国立国語研究所　創立70周年・移管10周年
公開講演会「言語史の計量的研究　――宮島達夫の言語史研究」
2019年6月22日（土）　国立国語研究所（2階）多目的室

『日本古典対照分類語彙表』および『雑誌用語の変遷』

石井　久雄

宮島さんの論文集『言語史の計量的研究』が，本日，笠間書院から刊行の運びになりました。宮島さんの研究全般についてというよりは，この論文集をめぐって，二三申します。

論文集では収載論文それぞれについて1ページ程度の解説（pp. 891～922）が添えられ，執筆に当たったのは専ら鈴木泰さん・安部清哉さんです。松本泰丈さんの序文（pp. 1～10），湯本昭南さんの跋文（pp. 923～927）も出色で，その依頼も鈴木さん・安部さんからなされました。つまり，論文集について述べるには，鈴木さん・安部さんが適任であって，それ以外にはありえません。私は先月1箇月間に脇から茶茶を入れただけですので，態度が大きいと重重自覚しつつ，開き直ります。

なお，宮島さんについてこの場では敬語を用いませんので，ぶっきらぼうで不愉快と感じられるかもしれませんが，よろしくお願いします。

――　1　『言語史の計量的研究』の概要

今回の論文集は，書名『言語史の計量的研究』も全体の構成も，宮島さんが自身で決めて遺したと承知しています。内容は，25年前の論文集『語彙論研究』（1994年，むぎ書房）のあとがきに予告されていました。次のように始まります。

> 本書は，現代日本語の語彙論的研究をあつめた論文集である。古代語・近代語や方言の語彙についてかいたものは，はぶいた。
>
> わたしの おもな研究の場は，２つあった。１つは奥田靖雄氏を指導者とする 言語学研究会・教科研（教育科学研究会）国語部会で，ここでは，言語についての基本的な考え方と，おおくの実例による研究法とをまなんだ。もう１つは，国立国語研究所であり，林大氏や水谷静夫氏のもとで大規模な語彙調査に参加したおかげで，統計的な方法をまなぶことができた。本書の論文で，『教育国語』『計量国語学』や国語研究所の刊行物に発表したものがおおいのは，そのためである。（『語彙論研究』p. 583）

この第1段落から，研究対象を歴史・方言とする論文が少なからずあると了解できます。この予告に添って今回の論文集をまとめ，書名に「言語史」を据えたのでしょう。一書の前半に，古代語・近代語にかかわる論を並べます。後半は，日本語の歴史を諸言語の歴史と比較対照するものが並び，言語史と言っては一般的には想像し難いのですが，宮島さんの歴史研究の際立った特色となっています。

『語彙論研究』あとがきの第2段落に言われる研究の場は，基盤としては変わっていないと理解します。統計的な方法というのを，論文集の書名では「計量的」と言い換えています。『語彙論研究』を刊行したときにはすでに大阪大学に在職し，定年退職を控えていました。仁田義雄さんや大学院生とのゼミ・研究会が充実していたと回顧し，その協同の成果を退職記念としてまとめています。

1995年　（宮島達夫・仁田義雄）　日本語類義表現の文法　（上）単文編・（下）複文・連文編　4＋合計709ページ　くろしお出版

大阪大学退職後は京都橘大学に迎えられて，日本語教育の指導にも携わり，客員教授としての期間も含めて10年以上を過ごします。このような広がりも，基盤に支えられたものでした。

今回の論文集『言語史の計量的研究』は，『語彙論研究』から25年を経ているにもかかわらず，25年間の論文ばかりを集めたのではないと，以上の事情から予想されます。二書とも，各章は元の論文一編にほぼ対応し，その章の数を，元の論文の年代で分けてみます。

	全章数	1993年まで	1994年以降
『語彙論研究』	27	27	
『言語史の計量的研究』	41	18	23（第5部を除いて19）

『語彙論研究』を刊行したときに，そこに収めた 2／3（＝ 18／27 ）の量の論文を言語史に関係してすでに発表していた。それと同量（ 19 ≒ 18 ）の言語史関係論文を，『語彙論研究』以後にも発表している，ということになります。

宮島さんは，研究を始めたときから，日本語・諸言語の歴史に関心を強くもっていました。宮島さんは論文・著書を1950年代から発表し，論文集二つそれぞれの最初の年次は，次になります。

1956年 「文法体系について――方言文法のために」 国語学会『国語学』25 pp. 57～66
＝『語彙論研究』第4部第9章 pp. 523～535

1958年 「近代日本語における単語の問題」 筑摩書房『言語生活』79 pp. 17～26
＝『言語史の計量的研究』第2部第1章 pp. 191～205

最初の著書も，日本語の歴史のものでした。

1957年 『国語と文学の教室 ことばの発展』 5＋136ページ 福村書店

中学生・高校生に向けたシリーズ「国語と文学の教室」の一冊で，シリーズは全32巻，うち国語で9巻，文学で高木市之助『万葉集』・稲垣達郎『夏目漱石』など23巻の構成です。藤原与一『ことばの歴史』（1952年）の一冊もあって，宮島書と藤原書との目次を，大項目の標題で並べてみます。

宮島達夫『ことばの発展』	藤原与一『ことばの歴史』
日本語は役にたたないか	ことばの今と昔
ことばつくり	ことばはうごく
中国語からはいったもの	単語の歴史
ヨーロッパにおいつくために	語法の歴史
近代日本語の成立 ――言文一致まで	発音の歴史
これからの日本語	文字の歴史
	国語の歴史

宮島さんのほうは，語彙体系の発展を中心とします。以後の宮島さんの歴史研究の骨格ともなります。また，序文に叙述の姿勢が語られ，後の研究の姿勢に通じます。

> この本の中には，おもに過去の日本語のことをかきました。それは，過去のことをしるのがおもしろいからではなくて，みなさんに現在の日本語について，さらに将来の日本語について考えてもらいたいからです。
>
> 日本語はたえず変化してきました。しかしその変化は，……，ひとりでにおこるものではありません。人間の努力があってはじめておこったものです。だから，ここでは，ことばの発展を日本人の生活，日本の社会のうごきから，なるべくきりはなさないであつかおう，とおもいました。このしごとはむずかしいので，なかなかうまくいきませんでしたが，ただ，みなさんが日本人と日本語とについてあらためて考えるきっかけになれば，とおもっています。 （はしがき p. 3）

—— 3 『現代雑誌九十種の用語用字』

宮島さんの国語研究所での活動は，現代雑誌九十種の用語の調査・報告で始まります。この調査・報告は，宮島さんの生涯に大きく影響します。しかし，論文集『語彙論研究』にも『言語史の計量的研究』にも，その報告書からの抜粋といったものがありませんので，一言します。

宮島さんは，次の報告書の半分を執筆しています。

1964年 『現代雑誌九十種の用語用字 第三分冊 分析』（国立国語研究所報告 25）
6＋337ページ 秀英出版

第3章 「助詞・助動詞の用法」 pp. 69〜239

この報告の冒頭近くの例えば次，この用例数・挙例を目にして，その大きさ・的確さに私などは圧倒されます。大量の事例を精細に処理する宮島さんの研究を，ここにすでに確かに知ることができます。

が（格助詞）	全体4901	一層 589	二層 918	三層 884	四層 657	五層1853
1）主格	4831	581	898	867	652	1833
副長の声が叫ぶ	（実話雑誌 四月号 30ページ）					
やっぱり重慶が相手だ	（人物往来 二月号 61ページ）					
ジャズピアノが弾ける	（中央公論 七月号 266ページ）					
征長総督の任命が手間どった	（サンデー毎日 十月7日号 24ページ）					
2）連体修飾格	8	2	6	−	−	−
（下略）						
が（接続助詞）	全体1150	一層 130	二層 222	三層 218	四層 166	五層 414
（下略）						（p. 74）

宮島さんは，報告の翌年に補遺を記します。1節末尾に紹介した『日本語類義表現の文法』の書名が，ここから容易に想起されます。

1965年 「いくつかの文法的類義表現について」
『ことばの研究 2』（国立国語研究所論集 2） pp. 75〜106 秀英出版

また，報告書全3冊で，漢字についての報告は『第二分冊 漢字表』（国立国語研究所報告 22，1963年，秀英出版）としてあったが，用語の表記については分析が進められていないとして，高木翠さんとともに一連の論考を発表します。

1974年 （宮島達夫・高木翠）「外来語の表記の変化とゆれ」
計量国語学会『計量国語学』71 pp. 1〜17

1978年 （宮島達夫・高木翠）「雑誌九十種資料の漢語表記」
『研究報告集 1』（国立国語研究所報告 62） pp. 53〜104 秀英出版

1984年 （宮島達夫・高木翠）「雑誌九十種資料の外来語表記」
『研究報告集 5』（国立国語研究所報告 79） pp. 43〜76 秀英出版

1991年 （宮島達夫・高木翠）「雑誌九十種資料の和語表記」
『研究報告集 12』（国立国語研究所報告 103） pp. 1〜82 秀英出版

こうしたものの集大成は，語彙表の再構築に至りました。雑誌九十種の用語の調査報告は，推計上有意義である出現頻度7以上のものを取り上げるに留めています。しかし，宮島さんは，出現した用語すべてを取り上げ，表記にゆれがあればその表記ごとに出現頻度を示す，という整理をすることにしました。整理を終えての刊行とともに，分析結果の公表もしています。

1997年 『現代雑誌九十種の用語用字 全語彙・表記 FD版』

（国立国語研究所言語処理データ集 7）　FD 1枚，別冊8ページ　三省堂

1997年　「調査報告　雑誌九十種表記表の統計」

国立国語研究所『日本語科学』1　pp. 92～104　国書刊行会

宮島さんが現代雑誌九十種調査に触れた論著は，もとより多数に上ります。数十年も経って「現代」とは言えないと断りながら，振り返り続けたのでした。

―― 4　『分類語彙表』

宮島さんの研究の底流には，『分類語彙表』（国語研究所資料集 6，1964年，秀英出版）があります。『分類語彙表』は，第一義的には，現代雑誌九十種の用語の調査の，五十音順語彙表・出現頻度順語彙表に並ぶ，意味順語彙表です。宮島さんにとっては，しかし，調査の成果に至る手法を学んだというよりは，以後の意味論的研究に徹底して活用する基準でした。一例は次です。

1980年　「意味分野と語種」　『研究報告集 2』（国立国語研究所報告 65）　pp. 1～16

秀英出版　＝『言語史の計量的研究』第2部第3章　pp.258～273

『分類語彙表』が広く諸方面に活用されてきた状況を追って，文献解題もまとめています。

1992年　（宮島達夫・小沼悦）「言語研究におけるシソーラスの利用」

『研究報告集 13』（国立国語研究所報告 104）　pp. 1～30　秀英出版

＝『語彙論研究』第5部　pp. 539～568

『分類語彙表』は担当者が林大さんであり，宮島さんの編著書ではありませんが，『分類語彙表増補改訂版』（国立国語研究所資料集 14，2004年，大日本図書）の編集では，林さんを支えたグループの中心にいました。それも相当の長期間にわたっています。中野洋さんを研究代表者として科学研究費補助金を得，『分類語彙表形式による語彙分類表』という成果報告を，1989年・1996年の2度出しています。2度めに際しては，増補改訂の理念・作業を次のように紹介してもいます。

1996年　（宮島達夫・中野洋）「ノート　『分類語彙表』の増補について」

計量国語学会『計量国語学』20. 6　pp. 257～264

『分類語彙表』元版では，語彙調査の結果を示す語彙表として，一語を二つ以上の分類項目に配することは控えめでした。しかし，増補改訂版では語彙調査から解放され，多義にはそれぞれに対応して，一語が意味分類項目数箇所に現れることを，いとわないこととしました。類義語を集めるようなときには，類語集に性格を大きく変えた『分類語彙表』が大いに有用であることになります。

―― 5　語彙の類似度と『古典対照語い表』

さて，今回の論文集『言語史の計量的研究』を集約する概念があるとするならば，「語彙の類似度」であると言うことができます。収載論文すべてが類似度にかかわっているわけではありませんが，第１部「方法論」の主題は語彙の類似度です。宮島さんがこの概念を提示した論文は，次です。

1970年　「語いの類似度」　国語学会『国語学』82　pp. 42～64

＝『言語史の計量的研究』第1部第2章　pp. 31～62

前の論文集『語彙論研究』は刊行がこの論文の24年後になりますが，論文集の索引に「（語彙の）類似度」を見出だすことはできません。「共通度」といったものも見られません。

類似度は，二つの作品の用語を一つずつ出現比率で比べ，大きくないほうを加算する，といったように説明されます。これがどのようなことを意味するか，直感的に分かるようにはなかなか説明しづらいのですが，私には，結局，この論文「語いの類似度」のものが最もなじめます。その説明という

のが，不一致の部分も含めていて，最初に接したときには意外でした。私流に書き改めます。

用語 *w* の出現比率を作品 a で r*w*a，作品 b で r*w*b とし，d*w* ＝｜r*w*a－r*w*b｜とする。また，

r*w*a と r*w*b との大きくないほうを s*w* とすれば，他方は s*w*＋d*w* であり，

r*w*a＋r*w*b ＝ s*w*＋s*w*＋d*w* である。全用語のこの関係を総計して，ra＋rb ＝ s＋s＋d である。

ただし，ra＝rb＝1 であって，2＝2s＋d であり，1＝s＋d／2 である。この s が類似度である。

これでも，実体的には了解し難く，水谷静夫さん『数理言語学』（1982年，培風館，現代数学レクチャーズ D-3）は，次のように解釈し，類似度の一般的な簡略な説明に素直であると見受けます。

作品 a，b に共通に使われている見出し語では，それぞれの出現比率の大きくないほうの分だけ用語情況が似ている（として，その総和が全体での似寄りの度合である）。 （p. 63）

作品の用語の近さ・遠さを数値で表そうとする手法は，見出し語数によるもの，単位語数によるもの，を初めとして，少なくないでしょう。水谷さんは，類似度について，考えかたそのものは宮島さんが最初に出したとは言い切れないようであると注意しています。宮島さんの論文「語いの類似度」でも，上の式の説明の注記に，樺島忠夫さんに教わったところが大きいとあります。種種の手法のうちで，類似度が一般にどのように活用されてきたか，またどのような手法が迎えられてきたか，実は私は承知していません。ただ，宮島さんは類似度を力強く用い続けた，それは確かです。

論文「語いの類似度」は，次のように古典文学作品の事情から切り出します。後に続く説明でも，古典を引き合いに出します。

われわれは，たとえば紫式部日記のことばは徒然草よりも源氏物語にちかい，と感じる。その理由はいくつもあるだろうが，その一つに，源氏と紫とでつかわれている語いが似ている，ということがあるはずだ。 （『言語史の計量的研究』p. 31）

現代語を研究する機関として設置された国立国語研究所で，古典文学作品の語彙を捉えようとするこの研究は，宮島さんに言わせればアンダー・ザ・テーブルで進められました。

成果は，各作品での用語の出現頻度を対比させた表を本体とし，類似度を含む種種の集計を示した表を掲げる形で示され，最初のものは，謄写版刷りで大学などに配布されました。

1969年 『古典対照語い表』（文部省科学研究費補助金一般研究（D）

「作品用語の類似度の研究」研究成果） 14＋129ページ 研究代表者・宮島達夫

ここで取り上げた作品は12，万葉集・竹取物語・伊勢物語・古今集・土左日記・後撰集・枕草子・源氏物語・紫式部日記・更級日記・方丈記・徒然草でした。次いで，2作品，蜻蛉日記・大鏡を加えて，市販品が普及します。宮島さんは茨城県水海道町の出身であり，笠間書院は創業者・池田猛雄さんが茨城県笠間市の出身であることから，当時の国立国語研究所長・岩淵悦太郎さんが間に入ってくれたと聞いています。

1971年 『古典対照語い表』 14＋340ページ 笠間書院，笠間索引叢刊 4

さらに，データを電子的に処理できるようにもしました。

1989年 （宮島達夫・中野洋・鈴木泰・石井久雄）

『フロッピー版古典対照語い表 および使用法』 FD 1枚，別冊48ページ 笠間書院

『古典対照語い表』の研究に着手した最初の厄介事の指摘・処理・提案は，論文「語いの類似度」に先立って発表され，今回の論文集『言語史の計量的研究』の巻頭に置かれています。

1969年 「総索引への注文」 国語学会『国語学』76（特集 国語史資料論） pp.110～120

＝『言語史の計量的研究』第1部第1章 pp.13～30

—— 6 『日本古典対照分類語彙表』

『古典対照語い表』は，フロッピー版の準備のころから増補が構想され，25年の後に実現します。

2014年 （宮島達夫・鈴木泰・石井久雄・安部清哉）

『日本古典対照分類語彙表』 1147ページ，別冊31ページ，CD 1枚 笠間書院

この版は，『古典対照語い表』に対して，大きい違いが3点あります。

- 3作品，新古今集・宇治拾遺物語・平家物語を追加して，合計17作品を対照する。
- 一語一語に『分類語彙表 増補改訂版』の意味分類項目番号を記す。
- 書籍本体の五十音順の語彙表に加えて，出現頻度順・意味分類順の語彙表，種種の統計，別冊の内容を，CD で提供する。語彙表はEXCELファイルである。

別冊には，宮島さんの解説「古典語の統計と意味」，および小木曽智信さんの「EXCELによるデータの活用」を収めています。

『日本古典対照分類語彙表』の刊行の直前に，田島毓堂さんに誘われて紹介をしました。研究発表の体裁をとって，標題にはカモフラージュがあります。

2014年 （石井久雄）「古典語彙の出現頻度の分布——『日本古典対照分類語彙表』を利用して」

語彙研究会定例研究会第100回 5月24日 愛知学院大学大学院栄サテライト

予稿所在 http://kotonoha.art.coocan.jp/goi_m100.pdf

万葉集は，依拠する索引を変更しました。宮島さんはその語彙表を作成し，巻の間の関係を見るのも意義があるとして，一書にまとめます。体裁は『日本古典対照分類語彙表』に倣い，作品は各巻が相当することになります。万葉集に現れない意味は削除するといった調整もしています。

2015年 『万葉集巻別対照分類語彙表』 253ページ，CD 1枚 笠間書院

こうした古典語彙表の今後のことを一言します。作品の追加としては，今昔物語集が夙に上がり，作業も開始しました。

1999～2001年度科学研究費補助金基盤研究（C）「『今昔物語集』語彙の計量的研究」

（研究代表者・宮島達夫，研究分担者・鈴木泰）

この成果報告書の内容は，論文などには全く見えませんので，紹介します。すなわち，馬渕和夫監修『今昔物語集文節索引』全28巻を基として，次のように要約できます。

		異なり	延べ		類似度			
全体		20326	358790		震旦	本朝	仏法	世俗
天竺	（巻第 1～ 5）	5834	66520	天竺	.666	.612	.635	.537
震旦	（巻第 6～10）	5416	51209	震旦	....	.640	.677	.544
本朝	（下記合計）	15991	241061					
仏法部	（巻第11～20）	10567	135459					
世俗部	（巻第21～31）	9746	105602					

『古典対照語い表』14作品の間91対の類似度は，古今集・後撰集.733 で最大であり，次いで源氏物語・蜻蛉日記.588 である。それと比べて，今昔物語集の内部は等質性が高いと言ってよい。

この成果を得たものの，論文「総索引への注文」にも指摘がある単位の調整の問題に，やはり突き当たり，あらためて索引から取り掛かることにしました。

2005～2007年度科学研究費補助金基盤研究（C）

「『今昔物語集』のKWIC作成の原理的問題とその意味・文法論的利用の可能性」

（研究代表者・鈴木泰，研究分担者・宮島達夫・石井久雄・安部清哉）

これも停滞していますが，いずれ本格的に再開しなければなりません。

古典語彙表の今後の展開には，意味ごとに出現頻度を挙げるという夢を見ることができます。そこでの類似度も算出してみたいものです。『日本古典対照分類語彙表』の企画の際に考えなかったわけではありませんが，全用語にわたる作業量を予想して断念し，作品いずれかに現れた意味を挙げるに留めることとしました。しかも，その最低限でも意味を記した，そのことが，今後，作品の追加を難しくする要因になっているとも思われます。新たな用語には意味を記述しなければならず，すでにある用語には，すでに記されている意味の範囲内に収まるか検討して，ときに新たな意味を記述しなければならない。といった手間が，結構大変であろうと想像します。『日本古典対照分類語彙表』の全面的な意味記述は，宮島さんという情熱・個性があったればこそ遂げられたものでした。

―― 7 『雑誌用語の変遷』

宮島さんは，論文「語いの類似度」の後半で，類似度を展開する事例を考え，品詞構成・語種構成，文体論，文法論，文字論，方言研究を挙げます。例ですから，挙げていないとあげつらうのも妙ですが，歴史的変遷をとらえることにも類似度を適用することができます。宮島さん自身が実行します。

1987年 『雑誌用語の変遷』（国立国語研究所報告 89） 427ページ 秀英出版

1. 5. 1 pp. 11～23 「単位語の認定 ――調査単位のながさの問題」

＝『語彙論研究』第2部第2章 pp.113～119

＝『言語史の計量的研究』第1部第5章 pp.108～117

2 pp. 35～37 「文体」 3 pp. 38～87 「語彙」

5 pp. 155～177 「文法」 6 pp. 178～193 「表記」

＝『言語史の計量的研究』第2部第5章 pp.287～360 「文体・語彙・文法・表記」

『雑誌用語の変遷』第1章からは論文集二つに重ねて収載し，第2～6章からは抜粋して論文集の一章に仕立て直しています。二つの論文集に重ねたのは，あるいはミスではないかと疑います。

類似度は，『日本古典対照分類語彙表』と『雑誌用語の変遷』とで，役割を異にします。古典では，歴史というものが過去の一静態であって，そのなかでの異同を見ていた。雑誌では，歴史は移り変わる動態であって，変化の大小を見ている。と言うことができます。しかも，『雑誌用語の変遷』で，類似度は共通度という概念に拡張されます。例えば，次のように計算されます。

16年・26年・36年の共通度は，一語一語について，16年の出現比率・26年の比率・36年の比率で最小のものを選び，それを合計した数値である。

これについて，次のように言っています。

さきに，類似度の方法を共通度に拡張した，といったが，じつは，この共通度のかんがえかたが，一般的な方式であって，類似度は，比較の対象が２つにかぎられた，共通度の特殊例だ，とみるべきだろう。したがって，指標のよびなとしても，おなじなまえをつかってもいいのだが，３つ以上のばあいについては，類似度というよりも，共通度のほうが自然におもわれるので，そうよぶことにした。 （『雑誌用語の変遷』p. 52＝『言語史の計量的研究』p. 294）

『雑誌用語の変遷』から抜粋した『言語史の計量的研究』第2部第5章は，その一章で74ページを占めます。『言語史の計量的研究』で最大の98ページを占める

1979年 「「共産党宣言」の訳語」 言語学研究会編『言語の研究』 pp. 425～517 むぎ書房

＝『言語史の計量的研究』第2部第11章 pp. 475～572

に，次ぐ分量です。ちなみに，ページ数第3位・第4位は，52ページ・44ページのものです。

1967年　「現代語いの形成」　『ことばの研究 3』（国立国語研究所論集 3）　pp. 1〜50
秀英出版　＝『言語史の計量的研究』第2部第2章　pp. 206〜257

1969年　「近代日本語における漢語の位置」　教育科学研究会国語部会『教育国語』16
pp. 17〜44　麦書房　＝『言語史の計量的研究』第2部第7章　pp. 376〜419

いずれも第2部「近代語」に置かれ，第2部は一書全体のページ数のちょうど半分を占めます。ちなみに，ページ数第5位は，５節冒頭に引いた「語いの類似度」でした。

『雑誌用語の変遷』については，抜粋されなかった部分についても記したいことがあり，以下，箇条書きのように書き留めます。

論文集に抜粋されなかった「第4章　語誌――意味分野ごとの記述」（pp. 88〜154）は，『分類語彙表』の意味項目番号が同じである語を集めて，状況を見ています。例えば，次のようです。

1. 1113　理由・目的・証拠

	新しさ	戦前	戦後	06年	16年	26年	36年	46年	56年	66年	76年
ゆえ	−3. 47	49	11	20	14	6	9	6	2	2	1
ゆえん（所以）	−3. 84	15	4	3	10	2	−	4	−	−	−
ため（為）	0. 00	126	137	29	35	34	28	46	32	23	36

ごく一般的な単語だが，「ゆえ」と「ゆえん」は文体的に古いことがはっきりしている。
（『雑誌用語の変遷』p. 93）

「戦前」は36年までの出現頻度の合計，「戦後」は46年からです。「新しさ」は，次の計算値を頻度合計で除した指標です。宮島さんは，このような指標を幾つも作っています。

7×76年出現頻度 ＋ 5×66年頻度 ＋ 3×56年頻度 ＋ 1×46年頻度
−（7×06年出現頻度 ＋ 5×16年頻度 ＋ 3×26年頻度 ＋ 1×36年頻度）　（同上 p. 66）

こうした語誌は，意味項目番号を利用した精細な記述の見本になります。

この調査は，各年の延べ語数を1万で一定させています。感覚的には抽出比率を一定にしそうですが，それを退けたところに，構造のありかたをふまえて歴史をとらえようとする姿勢が現れています。

> ここでは，用語の歴史的な変化をしろうとする立場にたって対象をながめている。漢語の比率にしても，ことなり語数についていおうとすると，標本の量がちがっていては不都合なのである。というのは，漢語は和語にくらべて基本的でないものがおおいから，標本の量をふやしていくと，ことなり語数における漢語の増加は，和語のそれを上まわり，のべにおける比率が一定であるにもかかわらず，のべ語数がおおいほど，ことなり語数における漢語の比率がたかくなるという可能性があるからである。　（同上 p. 8）

この調査には，宮島さん自身のものでなく，追跡があります。

1990年　（石井久雄）「『中央公論』1986年の用語」
『研究報告集 11』（国立国語研究所報告 101）　pp. 1〜40　秀英出版

2011年　（入江さやか）「『中央公論』101年の語彙」
同志社大学『同大語彙研究』13　pp. 9〜14

2015年　（石井久雄・入江さやか）「『中央公論』101年間語彙表」
同志社大学『同志社日本語研究』別刊2　pp. i〜x，1〜410

101年間というのは2006年まで，調査は『雑誌用語の変遷』と同じく10年ごとです。

—— 8 『言語史の計量的研究』の後に

論文集『言語史の計量的研究』の第3部・第4部について何かを言うことは，私からはもはや差し控えます。宮島さんの元もとの構想になかった第5部に関連して，妄想を膨らませます。

宮島さんがこの論文集を構想したときには，自身がかくも早く研究から退かなければならなくなるとは，予想していなかったでしょう。『語彙論研究』を編集しつつ恐らくは『言語史の計量的研究』を周到に準備していたように，今回，第三の論文集も企画していたのではなかろうかと想像するのです。例えば，「コーパス」という，キーワードになってよさそうなものが，論文集に見当たりません。宮島さんは，しかし，コーパスに早くから注目して，標題に明示した論文があります。

2007年 「語彙調査からコーパスへ」 国立国語研究所『日本語科学』22
（特集 コーパス日本語学の射程） pp. 29〜45 国書刊行会

2008年 「文章の文体と単語の文体 ——国研コーパスを利用して」
近代語学会『近代語研究』14 pp. 388〜375 武蔵野書院

2011年 「コーパスによる日本語と英語の対照」
大東文化大学海山研究所『対照言語学研究』21 pp. 1〜15

標題には謳わなくとも，キーワードに「コーパス」と挙げることがあります。

2006年 「「テイル」と「テル」」 土岐哲先生還暦記念論文集編集委員会編
『日本語の教育から研究へ』 pp. 239〜252 くろしお出版

このようなものと，今回の論文集の第5部の論文とを並べて，コーパスによる言語研究といった一書を成り立たせることは，できなかったであろうか。今回の論文集の編集に注文をつけるわけではありませんが，思わせられます。

宮島さんは，毎夏，古典文学を読むことにして，続けていた。老化防止のために，プログラミングを好んだが，スパゲッティ好きであった。……そのようなエピソードを思い起こすと，きりがなくなります。この講演会の場にはふさわしいとも思えませんので，遠慮します。

ご清聴ありがとうございました。

補記

飛田良文さんは，『哲学字彙訳語総索引』（1979年，笠間書院，笠間索引叢刊 72）を上梓するに当たって宮島さんがすばやく取り次いでくれたと，同書の跋文に記しています。宮島さんは，かつて岩渕悦太郎さんから蒙った恩を，他に及ぼして酬いていることになります。

付論　　宮島達夫の語彙の類似度について

語彙の類似度の実体的イメージは，次のようである。すなわち，延べ語数が等しい作品二つについて，一方の単位語と他方の単位語とが標目語を同じくするときに，その単位語を拾い上げる。ただし，一つの標目語には二つの作品で同数の単位語を拾うものとし，単位語の数に多寡の差があるときには，少ない方の分のみを拾い，多い分は捨てる。そうして，できる限り多く単位語を拾い上げると，その一団は，作品二つの間で堅く重なっているところであり，延べ語数に占めるその単位語の数の比率が，類似度である。

例えば，いま，延べ語数が 9 である作品が二つある。

年の　内に　春は　来にけり　一年を　去年とや　言はむ　今年とや　言はむ

春　来ぬと　人は　言へども　鶯の　　鳴かぬ　限りは　　あらじとぞ　思ふ

このうちで，標目語「春」「来」「言ふ」が，作品双方に重なって現れている。「春」「来」は単位語数それぞれ 1 であり，類似度算出のためにはそのまま拾い上げる。「言ふ」は，一方の作品で2回繰り返されて，他方で 1回であり，拾うのは 1回として，多いほうの 1回分は捨てる。二つの作品は，単位語数 3 が重なり，それぞれにおいて，単位語全体の 1／3（＝重なり単位語3／延べ9）を他方と共有する。類似度は 1／3 ＝ 0.3333… である。

しかしながら，延べ語数が等しい作品というものは，そうそうにあるものではなく，何らかの調整をすることになる。延べ語数が 11 である作品

世の中に　いづら　吾が　身の　ありて　なし　あはれとや　言はむ　あな　憂とや　言はむ

においても「言ふ」が 2回繰り返される。やはり「言ふ」が繰り返される「年の内に」歌との類似度を見ようとして，ともに 2回を拾い上げても，延べ語数が異なるために，単純には計算ができない。

処理する簡明な方法は，当の 2作品の延べ語数を公倍数に拡大し，「言ふ」の出現も相似的に増大させるものである。2作品の延べ語数をともに 99 ＝ 9×11 として，「言ふ」の出現は「年の内に」歌で 2×11，「世の中に」歌で 2×9，重なるところは 18 ＝ 2×9 の分であるとする。「言ふ」が類似度に寄与する数値は，18／99 ＝ 0.1818… である。下線は，類似度に寄与する数値である。

「春来ぬと」歌と「世の中に」歌との間では，類似度に寄与する「言ふ」の数値は，11 ＝ 1×11 と 18 ＝ 2×9 とを対比して 11／99 ＝ 0.1111… とする。この 2作品では，「あり」も 1回ずつ出現し，11 ＝ 1×11 と 9 ＝ 1×9 とを対比して 9／99 ＝ 0.0909… の数値を得，全体の類似度は，11／99 ＋ 9／99 ＝ 0.2020… である。また，延べ語数が 10 である作品

君が　名も　吾が　名も　立てじ　難波なる　御津とも　言ふな　逢ひきとも　言はじ

と「世の中に」歌との間では，類似度は，「言ふ」について 22 ＝ 2×11 と 20 ＝ 2×10 とを対比して 20／110 ＝ 0.1818… を得，「吾が」について 11 ＝ 1×11 と 10 ＝ 1×10 とを対比して 10／110 ＝ 0.0909… を得て，全体で 30／110 ＝ 0.2727… である。

以上は次のようにまとめられる。初めの 2首の分も掲げる。

	言ふ	言ふ	あり	言ふ	吾が	春	来	言ふ
「年の内に」歌	22／99					1／9	1／9	2／9
「春来ぬと」歌		11／99	11／99			1／9	1／9	1／9
「君が名も」歌				22／110	11／110			
「世の中に」歌	18／99	18／99	9／99	20／110	10／110			
類似度への寄与	2／11	1／ 9	1／11	2／ 11	1／ 11	1／9	1／9	1／9

「春来ぬと」歌と「世の中に」歌とでは，「言ふ」のもともとの出現頻度に 1回・2回の違いがある。「世の中に」歌の 18／99 を 9／99 ずつに分け，「春来ぬと」歌とは，11／99 でなくてその 9／99 で重なるとも扱いうるが，「春来ぬと」歌に 2／99 の余りが出て，「世の中に」歌に残った 9／99 との間の処理をまた考えなければならない。そこで単位語一つ一つの制約を離れ，標目語ごとに出現比率を求めて，二つ作品の出現比率の小さいほうの分を，大きいほうと重なると見，類似度の算出のために掬い上げることにする。出現比率が大きいほうに，掬わないところを残す。

ここまで下線を施した，類似度の算出に寄与する数値は，約分したものも示す。約分したものに，出現比率であることが明らかに見て取れ，対比されたものも出現比率にほかならないと理解される。「年の内に」歌と「春来ぬと」歌との間の数値も，つまりは出現比率である。

作品ごとに，単位語は標目語のもとで出現頻度にまとまって出現比率となり，延べ語数はすべての標目語の出現比率の合計として常に 1 となる。二つの作品で標目語のそれぞれの出現比率を対比し，大きくないほうの数値を合計して，その合計値が類似度である。――これが，論文「語いの類似度」における類似度の規定である。

延べ語数が同じくない二つの作品につき，類似度を算出するために何らかの操作をするとして，いま一つの方法を試み，本来の類似度との離合を見てみる。その方法というのは，延べ語数が大きいほうの単位語を，標目語の五十音順語彙表に添って一列に並べ，その列から等間隔に拾い上げて，総数が小さいほうと等しくなるようにするものである。

例えば，源氏物語（延べ語数 207788）について，五十音順語彙表の最初は，表Ａの左第1・2列のように見出し・出現頻度を示す。それを，第3列αのような並びとして理解し，1番から207788番までの番号を付ける。いま万葉集（延べ語数 50056）との対照のために調整するならば，αの番号に即して，4.1511番 ＝ 207788／50056 ごとに 1語ずつ取り上げて，万葉集の延べ語数と同数の単位語を得ることができる。そこで，αの番号を 4.1511 で除した商をβとし，小数を四捨五入した整数をγとする。γで等値であるものは，最初のものを取り上げて，δとする。δで数字 1～50056 を置いた単位語が，対照に用いるものであり，標目語で整理して，εの出現頻度になる。調整した語彙表を読むならば，「あ(彼)　出現頻度1」（原出現頻度3），「あ(案)」「ああ(嗚呼)」は標目語として失われ，「あいぎゃう(愛敬)　4」（原14），「あいぎゃうづく(愛敬付)　10」（原42），以下省略，である。なお，調整した源氏物語の語彙表では，異なり語数 6736 である。

表Ａ

語彙表見出し	出現頻度	α（随時省略）	β	γ	δ	ε
あ（彼）	3	1あ(彼)	0.240	0		1
		2あ(彼)	0.481	0		
		3あ(彼)	0.722	1	1	
あ（案）	1	4あ(案)	0.963	1		0
ああ（嗚呼）	1	5ああ(嗚呼)	1.204	1		0
あいぎゃう（愛敬）	14	6あいぎゃう(愛敬)	1.445	1		4
		7あいぎゃう(愛敬)	1.686	2	2	
		8あいぎゃう(愛敬)	1.927	2		
		11あいぎゃう(愛敬)	2.649	3	3	
		15あいぎゃう(愛敬)	3.613	4	4	
		19あいぎゃう(愛敬)	4.577	5	5	
あいぎゃうづく（愛敬付）	42	20あいぎゃうづく(愛敬付)	4.817	5		10
		23あいぎゃうづく(愛敬付)	5.540	6	6	
		61あいぎゃうづく(愛敬付)	14.694	15	15	

同様に他の諸作品に対しても，それぞれの延べ語数に応じて別別に源氏物語を調整する。表Ｂに，各作品の異なり語数・延べ語数，源氏物語との本来の類似度，また，その作品と源氏物語との双方に重なって出現する標目語の数，その出現頻度の合計すなわち単位語の数を示す。重なりの標目語・単位語の数は，類似度の計算には直接には関係せず，参考である。表Ｂに，続けて，各作品の延べ語数と同じくなるように単位語数を調整したときの，源氏物語の異なり語数を示す。また，調整した源氏物語と各作品で重なる標目語・単位語の数を，参考として示す。

調整した源氏物語と万葉集とのばあいは，重なって出現する標目語の表が，次のように始まる。

	調整した源氏物語の		万葉集の		類似度のために取り上げる	
あが(吾)	出現頻度	4	出現頻度	406	大きくないほうの数値	4
あかし(明石)		11		7		7
あかし(明・赤)		10		2		2
あかしかぬ(明不堪)		1		1		1
あかす(明)		7		7		7

延べ語数が二作品で等しいので，類似度を算出するためには，大きくないほうの出現頻度を記録して，

表Ｂ

作品	異なり	延べ	類似度	重なり		調整源氏	重なり		類似度	
源氏	11416	207788	対源氏	標目語	単位語	異なり	標目語	単位語	対調整	
平家	13173	100091	.405488	*3320*	*233728*	8922	*3005*	*144559*	.405361	−0
万葉	6601	50056	.293418	*1862*	*157674*	6736	*1466*	*63236*	.293131	−0
宇治	6540	49183	.489242	*2975*	*212777*	6722	*2418*	*80397*	.489112	−0
枕	5246	32904	.571628	*3292*	*206749*	5581	*2514*	*56776*	.571814	+0
大鏡	4819	29253	.527175	*2672*	*196509*	5274	*2054*	*47677*	.527467	+0
蜻蛉	3599	22400	.590332	*2639*	*191886*	4643	*1975*	*38399*	.590982	+1
新古	2543	17165	.354508	*1652*	*138304*	4056	*1034*	*23817*	.354442	−1
徒然	4240	17110	.510674	*2211*	*176488*	4045	*1559*	*26275*	.510461	−1
後撰	1923	11955	.377947	*1403*	*133797*	3376	*840*	*16507*	.374654	−3
古今	1993	10013	.353971	*1422*	*128726*	3022	*772*	*13267*	.351742	−2
紫	2468	8736	.569244	*1947*	*170493*	2829	*1243*	*13818*	.569139	−0
更級	1950	7243	.523984	*1576*	*157806*	2527	*964*	*11087*	.524920	+1
伊勢	1692	6931	.447681	*1304*	*147518*	2466	*795*	*10250*	.449141	+1
竹取	1310	5119	.455774	*1026*	*138159*	2038	*599*	*7077*	.454385	−2
土左	984	3496	.369344	*714*	*115682*	1590	*378*	*4231*	.367562	−1
方丈	1148	2527	.376336	*748*	*108886*	1294	*349*	*2726*	.376335	−0

表Ｃ

作品	異なり	延べ	類似度	重なり		調整作品	重なり		類似度	
方丈	1148	2527	対作品	標目語	単位語	異なり	標目語	単位語	対調整	
源氏	11416	207788	.376336	*748*	*108886*	1294	*349*	*2726*	.376335	−0
平家	13173	100091	.336297	*861*	*41916*	1651	*361*	*2502*	.334388	−2
万葉	6601	50056	.267700	*503*	*21933*	1314	*272*	*2188*	.268302	+0
宇治	6540	49183	.375082	*689*	*29012*	1268	*345*	*2814*	.375544	+1
枕	5246	32904	.340913	*566*	*18248*	1268	*315*	*2589*	.341115	+0
大鏡	4819	29253	.326667	*583*	*15717*	1252	*305*	*2507*	.325682	−1
蜻蛉	3599	22400	.372456	*537*	*14771*	1126	*331*	*2861*	.372378	−0
新古	2543	17165	.314857	*455*	*11035*	1086	*309*	*2646*	.312623	−2
徒然	4240	17110	.428680	*643*	*11579*	1358	*402*	*2975*	.434903	+6
後撰	1923	11955	.336771	*394*	*8151*	961	*294*	*2601*	.336762	−0
古今	1993	10013	.326055	*407*	*6815*	1022	*302*	*2576*	.328056	+2
紫	2468	8736	.317321	*418*	*5265*	1286	*292*	*2356*	.317768	+1
更級	1950	7243	.378315	*427*	*5656*	1133	*339*	*2791*	.382271	+4
伊勢	1692	6931	.352490	*412*	*5462*	1028	*320*	*2775*	.351009	−1
竹取	1310	5119	.356161	*362*	*4277*	943	*297*	*2704*	.354966	−1
土左	984	3496	.331794	*282*	*3232*	851	*256*	*2615*	.331222	−1

合計を延べ語数で除すればよい。調整した源氏物語と万葉集とで重なる標目語 1466 について，それぞれからの数値を合計すると 14673 であり，延べ語数 50056 で除して，類似度 0.293131 である。

表Ｂの最右列に，調整した源氏物語と諸作品と間の類似度を示す。小数の下位のほうは切り捨てる。また，調整した類似度を本来の類似度と比べて，大きければ + を記し，小さければ − を記す。さらに，二つの類似度を小数第4位で四捨五入し，その差，千分の幾つであるかを，正負号に続けて記す。源氏物語と万葉集との間では，本来の類似度 0.293418，調整した類似度 0.293131 であり，調整したほうが小さくて負号 − が付き，小数第4位の四捨五入でともに 0.293 であって差 0 である。二つの類似度の数値の差は，源氏物語と他の作品との間では，高だか千分の3 程度であることになる。

方丈記は作品中で延べ語数が最小であり，調整すると，他の作品いずれもがこれに合わせることになる。表Ｂに倣って表Ｃを掲げる。ただし，中央の異なりの列は，諸作品の，単位語数を 2527 に調整したときの標目語数である。本来の類似度と調整後との差は，千分の6 程度に広がる。

表Ｄ

		平家	宇治	方丈	新古	大鏡	更級	紫	源氏	
	徒然	.440582	.535947	.428680	.312719	.466623	.462330	.435328	.510674	徒然
平家	.4424+1	平家	.486336	.336297	.268747	.451088	.377858	.354630	.405488	平家
宇治	.5371+1	.4860−0	宇治	.375082	.303263	.507575	.495847	.426384	.489242	宇治
方丈	.4349+6	.3343−2	.3755+1	方丈	.314857	.326667	.378315	.317321	.376336	方丈
新古	.3126−0	.2690+0	.3030−0	.3126−2	新古	.255818	.413942	.286286	.354508	新古
大鏡	.4681+1	.4507−0	.5075−0	.3256−1	.2555−0	大鏡	.437159	.466398	.527175	大鏡
更級	.4611−1	.3782+0	.4945−1	.3822+4	.4125−1	.4390+2	更級	.463930	.523984	更級
紫	.4338−1	.3545−0	.4252−1	.3177+1	.2875+2	.4679+2	.4630−1	紫	.569244	紫
源氏	.5104−1	.4053−0	.4891−0	.3763−0	.3544−1	.5274+0	.5249+1	.5691−0	源氏	
枕	.5119+0	.3998+0	.5448−1	.3411+0	.3049−0	.5054+0	.5358+1	.5528−1	.5718+0	枕
蜻蛉	.4936−0	.3831−0	.5233−2	.3723−0	.3861+1	.4575−0	.5641+3	.4814+0	.5909+1	蜻蛉
後撰	.3468+0	.2761−1	.3296+1	.3367−0	.6388+1	.2732−1	.4368+2	.3097+1	.3746−3	後撰
土左	.3893−3	.3060−3	.4270+1	.3312−1	.3326+2	.3415−0	.4382−2	.3406+1	.3675−1	土左
古今	.3396+2	.2708+2	.3184−1	.3280+2	.6232−2	.2642−2	.4112+1	.3008+1	.3517−2	古今
伊勢	.4459+3	.3456−2	.4608−0	.3510−1	.4136+1	.3896+1	.4993+1	.3918+1	.4491+1	伊勢
竹取	.4537−0	.3737−1	.5196−3	.3549−1	.2883−1	.4305+1	.4696−0	.3936+1	.4543−2	竹取
万葉	.2917−0	.2711+0	.3106+1	.2683+0	.4527+1	.2531−0	.3452−0	.2532+0	.2931−0	万葉
	徒然	平家	宇治	方丈	新古	大鏡	更級	紫	源氏	

		枕	蜻蛉	後撰	土左	古今	伊勢	竹取	万葉	
	徒然	.511751	.494277	.346852	.391718	.338271	.443317	.454215	.292227	徒然
	平家	.399838	.383441	.277331	.309201	.269322	.347684	.375330	.271060	平家
	宇治	.545748	.524694	.329349	.426018	.318774	.461403	.522739	.310485	宇治
	方丈	.340913	.372456	.336771	.331794	.326055	.352490	.356161	.267700	方丈
	新古	.305369	.385068	.637668	.331276	.625481	.413282	.288514	.452296	新古
	大鏡	.504855	.457965	.274005	.342322	.266026	.388797	.429895	.253342	大鏡
	更級	.534538	.560961	.435252	.440149	.410496	.497909	.470323	.345315	更級
	紫	.554268	.480596	.308550	.340376	.300440	.391163	.392664	.252513	紫
	源氏	.571628	.590332	.377947	.369344	.353971	.447681	.455774	.293418	源氏
		枕	.579150	.337144	.408761	.325587	.460558	.468350	.296235	枕
	蜻蛉	.5784−1	蜻蛉	.424007	.452675	.404815	.507937	.471343	.337238	蜻蛉
	後撰	.3369−0	.4219−2	後撰	.371882	.735076	.454290	.333727	.492621	後撰
	土左	.4090+0	.4530+0	.3701−2	土左	.365714	.476234	.434419	.339487	土左
	古今	.3275+2	.4065+2	.7343−1	.3615−4	古今	.462329	.318408	.521621	古今
	伊勢	.4601−1	.5071−1	.4546+1	.4788+3	.4640+2	伊勢	.457624	.393001	伊勢
	竹取	.4711+3	.4733+2	.3340+0	.4342−0	.3151−3	.4588+1	竹取	.305506	竹取
	万葉	.2960−0	.3360−1	.4913−2	.3392−0	.5214−1	.3937+1	.3088+3	万葉	
		枕	蜻蛉	後撰	土左	古今	伊勢	竹取		

本来の類似度と，延べ語数を調整した類似度とを，併せて表Ｄとする。作品の配列は，延べ語数が多い順でなく，日本古典対照分類語彙表に添った，成立年次が新しい順とする。左上から右下への対角線を境にして，右上が本来の類似度であり，左下が調整後の類似度である。調整後の類似度は，小数第4位までを示し，第5位以下は切り捨てる。二つの類似度を小数第4位で四捨五入し，その差を調整後の類似度に示す。そうした表の全体を左右に切り分け，表Ｄでは左部分を上段に配し，右部分を下段に配する。

表Ｄの左上あたりを読むと，徒然草と平家物語との本来の類似度は 0.440582 であり，調整した類似度は 0.4424（小数第5・6位は示さないが 31）であって，調整した類似度のほうが大きく，二つの類似度それぞれの小数第4位を四捨五入した 0.441 と 0.442 との差は千分の1 である，のようになる。表の右下は，竹取物語と万葉集との類似度は 0.306 および 0.309 であって，調整したほうが千分の3 大きい，と読まれる。

本来の類似度と調整した類似度との差，すなわち調整した類似度に表Ｄで付けた ±n を，整理すると次のようである。作品の側から見ると，例えば徒然草では，−3・+2・+3・+6 がそれぞれ1対（対土左・古今・伊勢・方丈）あり，−1・+1 がそれぞれ 3対，−0 が 4対，+0 が 2対ある，ということになる。差の側から見ると，−4 は，土左・古今の 1対であり，ここの数値の縦の並びは，一つの対を二作品それぞれで数えているから，全体としては半分とすることになる。

＼差	−4	−3	−2	−1	−0	+0	+1	+2	+3	+4	+6	平均
徒然		1		3	4	2	3	1	1		1	+0.5
平家		1	2	2	5	4	1	1				−0.375
宇治		1	1	4	5		5					−0.25
方丈			2	4	3	2	2	1		1	1	+0.375
新古			2	3	4	1	4	2				+0.0625
大鏡			1	2	6	2	3	2				+0.1875
更級			1	4	2	1	4	2	1	1		+0.5625
紫				4	2	2	6	2				+0.375
源氏		1	2	3	5	2	3					−0.4375
枕				4	3	6	1	1	1			+0.125
蜻蛉			2	3	4	2	2	2	1			+0.125
後撰		1	3	3	2	2	4	1				−0.375
土左	1	2	2	2	3	2	2	1	1			−0.5625
古今	1	1	3	3			2	6				−0.125
伊勢			1	3	1		8	1	2			+0.6875
竹取		2	1	3	3	1	3	1	2			0
万葉			1	2	6	3	3		1			+0.125
全体	1	5	12	26	29	16	28	12	5	1	1	+0.058823

いま，徒然草における二つの類似度の差について，平均を

（−3×1−1×3−0×4+0×2+1×3+2×1+3×1+6×1）÷16 ＝ +8÷16 ＝ +0.5

として求め，他の作品についても同様に計算する。差の平均は，最小が土左日記 −0.5625，最大が伊勢物語 +0.6875 であり，全体では合計欄から求めて +0.058823 ＝ +8÷136 である。

類似度は，出現比率によっても，延べ語数が大きいほうの単位語を取捨しても，算出した数値の差は小さいと見られる。差が小さい理由は，類似度が取り上げる標目語・単位語が，作品二つのいずれにも出現するものであるからであると考えられる。一方の作品に偏って出現するものは，類似度の算出では取り上げられない。二つの作品で出現頻度が相応に大きければ，全体が圧縮されても同じ比率で圧縮され，全体が取捨されても掬われて，類似度の数値のしかるべき部分を占めることになる。

類似度の算出で二つの作品の延べ語数を合致させるということについて，議論は以上である。さらに立ち入ることは，類似度の問題であるよりは，抽出ということの問題になるであろう。

類似度についての付けたりとして，表Ｃの調整を顧みる。すなわち，方丈記の延べ語数 2527 に合わせて，他のすべての作品が単位語数を調整されている。それらの間で類似度を算出するとどのようになるか，というものである。表Ｄに倣うならば，表Ｅである。ただし，作品は延べ語数が大きい順に掲げる。本来の類似度との数値の差は，千分の数十であり，数字がやや混み合う。

本来の類似度と，単位語数 2527 に調整された類似度とは，調整されたほうが全体的に小さく，差が源氏物語・枕草子 −64 と 徒然草・方丈記 +6 との間にある。もとの延べ語数が小さい作品では，差の範囲も小さく，方丈記 −6〜+4，土左日記 −11〜+1，竹取物語 −17〜−1，伊勢物語 −28〜−1 である。和歌集でも，差の範囲が小さく，万葉集 −32〜+1，新古今集 −34〜−2，後撰集 −34〜−0，古今集 −34〜+2 である。差を幅 10 でまとめると，次のようである。

表Ｅ

		平家	万葉	宇治	枕	大鏡	蜻蛉	新古	徒然	
	源氏	.405488	.293418	.489242	.571628	.527175	.590332	.354508	.510674	
平家	.3593−46	平家	.271060	.486336	.399838	.451088	.383441	.268747	.440582	平家
万葉	.2758−17	.2504−21	万葉	.310485	.296235	.253342	.337238	.452296	.292227	万葉
宇治	.4364−53	.4226−63	.2940−16	宇治	.545748	.507575	.524694	.303263	.535947	宇治
枕	.5081−64	.3541−46	.2675−28	.4930−53	枕	.504855	.579150	.305369	.511751	枕
大鏡	.4697−57	.3957−55	.2330−20	.4570−51	.4523−53	大鏡	.457965	.255818	.466623	大鏡
蜻蛉	.5350−55	.3442−39	.3169−20	.4843−41	.5259−53	.4226−35	蜻蛉	.385068	.494277	蜻蛉
新古	.3359−19	.2497−19	.4242−28	.2853−18	.2849−20	.2354−21	.3620−23	新古	.312719	新古
徒然	.4693−42	.3842−57	.2793−13	.4835−52	.4681−44	.4178−49	.4558−38	.2940−19	徒然	
後撰	.3620−16	.2572−20	.4693−24	.3106−18	.3169−20	.2568−17	.4024−22	.6042−34	.3359−11	後撰
古今	.3363−18	.2504−19	.4903−32	.3039−15	.3062−20	.2512−15	.3802−25	.5916−33	.3233−15	古今
紫	.5294−40	.3272−28	.2429−10	.3992−27	.5124−42	.4352−31	.4542−27	.2734−13	.4052−30	紫
更級	.5017−22	.3577−20	.3288−16	.4752−21	.5045−30	.4202−17	.5346−26	.3937−20	.4416−20	更級
伊勢	.4281−20	.3276−20	.3727−20	.4428−18	.4329−28	.3692−20	.4815−26	.3937−19	.4218−21	伊勢
竹取	.4420−14	.3609−14	.2979−08	.5092−14	.4527−15	.4131−17	.4566−14	.2805−08	.4451−09	竹取
土左	.3688−00	.3055−03	.3395+01	.4190−07	.4016−07	.3371−05	.4432−10	.3252−06	.3929+01	土左
方丈	.3763−00	.3343−02	.2683+00	.3755+01	.3411+00	.3256−01	.3723−00	.3126−02	.4349+06	方丈
		平家	万葉	宇治	枕	大鏡	蜻蛉	新古	徒然	

		後撰	古今	紫	更級	伊勢	竹取	土左	方丈	
	源氏	.377947	.353971	.569244	.523984	.447681	.455774	.369344	.376336	源氏
	平家	.277331	.269322	.354630	.377858	.347684	.375330	.309201	.336297	平家
	万葉	.492621	.521621	.252513	.345315	.393001	.305506	.339487	.267700	万葉
	宇治	.329349	.318774	.426384	.495847	.461403	.522739	.426018	.375082	宇治
	枕	.337144	.325587	.554268	.534538	.460558	.468350	.408761	.340913	枕
	大鏡	.274005	.266026	.466398	.437159	.388797	.429895	.342322	.326667	大鏡
	蜻蛉	.424007	.404815	.480596	.560961	.507937	.471343	.452675	.372456	蜻蛉
	新古	.637668	.625481	.286286	.413942	.413282	.288514	.331276	.314857	新古
	徒然	.346852	.338271	.435328	.462330	.443317	.454215	.391718	.428680	徒然
		後撰	.735076	.308550	.435252	.454290	.333727	.371882	.336771	後撰
	古今	.7012−34	古今	.300440	.410496	.462329	.318408	.365714	.326055	古今
	紫	.2983−11	.2876−12	紫	.463930	.391163	.392664	.340376	.317321	紫
	更級	.4159−19	.3913−19	.4487−15	更級	.497909	.470323	.440149	.378315	更級
	伊勢	.4396−14	.4495−12	.3747−16	.4776−20	伊勢	.457624	.476234	.352490	伊勢
	竹取	.3260−08	.3047−13	.3802−13	.4582−12	.4483−10	竹取	.434419	.356161	竹取
	土左	.3688−03	.3609−05	.3324−08	.4305−09	.4653−11	.4309−03	土左	.331794	土左
	方丈	.3367−00	.3280+02	.3177+01	.3822+04	.3510−01	.3549−01	.3312−01	方丈	
		後撰	古今	紫	更級	伊勢	竹取	土左		

＼差	−69〜0	−59〜0	−49〜0	−39〜0	−29〜0	−19〜0	−09〜0	+00〜9	平均
源氏	1	3	3		2	5	2		−30. 1875
平家	1	2	2	1	5	3	2		−29. 5
万葉				1	7	5	1	2	−17
宇治	1	4	1		2	6	1	1	−29. 125
枕	1	3	3	1	5	1	1	1	−32. 6875
大鏡		4	1	2	3	4	2		−29
蜻蛉		2	1	3	7	2	1		−28. 375
新古				2	5	6	3		−18. 875
徒然		2	3	2	2	4	1	2	−25. 8125
後撰				2	4	7	3		−16. 9375
古今				3	2	9	1	1	−17. 8125
紫			2	2	3	7	1	1	−20. 125
更級				1	7	6	1	1	−17. 625
伊勢					8	7	1		−17. 25
竹取						10	6		−10. 8125
土左						2	12	2	−4. 75
方丈							9	7	+0. 375
全体	2	10	8	10	31	42	24	9	−20. 323529

類似度の最後に，もとの作品と，単位語数を 2527 に調整した作品との間のものを並べる。調整した作品は，もとの作品の部分集合である。類似度の計算は出現比率による。類似度は，もとの延べ語数が小さい作品，また和歌集で，大きい。

源氏物語	. 745871
平家物語	. 630167
万葉集	. 746287
宇治拾遺物語	. 741222
枕草子	. 754603
大鏡	. 759918
蜻蛉日記	. 797512
新古今集	. 832831
徒然草	. 738370
後撰集	. 860100
古今集	. 853290
紫式部日記	. 805878
更級日記	. 844305
伊勢物語	. 865994
竹取物語	. 902758
土左日記	. 923885
方丈記	1.

（記）

2019年 8月25日，講演会の発表原稿に手を加え，付論を添えて，ウェブに挙げた。

同年10月22日，城阪早紀さんの指摘を受けて，発表原稿・付論を修正した。

お読みくださったかたがたに感謝します。