Appendix C1

A List of AWK Programs and Data Files Used in the Study

N	File Name	Dated	Description
1	ad_sp_ed.awk	980628	Insert space before the return mark
2	add$0.awk	980820	Adds all the values contained in $1 through $n respectively.
3	bun_fre2.awk	980724	The main program of "Sentence Profiler (Ver.1)." Print sentence- length profile table and graph (=> See "bun_fre4. awk" for a revised version).
4	bun_fre4.awk	980730	Revised version of "bun_fre2.awk"
5	cnt_freq.awk		Counts the number of each tag sequence and to produce a list of modal verb-structures with frequency information (See Endnote 31, Chapter 4).
6	capital.awk	980622	Prints text lines beginning with a capital letter (for extracting proper nouns from a wordlist).
7	chikan.awk	980814	Compares an input file and a specified dictionary. If the words in $1 of the dictionary matches words in the input file, the latter will be replaced with the $2 data in the former. (See "fmatch. awk").
8	cleantag.awk	980818	Cleans up a file tagged with the Brill Tagger, and replaces the default slash symbol (/) with the underbar (_).
9	countme.awk	981002	Counts the number of words in a text , either as type or token.
10	del_hyph.awk	971117	Deletes line-end hyphens.
11	del_nbr.awk	980623	Deletes line-initial numbers and symbols.
12	del_null.awk	980205	Deletes excess blank lines, leaving only one blank line.
13	del_rtn.awk	980518	Deletes the return mark at the end of each record
14	del{_}.awk	981007	Deletes the "{…}_idm" mark from the output of "idmfreq. awk"(to be used as part of "fmatch3.bat").
15	delblank.awk	980601	Deletes all blank lines.
16	delkigou.awk	980721	Deletes all symbols and marks in $2.
17	delslash.awk	970831	Replaces the slash with a space.
18	ex_there.awk	980628	Extract all the "Ex-There" constructions.
19	f1_del.awk	980417	Print all the data except those in $1.
20	fmatch.awk	980814	The main program of "Collocation and Idiom Finder (Ver.1)." Marks all the matched strings in the format of "{ idiom }_IDM."
21	hv_vbn.awk	980628	Extracts all the present perfect constructions from a tagged corpus.
22	ichigyo.awk	980205	Same as "del_null.awk"
23	idmfreq.awk	981007	Produces a frequency comparison table of specified collocations and idioms. Used as part of "Collocation and Idiom Finder (Ver.1).
24	if$2none.awk	980821	Prints records whose $2 is not blank.
25	if_md.awk	980628	Extracts all the IF+MD constructions from a tagged corpus.
26	JJ.awk	980912	Extracts all the adjectives from a tagged wordlist.
27	kaihi-1.awk	980523	Prints the data as is, except for those marled by #.
28	kaihi-2.awk	980730	Prints the data as is if marked marked by #. If not, adds sentence ID numbers before printing.
29	karamoji.awk	980417	Deletes sentence-initail space.
30	kensaku.awk	980201	Regular expression search from the command line.
31	l_sp_del.awk	971004	Deletes excess line-initial space.
32	line_nbr.awk	980518	Adds sentence numbers.
33	makeline.awk	980201	Inserts a return code at the end of sentence-initial punctuation marks and symbols, except at specified abbreviations (used in conjunction with "txt_id.awk").

34	matching.awk	980620	Replaces each entry word in the input file with a corresponding WL tag as defined in the WL-tag dictionary file. Non-match strings are printed as is (used as part of "Word Level Checker").
35	matchnew.awk	980825	Replaces each entry word in the input file with a corresponding WL tag as defined in the WL-tag dictionary file (See endnote 8, Chapter 3).
36	merge$0.awk	980623	Merges two wordlists (Add FILE1 to FILE2, and prints FILE3, for $0).
37	merge$1.awk	980703	Merges two wordlists (Add FILE1 to FILE2, and prints FILE3, for $1).
38	nandoprn.awk	980624	Sorts and prints the results of "matching.awk" (used as part of "wlc.bat" and "w_nando.bat").
39	NN.awk	980915	Extracts all the words with NN tags.
40	non_cap.awk	980624	Prints all lines starting with a lower case letter (for extracting data other than proper nouns from a wordlist).
41	open_con.awk	980529	Opens contractions (e.g. I'm, we'd, we'll, couldn't, etc.) . Used before executing the Brill Tagger.
42	predcnt1.awk	980725	Counts the number of predications (mentioned in End note 19, Chapter 2)
43	prn_!tag.awk	980825	Prints text data only from a POS-tagged text.
44	prn_tag.awk	980601	Extracts POS tag data from a tagged text, and prints them onto a separate file (See Endnote 6, Chapter 2).
45	prn{_}.awk	981007	Prints lines that include strings marked "{…}_IDM " (used as part of "fmatch1.bat" and "fmatch2.bat").
46	prn_MD.awk	980601	Extracts all MD tags from a tagged text, and prints them onto a separate file (See Endnote 30, Chapter 4).
47	r_sp_del.awk	980207	deletes space between the return code and the last word of each sentence.
48	RB.awk	980912	Extracts all the words with RB tags.
49	rtn@}.awk	981007	Inserts a return code at the "{ " mark to the output of "prn{_}.awk." (used as part of "fmatch1.bat2" and "fmatch2. bat").
50	sentence.awk	980718	Main program of "Sentence Profiler (Ver.1)." Counts the numbers of words and sentences, and the average number of words per sentence, and print the result.
51	shiage.awk	980107	Deletes unnecessary data from the output of "prn{_{.awk => rtn@}.awk" and prints the result after sorting (used as part of "fmatch1.bat" and "fmatch2.bat").
52	sp_kigou.awk	980523	Adds space before and/or after specified punctuation marks and symbols (used in conjunction with the Brill Tagger).
53	tagme.awk	980623	Experimental tagging program.
54	TagToSyn.awk	980720	Extract syntactic information from POS tag data.
55	tokei.awk	980801	Calculates sum, mean, variance, SD, dispersion and usage.
56	txt_id.awk	980623	Adds "Sentence ID and Number" to a plain running text.
57	VB.awk	980830	Extracts all the words with VB tags (See Endnote 25, Chapter 3).
58	voc_lev1.awk	980823	Processed input data for "voc_lev2.awk"
59	voc_lev2.awk	980823	Prints the results of "matching.awk => nandoprn. awk" with a graph and a table (used as part of "wlc.bat").
60	mk_list.awk	980801	A multi-function wordlist compiler, mk_list.awk. Mentioned in Endnote 14, Chapter 3. See Appendix C2 for program source.
61	word.awk	980828	Produces a simple wordlist from a plain running text file.
62	wordlist.awk	980718	Produces a simple wordlist with frequency information from a plain running text file.
63	wrdlevel.awk	980623	Replaces entries in a wordlist with WL tags.

	Batch Files		
1	fmatch1.bat	980818	makeline.awk => del_null.awk => txt_id.awk =>kigou_3.awk =>fmatch.awk
2	fmatch2.bat	981007	makeline.awk => del_null.awk => kigou_3. awk => fmatch.awk => prn{_}.awk => rtn@}.awk => prn{_}.awk => shiage.awk
3	fmatch3.bat	981007	makeline.awk => del_null.awk => kigou_3. awk => fmatch.awk => prn{_{.awk => rtn@}.awk => prn{_}.awk => shiage.awk => idmfreq.awk => del{_}.awk
4	mk_brill.bat	980824	makeline.awk => sp_kigou.awk => open_con.awk
5	seikei.bat	980622	r_sp_del. awk => l_sp_del.awk => del_hyph.awk => ad_sp_ed. awk => l_sp_del.awk => ad_sharp.awk => del_rtn.awk => del_null.awk
6	sp_prep.bat	980801	makeline.awk => del_null.awk => txt_id.awk => kigou_2.awk
7	sp_prep2.bat	980801	makeline.awk => del_null.awk => txt_id.awk => kigou_2.awk => open_con.awk
8	w_nando.bat	980923	wordlist.awk => del_nbr.awk => matching.awk => nandoprn. awk
9	w_nando1.bat	980923	word.awk => del_nbr.awk => matching.awk => nandoprn.awk
10	wlc.bat	980923	Batch file for "Word Level Checker (Ver 1)." Programs are executed in the following order (words are treated as types): wordlist.awk => del_nbr.awk => matching.awk => nandoprn. awk => voc_lev1.awk => voc_lev2.awk
11	wlc1.bat	980923	Batch file for "Word Level Checker (Ver 1)." Programs are executed in the following order (words are treated as tokens): word.awk => del_nbr.awk => matching.awk => nandoprn.awk => voc_lev1.awk => voc_lev2.awk
	Dictionary and Other Data Files		
1	idiom-3dic	980823	Regular expression list of idioms and collocations, used in conjunction with "fmatch.awk.".
2	syntaxt-1.dic	980717	Rule files for "TagToSyn.awk."
3	wrdlvl-2.dic	980906	Master WL-tag dictionary for "Word Level Checker (Ver. 1)"
4	MD.dic	980927	Master dictionary of MD-verb combinations.
5	jargon.txt	980813	Master dictionary of "Business jargons."
6	d-marker.tag	980518	Master dictionary of discourse markers.
7	lemma-1.dic	980615	Master lemma dictionary, to be used in conjunction with WordSmith.
8	loblemma.dic	980826	POS tag lemmatization dictionary for LOB Corpus.