

Joel Chandler Harris 作「リーマスじいや物語」の英語に見られる特徴

—(1) コーパス構築と語彙研究 *—

五百蔵 高 浩

高知女子大学文化学部

ioroi@el.kochi-wu.ac.jp

1. はじめに

コンピュータに代表される情報機器は絶えず目覚ましいほどの発展を続けている。処理可能なデータ量も莫大に増加し、その速度も著しく向上しつつある。それに伴い、言語学の分野においてもコンピュータを利用した研究がますます活発になりつつある。大量の言語資料を機械可読なテキストファイルとしてデータベース（「コーパス」と呼ばれることが多い）化し、コンピュータ上で検索・分析することによって、ことばの持つ特徴を実証的に追っていくアプローチに关心を寄せる研究者および言語教育者の層はますます厚くなりつつある。

本論文の目的は、Joel Chandler Harris (1848-1908) の全作品を収めた本である *The Complete Tales of Uncle Remus* の本体物語部分を電子テキスト化し、コンピュータによる検索処理を通じて、この作品において用いられている方言が持っている特徴を明らかにしていく試みの一端を報告することである。

リーマスじいやの語る民話『ウサギどんキツネどん』として知られているこの物語集は、今からおよそ 100 年前に出版された。内容は、リーマスという黒人が、白人の農園である主人の息子に語り聞かせるという形式を取った物語で、話の中では、動物たちが人間と同じ語り口で会話を進めていく様子が描かれている。アメリカ英語の方言研究という観点から補足するならば、この物語は、当時のアメリカ南部の黒人が用いていた英語を忠実に記録したものであると言われている。(1) は、それぞれ、『全作集』の序文と Dillard (1972) から引用したものである。

(1) he [Harris] took great pains to spell the language as it sounded – with all the inflections and rhythms, word combinations, and expressions characteristic of the dialect.” (A Note on the Text; *The Complete Tales of Uncle Remus*, Richard Chase ed., Houghton Mifflin Company, 1983. p. xvii)

Before Emancipation, the children of the slave-owing class had plenty of opportunity to acquire Black English; individual families obviously differed as to whether the acquisition was welcomed or resisted. The knowledge of Negro dialect displayed by outstanding white dialect writers like Joel Chandler Harris, Ambrose E. Gonzales, and Charles Colcock Jones, Jr., must have been acquired in this way. (Dillard (1972: 193))
(Underlines added)

当然のことながら、この物語において用いられている方言の特徴についても、標準英語の研究と同じく、形態論、音韻論、統語論、文体論、その他様々な観点より分析することが可能であると思われる。しかし、残念なことに、この作品の言語面の特徴に対して、これまで十分な関心は払われておらず、従って、まとまった研究書というのも殆ど見られない。僅かに Sumner Ives の *Phonology of the Uncle Remus Stories* (絶版) と他の数編の論考が目にとまるのみである。Ives (1955) はこの方言の一般的特徴について次のように述べている。

(2) The best known of all the dialects which Harris used is the Negro dialect of the Uncle Remus stories. I have shown elsewhere some details of how this dialect differed from rustic white speech, as Harris represented it, and shall give here only a few generalizations. There are, it is true, many items which appear in both his Negro and his rustic white speech, but there are others which appear only in his representation of Negro speech. Among them are the use of *d* for *th* when it indicates a voiced sound, as in *the, that, them, and whether*; the use of final *f* for voiceless *th* in words like *mouth* and *tooth*; the omission of *r* in words like *before* and *sure*; and the omission of *h* in words like *what* and *why*. Also, the Uncle Remus speech shows far more assimilation, more loss of initial unstressed syllables, more leveling of preterit and past participle forms, and a greater density of archaic forms. Uncle Remus himself has a peculiarly ornate use of words, with such exotic formations as *sollumcholly*, *sustonished*, and *rekerremember*. He has a rich imagery, with such expressions as leg-bail for escape by flight and *you're thumpin' de wrong watermillion* for barking up the wrong tree. Furthermore, his speech is larded with such proverbs as 'oman tongue ain't got no Sunday and such comparisons as *ez ca'm ez a dead pig in de sunshine*. (Ives (1955: 416))

この物語の言語にコンピュータを用いた分析を適用してみると何が見えてくるか。それが本研究の関心である。以下、コーパス作成の過程と検索作業の手順と検索結果の一例を報告する。

2. コーパスの作成

2.1 データ源

データ源としては Richard Chase 編の *The Complete Tales of Uncle Remus*. Boston: Houghton Mifflin Company, 1955 を用いた。(3)に示したように、この物語集には 9 編の発行年が異なる物語群が収められている。最初の物語集である Uncle Remus: His Songs and His Saying が 1880 年出版、最後の物語集である Seven Tales of Uncle Remus の出版が 1948 年となっているので、その間 60 年以上の時間的広がりがある。さらに、Harris の没年が 1908 年であることから、最後の 2 作品はその後の出版ということになる。

物語集のタイトルは長いので、以下、行末の括弧に入れたアルファベットでそれぞれのテキストを呼ぶこととする。

- (3) 1. Uncle Remus: His Songs and His Sayings, 1880, 1883² (A)
2. Nights with Uncle Remus: Myths and Legends of the Old Plantation, 1883 (B)
3. Daddy Jake, the Runaway: And Short Stories Told after Dark, 1889 (C)
4. Uncle Remus and His Friends: Old Plantation Stories, Songs, and Ballads with Sketches of Negro Character, 1892 (D)
5. Told by Uncle Remus: New Stories of the Old Plantation, 1905 (E)
6. Uncle Remus and Brer Rabbit, 1907 (F)
7. Uncle Remus and the Little Boy, 1910 (G)
8. Uncle Remus Returns, 1918 (H)
9. Seven Tales of Uncle Remus, 1948 (I)

2.2 使用機器とテキストファイル化

最近では多くの文学作品が電子テキストファイル化されており、インターネットを通じて入手可能な場合が多いが、残念ながら Harris の作品は一部しか公開されていない。従って、データベースの作成にあたって、まず、およそ 860 ページある全作品集を電子テキスト化する作業を行った。作

業に用いたのはパソコン、スキャナ、および市販の文字認識ソフトウェアである。この作業及び後の分析において主に使用したハードウェア、ソフトウェア類は(4)に挙げてある。これ以外に利用したプログラムおよびソフトウェアについては、以下、必要に応じて触れることがある。

(4) パソコン：IBM互換機2台(OS: Windows95, Linux)

スキャナ：エプソン GT6500

OCRソフト：Win Reader Pro. ver. 3, Media Drive Corporation社製

表計算ソフト：Lotus 1-2-3

エディタ：秀丸エディタ for Windows 95, ver. 2.23

文字列検索ソフト：QGREP³

この装置を用いて、テキストを文字認識ソフトで読み取り、テキストファイルに変換した上で1ページごとに保存していった。最近の文字認識ソフトの速度および精度は数年前のものと比較にならないほど格段に向上しているが、印刷に使用されている書体の影響を大きく受けるので、小文字のIを大文字のIと誤認識しているケースが数多く見られる。こういった誤認識された文字はエディタを用いて、画面表示と元のテキストとを照合しながら手作業で修正していった。この後、1ページごとに保存されたテキストファイルを順次結合し、(3)に挙げた作品ごとにまとめた。

上記の手順で作成した基礎データは、(5)に示すような平テキスト(plain text)ファイルである。収録したのは物語のタイトル及び本文部分のみであり、著者による脚注部分は割愛した。1行あたりの文字数は、今後の検索作業における利便性を考慮し、70文字になるように改行記号を入れた。

(5)

Uncle Remus Initiates the Little Boy

1

One evening recently, the lady whom Uncle Remus calls "Miss Sally" nussed her little seven year old boy. Making search for him through the house and through the yard, she heard the sound of voices in the old man's cabin, and, looking through the window, saw the child sitting by Uncle Remus. His head rested against the old man's arm, and he was gazing With an expression of the most intense interest into the rough, Weather-beaten face, that beamed so kindly upon him. This Is what "Miss Sally" heard:

"Bimeby, one day, atter Brer Fox bin doin' all dat he could fer ter ketch Brer Rabbit, en Brer Rabbit bin doin' all he could fer to keep 'im fum it, Brer Fox say to hisse'f dat he'd put up a game on Brer Rabbit, en he ain't mo'n got de wuds out'n his mouf twel Brer Rabbit come a lopin' up de big road, lookin' des ez plump, en ez fat, en ez sassy ez a Moggin hoss in a barley-patch.

"'Hol' on dar, Brer Rabbit,' sez Brer Fox, sezee.

"'I ain't got time, Brer Fox,' sez Brer Rabbit, sezee, sorter mending' his licks.

"'I wanter have some confab wid you, Brer Rabbit sez Brer Fox, sezee.

"'All right, Brer Fox, but you better holler fum whar you stan'. I'm monstus full er fleas dis mawnin',' sez Brer Rabbit, sezee.

"'I seed Brer B'ar yistiddy,' sez Brer Fox, sezee, 'en he sorter rake me over de coals kaze you en me ain't make fr'en's en live neighborly, en I tole 'im dat I'd see you.'

"Den Brer Rabbit scratch one year wid his off hinefoot sorter jubously, en den he ups en sez, sezee:

3. データ処理

(6) に示したのが各作品群のデータサイズである。この処理は awk というスクリプト言語を用いて行った。⁴

(6) 作品群ごとのデータサイズ

行数	語数	バイト数	作品群
3220	33413	171714	A
10594	109482	560618	B
2312	24356	122500	C
3256	34650	173714	D
4822	54901	275827	E
388	4284	21391	F
797	8779	44084	G
2117	24234	122500	H
775	8235	41576	I
28281	302334	1533924	Total

この表によって各テキストのおおよその規模が比較できる。なお、これらの数値には幾種類かの不要な記号類の出現回数も多少含まれているので、あくまで概数であるということを断っておく。

このコーパスは、全体として文字数約 150 万、語数約 30 万の規模である。最大のものは、1883 年出版のテキスト B であり、約 10 万語からなっており、全体の約 3 分の 1 を占めている。最小のものは、1907 年出版のテキスト F である。僅か 4 千語しかない。他に、テキスト G、テキスト I がそれぞれ小規模である。語数 8 千台となっている。このように、データサイズという点において、それぞれの物語群の間にかなりのばらつきが見られることは、今後の比較検討作業において常に留意しておく必要があるだろう。

次の段階は語形頻度表の作成である。このような目的のために開発されたソフトウェアは幾種類もあるが、今回はあえてそのような特別なソフトウェアを使用しなかった。この作業を行うにあたっては、園田 (1988) を参考にし、テキスト処理に関する機能が充実している UNIX 機を利用した。

まず、(5) のような、各物語集ごとに作成した平テキストファイルを単語ごとに (7) のコマンドを打ち込むと (9) の結果ファイルが得られる。

(7) 単語切り出しのためのコマンド⁵

```
sed -f Word < 1880.txt | sort -f > 1880word.txt
```

(8) sed 用の Word スクリプトの内容 (cf. 園田 (1998: 107))

```
s/^ *//  
s/ *$/  
#  
s/[,. -?!]" / /g  
s/[,. ?!]" ]*$//  
s/ ["() / /g  
s/^["() //  
#  
s/ ¥+/¥  
/g
```

(7) のコマンドを解釈し実行してくれるプロセス(フィルタ)は UNIX には最初から備わっている。

(7) の命令によって、1880 年出版の物語集が入っているテキスト A である 1880.txt というファイルが入力として sed に渡される。その際に、Word という名のスクリプト(8)が実行され、それによって単語がひとつずつ切り出される。さらにその切り出された単語はパイプと呼ばれる受け渡し機能——縦棒によって表記——を通じて仕分け機能を持つ sort と呼ばれるフィルタに送られる。仕分けされた結果は最終的に 1880word.txt というファイル名で保存され、一連の操作が完了する。

こうして作成した結果ファイル(9)を見てみると、各単語がアルファベット順に、出た回数だけ列べられていることがわかる。

(9)

'Ain't	'Gi'
'Ain't	'Gi'
'ar	'Gi'
'bout	'Git'
'bout	'Gree'
'bout	'Gree'
'bout	'greein'
'bout	'gun'
'bout	'gun'
'bout	'gun'
'bout	'headed'
'bout	'Hello'
'bout	'Heyo'
'bout	'How'
'bout	'How'
'bout	'How'
'bunnunce	'how'
'but	'I'
'ceppin'	'I'
'Come	'I'
'Dat	'I'
'dat	'I'
'Dat's	'I'll'
'De	'I'll'
'De	'I'll'
'Den	'I'm'
'Des	'I'm'
'Des	'I'm'
'Dey's	'I'm'
'Don't	'im'
'Ef	'im'
'Ef	'im'
'em	'im'
'em	'im'
'er	'im'
'Eve'ybody	'im'
'Fetch	'im'
'fo'	'im'

4. 作品群毎の語数

(9) の冒頭にある 'Ain't を見て分かるように、前節で説明した方法では作成したファイルには不必要な句読点や空白が含まれている。また、大文字から始まる語と小文字から始まる語が混在している。従って、語数を求める際、例えば Ef と ef は別の語として数えられてしまい正確な値が求められなくなる。そのため、エディタの置換機能等を使用して、データに残っていた不必要的記号類を除去した後、全使用語を一旦大文字へと一括変換し、その上で sort 機能を再度実行した。⁶ UNIX には同一の文字列が何回出現しているかを算出してくれる uniq というフィルタも備わっているので、先ほどの sort の出力を uniq に通し、それぞれの語形の個数を求めた。このままの出力ではアルファベット順に列べられたものが出てくるので、出現回数順の一覧を求める場合には再度 sort フィルタを通してよい。(10) はその結果ファイルの一部を抜粋したものである。

(10)

2037	AN'
1962	HE
1771	DE
1021	YOU
1004	A
1000	TER
883	THE
799	DAT
719	IT
706	I
698	IN
648	BRER
587	HIS
569	DEY
519	WUZ
495	ER
441	WHAT
441	BUT
430	FER
413	SAY
389	ON
389	AIN'T

上の手順で(10)のような結果ファイルをそれぞれの物語群ごとに作成した。

5. 語の分布

次に、全物語群を通じて使用されている語がどれほど存在するのか見てみよう。この作業を行うためには作成した 9 個のファイルの各々に挙げられている語を全て照合する必要がある。そこで、任意に選んだ 2 つの単語リストファイル間でそれらに共通して含まれている語がある場合、その語 자체を返値として出すプログラムを利用した。このプログラムを用いて 9 個全てのテキストファイルを比較していくと(11)に挙げた語が抽出される。全部で 350 語あり、以下、これを Shared Words と呼ぶことにする。

(11) Shared Words

he, brer, de, rabbit, a, the, ter, you, fox, dat, in, his, dey, up, sezee, i, wuz, er, den, on, fer, but, out, did, little, 'im, dar, wid, come, man, sez, uncle, down, to, remus, got, it, ain't, say, and, boy, miss, of, no, one, like, time, me, gwine, all, git, she, so, at, des, old, done, off, ef, see, mighty, was, b'ar, dis, be, 'roun', bimeby, make, wolf, atter, is, go, my, don't, some, um, honey, had, ez, back, tell, un, yo', day, how, by, do, hit, holler, look, i'm, twel, kaze, now, 'long, der, put, with, 'bout, said, 'fo', whar, went, well, long, folks, i'll, sot, head, dem, hear, mo', way, big, house, water, who, run, asked, him, that, kin, know, take, year, fum, keep, road, dat's, sho, right, speck, tuck, hisse'f, tail, bad, gone, what, could, dan, home, nuff, as, ax, over, have, inter, nothin', can't, good, hatter, he'd, talk, 'low, her, ketch, mouf, you'll, an, here, settin', away, fin', let, made, seed, 'twa'n't, call, out'n, this, too, do', eve'y, great, loose, which, i'd, laugh, news, place, replied, better, days, eyes, for, into, much, never, wanter, 'gun, behime, knows, mought, when, 'oman, after, comin', creeturs, eye, fambly, night, not, roun', stan', thoo, want, yit, close, doin', set, 'uz, bag, child, dunner, kinder, sump'n', then, watch, we, 'ud, full, ha'r, hard, lef', mad, name, nigh, been, ever, from, hand, paused, runnin', tale, there, wa'n't, were, b'lieve, hide, if, story, two, while, would, 'twuz, bofe, exclaimed, fix, know'd, own, show, turn, walk, allers, broke, col', face, sight, tote, will, 'tain't, bed, black, couldn't, dey'd, legs, myse'f, skeer'd, try, wait, whole, wouldn't, about, balance, fence, front, fun, fur, maybe, ner, or, seem, stop, any, drapt, dry, he'll, he'p, laughed, makin', may, meat, neck, once, remarked, responded, roll, start, tryin', why, word, worl', yo'se'f, yuther, bit, fell, quick, sech, sence, ve'y, before, creetur, door, fat, feelin', fine, flung, foots, get, hongry, it's, live, sharp, short, stidder, sun, takin', think, took, work, you'd, 'gree, befo', foun', has, ready, room, she'll, sleep, them, uv, an', beat, best, hurt, kind, mornin', move, notion, pay, wuss

この shared words が各テキストにおいてどのような出現頻度を持っているか一覧表として示したのが Appendix A と Appendix B である。A はアルファベット順にソートしたものであり、B は頻度順にソートしたものである。この表の作成は市販の表計算ソフトである Lotus 1-2-3 を用いた。該当する語にはその右隣のセルにアステリスクを付与するようにした。Appendix A と Appendix B はアステリスクのついた語のみをソートした結果であり、これ以外の部分は Non-Shared Words の分布状態を表している。

Shared Words と Non-Shared Words の平均出現率をそれぞれのテキストごとに求めたのが(12)である。

(12) それぞれのテキストにおける Shared Words と Non-Shared Words の平均出現率

Text	A	B	C	D	E	F	G	H	I
SW	0.192	0.180	0.193	0.200	0.215	0.223	0.214	0.215	0.202
NSW	0.010	0.006	0.014	0.010	0.006	0.034	0.022	0.010	0.027

Shared Words と Non-Shared Words の出現頻度は、いずれのテキストにおいても一桁もの違いが見られることがわかる。

さて、以下では Shared Words の検討を別の機会に譲り、Non-Shared Words のみに目を向けたい。まず(13)の表を見てみよう。この表が示しているのは、テキスト A において用いられている各々の語が、他のテキストにおいて、どういった頻度で出現しているかを上位 200 語分について表示したものである。

(13) 使用語彙のテキスト間における共有関係と出現頻度（テキスト A）

1880 テキスト A 総語数 10742										
出現回数	出現語	出現頻度 (%)	B	C	D	E	F	G	H	I
1618	EN	4.868	3.603	4.458	4.478	0.000	0.000	0.000	0.000	2.801
180	W'AT	0.542	0.455	0.576	0.000	0.000	0.000	0.000	0.000	0.000
154	OLE	0.463	0.543	0.445	0.422	0.007	0.000	0.000	0.021	0.436
142	W'EN	0.427	0.402	0.515	0.000	0.000	0.000	0.000	0.000	0.199
109	TARRYPIN	0.328	0.216	0.182	0.000	0.000	0.000	0.000	0.000	0.000
103	YER	0.310	0.364	0.321	0.104	0.040	0.000	0.000	0.029	0.237
95	SORTER	0.286	0.133	0.119	0.119	0.013	0.000	0.012	0.025	0.087
86	MR	0.259	0.419	0.282	0.000	0.351	0.238	0.349	0.137	0.710
71	BIN	0.214	0.367	0.185	0.029	0.007	0.000	0.000	0.000	0.087
65	BUZZARD	0.196	0.020	0.169	0.000	0.069	0.000	0.035	0.004	0.187
55	POSSUM	0.165	0.008	0.000	0.066	0.002	0.048	0.000	0.000	0.012
53	COW	0.159	0.039	0.080	0.088	0.005	0.000	0.000	0.004	0.000
44	LAY	0.132	0.103	0.095	0.054	0.026	0.024	0.000	0.058	0.062
40	GALS	0.120	0.039	0.025	0.000	0.005	0.024	0.000	0.004	0.000
40	MEADOWS	0.120	0.032	0.054	0.000	0.000	0.000	0.000	0.004	0.000
34	JUNP	0.102	0.064	0.029	0.045	0.029	0.000	0.105	0.025	0.050
34	UP'N	0.102	0.119	0.016	0.029	0.000	0.024	0.000	0.004	0.012
34	SAME	0.102	0.079	0.062	0.056	0.060	0.167	0.000	0.054	0.025
31	CONTINUED	0.093	0.105	0.045	0.036	0.007	0.000	0.012	0.008	0.037
31	W'ICH	0.093	0.064	0.025	0.000	0.000	0.000	0.000	0.000	0.000
31	'ER	0.093	0.149	0.523	0.182	0.068	0.000	0.047	0.091	0.037
31	CHILLUNS	0.093	0.029	0.012	0.057	0.000	0.000	0.000	0.004	0.112
30	OH	0.090	0.041	0.025	0.059	0.038	0.000	0.012	0.008	0.037
30	NEX'	0.090	0.050	0.082	0.064	0.027	0.000	0.047	0.058	0.037
29	DISH	0.087	0.085	0.132	0.069	0.026	0.000	0.000	0.017	0.149
28	GROUN'	0.084	0.045	0.054	0.058	0.037	0.095	0.058	0.062	0.000
28	DIDN'T	0.084	0.015	0.016	0.021	0.037	0.000	0.035	0.029	0.000
28	UDDER	0.084	0.006	0.004	0.027	0.000	0.000	0.023	0.005	0.037
26	COON	0.078	0.026	0.152	0.000	0.005	0.119	0.000	0.000	0.000
26	BLACKSMIF	0.078	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	TREE	0.075	0.078	0.128	0.119	0.040	0.000	0.023	0.050	0.112
25	DEAD	0.075	0.050	0.029	0.013	0.015	0.000	0.047	0.025	0.000
25	NUDDER	0.075	0.006	0.016	0.029	0.007	0.000	0.000	0.004	0.000
24	LAS'	0.072	0.050	0.037	0.027	0.035	0.000	0.000	0.004	0.000
24	RABBITS	0.072	0.002	0.000	0.005	0.004	0.000	0.000	0.004	0.000

24	TUCK'N	0.072	0.266	0.177	0.029	0.004	0.000	0.023	0.000	0.000
23	HIT'S	0.069	0.050	0.070	0.050	0.000	0.000	0.000	0.000	0.012
23	BULL-FROG	0.069	0.005	0.000	0.000	0.069	0.000	0.000	0.000	0.000
23	DRAP	0.069	0.047	0.082	0.053	0.016	0.024	0.000	0.017	0.012
23	FEEL	0.069	0.099	0.045	0.074	0.055	0.000	0.047	0.046	0.050
23	FIER	0.069	0.055	0.037	0.029	0.044	0.048	0.012	0.083	0.000
21	GITTIN'	0.063	0.038	0.049	0.024	0.037	0.000	0.105	0.050	0.012
21	W'ILES	0.063	0.065	0.066	0.000	0.000	0.000	0.000	0.000	0.000
21	TAR-BABY	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.000
20	STAY	0.060	0.043	0.074	0.048	0.034	0.000	0.023	0.025	0.000
19	ROCK	0.057	0.029	0.004	0.037	0.005	0.000	0.125	0.000	0.050
19	BUTTER	0.057	0.025	0.000	0.003	0.007	0.000	0.000	0.004	0.000
19	HAN'S	0.057	0.045	0.041	0.032	0.033	0.045	0.000	0.054	0.100
19	CO'SE	0.057	0.025	0.029	0.021	0.004	0.000	0.000	0.000	0.012
19	MO'N	0.057	0.034	0.012	0.024	0.000	0.024	0.012	0.004	0.012
19	EEN'	0.057	0.022	0.016	0.016	0.016	0.045	0.012	0.005	0.000
19	FETCH	0.057	0.056	0.016	0.032	0.024	0.000	0.035	0.033	0.050
18	COTCH	0.054	0.000	0.000	0.027	0.015	0.000	0.035	0.012	0.037
18	SAYIN'	0.054	0.017	0.004	0.013	0.018	0.000	0.000	0.029	0.000
18	FUS'	0.054	0.047	0.074	0.053	0.022	0.000	0.000	0.000	0.000
18	PULL	0.054	0.034	0.025	0.021	0.011	0.000	0.047	0.017	0.025
17	NICE	0.051	0.031	0.054	0.011	0.009	0.000	0.000	0.012	0.062
17	TERGEDDER	0.051	0.011	0.000	0.013	0.013	0.000	0.012	0.034	0.025
17	CUT	0.051	0.034	0.078	0.032	0.026	0.000	0.000	0.012	0.012
17	ENNY	0.051	0.002	0.004	0.000	0.000	0.000	0.000	0.000	0.000
17	MOS'	0.051	0.057	0.074	0.053	0.042	0.072	0.000	0.021	0.037
17	MIN'	0.051	0.087	0.070	0.088	0.075	0.000	0.047	0.037	0.087
17	OPEN	0.051	0.061	0.078	0.050	0.042	0.072	0.055	0.050	0.000
16	NIGGER	0.048	0.042	0.070	0.101	0.011	0.000	0.035	0.004	0.149
16	W'ILE	0.048	0.050	0.054	0.000	0.000	0.000	0.000	0.000	0.000
16	TOLE	0.048	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.087
16	STUDY	0.048	0.052	0.049	0.066	0.031	0.000	0.047	0.029	0.000
16	SING	0.048	0.083	0.070	0.008	0.005	0.000	0.000	0.012	0.062
16	BLESS	0.048	0.067	0.041	0.024	0.013	0.000	0.000	0.004	0.062
16	MAWNIN'	0.048	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	BIRD	0.045	0.006	0.008	0.016	0.024	0.000	0.012	0.203	0.000
15	MONSTUS	0.045	0.036	0.008	0.016	0.018	0.000	0.000	0.029	0.000
15	FLING	0.045	0.025	0.012	0.011	0.004	0.000	0.023	0.000	0.050
15	LIT	0.045	0.014	0.033	0.008	0.013	0.000	0.012	0.004	0.000
15	GRAB	0.045	0.028	0.008	0.013	0.011	0.000	0.000	0.008	0.012
15	HOLE	0.045	0.023	0.012	0.035	0.002	0.000	0.012	0.008	0.000
15	DEZE	0.045	0.077	0.058	0.039	0.042	0.072	0.000	0.058	0.075
14	WOODS	0.042	0.035	0.054	0.077	0.034	0.072	0.000	0.025	0.025
14	KILL	0.042	0.022	0.021	0.061	0.015	0.000	0.012	0.021	0.012
14	TU'N	0.042	0.062	0.049	0.008	0.000	0.000	0.000	0.000	0.062
14	PASS	0.042	0.030	0.045	0.011	0.016	0.000	0.023	0.041	0.000
14	LOW	0.042	0.017	0.033	0.064	0.013	0.000	0.012	0.017	0.025
14	HOL'	0.042	0.075	0.058	0.029	0.024	0.000	0.012	0.041	0.012
14	DUZ	0.042	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000

14	DOGS	0.042	0.005	0.082	0.066	0.022	0.119	0.000	0.075	0.000
14	JACK	0.042	0.190	0.004	0.003	0.002	0.000	0.000	0.000	0.037
14	STILL	0.042	0.039	0.029	0.042	0.075	0.024	0.035	0.029	0.000
14	YEARS	0.042	0.015	0.021	0.032	0.016	0.072	0.035	0.025	0.000
14	AGIN	0.042	0.000	0.000	0.003	0.002	0.000	0.000	0.000	0.037
14	WUNNER	0.042	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	MARS	0.042	0.014	0.025	0.011	0.000	0.000	0.000	0.000	0.000
14	SIS	0.042	0.072	0.008	0.011	0.011	0.000	0.000	0.000	0.000
13	MINNIT	0.039	0.017	0.012	0.013	0.013	0.000	0.035	0.021	0.012
13	OUTER	0.039	0.028	0.008	0.0000	0.018	0.000	0.035	0.012	0.012
13	MEADOWS'S	0.039	0.000	0.000	0.0000	0.000	0.000	0.000	0.000	0.000
13	SALLY	0.039	0.057	0.029	0.040	0.040	0.000	0.035	0.050	0.075
13	NOBODY	0.039	0.036	0.041	0.045	0.060	0.000	0.047	0.066	0.062
13	HOWDY	0.039	0.017	0.000	0.018	0.007	0.048	0.023	0.012	0.037
13	GOES	0.039	0.013	0.021	0.013	0.009	0.000	0.023	0.000	0.000
13	DINNER	0.039	0.022	0.016	0.040	0.026	0.000	0.140	0.008	0.037

この表において、あるテキストにおいては高い頻度をもって出現しているにもかかわらず、別のテキストにおいては用いられることのない語があることがわかる。テキスト F, G, I のファイルサイズが他のファイルと比べてかなり小さいという難点はあるが、出現頻度が比較的高いと思われる語形が急に姿を見せなくなっている場合が目に付く。

例えば、en という語はテキスト Aにおいて 4.868% (1618 回) という高い頻度を出しているにも拘わらず、E, F, G, H の 4 テキストには現れてない。W'at についてはテキスト A, B, C, D, E には現れているが、それ以降のテキストからは姿を消している。Ole はテキスト Eにおいて極端に低い頻度をとり、F と G においては皆無である。また、Hにおいても5回しか用いられておらず、かなり低い数字 0.021 となっている。W'en はテキスト D からテキスト H まで姿を消している。そこで、類似した出現傾向を示す語を 16 語選んでみた。

(14)

出現語	gloss	A	B	C	D	E	F	G	H	I
EN	AND	4.868	3.603	4.458	4.478	0.000	0.000	0.000	0.000	2.801
W'AT	WHAT	0.542	0.445	0.576	0.000	0.000	0.000	0.000	0.000	0.000
OLE	OLD	0.463	0.543	0.445	0.442	0.007	0.000	0.000	0.021	0.436
W'EN	WHEN	0.427	0.402	0.515	0.000	0.000	0.000	0.000	0.000	0.119
W'ICH	WHICH	0.093	0.064	0.025	0.000	0.000	0.000	0.000	0.000	0.000
LAS'	LAST	0.072	0.050	0.037	0.027	0.035	0.000	0.000	0.004	0.000
HIT'S	IT'S	0.069	0.050	0.070	0.050	0.000	0.000	0.000	0.000	0.012
W'ILES	WHILE	0.063	0.065	0.066	0.000	0.000	0.000	0.000	0.000	0.000
FUS'	FIRST	0.054	0.047	0.074	0.053	0.022	0.000	0.000	0.000	0.000
ENNY	ANY	0.051	0.002	0.004	0.000	0.000	0.000	0.000	0.000	0.000
W'ILE	WHILE	0.048	0.050	0.054	0.000	0.000	0.000	0.000	0.000	0.000
TOLE	TOLD	0.048	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MAWNIN'	MORNING	0.048	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TU'N	TURN	0.042	0.062	0.049	0.008	0.000	0.000	0.000	0.000	0.000
DUZ	DOES	0.042	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BIN	BEEN	0.214	0.368	0.185	0.029	0.007	0.000	0.000	0.000	0.087

この表から Harris は、ほぼ 1892 年出版であるテクスト D (Old Plantation Stories, Songs, and Ballads with Sketches of Negro Character)あたりから当該使用語の綴りを変えていることがわかる。そしてこの変化に伴って何か別の事態が生じているように思われる。例えば、(14) の各出現語に対する異綴形とされるものおよび完全形の出現状況を Appendix A の表を検索してみると (15) に示すような結果となる。

(15) 異形の出現状況

異形	A	B	C	D	E	F	G	H	I
AN'	0.003	0.076	0.193	0.005	3.722	5.054	4.302	4.185	1.195
AND	0.042	0.585	0.374	0.350	0.657	0.238	0.453	0.534	0.473
WHAT	0.102	0.064	0.074	0.756	0.805	0.596	0.732	0.761	0.709
OL'	0.000	0.000	0.000	0.000	0.467	0.500	0.337	0.579	0.149
OLD	0.285	0.171	0.185	0.169	0.164	0.071	0.139	0.178	0.435
WHEN	0.048	0.057	0.617	0.499	0.626	0.739	0.558	0.770	0.453
WHICH	0.057	0.044	0.016	0.074	0.073	0.071	0.011	0.115	0.049
LAST	0.012	0.004	0.004	0.000	0.014	0.023	0.011	0.012	0.024
IT'S	0.009	0.011	0.024	0.042	0.122	0.119	0.046	0.111	0.037
WHILE	0.030	0.050	0.024	0.069	0.069	0.071	0.046	0.070	0.049
WHILES	0.000	0.000	0.000	0.098	0.063	0.007	0.069	0.053	0.037
FIRST	0.003	0.004	0.004	0.000	0.014	0.000	0.000	0.004	0.000
ANY	0.015	0.056	0.102	0.069	0.785	0.095	0.011	0.107	0.037
TOLD	0.009	0.006	0.016	0.013	0.029	0.000	0.046	0.033	0.024
TOL'	0.000	0.000	0.053	0.103	0.042	0.071	0.197	0.086	0.037
MORNING	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000
MORNIN'	0.003	0.005	0.032	0.045	0.032	0.095	0.046	0.033	0.062
TURN	0.027	0.004	0.016	0.047	0.038	0.023	0.058	0.049	0.012
DOES	0.000	0.054	0.075	0.055	0.098	0.119	0.069	0.070	0.062
BEEN	0.033	0.043	0.016	0.183	0.374	0.190	0.348	0.343	0.149

そして (14) に挙げた語形の頻度と (15) に挙げた異形の頻度を突き合わせてグラフにプロットしてみると (16) のようになる。

(16) 出現語形の頻度推移

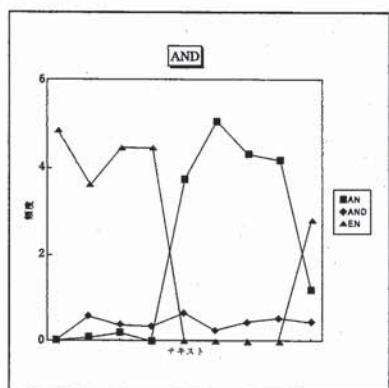


fig.1

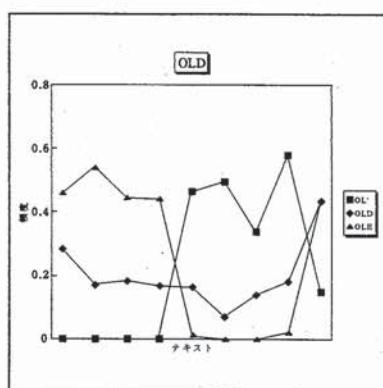


fig.2

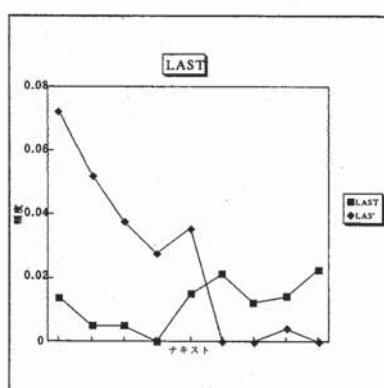


fig.3

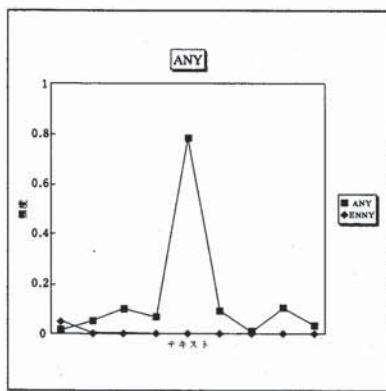


fig.4

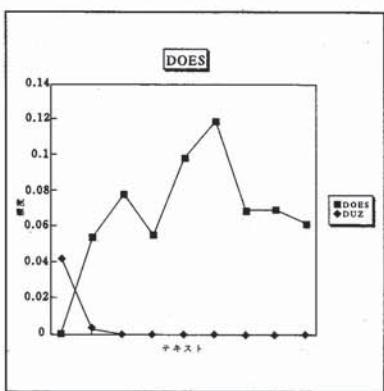


fig.5

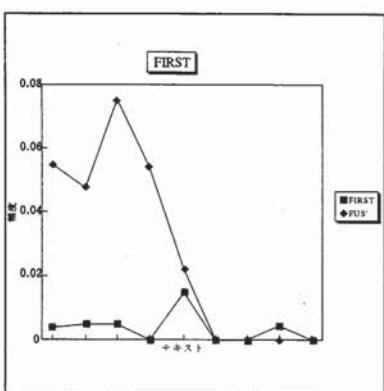


fig.6

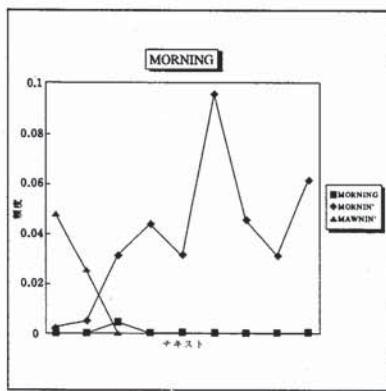


fig.7

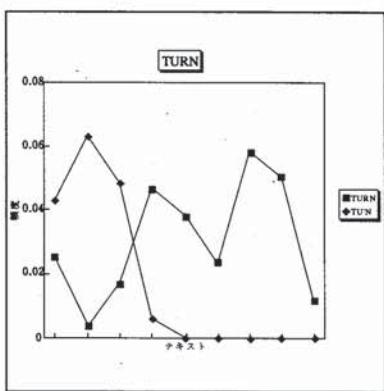


fig.8

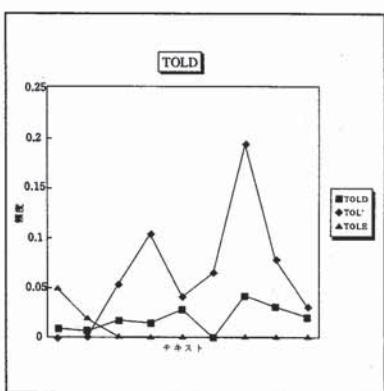


fig.9

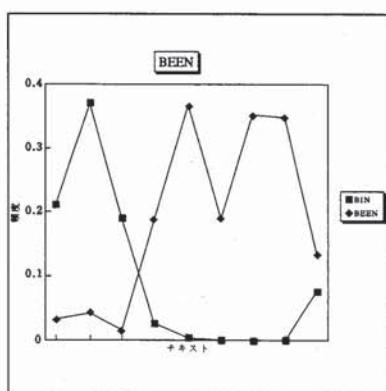


fig.10

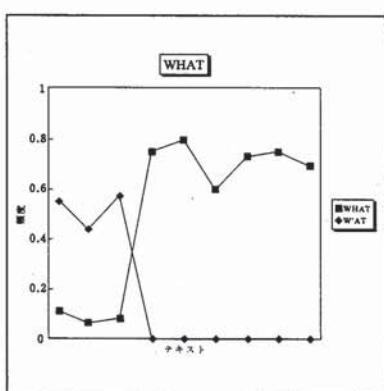


fig.11

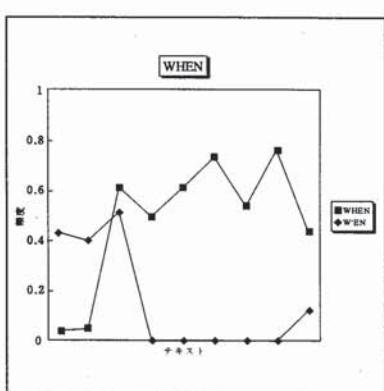


fig.12

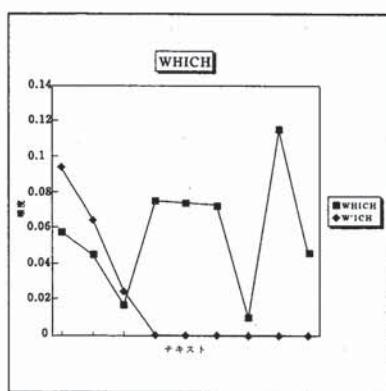


fig.13

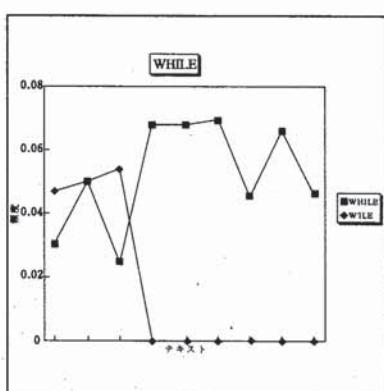


fig.14

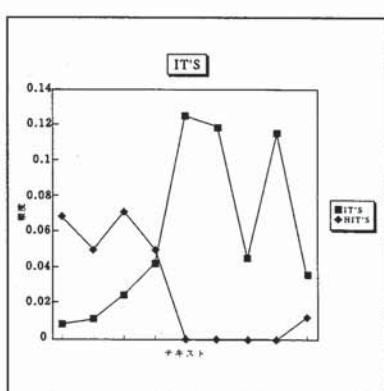


fig.15

いずれのグラフにおいても、横軸の9個のポイントはそれぞれのテキストに相当しており、左から A, B, C という順で並んでいる。縦軸は出現頻度となっている。

まず and の異綴形である en から見てみよう。完全形である and の頻度は急な変化も見られず全般的に低いといえるが、他方、en は最初の4テクストにおいては高い頻度を示しているにも拘わらず、その後テクスト I になるまで全く使用されていない。ここでそのギャップを埋めるように an' がテクスト E 以降出現数を伸ばしている。old についても同様の傾向が見られる。この語の頻度については、完全形の場合、最後のテキストを除きほぼ 0.2 である。ole と ol' に目を向けると、やはりテクスト E を境に状況が逆転している。次に las' についてであるが、グラフの上ではテクスト F の時期から、完全形と t が省略されている語形の頻度が逆転しているようである。しかし、頻度が低非常に低いこともあり、このデータからは決定的な傾向は見出せない。enny と any の分布についても enny 自体が使われる頻度が少なすぎることもあり、傾向が見られるような結果は出ていない。duz もまた最初の作品を除きほとんど用いられていない。fus' については、完全形の first がテクスト E まで比較的高い頻度で用いられているが、後のテクストにおいて、急に頻度が下がりほとんど使われなくなっている。この点では他の語形の示す傾向と異なっているように見えるが、fus' のほうも一貫して使用度も低いので判断材料にはできない。mawnin' についても母音の後の r を欠いた語形が初期の2つのテクストで使用されているものの、その後は途絶えてしまっている。完全形の morning は一貫してほとんど用いられていない。その空白を補うようにして出現しているのが、語末の g の省略以外標準的綴りをとっている mornin' という語形であり、テクスト C 以降で専ら用いられている。turn の場合も r が省略された tu'n という異綴形が見えるが、これも前期のテクストにおいては完全形よりも高い頻度で用いられているにも拘わらず、テクスト E から姿を消している。tole とその異綴形である tol' のについては mawnin'/mornin' の場合と同じく、テクスト C において頻度の交替が始まっている。完全形である told は似たような頻度を保っているが、これは語りの部分以外で用いられるためであると思われる。bin はテクスト D を境に完全形 been にその座を奪われている。

最後に、最も興味深いと思われる点、つまり、w'at /what, w'en/when, w'ich/which, w'ile/w'iles/while 及び hit's/it's といった語における氣息音の存在に相当する綴り字上の h の有無とテクストとの関係について触れておきたい。what の場合、初期のテクストにおいては完全形、h の無い語形とも使われており、w'at のほうが頻度が高いことは明らかである。にもかかわらずテクスト D になるとその語形の使用は止み、現れるのは完全形だけである。when, while についても、Harris はテクスト D から完全形のみの使用に切り替えている。似た傾向は hit's/it's の交替についても見られる。テクスト D では、両形式ともほぼ同じ頻度で用いられているにも拘わらず、テクスト E 以降は、帶氣音を綴りにもつ古い語形の hit's は消失し、最後のテクストにおいて僅かに見つかる程度となっている。⁷

ここで再度 Ives (1955) が当該方言の特徴として (2) で述べている "the omission of h in words like *what and why*" という点から考えてみると、確かに初期に書かれた4つのテクストを対象にするならばこの点は間違いなく Harris の用いた綴りに反映されている。しかし、その後の作品においては必ず

しも言語事実に忠実な表記法を用いているわけではないようである。例えば、Kenyon がアメリカ英語の音声について論じた研究書として *American Pronunciation* があり、1924年の初版となっている。これはリーマスじいや物語の出版時期とほぼ重なっており、当然ながら当時のアメリカ英語の音声がその考察の対象である。ここでは 30 年近く経過した 1951 年発行の第 10 版を参照したが、標準的なアメリカ英語の発音では、早口であっても h が保持されること、標準英語から教わっていない方言の話者が、保持すべき h を落として発音するといった旨の記述が見られる。しかし、it に対して hit という形がアメリカ南部方言では普通に見られたということも記されている。このことから、後半時期の物語を執筆していた時期でさえ、当該方言においては wh 語において h 消失は一般的であり、hit に関しては h を失わずに使用されていたと推測される。

このように、はっきりと見えてきたことは、何故か Harris の方言表記に対する忠実さというものがある時期から失われてしまい、特有の訛りを映し出していた綴りを標準的な綴りと一部取り入れる、または、すっかり置き換えてしまうという結果になっていることである。このことは、データの底本とした全作集に、Mr. Harris would experiment, searching for better ways of experimenting this elusive and fluid dialect. If it is possible to generalize, it would seem that he made more of an effort to spell phonetically in his early books, such words as *bawn* and *mawnin'* later becoming *born* and *mornin'*. — 'A Note on the Text', xvii. と記されていることに改めて注目する結果となった。つまり、上で実験的に行った数語の出現頻度推移についての検索結果は、小規模からも Harris の全作集の編纂者が抱いた直観を支持していると言えそうである。

6. まとめ

以上、本研究の前半では、Harris の「リーマスじいや物語」全文を機械可読であるテキストファイル化し、データとして活用する方法論を解説した。後半では、そのデータをもとに、各語形の出現頻度と執筆時期との関係を検討した。その結果、少なくとも音声学的観点からは、実際の言語状況を反映していないと思われる綴りが途中から急に導入されていることが判明した。同じ作者による形式的にもよく似た物語集であるが、作者の用いた言語に対する意識が変わってしまい標準英語の色合いを強めているといえるのではないだろうか。

今後の課題として、音声学以外の側面について、今回テキストファイル化した全作集がどれほどまで linguistic evidence として我々に教えてくれることがあるか検討を進めたいと考えている。この点については Bickley and Hugh (1997: 63) の示唆が興味深い。

(17) The speech of lower class whites and that of slaves in the mid-nineteenth century shared many similar elements. The only difference lies in the greater frequency of such elements in the speech of slaves. (Bickley and Hugh (1997: 63))

この点を探索するためには平テキストファイルに対して文法及びその他情報を記したタグを付与する作業が欠かせない。どのようなタグを設定しどのように割り付けするかいうことが次の課題への入り口であるといえるだろう。

注

*本論考は日本英文学会中四国支部第51回大会（於 松山東雲女子大学）での発表内容を加筆修正したものである。司会の山本和之先生（山口大学）ならびに貴重なコメントを下さった中尾佳行先生（広島大学）に対して感謝いたします。また、文字列の検索プログラムの作成にあたって、名和真一先生（高知女子大学）にお世話になりました。ここに記して感謝申し上げます。

¹<http://xroads.virginia.edu/~UG97/remus/contents.html>

²斎藤秀夫氏開発のシェアウェア。文字列検索機能が使いやすくできている。

<http://hidemaru.xaxon.co.jp> からダウンロード可能。

³文字検索において、当該行だけでなく、前後の行を含めて出力してくれる機能が備わっている。行数は任意に指定可能。

⁴この作業を行う awk スクリプトは植村・富永（1993：156）に掲載されているものを参考にした。もとのスクリプトは日本語の語数計算にも対応させたものであるが、英文のテキストファイルを処理するように簡略化した。

```
{  
    c[FILENAME] += length($0)  
    w[FILENAME] += NF  
    l[FILENAME]++  
}  
END {  
    for (f in c) {  
        print("%d\t%d\t%d\t%s\n", l[f], w[f], c[f], f)  
        tc += c[f]  
        tw += w[f]  
        tl += l[f]  
    }  
    printf("%d\t%d\ttotal\n", tl, tw, tc)  
}
```

⁵このコマンド及びスクリプトを解釈し実行してくれる sed や sort といったプログラムは最初から UNIX には最初から備わっている。このような機能を利用するとテキストファイルの加工等を特別なソフトウェアを購入しなくても比較的簡単な操作で行うことができる。MS-DOS 上で動く同様のツールもフリーウェアとして入手可能であるので、Windows ユーザであれば、DOS 窓を開いて同じ検索作業をすることも可能である。

⁶小文字を大文字に変換するために用いたコマンドは次のものである。

```
tr a-zA-Z <1880word.txt> 1880.WORD.TXT
```

tr コマンドは任意の文字（列）を別の任意の文字（列）に変換する。ここでは、1880word.txt というファイルを取り込み、アルファベットの小文字（a-z によって表される）を大文字（A-Z によって表される）に全て置き換え、その結果を 1880WORD.TXT というファイルに書き込むという操作が行われる。

⁷ テキスト I に関しては使用されている語の傾向が他のテキストと一致しない点が少なからず見受けられる。Harris の没後に出版された物語集であるため執筆時期の異なったテキストが混在している可能性があるように思われる。

References

- Bickley, R. Bruce and Hugh T. Keeman (1997) *Joel Chandler Harris: An Annotated Bibliography of Criticism, 1977-1996, with Suppliment, 1892-1976*. Westport, CT: Greenwood Press.
- Chace, Richard. (comp.) (1955) *The Complete Tales of Uncle Remus*. Boston: Houghton Mifflin Company.
- Dillard, Joey L. (1972) *Black English: Its History and Usage in the United States*. Vintage Books edition, 1973. New York: Vintage Books.
- Ellis, Michael. (1994) "Literary Dialect as Linguistic Evidence: Subject-Verb Concord in Nineteenth-Century Southern Literature." *American Speech*, 69.2. 128-144.
- ハリス, J. C. 作, 八波直則 [訳] (1953) 『ウサギどんキツネどん』 岩波書店.
- Ives, Sumner. (1955) "Dialect Differentiation in the Stories of Joel Chandler Harris." *American Literature*, 17. 88-96.
- Kennyon, John S. (1951¹⁰) *American Pronunciation*. Ann Arbor, Michigan: George Wahr Publishing Company. 竹林滋 [訳] 『アメリカ英語の発音』 大修館書店, 1973.
- Macaulay, Ronald K. S. (1991) "'Coz it Izny Spelt When They Say it': Displaying Dialect in Writing." *American Speech*, 66.3. 280-291.
- Miyoshi, Akihiko. (1998) *Grep, Sed, Awk: Manual & Reference*. Tokyo: Shuwa System Co., LTD.
- 長瀬眞理・西村弘之 (1986) 『コンピュータによる文章解析入門』 オーム社.
- 岡田毅 (1995) 『実践「コンピュータ英語学」』 鶴見書店.
- 斎藤俊雄 [編] (1992) 『英語英米文学研究とコンピュータ』 英潮社.
- 斎藤俊雄・中村純作・赤野一郎 [編] (1998) 『英語コーパス言語学』 研究社.
- 園田勝英 (1998) 「コーパスに基づく語彙研究」 斎藤・中村・赤野 (1998) 105-122.
- Stoller, Paul. (ed.) (1975) *Black American English: Its Background and Its Usage in the Schools and in the Literature*. New York: Dell Publishing.
- SE 編集部 (1992) 『MS-DOS テキスト料理学— sed, awk のある UNIX 流パソコン環境』 翔泳社.
- 植村富士夫・富永浩之 (1993) 『awk でプログラミング』 オーム社.

Appendix A (アルファベット順にソートしたもののリスト)

Appendix B (頻度順にソートしたもの)

(いおり たかひろ・本学助教授)