



Title	「学会抄録コーパス」にあらわれた概数をあらわす語
Author(s)	小林, ミナ; Kobayashi, Mina
Citation	北海道大学留学生センター紀要, 3, 55-67
Issue Date	1999-12
Doc URL	<a href="https://hdl.handle.net/2115/45577">https://hdl.handle.net/2115/45577</a>
Type	departmental bulletin paper
File Information	BISC003_005.pdf



# 「学会抄録コーパス」にあらわれた 概数をあらわす語<sup>1)</sup>

小 林 ミ ナ

## 要 旨

理系学会大会の抄録論文等を第一次資料とするコーパス (=学会抄録コーパス) を作成し、「概数をあらわす語 (約、ほぼ、およそ等)」の意味、用法について、特に後接する語の種類に注目して分析した。それにより、次の5点がわかった。1) コーパスにあらわれた7種類の「概数をあらわす語」は、「頻繁に使用される語」、「あまり使用されない語」、「ほとんど使用されない語」の3つの語群に分類できる。2) 「約」と「ほぼ」には「約+数値」、「ほぼ+数値以外」というすみわけがみられる。3) 「約+数値以外」の用例はすべて「半～」という語句であった。このことから、「約」に後接可能という意味では「半数、半生、半分」は「数値」と等しいと言える。4) 「ほぼ+数値」の半数以上の用例が「極限の意味を含む数値あるいは文脈」であり、これは「ほぼ+数値以外」の用例にも共通する特徴である。5) 「ほぼ+数値以外」が使われるのは、「ある基準あるいは定義に照らしてその状態である」といった文脈であることが非常に多い。

これらの知見は母語話者の内省や既存の国語辞書の記述からは得られないものである。また、日本語教育への還元という意味では、適切な産出のために必要な情報提供という点で有益なものと思われる。

〔キーワード〕 コーパス、学会抄録論文、概数をあらわす語、約、ほぼ、およそ

## 1. はじめに

コーパス (corpus) とは、広義には「紙に印刷された文字資料 (書籍、パンフレット、ちらし等) や録音資料等の第一次資料の総体」、狭義には「コンピュータ処理が可能な形で処理、蓄積された言語資料の総体」をいう (齊藤他1998、加藤1999、滝沢1999等)。広義、狭義のコーパスを明示的に区

別するために、前者を単なる「コーパス」、後者を「コンピュータ・コーパス (computer corpus)」あるいは「電子コーパス (electronic corpus)」と言い分けることもある。本稿では、「コーパス」という語を狭義のコーパスに限り論を進めることにする。

実際のコミュニケーションに役立つ日本語を教えるには、ことばが実際にどのように使われているかを先ず知ることが必要である。加藤 (1999) は、国定読本 (明治37年～昭和22年) を第一次資料とするコーパスを用いて、副助詞「くらい」と「ぐらい」の意味、用法を分析している。「くらい」と「ぐらい」は、国語辞典などでは等価のように記述されることが多い。しかし、両者には「後続しやすい品詞」と「使われやすい意味」の2点において違いがみられることを、コーパスを用いた分析により明らかにしている。

また、田野村 (1995) は、朝日新聞の6年分 (1987～1992年) の記事から動詞「そびえる」の用例を検索している。「そびえる」は、「いつも『ーている』の形で状態を表すのに用い (金田一1976: 8)」られることから、いわゆる「第四種の動詞」とされる。しかし、データから得られた実例によれば、「主節の言い切りの位置に限っても、『そびえる』が三〇例、『そびえている』が三五例と、内省による予想が完全に裏切られる結果」(田野村1995: 54) であることがわかったことが報告されている。

このように、コーパスを利用して大量の生データを扱うことにより、母語話者の内省を補完、実証するだけでなく、言語使用の実態についてより詳細な現象記述が可能になる。そしてその成果は、日本語教育にとってもきわめて有効な情報となるはずである。

筆者は、全国規模の理系学会大会で発行された抄録論文等 (詳細は後述) を第一次資料とするコーパス (以下、「学会抄録コーパス」) を作成し、いくつかの語<sup>2)</sup> の出現環境について分析した。本稿では、このうち「概数をあらわす語」の意味、用法について、特に後接する語の種類に注目して分析、考察する。

## 2. 「概数をあらわす語」

### 2.1 「概数をあらわす語」とは何か

本稿でいう「概数をあらわす語」とは、数値などに前接し「おおよその数」を示すのに用いられる語のことをいう。具体的には、1) ～ 3) にみ

られるような「約、およそ、ほぼ」等の語（下線部）を指す。

- 1) 浸出されてきた鉄イオンの約90%が二価鉄イオンとして存在していた。
- 2) 当施設における動物病院関連の鳥類事例の鑑定数はこの5年間に限るとおよそ310件で、これは漸増傾向にある。
- 3) 海外の報告では、上皮系腫瘍が60~70%を占め、腺癌が最も多く、間葉系腫瘍が約30%で、軟骨肉腫が多いといわれており、本病院における結果とほぼ一致していた。

## 2.2 内省による観察

「概数をあらわす語」は、次の4)、5)に見られるように、数値やある種の語に前接し、「おおよその値であることを示す」という点では共通している。

- 4) 浸出されてきた鉄イオンの {およそ／だいたい／ほぼ／約} 90%が二価鉄イオンとして存在していた。
- 5) 当施設における動物病院関連の鳥類事例の鑑定数はこの5年間に限ると {およそ／だいたい／ほぼ／約} 310件で、これは漸増傾向にある。

しかし、被修飾語の制限や副詞以外の用法の有無についてはいくつかの違いがある。

一つは、どのような語を修飾できるかという被修飾語の制限である。下の6)、7) からわかるように、「およそ、だいたい、ほぼ」は数値以外の語句（ここでは動詞や形容動詞などの用言）を修飾できるのに対して、「約」はそのような用法をもたない（\*は不適切な用法をあらわす）。

- 6) a 3 と b 3 の値は、気乾状態と湿潤状態とで {およそ／だいたい／ほぼ／\*約} 一致している。
- 7) また、軸ひずみ速度は時間の経過とともに {およそ／だいたい／ほぼ／\*約} 一定の割合で減少し、最小値に至った後、再び増加し、試験片が破断に至る傾向を示すことなどが認められた。

二つ目に、名詞<sup>3)</sup>としてふるまうか否かがある。8)、9) からわかるように、「およそ、だいたい」は「の」を介して名詞を修飾するという意味において、それ自身が名詞としてふるまえるのに対して、「約、ほぼ」はそうではない。

8) {およそ/だいたい/\*ほぼ/\*約} の見通しをたてる。

9) {およそ/だいたい/\*ほぼ/\*約} の計画が発表された。

最後に、それぞれの語が用いられる文体レベルがある。「約、およそ、ほぼ」と異なり、「だいたい」には口語的なニュアンスが感じられる。そこで、研究論文や新聞記事といったあらたまった書きことばには、「だいたい」はあらわれにくいことが予想できる。

## 2.3 国語辞書における記述

次に、これらの語が既存の国語辞書でどのように記述されているかを見してみる。例えば、『日本語大辞典第二版』（講談社）には、以下の記述がある（該当部分についての記述のみ掲げる）。

やく【約】⑤おおよそ。ほぼ。[用例]（連体）－半月。－五カ年。

およそ【凡そ】《「おおよそ」の約》[一]（名）あらましのこと。概略。大略。outline [用例]－の見当。－の見通しをつける。[二]（副）①だいたいのところ。about; some [用例]－百人ほど集まった。

おおよそ【大凡】[一]（名）物事のあらまし。大要。outline [用例]－のところを話せば。[二]（副）こまかい部分を省略すると。おおかた。だいたい。ふつう。generally [用例]－人間というものは。

ほぼ【略粗】（副）おおむね。おおかた。だいたい。almost [用例]－完成した。

だいたい【大体】[一]（名）全体のうちのおおよそのこと。主要な部分。あらましのこと。大略。outline [用例]準備の－ができた。－の説明を聞く。[二]（副）①大部分。総じて。たいがい。generally [用例]－書き終えた。－わかった。②およそ。ほぼ。約。about; roughly [用例]－いつごろまでかかるかな。

やく【約】、およそ【凡そ】の語義・用法説明からは、「約半月」、「およそその見通しをつける」といった語連鎖が日本語として可能であることはわかる。しかし、だいたい【大体】の項をみると、「②およそ。ほぼ。約。about; roughly [用例]－いつごろまでかかるかな。」とあり、この説明では「いつごろまでかかるかな」という文には、「だいたい／およそ／ほぼ／約」のどれでもが使えるとも解釈可能である。（しかし、実際は「\*ほぼ

／\*約いつごろまでかかるかな」はおかしい)。このように、辞書ではそれぞれの語が相互的に語義説明として使われており、各語の違いについては明らかではない。

### 3. 「学会抄録コーパス」について

#### 3.1 第1次資料

「学会抄録コーパス」の第1次資料としたのは、理系学会の全国大会で発行された予稿集に掲載された抄録論文（1～5）、講演要旨（6）、および、学会誌に掲載された研究論文（7）の7種類である。学会大会名、論文本数等の情報を表1に掲げる。

表1：第1次資料

大会（誌）名【本稿での略号】	論文本数	論文1本あたりの平均字数
1. 資源・素材学会平成5年度春季大会【資】	195	2200
2. 平成6年度日本小動物獣医学会年次大会【小】	80	1300
3. 平成6年度日本獣医公衆衛生学会年次大会【公】	32	1700
4. 平成6年度日本産業動物獣医学会年次大会【産】	49	1600
5. 第119回日本獣医学会（発表要旨）【獣】	404	700
6. 第119回日本獣医学会（講演要旨）【獣】	64	2200
7. 電子情報通信学会論文誌vol.J80-D・【電】	36	9000

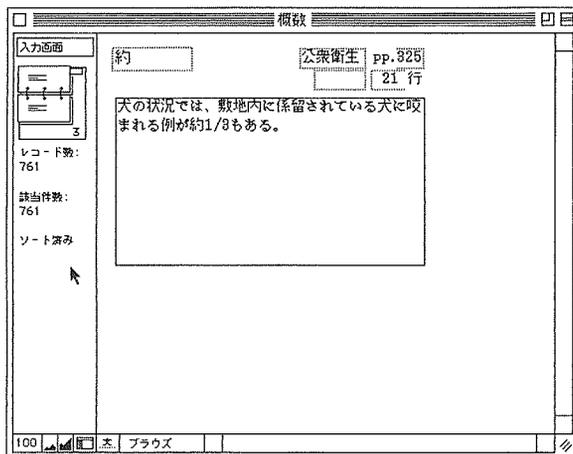
表1にみられるように、一口に抄録論文と言っても論文1本あたりの分量（文字数）は、学会によってさまざまである。また、6. は講演要旨、7. は研究論文（フルペーパー）のため1本あたりの文字数が1～5より多い。（コーパスの分量・規模を示すとき、英語のように分かち書きを用いる言語の場合は語数で示するのが一般的であるが、日本語の場合は語の認定単位が問題となる。そこで、本稿では分量を字数で示した。語数、文字数以外では、1.7MB等といったデータ・サイズで示す方法もある。）

#### 2.2 学会抄録コーパスの設計

表1に示した第1次資料から、「概数をあらかず語」を含む文をすべて抽出した。そして、それをカード型データベースソフトである「ファイルメーカーPro 3.0v1」（米国クラリス社製）を用いて、1文を1レコード

とするデータベースを作成した。データ画面の例を、図1に示す。

図1：データ画面



園田（1999：2）は、データ分析に使用可能なコンピュータ・ソフトウェア、プログラム言語の観点から、コーパスを次の5段階に分類している。

- (a) Excel などの表計算ソフトあるいはFile Maker, Access などの汎用のデータベースソフト
- (b) Wordsmith, Corpus Wizard などのテキストデータ分析専用ソフト
- (c) Grep 単独あるいはgrep, sort, sed の組み合わせ
- (d) AWK, Perl, Snobol, Icon などのテキスト処理が強化されたプログラム／スクリプト言語
- (e) C, Pascal Fortran などの汎用プログラム言語

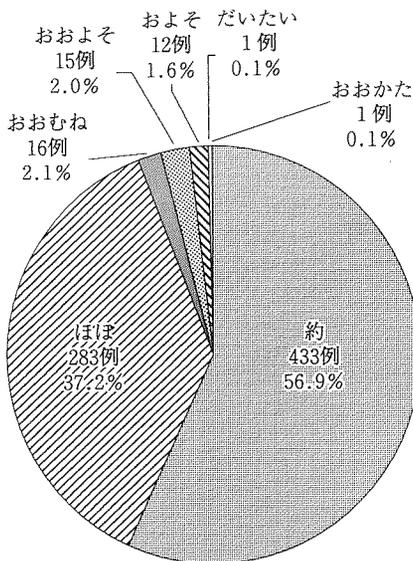
この分類にしたがえば、本稿の学会抄録コーパスは(a)の段階に位置づけられる。

## 4. 結果と考察

### 4.1 全体の傾向

得られた結果について、全体の傾向を把握する。使われていた語と用例数を、グラフ1として示す。

グラフ1：全体の傾向



グラフ1をみると、学会抄録コーパスにあらわれた「概数をあらわす語」は、用例数の観点から、「頻繁に使用される語（約、ほぼ）」、「あまり使用されない語（おおむね、おおよそ、およそ）」、「ほとんど使用されない語（だいたい、おおかた）」の3つの語群に分類できることがわかる。「約（433例）」と「ほぼ（283例）」の合計は716例であり、この2語だけで用例全体の約94%を占めている。

また、「だいたい、おおかた」の用例が少ない（各1例）ことは、使用される文体に関する内省と一致する。しかし、「おおむね、おおよそ、およそ」の用例が、「約、ほぼ」に比してこれほど少ない（それぞれ16、15、12例）ことは内省からは予想することはできない。

#### 4.2 後接する語からみた「約」と「ほぼ」の使い分け

ここでは、頻繁に使用されていた「約」と「ほぼ」の2語を取り上げ、両者の意味、用法にどのような違いが見られるかについて、特に後接する語に注目して分析する。それぞれの用例を「後接する語が数値か否か」で分類したものが表2である。表2をみると、「約」に後接する語には「数値（428例）＞数値以外（5例）」、「ほぼ」に後接する語には「数値（15例）＜数値以外（268例）」という傾向があることがわかる。

表2：後接する語からみた「約」と「ほぼ」の用例

( [ ] の数字は用例数をあらわす)

語	全用例数	数値が後接する例	数値以外が後接する例
約	433	428	5
		表面粗さ約 $4 \mu\text{m}$ 【資】 同一配合で約 $300\text{kgf}/\text{cm}^2$ のコンクリート強度差が生じ 【資】 K S 86 ( + ) 株接種後約 1 ヶ月及び約 10 ヶ月で 【獣】 約 1/2 の隠れユニット数 【電】 …他多数	約半数 【小】 【公】 約半数例 【獣】 約半生に近い 60 年余にわたり 【獣】 約半分 【資】
ほぼ	283	15	268
		ほぼ 0 (ゼロ) となった [ 4 ] 【資】 ほぼ 100% の個体 【獣】 その消失にはほぼ 10 週以上 【獣】 ほぼ 80 日齢で終了 【獣】 羽との距離がほぼ 20 m に達した時点 【資】 平均的にはほぼ 0.04 ~ 5% 程度 【獣】 ほぼ 0.04 ~ 5% 程度 【資】 貯留層となるほぼ 1500 m 以深の花崗閃緑岩 【資】 ほぼ 2.5% おきに設定した 【資】 常温付近でほぼ $5 \times 10^{-6} \text{C} - 1$ 前後であった線膨張率 【資】 アセトニトリル 3 ml の溶出でほぼ 90% 以上 【公】 ほぼ $5 \times 10^{-5} \sim 1 \times 10^{-4} \text{mol} \cdot \text{dm}^{-3}$ の A 1 濃度範囲 【資】	ほぼ一致 / 同じ / 同レベル [117] ほぼ直線的 / 水平 / 正常 [115] ほぼ全域 / 全長 / 全量 [27] ほぼ対応する / 判定がつく [9]

「約」について言えば、全433例のうち「数値以外が後接する例」がわずか5例(1.1%)しかみられなかったことは、「約+数値以外(\*約水平)」といった用法が不適切であるという内省から予想、説明が可能である。また、「約+数値以外」の5つの用例をみると、後接しているのは「半数(2例)、半数例、半生、半分(各1例)」のように、すべて「半~」とう語句である。このことから「半数、半数例、半生、半分」は、「約」との共起という点では数値と同じようにふるまうと言える。

これに対して「ほぼ」については、「ほぼ+数値(ほぼ50%)」、「ほぼ+数値以外(ほぼ水平)」のどちらも内省では十分あり得る語連鎖である。

したがって、「数値が後接する例」が少ない(15例、5.3%)という事実は、内省では予想、説明することができない。そこで、「ほぼ」が後接する語に注目してみる。まず、「ほぼ+数値」の15例を見ると、1/3にあたる5例が「ほぼ100%の個体」、「ほぼ0」のように「極限の値を示す語(下線)」と共起していることがわかる。さらに3例が「その消失にはほぼ10週以上」、「ほぼ80日齢で終了」、「ほぼ20mに達した」のように「極限の意味を含む語(下線)」と共起している。このデータを見る限り、「ほぼ」が数値と共起するのは何らかの形で極限の意味を含む文脈が多い(53.3%)と言える。もしこの指摘が正しければ、「ほぼ」が使われるためには、単に概数であるだけでなく極限の意味を含む文脈が整うことが必要であり、それが「ほぼ+数値」の用例数を少なくしている理由とも考えられる。そして、「約」については、このような文脈の要請はなく、単に概数であればよいので、数値との結びつきが「ほぼ」より多く見られると考えられる。

次に、「ほぼ+数値以外」の用例を見てみる。268の全用例のうち、117例(41.3%)が「一致、同じ、同～」といった「同一であることをあらわす語」、115例(40.6%)が「水平、直線的、正常」といった「状態をあらわす語」と共起していることがわかる。しかし、「状態をあらわす語」と言っても、例えば「斜め、曲線的」等と共起した用例はデータには見られず、また「ほぼ斜め、ほぼ曲線的」のような語連鎖は内省でも不自然に感じられる。「水平、直線的」といった語と「斜め、曲線的」といった語の違いを考えてみると、「水平」とは「地球の重力と直角に交わる方向」であり厳密に定義された用語である。これに対して「斜め」というのは、「正面/真っ直ぐ/水平でない」といった意味で日常的に使われる語である。また、「直線」とは「二点を結ぶ線(=曲線)」の中で「最短距離で結んだ線」だけをいう。すなわち、前者(水平、直線的)は、後者(斜め、曲線的)に比して、より厳密な定義や基準などに照らした狭い意味で用いられる語だと言える。このことから、「ほぼ」が数値以外の語と共起するのは、「ある基準あるいは定義に照らしてその状態であるとみなせる」という文脈であることが非常に多い(81.9%)と言えるだろう。

「ほぼ+数値以外」で次に多かったものが、「ほぼ全長」、「ほぼ全量」といった「全数/全量を示す語(下線)」と共起する用例で、全268例のうち27例(10%)がこれに該当する。この事実は、「ほぼ+数値」が極限の意味を含む文脈で使われることが多いことと通じるものと思われる。

る。すなわち、「100%」と「全量」は、(数値か語句かの違いはあるが)どちらも極限を示すという意味においては共通するからである。さらに、残る9例は「対応する」、「判定がつく」等、いずれも終点を含む意味の動詞または動詞句であった。このことも、「ほほ」と「極限の意味を含む文脈」の共起性で説明することが可能である。

## 5. 日本語教育への還元可能性

「概数」という概念は、その言語形式はさまざまであるが、おそらくどの言語にも普遍的に存在するものであろう。この前提が正しければ、日本語学習者がこれらの語(を含む文)に接したとき、おおよその意味を理解することはそれほど難しいことではない。しかし、2.3.でみた国語辞書の記述からもわかるように、各語の違いに目を向けてみると、網羅的な現象記述、適切な用法分類がなされているとはいいがたい。言い替えば、日本語学習者にとっては、適切な産出に必要な情報が準備されていないのが現状である。

本稿の学会抄録コーパスを用いた研究により、「概数をあらわす語」について次の5つの知見を得た。

知見1 使われる頻度から3つの語群に分類できる。

知見2 「約」と「ほほ」には「約+数値」、「ほほ+数値以外」というすみわけがみられる。

知見3 「約+数値以外」の用例はすべて「半～」という語句である。このことから、「約」に後接可能という意味では「半数、半生、半分」という語は「数値」と等しいと言える。

知見4 「ほほ+数値」の半数以上の用例が「極限の意味を含む数値あるいは文脈」であり、それは「ほほ+数値以外」の用例にも共通する特徴である。

知見5 「ほほ+数値以外」が使われるのは、ある基準あるいは定義に照らして「その状態である」といった文脈であることが非常に多い。

知見1については、「作文教育」あるいは「学術研究のための日本語教育(Academic Japanese)」のシラバス・デザインに役立つ情報である。2.2.で述べた内省による記述の実証的な裏付けにとどまらず、「約」と「ほほ」をまず最優先に取り上げるべきであることが示唆された。

知見2については、「約+数値以外」という用例がほとんど見られないことは、内省でも容易に指摘できる。一方、「ほぼ+数値」の用例が少ないことは、内省だけでは気づかない事象であり、「数値との共起」という点では「ほぼ」より「約」の導入、説明の方が優先されるべきと言える。

さらに知見3からは、一部の「半～」という語句が数値と等しく「約」と共起可能であることを紹介する必要がある。また、あわせて「半～」と同様に数値をあらわすとも思える「全～（全数、全量、全員）」は「約」と共起しないという否定的情報を与えることも有益と思われる。

知見4、5は、「ほぼ」そのものについての語義、用法記述をさらに精緻化するために有効な情報である。そしてこれは、日本語教育のみならず日本語学にも還元可能である。

## 6. おわりに

残された課題を3つあげる。一つは後接要素の問題である。「概数をあらわす語」には、本稿でとりあげた前接要素のほかに、「ぐらい（1ccぐらい）、前後（1.8前後）、ほど（60%ほど）」といった後接要素がある。本稿では考察の対象とはしなかったが、今後はそれらの意味や機能も含んだ考察が必要である。二つ目に、分野による違いがある。本稿では「理系学会大会の抄録論文」をひとしなみに扱った。しかし、研究領域やトピックによってそれぞれの語の使われ方に違いが見られる可能性もある。三つ目としては、抄録論文には限られた字数、スペースで研究成果をまとめるという必然的な制約がある。「約」は1文字だが「およそ」は3文字であるといったような字数のちがいが、語の選択に影響を与えているかもしれない。このようなことも含めたさらに詳しい考察は、今後の課題としたいと思う。

本稿で述べた学会抄録コーパスは、コーパスの設計としては発展途上の段階にあり、規模も決して大きなものではない。しかし、それに関わらず内省の補完以上の知見が得られたことの意義は大きい。それは、「概数をあらわす語」について詳細な現象把握ができたというだけでなく、コーパスを使った日本語研究の成果が日本語教育に具体的に貢献できることが示唆された、という2点においてである。

注：

- 1) 本研究は、平成9～10年度文部省科学研究費補助金奨励研究(A)「研究留学生のための作文教材開発に関する基礎的研究」(研究代表者小林ミナ、課題番号09780199)による研究成果の一部である。また本論文は、1999年3月に行われた「北海道大学留学生センター日本語・日本語教育講演会 コーパス言語学と日本語教育」で行った口頭発表の内容に加筆修正を加えたものである。発表に際し有益なコメントをくださった参加者の皆様に感謝の意を表します。
- 2) 「概数をあらわす語」以外に、「接尾辞『～化』」の前接関係と生産性、「結論、仮説、考察」の談話中でのあらわれかたを考察した。
- 3) 「約」のこのような用法については、「副詞」とみなす立場と「連体詞」とみなす立場の二つがみられる。

参考文献：

- 加藤安彦 1998 「辞典とコーパス」『日本語学』第17巻12号、PP.37-44、明治書院
- 金田一春彦 1950 「国語動詞の一分類」『言語研究』15号
- 金田一春彦 1976 「国語動詞の一分類」金田一春彦編『日本語動詞のアスペクト』、むぎ書房
- 斎藤俊雄他編 1998 『英語コーパス言語学－基礎と実践－』、研究社出版
- 園田勝英 1999 「コーパス言語学と日本語教育」北海道大学留学生センター日本語・日本語教育講演会 コーパス言語学と日本語教育」配布資料
- 滝沢直宏 1999 「コーパスと語法・文法研究」北海道大学留学生センター日本語・日本語教育講演会 コーパス言語学と日本語教育」配布資料
- 田野村忠温 1995 「パソコン利用の現状と課題 意味」『日本語学』第14巻8号、PP.53-62、明治書院

## Words expressing approximation : A study based on an academic proceedings corpus

KOBAYASHI, Mina

This paper investigates the usage of words expressing approximation. Careful observation based on an academic proceeding corpus suggests the following findings.

- 1) The following seven words expressing approximation are observed in the corpus: *yaku*, *hobo*, *oomune*, *ooyoso*, *oyoso*, *daitai* and *ookata*. They can be classified into three groups by frequency: frequent words (*yaku*, *hobo*), relatively infrequent words (*oomune*, *ooyoso*, *oyoso*), and rarely used words (*daitai*, *ookata*).
- 2) Both *yaku* and *hobo* are used frequently, but they differ in the kind of words which commonly follow them: in the vast majority of cases, *yaku* is used with numbers and *hobo* with non-numbers.
- 3) Number words which collocate with *yaku* include *hansuu*, *hansei*, *hanbun*.
- 4) Over half the combination of *hobo* + number carry the meaning of a limit. This meaning is also observed in combinations of *hobo* + non-number
- 5) Combinations of *hobo* + non-number are often used in sentences with the contextual meaning 'be in a certain condition accruing to some criterion'.

These findings could not be arrived at through the intuition of native speaker or Japanese dictionaries, and they provide useful information for acceptable Japanese production.