言語学: 日常のことば

言語のありさまを調べるための計算と考え方

ワークブック

山元啓史（やまもとひろふみ）

Ph. D. in Linguistics

言語学: 日常のことば

「システィーナの聖母（Madonna Sistina)」の下部に描かれた天使たち
「言語」と「思考」の関係は相互的か独立的か。
言語があるから思考が成立するのか、それとも思考があるからそういう言語が口から出るのか。

Hilofumi Yamamoto, Ph. D. in Linguistics, Tokyo Institute of Technology
© Hilofumi Yamamoto, 2017

はじめに

本書は東京工業大学リベラルアーツ研究教育院で行われている言語学Ａの教科書である。社会言語学を謳っているようであるが、社会言語学のみを取り上げるのが主眼ではない。社会的存在としての学生自らが自分の知識を通して言語を見つめることが、言語学の入門として有利と考えたまでである。

理論があって、ごく日常で使っていることばがあるのではない。まず、日常で使っていることばがあって、それを分析するとある理論が浮かび上がってくるのである。理論先行型の研究の手始めに、いいことなどないと信じている。その浮かび上がってくる理論を記述するのが言語学の仕事である。なので、日常を見つめることから言語学をはじめる。

山元啓史
東京工業大学教授
2018 年 2 月 5 日

目次

第 1 章　言語と社会 1
1.1 社会言語学 1
1.2 ことばの多様性 1
1.3 単語のカウント 1
1.4 χ^2 検定 2
1.5 さらに勉強するには 2

第 2 章　二言語・多言語の使用 5
2.1 サピア＝ウォーフ仮説 5
2.2 バイリンガリズム 5
2.3 British English と American English 5
2.4 文の長さの測定方法 6
2.5 言語の比較 6
2.6 さらに勉強するには 8

第 3 章　方言 9
3.1 方言 9
3.2 ビデオを見る前に意見交換 9
3.3 留学生と方言 11
3.4 蝸牛考 11
3.5 さらに勉強するには 11

第 4 章　言語の変化 15
4.1 文字数と文字種 15
4.2 昔の日本語と今の日本語 16
4.3 流行語とは何か 16
4.4 敬語の機能 17

第 5 章　使用域と文体 19
5.1 文体論 19
5.2 使用域 19
5.3 内容語と機能語 19
5.4 ピジンとクレオール 21

第 6 章　言語能力と言語運用 23
6.1 共出現パターン 23
6.2 プロトタイプ理論 23
6.3 発表 24
6.4 試験 24
6.5 レポート執筆 24

第 7 章　論文発表会 27
7.1 発表の実施 27
7.2 質問・コメントの方法 27
7.3 クリップボード発表会 27
7.4 ページスタイル 27

第 8 章　まとめ 29
8.1 ディスカッション 29
8.2 修正 29
8.3 投稿 30

第 9 章　口述試験 31
9.1 概要 31
9.2 目的 31
9.3 試験問題 31
9.4 合否判定 32
9.5 結果 32

第 10 章　落ち葉ひろい 33
10.1 プログラミング 33
10.2 統計 33

図目次

1.1 井上作品、筒井作品の品詞による χ^2 検定 3

2.1 井上ひさし「ブンとフン」、筒井康隆「エディプスの恋人」の文長 7

4.1 日本の小説 10,000 行のうちに見られた漢字数と飽和率 15

5.1 DOT 言語の例にしたがったグラフ出力 21

6.1 共出現パターンの作り方 23

表目次

3.1 広島弁ビデオ視聴後の留学生の感想 1 12
3.2 広島弁ビデオ視聴後の留学生の感想 2 13
3.3 広島弁ビデオ視聴後の留学生の感想 3 14

4.1 2012 年度ユーキャン流行語大賞 16

プログラム一覧

1.1　単語のカウント方法 1
2.1　文の長さの測定方法 6
4.1　漢字文字種のカウント 15
5.1　DOT 言語で記述された「春、声、鶯、鳴く」の関係 20

第 1 章

言語と社会

1.1 社会言語学

問 1　社会言語学とはどんな学問であるのかを話し合いなさい。

問 2　テキスト (田中・田中 2015) 中、pp. 2–3 のわからない単語にはどんな意味があるのかを考えなさい。

問 3　社会言語学としてことばの研究するのなら、どんなテーマがよいかを考えなさい。

問 4　テキスト (田中・田中 2015) 中、pp. 4–5 の中のわからない単語にはどんな意味があるのかを考えなさい。

1.2 ことばの多様性

問 5　ことばの多様性の例を日本語と英語で複数あげなさい。

1.3 単語のカウント

問 6　単語のカウント方法を考えなさい。

問 7　動詞、名詞別に単語のカウント方法を考えなさい。

Listing 1.1: 単語のカウント方法

```
% cat inoue.txt | mecab | grep "名詞" | wc
```

1.4　χ^2 検定

問 8　井上作品、筒井作品で動詞、名詞比率が異なるかどうかを調べなさい。

問 9　Easy Chi-Square Calculator を使って、χ^2 検定を実施してみなさい。

問 10　どんな比較をするとおもしろい結果が出るかを考えなさい。

1.5　さらに勉強するには

問 11　$m \times n$ の分割表による χ^2 検定をしてみなさい。

問 12　他の品詞や文の長さなどの計測をし、比較して有意な差があるかどうかを検討しなさい。

問 13　自分の興味にあった言語分析のテーマを選び、リサーチ・クエスチョン[*1]を作れ。

[*1] Research question: 研究上、追求すべき問題。Yes か No で応えられるような簡素な答が結論となるような問題文。「可能性があるかどうか」のような存在の有無でよいが、具体的に「…であることが根拠で…の可能性があるかどうかを検証する」のような文が望ましい。

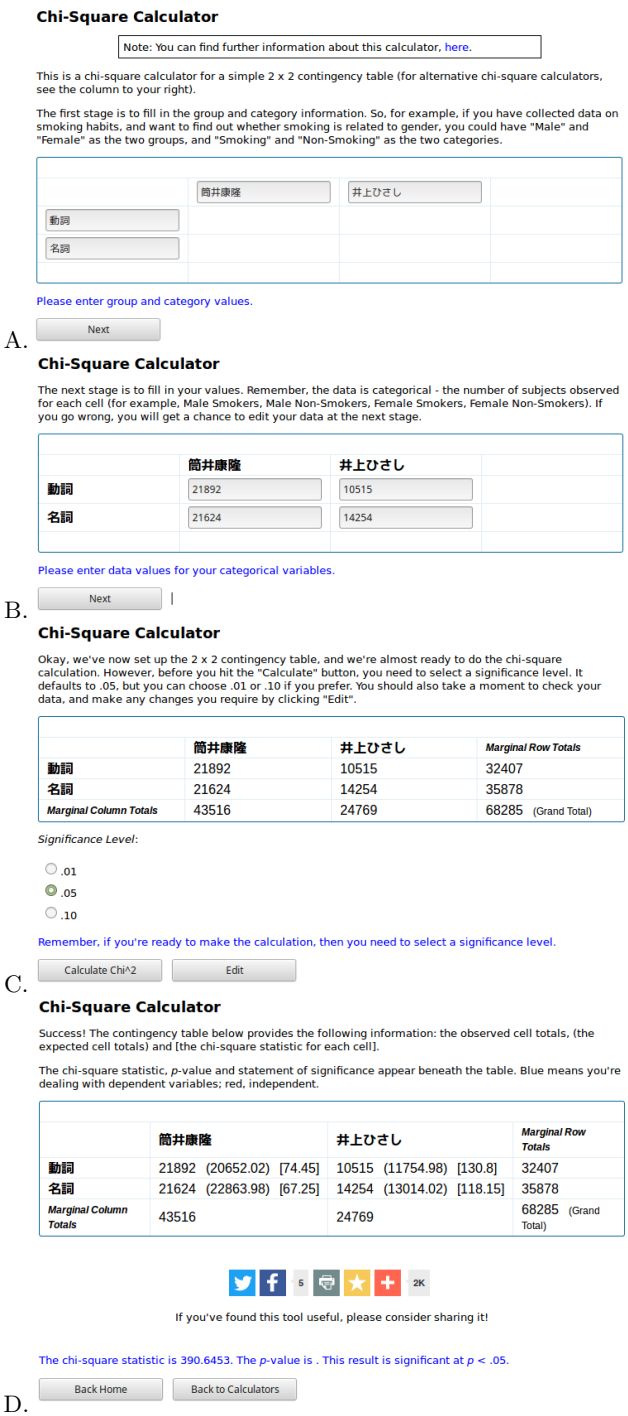

図 1.1: 井上作品、筒井作品の品詞による χ^2 検定: 両作品の品詞を数え、その比率を χ^2 検定してみると有意であった。

第 2 章

二言語・多言語の使用

2.1 サピア＝ウォーフ仮説

問 14　サピア＝ウォーフ仮説は教科書のどこに書いてあるか。

問 15　サピア＝ウォーフの仮説はどんな仮説かを説明しなさい。

問 16　サピア＝ウォーフの仮説が正しいかどうかはどうすれば証明できるかを考えよ。

2.2 バイリンガリズム

問 17　バイリンガリズムは教科書のどこに書いてありますか。

問 18　一般にバイリンガルと言われている人たちを示すバイリンガルと教科書で言われているバイリンガルとの違いについて話し合いなさい。

問 19　二言語・多言語の使用 不完全言語使用について調べ、ノートを作りなさい。

2.3 British English と American English

問 20　British English と American English のビデオを見て、どのように違うかをペアで確認しなさい。

問 21　British English と American English をまねてみなさい。

問 22　British English と American English の違いをノートに記しなさい。

問 23　変種の例をあげてみなさい。

問 24　なぜ言語に変種が現れるのかを考えなさい。

問 25　変種の研究をするにはどのような方法を使えばよいか考えなさい。

2.4　文の長さの測定方法

問 26　2 つの作品、あるいは 2 人の作家による作品、それぞれにおける文の長さに違いはあるかどうかを調べなさい。

問 27　どうすれば文の長さが測定できるかを考えよ。

問 28　頻度を調べたら、グラフに書いてみてどのような特性があるかを考えなさい。

問 29　頻度を対数に変換したら、グラフはどう変化するかを話しあいなさい。

Listing 2.1: 文の長さの測定方法

```
% cat inoue.txt | sed -e 's/。/。\n/g; s/　//g' | awk '{if(length($1) > 0){print $1,
    length($1)}}' | awk '{print $2}' | sturges -c | graph -TX
```

'sturges' はスタージェスの公式 $k = 1 + log_2 N$ によってヒストグラムの区間を計算している (Sturges 1926)。

問 30　対数変換するのはなぜか。対数変換する前後は、どんな分布型であるか。

問 31　文の長さの分布はどのような分布であるかを、話しあいなさい。

問 32　井上作品と筒井作品の文の長さの違いについて話しあいなさい。

問 33　2 つの分布型が同じか、違うかを判定するにはどうすればよいか。

2.5　言語の比較

アイスランド語とグリーンランド語は、いずれも数万人から数十万人によって現在も話されている言語である。2 つの言語の違いについて、以下の記述から考察せよ。

アイスランド語

アイスランド語は印欧語族、ゲルマン語派北ゲルマン語群に属する言語。使用範囲はアイスランドのみで、使用人口は約 30 万人。アイスランドを「氷島」と表記することから「氷島語」略して「氷語」や「氷」とも言う。9 世紀頃にノルウェーから移住したバイキングがもたらした言語であり、他の北ゲルマン語（デンマーク語・ノルウェー語・スウェー

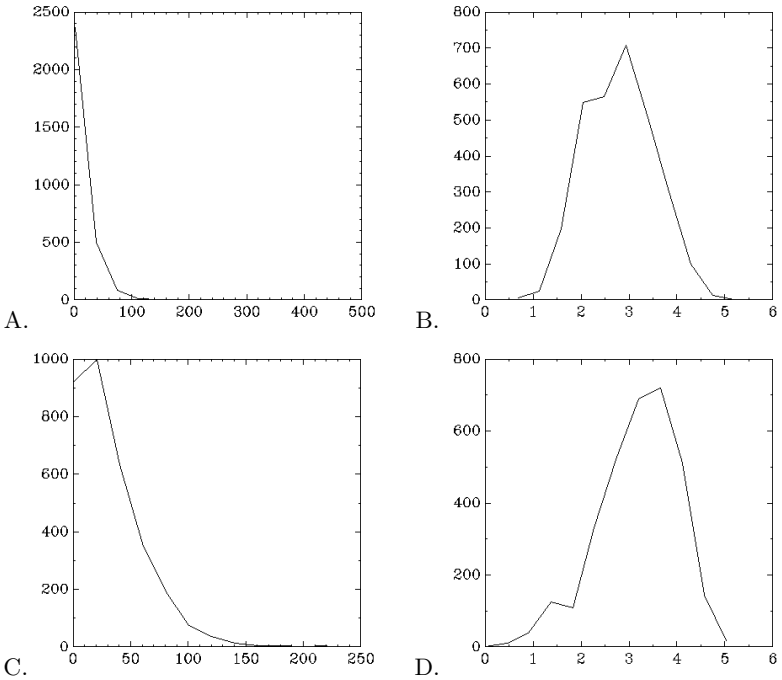

図 2.1: 井上ひさし「ブンとフン」A, B、筒井康隆「エディプスの恋人」の文長 C, D。BD はそれぞれの数値を対数変換したもの。縦軸は頻度、対数変換前の横軸は文字数を表す。

デン語）の中ではノルウェー語に一番近い。他の北ゲルマン語が失った 3 性、つまり、他の北ゲルマン語は両性名詞（共性名詞）と中性名詞だけになってしまっているが、この 3 性や格変化などを保持している。3 性・格変化が必要なため、英語やフランス語などからの借用語が排除されるためか、古風な色合いを残していると言われる。

グリーンランド語

グリーンランド語 (Greenlandic, Greenlandic Inuktitut) はエスキモー・アレウト語族の言語のひとつ。デンマーク領グリーンランドのカラーリット（Kalaallisut; グリーンランド・イヌイット）によって話され、カナダのイヌクティトゥット語などと同族である。イヌクティトゥット語のうちグリーンランドで話される諸方言をグリーンランド語、イヌクティトゥット語の一方言のカラーリット語 (Kalaallisut) をグリーンランド語と呼ぶこともある。話者数ではカラーリット語が約 5 万 7 千人で、学校などで学習するのはこの西方言である。

問 34　どちらの言語が安定しているか。それはなぜか。

2.6　さらに勉強するには

問 35　言語が類似しているとはどういうことかを考えよ。

問 36　言語の古い形が残る条件は何かを考えよ。

問 37　日本語はどの程度昔にまで遡れるかを考えよ。

第 3 章

方言

3.1　方言

問 38　日本語の方言を 2 つ取り上げて、その違いについて話し合いなさい。

問 39　(田中・田中 2015) の pp. 10–11 の中のわからない単語はどんな意味かを考えてみなさい。

問 40　電車のアナウンスやテレビ放送は方言で放送されるかどうかを話しあいなさい。

3.2　ビデオを見る前に意見交換

問 41　テレビなどのからの印象から大阪弁の特徴を話しあいなさい。

問 42　テレビなどのからの印象から広島弁の特徴を話しあいなさい。

3.2.1　大阪弁

大阪弁は近畿地域の方言の一つ、総称として呼ばれることもある。

以下はよしもと新喜劇の定番ギャグのひとつ「乳首ドリル」を一部文字化したものである。乳首ドリルは 2010 年頃からたびたびマスコミ、雑誌などで取り上げられている吉田裕・すっちーによるギャグで、

> 大阪の小学生の間で大流行！！小学生が選ぶ新喜劇イチオシギャグに吉田裕の “乳首ドリル” （芸能社会面 SANSPO.COM サンスポ）

に示すように言語を伴う文化は小学生のような若い世代に大きな影響を及ぼしていることがわかる。また、2014 年の「小学生が選ぶ新喜劇イチオシギャグ」に吉田裕の “乳首ドリ

ル” が選ばれている（SANSPO.COM 2014.6.2）。

問 43　アンダーラインの箇所を標準的な言い方にしなさい。

1. 吉田: おいコラ！なにお前 a.ぶつかっとんじゃ コラ！
2. すっちー: ごめんなさいね
3. 吉: オレ誰か b.知ってんのんかい？
4. す: マキバオーね
5. 吉: 誰がマキバオーじゃ、誠意みせんかい！
6. 吉: ワァーオイ！ワァーオイ！
7. 吉: c.こーへんのかい！
8. 吉: 来るか思たら来ーへんのかい！…何か喋れよお前も！ア〜？ d.ちゃうんかいこら！
9. 吉: つま先やめろ、お前！
10. 吉: あごやめろ！お前！
11. 吉: 脇やめろ！脇やめろ！毛細血管がいっぱい詰まってるとこ！わきー！！
12. 吉: 毛細血管がいっぱい詰まってるとこ、わきー！！！
13. 吉: なんで e.聞こえへんねん、お前！この距離やど！
14. す: 毛細血管がいっぱい詰まってるとこ、わきーという部分が聞き取りづらい…
15. 吉: そう f.言うたんや！
16. 吉: お前おかしいぞ…乳首ドリルすな！
17. 吉: 乳首ドリルすな…ドリルすな…すなすなすなすなすなすなすな…
18. 吉: すなすなすなすな…誰か止めてーーー！！！すなすなすな…
19. 吉: つま先、あご、脇やめろ…ドリルせんのかーい！！
20. 吉: ドリル g.せんのかい！
21. 吉: h.すんのかと思ったらせんのかい！すんのかと思ったらせんのかい！…すんのかい…すんのーかい

問 44　標準的な言い方にできたかどうかを隣の人と話し合いなさい。

問 45　方言の特徴を具体的に述べよ。

3.2.2　広島弁

広島弁を考察するために「Apple 動画の広島弁吹き替え」を用いる。このテレビコマーシャルは 2010 年ごろから Apple の商品説明コマーシャルが広島弁に翻訳されて YouTube をはじめとする動画サイトにアップされるようになった。その自然な翻訳が話題になり、一挙に Apple 以外に多数の翻訳広島弁がアップロードされている（`https://nyoho.jp/wiki/?`広島弁吹き替えシリーズ）。尚、本筆記は Jeffrey Pannell の協力によるものと述べている。

問 46　下線に相当することばをビデオを見て書き取りなさい。

1. こりゃ a._____ が信じ b._____ ことなんじゃけどね。

(This is what we believe.)

2. テクノロジだけ c._____ 足りんのんよ。
(Technology alone is not enough.)
3. より速う、より薄う、より軽うなる... d._____、みなあええことじゃああるわいの。
(Faster, thinner, lighter...those are all good things.)
4. e._____ けど、テクノロジがどっかしらんに f._____ とき、...
(But when technology gets out of the way,...)
5. すべてがもっとうれしい g._____ になる。もっと魔法みたいなもんになるんよ。
(everything becomes more delightful...even magical.)
6. それがあんたが躍進するときよ。(That's when you leap forward.)
7. それがあんたがたどり着くときよ、なんか・・・ h._____ もんにの。
(That's when you end up with something like this.)

問 47　書き取れたかどうかを隣の人と話し合いなさい。

問 48　方言の特徴を具体的に述べよ。

3.3　留学生と方言

留学生に広島弁のビデオを見せた後、広島弁の難易度と理解について質問した。

問 49　留学生は、広島弁の難易度はどのくらいと考えているか。推測可能なレベルか。

問 50　留学生にとって、広島弁は推測可能なレベルだと考えているか。

3.4　蝸牛考

問 51　「蝸牛考」の理論について説明しなさい。

問 52　「蝸牛考」に出てくるような例をあげてみなさい。

3.5　さらに勉強するには

問 53　大阪弁・広島弁と他の方言との類似性、相違点について述べよ。

問 54　大阪弁・広島弁フィルタ（あるいは方言変換 Proxy サーバ）というコンピュータ・プログラムあるいはサービスを探して、自分の文章がどのように変換されるかを試みなさい。

表 3.1: 広島弁ビデオ視聴後の留学生の感想 1: 難易度と理解

1. 外国人にとっては難しいと思う。理由は方言のなかに、標準語と一致しない文法的機能を果たす言葉があるから、それを知らなければ方言を理解するのが難しいから。特に広島は東京からかなり離れていて、距離の影響を考えると、標準語と結構違うところが多いので、広島弁は方言の中でもかなり難易度の高い方だと思う。
2. 関西弁より難しい。半分ぐらいしか聞き取れない。
3. 少し難しい。大体は推測できるが、推測できない言葉もある。
4. 語尾が変わるか元の表現の発音に近いものは問題なく推測できるが、特定な使い方が出たら全く推測不能でした...。
5. 私の考えでは、標準語より難しいのは当然で、大阪弁より難しい程度ではないかと思います。沖縄弁よりは地理的に首都に近いから易しいはずだと思います。
6. 広島弁は他の方言と比べれば、そんなに難しくない。文法の違いがあっても何となく意味が通じるが、言葉に関してはそうではないと思います。
7. 詳しい意味は把握し難いけど、予測はできる。
8. 単語の変形が多いため最初に聞いた時には理解できなかったが、私はアニメで広島弁を 30 分ほど聞いたことがあるため、変形された単語もなんとなく理解ができた。標準語が分かっていれば難しくないと考えた。推測可能なレベルだと思う。
9. やや難しい。言っていることの大体な意味を推測できる。
10. 中レベルだと思う。
11. 標準語と単語の使いかたが少し違うところはあるが、十分に推測可能なレベルだった。ただ、抑揚まで学ぶことは非常に難しいと思う。
12. 多くの場合はなんとなく分かるが、分からない場合もあります。

問 55　大阪弁変換プロジェクト（`http://www.geocities.co.jp/Hollywood/4078/osaka/`）にはさまざまな辞書と変換サーバがある。これらの基本技術について話し合いなさい。

表 3.2: 広島弁ビデオ視聴後の留学生の感想 2: あなたの感じ方はどうでしたか。

1.	わからないところが多いので、ちょっとびっくりした。
2.	北朝鮮の言葉、もしくは韓国の方言のように聞こえた。日本語だけど、あまり日本語ぽっくなかった。
3.	親しみやすい、面白い。
4.	弓道部にいたごろ飲み会の後先輩が不意に使った廣島弁は可愛かったが、アップルの CM のせいか洗脳を受けてる気がしました。
5.	全然洗練されてない感じ。iPad のコマーシャルとしては違和感を感じます。
6.	最初に聞いたとき、理解するのに少し時間がかかったが、聞けば聞くほどだんだん理解できるようになりました。
7.	韓国語の方言を聞くときと雰囲気が似ていると思った。
8.	最初に聞いた広島弁がアニメからのものだったせいか、可愛い方言だと思った。
9.	穏やか、古い、そして親切な感じである。
10.	広島人はきっと優しい性格を持っているんだよね。
11.	静かで、低い音調の方言の感じだった。関西弁とは重なるところもあれば、完全に違うところもあるのが結構面白かった。
12.	標準語より親切な言い方と思います。

表 3.3: 広島弁ビデオ視聴後の留学生の感想 3: 留学生が方言に遭遇した時、どうすればよいでしょうか。

1. わからないところがあったら、恥ずかしくて聞けない人はいるが、絶対聞いておいた方がいい。方言を使おうと思わないで、理解できればそれで十分だと思う。
2. 下手な真似はしないで、東京弁で、何回から繰り返し聞いてみる。完全言語に頼るよりは他のところから察知する。
3. わからない時はどんなに努力しても無駄なので、話し手のジェスチャーや表情など、言葉以外の表現をよく観察し、少しでも理解できるよう努力する。
4. 全く分からない表現が来たら一個一個の発音を聴き取って推測せずにまず空気を読みましょう！
5. 方言使う人も、標準語を知っていると思います。テレビなどいわゆるマスメディアに接しているはずなので。ゆえに、外国人であることを明かして標準語で話し合います。
6. 方言が聞き取れないことが多いと思うので、誰かに方言で話しかけられた時があったら、ゆっくり話してもらうのは一番いいと思う。しかし、言葉の意味で困っていたらちゃんとその言葉がわからないって言った方がいいです。
7. 特に他の対応はしなくてもいいと思う。もし知らなかったら聞いてみるぐらいでいいと思う。
8. 関西弁のアクセントを簡単に勉強しておく。広島の人たちも自分の方言以外の方言は分からないはずなので、下手な関西弁でも悪いとは思わなく、むしろ親しみを感じると思う。
9. 意味が分かれば、普通の日本語で話し合う。意味が分からない場合、標準語で話すことを要求するか、ほかに当たるかである。
10. まずは通用標準日本語の言葉に置き換えることができるならば一番いいだが、どうしても聞き取りにくい場合では相手に標準日本語で交流してくれないかとお願いするしかないだろう。
11. 分からなかったところは繰り返して聞いてみることがいい。積極的に接近すると、広島弁の使い手も喜んでおしえてくれるだろうと思う。
12. もう一回話を聞かせて、表情と動きなどを意味を推測する。

第 4 章

言語の変化

4.1 文字数と文字種

問 56　一般的に「文字種」というと何を想像するか、を隣の人と話し合いなさい。

問 57　漢字は何文字あるかを確認し、よく使われる漢字はどれぐらいの数であるかを話しあいなさい。

問 58　大体何文字ぐらい漢字（漢字熟語群）を知っていれば、本が読めるか、想像しなさい。

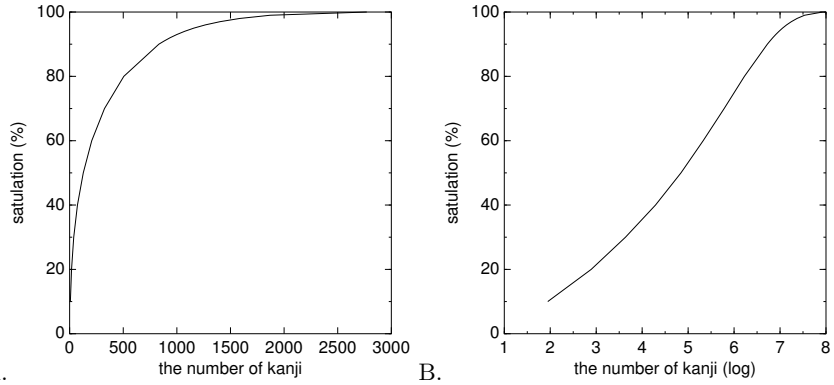

図 4.1: 日本の小説 10,000 行のうちに見られた漢字数と飽和率: 新潮文庫の 100 冊のうち、先頭から 10,000 行を取り、漢字のテキストに占める割合を計算した。B. は漢字数を対数変換したもの。

Listing 4.1: 漢字文字種のカウント

```
% cat bunko.db | head -10000 | ksf| tail -19 |awk '{print $2,$1}' | graph -TX --x-label "the␣number␣of␣kanji" --y-label "satulation␣(%)"
```

4.2　昔の日本語と今の日本語

問 59　「すばらしい」はどんな意味であるか。辞書を調べて、今の意味と同じかどうかを確認しなさい。

問 60　「網棚」「筆入れ」「下駄箱」と同じような例を探しなさい。

問 61　自分の知っているちょっと前の日本語と今の日本語の違いを見つけなさい。

問 62　「ら抜きことば」以外の日本語の変化の例を見つけ、説明せよ。

4.3　流行語とは何か

4.3.1　2017 年の流行語

問 63　流行語をあげてみなさい。

4.3.2　2012 年度ユーキャン流行語大賞

表 4.1: 2012 年度ユーキャン流行語大賞

	流行語: 詳細
1.	ワイルドだろぉ: お笑い芸人　スギちゃん　年間大賞
2.	ｉＰＳ細胞: 京都大学ｉＰＳ細胞研究所所長　山中伸弥さん
3.	維新: 日本維新の会代表代行　橋下徹さん
4.	ＬＣＣ: Peach Aviation 株式会社／エアアジア・ジャパン, ジェットスター・ジャパン
5.	終活: 週刊朝日編集部／流通ジャーナリスト　故・金子哲雄さん
6.	第３極: 第３極の皆さま
7.	近いうちに・・・: 内閣総理大臣　野田佳彦さん
8.	手ぶらで帰らせるわけにはいかない: ロンドン五輪競泳日本代表　松田丈志さん
9.	東京ソラマチ: 東武タウンソラマチ株式会社
10.	爆弾低気圧: 株式会社ウェザーニューズ

問 64　上記の流行語を覚えているものがいくつかあるか、を話しあいなさい。

問 65　なぜその語が流行したかを話しあい、流行語成立と衰退の一般的な過程をまとめなさい。

問 66　流行語は教科書の何章に関係するかを話しあいなさい。

4.4　敬語の機能

問 67　敬語の機能についてビデオを見て考えなさい。

問 68　敬語はだれに使うかを話しあいなさい。

問 69　上記、以外の人に敬語を使ったことがあるかどうかを話しあいなさい。

問 70　次の文を読んで敬語の機能について考えなさい。

敬語の機能

社会言語学の考えかたによれば、上品で礼儀正しいことばづかいをする理由は、話し手自身のフェイス (face /顔/面子/体面) を保つとともに、聞き手のフェイスをも尊重する態度を表明することにあるとされている。相手を汚くののしれば、相手の顔をつぶして気が済むかわりに、自分自身の人格 評価をも危険にさらす結果になる。相手が偉いから敬意を表明したり謙ったりするという説明原理よりも、このほうがずっと説得力がある (Brown=Levinson)。(小松 2013: 150): 小松の引用 (Brown=Levinson) は Brown and Levinson (1987) のこと。

問 71　自分が考えていた敬語と同じ考えであったかどうかを話しあいなさい。

問 72　敬語は教科書ではどのような用語で取り扱われているかを探し、話しあいなさい。

第 5 章

使用域と文体

5.1 文体論

問 73　メールの文体と論文の文体について比較せよ。

問 74　メールの文体はどのような文体なのかを考えなさい。

問 75　論文の文体はどのような文体なのかを考えなさい。

問 76　文体はなぜ変わるのかを考えなさい。

問 77　なぜそのような違いが出ているのかを考えなさい。

5.2 使用域

問 78　「話しことば」と「書きことば」は同じかどうかを考えなさい。

問 79　言語の使用域は何によって変わるのかを考えなさい。

5.3 内容語と機能語

問 80　内容語と機能語はどのようなものかを話しあいなさい。

問 81　思い出すことができるのは内容語か、機能語かを考えよ。

問 82　内容語と機能語はコンピュータでどのように分類できるかを考えよ。

問 83　$tf \cdot idf$ について考えよ。

$$tf \cdot idf(t) = tf \cdot (1 + \log \frac{N}{df(t)})$$

ただし、N は全文書数、$df(t)$ は単語 t が見られる文書の数、tf は単語 t の出てきた回数である。

問 84　単語頻度（term frequency）と文書頻度（document frequency）の違いを述べよ。

問 85　文書頻度（document frequency）idf の特徴について考えよ。

Listing 5.1: DOT 言語で記述された「春、声、鶯、鳴く」の関係

```
graph G {
graph[fontname="mincho"]
graph[center=1]
node[height=0, width=0,style="setlinewidth(.4)"]
edge[style="setlinewidth(.4)"]
edge[fontname="mincho"]
"春" [shape=ellipse fontsize=15];
"声" [shape=ellipse fontsize=12];
"鶯" [shape=ellipse fontsize=26];
"鳴く" [shape=ellipse fontsize=20];
"春" -- "鶯" [label=121 fontsize=8];
"春" -- "鳴く" [label=87 fontsize=8];
"声" -- "鶯" [label=69 fontsize=8];
"声" -- "鳴く" [label=55 fontsize=8];
"鶯" -- "鳴く" [label=190 fontsize=8];
}
```

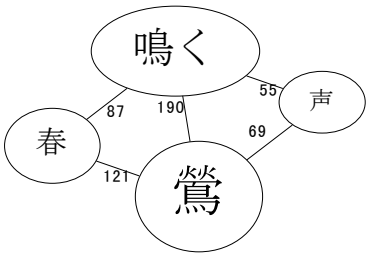

図 5.1: DOT 言語の例にしたがったグラフ出力: このグラフは４つのノードと５つのエッジを含む。それぞれのエッジに付した数字は共出現パターンの頻度を示す。それぞれのノードのフォントの大きさは当該データにおけるその語の頻度を相対的に示す。

5.4　ピジンとクレオール

問 86　ピジンとクレオールの違いについて、話し合いなさい。

問 87　混声言語の例を見て、ピジンとクレオールについて話し合いなさい。

問 88　どんな単語が聞こえたか。聞こえた単語に共通することは何か。

問 89　なぜピジンの現象が起こるのかについて、話し合いなさい。

問 90　脱クレオールとは何かを、話し合いなさい。

第 6 章

言語能力と言語運用

6.1　共出現パターン

共出現パターン（あるいは共起パターン: co-occurrence pattern）は、ある表現域に共に出現した任意 2 語の組み合わせのことである。

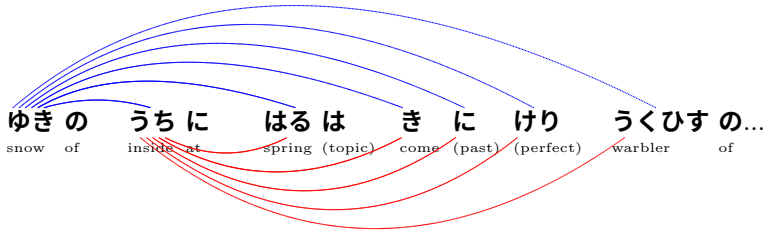

図 6.1: 共出現パターンの作り方

問 91　共出現パターンの重みを計算するとどのような分布となるかを考えなさい。

6.2　プロトタイプ理論

問 92　鳥のイメージを隣の人と交換しなさい。

問 93　教科書の図（p. 125）を見て、自分の考えていたことと同じかどうかを話し合いなさい。

問 94　この理論から考えて、人間の言語の使い方と対象のとらえ方について、話し合いなさい。

6.3 発表

来週の方法論を予習しておきます。

問 95　それぞれのテーマについて発表準備をしなさい。

問 96　ペアになり、時間 10 分、5 分ずつで互いの研究を紹介し、コメントと質問を聞いて、メモを残し、提出論文の改善に役立てなさい。

問 97　発表を効果的に行うにはどうすればよいかを考えよ。

6.4 試験

来週提出する試験問題の提出方法を予習しておきます。

問 98　まず、試しに試験問題を 1 つ作ってみなさい。試験問題には問題文、参照ページ、模範解答の 3 種のデータセットを 1 問とする。

問 99　試験問題をペアで確認し、良い答案としてどう書けばよいかを考えなさい。

問 100　問題文を読み、参照ページを見ると模範解答と同じと判断できる回答が可能かを確認しなさい。

問 101　模範解答には言語現象の説明と例が書かれていなければならない。例が書ける問題かどうかをペアで検討しなさい。

問 102　回答可能な問題のみが提出できる。

問 103　試験問題を作成し、フォームにて送りなさい。締め切りは授業中に発表する。

1. 試験問題提出 URL はいつものページに設置される。いつ設置されるかは、授業中に担当教員より伝える。
2. 口述試験予約 URL はいつものページに設置される。いつ設置されるかは、授業中に担当教員より伝える。

6.5 レポート執筆

問 104　レポートを提出するときの形式について、どこに何を書けばよいかを確認しなさい。日本語 1（留学生科目）、ライティングスキル（教養特論）の単位を取得済

みの人はすでに知っているので、この問を免除する。

問 105　自分のリサーチクエスチョンについてそれぞれ意見交換しなさい。

問 106　実験は比較可能な 2 カテゴリ以上の材料からデータが得られるようになっているかを確認しなさい。

問 107　数値の多い少ないは、簡単でも良いので統計検定を行っているかを確認しなさい。

レポート提出 URL はいつものページに設置される。いつ設置されるかは、授業中に担当教員より伝える。

第 7 章

論文発表会

7.1　発表の実施

問 108　グループになり、10 分で発表・質疑応答を行いなさい。

問 109　コメントと質問を交換し、メモを残し、提出論文の改善に役立てなさい。

7.2　質問・コメントの方法

問 110　コメントと質問は Web フォームに記録しなさい。

問 111　質問とコメントを書いてフォームにて送りなさい。ひとつひとつ意見を取り入れていけば、レポートが完成するように、具体的な改善作業について、コメントを書きなさい。

問 112　コメントはいつもの Web フォームから 4 桁のパスコードで閲覧できるので、内容を吟味しレポート執筆の改善に役立てなさい。

7.3　クリップボード発表会

人数が多いので、ポスターで発表することはむずかしい。そこで、クリップボードに A4（1 枚）で表現した論文概要を使って、意見交換を行う。

7.4　ページスタイル

テンプレートは Web ページに掲載してあるので、その書式を遵守すること。

第 8 章

まとめ

8.1 ディスカッション

問 113　言語学のテストに向けて。プレゼンテーションの良かった点、失敗した点について話し合え。

問 114　質問とコメントを見てどのように修正すればよいかを話しあいなさい。

問 115　レポートに盛り込むべきことを話しあいなさい。

問 116　オリジナルペーパーであるためには何に注意しなければならないかを話しあいなさい。

問 117　レポートを期日までに提出しなさい。

問 118　試験問題を検討し、どのような回答をすればよいかを検討しなさい。

問 119　テスト対策として何をすればよいかを話しあいなさい。

8.2 修正

ポスター発表で行った内容をレポートにして提出すること。

問 120　ポスターのフォントサイズをタイトル 14 ポイント、本文 11 ポイントないし、10.5 ポイントに修正し、文章をくわえ、論文の体裁にしなさい。

問 121　LaTeX で提出したい人はそれでも構わないが、2 段組 1 ページ以内にまとめなさい。図、表などを入れるときには、wrapfigure.sty を使いなさい。

問 122　図の場合には下に、表の場合には上にキャプションを必ずつけること。

問 123　文献を本文中で引用すること。引用のない参考文献リストは引用として認めない。URL は引用として認めない。書式、引用のルール違反は減点の対象とするので注意すること。

8.3　投稿

1. 論文は後日、pdf に変換して、1 ページで Web サイトから投稿すること。
2. 投稿の内容はパスコードでクラスメートの pdf が閲覧できる。
3. 投稿の締め切りは授業中に述べる。締め切り厳守。

第 9 章

口述試験

9.1 概要

1. 試験時間の予約は Web にて。
2. 2018 年 2 月 6, 7, 8 日の 3 日間、9:00 から 19:00 まで（途中、会議等で中断）。
3. 個別（学生 1 人）につき、およそ 10 分間実施。
4. 公正さを確保するために、試験中の会話はすべて録音。

9.2 目的

指定の教科書は言語の一般的な特徴を考える上で非常によくまとまったリソースブックであり、独習が可能な内容となっている。ただ読むだけでなく、説明ができることが望ましい。試験問題作成の目的は、本講座のまとめとして、試験問題を作成することにより、教科書の理解を深めることである。

9.3 試験問題

試験問題は学生が作成する。学生 1 人が 3 問作成する。作成した試験問題は指定した Web にアップロードする。アップロードした試験問題は、公開される。試験問題の執筆は、1）問い、2）模範解答、3）参照すべき教科書のページ数、の 3 点である。

問 124　試験問題を 3 問作成しなさい。

問 125　作成した試験問題から 3 問口述試験を行う。うち、1 問は自分の作った問題を行う。

問 126　教科書の持ち込みはない。速やかに答えること。

問 127　試験問題の実施状況の録音を聞き、試験の状況を感じ取りなさい。

問 128　口述試験の練習をクラスメートと行いなさい。

9.4　合否判定

1. 合否の判定は試験時間内では行わない。

9.5　結果

1. 結果は成績発表で総合的に判定する。

第 10 章

落ち葉ひろい

プログラミングや統計に関する便利な情報

10.1 プログラミング

10.2 統計

参考文献

Brown, Penelope and Stephen. C. Levinson (1987) *Politeness: Some universals in language usage*, Cambridge: Cambridge University Press.

小松英雄 (2013) 『日本語はなぜ変化するか: 母語としての日本語の歴史』，笠間書院.

Sturges, Herbert (1926) "The choice of a class-interval", *Journal of American Statistics Association*, Vol. 21, pp. 65–66.

田中春美・田中幸子 (2015) 『よくわかる社会言語学』，やわらかアカデミズム・「わかる」シリーズ，ミネルヴァ書房.

裏表紙について

クィントゥス・ホラティウス・フラックス
Quintus Horatius Flaccus
BC.65.12.8–BC.8.11.27
古代ローマ時代の南イタリアの詩人

裏表紙の写真は日時計です。そこにはラテン語で “Carpe Diem”（カルペ・ディアム）と彫ってあります。その意味は「その日を摘め（そのひをつめ）」で、紀元前 1 世紀の古代ローマの詩人ホラティウスの詩に登場する語句です。「一日の花を摘め」、「一日を摘め」などとも訳されますが、「その日を楽しみ、精一杯いきること」と考えることができます。このことばは、映画「いまをいきる」（ロビン・ウィリアムズ主演）の中に出てきます。

言語のありさまを調べるための計算と考え方
ワークブック
2017 年 12 月 4 日 第 1 版
著者: 山元啓史
©2015, 2016, 2017 Hilofumi Yamamoto

Carpe Diem

東京工業大学
Tokyo Institute of Technology

科研費
KAKENHI

滝久雄日本語プロジェクト
滝久雄日本教育支援基金