

書誌情報:

大森晃(2008):日本語ウェブページに出現するムードの収集,および拡充したムード体系の提案, 自然言語処理, Vol.15, No.1, pp. 53-79.

以下は上記論文の最終投稿原稿です.

論文

Paper

日本語ウェブページに出現するムードの収集, および拡充したムード体系の提案

Collection of moods in Japanese web pages, and a proposal of an expanded mood system

大森晃 *

Akira Ohmori *

* 〒162-8601 東京都新宿区神楽坂1-3, 東京理科大学工学部第二部経営工学科

* Dept. of Management Science, Faculty of Engineering, Tokyo University of Science, 1-3 Kagurazaka, Shinjuku-Ku, Tokyo, JAPAN

キーワード

日本語ウェブページ, 文末語, 品詞, ChaSen, NTCIR-3 WEB

Keywords

Japanese web pages, sentence-end-word, part of speech, ChaSen, NTCIR-3 WEB

Summary

Some systems of moods of Japanese sentences are presented. However, it is not definitely shown what kind of method those known mood systems are constituted with. In addition, it does not seem that they are constituted through analyzing sentences included in various Japanese web pages. Therefore, in linguistic information processing for Japanese web pages, it is very likely that those known mood systems are insufficient at a point of exhaustiveness of moods.

In this article, we describes a systematic method, in detail, with which we analyze sentences included in 11,034,409 Japanese web pages collected by NTCIR project and we collect new moods with known moods. A basic procedure of the method is as follows: (1) Divide a Japanese sentence into words with ChaSen; (2) Focus on a sentence-end-word that probably has various kinds of moods; (3) Assign a proper mood to the sentence-end-word manually. We show the collected new moods, and compare the collected moods and the other known moods in order to clarify what moods can not be collected and what moods have been already presented among the collected new moods. Based on findings from the comparison, we propose an expanded mood system such that more exhaustiveness of moods is provided.

概要

日本語文のムードについて、いくつかの体系が提示されている。しかしながら、既知のムード体系がどのような方法によって構成されたかは明確に示されてはいない。また、多種多様な日本語ウェブページに含まれるような文を対象にして、ムード体系を構成しているとは思われない。したがって、日本語ウェブページを対象にした言語情報処理において、既知のムード体系は網羅性という点で不十分である可能性が高い。本論文では、NTCIRプロジェクトによって収集された11,034,409件の日本語ウェブページに含まれる文を分析して既知のムードとともに新しいムードを収集するための系統的方法について詳述する。その方法の基本的手順は、(1)日本語文をChaSenによって単語に分割し、(2)様々な種類のムードを表出すると予想される文末語に着目し、(3)文末語に手作業でムードを割り当てる、というものである。そして、収集した新しいムードを示し、収集したムードとその他の既知ムードとの比較を行い、収集できなかったムードは何か、新しく収集したムードのうちすでに提示されているものは何か、を明らかにする。比較によって得た知見をもとに、より網羅性を高めるように、拡充したムード体系の構成を提案する。

1. まえがき

日本語文のムードについて、いくつかの体系が提示されている(益岡, 田窪 1999; 仁田 1999; 加藤, 福地 1989)¹. 益岡ら(益岡, 田窪 1999)は, 述語の活用形, 助動詞, 終助詞などの様々な文末の形式を対象にして, 「確言」, 「命令」, 「禁止」, 「許可」, 「依頼」などからなるムード体系を提示している. 仁田(仁田 1999)は, 述語を有するいわゆる述語文を中心に, 日本語のモダリティを提示している. 仁田の研究成果は益岡らによって参考にされており, 仁田が提示しているモダリティのほとんどは益岡らのムード体系に取り込まれている. 加藤ら(加藤, 福地 1989)は, 助動詞的表現(助動詞およびそれに準じる表現)に限定して, 各表現が表出するムードを提示している. 提示されているムードには, 益岡らのムード体系に属するものもあるが, 「ふさわしさ」, 「継続」など属さないものもある.

既知のムード体系がどのような方法によって構成されたかは明確に示されてはいない. また, どのようなテキスト群を分析対象にしてムード体系を構成したかが明確ではない. おそらく, 多種多様な文を分析対象にしたとは考えられるが, 多種多様な日本語ウェブページに含まれるような文を対象にして, ムード体系を構成しているとは思われない. そのため, 情報検索, 評判分析(乾, 奥村 2006), 機械翻訳などウェブページを対象にした言語情報処理がますます重要になっていくなか, 既知のムード体系は網羅性という点で不十分である可能性が高い.

本論文では, 多種多様な日本語ウェブページに含まれる文を分析して標準的な既知のムードとともに新しいムードを収集するために用いた系統的方法について詳述し, 新しいムードの収集結果を示す.

¹ 益岡ら(益岡, 田窪 1999)および加藤ら(加藤, 福地 1989)はムードという用語を用いているのに対して, 仁田(仁田 1999)はモダリティという用語を用いている. 彼らによるムードあるいはモダリティの概念規定は表面的には異なるが, 本質的には同様であると考えてよい.

また、収集したムードとその他の既知ムードとの比較を行い、収集できなかったムードは何か、新しく収集したムードのうちすでに提示されているものは何か、を明らかにする。そして、より網羅性のあるムード体系の構成について、ひとつの案を与える。

ここで、ムードの収集にあたって本論文で用いる重要な用語について説明を与えておく。文末という用語は、文終了表示記号(句点など)の直前の単語が現れる位置を意味する。文末語という用語は文末に現れる単語を意味する。POSという用語は単語の品詞を意味する。例えば、「我が家へようこそ。」という文において、文終了表示記号は「。」である。文末は下線部の位置であり、文末語は「ようこそ」であり、そのPOSは感動詞である。また、ムードの概念規定としては益岡ら(益岡, 田窪 1999)のものを採用する。彼らによれば「話し手が、文をコミュニケーションの道具として使う場合、ある特定の事態の表現だけではなく、その事態や相手に対する話し手の様々な判断・態度が同時に表現される」。この場合、事態や相手に対する話し手の判断・態度がムードである。ただし、本論文ではウェブページに記述された文を対象にすることから、文の書き手も話し手と見なすこととする。例えば、「毎日、研究室に来い。」という文は、相手に対して命令する態度を表現しており、「命令」というムードを表出している。また、「妻にはいつまでも綺麗でいて欲しい。」という文は、「妻がいつまでも綺麗である」という事態の実現を望む態度を表現しており、「願望」というムードを表出している。

以下、2節では日本語ウェブページからムードを収集する際の基本的方針について述べる。3節ではムードを収集する具体的方法を与える。4節ではムード収集において分析対象とした文末語の網羅性について議論する。5節ではムードの収集結果を示す。6節では収集したムードと既知ムードとの比較を行う。7節では、より網羅性のあるムード体系の構成について一案を示す。8節では本論文のまとめと今後の課題について述べる。

2. ムード収集の基本的方針

本節では，日本語ウェブページに含まれる文を分析して，文が表出するムードを収集するにあたっての基本的な方針について述べる．具体的には，文が表出するムードを収集する際に，どのような日本語ウェブページを利用して，どのような文を分析対象とするのか，文中のどのような単語に着目するのか，どのようなムード体系を標準として利用するのか，について述べる．

2.1 分析対象

ムードを収集するにあたって，本論文ではNTCIRプロジェクトによって収集された11,034,409件の日本語ウェブページ(以下，NTCIR日本語ウェブページと呼ぶ)から成るデータセットであるNTCIR-3 WEB²を利用する．これは，各ウェブページを，HTMLタグを取り除いたプレーンテキストとして提供している．データセットの規模から，多種多様なウェブページを網羅し，それ故に多種多様な文を網羅していると考えられる．

ムード収集のために分析対象とする文は，NTCIR日本語ウェブページに含まれる文のうち，文終了表示記号として全角の「。」，「。」，「！」，「？」，あるいは半角の「!」，「?」を有するものである．例えば，NTCIR日本語ウェブページには，図1に示すようなプレーンテキストが含まれる．このテキストでは改行は1行目，2行目および5行目にある．これらは不可視であるため，図中では(改行)として明示してある．この場合，「・・・に100Mbpsで接続。」という文字列は分析対象となる文である．しかしながら，「無停電・・・ご提供いたします」という文字列は，所望の文終了表示記号が無いことから，分析対象から除外される．

² NTCIR-3 WEBについては以下のURLを参照．

<http://research.nii.ac.jp/ntcir/permission/perm-ja.html#ntcir-3-web/>

インターネットのバックボーンである、(改行)
NSPIXP2, NSPIXP3, GlobalCrossong, JPI(改行)
Xなどに直結する MEX(メディアエクスチェンジ)に 100Mbps で接続。
無停電電源装置・室温制御・自動消火設備安定した運用環境をご提供
いたします(改行)

図 1 NTCIR 日本語ウェブページの 1 例

分析対象とする文をこのように限定するにしても、NTCIR日本語ウェブページからは多種多様な文を得られると期待でき、それ故に新しいムードを発見する可能性は高いと期待できる。この意味で、NTCIR日本語ウェブページは、日本語ウェブページに含まれる文が表出するムードを収集する対象として適したデータセットであると言える。

2.2 着目する単語

日本語文では、文末語が様々な種類のムードを表出すると予想できる。もちろん、文末以外に出現する単語がムードを表出する場合もある。しかしながら、そのような場合を扱うと、文の非常に複雑な解析が必要になる。本論文では、原則として文末語のみに着目する。文末語のみに着目してムードを研究することにも、それなりの意義があることを以下に述べる。

乾ら(乾, 内元, 村田, 井佐原 1998)は、アンケート調査における自由回答文の自動分類を試みている。自動分類の目的は回答者の意図³(賛成, 反対, 要望・提案など)を把握することであり、機械学習において自由回答文の文末表現⁴を素性として重要視している点だが、彼らの研究において特徴的である。結果として、彼らは文末表現が自由回答文の意図を分類することに貢献し得ることを示している。一方、人間との円滑なコミュニケーションを図るシステムの構

³ 本論文におけるムードに相当する。

⁴ 本論文における文末語に相当する。

築において、発話文における話し手の情緒を理解する機構を実現することは重要な問題である。横野(横野 2005)は、発話文における述語に加えて、話し手の発話内容に対する態度が表現される文末に着目した情緒推定法を提案している。彼は情緒推定のために文末表現⁵をいくつかの情緒カテゴリ⁶に分類している。そして、文末表現に着目した情緒推定法が有効であり得ることを示している。

これらの研究は、一般的な視点から捉えなおすと、文が表出するムードを分類する際に文末語が有用になり得ることを示唆している。したがって本論文は、文末語のみに着目してムードを研究するものではあるが、工学的に意義のある研究を行うための基礎になると考える。

2.3 標準とするムード体系

NTCIR日本語ウェブページに出現する文を分析して文が表出するムードを収集する際に、言い換えれば、文中の文末語にムードを割り当てる際に標準とするムード体系として、益岡ら(益岡, 田窪 1999)が提示しているムード体系を利用する。理由は、比較的うまく整理されているムード体系であると考えられるからである。彼らのムード体系は以下の通りである。各ムードについての簡単な説明と文例を付録に示す。

- (a)確言, (b)命令, (c)禁止, (d)許可, (e)依頼, (f)当為,
- (g)意志, (h)申し出, (i)勧誘, (j)願望, (k)概言, (l)説明,
- (m)比況, (n)疑問, (o)否定。

なお、これらのムードは意味的に必ずしも独立しているわけではない。益岡らによれば、例えば、「禁止」には「ある動作をしないことを命令する場合と、ある事態が生じないように努力することを命令する場合」がある。いずれの場合でも、「禁止」は「命令」のムードを含意する。

3. ムードの収集方法

⁵ 本論文における文末語に相当する。

⁶ 本論文におけるムードに相当する。

NTCIR日本語ウェブページを利用する場合、非常に多くの文を分析対象にしなければならない。ムード収集においては、文から文末語を抽出し、文末語にムードを割り当てるという作業が必要である。その際、文末語の抽出は自動化できるが、文末語にムードを割り当てる作業は人間の手作業に頼らざるを得ない。

本論文では、まず暫定的に、文末語へのムード割り当てを3人の作業員X、Y、Zによる議論と合意に基づいて行うこととし、作業の大きな流れを以下のように計画し、実行することとした。

(1)作業員XとYがムード割り当て案を作成する。作成中、新しいムードを設定する必要があるかどうか、設定するとすればどんな名称、意味が適切かなど、疑問点があれば作業員3人で議論し、合意によって疑問点を解消する。

(2)ムード割り当て案を作業員Zがレビューし、割り当てられたムードの不適切さなどの問題点があれば、作業員3人で議論し、合意によって問題点を解消する。

(3)3人の合意によってムード割り当て案を承認し、それを暫定的ムード割り当て結果とする。

また、文末語へのムード割り当て手続きは以下のように設定し、実行することとした。

(a)文末語へムードを割り当てる際には、第一に2.3節で示した益岡ら(益岡, 田窪 1999)のムード体系を標準として利用し、適当なムードを割り当てる。

(b)標準としたムードを割り当てるのは適当ではないが、文末語がなんらかのムードを表出していると考えられる場合には、独自に新しい暫定的ムードの名称と意味を設定し、それを割り当てる。

(c)新たに設定した暫定的ムードは、標準としたムードと併せて、その後のムード割り当て作業で利用する。

(d)標準としたムードあるいはすでに設定してある暫定的ムードを割り当てることが明らかに不適切であるような文末語に直面した場合には、さらに暫定的に新しいムードを設定し、それを割り当て

る。

(e)すでに設定してある暫定的ムードに類似するが、そのムードを割り当てるには若干不適切であるような文末語に直面した場合には、可能な限り、ムードの意味をそれまでの意味を包含するように文末語にあわせて変更し、必要があればムードの名称を意味にあわせて変更し、それを割り当てる。

(f)ムード名称を変更した場合には、変更前のムード名称が割り当てられている文末語に対して、変更後のムード名称を改めて割り当てる。

そして、最終的には、できるだけ妥当なムード割り当て結果を得るために、作業員X, Y, Zが承認した暫定的ムード割り当て結果を、作業員Wがレビューし、問題点がある場合にはそれを解消することとした。その実行内容の詳細については3.4節で述べる。

以上のような手作業によるムード割り当ての負荷が大きいと、不適切なムードの割り当てが起こる可能性が高くなる。したがって、文末語にムードを割り当てるという手作業の負荷をできるだけ軽減する必要がある。負荷を軽減するために作業員の数を増やし、抽出した文末語を例えばPOSごとにグループ分けし、文末語のグループごとにムード割り当て作業を分担するという案もあるが、これは適当ではない。なぜならば、割り当てるムードが異なるグループを担当する作業員の間で大きくゆれる可能性が高く、ムード割り当てが混乱し、後で調整するにしても、それが大変な作業になることが容易に予想できるからである。負荷の軽減は、ムードを収集する方法を工夫することによって行うのが適当であると考えられる。

以下3.1節で示す「基本的な方法」では、2節で述べたムード収集の基本的方針に従ってムードを収集するために一般に必要なと考えられるステップを与えている。この方法の実行途中で、POSが名詞である文末語(言い換えれば、体言止めの文)がかなり多く抽出されていることが分かった。これらの文末語へのムード割り当て作業の負荷を軽減するために実行したのが、3.2節で示す「名詞に特化した方

法」である。

一方、POSが助詞または助動詞である文末語の場合、助詞・助動詞の接続関係が複雑であることから、文末語の1語(助詞あるいは助動詞)だけに着目してムードを収集することは適切でないと考えた。また、「基本的な方法」の実行途中でPOSが助詞あるいは助動詞である文末語が非常に多く抽出されていることが分かった。このことから、POSが助詞あるいは助動詞である文末語からムードを収集する負荷を軽減するとともに、適切にムードの収集を行えるような別の系統的方法が必要であった。そのような方法として実行したのが、3.3節で示す「助詞と助動詞に特化した方法」である。

3.1 基本的な方法

まず、2節で述べたムード収集の基本的方針に従って日本語文が表出するムードを収集するために一般に必要なと考えられる以下のステップを実行した。

(1)NTCIR日本語ウェブページの各々に含まれる各文について、形態素解析システムChaSen⁷を利用して、その文を単語に分割するとともに、各単語にPOSを割り当てた。つまり、各文を2項組(単語, POS)の系列に変換した。以下、そのような2項組の系列をChaSen出力⁸と呼ぶ。

ここで、ChaSenが解析対象とする文について注意を要する。ChaSenは、テキストにおける1行(改行で区切られた文字列)を1つの文として解析する。例えば、図1に示したテキストについて言えば、ChaSenにとっての文は「インターネットのバックボーンである、」、
「NSPIX2, NSPIX3, GlobalCrossong, JPI」、
「Xなどに直結するMEX(メディアエクスチェンジ)に100Mbpsで接続。無停電電源装

⁷ 形態素解析システムであるChaSen(chasen-2.3.3)については以下のURLを参照。

<http://chasen.naist.jp/hiki/ChaSen/>

⁸ ChaSenによる解析結果を全面的に信用するわけではないが、形態素解析システムとしてある程度実用的な水準にあるものと考え、本論文ではChaSen出力に基づいたムード収集方法を採用した。

置・室温制御・自動消火設備安定した運用環境をご提供いたします」である。そのため、ChaSen出力が不可解な解析結果を含む可能性はある。しかしながら、次のステップ(2)で述べるように、我々は所望の文終了表示記号の直前にある単語(文末語)のみに着目するため、文全体として正しいChaSen出力を必ずしも必要としない。これは通常の形態素解析では好ましいことではないが、文末語のみに着目したムード収集を目的とする場合には大きな問題にはならないと考えられる。

(2)文終了表示記号として全角の「。」, 「.」, 「!」, 「?」, あるいは半角の「!」, 「?」を有する各文について、その文末語に着目した。そして、文末語と、そのPOSを2項組(文末語, POS)として収集した。結果として、2項組(文末語, POS)のバッグ(bag)を作成した。ここで、バッグという用語は要素の重複が許されるものの集まりという意味で用いている。

例えば、図1に示したテキストについて言えば、「Xなどに直結するMEX(メディアエクステンジ)に100Mbpsで接続。無停電電源装置・室温制御・自動消火設備安定した運用環境をご提供いたします」という文が、所望の文終了表示記号「。」を有しており、文末語として「接続」(POSは名詞-サ変接続)に着目することになる。この場合、人間にとっては文として認識できる「無停電電源装置・室温制御・自動消火設備安定した運用環境をご提供いたします」という文字列は、実質的には、ムード収集のための分析対象からは除外されることになる。

(3)ChaSen文法⁹で用いられる以下のような12種類のPOSに基づいて、上記で作成したバッグから12種類のサブ・バッグを構成した。

1. 名詞, 2. 助詞, 3. 助動詞, 4. 副詞, 5. 感動詞,
6. 形容詞, 7. 動詞, 8. 連体詞, 9. 接頭詞,

⁹ ChaSenのための日本語辞書IPADICで用いられている品詞体系であり以下のURLを参照。

<http://hal.yh.land.to/manual/ipadic/ipadic-ja.html#SECTop/>

10. 接続詞, 11. フィラー, 12. その他.

(4)各サブ・バッグについて, その要素である各2項組(文末語, POS)の出現頻度を計数しつつ, 当該サブ・バッグを, 3項組(文末語, POS, 頻度)を要素とする集合に変換し, さらに頻度に基づいて降順にソートし, ソート済み集合を作成した. このステップが終了するまでは, POSごとに分析対象となる文末語の数は不明であった. POSによっては, 非常に多くの文末語を分析対象としなければならない状況が生じる可能性があった. そのような状況が生じた場合, 手作業の負荷を考慮して, 頻度の小さい文末語を分析対象から容易に一括除外できるようにするために, このステップで頻度に基づくソートが必要であった¹⁰.

(5)上記で作成したソート済み集合の各々について, その要素である3項組(文末語, POS, 頻度)すべての文末語に, 可能である場合限りムードを手作業で割り当て, 4項組(文末語, POS, 頻度, ムード)を作成した. ここで, 同じ3項組(文末語, POS, 頻度)の文末語に複数のムードを割り当てることができる場合には, すべての場合について4項組(文末語, POS, 頻度, ムード)を作成した.

文末語だけを見てムード割り当てが可能であると判断できるような文末語に対してムードを割り当てる場合, 当該POSを有する文末語が文末に位置するような典型的と考えられる文例を書き手/話し手の立場になって内省によって検討し, その文末語が表出し得るムードを割り当てた. 例えば, 文末語「下され」(POSは動詞-非自立)について言えば, 相手の動作に関連して「・・・して下され。」(例えば, 「結婚して下され。」)など, ある事態に関連して「・・・であって下され。」(例えば, 「健康であって下され。」)などの文例を検討した. そして, ムードとして「依頼」と「願望」を割り当てた.

ムード割り当てに際して, 文末語に対応する元の文をNTCIR日本

¹⁰ 結果論ではあるが, ソート結果を役立てられたのはPOSを名詞とする文末語を分析対象とした場合(3.2節)である.

語ウェブページに遡って参照し、その文末語が本来表出するムードを割り当てるという方法ではなく、内省による文例検討に基づくムード割り当て法を採用したのは、作業による手作業の負荷をできるだけ軽減するためである。また、内省による文例検討に基づくムード割り当て法には、文末語が表出する本来のムードを見逃す可能性はあるが、大きなメリットもあるためである。文末語に対応して元の文をNTCIR日本語ウェブページに遡って参照すれば、その文末語が本来表出するムードを割り当てることができるが、それ以外のムードを割り当てない可能性が出てくる。例えば、前述した文末語「下され」(動詞-非自立)に対応する元の文が「結婚して下され。」といった相手に動作を頼むようなものだけである場合には、ムード「依頼」のみが割り当てられ、ムード「願望」を割り当てる機会を失う。さらに、この方法は作業による手作業の負荷を非常に大きくする。内省による文例検討に基づくムード割り当て法は、手作業の負荷を軽減するだけでなく、それを注意深く実行すれば、ムードを収集する機会を失うという問題を起こしにくく、網羅的にムードを収集する方法としては比較的良いと考えられる。

文末語のなかには、ムードの割り当てが可能であるかどうかの判断に困るものがあった。例えば「さ」(副詞-助詞類接続)、「う」(感動詞)、「よう」(形容詞-非自立)、「す」(動詞-自立)、「この」(連体詞)、「ノン」(接頭詞-名詞接続)、「ァ」(その他-間投)のような、1文字あるいは2文字からなる文末語が特にそうであった。このような文末語については、安易に切り捨てることはせず、可能な限り網羅的にムードを収集するために、元の文をNTCIR日本語ウェブページに遡って明らかにし、ムードの割り当て可否を判断し、可能な場合にはムードを割り当てた。標準としたムードを割り当てればよい場合もあれば、新たに暫定的ムードを設定し、それを割り当てるのが適当な場合もあった。

例えば、「この」(連体詞)は「なんだって、この！」のように使

われ、その暫定的ムードとして「失礼」(無礼な気持ちを表す)¹¹を割り当てた。また、「ノン」(接頭詞-名詞接続)は「正直高Lvになればなるほどモラルやマナーなんてノンノンノン。」のように使われ、その暫定的ムードとして「肯否」(ある動作や事態を肯定するか否定するかを伝える)を割り当てた。「ア」(その他-間投)は「おばちゃん、教えてあげなさいよおおお(#° ㇏°)ドルァ!!」のように使われ、怒りの叫び声に匹敵する顔文字「(#° ㇏°)ドルァ!!」の文末語となっている。そのため、暫定的ムード「叫声」(叫び声に匹敵する言葉を述べる)を割り当てるのが適当であった。その他、「う」(感動詞)と「よう」(形容詞-非自立)には暫定的ムード「強調」(伝えたいことに、強い調子を加味する)を割り当てるのが適当であった。

なお、POSを名詞とする3項組み、つまり(文末語, POS=名詞, 頻度), を要素とするソート済み集合(以下, S_n と表記する)については、ステップ5を実行しなかった。 S_n の要素数は80,200であり、手続き的に何も工夫しないで80,200個の文末語に手作業でムードを割り当てることは、非常に困難な作業と思われた。 S_n からムードを収集する別の方法については、3.2節で述べる。

また、以下の2つのサブ・バッグについてはステップ4を含めて、それ以降を実行しなかった。これらのサブ・バッグからムードを収集する別の方法については、別法が必要である理由も含めて、3.3節で述べる。

(a)POSを助詞とする2項組、つまり(文末語, POS=助詞), を要素とするサブ・バッグ(以下, B_{p1} と表記する)。

(b)POSを助動詞とする2項組、つまり(文末語, POS=助動詞), を要素とするサブ・バッグ(以下, B_{p2} と表記する)。

3.2 名詞に特化した方法

¹¹ 暫定的にムード「失礼」を割り当てたが、3.4節で分かるように、最終的にはこのムードは「非難」(過失, 欠点などを責めとがめる気持ちを表す)に変更される。

前節で述べたように、POSを名詞とする文末語(言い換えれば、体言止めの文)が非常に多く抽出された。以下に体言止めの文例を示す。

- (1)「これは私たちの仕事。」
- (2)「うふふ、から、へ。」
- (3)「(鼻。華。涙。) 花。」
- (4)「とにかく、一心不乱！」
- (5)「南無阿弥陀仏・・・南無阿弥陀仏。」

一般に、体言止めの文は話し手の主観を抑えて、ただ単に情報を伝えるだけである。この「ただ単に情報を伝えるだけ」を、本論文では「陳述」のムードとする。上記5つの文では、それぞれ「仕事」、「へ」、「花」、「一心不乱」、「南無阿弥陀仏」を文末語としており、これらは陳述のムードを表出する。なお、第1の文例は、益岡らによる「確言」のムードを表出すると考えられるかもしれない。しかしながら、「仕事です」とは断定していないので、「話し手が真であると信じていることを相手に知らせる」という「確言」のムードを表出すると考えるのは適当でない。

以上から、POSを名詞とする文末語には「陳述」のムードを割り当てることができ、そうすることが適当であると考えられる。しかしながら、「陳述」のムードを割り当てることが可能であるが、その他のムードを割り当てる方が適当である場合がある。例えば、「なんとまー、久し振り。」の文末語は「久し振り」(名詞-一般)である。「陳述」のムードを割り当てることが可能であるが、むしろムード「挨拶」(友好的な気持ちを表す)を割り当てる方が、ムードを網羅的に収集するためには適当と考える。

そこで、POSを名詞とする文末語へのムード割り当てについては、最初に以下の手続きAを実行した。

(A-1) 集合 S_n から、「陳述」のムードを割り当てるのが適当である可能性が高い文末語を有する3項組(文末語、POS、頻度)を自動的に抽出した。これを実行するにあたり、分類語彙表(国立国語研究所2004)を利用した。分類語彙表は、ある種の日本語シソーラスであり、

単語は以下の4つのカテゴリに分類されている¹²。

- (a)体の類：POSが名詞である単語を主として含む。
- (b)用の類：POSが動詞である単語を主として含む。
- (c)相の類：POSが形容詞あるいは副詞である単語を主として含む。
- (d)その他の類：POSが上記以外の単語を主として含む。

文末語が「体の類」のみに属する場合、それに「陳述」のムードを割り当てるのが適当である可能性が高いと予想できた。そこで、集合 S_n の要素と分類語彙表を照合し、「体の類」のみに属する文末語を有する3項組(文末語, POS, 頻度)を抽出した。その数は32,538個であった。

(A-2)抽出した3項組(文末語, POS, 頻度)のすべてに対して自動的に、とりあえずのムードとして「陳述」を割り当て、4項組(文末語, POS, 頻度, ムード=陳述)を作成した。

(A-3)「陳述」以外のムードを割り当てることが適当である文末語を有する4項組(文末語, POS, 頻度, ムード=陳述)を見つけ出し、適当なムードを割り当て直した。ムードを割り当て直す対象になった4項組は199個であった。このことは、「体の類」のみに属する文末語に「陳述」のムードを割り当てるのが適当である可能性が高いと予想したことが、結果としてかなりの的確であったことを意味している。

上記手続きAの実行において、最終的には32,538個の文末語に割り当てられたムード「陳述」の適否を手作業で確認する必要があり、手作業の負荷は大きかった。しかしながら、この手作業の負荷は、最初から個々の文末語に適当なムードを割り当てる手作業の負荷に比べると、かなり軽減されたものと考えられる。

続いて、「相の類」と「その他の類」に名詞として利用可能な単語が登録されていることに着目して、以下の手続きBを実行した。

(B-1)分類語彙表の「相の類」と「その他の類」から名詞として利

¹² ひとつの単語が必ずただひとつのカテゴリに分類されているわけではなく、複数のカテゴリに分類されている単語もある。

用可能な単語を手作業で選定した。

(B-2)集合 S_n の要素と選定した単語とを照合し、自動的に、当該単語を文末語として有する3項組(文末語, POS, 頻度)を抽出し、とりあえずのムードとして「陳述」を割り当てた4項組(文末語, POS, 頻度, ムード=陳述)を作成した。その数は2,945個であった。

(B-3)手続きAの(A-3)と同様のことを行った。ムードを割り当て直す対象になった4項組は402個であった。ここでも、手続きAと同様、手作業の負荷はある程度軽減されたものと考えられる。

以上のように、集合 S_n の3項組のうち、分類語彙表を利用して、その文末語に「陳述」あるいは「陳述」以外のムードを割り当てたのは35,483個であった。最後に、集合 S_n においてまだムードを割り当てていない文末語を有する3項組が44,717個残っており、これらについては、その文末語に手作業でムードを割り当てた。ただし、小さい頻度(頻度の範囲が1~5)を有する3項組は除外した。結果として、「陳述」あるいは「陳述」以外の何らかのムードを割り当てることが出来る文末語を有する250個の3項組を得た。それらについては、4項組(文末語, POS, 頻度, ムード)を作成した。内訳として、ムードを「陳述」とする11個の4項組と、ムードを「陳述」以外とする239個の4項組を作成した。

なお、「陳述」のムードを割り当てた文末語のなかに、他のムードも割り当てることができるものが明らかに幾つか存在していた。それらは、POSが名詞-非自立-一般である文末語「こと」と「事」であり、POSが名詞-一般である文末語「こと」である。POSが名詞-非自立-一般である文末語「こと」(あるいは「事」)については、「勉強をすること。」というような文が考えられる。この文における「こと」には「陳述」のムードを割り当てることができるし、「命令」のムードを割り当てることができる。また、POSが名詞-一般である文末語「こと」については、「苦しいっていう、こと!」、「すぐに行けって、こと!」のような文が考えられる。前者の文における「こと」には「陳述」のムードを、後者の文における「こと」には

「命令」のムードを割り当てることができる。このように、「陳述」のムードを表出する文末語には、「事」と「こと」のように例外的に「命令」のムードを表出し得るものもあり、それらに対しては「命令」のムードも割り当てた。

3.3 助詞と助動詞に特化した方法

3.1 節後半で言及したサブ・バッグ B_{p1} と B_{p2} に含まれる 2 項組(文末語, POS)のすべての集まりであるサブ・バッグを, B_p と表記する。本節では, このサブ・バッグ B_p からムードを収集する方法を述べる。

助詞, 助動詞の接続関係は複雑である。その複雑さは, 表 1 を使って説明できる(野田 1995)。表 1 に示した日本語表現「かけられてなかったみたいだね」(句点は省略)を構成する単語には連番を付与してある。各単語の役割は, 付与された連番に応じて, 以下に示す通りである。なお, 括弧内には ChaSen 文法による POS を示してある。

- (1)活用語幹(動詞)
- (2)受動表現(動詞)
- (3)進行表現(動詞)
- (4)否定表現(助動詞)
- (5)過去表現(助動詞)
- (6)事態に対する話し手の判断・態度(助動詞)
- (7)相手に対する話し手の判断・態度(助詞)

表1 日本語における助詞, 助動詞の系列

1	2	3	4	5	6	7
かけ	られ	て	なかっ	た	みたいだ	ね

この日本語表現の場合, 7番目の単語(つまり, 文末語)だけでなく6番目の単語もムードを表出している。ムードの収集という観点から, 6番目の単語(助動詞)と7番目の単語(助詞)に着目する必要がある。言い換えれば, サブ・バッグ B_{p1} と B_{p2} の各々から, 別々にムードを収

集することは適切でないと言える。

そのため、サブ・バッグ B_{p1} と B_{p2} からムードを収集するために、以下のような方法を採用した。

(1) サブ・バッグ B_p に含まれる2項組(文末語, POS)の各々について、その生成に利用した元の文を、NTCIR日本語ウェブページに遡って明らかにした。

(2) 上記で明らかにした元の文の各々について、ChaSen出力すなわち2項組(単語, POS)の系列を分析し、ChaSen出力から以下の条件を満たすようなサブ系列(以下、 q と表記する)を生成した。

- (a) その最後に位置する2項組の単語が、ChaSen出力における文末語(文終了表示記号の直前にある単語)に対応する。
- (b) ChaSen出力における連続した2項組から成る。
- (c) ChaSen出力において助詞または助動詞が連続して出現する部分をすべて含む。

図2に、このようなサブ系列の例を示した。サブ系列 q がサブ・バッグ B_p の要素、つまり2項組(文末語, POS)を含むことは明らかである。しかしながら、サブ系列 q に含まれる単語の系列(以下、 q -単語系列と呼ぶ)には、文末語ではない単語が含まれる場合もある。これは用語上の問題を引き起こす。本論文では便宜上、 q -単語系列を文末語と見なすことにした。

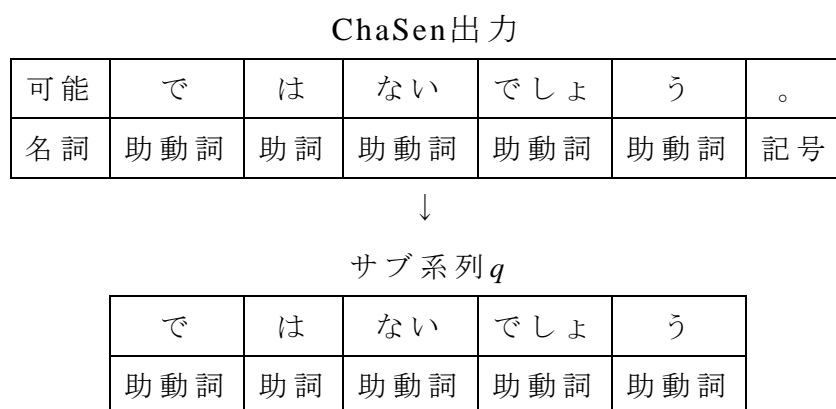


図2 サブ系列 q の例

(3)サブ系列 q を要素とするバッグを作成し、そのバッグを集合 Q に変換した。

(4)集合 Q に含まれるサブ系列 q の各々について、 q -単語系列のモードが確定できるかどうかを手作業で分析した。確定できる場合には q -単語系列にモードを割り当て、確定できない場合にはモードの割り当てを次のステップまで保留した。

(5)残ったサブ系列 q の各々について、その生成に利用した元の文を、NTCIR日本語ウェブページに遡って明らかにした。そして、その文の文脈を手作業で分析し、可能である場合に限り q -単語系列にモードを割り当てた。

3.4 暫定的モード割り当て結果のレビュー

以上の3つの方法を利用して、3人の作業者の合意に基づいて承認された暫定的モード割り当て結果を得た。最終的にできるだけ妥当なモード割り当て結果を得るために、その暫定的モード割り当て結果を、作業者Wが以下の3点についてレビューした。

(1)標準としたモードの割り当てが妥当であるか？

モード「申し出」あるいは「否定」を割り当てることができる文末語があるにも関わらず、まったくそれらのモードが割り当てられていなかった。文末語全件をチェックし、こうしたモードの割り当て漏れの問題を可能な限り解消した。この作業の過程で、標準とした他のモードの割り当てについても、その妥当性について可能な限り確認した。

(2)新しく設定した暫定的モードの名称と意味が妥当であるか？

モード自体の意味が文脈によって変わるという点で妥当とは考えられないものが、3種類設定されていた。モード名称も「～?」、「～!」、「～が」と設定されており、不適當と思われた。そこで、暫定的モード「～?」と「～!」は除外した。一方、暫定的モード「～が」の主たる意味はすでに設定されていた暫定的モード「逆接」(先行する言明について、それとの対比、それへの反対、その否定、それから予想される事態や動作の否定、あるいはその話題に関連した対

比的な評価というような、対立する何かを述べたいことを伝える)に相当するため、ムード「～が」をムード「逆接」に吸収した。

その他の暫定的ムードについては、ムードが割り当てられている文末語に照らして、その名称と意味が妥当であるかどうかを吟味した。結果として、暫定的ムード「失礼」(無礼な気持ちを表す)はその意味が理解しにくく、ムード「失礼」が割り当てられている文末語は相手あるいは事態を非難する語であることから、ムード「失礼」をムード「非難」(過失、欠点などを責めとがめる気持ちを表す)に変更した。

(3) 暫定的ムードの割り当てが妥当であるか？

この点に関するレビューは、第二点目に関するレビューと同時にを行い、暫定的ムードの割り当てが妥当であることを可能な限り確認した。また、除外した暫定的ムード「～？」と「～！」が割り当てられていた文末語のムードを再考し、適切なムードを割り当てた。さらに、暫定的ムード「～が」が割り当てられていた文末語に対してムード「逆接」を割り当て直し、暫定的ムード「失礼」が割り当てられていた文末語に対してムード「非難」を割り当て直した。

以上のようなレビューを通じて、暫定的ムード割り当て結果よりも妥当と考えられる最終的ムード割り当て結果を得た。

4. 文末語の網羅性

本論文で分析対象にした文末語は、NTCIR日本語ウェブページ(11,034,409件の日本語ウェブページ)に含まれる文のうち、文終了表示記号として全角の「。」「.」「!」「?」、あるいは半角の「!」「?」を有する文から抽出したものである。表2に、POSごとに文末語に関する3種類の数値情報を示す。各欄A, B, Cは以下を意味している。

A: 分析対象とした文末語の延べ数。

B: 分析対象とした文末語の異なり数¹³。

¹³ 字面が同じ文末語であっても、POSが異なれば異なる文末語として計数している。

C: ムードが割り当てられた文末語の異なり数.

例えば動詞について言えば, 延べ数で22,314,847個の文末語を抽出し, そのうち相異なる文末語が17,881個あり, 最終的に何らかのムードを割り当てることができた文末語は2,820個であるということが読み取れる.

表2 文末語の数

	A	B	C
動詞	22,314,847	17,881	2,820
副詞	504,512	2,041	1,094
感動詞	798,140	246	232
形容詞	1,717,386	2,367	325
連体詞	51,383	52	13
接続詞	53,256	137	95
接頭詞	35,618	109	7
フィラー	249	2	2
名詞	17,158,212	80,200	35,733
助詞・助動詞	78,240,368	61,538	968
その他	11,399	2	2
合計	120,885,370	164,575	41,291

表2から, 分析対象として抽出した文末語の延べ数は120,885,370個である. この数値は, 文末語の抽出対象とした文の数が120,885,370個であることを意味している. 言い換えれば, 膨大な数の文から文末語を抽出したことになる. また, 分析対象とした文末語の異なり数は164,575個であり, 極めて多くの種類の文末語を分析したことになる. さらに, ムードが割り当てられた文末語の異なり

数は41,291個であり，ムードを表出する多くの文末語を得たことになる¹⁴。

一方，NTCIR日本語ウェブページは，主として.jpドメインから広範囲に収集されたもので，我々の調査によれば批評，解説，報道，感想，Q&A，記録，商品広告，マニュアル，用語説明，案内・紹介など，様々なジャンルに属するウェブページを含んでいる．このように，NTCIR日本語ウェブページに含まれるテキストの多様性は極めて高く，極めて多様な文から文末語を抽出したと言える．

以上のことから，本論文で分析対象とした文末語について以下の4点を主張することができ，日本語文に現れる文末語を，多様性と数の両面から相当程度に網羅していると言ってよかろう．

- (1)極めて多様な文から文末語を抽出した．
- (2)膨大な数の文から文末語を抽出した．
- (3)極めて多くの種類の文末語を分析した．
- (4)ムードを表出する多くの文末語を得た．

5. ムードの収集結果

2.3節で示した標準とするムード体系に含まれない23種類の新しいムードを収集することができた．標準としたムードとともに，それらを表3に示す．「ムード」の欄では，標準としたムードには付録で用いたアルファベット記号をつけ，新しいムードには連番をつけた．「例」の欄には，当該ムードを表出する文末語の例を示した．「数」の欄には，当該ムードを表出する文末語の異なり数を示した．

なお，新しいムード「可能」を表出する文末語の異なり数は4であり，他のムードに比べてかなり小さい．この場合，「可能」を意味のあるムードとして取り上げる価値はないのではないかと，という疑問が生じるかもしれない．ムード「可能」を表出する文末語は以下の4種類である．括弧内数値は出現頻度を示す．さすがに接頭詞とし

¹⁴ 41,291個中，「陳述」のムードを表出するものが34,895個ある(表3を参照)．これは日本語ウェブページの特徴のひとつであると考えられる．残りの6,396個は「陳述」以外のムードを表出する．

て出現する頻度は小さいが、他の場合の出現頻度は大きい。文末語の種類は現段階では少ないが、「可能」を新しいムードとして取り上げる価値は十分にある。

表3 標準としたムードと新しいムード

ムード	例	数	ムード	例	数
a.確言	する	1,751	5. 非難	あほらし	361
b.命令	来い	1,112	6. 賛辞	おめでとう	126
c.禁止	な	42	7. 謝罪	ごめん	38
d.許可	よい	50	8. 感謝	ありがとう	43
e.依頼	ちょうだい	82	9. 可能	可	4
f.当為	べきだ	74	10. 歓喜	うれしい	42
g.意志	ます	1,301	11. 感心	なるほど	36
h.申し出	ますよ	11	12. 叫声	おー	68
i.勧誘	ましょう	87	13. 自聞	はてな	14
j.願望	ほしい	58	14. 擬音擬態	ぼたぼた	1,072
k.概言	らしい	132	15. 逆接	しかし	79
l.説明	たので	103	16. 順接	そこで	42
m.比況	の如し	33	17. 並立	および	10
n.疑問	何かしら	145	18. 選択	または	14
o.否定	じゃない	104	19. 添加	そのうえ	19
1. 陳述	一緒	34,895	20. 転換	ところで	10
2. 肯否	はい	139	21. 意外	愕然	37
3. 挨拶	おはよう	113	22. 強調	必ず	623
4. 残念	後の祭り	178	23. 確認	ですね	96

(1) 「可」(名詞-接尾-一般) [19,229].

(2) 「可」(名詞-一般) [14,747].

(3)「可」(接頭詞-名詞接続)¹⁵ [38].

(4)「可能」(名詞-形容動詞語幹) [155,637].

以下では，新しいムードの意味を簡潔に述べる．また，新しいムードの理解を助けるために，文末語を文例とともに示す．文例中の下線付き単語が文末語である．文末語は，本論文で実際に分析対象としたものである．

(1)陳述：ただ単に情報を伝えるだけ．

<文例1>これが私の青写真．

<文例2>信号は，あのとき赤．

<文例3>大事なのは勉強すること．

(2)肯否：ある動作や事態を肯定するか否定するかを伝える．

<文例1>それはあかん．

<文例2>サボリもありあり．

<文例3>卒業は無理．

(3)挨拶：友好的な気持ちを表す．

<文例1>メリークリスマス．

<文例2>かっちゃん，久し振り．

<文例3>みんな，おす．

(4)残念：悔しい気持ちを表す．

<文例1>失敗したとは，あいた．

<文例2>2着とは惜しい．

<文例3>まじ，凹む．

(5)非難：過失，欠点などを責めとがめる気持ちを表す．

<文例1>あいつは浅はか．

<文例2>とっちゃん，おもろない．

<文例3>やっぱり奴はみっともない．

¹⁵ 文例「登校は可．」(ただし「．」は全角)に対して，ChaSenは「可」を接頭詞-名詞接続，句点「．」を名詞-数として解析する．これは人間にとっては不可解な解析結果である．しかしながら，ChaSenを利用して，機械処理によって文末語のムードを推定しようとする場合には，このことは問題にはならない．

(6)賛辞：褒め言葉を述べる。

<文例1>めでたし，めでたし。

<文例2>合格するとは立派。

<文例3>おかんは，すごい。

(7)謝罪：罪や過ちを詫びる気持ちを表す。

<文例1>なんとも，かたじけない。

<文例2>まだ生きてます，恥ずかしながら。

<文例3>ほんとに，悪しからず。

(8)感謝：ありがたく感じる気持ちを表す。

<文例1>皆様に多謝。

<文例2>皆さんのお陰さま。

<文例3>どうも，おおきに。

(9)可能：ある動作の実行が可能であること，あるいは，ある事態の成立が可能であることを伝える。

<文例1>入ることは可能。

<文例2>入室可。

<文例3>登校は可。

(10)歓喜：喜びの気持ちを表す。

<文例1>くじやって，なんと大当たり。

<文例2>今日の気分は上々。

<文例3>採用だ，よっしゃ。

(11)感心：深く感じて心を動かされたということを伝える。

<文例1>満点をとったとは，さすが。

<文例2>合格したのか，ふうん。

<文例3>一人でやったなんて，へええ。

(12)叫声(きょうせい)：叫び声に匹敵する言葉を述べる。

<文例1>蜘蛛がいたんだ，ワーッ。

<文例2>財布落としちまった，ガーン。

<文例3>明日までにやるのか，ひゃー。

(13)自問(じぶん)：自分自身への疑いの気持ちを表したり，ある動

作を開始する時に自分に言い聞かせる言葉を述べる。

<文例1>こんな時間に腹がすくなんで、はて。

<文例2>起きる時刻だし、さて。

<文例3>よっこらしよ。

(14)擬音擬態：擬音語や擬態語を述べる。

<文例1>最近仲良しこよしでウキウキ。

<文例2>なんと仕事中にうとうと。

<文例3>とにかく部屋は、がらん。

(15)逆接：先行する言明について、それとの対比、それへの反対、その否定、それから予想される事態や動作の否定、あるいはその話題に関連した対比的な評価というような、対立する何かを述べたいことを伝える。対比的な評価の例は、「彼女は綺麗だが。」における文末語「だが」が、彼女という話題に関連して「おっちょこちよいでもある。」などの発言を予想させる場合である。

<文例1>みんな、勉強はするが。

<文例2>あいつ、バカはするまいが。

<文例3>あいつ、今夜はここにいねーずが。

(16)順接：先行する言明を前提にして何かを続けて述べたいことを伝える。

<文例1>ノートをコピるなら。

<文例2>リハビリするにつれて。

<文例3>私はもう若くはなく、したがって。

(17)並立：先行する言明に並立させて何かを述べたいことを伝える¹⁶。

<文例1>祝卒業および。 . . .

<文例2>彼はカツ丼を、一方。 . . .

<文例3>参院選圧勝ならびに。 . . .

(18)選択：明示した選択肢以外にも選択肢があることを伝える。

¹⁶ 「並立」のムードは通常の文には現れにくいだが、日記などの表題において出現する。

<文例1>生きるべきか，それとも。

<文例2>飲み物はコーヒーでもよし，何でも。

<文例3>デートの時は，映画など。

(19)添加：先行する言明に続いて，それに関連する何かを付け加えたいことを伝える。

<文例1>ドアを開けてもいいよ，ただし。

<文例2>あそこに行けば星が見える，しかも。

<文例3>君が悪いと思うんだよね，なぜなら。

(20)転換：話題を変更したいことを伝える。

<文例1>今食べたところ，そういえば。

<文例2>別れる気はないし，てか。

<文例3>それはいいとして，取りあえず。

(21)意外：期待したことが達成されなかったことに対する不満，期待していなかったことが達成されたことに対する驚き・喜びの気持ちを表す。

<文例1>君がいるとはもっけの幸い。

<文例2>今日は誕生日だっけか，あれれ。

<文例3>なんとなんと，あんどり。

(22)強調：伝えたいことに，強い調子を加味する。

<文例1>受付を通してね。

<文例2>そのうちミスるだろ。

<文例3>なんでやねん。

(23)確認：伝えたいことを相手が理解していることを確かめる。

<文例1>ビールが飲みたいもんだよ，おい。

<文例2>それはさっき言ったろ。

<文例3>学校へ行くよね。

6. 既知ムードとの比較

本節では，標準とした益岡ら(益岡，田窪 1999)によるムードおよび以上のように収集した新しいムード(以下しばしば，単に「新しいムード」と略称する)と，加藤ら(加藤，福地 1989)および仁田(仁田

1999)によって提示されている既知ムードとの比較を行う。比較によって、加藤らによるムード、および仁田によるムードのなかから、本論文では収集できなかつたムード¹⁷を明らかにする。また、新しいムードのうち、すでに提示されているものを明らかにする。

加藤らはまず、ムードを表出する助動詞的表現(助動詞およびそれに準じる表現)に限定し、それらを大きく「主観的推量」(自分自身の経験、直観あるいは他から得た情報に基づいて、話し手が推量した結果を主観的に述べるムード)と「推論ないしは背景の説明」(ある事柄の背景やその意義づけ、評価などを示し、相手に理解させることを主眼とするムード)という2つの類に分けている。そして各類に含まれる助動詞的表現の意味を説明している。我々は、加藤らに取り上げた助動詞的表現の類ごとに、各表現に対して彼らと与えている意味説明を解釈し、各表現が本論文で収集したムードのどれに相当するかを考察した。その結果、以下のような知見を得た。

(1)主観的推量：この類に含まれる助動詞的表現に関する知見を以下に示す。本論文で収集できなかつたムードはあるが、新しいムードに相当するものはない。

(1-1)「ダロウ」：これは益岡らによる「概言」の下位ムード「断定保留」に相当する。

(1-2)「ソウダ」：これは益岡らによる「概言」の下位ムード「伝聞」あるいは「様態」に相当する。

(1-3)「ヨウダ」：これは益岡らによる「比況」、あるいは「概言」の下位ムード「証拠のある推定」に相当する。

(1-4)「ラシイ」：これは益岡らによる「概言」の下位ムード「証拠のある推定」に相当する。加藤らは、この表現が「彼の態度は、いかにも教師らしい」というように「ふさわしさ」も表出すると指

¹⁷ 益岡らによるムードと新しいムードが本論文で収集したムードであり、それ以外は本論文で収集できなかつたムードである。なお、益岡らによる「概言」というムードにはいくつかの下位ムードがある。それら下位ムードは明示的に収集してはいないが、「概言」として一括して収集したものとする。

摘している。「ふさわしさ」というムードは本論文では収集できなかったものである。

(1-5)「マイ」：これは益岡らによる「意志」、あるいは「概言」の下位ムード「断定保留」に相当する。

(1-6)「カモシレナイ」：これは益岡らによる「概言」の下位ムード「可能性」に相当する。

(1-7)「ニチガイナイ」：これは益岡らによる「概言」の下位ムード「直感的確信」に相当する。

(2)推論ないしは背景の説明：この類に含まれる助動詞的表現に関する知見を以下に示す。本論文で収集できなかったムードがあり、さらに新しいムードに相当するものもある。

(2-1)「ハズダ」：これは益岡らによる「概言」の下位ムード「証拠のある推定」に相当する。

(2-2)「コトニナル」：これは益岡らによる「概言」の下位ムード「証拠のある推定」に相当する。

(2-3)「ワケダ」：これは益岡らによる「説明」に相当する。

(2-4)「モノダ」：これは益岡らによる「当為」あるいは「説明」だけでなく、新しいムード「感心」あるいは「意外」に相当する。加藤らは、例えば「子供というのは、(中略)、急に熱を出したりするものだ」のように、この表現が「ものの本来の性質・一般的性向」も表出すると指摘している。「ものの本来の性質・一般的性向」というムードは本論文では収集できなかったものである。

(2-5)「コトダ」：これは益岡らによる「当為」、あるいは新しいムード「意外」に相当する。

(2-6)「ノダ」：これは益岡らによる「説明」に相当する。

(2-7)「トコロダ」：加藤らは、継続中の動作・出来事について、直前にくる動詞の形により、この表現が「継続」、「未然」、「已然」を表出すると指摘している。これらは本論文では収集できなかったムードである。

一方、仁田は述語を有するいわゆる述語文を中心に、日本語のモ

ダリティを考察している。彼は、文は大きく質的に異なった2つの層、つまり「言表事態」(話し手が現実との関わりにおいて描き取った一片の世界、文の意味内容のうち客体的な出来事や事柄を表した部分)と「言表態度」(話し手の言表事態を巡っての把握の仕方や発話・伝達的な態度のあり方を表した部分)から成り立っているとし、モダリティを大きく「言表事態めあてのモダリティ」(発話時における話し手の言表事態に対する把握の仕方の表し分けに関わる文法表現)と「発話・伝達のモダリティ」(発話時における話し手の発話・伝達的な態度のあり方の表し分けに関わる文法表現)の2種に分けている。そして、それぞれのモダリティについて、下位のモダリティを提示している。我々は、それらのモダリティに関する記述内容を解釈し、各モダリティが本論文で収集したムードのどれに相当するかを考察した。その結果、以下のような知見を得た。

(a)発話・伝達のモダリティ：この種の下位モダリティに関する知見を以下に示す。本論文で収集できなかったムードはあるが、新しいムードに相当するものはない。

(a-1)「命令」：これは益岡らによる「命令」に相当する。

(a-2)「依頼」：これは益岡らによる「依頼」に相当する。

(a-3)「禁止」：これは益岡らによる「禁止」に相当する。

(a-4)「誘いかけ」：これは益岡らによる「勧誘」に相当する。

(a-5)「意志」：これは益岡らによる「意志」に相当する。

(a-6)「希望」：これは話し手の話し手自身あるいは聞き手への願いであり、益岡らによる「願望」(自分自身の動作・状態を望む場合、他人の動作・状態を望む場合)に相当する。

(a-7)「願望」：これも話し手の願いであり、その願いを遂行する聞き手が不在である点で希望と区別されているに過ぎず、益岡らによる「願望」(特に、「早く月が出てほしい。」のような、ある事態の成立を望む場合)に相当する。

(a-8)「現象描写」：これは、話し手の感覚を通して捉えられたある時空のもとに存在する現象を、主観を加えないで述べるものであ

る。「主観を加えない」ということから新しいモード「陳述」に相当しそうであるが、「信号が赤だ(である, です)。」のような述語文に対して、仁田はこのモードを提示している。したがって、この「現象描写」は文末語の表現形式からすれば、益岡らによる「確言」に相当すると考えるのが妥当である。

(a-9)「判定」：これは、益岡らによる「確言」と「概言」を内包し、双方とは異なるモードである。ただし、モードとして未分化であり有用とは考えられないため、このモードを、本論文で収集できなかったモードとしては取り上げないこととする。

(a-10)「疑い」：これは、益岡らによる「疑問」(特に、「自問型の疑問」)に相当する。

(a-11)「判断の問いかけ」：これは、言表事態の成立について判定を下せないため、判定を下すために必要な情報を聞き手に問いかけるもので、益岡らによる「疑問」(特に、相手に未知の部分の情報を求める「質問型の疑問」)に相当する。ここで、「未知の部分の情報」を「判定を下すために必要な情報」と具体化しているという点で、これは「質問型の疑問」に内包される下位モードであるとも考えられる。「判断の問いかけ」については、益岡らによる「疑問」の説明において具体的に言及されていないことから、本論文で収集できなかったモードであると言える。

(a-12)「情意の問いかけ」：これは、言表事態に対する聞き手の心的態度が不明であることから、それを問いかけるもので、益岡らによる「疑問」(特に、相手に未知の部分の情報を求める「質問型の疑問」)に相当する。ただし「情意の問いかけ」についても、「判断の問いかけ」と同様の理由から、本論文で収集できなかったモードであると言える。

(b)言表事態めあてのモダリティ：この種の下位モダリティとして、「意志」, 「希望」, 「願望」, 「話し手の把握・推し量り作用を表すもの」, 「推し量りの確からしさを表すもの」, 「徴候の存在のもとでの推し量りを表すもの」, 「推論の様態に関わるもの」が

提示されている。「意志」，「希望」，「願望」に関する知見についてはすでに言及した。その他の下位モダリティについては以下の通りであり，本論文で収集できなかつたムードも，新しいムードに相当するものもない。

(b-1)「話し手の把握・押し量り作用を表すもの」：これは，言表事態に対する話し手の把握・押し量り作用を表すもので，「～スル」形の「断定」，「～スルダロウ」形，「～スルマイ」形の「推量」がある。「断定」は益岡らによる「確言」に相当し，「推量」は「概言」(特に，その下位ムードである「断定保留」)に相当する。このことから，これは益岡らによる「確言」と「概言」を内包し，双方とは異なるムードである。ただし，ムードとして未分化であり有用とは考えられないため，このムードを，本論文で収集できなかつたムードとしては取り上げないこととする。

(b-2)「押し量りの確からしさを表すもの」：これは，言表事態がどれ位の確からしさもって成立するのかを表すもので，「～ニチガイナイ」形の「必然性」，「～カモシレナイ」形の「可能性」がある。「必然性」は益岡らによる「概言」の下位ムード「直感的確信」に相当し，「可能性」は「概言」の下位ムード「可能性」に相当する。いずれにしても「概言」に相当する。

(b-3)「徴候の存在のもとでの押し量りを表すもの」：これは存在する徴候から引き出された押し量りを表すもので，代表的な表現形式として「ラシイ」，「ヨウダ」，「ミタイダ」，「ソウダ」がある。これらは益岡らによる「概言」の下位ムード「証拠のある推定」に相当する。

(b-4)「推論の様態に関わるもの」：これは言表事態がある推論によって引き出されたものであることを表すもので，代表的な表現形式として「ハズダ」がある。これは益岡らによる「概言」の下位ムード「証拠のある推定」に相当する。

以上をまとめると，加藤らの「ふさわしさ」，「ものの本来の性質・一般的性向」，「継続」，「未然」，「既然」というムード，

仁田の「判断の問いかけ」，「情意の問いかけ」というムードは，本論文では収集できなかつた．一方，益岡らのムード体系を標準にして本論文で新しいムードとして収集した「感心」と「意外」は，加藤らによってすでに提示されていた．この結果，本論文で新しいムードとして収集した23種類のムードのうち，本当の意味で新しいと言えるムードは「感心」と「意外」を除く21種類であることが分かった．

7. ムード体系の拡充

本節では，より網羅性のあるムード体系の構成について，ひとつの案を示す．それは，益岡ら(益岡，田窪 1999)のムード体系を，本論文で収集した新しいムードと，加藤ら(加藤，福地 1989)のムード，仁田(仁田 1999)のムードのうち本論文では収集できなかつたムードを加えて拡充したものである．拡充したムード体系を表4に示す．

表4 拡充したムード体系

益岡ら (15種類)	加藤ら (7種類)	新しいムード (21種類)
確言，命令，禁止，許可，依頼，当為，意志，申し出，勧誘，願望，概言，説明，比況，疑問，否定	感心，意外，ふさわしさ，ものの本来の性質・一般的性向，継続，未然，既然	陳述，肯否，挨拶，残念，非難，賛辞，謝罪，感謝，可能，歓喜，叫声，自聞，擬音擬態，逆接，順接，並立，選択，添加，転換，強調，確認

ここで，本論文で収集できなかつた加藤らの「ふさわしさ」，「ものの本来の性質・一般的性向」，「継続」，「未然」，「既然」というムードは，本論文で収集したムードと比べて明らかに異質であることから，拡充したムード体系に採用した．一方，本論文で収集できなかつた仁田の「判断の問いかけ」，「情意の問いかけ」というムードは，益岡らによるムード「疑問」の下位ムードとして位置づけられる．これらについては，拡充したムード体系に明示的に採

用するよりも、益岡らによるムード「概言」の説明と同様に、ムード「疑問」の説明に下位ムードとして明記する方が適切と考えた。

結果として、43種類のムードからなる拡充したムード体系を構成した。言うまでもなく、この拡充したムード体系は、既知のムードだけでなく、多種多様な日本語ウェブページを網羅し、それ故に多種多様な文を網羅していると考えられるNTCIR日本語ウェブページから収集した本当の意味での新しいムードを21種類含んでいる。それらは全体のほぼ半分を占める。そのため、拡充したムード体系は日本語ウェブページをも対象にした今後の言語情報処理の基礎として役立つものと考えられる。

8. むすび

本論文では、極めて多様性の高いNTCIR日本語ウェブページに含まれる多数の日本語文から、ムードを収集する方法を詳説した。収集方法の基本的な手順は、(1)日本語文をChaSenによって単語に分割し、(2)様々な種類のムードを表出すると予想される文末語に着目し、(3)益岡ら(益岡, 田窪 1999)のムード体系を標準として利用し、文末語に手作業でムードを割り当てる、というものである。20種類程度の新しいムードを収集することができたという点で、その方法は効果的であったと考えられる。

また、収集したムードと、加藤ら(加藤, 福地 1989)および仁田(仁田 1999)によって提示されている既知ムードとの比較を行った。これによって、NTCIR日本語ウェブページからは収集できなかった既知ムードがいくつか存在していること、新しく収集した23種類のムードのうち2種類は既知ムードとして存在していること、を明らかにした。そして、比較によって得た知見をもとに、より網羅性のあるムード体系の構成について一案を示した。それは、既知ムードと新しいムードとを併せて43種類のムードから構成したものである。約半分が、NTCIR日本語ウェブページから収集した既知ではないムードであり、日本語ウェブページをも対象にした今後の言語情報処理の基礎として役立つものと考えられる。

本論文は，日本語ムード辞書を整備するための基礎を与えているとも言える．工学的に望まれる日本語ムード辞書にはいろいろな様式が考えられる．もっとも簡潔な辞書は，実際の日本語ウェブページに出現する文末語とそのPOS，それが表出し得るムードを1つのレコードとしてエントリしたものである．これにより，分析対象とするテキスト内の各文から文末語とそのPOSを抽出し，辞書を参照して，その文末語が表出し得るムードを分析できる．あるいは，所望のムードを表出し得る文末語を有する文の抽出も可能であろう．このような辞書を整備するための基礎として，できるだけ網羅性の高いムード体系が必須である．7節で提案したムード体系はそうしたものとして役立つと考えられる．そのようなムード体系を利用して，なんらかの日本語ムード辞書を整備していくことは今後の課題である．

謝辞

NTCIR-3 WEBは国立情報学研究所の許諾を得て使用させていただきました．この場を借りて深謝いたします．また，ムード収集方法について有益な助言をして頂いた土井晃一工学博士に感謝します．さらに本論文の完成度を高めるために非常に参考となるコメントを，査読者から多く頂きました．この場を借りてお礼申し上げます．

参考文献

- 乾孝司，奥村学 (2006)．“テキストを対象とした評価情報の分析に関する研究動向．” 自然言語処理，13(3)，pp.201-241．
- 乾裕子，内元清貴，村田真樹，井佐原均 (1998)．“文末表現に着目した自由回答アンケートの分類．” 情報処理学会自然言語処理研究会報告，98(99)，pp.181-188．
- 加藤泰彦，福地務 (1989)．テンス・アスペクト・ムード．荒竹出版．
- 国立国語研究所 (2004)．分類語彙表増補改訂版ーデータベース．国立国語研究所．
- 仁田義雄 (1999)．日本語のモダリティと人称．ひつじ書房．
- 野田尚史 (1995)．“文の階層構造からみた主題ととりたて．” 益

岡隆志，野田尚史，沼田善子(編)「日本語の主題と取り立て」，
pp.1-35，くろしお出版．

益岡隆志，田窪行則 (1999)．基礎日本語文法．くろしお出版．

横野光 (2005)．“情緒推定のための発話文の文末表現の分類．” 情報処理学会自然言語処理研究会報告，2005(117)，pp.1-6．

付録：益岡らのムード体系

益岡ら(益岡，田窪 1999)のムード体系について，以下に簡単な説明を示すとともに，括弧内に文例を示す．文例は，概ね彼らが用いているものを参考にした．

- (a) **確言**：話し手が真であると信じていることを相手に知らせたり，同意を求めたりする．(変な音がする．)
- (b) **命令**：相手が意志的に制御できる動作を，相手に強制する．(早く勉強すること．)
- (c) **禁止**：ある動作をしないこと，ある事態が生じないように努力することを命令する．(こっちに來るな．)
- (d) **許可**：ある動作が他の動作と同じく容認可能であることを相手に指摘する．(食べてもいいよ．)
- (e) **依頼**：相手の意志を尊重して，ある動作をするよう頼む．(水をまいておいてちょうだい．)
- (f) **当為**：ある事態が望ましいとか，必要だ，というように事態の当否を述べる．(貿易黒字を減らすべきだ．)
- (g) **意志**：ある動作を行う意志を表す．(先に行きます．)
- (h) **申し出**：相手に対する自分の動作を申し出る．(荷物を持ちましょう．)
- (i) **勧誘**：共同動作の申し出を表す．(出かけましょう．)
- (j) **願望**：事態の実現を望んでいることを表す．このムードには，自分自身の動作・状態を望む場合，他人の動作・状態を望む場合，ある事態の成立を望む場合がある．(宇宙飛行士になりたい．)
- (k) **概言**：真とは断定できない知識を述べる．概言は下位ムードとして，「断定保留」(「だろう，まい」)，「証拠のある推定」(「らし

い, ようだ, みたいだ, はずだ」), 「可能性」(「かもしれない」), 「直感的確信」(「にちがいない」), 「様態」(「そうだ」), 「伝聞」(「そうだ, という, とのことだ」)を内包するが¹⁸, 本論文ではこれらを一括して「概言」としている。(来年はきっと不景気になるだろう.)

(l)説明: ある事態の説明として, 別の事態を述べる。(「遅かったじゃないですか。」 渋滞に巻き込まれたんです.)

(m)比況: ある事態を性質の類似した別の事態で特徴づける。(この絵は写実的で, 写真のようだ.)

(n)疑問: 話し手が相手に未知の部分の情報を求めたり(質問型の疑問), 自分自身に問いかけたりする(自問型の疑問). 本論文では, 質問型も自問型も一括して「疑問」としている。(きのう, 誰に会ったのですか.)

(o)否定: 対応する肯定の事態や判断が成り立たないことを意味する。(雨が降らなかった.)

¹⁸ 「断定保留」など下位ムード名称の後に, 括弧つきで代表的な表現形式を例示した.