

## 論文

# 日本語同音語のアクセントと音節構造

徳 弘 康 代

A statistical analysis on the relationship between  
accentuation and syllabic structure of Japanese homophonous words  
TOKUHIRO Yasuyo

### 目次

#### 要旨

1. 統計分析のための資料
2. 日本語の同音語の性質
3. 単語のアクセントと音節構造
4. まとめ

#### 参考文献

## 要約

日本語の同音語のアクセント型と音節構造について、統計的データにもとづいて解析した。

EDR（日本電子辞書研究所）が編集した日本語単語辞書のうちの、約30万語の普通名詞を統計的分析に使った。そのうち、外来語を含む単語は削除した。そして、同じ発音と意味を持つ単語の候補は、代表する単語一語に統合した。それら124,000語に統合された単語集合が、統計分析のために使われた。

統合された単語集合の約35%の単語が、同音語の組に属していた。同音語の組の数は、組数が増すほど、その組内の同音語数は減少し、それらはお互いに正確に反比例していた。

この分析において、日本語の音節は、先行子音（+半母音）+後続母音に、モーラ音素が付属する/しないものと定義した。ここで採用したモーラ音素は、長母音/H/、促音/Q/、および音節鼻子音（撥音）/N/である。

分析の結果、アクセントが付く確率は、母音型の音節（Vs）には平均13%、半母音+母音型の音節（yVs）には16%であった。どちらのケースにも、後続母音の種類による際立った差はなかった。モーラ音素の影響については、モーラ音素が付いた音節にアクセントが付く確率は、後続する半母音や母音の種類によらず、平均より数%高くなった。母音型の音節と半母音+母音型の音節を合わせると、アクセントの付く平均的な確率は14%であったが、どの種類の先行子音の音節でも、モーラ音素が後続する音節の確率は、平均より数%高かった。分析に使った単語集合に含まれる約49万音節のうちでモーラ音素が後続する音節は30%以上もあった。さらに、モーラ音素が後続する音節は、組の中の語数が非常に多い同音語の音声に多く現れた。日本語の同音語は、全体の五十音順の表と、同音語の組の中の単語数や単語を構成する音節数に分けて五十音順に配列した表に編集した。モーラ音素を含む単語は、モーラ音素の種類別に、五十音順の表と、

単語を構成する音節数に分けて五十音順に配列した表に編集した。統計分析のこれらの結果は、言語処理の基礎研究だけでなく語学教育および会話訓練のための有益なデータとなる。

## 1. 統計分析のための資料

EDR（日本電子辞書研究所）により編集された日本語単語辞書を、この統計分析のために使った。

### 1.1. EDR日本語単語辞書の普通名詞

EDR日本語単語辞書の、全体の「単語見出し」の約41万語から、約70%を占める約30万語の普通名詞を抽出した。次に、その15%を占める外来語を除いて、残りの「単語見出し」約256,000語を統計分析の元の資料として使った。

### 1.2. 同音語の統合の過程

約256,000語の「単語見出し」の約80%は、同じ発音の単語のグループに属していた。しかし、それらのグループの単語のすべてが、いわゆる同音語であるわけではない。見出し語5～7万語程度の一般的な国語辞典では、一つの見出し語として扱われている単語が、EDRの「単語見出し」では、意味や表記の違いによって、細かく分かれている。これはEDRの日本語単語辞書が自然言語処理関連の需要により作成されているため、意味の違いによって語を一つ一つ詳細に区別しているからである。そこで、次のような段階を踏んで「単語見出し」の統合の処理をした。

段階1. 「発音」（「かな表記」ではない）が重複した「単語見出し」のグループに、同じ「概念識別子」（単語の概念の最小単位）の部分組があるならば、代表として「概念別頻度」が最も高いものを残す。「概念別頻度」が同じ場合には、「単語見出し」の最後の行を

残す。

段階 2. 「発音」が重複している「単語見出し」の組の中で、「概念見出し」（「概念識別子」よりは大きいまとまりであるが、一般的な辞書の見出しよりは細かい単位）が同じ部分組があれば、段階 1 と同じ方法で代表を残す。

段階 3. 「発音」と「単語見出し」が同じ組の中から、代表として「概念別頻度」が最も高いものを残す。「概念別頻度」が同じ場合には、EDRの「レコード番号」の順で最後の行を残す。

統合の経過を図 1 に示す。段階 1 で元の「単語見出し」の 256,000 語が約 30% 減って 178,000 語に、段階 2 で 18% 減って 147,000 語に、段階 3 で 16% 減って、約 124,000 語になった。これは元の「単語見出し」の約 50% である。削除過程の具体例を表 1 に示す。このように統合された単語集合が以下の統計分析のために使われた。

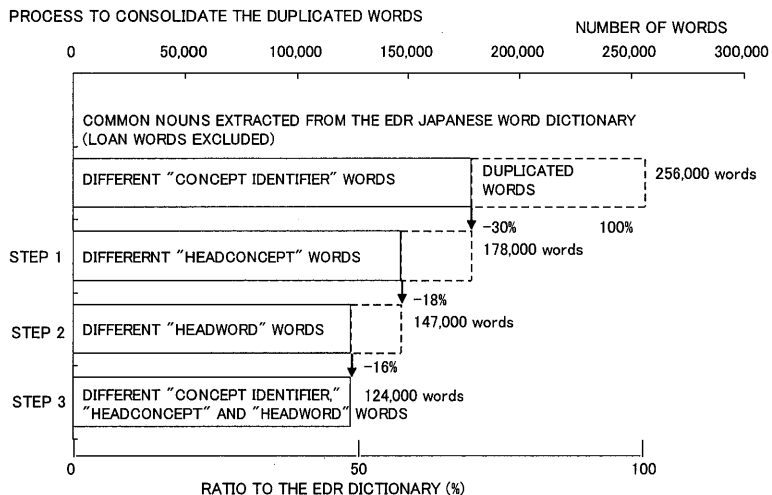


図 1. 見出し語重複の削除過程

表 1. 見出し語重複削除の例

AN EXAMPLE OF THE PROCESS OF CONSOLIDATING THE EDR HEADWORDS

RECORD NO.	HEADWORD	PRONUNCIATION		HEADCONCEPT JAPANESE	ENGLISH	PROCESS				
						MAIN	STEP1	STEP2	STEP3	
JWD0257498	買い子	カイク	0	1f1dff	買い子[カイク]	PURCHASING AGENT	○			
JWD0257499	買子	カイク	0	1f1dff	買い子[カイク]	PURCHASING AGENT	○	○		
JWD0425223	買い子	カイク	0	3c62c9	買い子[カイク]	PURCHASING AGENT	○			
JWD0425224	買子	カイク	0	3c62c9	買い子[カイク]	PURCHASING AGENT	○	○	○	○
JWD0548703	回顧	ガ'イコ	1	3d0c34	回顧[カイク]	REVIEW	○	○	○	
JWD0517475	回顧	ガ'イコ	9	3c6da6	回顧する[カイク・スル]	REVIEW	○	○	○	○
JWD0036237	解雇	ガ'イコ	0	0e8908	解雇する[カイク・スル]	FIRE	○	○	○	
JWD0543791	解雇	ガ'イコ	62	3d032f	解雇する[カイク・スル]	FIRE	○	○	○	○
JWD0036263	懐古	ガ'イコ	4	0e890a	懐古する[カイク・スル]	REMINISCENCE	○	○	○	○
JWD0036175	蚕	カ'イコ	16	0e88cf	カイク[カイク]	SILKWORM	○	○	○	○
CONCEPT IDENTIFIER							10	8	7	5
CONCEPT OCCURRENCE FREQUENCY							NUMBER OF HEADWORDS			

## 2. 日本語の同音語の性質

### 2.1. 同音語の組数と組の中の単語数

統合された単語集合に含まれている同音語の組数と、組の中の単語数との関係を調べると、図2に示すようになる。同音語の組の中の単語数は、この単語集合では、2語から37語にまで及んでいる。組数は組の中のそれぞれの単語の数に応じて減少するが、このような多量の単語集合では、それらが正確に反比例していることを表している。同音語の組内の単語数の多い単語集合を表2に示し、表3にその一例を示す。同音語の組内の単語数が10語以上の語の1、2音節語の音節別の単語数を表4に示す。

同音語を持たない単語は81,000語で、全体の124,000語の約65%である。したがって、残りの約35%の単語が同音語の組に属していることになる。この割合は、国立国語研究所1989の調査の、助辞を除いたM単位の同音語の異なり語数の割合の34.9%に近いものである。

GROUPS OF HOMOPHONOUS WORDS

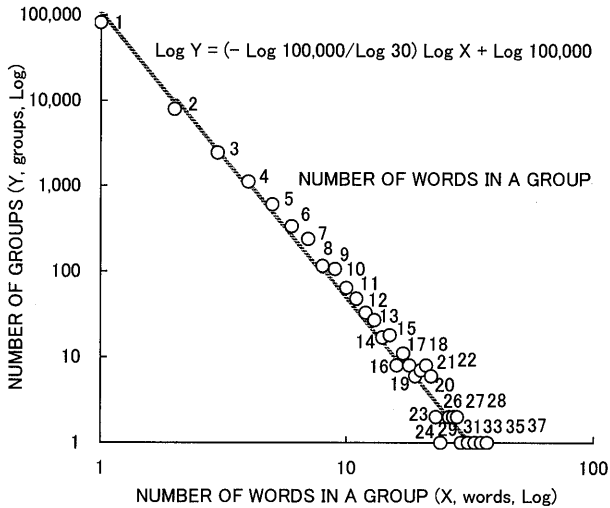


図 2. 同音語の組数と、組の中の単語数との関係

表 2. 組内の単語数の多い同音語

EXAMPLES OF GROUPS HAVING A LARGE NUMBER OF HOMOPHONOUS WORDS

RANKING	KANA NOTATION	PHONEMIC SYMBOL	NUMBER OF HOMOPHONOUS WORDS
1	コーシ	koH si	37
2	コーショー	koH syoH	35
3	カン	kaN	33
4	コー	koH	31
5	セーシ	seH si	29
6	ショー	syoH	28
7	シセー	si seH	28
8	セーコー	seH koH	27
9	センコー	seN koH	27
10	キコー	ki koH	26
11	ショーカ	syoH ka	26
12	シコー	si koH	24
13	カシ	ka si	23
14	コーキ	koH ki	23
15	シ	si	22
16	コーカ	koH ka	22
17	コーシン	koH siN	22
18	コーセン	koH seN	22
19	ショーシ	syoH si	22
20	シンコー	siN koH	22
21	コーコー	koH koH	21
22	コージ	koH zi	21
23	コーゾー	koH soH	21
24	コーテー	koH teH	21
25	シショー	si syoH	21
26	セーカ	seH ka	21

表 3. 組内の単語数の多い同音語の例

EXAMPLE OF A GROUP OF HOMOPHNOUS WORDS (キコー ki koH)

KANJI	ENGLISH MEANING
貴公	YOU
貴校	YOUR SCHOOL
機構	THE STRUCTURE OR FRAMEWORK OF SOMETHING
奇功	AN ACHIEVEMENT THAT WAS NOT EXPECTED
奇効	OF SOMETHING, TO HAVE A REMARKABLE EFFECT
奇巧	OF HANDIWORK, A CHARACTERISTIC OF BEING RARE AND INGENIOUS
奇行	AN ECCENTRIC ACTION
奇香	A UNIQUE SMELL
寄港	TO STOP BY SOMEWHERE ON THE WAY TO ONE'S FINAL DESTINATION
寄稿	A CONTRIBUTED MANUSCRIPT
機巧	A DEVICE THAT IS MOST INGENIOUS
機甲	THE ACT OF REARMING WITH NEW WEAPONS
帰校	OF A PERSON, TO COME BACK TO SCHOOL
帰港	OF A SHIP, TO RETURN TO PORT
帰耕	THE ACT OF RETURNING TO WORK AT THE FARM
帰航	THE RETURN ROUTE OF ONE'S VOYAGE
氣候	THE LONG-TERM WEATHER CONDITIONS IN AN AREA
氣孔	STOMA
季候	THE WEATHER CONDITIONS OF A PARTICULAR SEASON
稀覯	A RARE SIGHT
紀綱	THE FUNDAMENTAL PRINCIPLES UNDERLYING A GOVERNMENT
紀行	A RECORD ONE'S TRAVELS
起工	THE ACT OF BREAKING GROUND IN ORDER TO BEGIN BUILDING
起稿	TO WRITE OUT A ROUGH COPY
輝光	SPARKLING LIGHT
騎行	TO GO SOMEWHERE ON HORSEBACK

表 4. 同音語の組内単語数が10語以上の語の音節別の単語数

ONE AND TWO SYLLABLE WORDS OF GROUPS OF HOMOPHONOUS WORDS, AND THEIR NUMBER OF HOMOPHONOUS WORDS IN THE GROUP (GROUPS CONSISTED OF MORE THAN 10 HOMOPHONOUS WORDS ARE SHOWN)

\_\_\_\_ : 1ST AND 2ND SYLLABLES ARE THE SAME

	1ST SYL	ko(ə)H	si(ə)	syo(ə)H	ka(ə)N	se(ə)H	ka(ə)	ka(ə)i	so(ə)H	ki(ə)	se(ə)N	si(ə)N	ta(ə)i	to(ə)H
2ND SYL GROUPS		38	18	14	12	12	11	9	9	8	7	7	6	6
WORDS		684	305	232	198	194	171	128	124	134	113	111	87	76
NONE	25	koH	31	siə	syoH	ka@N	se@H	ka@	15	19	12	19	15	15
	443				28	33	14	15	19	12	19	15	15	14
si(ə)	22	koH si	siə si	syoH si	ka@N si	se@H si	ka@si		soH si		se@N si	si@N si		
	361	37	19	22	14	29	23		13		15	19		
ko(ə)H	20	koH koH	si koH	syoH koH	kaN koH	seH koH	ka@ koH	kai koH	soH koH	ki koH	seN koH	si@N koH	tai koH	toH koH
	352	21	24	17	18	27	20	18	20	26	27	22	20	11
se(ə)H	15	koH seH	si seH	syoH seH	kaN seH		ka seH		soH seH	ki seH		siN seH	tai seH	
	231	19	23	12	18			19		16	17		18	13
ka(ə)	15	koH ka	si ka@	syoH ka	ka@N ka	se@H ka			soH ka	ki@ ka	se@N ka		ta@ ka	to@H ka
	230	22	16	26	16	21			12	11	14		13	17
syo(ə)H	14	koH syoH	si syoH		kaN syoH		ka syoH	kai syoH		ki syoH	seN syoH	siN syoH	tai syoH	
	234	35	21		19		14	14		15	14	14	16	
ki(ə)	9	koH ki	si@ ki	syoH ki		se@H ki	ka@ ki			ki@ ki		si@N ki		to@H ki
	141	23	20	15		19	14				17		11	11
ka(ə)N	9	koH kaN	si kaN	syoH kaN		seH kaN			soH kaN	ki kaN			tai kaN	
	132	18	20	16		11			12	18			13	
to(ə)H	9	koH toH		syoH toH	kaN toH	seH toH		kai toH	soH toH			siN toH		
	125	20		12	14	12		14	15			12		
zyo(ə)H	7	koH zyoH	si zyoH	syoH zyoH	kaN zyoH		ka zyoH	kai zyoH			seN zyoH			
	101	20	14	14	15		11	12			15			
si(ə)N	7	koH siN	si siN	syoH siN				kai siN						toH siN
	95	22	12	13					13					12
so(ə)H	6	koH soH	si soH		kaN soH	seH soH		kai soH		ki soH				
	94	21	19		15	13		15		11				

GROUPS: NUMBER OF GROUPS OF HOMOPHONOUS WORDS WHICH INVOLVE THE SYLLABLE (SYLLABLES INVOLVED IN MORE THAN 5 GROUPS ARE SHOWN)  
 WORDS: SUM OF THE NUMBERS OF HOMOPHONOUS WORDS IN EACH OF THE GROUPS

## 2.2. 同音語の単語の音節構造

### 2.2.1. 日本語音節の定義の修正

この統計分析において、日本語の音節は、伝統的な「先行子音（＋半母音）＋後続母音」ではなく、「先行子音（＋半母音）＋後続母音、の後ろにモーラ音素が付く／付かない」として定義した。日本語の東京方言の単語アクセントはモーラ音素には付かないので、この定義による音節は、東京方言の単語アクセントがひとつ付加され得る最小区分となる。この区分は、後の、モーラの単語アクセントへの影響についての分析に役立つ。ここで使われた子音の種類は、/#/（先行子音なし）と、/s/、/k/、/z/、/r/、/g/、/h/、/b/、/n/、/m/、/p/、/d/（ダデドの音節のみ）とした。母音は、/a/、/i/、/u/、/e/、/o/、半母音は/y/（ヤユヨ行）と/w/（ワ）である。

モーラ音素は、/H/（母音の長音化した部分）、/N/（音節鼻子音（撥音部分））、/Q/（声に出さない子音に進む促音化した部分）である。二重母音化された部分/I/は、この分析では扱われていない。

### 2.2.2. 単語の音節の数

単語の音節の数は、同音語の組の単語の数が増えるのに従って減少した。22語を超える単語を持つ20組では、それらの80%は2音節語であり、残り20%は1音節語であった。

### 2.2.3. 単語の先行子音

この単語集合の音節数は、母音のみが続くもの（Vs）が436,000音節、半母音＋母音が続くもの（yVs）が55,000音節で、合わせると491,000音節あった。

先行子音の種類の出現頻度を、VsとyVsに分けて、図3に示す。



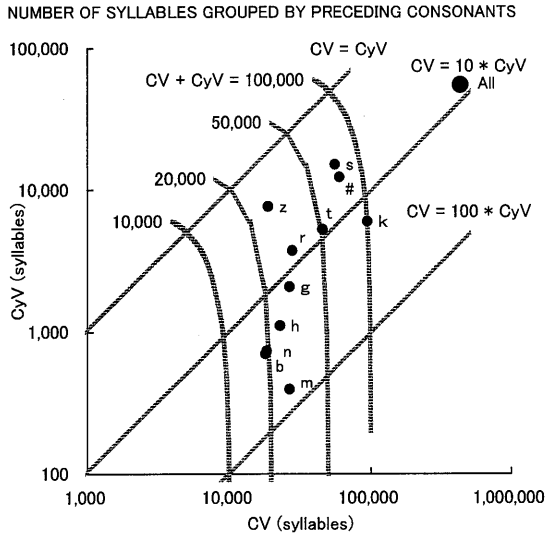


図3. 母音型の音節と半母音+母音型の音節の出現頻度の関係、先行子音による分類

yVsは平均でVsの約1/10である。yVsの値は、先行子音が/z/の場合Vsに近かったが、/m/では約1/100であった。子音の種類による出現頻度は、/k/が100,000音節、/#/が72,000音節、/s/が71,000音節、次いで/t/が51,000音節であった。/z/、/r/、/g/、/h/、/b/、/n/、/m/などの出現頻度は、/k/の1/5程度に少ない。

多くの同音語を持つ20組に含まれている36種類の音節では、先行子音は/k/と/s/だけしかなく、両方50%であった。モーラ音素の/h/は1/2の音節についており、/N/は1/6の音節についていた。これらの/k/と/s/の出現頻度は非常に大きい。なぜなら、統合された単語集合の約491,000音節の中では、両方とも1/5ぐらいだからである。また、/k/と/s/とほとんど同ぐらい高い出現頻度を持つ先行子音の/#/や/t/が含まれていないことは注目に値する。また、モーラ音素/h/の出現頻度の1/2も、平均出現頻度である1/9に比べて高すぎる。

さらに、これらのすべてが漢字（日本語の正書法における漢字）であった。この割合は、統合された単語集合の中の漢字だけの単語の割合よりずっと多いことから、中国語の発音の影響を受けているものと考えられる。

#### 2. 2. 4. 単語アクセント型による区別

これらの同音語の組の中では2つ（または3つ）の単語アクセント型による部分組が見られた。しかし、大きい方の部分組が組の同音語の単語の80%より多くを占めていたので、アクセント型による区別は効果的ではなかった。

### 3. 単語のアクセントと音節構造

#### 3. 1. アクセントの付く確率

アクセントの付いた音節と付かない音節との割合を後続母音の種類別に分類したものを、図4に示す。Vsの出現頻度で/a/、/i/、および/o/が多く、/e/と/u/は少なかったが、違いは1/2以内であった。一方、yVsでは、違いはより大きく、/yo/、/yu/、/ya/の順に1/2ずつ減少した。アクセントの付くVsの平均は13%で、yVsは16%であった。いずれの場合でも、後続母音の種類による差は顕著ではなかった。

#### 3. 2. モーラ音素を伴う音節

モーラ音素が付属する音節の全音節に対する比率を母音の種類別と子音の種類別に図5に示す。次に、アクセントが付くか付かないかによる音節の出現頻度の関係をモーラ音素が付くか付かないかに分けて、VsとyVsの音節の種類別に、図6に示す。

NUMBER OF SYLLABLES GROUPED BY FOLLOWING VOWELS

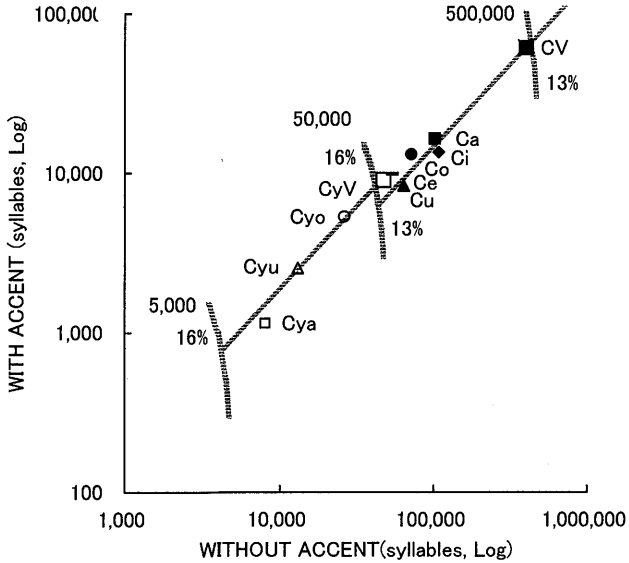


図 4. 母音型の音節と半母音+母音型の音節の出現頻度の関係, 母音による分類

RATIO OF NUMBER OF SYLLABLES ACCOMPANIED BY MORA PHONEMES TO ALL

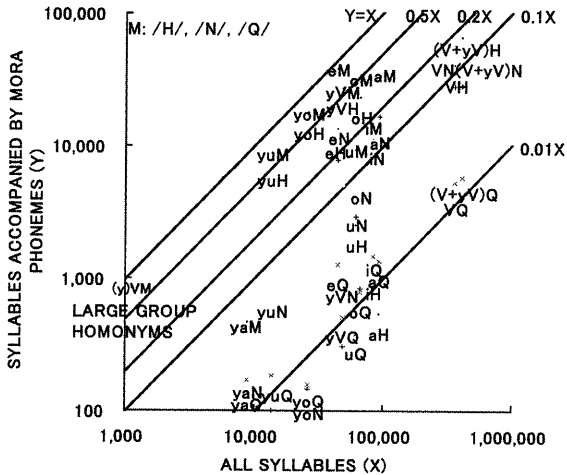


図 5 - 1. モーラ音素が付属する音節の全音節に対する比率, 母音の種類別



NUMBER OF SYLLABLES  
GROUPED BY FOLLOWING VOWELS

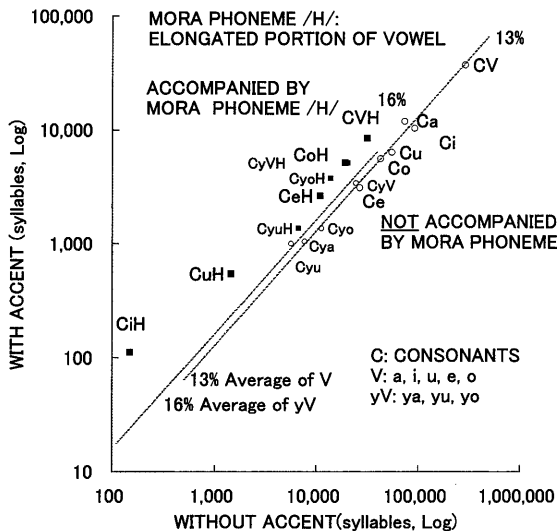


図 6 - 2. /H/

NUMBER OF SYLLABLES  
GROUPED BY FOLLOWING VOWELS

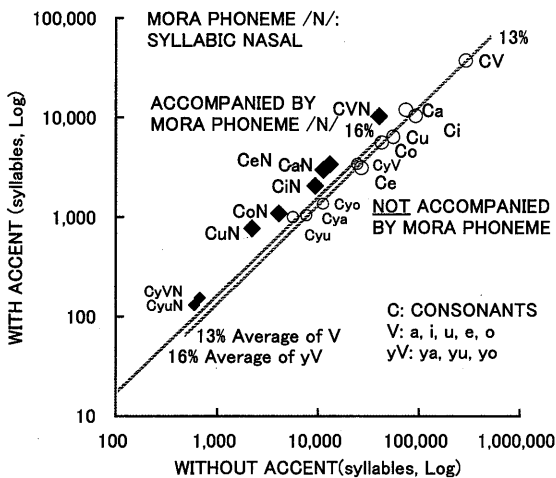


図 6 - 3. /N/

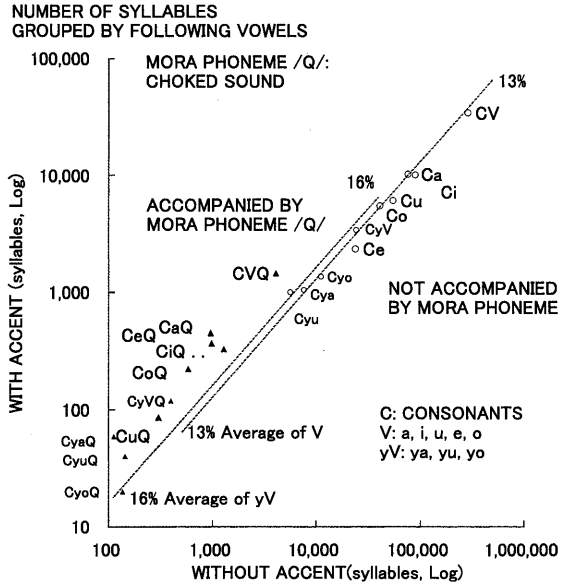


図 6 - 4 . /Q/

この統合された単語集合では、モーラ音素の付く音節の数は、Vsでは、/N/が61,000音節、/H/が47,000音節、/Q/が5,500音節、yVsでは、/N/が1,000音節、/H/が26,000音節、/Q/が500音節であった。モーラ音素の付いたyVsの出現率は、その型の音節の平均的な出現率が1/10であるのに対し、ほぼ1/2と高い。そのうちでは/yo/が3/4を、/yu/が1/4を占めていて、Vsに匹敵する。

モーラ音素の影響については、Vsの平均13%およびyVsの平均の16%より、モーラ音素が付いた音節に付くアクセントの出現頻度が数%高かった。また、モーラ音素を伴わない場合は、すべての種類の後続半母音と母音で数%低かった。これは、モーラ音素を伴う音節が持続時間が長いので、声のピッチを下降させやすく、同時にその変化が開き取りやすいということに関係していると考えられる。

3. 3. 先行子音の影響

アクセントが付くか付かないかによる音節の出現頻度の関係を、モーラ音素が付くか付かないかに分けて、先行子音の種類別に、図7に示す。

VsとyVsのアクセントの付く平均的な比率は14%であり、どのような種類の先行子音の音節でも、モーラ音素の付く音節は、平均より大きく、モーラ音素の付かないものは、平均よりも小さかった。先行子音の種類別に見ると、/q/の付く音節は2倍の28%にも及び、/g/、/k/、/s/、/b/などが先行する音節にアクセントがよりつきやすい傾向が見られる。

図7. アクセントが付くか付かないかによる音節の出現頻度の関係、先行子音の種類別。モーラ音素/H/, /N/, /Q/別。

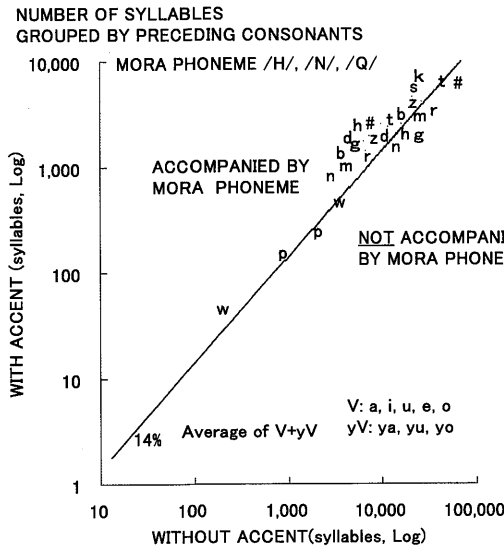


図7-1. /H/, /N/, /Q/

NUMBER OF SYLLABLES  
GROUPED BY PRECEDING CONSONANTS

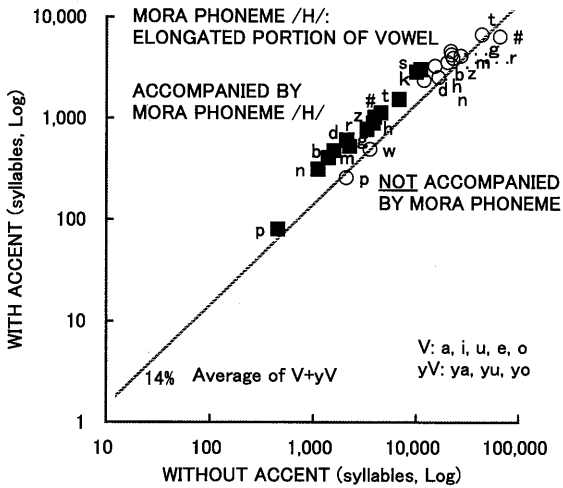


図 7 - 2. /H/

NUMBER OF SYLLABLES  
GROUPED BY PRECEDING CONSONANTS

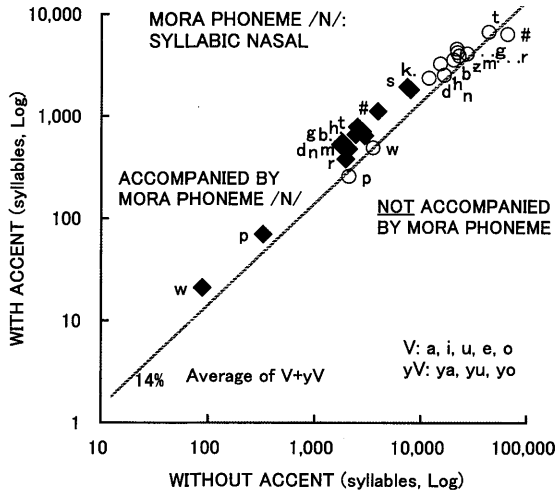


図 7 - 3. /N/



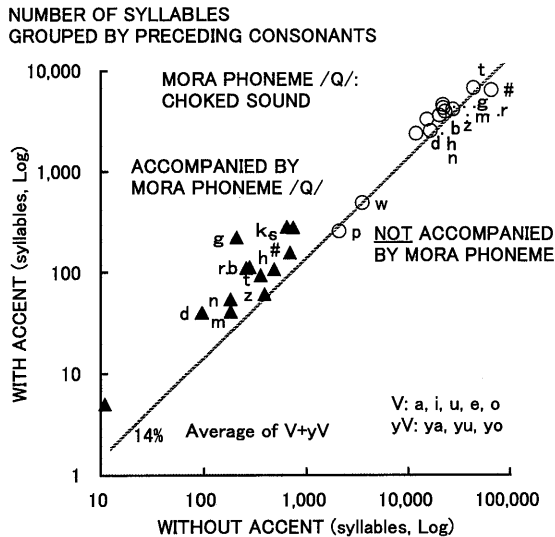


図7-4. /Q/

#### 4. まとめ

日本語の同音語のアクセント型と音節構造との関係について、統計的データにもとづいて解析した結果、統合された単語集合の約35%の単語が、同音語の組に属していた。同音語の組の数は、組数が増すほど、その組内の同音語数は減少し、それらはお互いに正確に反比例していた。

アクセントは、母音型の音節 (Vs) には平均13%、半母音+母音型の音節 (yVs) には16%付いた。どちらのケースにも、後続母音の種類による際立った差はなかった。モーラ音素の影響については、モーラ音素が付いた音節にアクセントが付く確率は、後続する半母音や母音の種類によらず、平均より数%高くなった。

日本語の同音語は、全体の五十音順の表と、同音語の組の中の単語数や単語を構成する音節数に分けて五十音順に配列した表に編集した。モーラ音素を含む単語は、モーラ音素の種類別に、五十音順の表と、単語を構成

徳 弘 康 代

する音節数に分けて五十音順に配列した表に編集した。

統計的分析に使った単語集合に含まれる約49万音節のうちでモーラ音素が後続する音節は30%以上もあった。さらに、モーラ音素が後続する音節は、組の中の語数が非常に多い同音語の音声に多く現れた。モーラ音素の音節での高い出現率、単語アクセントへのそれらの有意な影響、そして、それらの大量の同音語を作る働きは、日本語に特有の重要な性質である。このことから、日本語の同音語の性質や日本語の単語アクセントへのモーラ音素の影響についての統計的な分析結果を踏まえて編集したこれらの資料は、言語処理の基礎研究だけでなく、語学教育に有用なデータベースとなる。

## 参考文献

- Communications Research Laboratory (1995) The EDR Electronic Dictionary. Japan Electronic Dictionary Research Institute, Ltd
- Tokuhiro, Y. and Hiki, S. (2004) Effects of Mora Phonemes on Japanese Word Accent. Proceedings of the 18th Pacific Asia Conference on Language, Information and Computation
- 国立国語研究所 (1989) 『高校・中学校教科書の語彙調査分析編』 秀英出版 p.80

(本学経営学部兼任講師)