United States Patent [19]

Kucera et al.

[11] **Patent Number: 4,868,750**

[45] **Date of Patent: Sep. 19, 1989**

[54] COLLOCATIONAL GRAMMAR SYSTEM

[75] Inventors: **Henry Kucera**, Providence, R.I.; **Alwin B. Carus**, Newton, Mass.; **Jeffrey G. Hopkins**, Pawtucket, R.I.

[73] Assignee: **Houghton Mifflin Company**, Boston, Mass.

[21] Appl. No.: **106,127**

[22] Filed: **Oct. 7, 1987**

[51] **Int. Cl.**[4] **G06F 15/02**

[52] **U.S. Cl.** **364/419**; 364/900

[58] **Field of Search** ..... 364/419, 900 MS File, 364/200 MS File

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,661,924	4/1987	Okamoto et al.	364/900
4,703,425	10/1987	Muraki	364/419
4,724,523	2/1988	Kucera	364/419
4,742,481	5/1988	Yoshimura	364/419
4,747,053	5/1988	Yoshimura	364/419
4,750,122	6/1988	Kaji	364/419
4,760,528	7/1988	Levin	364/419
4,773,009	9/1988	Kucera	364/419

OTHER PUBLICATIONS

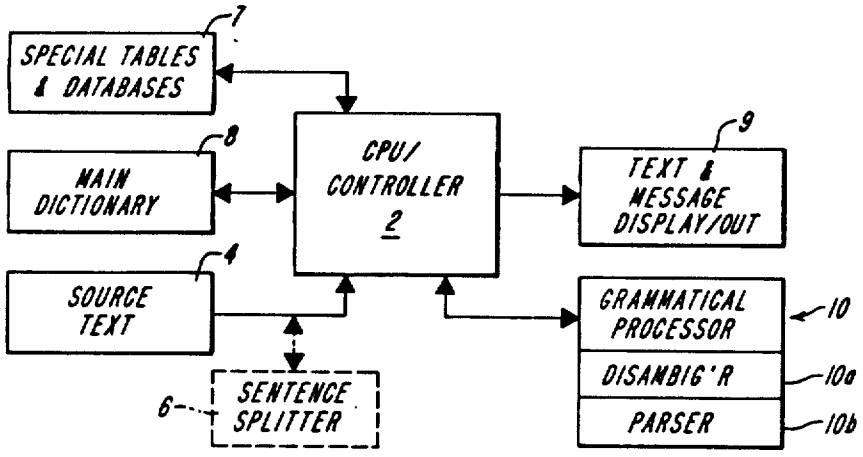
Choice of Grammatical Word-Class Without Global Syntactic Analysis: Tagging Words in the Lob Corpus, Marshall, Ian in *Computers and the Humanities* 17 (1983) 139-150.

Primary Examiner—Michael R. Fleming
Attorney, Agent, or Firm—Lahive & Cockfield

[57] **ABSTRACT**

A system for the grammatical annotation of natural language receives natural language text and annotates each word with a set of tags indicative of its possible grammatical or syntactic uses. An empirical probability of collocation function defined on pairs of tags is iteratively extended to a selected set of tag sequences of increasing length so as to select a most probable tag for each word of a sequence of ambiguously-tagged words. For listed pairs of commonly confused words a substitute calculation reveals erroneous use of the wrong word. For words with tags having abnormally low frequency of occurrence, a stored table of reduced probability factors corrects the calculation. Once the text words have been annotated with their most probable tags, the tagged text is parsed by a parser which successively applies phrasal, predicate and clausal analysis to build higher structures from the disambiguated tag strings. A voice/text translator including such a tag annotator resolves sound or spelling ambiguity of words by their differing tags. A database retrieval system, such as a spelling checker, includes a tag annotator to identify desired data by syntactic features.

17 Claims, 13 Drawing Sheets



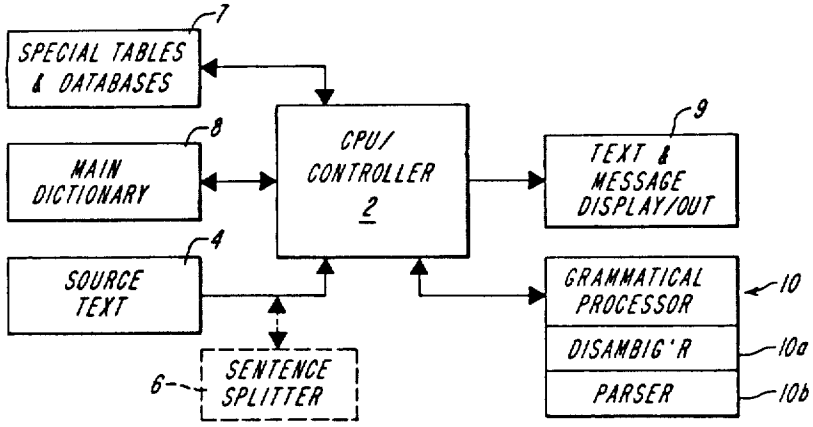


FIG. 1

.	SENTENCE CLOSER (PUNCTUATION TAG)	01
(	OPEN PARENTHESIS (PUNCTUATION TAG)	02
*	"NOT" (ADVERBIAL TAG)	03
)	CLOSE PARENTHESIS (PUNCTUATION TAG)	04
;	SEMI-COLON (PUNCTUATION TAG)	05
-	DASH (PUNCTUATION TAG)	06
,	COMMA (PUNCTUATION TAG)	07
:	COLON (PUNCTUATION TAG)	08
ABL	PRE-QUALIFIER (PRE-NOMINAL TAG)	09
ABN	PRE-QUANTIFIER (PRE-NOMINAL TAG)	10
ABX	PRE-QUANTIFIER/DOUBLE CONJUNCTION (PRE-NOMINAL TAG)	11
AP	POST-DETERMINER (PRE-NOMINAL TAG)	12
AT	ARTICLE (PRE-NOMINAL TAG)	13
BED	"WERE" (VERBAL TAG) <F>	14
BEDZ	"WAS" (VERBAL TAG) <F>	15
BEG	"BEING" (VERBAL TAG) <N>	16
BEI	"BE"-BASE FORM (VERBAL TAG) <N>	17
BEM	"AM" (VERBAL TAG) <F>	18
BEN	"BEEN" (VERBAL TAG) <N>	19
BER	"ARE" (VERBAL TAG) <F>	20
BEZ	"IS" (VERBAL TAG) <F>	21
CC	COORDINATING CONJUNCTION (SENTENTIAL TAG)	22
CD	CARDINAL NUMBER (PRE-NOMINAL TAG)	23
CS	SUBORDINATING CONJUNCTION (SENTENTIAL TAG)	24
DOD	"DID" (VERBAL TAG) <F>	25
DOG	"DOING" (VERBAL TAG) <N>	26
DOI	"DO"-BASE FORM (VERBAL TAG) <N>	27
DON	"DONE" (VERBAL TAG) <N>	28
DOP	"DO" (VERBAL TAG) <N>	29
DOZ	"DOES" (VERBAL TAG) <F>	30
DT	SINGULAR DETERMINER (PRE-NOMINAL TAG)	31
DTI	SINGULAR/PLURAL DETERMINER (PRE-NOMINAL TAG)	32
DTS	PLURAL DETERMINER (PRE-NOMINAL TAG)	33
DTX	DETERMINER/DOUBLE CONJUNCTION (PRE-NOMINAL TAG)	34
EX	EXTISTENTIAL "THERE" (NOMINAL TAG) <P>	35
HVD	"HAD"; PAST TENSE (VERBAL TAG) <F>	36
HVG	"HAVING" (VERBAL TAG) <N>	37
HVI	"HAVE"-BASE FORM (VERBAL TAG) <N>	38
HVN	"HAD"; PAST PARTICIPLE (VERBAL TAG) <N>	39
HVP	"HAVE" (VERBAL TAG) <N>	40
HVZ	"HAS" (VERBAL TAG) <F>	41
IN	PREPOSITION (SENTENTIAL TAG)	42
JJ	ADJECTIVE (PRE-NOMINAL TAG)	43
JJR	COMPARATIVE ADJECTIVE (PRE-NOMINAL TAG)	44
JJS	SUPERLATIVE ADJECTIVE (PRE-NOMINAL TAG)	45
JJT	SEMANTIC SUPERLATIVE (PRE-NOMINAL TAG)	46
MD	MODAL AUXILARY (VERBAL TAG) <F>	47

FIG. 2
(01 THRU 47)

NN	SINGULAR COMMON NOUN - BASE FORM (NOMINAL TAG) <N>	48
NN$	POSSESSIVE SINGULAR COMMON NOUN (PRE-NOMINAL TAG)	49
NNS	PLURAL COMMON NOUN (NOMINAL TAG) <N>	50
NNS$	POSSESSIVE PLURAL COMMON NOUN(PRE-NOMINAL TAG)	51
NNX	NON-POSSESSIVE COMMON NOUN - BASE FORM (NOMINAL TAG) <N>	52
NP	SINGULAR PROPER NOUN - BASE FORM (NOMINAL TAG) <N>	53
NP$	POSSESSIVE SINGULAR PROPER NOUN (PRE-NOMINAL TAG)	54
NPS	PLURAL PROPER NOUN (NOMINAL TAG) <N>	55
NPS$	POSSESSIVE PLURAL PROPER NOUN (PRE-NOMINAL TAG)	56
NR	SINGULAR ADVERBIAL NOUN - BASE FORM (NOMINAL TAG) <P>	57
NR$	POSSESSIVE SINGULAR ADVERBIAL NOUN (PRE-NOMINAL TAG)	58
NRS	PLURAL ADVERBIAL NOUN (NOMINAL TAG) <P>	59
NRS$	POSSESSIVE PLURAL ADVERBIAL NOUN (PRE-NOMINAL TAG)	60
OD	ORDINAL NUMBER (PRE-NOMINAL TAG)	61
PN	NOMINAL PRONOUN (NOMINAL TAG) <P>	62
PN$	POSSESSIVE NOMINAL PRONOUN (PRE-NOMINAL TAG)	63
PP$	POSSESSIVE PRONOUN (PRE-NOMINAL TAG)	64
PP$$	SECOND POSSESSIVE PRONOUN (NOMINAL TAG) <P>	65
PPL	SINGULAR REFLEXIVE PRONOUN (NOMINAL TAG) <P>	66
PPLS	PLURAL REFLEXIVE PRONOUN (NOMINAL TAG) <P>	67
PPO	OBJECTIVE PERSONAL PRONOUN (NOMINAL TAG) <P>	68
PPS	THIRD PERSON NOMINATIVE PERSONAL PRONOUN(NOMINAL TAG <P>	69
PPSS	NON-THIRD PERSON NOMINATIVE PERSONAL PRONOUN (NOMINAL TAG) <P>	70
PPX	NON-POSSESSIVE PERSONAL PRONOUN (NOMINAL TAG) <P>	71
QL	QUALIFIER (PRE-NOMINAL TAG)	72
QLP	POST-QUALIFIER (PRE-NOMINAL TAG)	73
RB	ADVERB (ADVERBIAL TAG)	74
RBR	COMPARATIVE ADVERB (ADVERBIAL TAG)	75
RBT	SUPERLATIVE ADVERB (ADVERBIAL TAG)	76
RN	NOMINAL ADVERB (ADVERBIAL TAG)	77
RP	ADVERB/PARTICLE (ADVERBIAL TAG)	78
TO	INFINITIVAL "TO" (VERBAL TAG) <(IN)>	79
UH	EXCLAMATION (SENTENTIAL TAG)	80
VBD	VERB PAST TENSE FORM (VERBAL TAG) <F>	81
VBG	VERB PRESENT PARTICIPLE (VERBAL TAG) <N>	82
VBI	VERB INFINITIVE - BASE FORM (VERBAL TAG) <N>	83
VBN	VERB PAST PARTICIPLE (VERBAL TAG) <N>	84
VBP	VERB NON-THIRD PERSON SINGULAR, PRESENT TENSE (VERBAL TAG) <N>	85
VBZ	VERB THIRD-PERSON SINGULAR, PRESENT TENSE (VERBAL TAG) <F>	86
WDT	WH-DETERMINER (PRE-NOMINAL TAG)	87
WP$	PERSONAL WH-PRONOUN (PRE-NOMINAL TAG)	88
WPO	OBJECTIVE WH-PRONOUN (NOMINAL TAG) <P>	89
WPS	NOMINATIVE WH-PRONOUN (NOMINAL TAG) <P>	90
WQL	WH-QUALIFIER (PRE-NOMINAL TAG)	91
WRB	WH-ADVERB (ADVERBIAL TAG)	92
XX	-PARSER INTERNAL - NO TAG ASSIGNED-	93
ZZZZ	END-OF-FILE MARKER	94

FIG. 2
(48 THRU 94)

EXAMPLES OF "NORMAL" MAIN DICTIONARY RECORDS

/1	/26	/30	/34	/80
.			.	
!			.	
a			AT	
aback			RB	
abandonment	NI			
abase		VI		
abasement	NI			
abashed			JJ	
as			CS IN QL RB	
back	NI	VI	JJ RB	
backing	NI			

FIG. 3A

EXAMPLES OF "EXCEPTION" MAIN DICTIONARY RECORDS

/1	/26	/30	/34	/80
adieu	NIs		(NN: adieus/adieux)	$
adieus	N4p		(NNS: adieu)	$
adieux	N4p		(NNS: adieu)	$
am			BEM (xref: be)	$
are			BER (xref: be)	$
arise			(VBI-1: arose/arisen)	$
arisen			(VBN: arise)	$
arose			(VBD: arise)	$
ate			(VBD: eat)	$
awake			JJ RB (VBI-1: awoke/awoken)	$
awoke			(VBD: awake)	$
awoken			(VBN: awake)	$
bacilli			(NNS: bacillus)	$
bacillus			(NN: bacilli)	$
bade			(VBX: bid)	$
base			(NN: bases)	$
bases			(NNS: base/basis)	$
basis			(NN: bases)	$
be			BEI (xref: is)	$
bear			(VBI-1: bore/borne)	$

FIG. 3B

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.