

TR-I-0157

対話データベースからの頻度情報の抽出
Token Count Statistics Extracted from
the ATR Dialogue Database

江原暉将
Terumasa EHARA

1990.4

概 要

対話に関する言語モデルを作成するための基礎資料を得ることを目的として、ATR対話データベース(ADD)から、種々の言語事実に関する頻度情報を抽出した。抽出データの内容としては、記録情報の単純な度数、言語単位の各階層における構成要素数、日英対応の対応/非対応の度数、日英対応の対応要素数、言語単位の内部構造である。使用したデータは、国際会議に関する電話会話(54, 931語)とキーボード会話(39, 606語)である。

抽出データから次の様なことが明らかになった。出現単語について、電話会話とキーボード会話を比較したところ、間投詞や、言い淀み、言い直しが電話会話に特有であり、また、電話会話の語彙がキーボード会話のそれを包含する傾向にある。文節パターンについては、間投詞などを除き、電話とキーボードで、あまり差がない。文パターンについては、電話会話の方が文の長さのばらつきが大きい。しかし、文型については、パターンの度数が少ないために電話とキーボードの差を明らかにすることが出来なかった。日英対応では、言語単位が長くなるほど対応率が増す。特に、キーボード会話の文節対応は67%と単語対応の3.2倍に達しており、文節対応を利用することの有効性が示唆される。

TR-I-0157

Token Count Statistics Extracted from
the ATR Dialogue Database

Terumasa EHARA

1990.4

Abstract

ATR is constructing a Dialogue Database named ADD (ATR Dialogue Database). It is built from simulated telephone and keyboard conversations. This report shows the token count statistics of several linguistic phenomena of ADD and makes a comparison between telephone and keyboard conversations. The sample sizes are 54,931 words from the telephone conversations and 39,606 words from the keyboard conversations.

The following computations have been included in this report : counts of words, word categories, interjections, restatements, phrases, phrase patterns, clause patterns, sentence patterns, syntactic dependencies, semantic dependencies, word correspondences between Japanese and English, phrase correspondences, clause correspondences, sentence correspondences, utterance correspondences, word bigrams and word trigrams.

From the statistical data, we conclude the following. Words used in telephone conversations are a superset of those used in keyboard conversations. The phrase patterns of telephone conversations are nearly the same as those of keyboard conversations. Telephone conversations tend to be more varied in sentence length than keyboard conversations. We were unable to extract the statistics for sentence patterns because the sample size was insufficient.

目次

1. はじめに	1
2. データベースの内容	1
3. 抽出データの内容	1
4. 抽出方法	2
5. 抽出結果	2
6. 抽出結果の考察	2
7. 結論	4
参考文献	5
付録1 使用したデータの会話ID	6
付録2 抽出データの詳細内容	10
付録3 抽出プログラム例	27
付録4 抽出結果の一部	30
付録4.1 キーボード会話	31
付録4.2 電話会話	50
付録4.3 抽出結果のグラフ	74

1. はじめに

A T R で構築している対話データベースから各種言語現象に関する頻度情報を抽出し、結果を考察した。これは、対話に関する言語モデルを作成するための基礎資料を得ることを目的として行ったものである。

2. データベースの内容

頻度情報抽出に用いたデータベースはA T R 対話データベース (A D D) の中の付録1に示す会話IDを持つものである。領域は国際会議であり、メディアは電話 (t e l) とキーボード (k e y) である。言語単位の総数を表2. 1に示す。

表2. 1 データベースの大きさ

	t e l	k e y
会話数	4 7	9 5
発話数	6 2 4	9 0 4
文数	1 8 5 7	2 5 0 8
文節数	1 9 9 2 3	1 3 8 6 4
形態素数	5 4 9 3 1	3 9 6 0 6

3. 抽出データの内容

抽出した言語現象は以下の項目である。詳細な説明を付録2に示す [注1]。

(1) 記録情報の度数

- ・ 間投詞 (SURF-j_inter, SURF-e_inter)
- ・ 言い淀み、言い直し (SURF-j_ref, SURF-e_ref)
- ・ 出現単語 (JWRD-word, EWORD-word)
- ・ 読み (JWRD-yomi)
- ・ 標準表現 (JWRD-regu)
- ・ 標準表現と品詞 (JWRD-regu_parts)
- ・ 品詞 (JWRD-parts)
- ・ 係り受け構文関係コード (JREL-syn)
- ・ 係り受け意味関係コード (JREL-sem)

(2) 言語単位の各階層における構成要素数

- ・ 1文節中の単語数 (COMP-word_in_phrase)
- ・ 1格構造中の単語数 (COMP-word_in_kaku)
- ・ 1格構造中の文節数 (COMP-phrase_in_kaku)
- ・ 1文中の単語数 (COMP-word_in_sent)
- ・ 1文中の文節数 (COMP-phrase_in_sent)
- ・ 1文中の格構造数 (COMP-kaku_in_sent)
- ・ 格構造の埋め込み数 (COMP-kemb)
- ・ 1発話中の文数 (COMP-sent_in_utter)
- ・ 1会話中の文数 (COMP-sent_in_conv)

- ・ 1 会話中の発話数 (COMP-utter_in_conv)
- (3) 日英対応の対応／非対応の度数
- ・ 単語対応 (JE-word)
 - ・ 文節対応 (JE-phrase)
 - ・ 格構造対応 (JE-kaku)
 - ・ 文対応 (JE-sent)
 - ・ 発話対応 (JE-utter)
- (4) 日英対応の対応要素数
- ・ 単語対応の単語数 (JECOMP-word)
 - ・ 文節対応の単語数 (JECOMP-word_in_phrase)
 - ・ 格構造対応の単語数 (JECOMP-word_in_kaku)
 - ・ 格構造対応の格要素数 (JECOMP-kelem)
 - ・ 格要素対応の単語数 (JECOMP-word_in_kelem)
 - ・ 文対応の文数 (JECOMP-sent)
 - ・ 文対応の単語数 (JECOMP-word_in_sent)
 - ・ 発話対応の発話数 (JECOMP-utter)
 - ・ 発話対応の文数 (JECOMP-sent_in_utter)
- (5) 言語単位の内部構造
- ・ 文節中の単語の並び (STRUCT-word_in_phrase)
 - ・ 単語バイグラム (STRUCT-bigram)
 - ・ 単語トライグラム (STRUCT-trigram)
 - ・ 文中の単語の並び (STRUCT-word_in_sent)
 - ・ 文末の文節中の単語の並び (STRUCT-word_in_last_phrase)

[注1] 本抽出作業の実際は東洋情報システムが行った [1]。

4. 抽出方法

抽出はRDBからsqlおよびシェルスクリプトによって行った。 [1]

5. 抽出結果

抽出結果を付録4に示す。

6. 抽出結果の考察

6. 1 記録情報の度数

6. 1. 1 間投詞 (SURF-j_inter, SURF-e_inter)

出現がtelに限られている。keyでは、uh,が2例あるのみである。telでの間投詞の使用は、かなり語彙が限られている。

6. 1. 2 言い淀み、言い直し (SURF-j_ref, SURF-e_ref)

()で句繰られた部分を抽出したので、keyでは言い淀み・言い直しとは異なる意味で用いられている場合も考えられる。

6. 1. 3 出現単語 (JWRD-word, EWORD-word, JWRD-yomi, JWRD-regu, JWRD-regu_parts)

JWRD-regu_partsとEWORD-wordのカバー率に関するグラフから出現単語はやや飽和傾向にある。上位500位で84%ないし92%のカバー率である。高頻度の語には、助詞や助動詞に混じって、t e lでは間投詞および、間投詞や言い淀みなどを表す括弧が目だっている。また、終助詞「ね」や形式名詞「方」が上位に来ているなどt e lがk e yを包含する傾向にある。

t e lでの人称の表現例は次の通りである。「私」は300例あるが、「あなた」は1例である。相手を表現すると思われる「先生」が121例、「そちら」が115例である。一方、英語では、<I>が708例、<you>が867例である。日本語での名詞や代名詞での人称表現が少ないことが分かる。

6. 1. 4 品詞 (JWRD-parts)

t e lで間投詞、記号が上位にある以外は、t e lとk e yで大きな差はない。

6. 2 言語単位の各階層における構成要素数

6. 2. 1 1文節中の単語数 (COMP-word_in_phrase)

1文節中の平均単語数はk e yで2. 8、t e lで2. 7とあまり変わらない。

6. 2. 2 1文中の文節数 (COMP-phrase_in_sent)

1文中の平均文節数は、k e yで5. 5であるが、t e lでは、10. 7である。これは、t e lで非常に長い文が存在するため、最長65文節のものが存在する。k e yでは、最長26文節である。しかし、むしろt e lでは短い文にも度数が集中する傾向があり、短い文と非常に長い文が混ざって存在していることが分かる。つまり、t e lはk e yに比較して、長さの分散が大きく、文のバラエティが富んでいると思われる。これは、計算機処理にとって、不利なことであろう。

6. 2. 3 1文中の単語数 (COMP-word_in_sent)

1文中の単語数も文節数と同様の傾向にある。平均単語数は、k e yで15. 8、t e lで29. 5である。

6. 2. 4 1発話中の文数 (COMP-sent_in_utter)

1発話中の平均文数は、k e yが2. 7、t e lで2. 9である。ここでも、t e lで、長い発話が存在する傾向にある。

6. 3 日英対応での対応／非対応の度数

k e y t e lともに単位長が長くなるほど対応の度数が増える傾向にある。ただし、格構造についてはそうになっていない。これは、格構造の定義が日本語に独特のものであるためと思われる。文節対応がk e yで65%に達しており、日英対応に基づく研究に文節対応が有効である可能性がある。ただし、t e lでは、文節対応でも、38%である。文や発話の対応率が高いのは当然であろう。

6. 4 日英対応の対応要素数

6. 4. 1 単語対応の単語数 (JECOMP-word)

ほとんどの対応が1対1である。k e yで87%、t e lで84%の対応が1対1である。

6. 4. 2 文節対応の単語数 (JECOMP-word_in_phrase)

日本語の1文節に対応する英語の単語数は単調に減少している。keyとtelでの著しい差はない。

6. 4. 3 文対応の文数 (JECOMP-sent)

ほとんどの対応が1対1である。keyで96%、telで87%が「日本語の文数」対「英語の文数」が1対1、1対2、2対1のいずれかの対応である。

6. 5 言語単位の内部構造

6. 5. 1 文節中の単語の並び (STRUCT-word_in_phrase)

文節中の単語の並びのパターンに関するカバー率のグラフから、ランク対頻度の分布はほぼ $1/r$ に比例していることが分かる。出現単語の場合と異なり、飽和傾向にはない。特に、ランクの大きな所でカーブが立つ傾向にあり、文節パターンが尽くされていないことを示している。度数1のパターンがkeyで63%、telで64% (共に異なり) を占めている。ここで、文節パターンとは、内容語については、品詞を、機能語については、読みを、記号については、語形を用いたパターンである。それぞれの語種の範囲は、次の通りである。

・内容語

形容詞	形容動詞ダ	形容動詞タ	普通名詞
サ変名詞	代名詞	数詞	副詞
自動詞	他動詞	固有名詞	形容名詞
本動詞			

・機能語

連体詞	接続詞	感動詞	助動詞
副助詞	接続助詞	格助詞	終助詞
接尾語	接頭語	補助動詞	間投詞
準体助詞	並立助詞	係助詞	慣用句
その他			

高頻度の文節パターンとしては、telで間投詞の文節が多い他は、keyとtelであまり差がない。

6. 5. 2 文中の単語の並び (STRUCT-word_in_sent)

文中の単語の並びパターンを文節の場合と同様に調べたが、度数1のパターンがkeyで95%、telで98% (共に異なり) であり、パターンが尽くされていない。これは、カバー率のグラフからも分かる。

6. 5. 3 文末の文節中の単語の並び (STRUCT-word_in_last_phrase)

keyとtelで文末文節のパターンには、大きな差がない。

7. 結論

ATR対話データベースを用いて、各種の統計情報を抽出し、結果を考察した。その結果次の様なことが明らかになった。

出現単語について、電話会話とキーボード会話を比較したところ、間投詞や、言い淀み、言い直しが電話会話に特有であり、また、電話会話の語彙がキーボード会話のそれを包含する傾向にある。間投詞などを除き、文節パターンについては、電話とキーボードで、あまり差がない。-文パターンについては、電話会話の

方が文の長さのばらつきが大きい。しかし、文型については、パターンの度数が少ないために電話とキーボードの差を明らかにすることが出来なかった。日英対応では、言語単位が長くなるほど対応率が増す。特に、キーボード会話の文節対応は67%と単語対応の3.2倍に達しており、文節対応を利用することの有効性が示唆される。

今後の課題として、文節をカテゴライズすることによって、文のパターンを考察すること〔2〕、領域を国際会議だけでなく、旅行などに広げ、領域による差を明らかにすることなどがあげられる。

参考文献

- [1] 東洋情報システム：言語データベースからの情報抽出作業 納品書類、1990. 3。
- [2] 国立国語研究所：談話語の実態、国研報告 8、1955。

付録1 使用したデータの会話ID

3,000番未満は電話会話であり、3,000番以上はキーボード会話である。

81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
108
109
110
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129

3045
3046
3048
3049
3050
3051
3052
3053
3054
3055
3056
3057
3058
3059
3062
3063
3064
3065
3066
3067
3068
3069
3070
3071
3072
3073
3074
3075
3076
3077
3078
3079
3080
3081
3082
3083
3084
3085
3086
3087
3088
3089
3090
3091
3092
3093
3094
3095
3096
3097
3098
3099
3100
3101
3102
3103

3104
3105
3106
3107
3108
3109
3110
3111
3112
3113
3114
3115
3116
3117
3118
3119
3120
3121
3122
3123
3124
3125
3126
3127
3128
3129
3130
3131
3132
3133
3134
3135
3136
3137
3138
3139
3140
3141
3142

付録2 抽出データの詳細内容 [1]

4-1-3. 統計情報ファイル名定義内容

検索項目	検索内容細目	ファイル名
会話文字化データ (4項目)	{ 間投詞 (日本語) 間投詞 (英語) 言いよどみ (日本語) 言いよどみ (英語)	SURF-j_inter SURF-e_inter SURF-j_ref SURF-e_ref
文節データ (5項目)	{ 出現単語 読み 標準表現 品詞 標準表現と品詞	JWRD-word JWRD-yomi JWRD-regu JWRD-parts JWRD-regu_parts
係り受けデータ (2項目)	{ 構文関係コード 意味関係コード	JREL-sy ^m JREL-sem
対訳対応データ (2項目)	{ 出現単語 (英語) 単語対応	EWRD-word EWRD-wcorr
言語単位の各階層に於ける構成要素数 (10項目)	{ 1 文節中の単語数 1 格構造中の単語数 1 格構造中の文節数 格構造中の埋め込みの深さ 1 文中の単語数 1 文中の文節数 1 文中の格構造数 1 発話中の文数 1 会話中の文数 1 会話中の発話数	COMP-word_in_phrase COMP-word_in_kaku COMP-phrase_in_kaku COMP-kemb COMP-word_in_sent COMP-phrase_in_sent COMP-kaku_in_sent COMP-sent_in_utter COMP-sent_in_conv COMP-utter_in_conv
日英対応の可能性 (5項目)	{ 単語対応 文節対応 文対応 発話対応 格構造対応	JE-word JE-phrase JE-sent JE-utter JE-kaku
日英対応の要素数 (9項目)	{ 単語対応の単語数 文節対応の単語数 格構造対応の格要素数 格構造対応の単語数 格要素対応の単語数 文対応の文数 文対応の単語数 発話対応の発話数 発話対応の文数	JECOMP-word JECOMP-word_in_phrase JECOMP-kelm JECOMP-word_in_kaku JECOMP-word_in_kelm JECOMP-sent JECOMP-word_in_sent JECOMP-utter JECOMP-sent_in_utter
言語単位の内部構造 (5項目)	{ 文節中の語の並び b i g r a m t r i g r a m 文中の語の並び 文末の文節中の語の並び	STRUCT-word_in_phrase STRUCT-bigram STRUCT-trigram STRUCT-word_in_sent STRUCT-word_in_end_phrase last

3. 抽出データ内容

☑出力例についての注意:

出力はすべて下の形式に従っている。

出力形式	
・第1フィールド	件数
・第2フィールド以下	データ(パターン)
・フィールド区切り	
*件数 データ[データ]	

3-1. 記録情報の単純な統計

以下の項目についての異なりパターン毎の延べ出現度数を求める。

3-1-1. 会話文字化データ

(1) 間投詞/日本語 (* SURF-j_inter/surf-j_inter)

内容: 日本語の間投詞。

[]ではさまれたデータを1パターンとして扱う。
[えー、あの] 郵便で送りたいと思います。

出力例: 584|えー
489|あの
121|えー、あの

(2) 間投詞/英語 (SURF-e_inter/surf-e_inter)

内容: 英語の間投詞。

[]ではさまれたデータを1パターンとして扱う。

出力例: 211|ah,
129|uh,

*印 ()内は'出力ファイル名/項目別検索プログラム名'を示す。

(3) 日本語／言い直し、言い淀み(SURF-j_ref/surf-j_ref)

内容：日本語の言い直し、言い淀み。
()ではさまれたデータを1パターンとして扱う。

出力例：6|こ
6|を

(4) 英語／言い直し、言い淀み(SURF-e_ref/surf-e_ref)

内容：英語の言い直し、言い淀み。
()ではさまれたデータを1パターンとして扱う。

出力例：15|I
15|the

3-1-2. 文節データ

(1) 出現単語(JWRD-word/jwr-d-word)

内容：日本語の出現単語。

出力例：170|一応
152|頂き

(2) 読み(JWRD-yomi/jwr-d-yomi)

内容：日本語単語の読み。

出力例：170|いちおう
152|いただき

(3) 標準表現(JWRD-regu/jwr-d-regu)

内容：日本語単語の標準表現。

出力例：229|いただく
178|いちおう

(4) 品詞(JWRD-parts/jwr-d-parts)

内容：日本語単語の品詞。

出力例：376|記号
222|普通名詞

(5) 標準表現と品詞(JWRD-regu-parts/jwr-d-regu-parts)

内容：日本語単語の標準表現とその品詞。

出力例：31|いただく|補助動詞
16|一応 |副詞

3-1-3. 係り受けデータ

(1) 構文関係コード(JREL-syn/jrel-syn)

内容：構文関係名。

出力例：4583|連用格
1445|連体修飾

(2) 意味関係コード(JREL-sem/jrel-sem)

内容：意味関係名。

出力例：1257|対象
1174|内容

3-1-4. 対訳対応データ

(1) 英語、出現単語(EWRD-word/ewrd-word)

内容：英語の出現単語。

出力例：594|to
573|you

(2) 単語対応(EWRD-wcorr/ewrd-wcorr)

内容：相対応する日英の出現単語。

出力例：33|ありがとう|Thank you
31|から |from

3-2. 言語単位の各階層間における構成要素数

次に示す言語単位の階層関係において、上位の言語単位を構成する下位の言語単位の要素数を算出し統計をとる。

単語<句(文節)<節(格構造)<文<発話<会話

(1) 1文節中の単語数 (COMP-word_in_phrase/comp-word_in_phrase)

内容：文節を構成する単語の数。

誰でも → 2単語
(一文節)

出力例：4023|3
3954|2

(2) 1格構造中の単語数 (COMP-word_in_kaku/comp-word_in_kaku)

内容：格構造を構成する単語数。

その点についてちょっとお伺い → 6単語
(一格構造)

出力例：408|3
278|4

(3) 1 格構造中の文節数 (COMP-phrase_in_kaku/comp-phrase_in_kaku)

内容：格構造を構成する文節数。

その点について ちょっと お伺い → 3 文節
(一格構造)

出力例：702|2
528|3

(4) 格構造の埋め込みの深さ (COMP-kemb/comp-kemb)

内容：埋め込まれた格構造数と埋め込まれていない格構造数。

その点について ちょっと お伺い したい んです けれど も。
┌──────────────────────────────────┐
│ (埋め込まれている) │
└──────────────────────────────────┘
┌──────────────────────────────────┐
│ (埋め込まれていない) │
└──────────────────────────────────┘

出力例：9183|not-embedded
8845|embedded

(5) 1 文中の単語数 (COMP-word_in_sent/comp-word_in_sent)

内容：文を構成する単語数。

申し訳 ございませ ん でした 。 → 6 単語
(1 文)

出力例：14|6
11|4

(6) 1文中の文節数 (COMP-phrase_in_sent/comp-phrace_in_sent)

内容：文を構成する文節数。

それは どうも、 ご丁寧な ありがとうございます。 → 4文節
(1文)

出力例：702|2
528|3

(7) 1文中の格構造数 (COMP-kaku_in_sent/comp-kaku_in_sent)

内容：文を構成する格構造数。

国際コンピューター会議の事務局の者ですけれども、
先生にはいつもお世話になっております。 → 2格構造
(一文)

出力例：20|2
13|4

(8) 1発話中の文数 (COMP-sent_in_utter/comp-sent_in_utter)

内容：発話を構成する文の数。

1発話 { はい、どうもありがとうございます。
それでは、先生からのお手紙の方、お待ちしておりますので。 → 3文
どうもありがとうございました。

出力例：312|2
117|1

(9) 1 会話中の文数 (COMP-sent_in_conv/comp-sent_in_conv)

内容：会話を構成する文の数。

出力例：2|47
1|64

(10) 1 会話中の発話数 (COMP-utter_in_conv/comp-utter_in_conv)

内容：会話を構成する発話の数。

出力例：3|12
2|17

3-3. 日英対応の可能性

日英の各データの対応について、対応しているデータ個数と対応していないデータ個数を求める。

(1) 単語対応 (JE-word/je-word)

内容：日英の単語の対応個数と非対応個数。

出力例：34582|taiounasi
5843|taiouari

(2) 文節対応 (JE-phrase/je-phrase)

内容：日英の文節の対応個数と非対応個数。

出力例：7750|taiounasi
4952|taiouari

(3) 格構造対応 (JE-kaku/je-kaku)

内容：日英の格構造の対応個数と非対応個数。

出力例：13952|taiouari
4076|taiounasi

(4) 文対応 (JE-sent/je-sent)

内容：日英の文の対応個数と非対応個数。

出力例：1102|taiouari
66|taiounasi

(5) 発話対応 (JE-utter/je-utter)

内容：日英の発話の対応個数と非対応個数。

出力例：419|taiouari
1|taiounasi

3-4. 日英対応の要素数

日英の各言語単位の対応における、構成要素数の対比について統計を求める。

例： 日本語1発話（3文）：英語1発話（2文） → 3：2
（発話対応における文の数の対比）

(1) 単語対応の単語数(JECOMP-word/jecomp-word)

内容：日本語1単語に対応する英単語数。

単語対応

ありがとう：thank you → 英語2単語

出力例：382|3
 281|1

(2) 文節対応の単語数(JECOMP-word_in_phrase/jecomp-word_in_phrase)

内容：日本語1文節に対応する英語文節内の単語数。

文節対応

国際コンピューター会議の：International Conference on Computer Science

→ 英語5単語

出力例：1660|1
 1252|2

(3) 格構造対応の英語格要素数(JECOMP-kelm/jecomp-kelm)

内容：日本語1格構造に対応する英語格構造内の格要素数。

格構造対応

レイノルズです： is Dr. Reynolds office → 格要素数3

出力例：382|3
281|1

(4) 格構造対応の英語単語数(JECOMP-word_in_kaku/jecomp-word_in_kaku)

内容：日本語1格構造に対応する英語格構造内の英単語数。

格構造対応

レイノルズです： is Dr. Reynolds office → 英語5単語。

出力例：167|4
152|7

(5) 格要素対応の英語数(JECOMP-word_in_kelm/jecomp-word_in_kelm)

内容：日本語1格要素に対応する英語格要素内の英単語数。

格要素対応

レイノルズ： Reynolds office → 英語3単語

出力例：8596|1
672|2

(6) 文対応の文数(JECOMP-sent/jecomp-sent)

内容：対応する日英の文において、日本語文の数と英語文の数。

文対応

どうもありがとうございます。 : OK Thank you very much.

→ 日本語文1 : 英語文2

出力例 : 583|1|1
209|1|2

(7) 文対応の単語数(JECOMP-word_in_sent/jecomp-word_in_sent)

内容：対応する日英の文において、日本語文内の単語数と英語文内の単語数。

文対応

どうも ありがとう ございます。 : OK Thank you very much.

→ 日本語4単語 : 英語7単語

出力例 : 2|13|25
2|28|4

(8) 発話対応の発話数(JECOMP-utter/jecomp-utter)

内容：対応する日英の発話において、日本語の発話数と英語の発話数。

発話対応

そうですか。	OK	
わかりました。	:	→ 日本語 1 発話 : 英語 1 発話
どうもありがとうございます。	Thank you.	

出力例：419|1|1

(9) 発話対応の文数(JECOMP-sent_in_utter/jecomp-sent_in_utter)

内容：対応する日英の発話において、日本語発話内の文の数と英語発話内の文の数。

発話対応

そうですか。	OK	
わかりました。	:	→ 日本語文 3 : 英語文 2
どうもありがとうございます。	Thank you.	

出力例：63|1|1
43|2|3

3-5. 言語単位の内部構造

内部構造を持つ言語単位についてその構造のパターンと延べ度数を求める。

3-5-1. 文節の構造

(1) 文節中の語の並び (STRUCT-word_in_phrase/struct-word_in_phrase)

内容：文節内における接頭語・接尾語・付属語および自立語属性の配列パターン。
各パターンの表示は次の形式で行なう。

パターン表示形式

接頭語・接尾語・付属語	→	単語の読みを表示
自立語	→	品詞を表示
例：申しますが、	→	本動詞 ますが、

出力例：188|普通名詞 が
185|本動詞 ます。

(2) b i g r a m (STRUCT-bigram/struct-bigram)

内容：単語レベルのバイグラム。

出力例：214|と いう
208|です ね

(3) t r i g r a m (STRUCT-trigram/struct-trigram)

内容：単語レベルのトライグラム。

出力例：181|と 思います
156|て おります

3-5-2. 文の構造

(1) 文中の語の並び (STRUCT-word_in_sent/struct-word_in_sent)

内容：文内における接頭語・接尾語・付属語および
自立語属性の配列パターン。
各パターンの表示形式は 3-5-1(1)文節内の語の並びと同様。

出力例：4|副詞 ありがとう ございます。
3|はい、代名詞 です が。

(2) 文末の文節中の語の並び (STRUCT-word_in_last_phrase/ struct-word_in_last_phrase)

内容：文末の文節内における接頭語・接尾語・付属語および
自立語属性の配列パターン。
各パターンの表示形式は 3-5-1(1)文節内の語の並びと同様。

出力例：185|本動詞 ます。
68|本動詞 て おり ます。

付録3 抽出プログラム例

プログラムの全体は、atr-dpの/data6/XLDB/pgmの下にある。

```
#!/usr/jsy/bin/csh -f
#####
(surf-j_inter)
```

SubSCRIPT LIBRARY OF THE INFORMATION EXTRACTION FROM THE LDB No.1-1

- 1.ITEM : 会話文字化データ/間投詞 (日本語)
- 2.CONTENTS : 間投詞。
- 3.INPUT : 会話 I D のリスト・ファイル。
- 4.OUTPUT : "件数"1"出現パターン"
- 5.USAGE : surf-j_inter [-h] [会話 I D リスト・ファイル]
-h オプションはインフォメーションの表示。
- 6.COMMENT : j_word.出現単語のうち、" ["と"] "ではさまれたデータを
取りだしています。
環境変数SQLNODECNT,SQLCONCNTはそれぞれ、1000,300に
設定しています。検索できる会話数の上限はSQLCONCNTの値と
同じになるため、検索する会話数により検索不能の場合は、
この設定を変更して下さい。
- 7.AUTHOR : Makoto Takahashi (TIS)
- 8.HISTORY : 89-11-21

```
#####
if ($#argv == 0 || $1 == "-h") then
    egrep '^#X' $0 | sed 's/#X//'; exit 0;
endif
```

```
### INITIALIZE ###
```

```
onintr clean
setenv SQLNODECNT 1000 # SQL記述子ノード設定 #
setenv SQLCONCNT 300 # SQL定数設定 #
set tmp1 = /tmp/xldb-jinter1.$$
set tmp2 = /tmp/xldb-jinter2.$$
set id_set = `cat $1 | tr -s "Y012" ",,"` # 会話 I D リスト #
```

```
### END OF INIT ###
```

```
### PROCEDURE 1 #####
# 処理内容: j_wordより出現単語を検索出力する(単語 I D の昇順)。 #
# 入力形式: 会話 I D, 会話 I D, 会話 I D, . . . #
# 出力形式: 出現単語 #
```

```
#####
SQL <<_EOS_>& /dev/null
    lines      0
    select     出現単語
    from       j_word
    where      会話 I D in <${id_set}>
    order by   会話 I D, 単語 I D
    into       '${tmp1}'/
```

```
_EOS_
### END OF PROCEDURE 1 ###
```

```
## PROCEDURE 2 #####
# 処理内容：全ての出現単語の中から"["と"]"ではさまれたデータ(間投詞)を #
#   取りだす。 #
# 入力形式：出現単語 #
# 出力形式：間投詞 #
# 注意："["と"]"の間の語は1レコードで出力する。 #
#   1つの単語の中に間投詞が含まれたデータ(ex.)についての処理を含む。 #
#   ex.堺、[いー]俊幸先生 #
#####
# 間投詞が1単語の中に含まれる場合 #Y
grep '[.]' $tmp1 |
sed 's/^.*[Y(.*)] .*$/Y1/' > $tmp2
grep -v '[.]' $tmp1 |Y
sed -n '/ [/,/] /p' |Y
sed 's/ [/[/' | sed 's/] /]/' | # awkが全角文字を処理しないため #Y
awk '{Y
    if ($0 != "[" && $0 != "]")Y
        printf ("%s+", $0);Y
    else if ($0 == "]")Y
        printf ("\n");Y
}' |Y
sed 's/Y(.*)+$/Y1/' >> $tmp2
### END OF PROCEDURE 2 ###
## PROCEDURE 3 #####
# 処理内容：間投詞データの統計をとる。 #
# 入力形式：間投詞データ #
# 出力形式：件数、間投詞データ #
#####
sort $tmp2 | uniq -c | sort -nr |Y
awk '{ printf ("%8d%s\n", $1, $2); }' |Y
sed 's/+/ /g'
### END OF PROCEDURE 3 ###
clean:
    onintr -
    set nonomatch
    rm -f $tmp1 $tmp2
```

付録4 抽出結果の一部

抽出結果の全体は、atr-dpの/data6/XLDB/data/keyの下および/data6/XLDB/data/telの下にある。

付録4. 1 キーボード会話 (k e y)

SURF-j_inter

SURF-e_inter

2|uh ,

SURF-j_ref

2|社

1|\$ 800 ですね

1|10 / 27

1|10 / 30

1|411

1|604

1|672

1|エコノミークラス

1|シングル

1|ホテル代も含む

1|一般公募はありません

1|最終レイアウトの終わった原稿

1|住所、名前

1|日本時間

SURF-e_ref

26|double

10|yen

2|ICCS

1|20,000 for a copy

1|411

1|604

1|C.O.D.

1|DM to yen

1|Japan time

1|economy class

1|estimated guess

1|facilities such as the meeting place , printing of the related literature and managemen

1|handling fee will be deducted

1|probably within a month .

1|the cost of the proceedings will be deducted

1|the welcome reception will be held in the evening of this day

JWRD-word

2442|。

2236|、

1893|の

1212|ます

1158|て

1127|に

917|が

913|は

818|を

761|お

729|です

678|し

678|と

625|か

603|で

561|た

371|まし

355|う

312|でしょ

309|会議

275|ん

272|おり

246|こと

JWRD-yomi

2442|。

2236|、

1893|の

1212|ます

1169|に

1160|て
 917|が
 913|は
 818|を
 762|お
 729|です
 694|し
 678|と
 673|か
 605|で
 568|た
 371|まし
 355|う
 316|ご
 312|でしよ
 309|かいぎ
 292|こと
 275|おり

JWRD-regu

2442|。
 2236|、
 1893|の
 1794|ます
 1158|て
 1103|です
 1045|御
 1020|に
 917|が
 913|は
 818|を
 799|する
 678|と
 625|か
 564|た
 449|で
 425|だ
 355|う
 309|会議
 292|事
 275|ん
 274|おる
 243|私

JWRD-regu_parts

2441|
 2234|
 1794|
 1535|
 1158|
 1103|
 1045|
 1019|
 913|
 818|
 797|
 573|
 564|
 531|
 492|
 425|
 424|
 421|
 358|
 354|
 309|
 292|
 274|

。|記号
 、|記号
 ます|助動詞
 の|格助詞
 て|接続助詞
 です|助動詞
 御|接頭語
 に|格助詞
 は|係助詞
 を|格助詞
 する|補助動詞
 と|格助詞
 た|助動詞
 か|終助詞
 が|格助詞
 だ|助動詞
 が|接続助詞
 で|格助詞
 の|準体助詞
 う|助動詞
 会議|普通名詞
 事|普通名詞
 おる|補助動詞

JWRD-parts

6502|普通名詞
 5270|助動詞
 5252|格助詞
 4855|記号
 3062|本動詞
 2436|補助動詞
 2204|接続助詞
 1154|接頭語
 1114|係助詞
 946|接尾語
 933|副詞
 793|代名詞
 732|固有名詞
 713|サ変名詞
 616|終助詞
 453|感動詞
 439|数詞
 417|準体助詞
 390|接続詞
 338|形容詞
 319|形容名詞
 288|副助詞
 280|連体詞
 90|並立助詞
 9|その他
 1|間投詞

JREL-syn

6633|連用格
 2074|連体修飾
 1098|文連用修飾
 775|連体格
 684|述語連用修飾
 247|並列

JREL-sem

2119|対象
 1384|内容
 729|条件
 537|程度1
 532|陳述
 488|時点
 460|対象/話題提示
 439|範囲規定、関係
 435|意味並列
 372|限定
 363|場所
 339|原因、理由
 304|動作主
 299|外格
 239|補文標識
 187|状態変化の終点
 174|対象属性値
 168|方式、様態
 139|受け手
 135|準補文標識
 125|関係の基準
 124|目的
 106|所有者
 92|相手1
 90|与え手
 82|価値判断
 73|その他
 67|譲歩
 67|道具、手段
 64|命名
 63|役割

58|経験者
 55|場所、終点
 55|属性対象
 54|範囲規定、関係/話題提示
 51|同格
 44|連接
 39|動作主/話題提示
 37|全体
 32|選択
 30|選言
 26|場所、起点
 23|条件/話題提示
 22|時点/話題提示
 20|作成者
 19|無意思主体
 18|時、継続
 17|例示
 12|属性値
 12|付帯状況
 11|時、起点
 10|連接/話題提示
 8|相手2
 8|比較
 7|経験者/話題提示
 6|程度1/話題提示
 5|時、終点
 5|状態変化の始点
 5|無意思主体/話題提示
 5|無意思動作主
 4|属性
 3|場所、通過
 3|部分
 3|目的/話題提示
 2|原因、理由/話題提示
 1|価値判断/話題提示
 1|材料
 1|時、継続/話題提示
 1|場所、通過/話題提示
 1|全体/話題提示
 1|役割/話題提示
 1|与え手/話題提示

EWRD-word

2263|.

1678|the

1343|.

1078|to

995|you

824|I

692|of

479|for

455|is

445|?

428|and

416|a

376|will

373|in

368|your

354|be

312|have

300|we

289|that

272|conference

256|it

242|on

216|this

EWRD-wcorr

202	の	of
165	会議	conference
103	です	is
86	ありがとう	Thank you
74	私	I
61	登録	registration
59	に	in
58	会議	Conference
58	に	to
53	が	but
50	はい	Yes
49	送り	send
48	と	and
47	で	in
44	から	from
44	で	at
42	申し	is
38	こちら	This
38	たい	like
37	に	on
36	料	fee
35	ので	so
31	大阪	Osaka
31	円	yen
30	の	in
29	の	for
28	ん	not
27	銀行	bank
27	で	by
26	こちら	we
25	時間	time
25	先生	you
25	て	and
24	この	this
23	そちら	you
23	名前	name
23	ば	if
22	たら	If
22	ない	not
22	番号	number
22	と	that
22	に	at
21	こちら	this
21	どうも	very much
21	あり	have
21	で	is
20	もしもし	Hello
20	講演	speech
20	参加	attend
20	3	three
19	について	about
19	ホテル	hotel
19	部屋	room
19	ば	If
19	ん	don't
18	広告	ad
17	それ	that
17	日本	Japan
17	問題	problem
17	と	with
17	私	My name
17	私	This
17	日	days
16	プログラム	program
16	申し込み	application
16	支払い	pay

さん|Mr.
たら|if
住所|address
電話|call

16|
16|
16|
16|
COMP-word_in_phrase
5483|2
2828|3
1922|1
1686|4
965|5
512|6
252|7
149|8
41|9
21|10
4|11
1|14

COMP-word_in_kaku
553|4
498|3
485|5
461|6
422|7
370|8
339|9
290|10
259|11
199|12
167|2
156|13
152|14
114|15
111|16
85|17
75|18
63|19
60|21
56|20
41|24
37|23
34|22
24|26
18|25
16|27
13|28
12|30
11|29
9|31
9|33
7|38
6|32
5|36
4|34
4|35
4|42
3|40
2|37
2|44
2|49
1|43
1|45
1|47
1|48
1|65

COMP-phrase_in_kaku
1168|2

1105|3
793|4
593|5
402|6
358|1
243|7
190|8
115|9
66|10
48|11
43|12
21|13
15|15
11|14
4|18
3|17
2|22
1|16
1|19
1|21

COMP-word_in_sent

174|4
136|7
128|6
119|5
111|10
110|12
108|8
96|9
93|14
90|11
88|16
85|13
83|15
76|17
75|18
73|2
65|23
64|19
57|25
55|21
55|3
44|26
43|20
43|24
42|28
39|22
39|29
34|30
33|27
27|32
23|31
21|34
20|33
17|35
15|38
12|37
9|36
9|40
9|42
8|47
7|41
7|43
7|46
6|39
6|49

5|44
5|45
5|53
4|52
3|58
2|50
2|51
2|54
2|55
2|59
2|60
2|62
2|65
2|67
1|48
1|57
1|61
1|63
1|74
1|76
1|78

COMP-phrase_in_sent

356|2
335|3
286|1
270|4
236|5
214|6
164|7
161|8
116|9
85|10
67|11
53|12
41|13
35|14
24|15
13|18
12|16
12|17
8|19
5|20
5|24
3|22
3|23
2|21
2|26

COMP-kaku_in_sent

744|1
502|2
339|3
215|4
111|5
69|6
41|7
17|8
9|9
5|10
2|11
1|13

COMP-kemb

10895|not-embeded
7374|embeded

COMP-sent_in_utter

273|2
240|3

157|1
 134|4
 63|5
 29|6
 6|7
 2|9
 COMP-sent_in_conv
 10|21
 8|23
 6|29
 6|34
 5|22
 5|31
 4|25
 4|28
 3|12
 3|19
 3|24
 3|26
 3|30
 3|32
 3|35
 2|13
 2|15
 2|16
 2|17
 2|18
 2|20
 2|27
 2|39
 2|40
 1|11
 1|33
 1|36
 1|41
 1|42
 1|44
 1|51
 1|57
 COMP-utter_in_conv
 20|8
 17|10
 9|11
 9|7
 9|9
 8|12
 7|6
 6|13
 3|14
 3|4
 2|17
 1|22
 1|5
 JE-word
 39606|taiounasi
 10627|taiouari
 JE-phrase
 9347|taiouari
 4517|taiounasi
 JE-kaku
 10589|taiounasi
 7680|taiouari
 JE-sent
 2477|taiouari
 31|taiounasi
 JE-utter

904|taiouari
0|taionasi
JECOMP-word
9242|1
900|2
286|3
83|4
48|5
18|8
18|9
17|6
11|7
1|10
1|11
JECOMP-word_in_phrase
2987|1
2459|2
1813|3
939|4
505|5
299|6
159|7
64|8
38|10
38|9
14|11
3|14
2|12
1|17
JECOMP-word_in_kaku
346|7
335|5
334|6
319|8
311|4
283|9
231|10
220|3
203|11
165|12
146|2
144|13
114|15
106|14
84|16
62|1
50|17
48|19
42|18
35|20
31|21
27|22
18|23
15|25
15|26
13|24
9|31
8|28
7|27
7|30
5|37
4|32
3|33
2|36
1|29
1|38

1|39
1|42
1|45
1|47
1|48
1|49
1|51
1|52

JECOMP-kelm

777|3
699|4
503|5
463|2
368|6
291|1
253|7
147|8
66|9
53|10
32|11
20|12
19|13
15|14
13|15
8|16
5|20
3|17
3|18
3|19
3|24
2|21
2|28
1|26
1|29
1|31
1|42

JECOMP-word_in_kelm

12759|1
1285|2
924|3
460|4
373|5
276|6
207|7
173|8
148|9
101|10
89|11
50|12
44|13
35|14
25|15
25|16
18|17
16|18
12|21
9|23
6|22
6|24
5|19
5|20
4|26
2|25
2|27
2|31
2|35

1|28
1|29
1|33
1|34
1|36
1|45
1|49

JECOMP-sent

1754| 1|1
270| 1|2
150| 2|1
43| 1|3
30| 2|2
8| 2|3
4| 3|1
3| 1|4
2| 3|2
2| 4|1
1| 2|4
1| 3|3

JECOMP-word_in_sent

25| 4|3
23| 2|2
20| 4|5
20| 7|7
19| 5|5
18| 12|10
17| 7|6
16| 4|2
16| 6|5
14| 10|8
14| 3|3
14| 8|5
13| 4|6
12| 6|7
12| 7|5
12| 9|10
11| 10|6
11| 9|8
10| 11|7
10| 13|10
10| 14|8
10| 15|10
10| 15|14
10| 16|12
10| 17|14
10| 5|4
10| 8|8
9| 11|11
9| 11|8
9| 3|5
9| 4|4
9| 5|8
9| 6|3
9| 6|6
9| 6|8
9| 7|9
9| 8|4
9| 9|12
9| 9|6
9| 9|7
8| 10|11
8| 11|6
8| 12|12
8| 12|7
8| 14|13

8	17 17
8	17 20
8	18 11
8	22 18
8	5 2
8	5 3
8	5 7
8	6 2
8	8 7
8	9 9
7	10 12
7	10 7
7	13 11
7	13 14
7	13 15
7	14 11
7	16 11
7	16 15
7	17 10
7	19 15
7	5 6
7	7 4
7	7 8
7	8 10
7	8 12
7	8 6
6	10 13
6	11 14
6	11 5
6	12 11
6	12 8
6	13 12
6	13 13
6	13 16
6	14 10
6	15 12
6	17 16
6	20 17
6	22 19
6	29 23
6	3 2
6	3 4
6	4 1
6	5 13
6	6 12
6	8 9

JECOMP-utter

904	1 1
-----	-----

JECOMP-sent_in_utter

157	2 2
121	3 3
71	1 1
67	2 3
64	4 4
59	1 2
57	3 4
41	4 3
32	3 2
29	5 5
25	2 1
21	1 3
20	2 4
20	3 5
19	4 5
13	5 4
8	5 6

8	6 6
7	6 5
6	4 6
6	5 3
5	3 1
4	5 7
4	6 3
4	6 7
3	1 4
3	3 6
3	4 1
2	1 5
2	2 5
2	5 2
2	5 8
2	6 8
2	7 4
2	7 6
1	1 6
1	3 7
1	4 2
1	4 7
1	5 9
1	6 1
1	6 2
1	6 4
1	7 3
1	7 7
1	8 5
1	9 12
1	9 6

STRUCT-word_in_phrase

690|普通名詞 の
652|副詞
496|普通名詞 に
489|普通名詞 を
375|普通名詞
318|本動詞 ます。
314|普通名詞 が
282|普通名詞 は
245|本動詞
171|普通名詞、
168|普通名詞 で
162|固有名詞 の
130|普通名詞 普通名詞 の
128|はい、
127|本動詞 ました。
124|形容詞
120|代名詞 の
116|その
107|普通名詞 と
104|副詞、
102|普通名詞 普通名詞 を
96|本動詞 て おります。
93|ありがとう ございました。
90|形容名詞 に
83|この
83|代名詞
70|固有名詞 と
70|代名詞 は
68|普通名詞 は、
68|本動詞 た
67|固有名詞 です。
67|普通名詞 普通名詞 に
59|普通名詞 も
59|本動詞 て

58|ありがとう ございます。
58|サ変名詞 する
57|お 本動詞 いたします。
57|本動詞 ません。
56|普通名詞 です。
54|形容名詞 です。
54|形容名詞 な
54|代名詞 で
53|代名詞 か
52|では、
52|代名詞 が
51|お 本動詞 します。
50|固有名詞、
50|普通名詞 普通名詞 は

STRUCT-bigram

814|ます。
362|か。
318|が、
315|して
312|でしょ う
300|た。
297|う か
287|です。
264|て おり
256|ですが
235|ました
229|おります
217|は、
178|ません
176|ので、
172|のです
171|思います
168|と 思い
149|をお
128|はい、
123|します
114|の方
113|ますので
110|てい
103|ますが
103|会議の
102|が。
101|ている
101|まして
101|ん。
99|ございます。
98|お願い
97|ございました
97|という
91|、私
91|したい
89|お 送り
86|、お
86|て、
86|事務局
83|ありがとう ございまし
83|せて
81|では

STRUCT-trigram

283|でしょ う か
247|う か。
223|て おります
184|ました。
165|ですが、
160|と 思います
159|おります。

140| 思います。
 131| のですが
 99| ますので、
 97| しており
 95| ございました。
 95| します。
 86| ません。
 83| ありがとう ございました
 80| ですが。
 78| わかりました
 77| ますが、
 73| ています
 72| 致します。
 66| させて
 61| たいのです
 58| ありません
 56| のでしょう
 55| たいと思い
 55| ますでしょう
 54| います。
 52| お願いし
 51| ありがとう ございます。
 49| ますか。
 46| ですか。
 45| ということ
 45| まして、
 44| ですね。
 43| になって
 41| と申します
 39| したいの
 39| んでしょう
 38| したいと
 38| 結構です。
 36| うか？
 35| 申します。
 34| いたします。
 34| ているの
 34| なっており
 33| にして
 32| 願いします
 31| せて頂き
 31| んが、
 30| 事務局の

STRUCT-word_in_sent

65| 本動詞 ました。
 26| ありがとう ございました。
 26| さようなら。
 17| 副詞 ありがとう ございました。
 15| こんにちは。
 14| はい、本動詞 ました。
 13| ありがとう ございます。
 10| 副詞 ですか。
 9| おはよう ございます。
 9| サ変名詞 いたします。
 8| はい、代名詞 固有名詞 です。
 8| 代名詞 で 形容名詞 です。
 7| 副詞 です。
 6| こんにちは。
 6| ご 普通名詞 ありがとう ございました。
 6| では、さようなら。
 6| はい。
 6| 形容名詞 です。
 6| 代名詞、固有名詞 と 本動詞 ます。
 6| 副詞 本動詞 ます。
 5| では、サ変名詞 いたします。
 5| サ変名詞 します。

- 5|固有名詞 です。
- 5|代名詞 か ご 普通名詞 でしょう か。
- 5|代名詞 は、固有名詞 です。
- 5|代名詞 固有名詞 です。
- 5|副詞。
- 5|副詞 ありがとう ございます。
- 4|お 普通名詞 副詞 ありがとう ございました。
- 4|それでは、サ変名詞 いたします。
- 4|どういたしまして。
- 4|はい、代名詞 で 形容名詞 です。
- 4|はい、代名詞 固有名詞 で ございます。
- 4|サ変名詞 しました。
- 4|代名詞 は 形容名詞 です ね。
- 4|代名詞 は 固有名詞 です。
- 4|代名詞 は 固有名詞 と 本動詞 ます。
- 4|副詞 お 本動詞 いたします。
- 4|副詞 お 本動詞 します。
- 4|副詞 副詞 お 本動詞 いたします。
- 3|お 普通名詞 ありがとう ございました。
- 3|お 普通名詞 お 本動詞 しておりました。
- 3|ご 普通名詞 ありがとう ございます。
- 3|ご 普通名詞 は。
- 3|では、サ変名詞 します。
- 3|どのような ご 普通名詞 でしょう か。
- 3|固有名詞 さまをお 本動詞 します。
- 3|代名詞 か。
- 3|代名詞 で 形容詞 でしょう か。
- 3|代名詞 は 形容詞 た。
- 3|普通名詞 本動詞 ん。
- 3|副詞 です ね。
- 3|副詞 本動詞 ました。

STRUCT-word_in_last_phrase

- 318|本動詞 ます。
- 127|本動詞 ました。
- 96|本動詞 て おります。
- 93|ありがとう ございました。
- 67|固有名詞 です。
- 58|ありがとう ございます。
- 57|お 本動詞 いたします。
- 57|本動詞 ません。
- 56|普通名詞 です。
- 54|形容名詞 です。
- 51|お 本動詞 します。
- 36|さようなら。
- 34|お 本動詞 ください。
- 33|サ変名詞 いたします。
- 32|本動詞 ています。
- 28|サ変名詞 します。
- 27|本動詞 てください。
- 24|本動詞 ます でしょう か。
- 22|お 本動詞 しております。
- 22|形容詞 でしょう か。
- 16|こんにちは。
- 15|お 本動詞 したい の です が。
- 15|ご サ変名詞 ください。
- 15|副詞。
- 15|副詞 です か。
- 15|本動詞 ん。
- 13|ご 普通名詞 でしょう か。
- 13|サ変名詞 しております。
- 13|本動詞 させていただきます。
- 12|サ変名詞 しました。
- 12|数詞 です。
- 12|普通名詞 でしょう か。
- 12|普通名詞 です ね。
- 12|本動詞 ます が。

- 12|本動詞 ませ んでしょ う か 。
- 11|お 本動詞 し ましょ う 。
- 11|サ変名詞 さ せて いた だき ます 。
- 11|副詞 でしょ う か 。
- 11|本動詞 ます か 。
- 10|形容詞 の でしょ う か 。
- 10|形容詞 の です が 。
- 10|固有名詞 で ござ い ます 。
- 10|普通名詞 普通名詞 です 。
- 10|本動詞 れ ます 。
- 9|おはよう ござ い ます 。
- 9|形容詞 です 。
- 9|副詞 です 。
- 9|本動詞 の でしょ う か 。
- 8|ありがとう 。
- 8|固有名詞 です が 。
- 8|本動詞 て いた だけ ます か 。
- 8|本動詞 て おり ませ ん 。
- 7|こんにちわ 。
- 7|ご サ変名詞 いた し ます 。
- 7|サ変名詞 いた し ました 。
- 7|サ変名詞 し て くだ さ い 。
- 7|形容名詞 でしょ う か 。
- 7|形容名詞 で ござ い ます 。
- 7|副詞 でしょ う 。
- 7|本動詞 。
- 7|本動詞 ん で し た 。
- 6|お 本動詞 いた だけ ます か 。
- 6|はい 。
- 6|サ変名詞 し て い ます 。
- 6|代名詞 でしょ う か 。
- 6|本動詞 て い ませ ん 。
- 6|本動詞 の です が 。
- 6|本動詞 ま しょ う 。
- 6|本動詞 ん でしょ う か 。

付録 4. 2 電話会話 (t e l)

SURF-j_inter

1183|えー
 576|あの
 413|え
 376|あの一
 116|おー
 92|あー
 85|あ
 60|ま
 49|うー
 43|いー
 43|えーと
 42|その
 36|お
 28|えーっと
 28|えっと
 22|あっ
 16|い
 14|え、あの
 14|まあ
 12|う
 12|ん
 11|んー
 7|その、おー
 7|ま、あの
 7|ま、あの一
 6|あの一、えー
 5|あの一、え
 5|えー、ま
 5|えっ
 5|その一
 5|まあ
 4|あ、あの一
 4|えー、あの
 4|で
 4|と
 4|ま、
 3|あーん
 3|あの一、うー
 3|あの、えー
 3|えー、あっ
 3|えー、え
 3|え、あの一
 3|で、あの一
 2|ー
 2|あ、えー
 2|うーん
 2|うん
 2|えーん
 2|えー、あ
 2|えー、で
 2|えと
 2|じゃ、あの
 2|っと
 2|つ
 2|まあ、あの一
 2|ん、えー

SURF-e_inter

640|ah ,
 503|uh ,
 163|Uh ,
 155|Ah ,
 98|um ,
 82|Um ,
 32|Oh ,
 14|ah , ah ,

12|er ,
11|ah , da ,
6|eh ,
4|ah , de ,
4|uum ,
3|Eh ,
3|Er ,
2|Mm ,
2|ah ... ah ,
2|ei ,
2|oh ,
2|uh , uh ,

SURF-j_ref

82|
8|を
6|お
6|こ
5|か
4|い
4|す
4|て
4|に
4|の
4|コンピューター
3|で
2|2
2|あ
2|いい
2|が
2|く
2|こう いった
2|ご
2|さ
2|し
2|せん
2|その
2|た
2|っ
2|なん
2|も
2|り
2|イン
2|ホ
2|何
2|会
2|開

SURF-e_ref

35|the
21|I
18|double
9|to
7|a
7|if
7|is
5|I'm
5|you
4|Mr.
4|in
4|my
4|we
3|078
3|That's
3|about
3|and
3|as
3|con

3|for
3|m
3|of
3|re
3|to , to
3|would
2|03
2|06
2|415
2|I just
2|I'll
2|I've
2|The
2|You
2|a , a
2|all the
2|an
2|co
2|computer
2|for the
2|how
2|it
2|on
2|our
2|perhaps
2|please
2|se
2|sh
2|so
2|ss
2|that
2|the a
2|this
2|we are
2|we have
2|wh
2|wi
2|you are
2|your

JWRD-word

3471| [
3471|]
3172|、
2018|の
1856|。
1481|て
1238|です
1236|えー
1211|に
1181|と
1147|ます
904|で
779|が
736|は
730|を
688|し
625|あの
622|ん
573|お
553|か
552|た
550|方
492|まし

JWRD-yomi

3471| [
3471|]

3172|、
2018|の
1856|。
1494|て
1258|に
1238|です
1236|えー
1181|と
1147|ます
915|で
779|が
736|は
730|を
702|し
625|あの
622|ん
586|か
576|お
559|た
492|まし
484|ね

JWRD-regu

3471| [
3471|]
3172|、
2018|の
1856|。
1743|ます
1481|て
1463|です
1236|えー
1181|と
1065|に
779|が
736|は
730|を
701|御
688|で
624|あの
624|する
622|ん
598|だ
555|言う
553|か
552|た
551|方
483|ね
452|思う
445|え
423|あの一
412|(
412|)
376|其
362|事
357|いただく
320|ので
307|私
295|たい
278|おる
268|から
237|けれども
237|為る
221|う
210|ございます
205|も
204|此方

193|此
 187|一応
 184|ない
 JWRD-regu_parts
 3471|
 3470|
 3172|
 1856|
 1837|
 1743|
 1478|
 1462|
 1236|
 1063|
 1048|
 736|
 729|
 701|
 624|
 622|
 598|
 555|
 550|
 537|
 490|
 487|
 483|
 470|
 452|
 449|
 445|
 423|
 412|
 412|
 362|
 356|
 320|
 308|
 307|
 295|
 277|
 237|
 236|
 226|
 208|
 205|
 204|
 202|
 202|
 187|
 183|

]|記号
 [|記号
 、|記号
 。|記号
 の|格助詞
 ます|助動詞
 て|接続助詞
 です|助動詞
 え|間投詞
 に|格助詞
 と|格助詞
 は|係助詞
 を|格助詞
 御|接頭語
 する|補助動詞
 あ|間投詞
 だ|助動詞
 言|本動詞
 た|助動詞
 方|普通名詞
 で|格助詞
 ん|準体助詞
 ね|終助詞
 が|格助詞
 思|本動詞
 か|終助詞
 え|間投詞
 あ|間投詞
 (|記号
)|記号
 事|普通名詞
 いただく|補助動詞
 の|接続助詞
 が|接続助詞
 私|代名詞
 たい|助動詞
 お|補助動詞
 けれども|接続助詞
 為|本動詞
 其|連体詞
 ございます|補助動詞
 う|助動詞
 此方|代名詞
 から|格助詞
 も|係助詞
 一応|副詞
 ない|助動詞

JWRD-parts
 12839|記号
 6507|普通名詞
 6069|格助詞
 5569|助動詞
 3748|本動詞
 3625|間投詞
 2984|接続助詞
 2230|補助動詞
 1541|副詞
 1035|接尾語
 1009|代名詞
 948|終助詞
 938|係助詞
 778|接頭語

692|固有名詞
665|準体助詞
657|接続詞
550|サ変名詞
491|数詞
413|感動詞
401|連体詞
351|副助詞
316|形容詞
294|形容名詞
141|並立助詞
140|その他

JREL-syn

6773|連用格
2264|連体修飾
1159|文連用修飾
909|連体格
859|述語連用修飾
594|並列

JREL-sem

1824|対象
1763|内容
784|条件
669|程度1
588|意味並列
540|限定
526|準補文標識
489|時点
474|陳述
472|範囲規定、関係
395|外格
383|動作主
377|場所
305|補文標識
297|対象/話題提示
240|原因、理由
219|対象属性値
170|状態変化の終点
148|受け手
132|方式、様態
130|目的
111|関係の基準
105|価値判断
103|同格
94|与え手
92|全体
69|場所、終点
69|命名
64|その他
64|相手1
64|連接
63|経験者
63|作成者
63|所有者
63|道具、手段
62|属性対象
47|役割
42|譲歩
39|範囲規定、関係/話題提示
38|選言
31|場所、起点
29|時、継続
28|付帯状況
25|時点/話題提示
24|例示
22|時、起点

22|動作主/話題提示
 22|無意思主体
 20|選択
 20|連接/話題提示
 14|時、終点
 10|条件/話題提示
 9|比較
 8|属性値
 5|経験者/話題提示
 4|属性
 2|原因、理由/話題提示
 2|場所/話題提示
 2|随伴
 2|相手1/話題提示
 2|相手2
 2|程度1/話題提示
 2|無意思動作主
 1|時、起点/話題提示
 1|時、継続/話題提示
 1|場所、終点/話題提示
 1|場所、通過
 1|状態変化の終点/話題提示
 1|比較/話題提示
 1|部分
 1|無意思主体/話題提示
 1|目的/話題提示
 1|役割/話題提示
 1|与え手/話題提示

EWRD-word

3384|,
 1936|.
 1779|the
 1774|[
 1774|]
 996|to
 867|you
 708|i
 698|ah
 611|of
 514|uh
 510|and
 489|(
 489|)
 479|a
 474|we
 438|that
 430|for
 414|be
 383|is
 331|in
 321|have
 319|will
 307|your
 284|>
 283|<
 266|as
 261|on
 246|?
 240|like
 238|would
 228|it
 225|And
 223|this
 222|i'm
 193|are
 187|conference

172|very
 166|can
 166|if
 164|Uh
 158|Ah
 145|at
 139|from
 129|with
 128|so
 124|Conference

EWRD-wcorr

120| の|of
 76| 会議|conference
 60| たい|like
 60| に|to
 54| さん|Mr.
 53| から|from
 52| です|is
 52| 私|I
 51| はい|Yes
 46| ありがとう|Thank you
 42| 先生|you
 41| に|in
 40| 申し|is
 39| と|and
 38| どうも|very much
 34| それから|and
 34| こちら|we
 34| で|in
 34| 私|we
 33| 送り|send
 33| に|on
 30| 国際コンピューター会議|the International Conference on Computer Scie

ce

30| ので|so
 30| で|And
 29| 岡本|Okamoto
 29| 番号|number
 29| の|for
 28| ファックス|fax
 28| ば|if
 25| あるいは|or
 25| 大阪|Osaka
 24| 午後|afternoon
 24| 思い|think
 24| の|in
 24| 日|days
 23| わかり|understand
 23| で|at
 22| たら|if
 22| 企画|Planning
 21| そちら|you
 21| 時間|time
 21| 住所|address
 21| 2|two
 21| と|that
 20| 名前|name
 20| で|by

COMP-word_in_phrase

6517|3
 6115|2
 3392|1
 1658|4
 1146|5
 572|6
 311|7

150|8
44|9
12|10
3|11
2|12
1|14
COMP-word_in_kaku
570|3
398|4
345|6
337|5
283|8
259|7
242|9
217|10
204|11
187|12
149|13
147|14
132|15
131|2
118|17
107|16
97|18
85|19
80|22
77|21
74|20
65|23
63|25
61|26
60|24
47|27
43|29
43|30
39|28
35|31
35|32
23|33
20|35
20|37
19|34
19|40
17|39
16|36
15|41
13|43
13|47
12|46
11|38
11|48
10|42
9|57
7|44
7|45
6|52
6|55
6|59
6|63
5|49
5|51
5|53
5|61
4|54
4|56
4|62

4|68
4|71
3|50
3|58
3|66
3|70
2|116
2|69
2|82
2|84
2|88
2|91
1|112
1|113
1|114
1|133
1|60
1|64
1|65
1|67
1|73
1|76
1|80
1|94

COMP-phrase_in_kaku

999|2
763|3
645|4
544|5
389|6
311|7
246|8
212|9
160|10
151|11
150|11
120|12
68|13
62|14
37|16
30|17
29|15
27|18
25|19
18|21
16|20
10|22
10|24
9|25
9|26
7|23
3|27
3|28
3|30
2|29
2|32
2|33
2|34
2|49
2|50
1|35
1|36
1|37
1|38
1|51

COMP-word_in_sent

88|6
73|4
55|7
47|11
46|9
45|10
45|17
45|8
42|12
42|15
41|5
40|14
39|13
39|22
38|16
38|2
36|18
35|24
34|19
32|23
32|29
29|28
29|35
29|37
29|41
28|27
28|31
28|33
27|20
27|34
26|26
25|25
23|3
23|30
23|42
22|21
22|32
21|36
20|38
20|40
20|43
19|49
18|44
18|53
17|47
16|45
16|46
15|39
14|48
14|50
14|51
13|52
13|70
10|60
9|54
9|62
9|67
8|59
8|68
8|97
7|56
7|57
7|61
7|65
7|66
7|71

7|74
6|78
6|81
5|55
5|63
5|75
5|77
4|58
4|64
4|72
4|73
4|76
3|100
3|105
3|109
3|69
3|80
3|83
3|84
3|87
3|91
3|96
2|101
2|111
2|112
2|114
2|122
2|127
2|82
2|85
2|89
2|90
2|95
2|98
2|99
1|102
1|103
1|106
1|110
1|117
1|118
1|120
1|121
1|126
1|132
1|134
1|147
1|148
1|153
1|160
1|172
1|173
1|79
1|88
1|92
1|94

COMP-phrase_in_sent

196|2
130|3
115|6
114|1
108|4
106|5
90|8
89|7
81|9

76|12
76|13
72|11
63|10
60|15
55|16
52|14
41|17
37|18
31|20
28|19
24|22
24|24
23|21
17|23
16|25
15|28
14|29
12|27
11|26
8|30
8|35
7|33
6|34
6|39
5|31
5|32
4|36
4|40
4|45
4|49
3|38
3|41
2|43
2|53
1|44
1|47
1|48
1|50
1|51
1|52
1|56
1|57
1|61
1|65

COMP-kaku_in_sent

315|1
275|2
233|3
171|4
128|5
93|6
67|7
36|8
33|9
13|10
11|11
9|12
4|14
3|13
3|15
2|16
1|17
1|19

COMP-kemb

14007|not-embedded

12909|embedded
COMP-sent_in_utter
179|2
152|1
102|3
81|4
42|5
28|6
16|8
11|7
5|9
4|11
2|10
1|12
1|13

COMP-sent_in_conv
3|24
3|29
3|32
3|37
3|49
2|23
2|28
2|30
2|34
2|38
2|43
2|47
2|48
2|55
1|19
1|26
1|27
1|33
1|40
1|41
1|42
1|44
1|45
1|50
1|62
1|63
1|64
1|96

COMP-utter_in_conv
10|11
9|12
6|14
3|8
2|10
2|13
2|15
2|16
2|17
2|20
2|9
1|21
1|22
1|23
1|26
1|6

JE-word
54931|taiounasi
8995|taiouari
JE-phrase
12280|taiounasi

7643|taiouari
 JE-kaku
 20881|talounasi
 6035|taiouari
 JE-sent
 1755|taiouari
 102|talounasi
 JE-utter
 621|taiouari
 3|talounasi
 JECOMP-word
 7579|1
 803|2
 250|4
 228|3
 58|6
 46|5
 12|7
 6|8
 6|9
 5|10
 JECOMP-word_in_phrase
 2710|1
 1919|2
 1307|3
 711|4
 429|5
 285|6
 147|7
 55|8
 30|9
 20|10
 7|11
 4|12
 2|13
 2|14
 1|16
 1|21
 JECOMP-word_in_kaku
 243|4
 228|5
 227|6
 218|7
 209|3
 202|2
 191|8
 185|9
 181|1
 160|10
 133|12
 132|11
 124|13
 95|14
 87|15
 74|16
 64|18
 59|17
 54|19
 53|20
 50|23
 45|22
 38|21
 29|24
 27|26
 26|25
 20|28

20|31
17|35
16|27
14|29
12|32
11|30
8|34
6|37
6|40
6|41
6|43
5|33
5|39
5|45
5|46
4|42
4|52
4|63
3|36
3|38
3|44
3|53
2|48
2|50
2|59
2|65
2|68
2|81
2|87
1|134
1|136
1|49
1|54
1|55
1|57
1|62
1|67
1|73
1|78
1|88
1|96

JECOMP-ke|m

578|3
464|1
413|2
407|4
328|5
230|6
206|7
151|8
95|9
87|10
62|11
48|12
40|15
39|13
32|14
22|18
19|16
15|17
15|21
14|20
10|19
9|23
9|24
7|27

5|22
5|26
5|28
4|30
3|25
3|33
2|31
2|32
2|35
2|37
2|41
2|45
1|106
1|29
1|44
1|57
1|59
1|60
1|61
1|66
1|81

JECOMP-word_in_kelm

15366|1
1028|2
642|3
303|4
249|5
241|6
173|7
130|9
128|8
111|10
69|11
59|12
56|13
43|14
40|15
38|16
31|17
27|18
20|20
18|19
18|22
13|24
11|23
9|21
6|25
6|26
6|31
4|27
4|28
4|29
4|32
4|33
4|35
3|30
3|37
3|39
2|34
2|40
2|44
1|104
1|38
1|41
1|46
1|52

1|53
1|54
1|60
1|62
1|63
1|66

JECOMP-sent

927| 1|1
294| 1|2
92| 2|1
63| 1|3
62| 2|2
23| 2|3
12| 3|1
10| 1|4
5| 2|4
5| 3|2
3| 4|1
2| 2|5
2| 2|6
2| 3|3
2| 3|4
1| 1|5
1| 4|3
1| 4|4
1| 5|5

JECOMP-word_in_sent

15| 2|2
11| 6|5
8| 4|5
7| 5|5
6| 6|3

JECOMP-utter

621| 1|1

JECOMP-sent_in_utter

71| 1|1
70| 2|2
53| 1|2
51| 2|3
28| 3|3
28| 3|4
21| 2|4
20| 4|4
19| 2|1
19| 3|2
18| 1|3
14| 4|3
14| 4|5
12| 2|5
12| 3|5
9| 3|1
9| 5|7
8| 3|6
8| 4|2
8| 5|5
7| 4|6
6| 5|8
5| 5|4
5| 5|9
5| 6|6
5| 6|8
4| 5|6
4| 6|2
4| 6|7
4| 8|6
3| 1|4

3	2 6
3	4 7
3	4 9
3	5 2
3	5 3
3	6 4
3	7 6
3	8 5
2	11 13
2	1 6
2	1 7
2	4 1
2	6 3
2	6 5
2	7 4
2	7 9
2	8 12
2	9 8
1	10 15
1	10 9
1	11 12
1	11 17
1	11 5
1	12 9
1	13 16
1	1 5
1	3 7
1	3 8
1	3 9
1	4 8
1	5 1
1	6 1
1	6 10
1	6 11
1	6 13
1	6 9
1	7 11
1	7 2
1	7 5
1	7 7
1	8 10
1	8 11
1	8 15
1	8 17
1	8 2
1	8 4
1	8 7
1	8 8
1	8 9
1	9 11
1	9 4
1	9 6
1	9 9

STRUCT-word_in_phrase

1178| [えー]

1165| 副詞

737| 普通名詞 の

572| [あの]

560| 普通名詞

559| 普通名詞 に

514| 本動詞

436| 普通名詞 を

413| [え]

375| [あのー]

291| 本動詞 ます。

271| 普通名詞 が

247|普通名詞 で
 211|代名詞 の
 209|その
 204|固有名詞 の
 202|普通名詞 は
 188|普通名詞、
 152|はい、
 145|普通名詞 普通名詞 の
 126|普通名詞 と
 126|本動詞 た
 125|で、
 120|この
 119|副詞、
 118|形容名詞 に
 114|[おー]
 110|代名詞
 101|本動詞 ました。
 99|形容詞
 95|本動詞 と
 92|[あー]
 89|本動詞 て おります。
 85|[あ]
 82|()
 80|それから
 75|固有名詞 と
 70|代名詞 は
 70|本動詞 て
 65|普通名詞 も
 64|代名詞 どもの
 62|普通名詞 から
 60|[ま]
 60|形容名詞 な
 57|ありがとう ございました。
 56|本動詞 ます ので、
 54|ありがとう ございます。

STRUCT-bigram

1214|[えー
 1202|えー]
 1000|、 [

614	ます。
608	あの]
589	[あの
491	の方
451	ですね
432	[え
426	え]
419	んです
403	あのー]
396	[あのー
370	して
348	ね、
316	という
301	思います
289	と 思い
254	て おり
222	おります
218	か。
216	まして
206	ます ので
204	いう こと
204	で、
204	ました
202	ですが
196	たいと
192	ので、
188	て 頂き

187|た。
184|ですけれども
178|うか
174|でしょ う
170|が、
156|の [
155|] (
152|はい、
144|です。
142|は、
141|て、
137|が。
129|) [
124|おー]
124|かと
119|けれども、
117|[おー

STRUCT-trigram

1181|[えー]
576|[あの]
452|、 [えー
413|[え]
376|[あのー]
343|です ね、
286|と 思 い ま す
213|て お り ま す
209|思 い ま す。
176|ね、 [
169|と い う こ と
160|う か。
160|でしょ う か
158|ん です が
138|、 [あの
136|ん です け れ ど も
127|、 [あのー
127|、 [え
122|ま し た。
122|ま す の で、
120|お り ま す。
118|た い と 思 い
115|[おー]
111|の 方 で
96|の 方 に
95|な ん です
93|です が。
92|[あー]
91|です け れ ど も、
89|て 頂 き ま す
88|わ か り ま し た
86|ま し て、
85|[あ]
84|し て 頂 き
83|です が、
78|し て お り
77|の で、 [
74|て 頂 き た い
71|です け れ ど も。
69|こ ち ら の 方
69|し た い と
69|で、 [
68|思 う ん です
66|か と 思 い
66|致 し ま す。
65|頂 き た い と
64|た い ん です

STRUCT-word_in_sent

- 30|本動詞 ました。
- 29|はい、本動詞 ました。
- 21|副詞 ありがとうございます。
- 20|さようなら。
- 15|副詞 ありがとうございます。
- 11|ありがとうございます。
- 8|[あ] 副詞 ですか。
- 6|ありがとうございます。
- 6|おはよう ございます。
- 6|さよなら。
- 5|[えー] 本動詞 ました。
- 5|副詞。
- 5|副詞 です。
- 4|[え] 本動詞 ました。
- 4|はい、副詞 ありがとうございます。
- 4|副詞 お本動詞 いたします。
- 3|[あ] ありがとうございます。
- 3|[あ] 副詞 ありがとうございます。
- 3|[あ] 本動詞 ました。
- 3|[えー] ありがとうございます。
- 3|[えー] こんにちは。
- 3|はい、代名詞 ですが。
- 3|はい。
- 3|副詞 副詞 お本動詞 いたします。
- 2|[あっ] ありがとうございます。
- 2|[あ] おはよう ございます。
- 2|[えー] おそれいます。
- 2|[えー] おはよう ございます。
- 2|[えー] 固有名詞 さん ですね。
- 2|[えー] 固有名詞 です。
- 2|[えー] 副詞 ですね。
- 2|[えー] 副詞 本動詞 たご普通名詞 でしょうか。
- 2|[えっと] 本動詞 ました。
- 2|ええ、代名詞 ですが、代名詞 か。
- 2|ええ、本動詞 ました。
- 2|はい、固有名詞 です。
- 2|はい、固有名詞 ですけれども。
- 2|はい、副詞。
- 2|はい [あの] 本動詞 ました。
- 2|もしもし、はい、代名詞 が固有名詞 ですが。
- 2|もしもし。
- 2|サ変名詞 いたします。
- 2|サ変名詞 します。
- 2|形容名詞 です。
- 2|副詞 お本動詞 します。

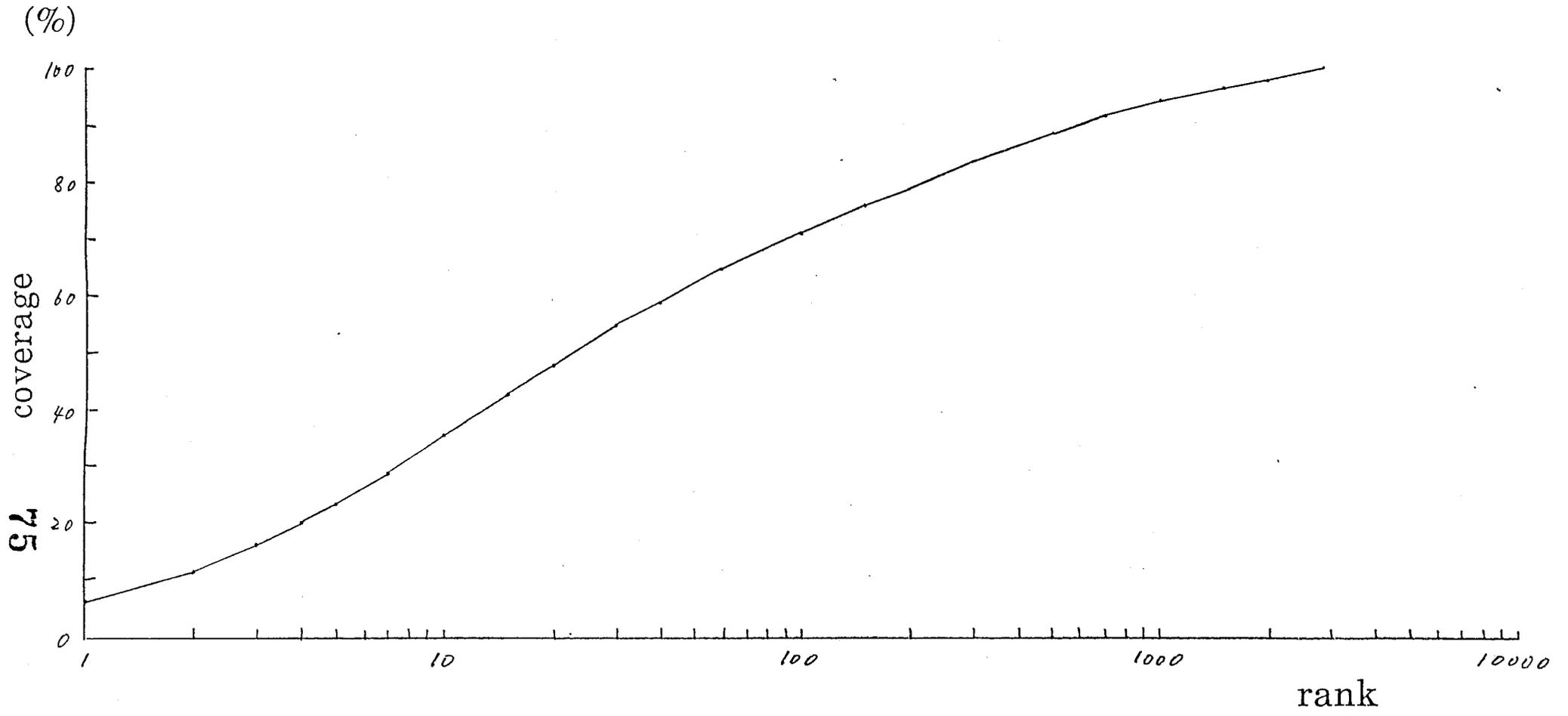
STRUCT-word_in_last_phrase

- 291|本動詞 ます。
- 101|本動詞 ました。
- 89|本動詞 ております。
- 57|ありがとうございます。
- 54|ありがとうございます。
- 49|お本動詞 いたします。
- 25|さようなら。
- 23|普通名詞 です。
- 22|本動詞 ています。
- 21|形容名詞 です。
- 21|普通名詞 ですね。
- 20|本動詞 ん ですが。
- 19|本動詞 ますが。
- 18|お本動詞 します。
- 18|本動詞 ます でしょうか。
- 18|本動詞 ます ので。
- 16|副詞。
- 16|本動詞 ません。
- 16|本動詞 ん ですけれども。
- 14|おはよう ございます。

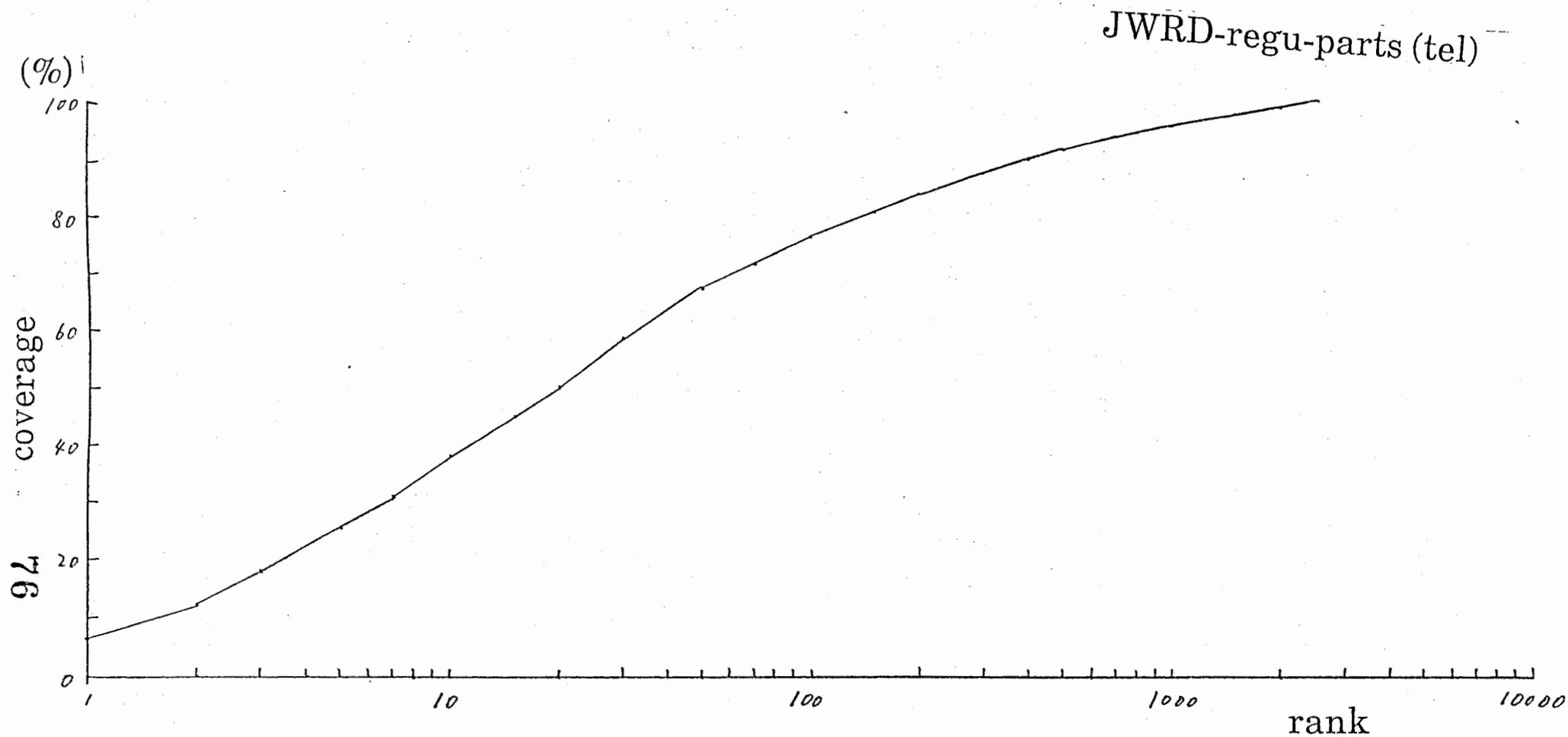
- 12|お 本動詞 しております。
- 12|サ変名詞 いたします。
- 12|副詞 ですか。
- 12|本動詞 んです。
- 11|お 本動詞 したい んです けれども。
- 11|お 本動詞 もうしあげます。
- 11|サ変名詞 さして いただきます。
- 11|形容詞 でしょうか。
- 11|固有名詞 です。
- 10|サ変名詞 いたしました。
- 10|サ変名詞 しております。
- 10|形容名詞 でございます。
- 10|副詞 です。
- 10|本動詞。
- 10|本動詞 せて いただきます。
- 10|本動詞 ン。
- 9|お 本動詞 したい んですが。
- 9|本動詞 ます けれども。
- 9|本動詞 ン でしょうか。
- 8|サ変名詞 させて いただきます。
- 8|普通名詞 でございます。
- 8|普通名詞 なん んですが。
- 8|副詞 でしょうか。
- 7|本動詞 して いただきます。
- 7|本動詞 て いただけます でしょうか。
- 7|本動詞 て ください。
- 6|さよなら。
- 6|固有名詞 ですが。
- 6|数詞 です。
- 6|代名詞 か。
- 6|普通名詞 でしょうか。
- 6|普通名詞 です から。
- 6|普通名詞 で。
- 6|普通名詞 なん です。
- 6|本動詞 てる んですが。
- 6|本動詞 でしょうか。

付録 4. 3 抽出結果のグラフ

JWRD-regu-parts (key)

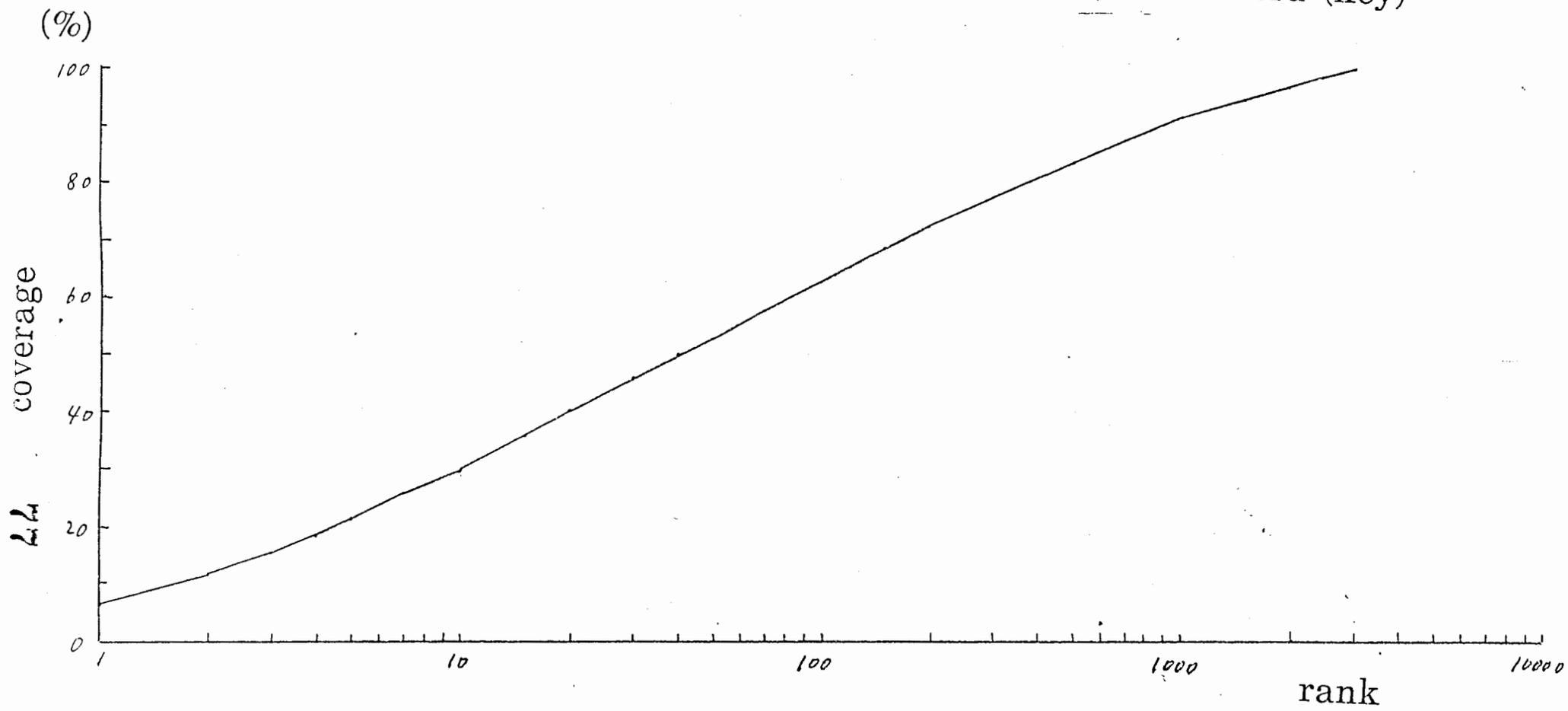


標準表現と品詞のランク対カバー率 (キー)

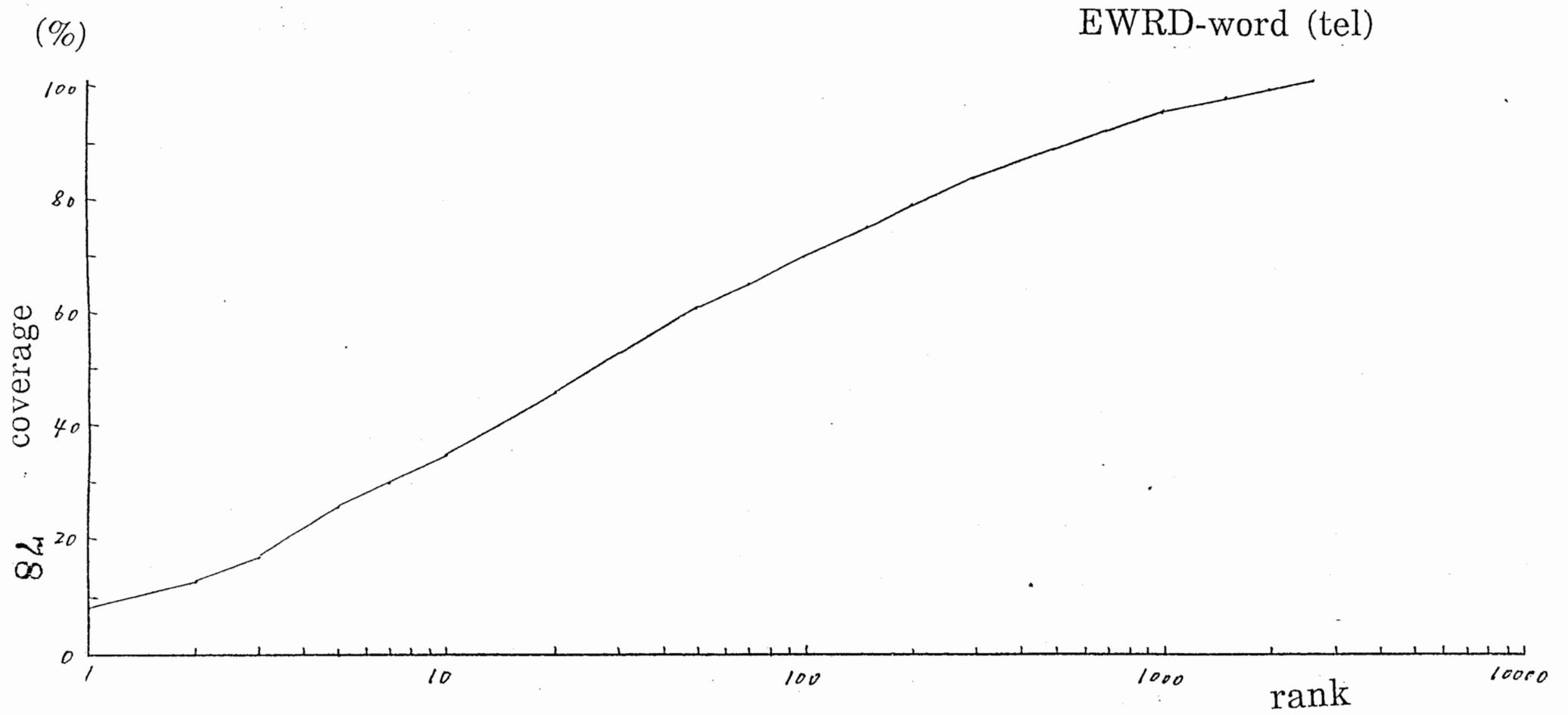


標準表現と品詞のランク対カバー率 (電話)

EWRD-word (key)

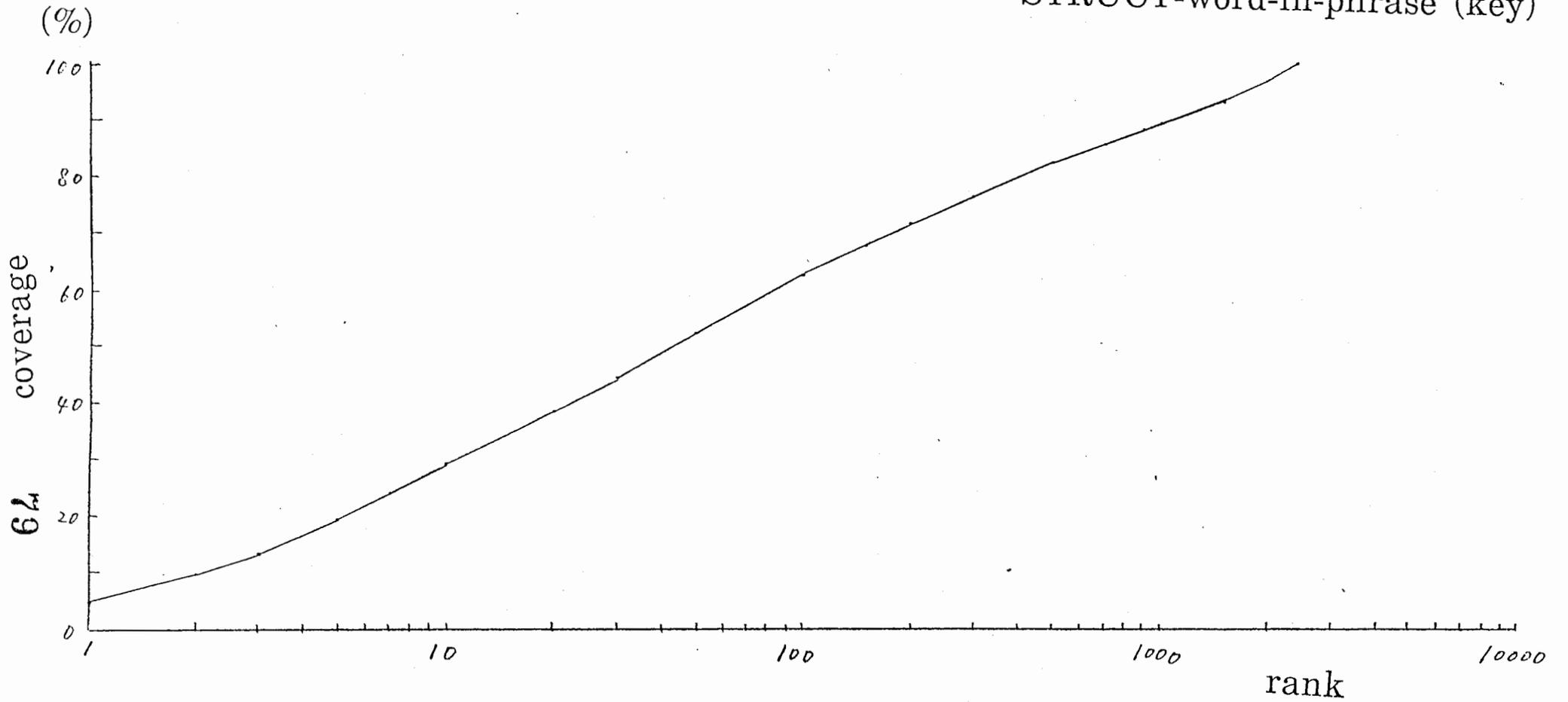


英語出現単語のランク対カバー率 (キー)

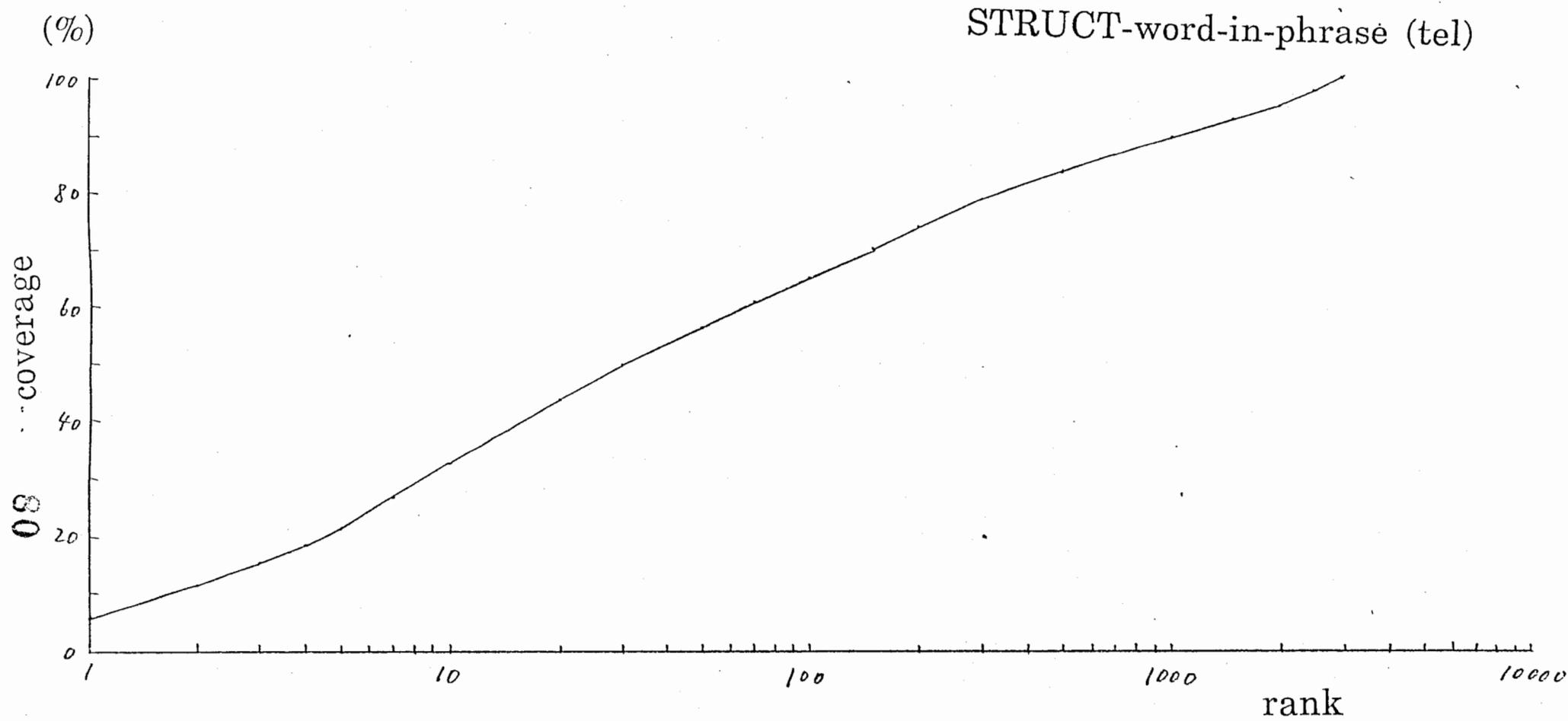


英語出現単語のランク対カバー率 (電話)

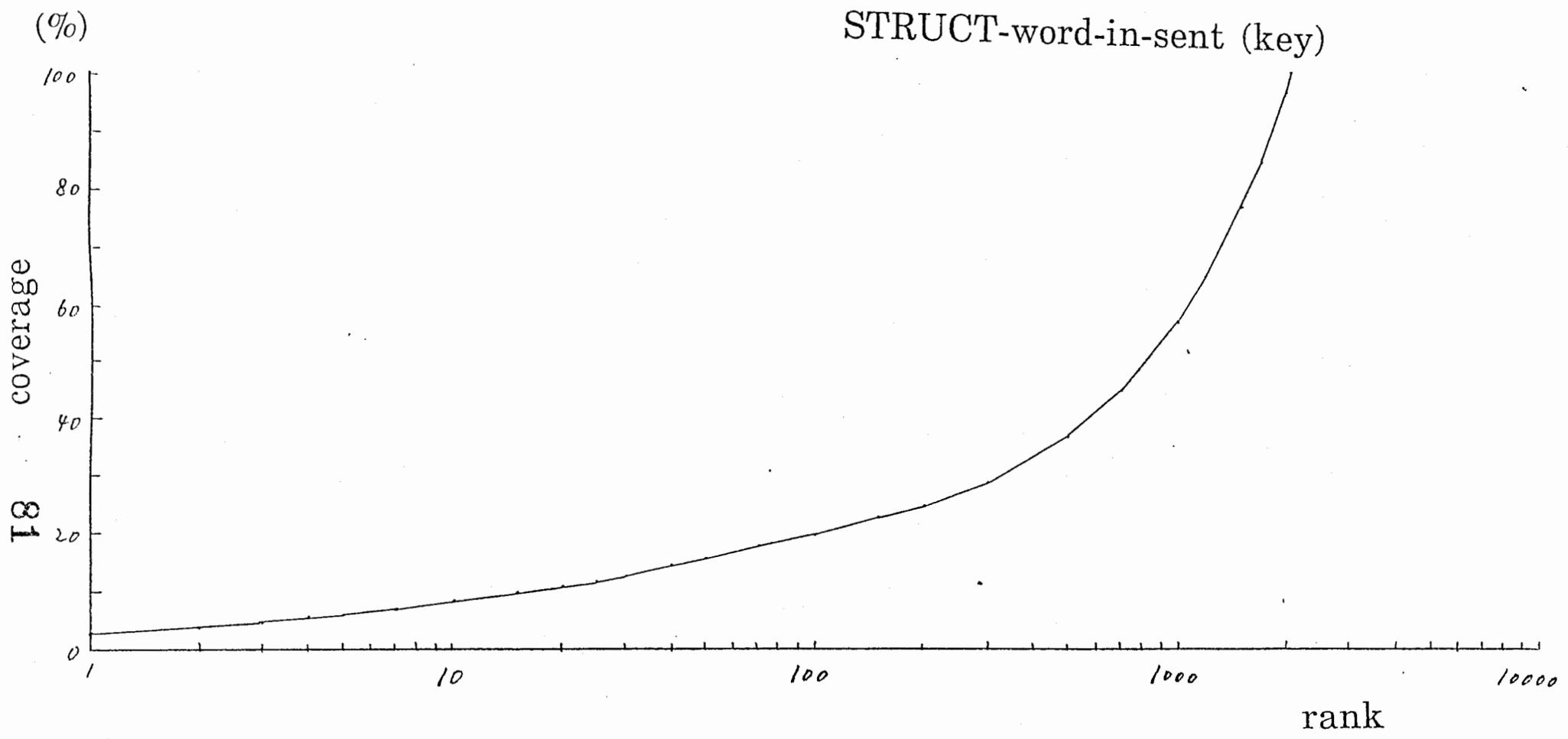
STRUCT-word-in-phrase (key)



文節中の単語並びパターンのランク対カバー率(キー)



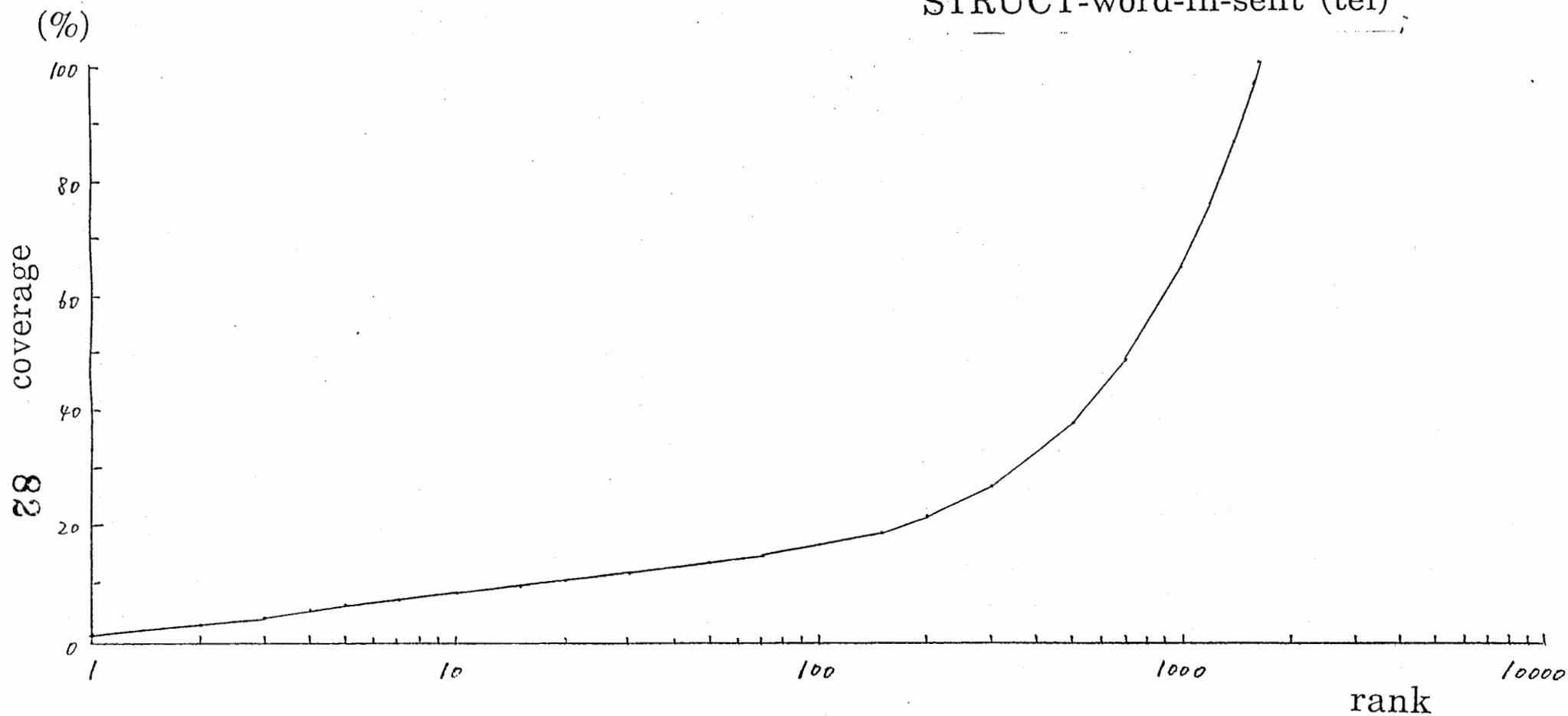
文節中の単語並びパターンのランク対カバー率(電話)



18

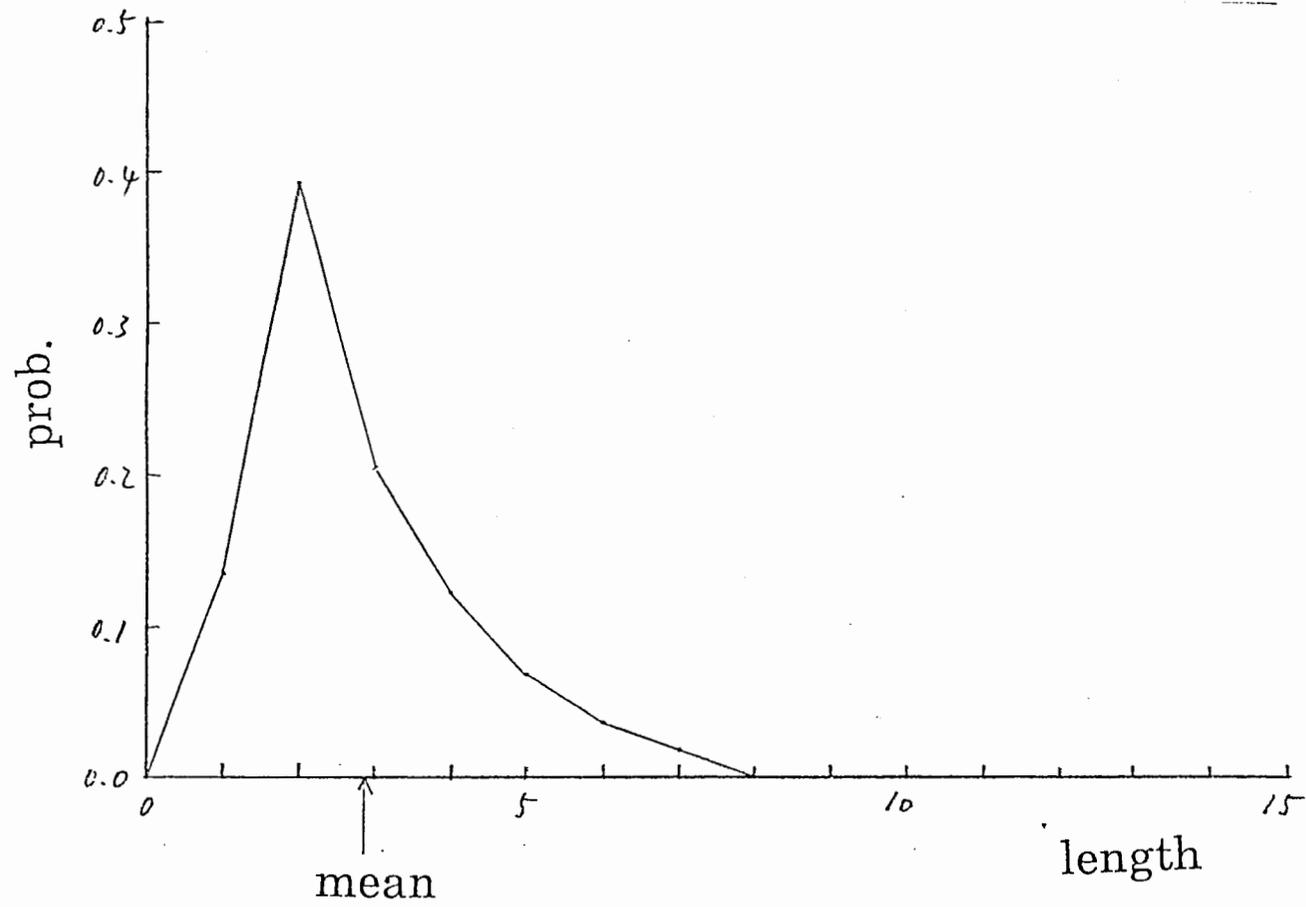
文中の単語並びパターンのランク対カバー率(キー)

STRUCT-word-in-sent (tel)



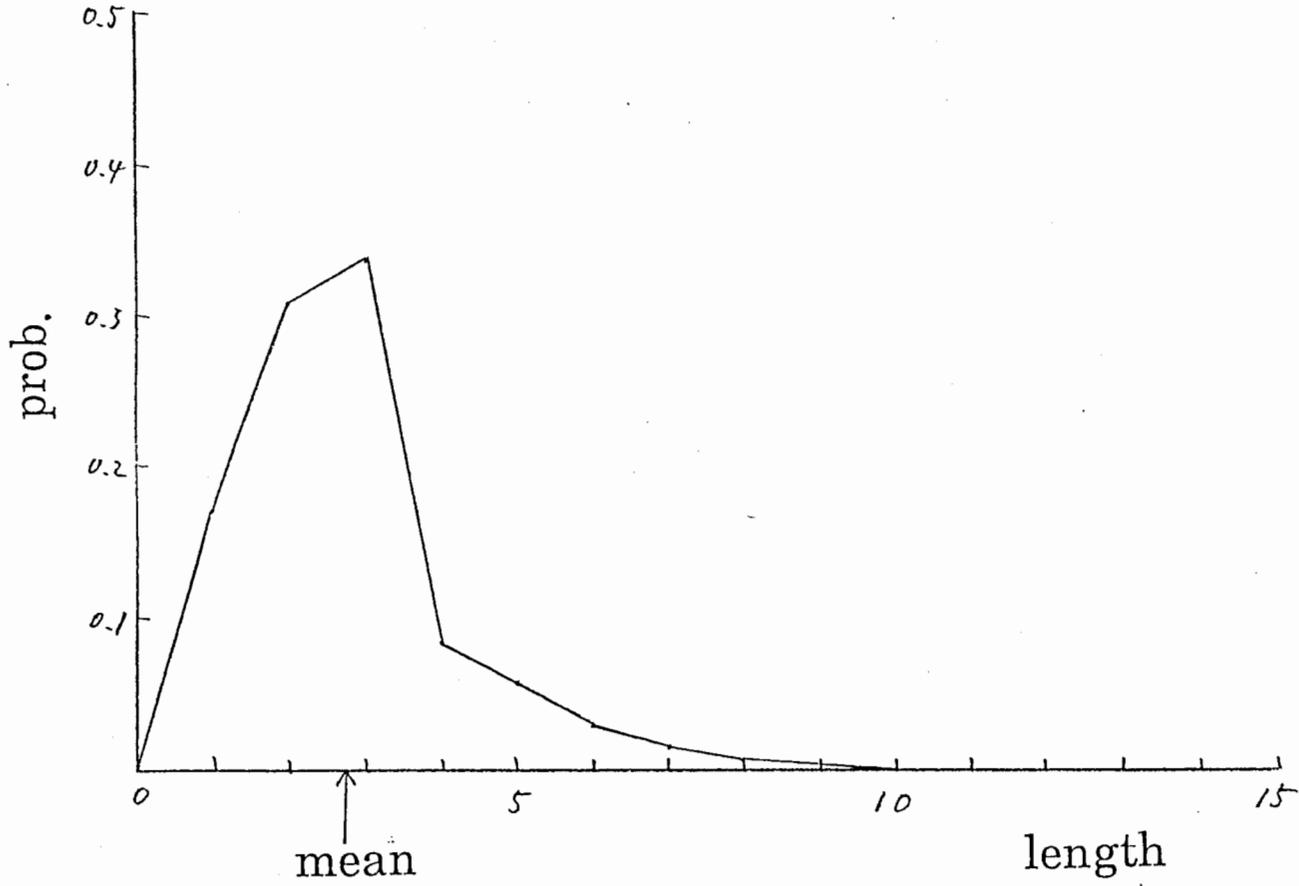
文中の単語並びパターンのランク対カバー率(電話)

COMP-word-in-kaku (key)

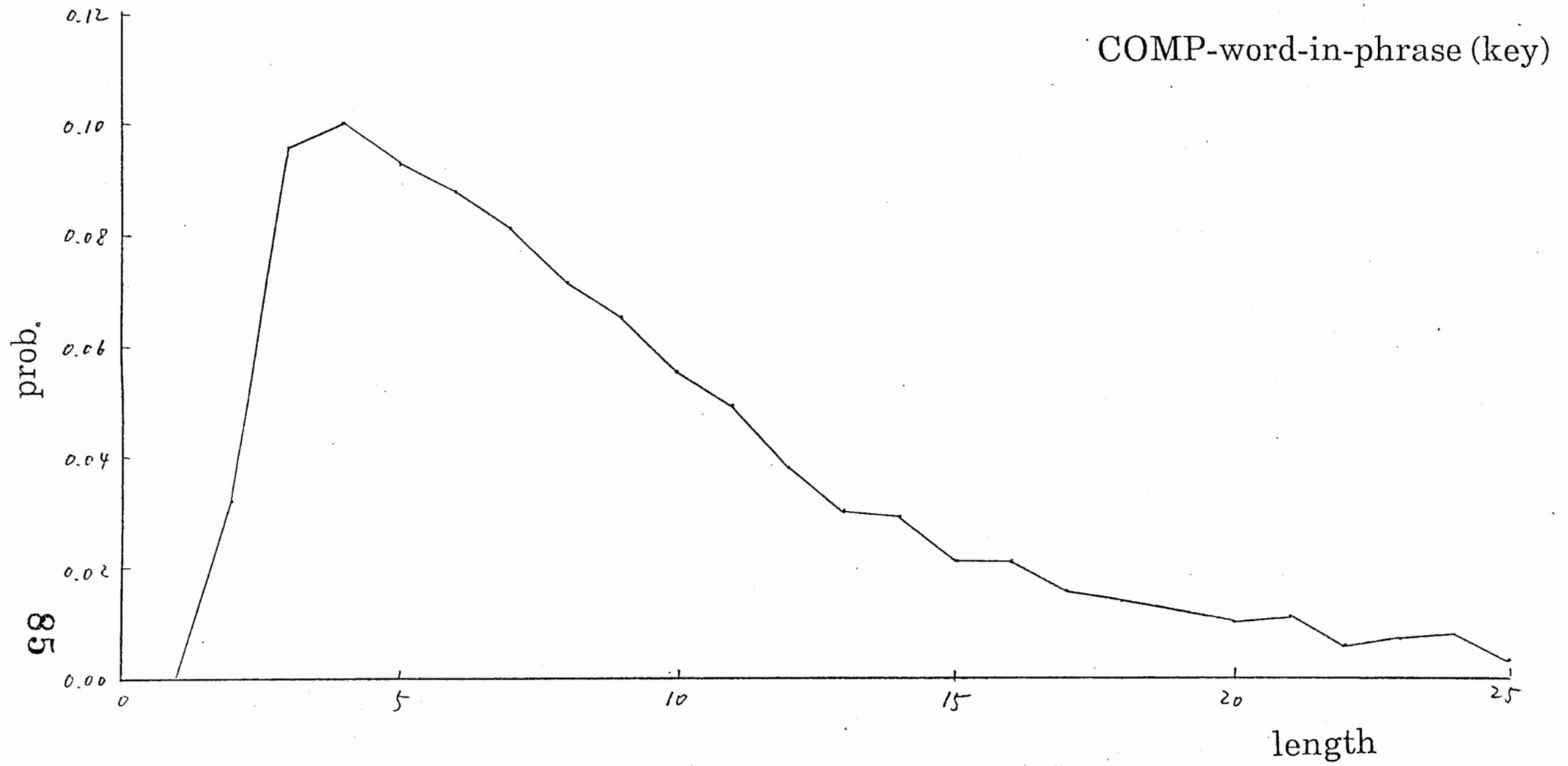


1文節中の単語数 (キー)

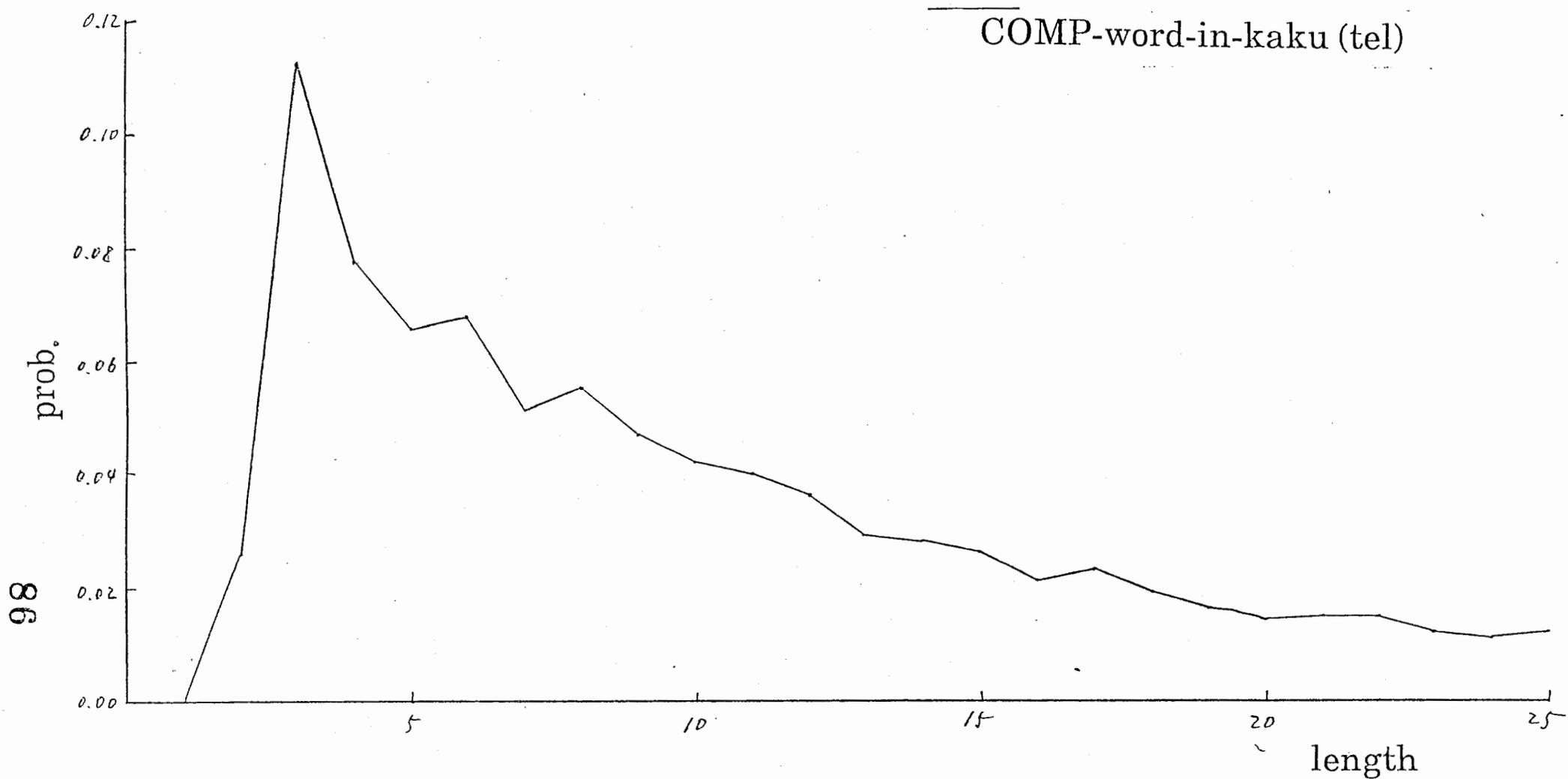
COMP-word-in-phrase (tel)



1文節中の単語数 (電話)

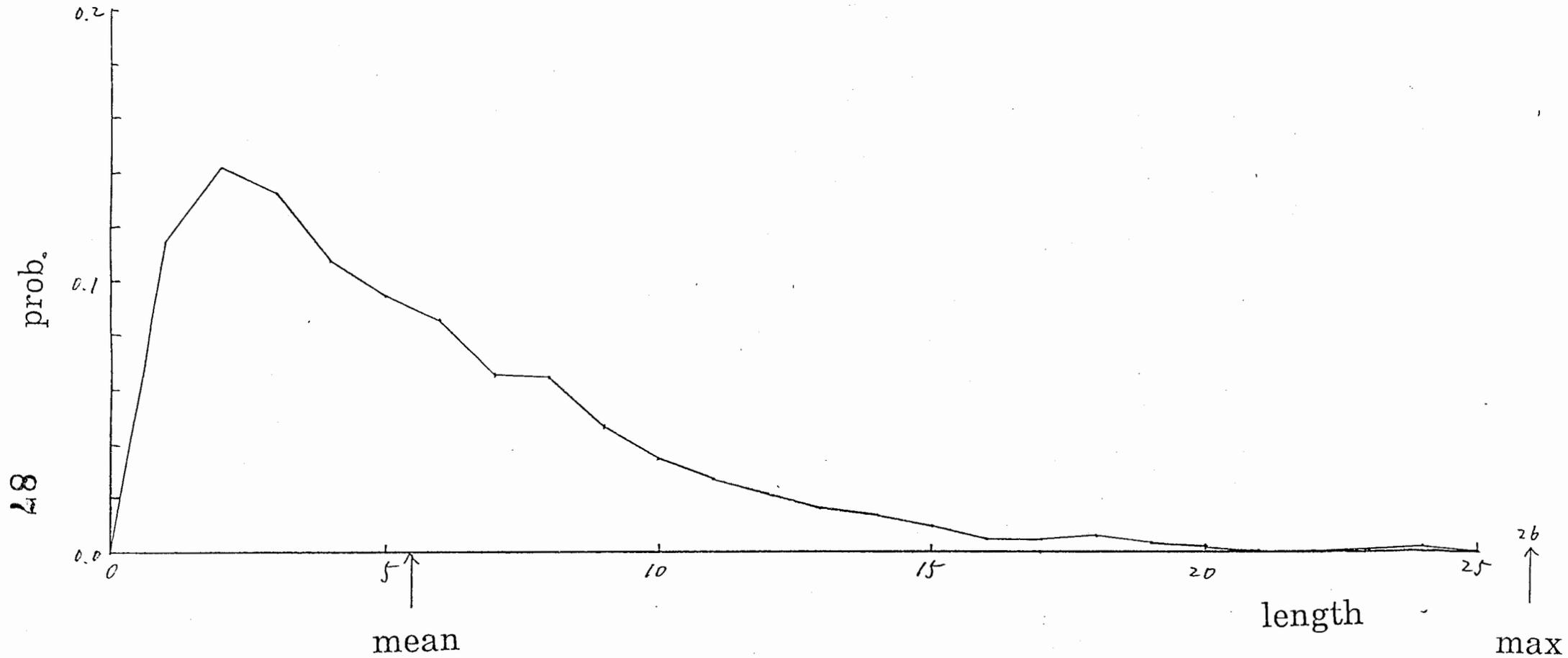


1格構造中の単語数 (キー)



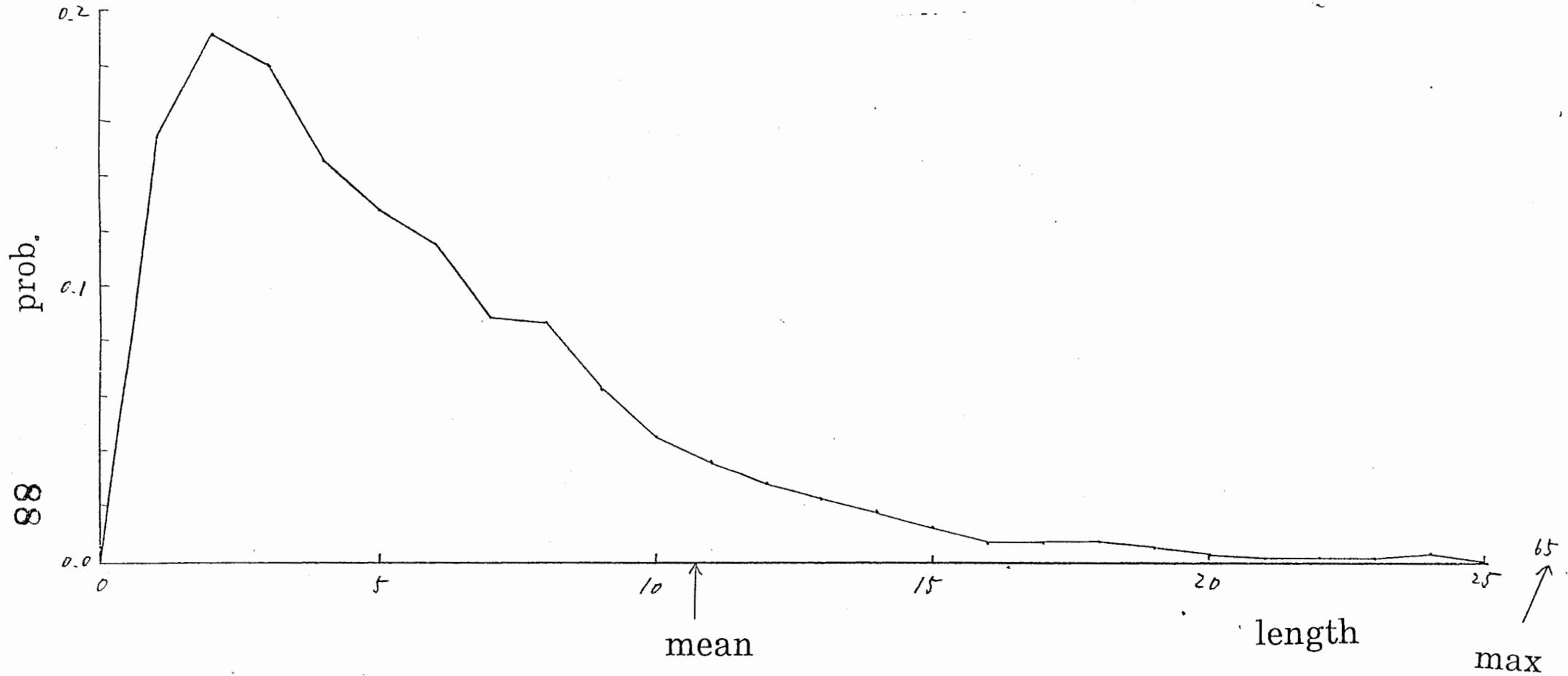
1格構造中の単語数 (電話)

COMP-phrase-in-sent (key)



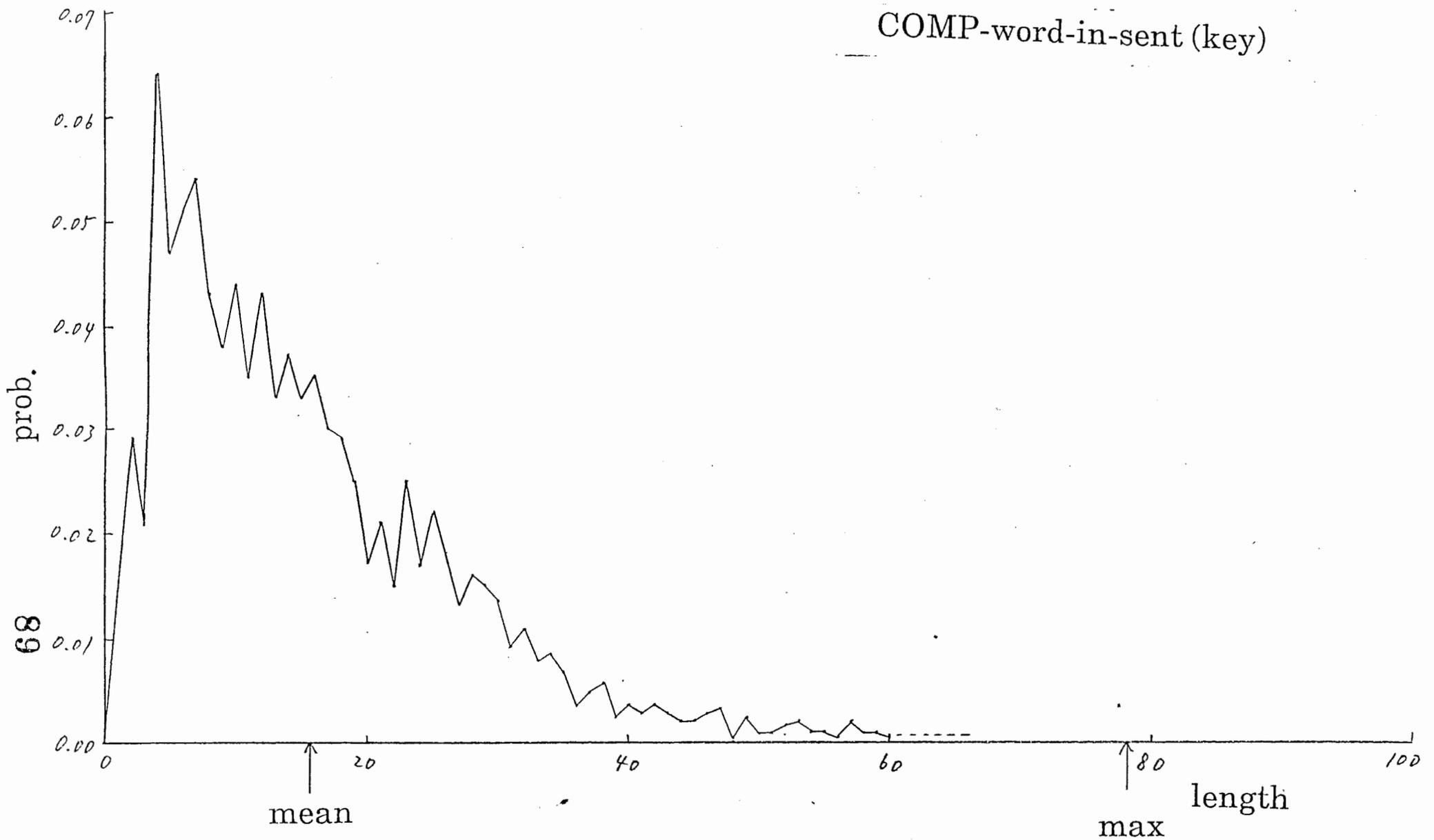
1文中の文節数 (キー)

COMP-phrase-in-sent (tel)



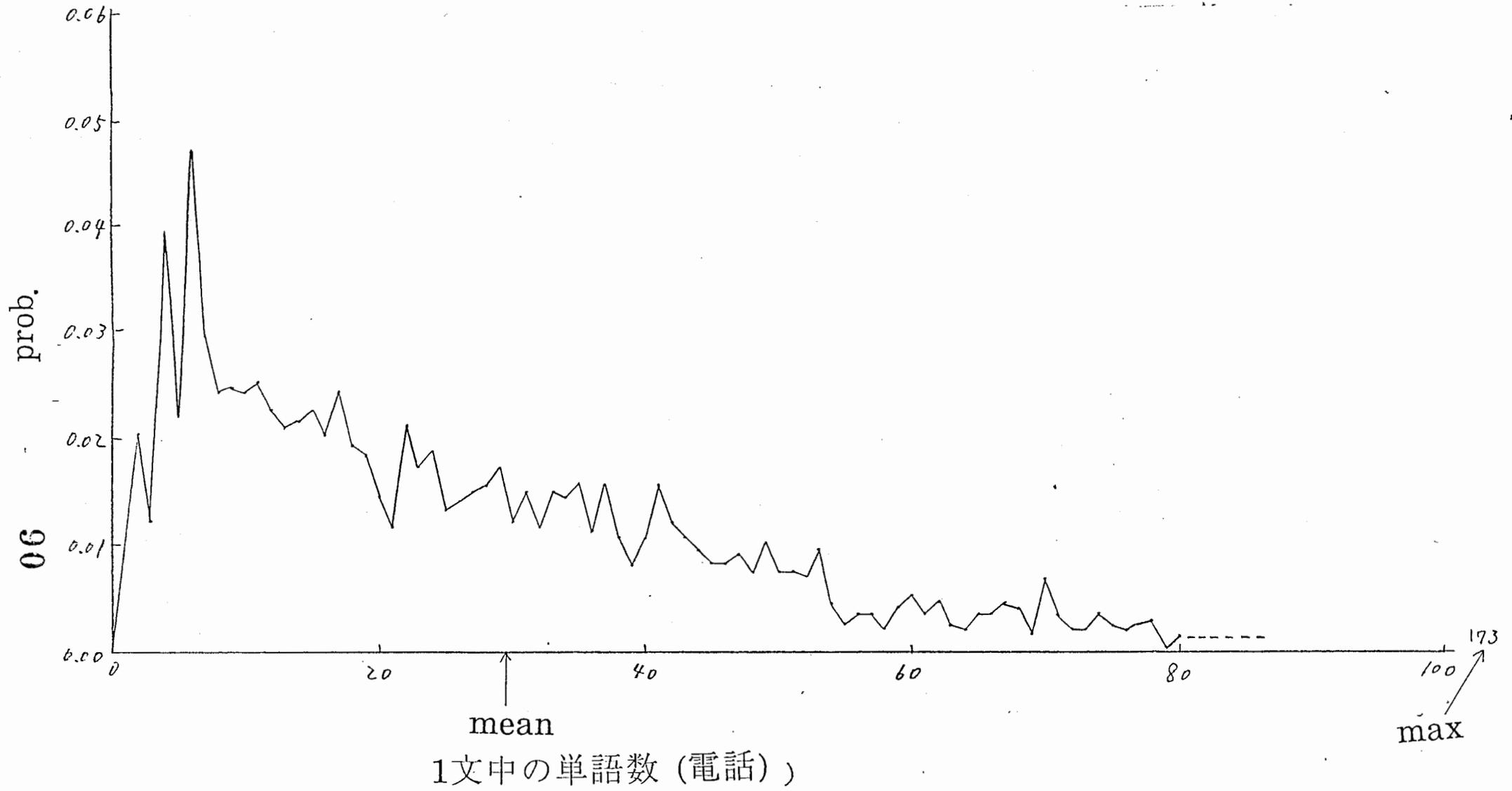
1文中の文節数 (電話)

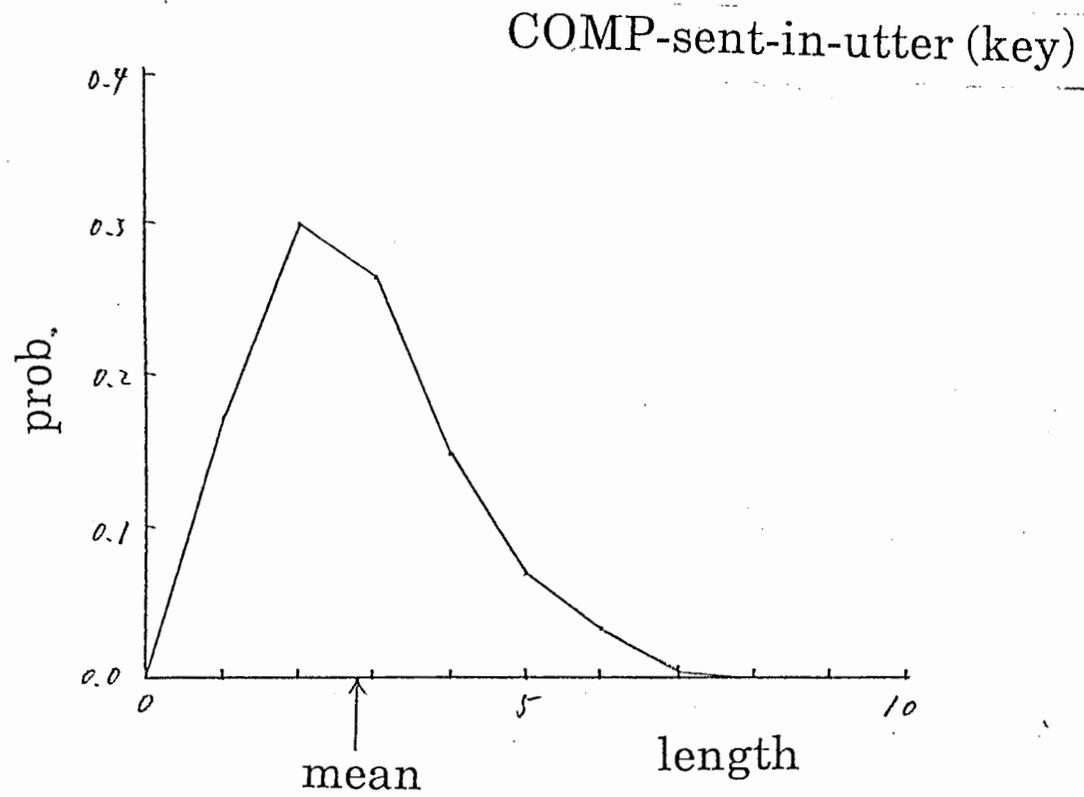
COMP-word-in-sent (key)



1文中の単語数 (キー)

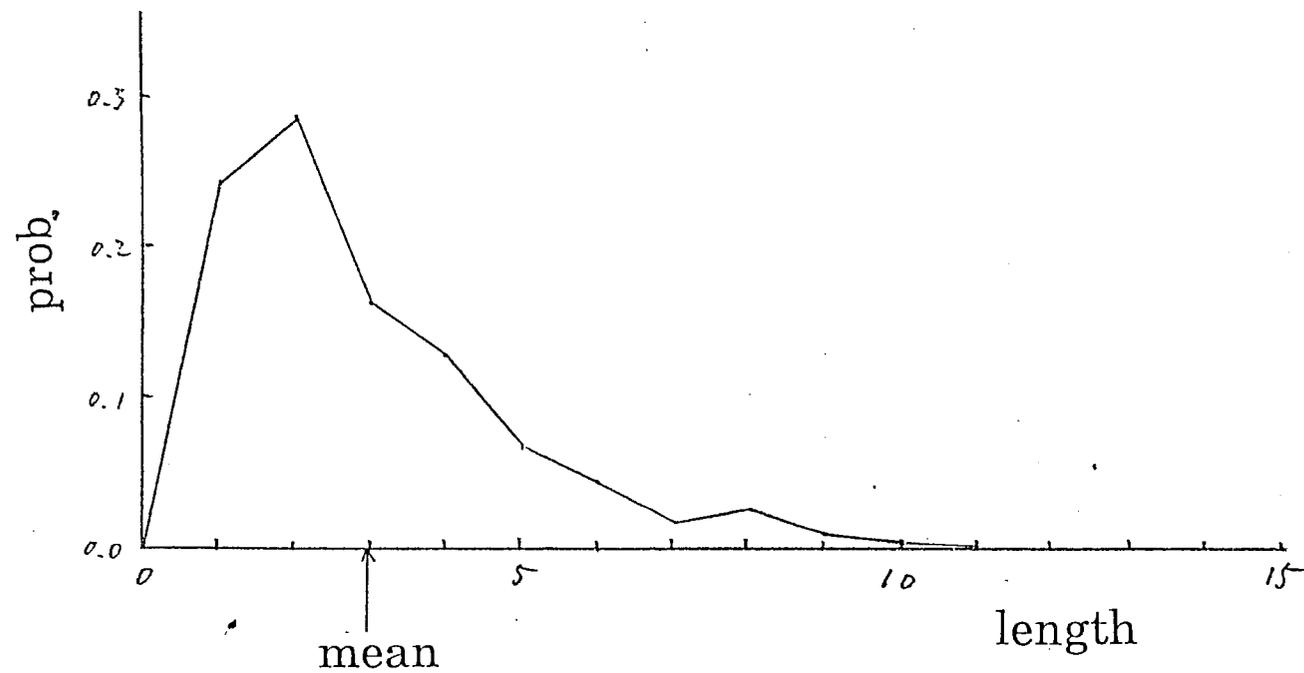
COMP-word-in-sent (tel)



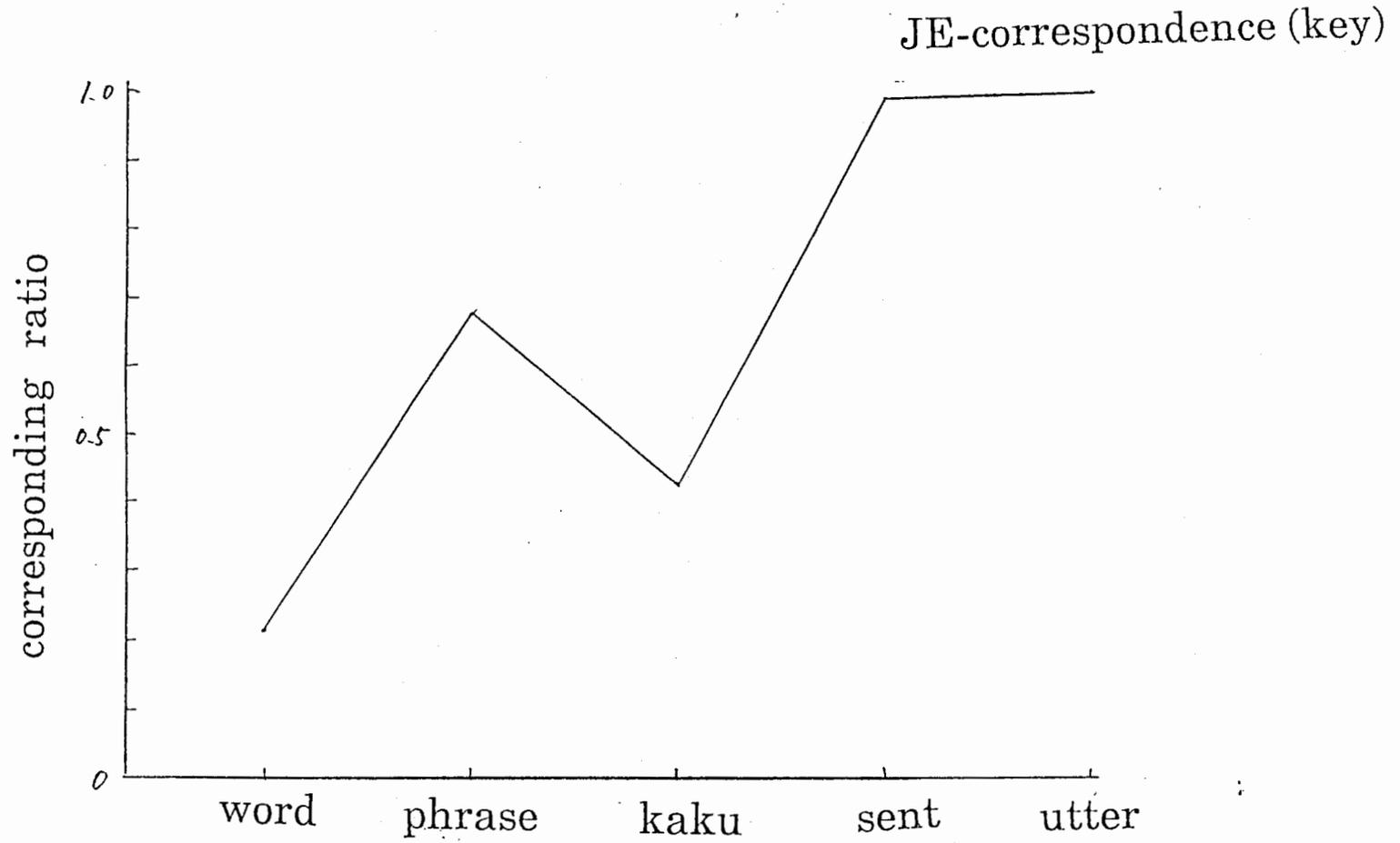


1発話中の文数 (キー)

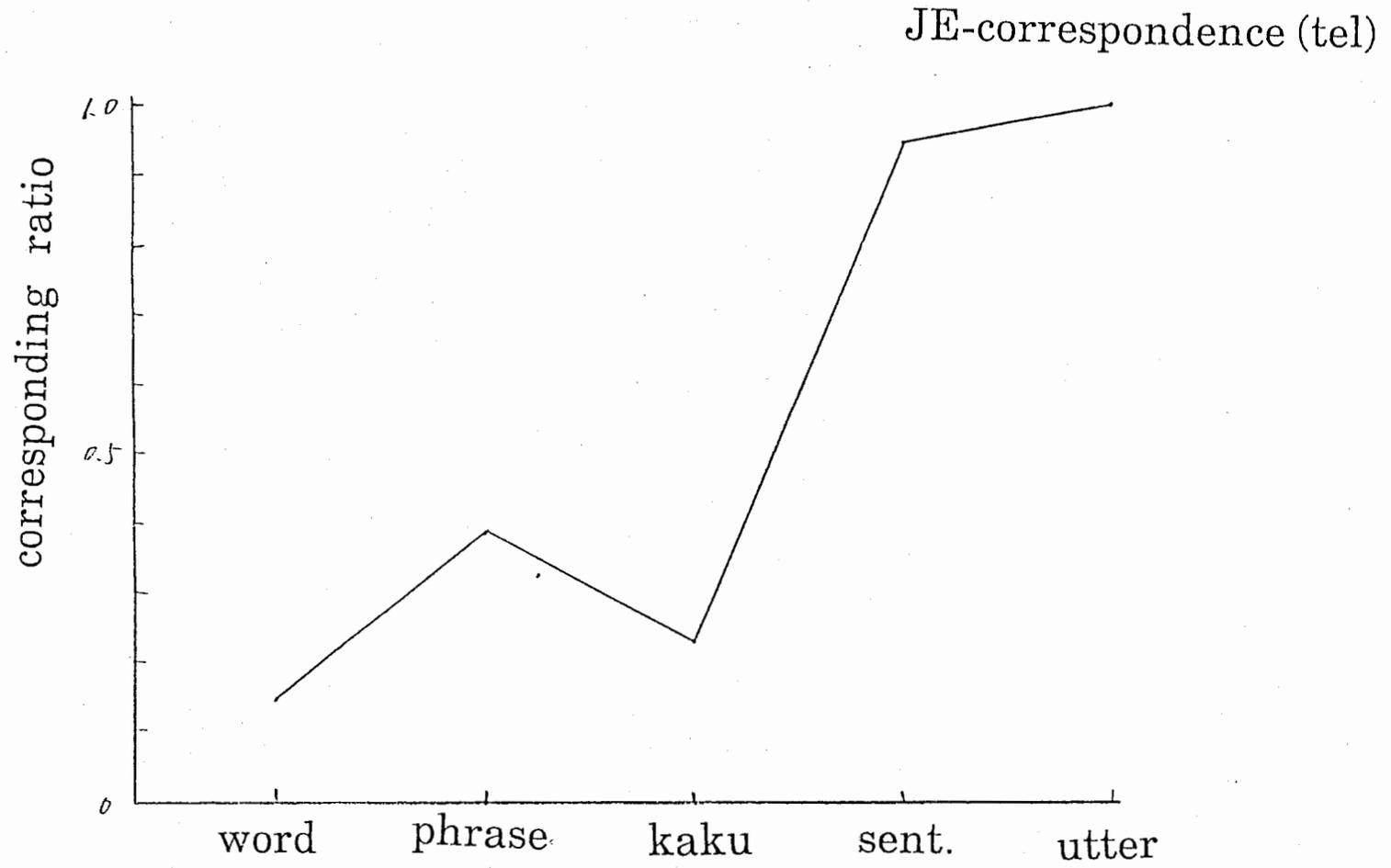
COMP-sent-in-utter (tel)



1発話中の文数 (電話)

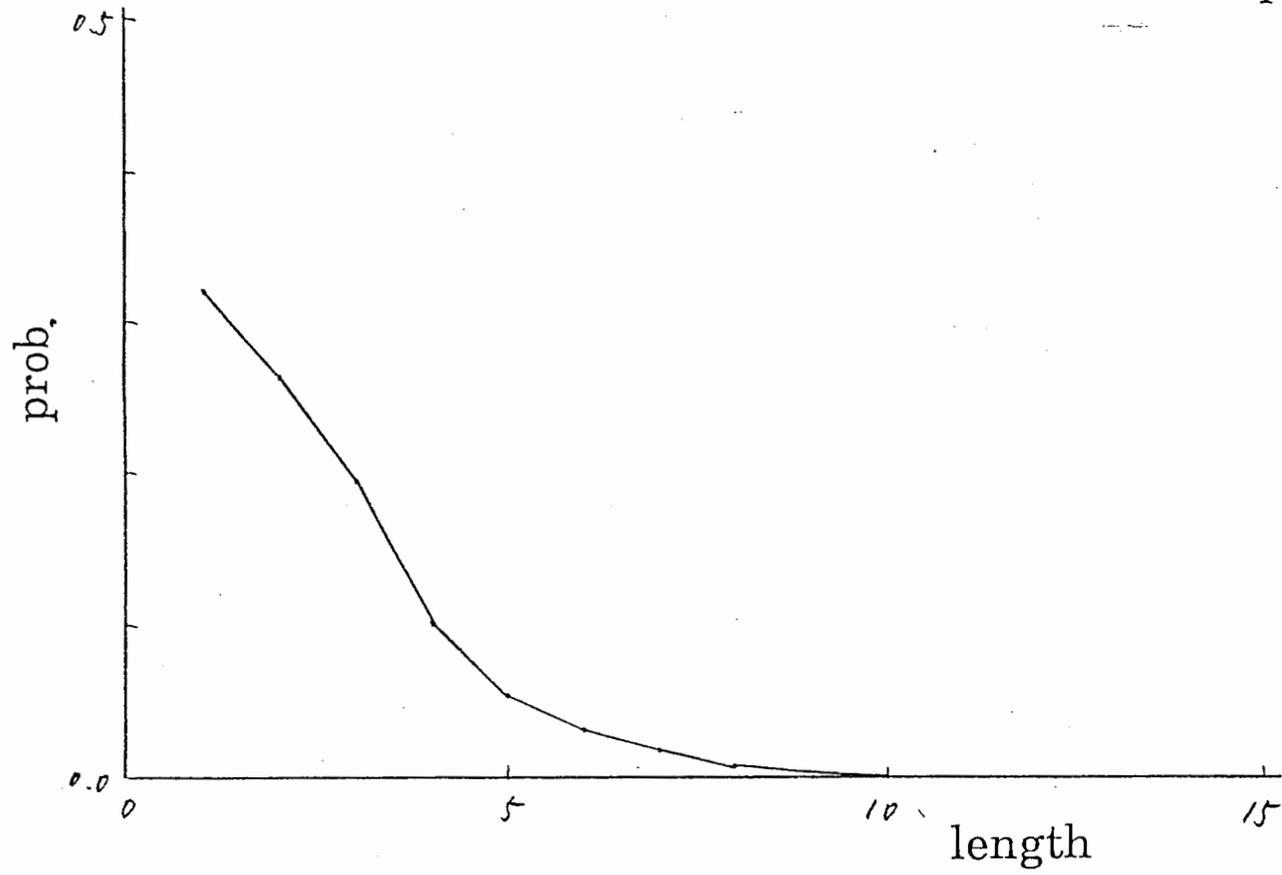


日英対応の対応率 (キー)



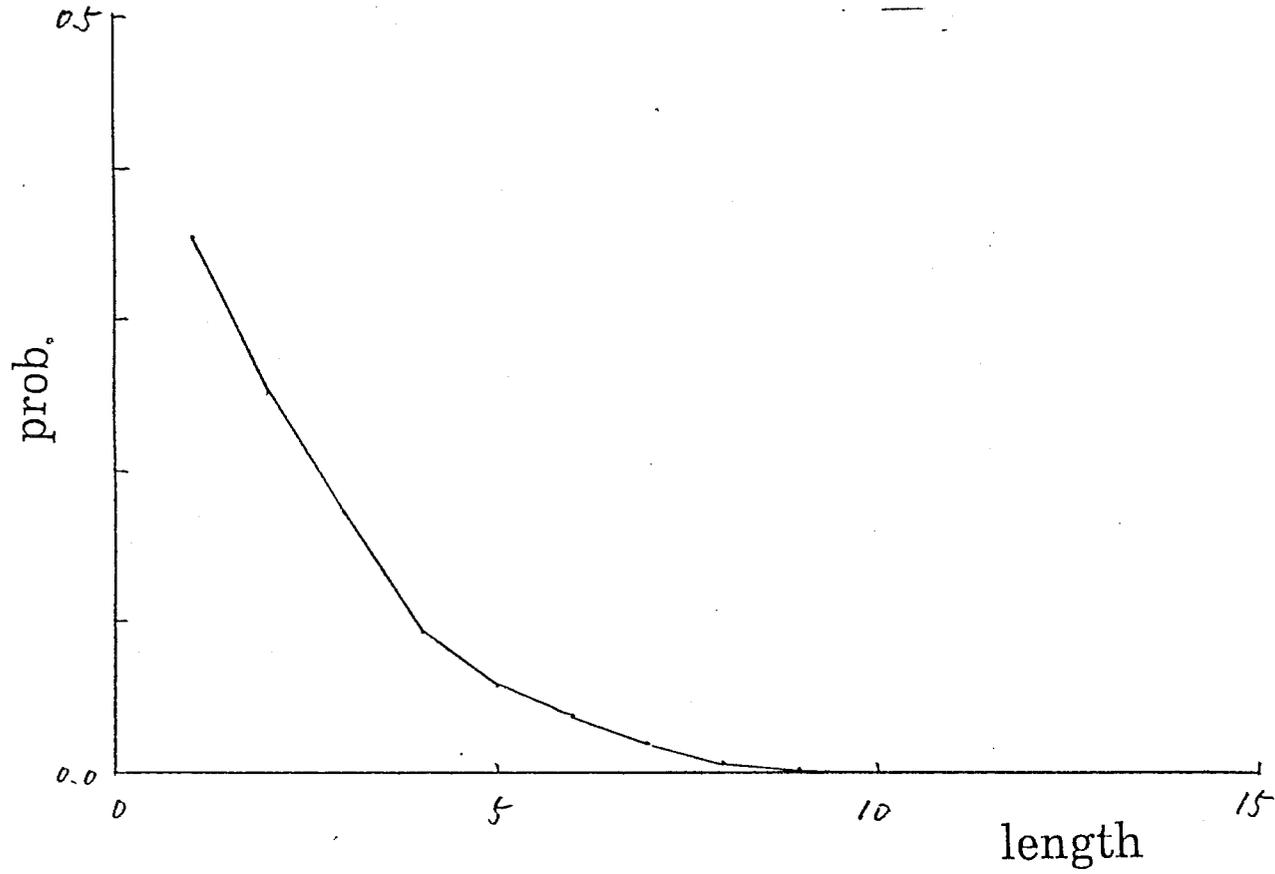
日英対応の対応率 (電話)

JECOMP-word-in-phrase (key)

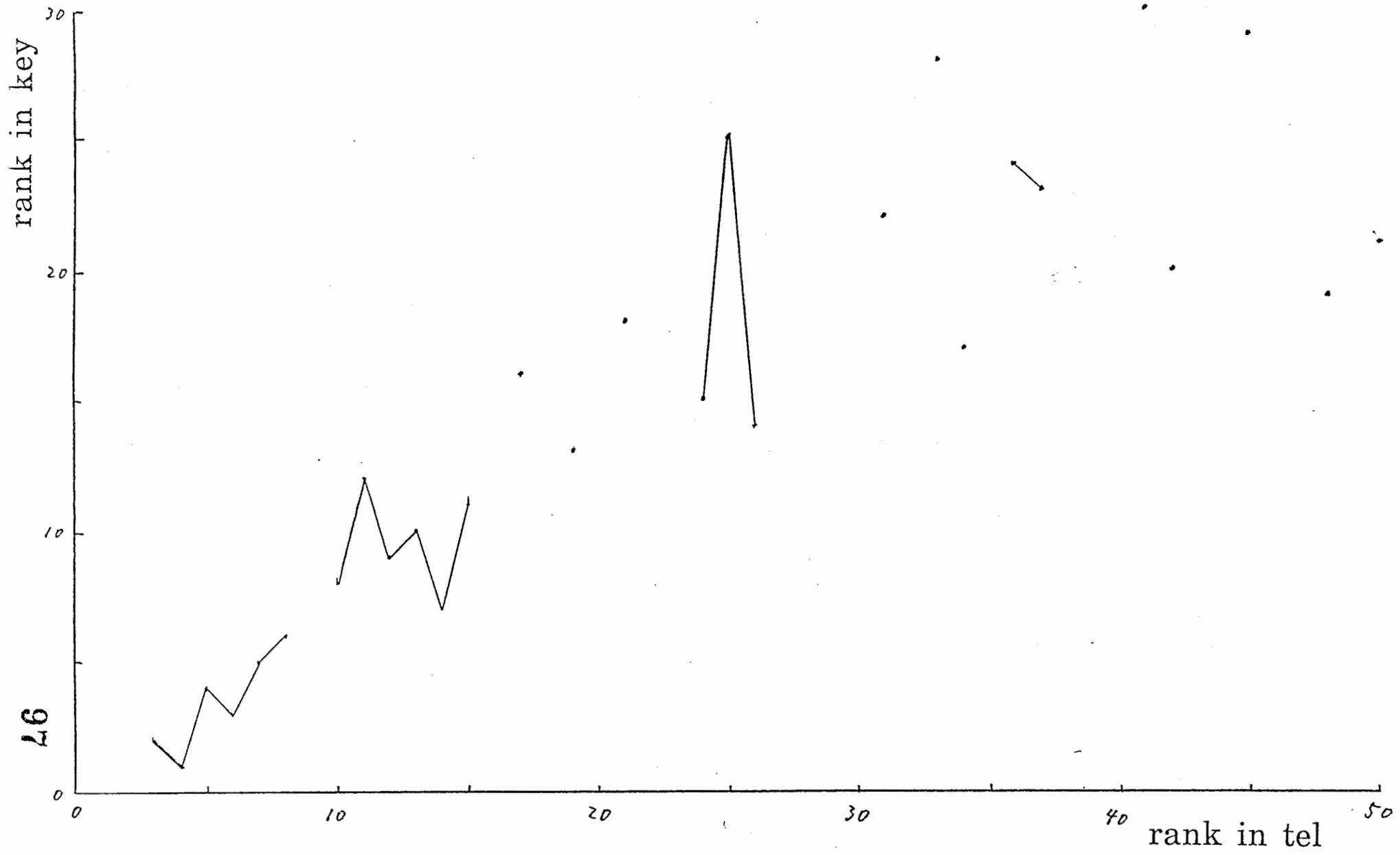


文節対応の単語数 (キー)

JECOMP-word-in-phrase (tel)



文節対応の単語数



46

JWRD-regu-parts

標準表現と品詞のキーと電話の対比