers</u> , <u>yours</u> , <u>mine</u> , <u>ours</u> , <u>myself</u> , <u>themselves</u> ...).
Ve	containing NOT		
Vg	beginning with an <u>-ing</u> participle	P*A	Subject Pronoun ( <u>I</u> , <u>we</u> , <u>he</u> , <u>she</u> , <u>they</u> ).
Vi	with infinitive head		
Vm	beginning with AM	R*	Adverb (including comparative, superlative and nominal adverbs: <u>galore</u> , <u>delicately</u> , <u>better</u> , <u>least</u> , <u>downstairs</u> , <u>indoors</u> , <u>now</u> , <u>then</u> , <u>to-day</u> , <u>here</u> ...).
Vn	beginning with a past participle	RI*	Adverb which can also be a particle or a preposition ( <u>above</u> , <u>between</u> , <u>near</u> , <u>across</u> , <u>on</u> , <u>about</u> , <u>back</u> , <u>out</u> ...).
Vo	Separate verb operator		
Vp	Passive verb phrase	VB*	Verb form (base form, past tense, present participle, past participle, 3rd person singular forms).
Vr	Separate verb remainder		
Vz	with distinctive 3rd person tense	WD*	Wh-determiner ( <u>which</u> , <u>what</u> , <u>whichever</u> ...).
W	WITH clause		
X	NOT separate from the verb	WP*	Wh-pronoun ( <u>who</u> , <u>whoever</u> , <u>whosoever</u> , <u>whom</u> , <u>whomever</u> , <u>whomsoever</u> ...).
Y	'Wild card'	*S	Plural form (of common nouns, abbreviated units of measurement, locative nouns, titular nouns, adverbial nouns, post determiners and cardinal numbers).
ℓ,+,,-,=	TAG-SUFFIXES for co-ordinated constructions and 'idiom phrases'.	*\$	Genitive form (of singular and plural common nouns, locative nouns with word initial capitals, titular nouns with word initial capitals, adverbial nouns, ordinals, adverbs, abbreviated units of measurement, nominal pronouns, post-determiners, cardinal numbers, determiners and wh-pronouns).

APPENDIX B  
Cover Symbols

AB*	Pre-qualifier or pre-quantifier ( <u>quite</u> , <u>rather</u> , <u>such</u> , <u>all</u> , <u>half</u> , <u>both</u> ...).
AP*	Post-determiner ( <u>only</u> , <u>other</u> , <u>little</u> , <u>much</u> , <u>few</u> , <u>several</u> , <u>many</u> , <u>next</u> , <u>last</u> ...).
BE*	Grammatical forms of the verb BE ( <u>be</u> , <u>were</u> , <u>was</u> , <u>being</u> , <u>am</u> , <u>been</u> , <u>are</u> , <u>is</u> ).
CD*	Cardinal ( <u>one</u> , <u>two</u> , 3, 195* - 60).
DO*	Grammatical forms of the verb DO ( <u>do</u> , <u>did</u> , <u>does</u> ).
DT*	Determiner or Article ( <u>this</u> , <u>the</u> , <u>any</u> , <u>these</u> , <u>either</u> , <u>neither</u> , <u>a</u> , <u>no</u> ;