# テキストの硬さと軟らかさの考察
## —『現代日本語書き言葉均衡コーパス』の収録書籍を対象に—

柏野 和佳子 *（国立国語研究所 言語資源研究系）
立花 幸子（国立国語研究所 コーパス開発センター）
保田 祥（国立国語研究所 コーパス開発センター）
丸山 岳彦（国立国語研究所 言語資源研究系）
奥村 学（東京工業大学 精密工学研究所）
佐藤 理史（名古屋大学 大学院工学研究科）
徳永 健伸（東京工業大学 大学院情報理工学研究科）
大塚 裕子（はこだて未来大学 メタ学習センター）
佐渡島 紗織（早稲田大学 留学センター）

# Analysis of Textual Formality and Informality:
# In the Case of the Book Samples
# in the Balanced Corpus of Contemporary Written Japanese

Wakako Kashino (Dept. Corpus Studies, NINJAL)
Sachiko Tachibana (Center for Corpus Development, NINJAL)
Sachi Yasuda (Center for Corpus Development, NINJAL)
Takehiko Maruyama (Dept. Corpus Studies, NINJAL)
Manabu Okumura (Precision and Intelligence Laboratory, Tokyo Institute of Technology)
Satoshi Sato (Graduate School of Engineering, Nagoya University)
Takenobu Tokunaga (Dept. Computer Science Graduate School of Information Science and Engineering, Tokyo Instiute of Technology)
Hiroko Otsuka (Center for Meta-Learning, Future University Hakodate)
Saori Sadoshima (Center for International Education, Waseda University)

## １．はじめに

我々は大規模なコーパスを様々な学術研究や教育に活用するためには，テキストを所望の目的で分類するための分類指標が必要だと考え，国立国語研究所の共同研究プロジェクト「テキストの多様性を捉える分類指標の策定」において，書籍テキストの多種多様な形式，内容，表現に関わる特徴を捉えるための分類指標の設計と検証とを進めている。

これまでに，テキストの内容や表現に関わる分類指標として「対象読者（難易），主観的・客観的，硬軟，丁寧さ，直接的な語り性の有無」の 5 つの分類指標を設計し，『現代日本語書き言葉均衡コーパス』（BCCWJ）[1]の図書館サブコーパスに収録される書籍テキストを対象に，人手付与を試行した（柏野・奥村 2012）。

それらのうち，「硬軟」と「丁寧さ」の分類指標の付与は，硬い印象を与えるテキスト，軟らかい印象を与えるテキスト，丁寧さを感じるテキスト，くだけた印象を与えるテキストの分類と抽出を目的とするものである。

本稿では，まず，試行したアノテーション作業の概要を述べ，「硬軟」と「丁寧さ」の分類指標の付与結果を取り上げて報告する。そして，付与結果に基づき，テキストの硬軟，丁寧さについて考察する。

* waka@ninjal.ac.jp

[1] http://www.tokuteicorpus.jp/を参照。

## ２．アノテーション作業

### ２.１　分類指標の設計

BCCWJ には，出版サブコーパス（10,117 サンプル），図書館サブコーパス（10,551 サンプル），特定目的サブコーパスの一つであるベストセラー(1,390 サンプル), あわせて約 22,000 の書籍サンプルが収録されている。それらには，NDC（日本十進分類法）によるジャンルや，C コード（日本図書コード）による販売対象，発売形態，また，著者情報，形態論情報などが付与されており，それらを利用して，半自動的に種々の観点から分類することは可能である。しかしながら，EAGLES(1996)がコーパスへ付与することが望ましいと挙げる，(A)対象読者に想定される読解レベル（難易度），(B)テキストの作成意図，(C)さまざまな文体情報の 3 種に関する情報は C コード以外には与えられておらず，それらの観点によるテキストの分類や抽出は困難である。そこで，(A)を補う「対象読者（難易）」，(B)を補う「主観的・客観的」，(C)を補う「硬軟」「丁寧さ」「直接的な語り性の有無」という，あわせて 5 つの分類指標を新たに設計した（柏野・奥村 2012）。

「(C)さまざまな文体情報」とは，EAGLES(1996)では定義が困難だと述べられている。複数のパラメータが議論されているが，標準は定まっていないと言う。しかし，たとえば学習者にとって重要な情報になり得るものであり，Joos(1961)の提案（"frozen"，"formal"，"informal"，"colloquial"，"intimate"）や，Halliday et al. (1964)の提案（"colloquial"，"polite"，"casual"，"intimate"，"deferential"）が紹介され，語レベルの文体情報（どのような文体で用いられる語であるか）は各種辞書に工夫されて記載されていると述べられている。

つまり，ここでコーパスに備えることが望ましいと議論されている「文体情報」とは，形式性，親疎性，口語性に関わる文体情報だと言える。よって，その形式性，親疎性を問うものとして「硬軟」と「丁寧さ」の指標を，口語性を問うものとして「直接的な語り性の有無」という指標を設けた。

指標の付与に際しては，「硬軟」は「硬いか軟らかいか」という選択肢にて，「丁寧さ」は「丁寧かくだけているか」という選択肢にて判断することとした。「硬い」とは，かしこまっている感じ，堅苦しい感じであり，「軟らかい」とは，かしこまっていない感じ，親しみやすい感じである。また，「丁寧」とはフォーマルな感じであり，その反対のフォーマルではない感じを「くだけている」という言葉で表すこととした。「くだけている」は「丁寧」の対義語であるという印象を持ちにくいが，「丁寧」の対義語として浮かびやすい「ぞんざい，乱暴，粗雑」といった語はネガティブな印象のみが強いため，「くだけている」を用いることとした。この時，「硬くてくだけている」というテキストは想定できなかったため，次のとおり「硬軟」と「丁寧さ」を組み合わせて同時に問う選択肢を設けた。しかしながら，「硬軟」と「丁寧さ」は関連性は強いが異なる軸として，次の付与作業段階では選択肢を分ける計画でいる。

1 とても硬くて丁寧
2 どちらかといえば硬くて丁寧
3-1 どちらかといえば軟らかくて丁寧
3-2 どちらかといえば軟らかくてくだけている
4-1 とても軟らかくて丁寧
4-2 とても軟らかくてくだけている

### ２.２　アノテーション作業の概要

アノテーション作業の概要は，次のとおりである。

- 作業目的：人手付与の作業上の問題点の検討，典型例の抽出，分類指標の検証及び基準の検討。
- 対象テキスト：BCCWJに収録されている図書館サブコーパス（10,551サンプル）よりランダムに抽出したサンプルのテキスト。（本稿作成時，合計3,324テキストへ付与済み。）
- 1テキストの範囲と長さ：コーパス収録テキストの分類指標とするため，その一部を字数を揃えて抽出することはせず，1サンプル全体を範囲とする。1テキストの平均はおよそ

3,000語。

●作業ファイル：サンプルを取得した書籍の紙面コピーの電子化ファイルを参照する。

●作業態勢：判断のゆれを検証するために1作業につき，作業者3人を確保した。同一の判定作業を3人がそれぞれ独立して行う。

●作業量：１セット約400～500の書籍テキストに対する指標付与を延べ約10日で行う。

●作業指示：付与すべき指標の種類をごく簡単な説明のみで指示。

また，作業手順は次のとおりである。

①形式による判定を行う。構造的に単純なテキストタイプ（例：章節構造）であれば細分類の対象とする[2]。

②細分類をする。「対象読者」「主観的・客観的」「硬軟」「丁寧さ」「直接的な語り性の有無」の分類指標を付与する。

### ２.３　アノテーション作業の結果

対象とした 3,324 テキストのうち，さらに細分類の対象となる「構造的に単純なテキストタイプ」と判断されたものは，2,672 テキストであった。このうち，「硬軟」「丁寧さ」の分類指標付与については，2 テキストに付与の欠落があったため，合計 2,670 テキストが付与済みのものとして得られた。付与結果の例を表 1 に示す。

表 1 「硬軟」と「丁寧さ」の付与結果例

| タイトル | 硬軟 | | |
|---|---|---|---|
| | A | B | C |
| 犬がころんだ | 4-2 とても軟らかくてくだけている | | |
| 天才の法則 | 2 どちらかといえば硬くて丁寧 | | |
| 夢のハワイ暮らしが実現できる本 | 3-1 どちらかといえば軟らかくて丁寧 | | |
| 金田一京助全集 | 1 とても硬くて丁寧 | 2 どちらかといえば硬くて丁寧 | 2 どちらかといえば硬くて丁寧 |
| 国民の文明史 | 1 とても硬くて丁寧 | 2 どちらかといえば硬くて丁寧 | 3-2 どちらかといえば軟らかくてくだけている |

表 1 にみられるように，3 人の判断が一致するもの，3 人のうち 2 人の判断が一致するもの，3 人ともに一致しないものとがあった。これら判断の一致率，カッパ係数（一致率から偶然の一致率をひいたもの），相関関数については Kashino and Okumura(2010)，柏野・奥村(2012)で報告した。本試行作業においては中～低度の一致であったが，今後のマニュアルの整備等でその一致度が改善する見通しを得ている。

また，判断の一致度をみるために，柏野・奥村(2012)では，各選択肢別に一致した人数とそのテキスト数を示した。「硬軟」と「丁寧さ」に関しては，付与済み 2,670 テキストのうち，全員一致が 387 テキスト，2 人一致が 1,383 テキスト，非一致が 900 テキストであった。各選択肢別の全員一致数，2 人一致数は表 2 のとおりである。

表 2 「硬軟」と「丁寧さ」の選択肢別一致数の内訳

| | 1.とても硬くて丁寧 | | 2.どちらかといえば硬くて丁寧 | | 3-1.どちらかといえば軟らかくて丁寧 | |
|---|---|---|---|---|---|---|
| 全員一致 | 2 | 0.1% | 250 | 9.4% | 60 | 2.2% |
| 2人一致 | 50 | 1.9% | 707 | 26.5% | 231 | 8.7% |
| 小計 | 52 | 1.9% | 957 | 35.8% | 291 | 10.9% |
| | 3-2.どちらかといえば軟らかくてくだけている | | 4-1.とても軟らかくて丁寧 | | 4-2.とても軟らかくてくだけている | |
| 全員一致 | 53 | 2.0% | 4 | 0.1% | 18 | 0.7% |
| 2人一致 | 326 | 12.2% | 34 | 1.3% | 35 | 1.3% |
| 小計 | 379 | 14.2% | 38 | 1.4% | 53 | 2.0% |

[2] 対象外とした形式が特徴的なテキスト（例：対談，Q&A 形式，図解，用語解説）については，一定量が分類されてから細分類を検討する予定でいる。

表 2 から「1 とても硬くて丁寧」，「4-1 とても軟らかくて丁寧」，「4-2 とても軟らかくてくだけている」の 3 つの選択肢において全員一致テキストが少なかったことがわかる。

### ２.４　アノテーション付与済みテキストの NDC 別特徴

「硬軟」と「丁寧さ」の分類指標を付与した 2,670 テキストの NDC 別の特徴を分析した。作業者 3 人の判断一致，不一致に関わらず，各人の選択結果 1 つを 1 点として，各テキストの分類指標の選択肢を点数化した。それを割合になおし，平均との差分を求めた。さらに，各分類指標がどちらに触れているかの尺度を次のとおり求めた。なお，判断のゆれを考慮し，選択肢別に重みづけは行わなかった。

「硬度」（選択肢 1～2 の和と 3～4 の和との差分）

「丁寧度」（選択肢 1,2,3-1,4-1 の和と 3-2,4-2 の和との差分）

この「硬度」を横軸，「丁寧度」を縦軸として NDC ごとの特徴をプロットした結果が，次の図 1 である。

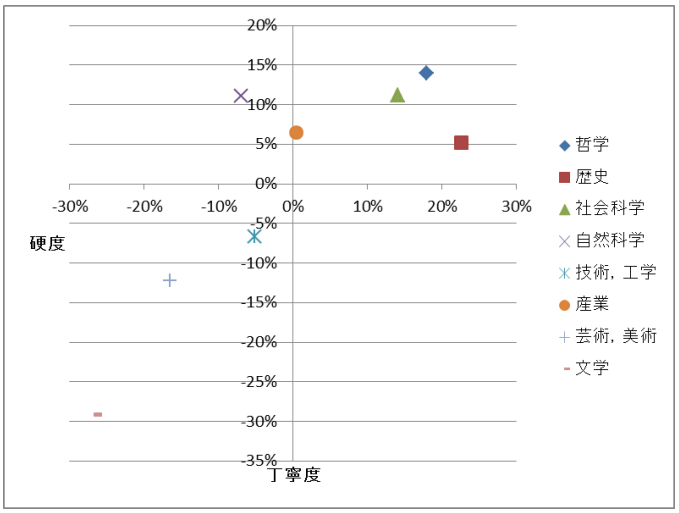



図 1「硬度」と「丁寧度」でみる NDC 別テキストの特徴

図 1 より，各 NDC 別には次の特徴のあることがわかる。

硬くて丁寧：「哲学」，「社会科学」，「歴史」
軟らかくて丁寧：「自然科学」
硬くも軟らかくもなくて丁寧：「産業」
平均的：「技術，工学」
軟らかくてくだけている：「文学」，「芸術，美術」

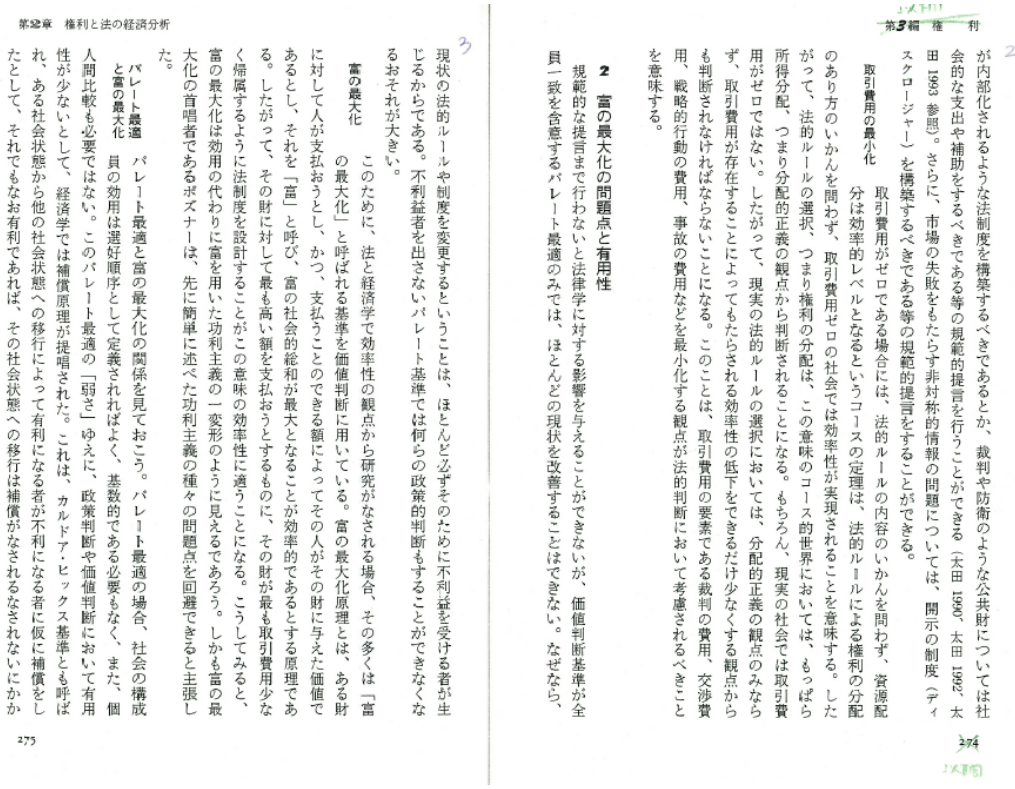
アノテーションの対象を，図書館サブコーパスに収録されたテキストとしたため，この特徴は，BCCWJ の図書館サブコーパスに収録されている書籍の場合の特徴をみていることになる。特に，「自然科学」が「軟らかくて丁寧」なテキストに位置する点が目立つ。また，「文学」だけでなく「芸術，美術」も「軟らかくてくだけている」傾向の強い点も改めて確認することができた点である。アノテーションをすることによって，コーパスに収録されているテキストの特徴分析が可能になる。

## ３．テキストの硬軟，丁寧さについての考察

### ３.１　典型例の抽出

テキストの硬軟，丁寧さについて考察するために，アノテーション結果を用いて典型例を抽出した。先の表 2 において全員一致数の少ないところで一致しているものが検討すべき典型例であることがわかった。すなわち，「1 とても硬くて丁寧」からは「硬い印象を与える典型例」として 2 例，「4-1 とても軟らかくて丁寧」からは「軟らかい印象を与える典型例」として 4 例，「4-2 とても軟らかくてくだけている」からは「くだけた印象を与える

典型例」として 18 例を抽出した。なお，「丁寧な印象を与える典型例」としたいものは「軟らかい印象を与える典型例」と重なると考え，典型例は先の 3 種に定めた。次に 1 例ずつ図示する（出典はサンプル ID と書名で記す，色鉛筆は電子化入力の際の指示）。



第2章　権利と法の経済分析

現状の法的ルールや制度を変更するということは、ほとんど必ずそのために不利益を受ける者が生じるからである。不利益者を出さないパレート基準では何らの政策的判断もすることができなくなるおそれが大きい。

**富の最大化**

このために、法と経済学で効率性の観点から研究がなされる場合、その多くは「富の最大化」と呼ばれる基準を価値判断に用いている。富の最大化原理とは、ある財に対して人が支払おうとし、かつ、支払うことのできる額によってその人がその財に与えた価値であるとし、それを「富」と呼び、富の社会的総和が最大となることが効率的であるとする原理である。したがって、その財に対して最も高い額を支払おうとするものに、その財が最も取引費用少なく帰属するように法制度を設計することがこの意味の効率性に適うことになる。こうしてみると、富の最大化は効用の代わりに富を用いた功利主義の一変形のように見えるであろう。しかも富の最大化の首唱者であるポズナーは、先に簡単に述べた功利主義の種々の問題点を回避できると主張した。

**パレート最適と富の最大化**

パレート最適と富の最大化の関係を見ておこう。パレート最適の場合、社会の構成員の効用は選好順序として定義されればよく、基数的である必要もなく、また、個人間比較も必要ではない。このパレート最適の「弱さ」ゆえに、政策判断や価値判断において有用性が少ないとして、経済学では補償原理が提唱された。これは、カルドア・ヒックス基準とも呼ばれ、ある社会状態から他の社会状態への移行によって有利になる者が不利になる者に仮に補償をしたとして、それでもなお有利であれば、その社会状態への移行は補償がなされるなされないにかかが内部化されるような法制度を構築するべきであるとか、裁判や防衛のような公共財については社会的な支出や補助をするべきである等の規範的提言を行うことができる（太田 1990、太田 1993、太田 1993 参照）。さらに、市場の失敗をもたらす非対称的情報の問題については、開示の制度（ディスクロージャー）を構築するべきである等の規範的提言をすることができる。

**取引費用の最小化**

取引費用がゼロである場合には、法的ルールの内容のいかんを問わず、資源配分は効率的レベルとなるというコースの定理は、法的ルールによる権利の分配のあり方のいかんを問わず、取引費用ゼロの社会では効率性が実現されることを意味する。したがって、法的ルールの選択、つまり権利の分配は、この意味のコース的世界においては、もっぱら所得分配、つまり分配的正義の観点から判断されることになる。もちろん、現実の社会では取引費用がゼロではない。したがって、現実の法的ルールの選択においては、分配的正義の観点のみならず、取引費用が存在することによってもたらされる効率性の低下をできるだけ少なくする観点からも判断されなければならないことになる。このことは、取引費用の要素である裁判の費用、交渉費用、戦略的行動の費用、事故の費用などを最小化する観点が法的判断において考慮されるべきことを意味する。

**2　富の最大化の問題点と有用性**

規範的な提言まで行わないと法律学に対する影響を与えることができないが、価値判断基準が全員一致を含意するパレート最適のみでは、ほとんどの現状を改善することはできない。なぜなら、

図 2　「硬い印象を与える典型例」（LBi3_00033『現代法社会学入門』）

図86　テコドント

図87　コチロサウルス

恐竜のさいご

恐竜が滅亡したわけや、恐竜たちのさいごのようすをしり、その原因をきわめるためには、恐竜の先祖のことをしらなくては、ほんとうのことがわかりません。

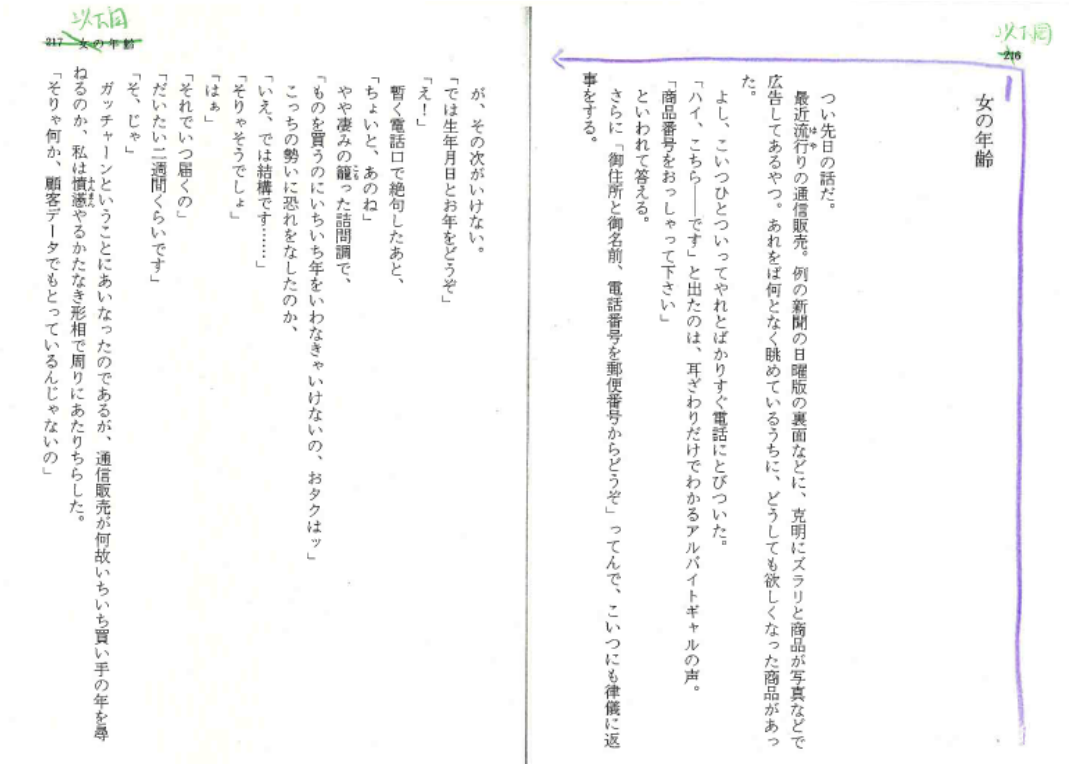
恐竜の先祖をしらべるには、ふるい時代につもった地層を、一枚一枚、したへとしらべていかなければなりません。

このようにして恐竜の先祖をたずねていくと、中生代の三畳紀のはじめにいた、「テコドント」（図86）という、からだの長さが一メートルあまりの爬虫類にいきあたります。テコドントは、四本足であるき、走るときは二本足だったことがわかっています。恐竜の先祖は、このころから四本足または二本足の動物だったわけですね。

ではつぎに、テコドントの先祖は、なんだったでしょうか。

古生代のおわりごろ（石炭紀）から中生代のはじめにかけての地層から、「コチロサウルス」（図87）という爬虫類の化石がみつかっています。コチロサウルスのなかまは、四本の足であるき、

図 3　「軟らかい印象を与える典型例」（LBa4_00010『恐竜の世界をたずねて』）



女の年齢

つい先日の話だ。
最近流行りの通信販売。例の新聞の日曜版の裏面などに、克明にズラリと商品が写真などで広告してあるやつ。あれをぼ何となく眺めているうちに、どうしても欲しくなった商品があった。
よし、こいつひとついってやれとばかりすぐ電話にとびついた。
「ハイ、こちら――です」と出たのは、耳ざわりだけでわかるアルバイトギャルの声。
「商品番号をおっしゃって下さい」
といわれて答える。
さらに「御住所と御名前、電話番号を郵便番号からどうぞ」ってんで、こいつにも律儀に返事をする。

が、その次がいけない。
「では生年月日とお年をどうぞ」
「え！」
暫く電話口で絶句したあと、
「ちょいと、あのね」
やや凄みの籠った詰問調で、
「ものを買うのにいちいち年をいわなきゃいけないの、おタクはッ」
こっちの勢いに恐れをなしたのか、
「いえ、では結構です……」
「そりゃそうでしょ」
「はぁ」
「それでいつ届くの」
「だいたい二週間くらいです」
「そ、じゃ」
ガッチャーンということにあいなったのであるが、通信販売が何故いちいち買い手の年を尋ねるのか、私は憤懣やるかたなき形相で周りにあたりちらした。
「そりゃ何か、顧客データでもとっているんじゃないの」

図 4 「くだけた印象を与える典型例」（LBf9_00067『男はオイ！女はハイ…』）

### ３.２　典型例の目視による考察

まず，「硬い印象を与える典型例」，「軟らかい印象を与える典型例」「くだけた印象を与える典型例」の各 2 例の本文を読み，目視から得られる特徴を列挙した。次のとおりである。

a.硬い印象を与える特徴
- •文末が「だ・である」の文である
- •断定，定義の文が多い
- •抽象物が主語の受身文が多い
- •接続詞・副詞が硬いものが多い（「すなわち，あるいは，ないし，一方で，にもかかわらず」，など）
- •親密度の低そうな語（学術・専門用語）が多い
- •難解な内容や説明である
- •疑問・回答が対応している

b.軟らかい印象を与える特徴
- •文末が「です・ます」の文である
- •直接的に語りかけてくるような文がある（文末が「ね，よ，でしょう，でしょうか」など）
- •接続詞・副詞に軟らかいものがある（「ですから」など）
- •平易な語がほとんどである
- •平易な内容や説明である

c.くだけた印象を与える特徴
- •体言止めや述語省略がある
- •一人称が主語の文が多い
- •平易な語に加え，俗語がある

•音変化（拗音化，撥音化など）の語がある
•オノマトペが多い
•感覚や感情表現が多い
•卑近な内容や説明である
•回答のない，いいっぱなしの疑問文がある

上記の特徴のうち，目視にて計測した結果を表 3 に示す。特徴として数が多かったところを太字にして表す。

表 3 「硬軟」と「丁寧さ」の特徴の計測

| 類型 | | 硬い | | 軟らかい | | くだけた | |
|---|---|---|---|---|---|---|---|
| サンプルID | | LBi3_00033 | LBr3_00018 | LBa4_00010 | LBo3_00103 | LBg5_00016 | LBf9_00067 |
| 書籍名 | | 現代法社会学入門 | 国民の天皇 | 恐竜の世界をたずねて | ストレスから子どもを守る本 | パパはごきげんななめ | 男はオイ！女はハイ… |
| NDC | | 321 | 313 | 457 | 379 | 599 | 914 |
| 対象とした文数 | | 96 | 140 | 47 | 29 | 183 | 45 |
| 文末表現 | 断定 | **34** | **24** | 3 | 2 | 18 | 7 |
| | 定義 | **8** | **1** | | | | |
| | です・ます | | | **47** | **29** | **66** | |
| | 語りかけ | | | **7** | **4** | | |
| | 体言止め・述語省略 | | | | | **3** | **12** |
| 文 | 受身（される・された） | **7** | **9** | 4 | | 2 | |
| | 一人称主語 | | 1 | | | **30** | **4** |
| 語 | 俗語 | | | | | **39** | **6** |
| | 音変化の文字再現・音表記 | | | | | **89** | **9** |
| | 感情（うれしい・かなしい） | | | | | **5** | **2** |

### ３.３　典型例の計量的考察に向けて

BCCWJ に収録されるテキストの文体を計量的に考察する試みはすでにいくつか行われている（小磯ほか 2008，間淵ほか 2010，小磯ほか 2011）。我々も，今回，抽出した典型例を用いて，計量的考察に着手したところである（保田ほか 2012）。

テキスト数は不揃いであるが，「硬い印象を与える典型例」2 例，「軟らかい印象を与える典型例」4 例，「くだけた印象を与える典型例」として 18 例と，比較用に，アノテーションの作業セット 1 つ分の 463 テキストを形態素解析し，おおよその比較を試みた。その結果，語種については，予測通り，「硬い」ものは漢語率が高く，「軟らかい」「くだけた」ものは和語率が高いことが確認できた。平均との差分を図示したものを図 5 に示す。

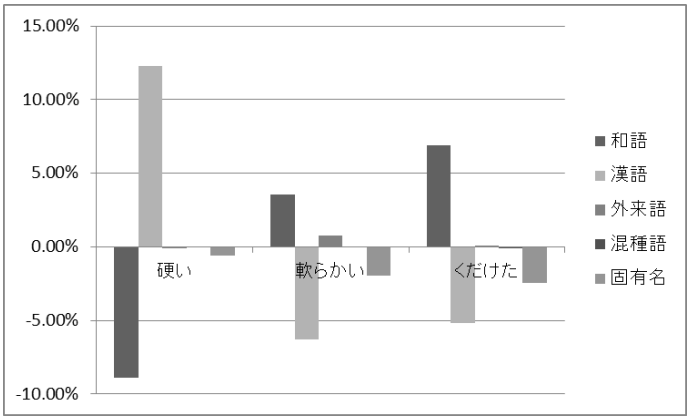



図 5 テキストタイプ別語種の比率の違い

そのほか，形態素解析結果より，品詞に関しては終助詞の比率が「軟らかい」「くだけた」もので高いこと，感動詞は「硬い」ものにはほとんど出現せず，特に「くだけた」もので

比率が高いことがわかった。

BCCWJ には形態論情報だけでなく，間淵ほか(2010)が利用した文書構造情報も付与されており，それらを用いた計量的考察を進めることは今後の課題の一つである。

## ４．まとめ

BCCWJ に収録する書籍コーパスの有効活用を可能とするための分類指標の人手付与作業の概要を報告した。分類指標のうち，「硬軟」と「丁寧さ」を取り上げ，その付与作業の結果から得られた対象テキストの NDC 別の特徴，典型例の特徴を述べた。

今後，抽出できた典型例の分析を進め，人手及び機械処理で付与する分類指標の正確さの向上を目指す。そして，少なくとも BCCWJ の図書館サブコーパスに収録される 10,551 サンプルの全てに分類指標を付与し，コーパスの研究や教育の利用価値を高めることを目指す。

さらに文体的な特徴を支える言語表現の分析を進め，辞書記述への応用を考えている。

## 謝　辞

本研究は，国立国語研究所の共同研究プロジェクト「テキストの多様性を捉える分類指標の策定」に基づくものです。また，BCCWJ の構築は，文部科学省科学研究費補助金特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築：21 世紀の日本語研究の基盤整備」（平成 18～22 年度，領域代表者：前川喜久雄）による補助を得たものです。

## 文　献

EAGLES. 1996. EAGLES Preliminary recommendation on Text Typology, *EAGLES Document EAG–TCWG–TTYP/P*, Version of Jun 1996. (http://www.ilc.cnr.it/EAGLES/texttyp/texttyp.html)

Halliday, M.A.K., A, McIntosh and P, Strevens. 1964 *The linguistic sciences and language teaching*. London: Longman.

Joos, M. 1961. *The five clocks*. New York: Harcourt Brace.

Wakako Kashino and Manabu Okumura(2010),An Approach toward Register Classification of Book Samples in the Balanced Corpus of Contemporary Written Japanese, *Proc. of PACLIC24*, pp.433-438.

柏野和佳子，奥村学(2012 予定)「書籍テキストへの分類指標人手付与の試み―『現代日本語書き言葉均衡コーパス』の収録書籍を対象に―」『言語処理学会第 18 回年次大会予稿集』B5-6.

小磯花絵，小木曽智信，小椋秀樹，冨士池優美，宮内佐夜香(2008)「『現代日本語書き言葉均衡コーパス』にもとづくジャンル間の文体差に関わる要因の分析」『社会言語科学会第 22 回研究大会発表論文集』pp.192-195.

小磯花絵，田中弥生，小木曽智信，近藤明日子(2011)「評定実験に基づくテキスト分類尺度の体系化の試み」『現代日本語書き言葉均衡コーパス』完成記念講演会予稿集，pp.47-52.

間淵洋子，柏野和佳子，山口昌也，高田智和(2010)「コーパスを用いたテキスト分類指標の検討―BCCWJ の文書構造情報分析を中心に―」『言語処理学会第 16 回年次大会予稿集』PA1-11.

保田祥，柏野和佳子，立花幸子，丸山岳彦(2012)「「語り性」を有する書きことばの典型例の分析」本予稿集収録.