

プログラムのページ

担当和田英一

6901. 発音別に配列した辞書の作成*

西村 恕彦 (通産省工業技術院電気試験所)

問題 普通の英語の辞書では、見出し語はその

もとの語 母音をそろえた位置

CONGENIAL	CONGENIAL
ASSOCIATE	ASSOCIATE
DEVIATE	DEVIATE
ASCRIBE	ASCRIBE
COMPATIBLE	COMPATIBLE
ATTRIBUTE	ATTRIBUTE
CONTRIBUTE	CONTRIBUTE
CHARACTERIS	CHARACTERISTIC
DEDICATE	DEDICATE
DEFICIENT	DEFICIENT
ADDICTED	ADDICTED
AFRAID	AFRAID
BID	BID
DEVOID	DEVOID
BESIDE	BESIDE
COINCIDE	COINCIDE
CONFIDE	CONFIDE
CONFIDENT	CONFIDENT
CONSIDERING	CONSIDERING
DIE	DIE
CONVENIENT	CONVENIENT
DEFICIENT	DEFICIENT
ACQUIESCE	ACQUIESCE
BELIEVE	BELIEVE
DIFFER	DIFFER
DELIGHT	DELIGHT
DELIGHTED	DELIGHTED
AVAIL	AVAIL
BUILD	BUILD
AIM	AIM
BRIMFUL	BRIMFUL
ABSTAIN	ABSTAIN
AGAIN	AGAIN
AKIN	AKIN
BEGIN	BEGIN
CERTAIN	CERTAIN
COMPLAIN	COMPLAIN
CONTAIN	CONTAIN
CONVINCE	CONVINCE
COINCIDE	COINCIDE
DISTINCT	DISTINCT
BEHIND	BEHIND
ACQUAINED	ACQUAINED
DESTINED	DESTINED
ACCORDING	ACCORDING
ANYTHING	ANYTHING
BRING	BRING
CLING	CLING
CONSIDERING	CONSIDERING
DISTINGUISH	DISTINGUISH
AGAINST	AGAINST

第1図 分類ずみの最終出力

つづり字の ABC 順に配列されている。一般的な用途では、この配列が十分にはやい、一意な検索を保証する。しかし、普通の配列の辞書ではあまり役に立たないような問題の例として、次のような事項の調査をあげることができる。すなわち、脚韻、語尾・語中の子音や母音、二重母音の発音、複合語などである。また文中の分詞 (-ing や -ed) の機能の調査もそうである。

語尾配列辞書 (reverse sorted dictionary) や文脈つき語尾配列索引 (reverse sorted KWIC index) が、これらの問題への一つの解答である。いま一つ考えられる方法を、ここに述べる。

このプログラムは各単語を、語頭からのつづり字の ABC 順ではなく、各単語中に含まれるつづり字の ABC 順に配列するためのものである。このプログラムは一般的な能力を有するものであるが、一例として、単語中の母音 (a, e, i, o, u, y) をキーとして、英語の単語を再配列するのに用いた例を示す。第1図はテストデータについての最終出力の一部であって、母音“i”で始まる文字列を各単語中で識別して配列した部分である。

プログラム プログラム“ROTATION”は、長さ1ないし67字の単語中の母音を識別し、その文字が出力レコード中の一定のけた位置にくるように、単

もとの語 母音をそろえた位置

AKIN	AKIN
AKIN	AKIN
ALIVE	ALIVE
ALIVE	ALIVE
ALIVE	ALIVE
ALL	ALL
ALLUDE	ALLUDE
ALLUDE	ALLUDE
ALLUDE	ALLUDE
ALONG	ALONG
ALONG	ALONG
B.C.	B.C.
BE	BE
BEAR	BEAR
BEAR	BEAR
BECAUSE	BECAUSE
BECAUSE	BECAUSE
BECAUSE	BECAUSE
BECAUSE	BECAUSE

第2図 中間出力

* Infix Arrangement Dictionary, by Hirohiko Nisimura (Electrotechnical Laboratory)

010100	IDENTIFICATION DIVISION.			
010200	PROGRAM-ID.			
010300	ROTATION.		NISIMURA	
010400	AUTHOR.			
010500	NISIMURA-H. ELECTROTECHNICAL LABORATORY.			
010600	JAPANESE GOVERNMENT.			
010700	DATE-WRITTEN.			
010800	1967-7-12.			
010900	DATE-COMPILED.			
011000	SECURITY.			
011100	MAY NOT BE COPIED WITHOUT A PERMISSION.			
011200	REMARKS.			
011300	TANGO NONAKANO MOZI NO SYURUI O SIRABERU.			
011400	ENVIRONMENT DIVISION.			
011500	CONFIGURATION SECTION.			
011600	SOURCE=COMPUTER.			
011700	NEAC-2200.			
011800	OBJECT=COMPUTER.			
011900	NEAC-2200.			
012000	INPUT-OUTPUT SECTION.			
012100	FILE-CONTROL.			
012200	SELECT I=FILE	ASSIGN TO		
012300	CARD-READER E.		NISIMURA	
012400	SELECT O=FILE	ASSIGN TO		
012500	PRINTER B.		NISIMURA	
012600	NOTE			
012700	CARD-PUNCH A.		NISIMURA	
012800	DATA DIVISION.			
012900	FILE SECTION.			
013000	FD I-FILE			
013020	LABEL RECORD IS		NISIMURA	
013040	OMITTED		NISIMURA	
013100	DATA RECORD I-RECORD.			
013200	01 I-RECORD.			
013300	10 FILLER	PICTURE X(12).		
013400	10 GEN=GO.			
013500	20 GEN=ZI	OCCURS 67	PICTURE X(1).	
013700	10 OWARI=ZI		PICTURE X(1).	
013800	FD O-FILE			
013820	LABEL RECORD IS		NISIMURA	
013840	OMITTED		NISIMURA	
013900	DATA RECORD O-RECORD.			
014000	01 O-RECORD.			
014100	10 O=AREA	PICTURE X(80).		
014400	WORKING-STORAGE SECTION.			
014700	77 SIRUSI	PICTURE X.		
014800	77 ARI	VALUE "A"	PICTURE X.	
014900	77 NASI	VALUE "N"	PICTURE X.	
015000	77 HYO-I		PICTURE 99.	
015200	77 BOIN-SU	VALUE 6	PICTURE 99.	
015600	77 MOZI		PICTURE X.	
015700	77 KUUZI	VALUE " "	PICTURE X.	
015800	77	GEN-I	PICTURE 9(3).	
015850	77	GEN-K	PICTURE 9(3).	
015900	77	BOIN-ITI	VALUE 117	
015950	77	KOTAE=K	PICTURE 9(3).	
016000	01 W-AREA.			
016100	05 A-AREA.			
016150	20 KOTAE=ZI	OCCURS 240	PICTURE X(1).	
016200	05 B-AREA	REDEFINES A=AREA.		
016250	10 FILLER		PICTURE X(80).	
016300	10 W=RECORD		PICTURE X(80).	
016350	10 FILLER		PICTURE X(80).	
016400	05 C=AREA	REDEFINES A=AREA.		
016450	10 FILLER		PICTURE X(115).	
016500	10 AKI		PICTURE X(1).	
016550	10 KOTAE=GO		PICTURE X(124).	
016600	01 BOIN-HYO.			
016700	10 ATAI	VALUE "AEIOUYVVV"	PICTURE X(9).	
016800	10 HYO	REDEFINES ATAI.		
016900	20 BUIN	OCCURS 9	PICTURE X(1).	

第3図 発音配列プログラム(1)

```

018000 PROCEDURE DIVISION.
018100 HAZIME.
018200 OPEN INPUT I=FILE.
018300 OPEN OUTPUT O=FILE. NISIMURA
018400 YOMIKOMU. NISIMURA
018500 READ I=FILE AT END GO TO OWARI.
018600 IF "YPRINTS" = I=RECORD GO TO OWARI.
018700 MOVE KUUZI TO OWARI-ZI. NISIMURA
018800 MOVE I TO GEN=I.
018900 MOVE NASI TO SIRUSI.
019000 MOZI=O=TOKU.
019100 MOVE GEN-ZI (GEN=I) TO MOZI.
019400 IF MOZI = KUUZI GO TO GEN=OWARI. NISIMURA
019500 MOVE I TO HYO=I.
019600 BOIN=HYO=HIKI.
019700 IF HYO=I GREATER BOIN=SU
019800 GO TO BOIN=DE=NAI.
019900 IF MOZI = BOIN (HYO=I)
020000 GO TO BOIN=ARI.
020100 ADD I TO HYO=I.
020200 GO TO BOIN=HYO=HIKI.
020300 BOIN=DE=NAI.
020400 ADD I TO GEN=I.
020500 GO TO MOZI=O=TOKU.
020600 BOIN=ARI.
020650 MOVE I=RECORD TO W=RECORD. NISIMURA
020700 SUBTRACT GEN=I FROM BOIN=ITI GIVING KOTAE=K. NISIMURA
020900 MOVE I TO GEN=K.
021000 MOVE SPACE TO KOTAE=ZI (KOTAE=K). NISIMURA
021050 ADD I TO KOTAE=K. NISIMURA
021100 KOTAE=UTUSU.
021200 MOVE GEN-ZI (GEN=K) TO MOZI. NISIMURA
021250 MOVE MOZI TO KOTAE=ZI (KOTAE=K). NISIMURA
021300 IF MOZI = KUUZI GO TO KOTAE=UTUSI=OWARI. NISIMURA
021600 ADD I TO KOTAE=K.
021700 ADD I TO GEN=K.
021800 GO TO KOTAE=UTUSU.
021900 KOTAE=UTUSI=OWARI.
022000 MOVE W=RECORD TO O=AREA.
022100 WRITE O=RECORD.
022200 MOVE ARI TO SIRUSI.
022300 TUGI=NO=ZI.
022400 ADD I TO GEN=I.
022500 GO TO MOZI=O=TOKU.
022800 GEN=OWARI.
022900 IF SIRUSI = ARI GO TO GO=SYORI=OWARI.
022930 MOVE I=RECORD TO W=RECORD.
022960 MOVE SPACE TO AKI.
023000 MOVE GEN=GO TO KOTAE=GO. NISIMURA
023100 MOVE W=RECORD TO O=AREA.
023200 WRITE O=RECORD.
023300 GO=SYORI=OWARI.
023400 GO TO YOMIKOMU.
023600 OWARI.
023700 CLOSE I=FILE O=FILE.
023800 STOP RUN.
999000 NOTE END. NISIMURA
999200 NOTE END. NISIMURA

```

第3図 発音配列プログラム(2)

語の文字列全体をずらす。このフェーズの中間出力は第2図のようなものである。これを次のフェーズで普通にソーティングすると、第1図の最終出力が得られる。

テストにおいては、ソーティングは別途にソートルーチンを利用したが、コボル語で書くとすれば、主プログラムの手続き部は右のようになるだろう。

```

CALL "ROTATION".
SORT S-FILE
ASCENDING KEY MOZIRETU
USING O-FILE
GIVING I-FILE.
STOP RUN.

```

入出力の仕様は次のとおりである。

(1) 入力レコード上の単語は、コラム 13 から始まって、最初に現われる空白の前の文字、またはコラム 79 で終る。これ以外の位置の文字はすべて注記とみなされる。

(2) 入力語の文字列は順次に検査されて、登録されている母音があると、その母音コラム 37 にくるように語の文字列がシフトされて、出力レコードに写される。出力レコード中の語の左右には空白がそう入される。

(3) 入力レコードのコラム 13 が空白であると、入力レコードがそのまま出力レコードとされる。

(4) 入力語に母音が含まれていない場合には、入力語の語頭の文字が出力レコードのコラム 37 にくるようにシフトされる。

参考文献

- 1) 郡司利男：英語学習逆引辞典，開文社，1967.
- 2) E. Mater：Rückläufiges Wörterbuch der Deutschen Gegenwartssprache，Veb Verlag Enzyklopädie，1965.
- 3) 蓼沼良一，玉井陽子：基本英語語彙集，情報処理学会機械翻訳研究委員会，1967-1.

(昭和 43 年 4 月 19 日受付)