

シソーラスを組み込んだ意味解析システム

国分 芳宏[†]・梅北 浩二[†]・松下 栄一[†]・末岡 隆史[†]

CGM(消費者生成メディア)が普及してきたため,そのための言語処理技術が必要になってきた.このような文章データの自然文による検索や翻訳のために,解析精度の向上が求められている.解析誤りの発生原因である,用語の異なり,構文構造の異なりに対処できる処理方式を実現する.この両者への対策として,シソーラスを用いて用語間の意味的な距離を決定する方式を提案する.具体的には,用語の標準化や係り受けの正規化をするシステムを実現し,さらに,付属語を調べて,省略された主語を復元すること,「文節意図」を付与することを試みた.「Yahoo!知恵袋」のデータを用いて解析実験をした結果,シソーラスを用いない場合に比較して約1%の精度の向上がみられた.システムが用いている辞書の内容について概要を述べる.

キーワード: 意味解析, シソーラス, 意味的な距離, 用語の標準化, 文節意図, 良しあし

A Semantic Analysis System of Japanese Assisted by a Thesaurus.

YOSHIHIRO KOKUBU[†], KOUJI UMEKITA[†], EIICHI MATSUSHITA[†] and TAKASHI SUEOKA[†]

Because Consumer Generated Media have spread, language processing technologies for that purpose are necessary. The improvement of parsing precision is demanded for both retrieval by a natural sentence and translation of such text data. We realize processing methods which can deal with analysis errors caused by fluctuating terms and ambiguous sentence structures. Specifically, we propose using a thesaurus to decide semantic distance between the terms. We have realized a system which standardizes the terms and normalizes the syntactic dependencies. Further, we examine the internal structure of predicates to recover omitted subjects and determine the "intention of a predicate". When we analyze texts of "Yahoo! Chiebukuro", the precision improves by about 1% compared with when the thesaurus is not used. We summarize the contents of the dictionaries our system uses.

Key Words: semantic analyzer of Japanese, Thesaurus, Semantic Distance, Term Standardization, intention of a predicate, Value Judgment of Proposition

1 はじめに

自然文検索や翻訳,レコメンデーションなどに使用可能な解析システムを実現した.

2000年に(南 1974; 白井 1995)を参考にして文節に強さを決めて,同じ強さの文節では,連用

[†]株式会社言語工学研究所, Institute of Language Engineering

修飾格は直後の用言に，連体修飾格は直後の体言に係るという規則を用いて構文解析プログラムを開発した．しかし実際の構文構造は，文節を飛び越して係る場合が見受けられた．文法的な情報だけでは不十分だと考え，意味的な情報の導入を検討した結果，シソーラスを組み込んで用語同士の意味的な距離を測って，その距離によって係り先を決定する手法を開発した．

この解析システムを自然文検索に用いる場合，同じ内容のことを言っているのにいくつかの書き方が許されていることからしばしば検索漏れが発生する．この異形式同内容に対応するため，用語の標準化，係り受けの正規化を実現した．さらに，翻訳などで使用することを考えて，文節意図(4.1で述べる)を把握しやすくするために係り受けとそれに続く付属語の並びをまとめた形で管理した．

手作業で収集した辞書に手作業でいろいろな情報を付加して機能を実現するという方式で開発した．統計的な手法は用いていない．

文末に試用サイトの URL を示したので試用していただきたい．

2 構文構造の決定

(白井 1995)では等しい階層的認識構造のあいだでは，構文構造を文節の連体修飾格は体言に，連用修飾格は用言にそれぞれ最初の文節に係るというアルゴリズムで決定していた．しかし複数の受けの候補があるときに，文法情報だけでは正しく決定されないという問題がある．(藤尾 2000)では統計的な手法で決定しているが，様々な分野に対して大量の解析済みのデータを準備するのは容易ではない．そこで筆者らはシソーラス(5.1で述べる)を用いて用語同士の意味的な距離を計算して，その距離の近いところに係るという方式を採用した．

2.1 意味的な距離によって構文構造を決定する

複数の受けの候補があるときに，図1のように意味的な近さ(2.2で述べる)で直後の文節を飛び越して係ることがある．

「ネットで」という連用修飾格は，「行く」か「調べる」という用言に係る可能性がある．これまで，直後にあるということで「行く」という文節に係るようになっていた．意味的な距離を測って比較すると次のようになる．

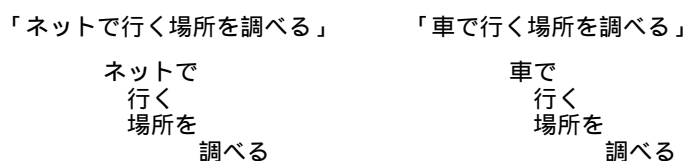


図1 連用修飾格での例

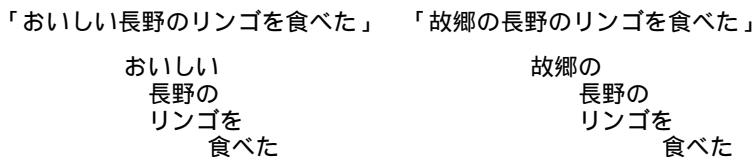


図 2 連体修飾格での例

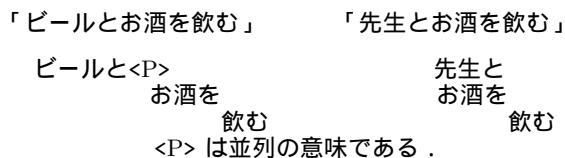


図 3 並列構造の例

意味的な距離

ネットで - 行く

ネットで - 調べる 1 (筆者らのシステムでの距離)

「ネットで」 - 「調べる」の方が意味的な距離が近いので、「ネットで」という文節は「調べる」という文節に係るようにした。

連体修飾格での例

同様に連体修飾格でも意味的な距離で係り先を決める。

図 2 では、「おいしい」という形容詞が文法情報だけで評価すると「長野の」か、「リンゴを」かどちらかの名詞の文節に係る可能性がある。同様に用語同士の意味的な距離を測って「リンゴを」という文節に係ける。

並列構造での例

(黒崎 1992) では、並列構造を付属語の類似性で決めている。筆者らのシステムではシソーラスを用いて名詞の意味的な距離で決めている。

「ビール」と「お酒」とは意味的な距離が近いので並列構造になるが、「先生」と「お酒」は並列構造にはならない。

2.2 用語同士の意味的な距離の定義

係り受け解析で係り先を決めるために 2 つの用語間の意味的な距離を定義した。本来、用語同士の意味的な距離はアナログ的なものである。極端な場合は人によっても異なるが、シソーラス上の用語同士の関係から表 1 のように定義した。

係り受け語は慣用的によく係り受けを構成する用語の組み合わせをネットなどから手作業で

表 1 意味的な距離の表

関係の種類	意味的な距離	組み合わせの例	
同義語	0	パソコン	コンピューター
反義語	1	明るい	暗い
狭義語	1	犬	土佐犬
広義語	3	犬	ほ乳類
関連語	4	犬	キツネ
		犬	犬小屋
係り受け語	1	寿命	延びる
未定義		シソーラスに未登録	

収集した．間に挟まる助詞と良しあしの情報も持っている．また係り受け語には，係の用語に意味が指定できる．(5.2 に示す)

例 (人)が 飲む

(人)は人の意味で人の意味の用語すべてを指定できる．

筆者らのシソーラスでは直近の関係語との関係しか持っていない．関係語とさらにその関係語との意味的な距離はそれぞれの意味的な距離を加算することにした．狭義語のさらに狭義語との意味的な距離は $1 + 1$ で 2 であると定義した．こうすることで，シソーラスにお互いの関係が登録されていない用語間の意味的な距離を定義した．

経験的にあまり遠い関係の用語同士の距離は評価しても意味がないので一定の距離で足切りをしている．足切りの値は係り先を決めるときと，並列構造を決めるときとは異なる．並列構造を決めるときの方が，広く関係を評価している．並列構造を決めるときには，係り受け語は考慮しない．

意味的な距離を測るときに多義語を区別している

意図したのと異なる意味の用語との距離を測ってしまうことが問題になることがある．例えば「お稲荷さん」には 2 つの意味がある．

意味的な距離

お稲荷さん - 稲荷神社 0 (同義語)

お稲荷さん - いなりずし 0 (同義語)

多義語をそれぞれの意味で区別しないで計算すると，「稲荷神社」と「いなりずし」とが 0 (同義語) になってしまう．

稲荷神社 - お稲荷さん - いなりずし

このことを防ぐために，我々のシステムでは「お稲荷さん」の 2 つの意味を区別して別の用

語として管理している．その結果「稲荷神社」と「いなりずし」との意味的な距離は未定義（無限大）になるので「いなりずしに参拝する」などという無意味な係り受けは排除される．

3 係り受けデータの整理

日本語では同じことを言うのにいくつかの書き方が許されている．自然文検索などで漏れを少なくするために形を整理する．

3.1 用語の標準化

日本語は表記の揺れを含めて同義語が多い．(国分, 岡野 2010) 著者と検索者とで異なる表記が使われることが検索漏れの一因になっている．検索対象データベース, 検索文ともに係り受けにしたあとシソーラスを用いてなかの用語を言語工学研究所が推奨する用語に標準化する．誤った表記, 差別語も標準の表記に置き換える．

例 1

インタフェイス
 インタフェース (J I S) ⇒ インターフェース (言語工学研究所推奨)
 インターフェイス (学術用語)
 インターフェース (新聞)

例 2

米, 米国
 U S A , U . S . A . ⇒ アメリカ (言語工学研究所推奨)
 合衆国, アメリカ合衆国
 アメリカ

3.2 係り受けを正規化

用言の使い方には限定用法と叙述用法がある．自然文検索で「青いリンゴ」(限定用法)と書いてある記事を「リンゴが青い」(叙述用法)という係り受けで検索しても検索できない．検索できるようにするために, 原記事, 検索文ともに係り受けは限定用法のものをすべて叙述用法に統一して, 正規化する．

例 青, い, リンゴ リンゴ, が, 青, い

用言が動詞の場合は, 動詞の性質によって名詞との間に挟む助詞が異なる．

例 食べたリンゴ リンゴを食べた (他動詞)

落ちたリンゴ リンゴが落ちた (自動詞)

当面, 間に挟む助詞も, 表 2 の 4 種類に限定している．

表 2 係り受けの関係の種類

が, は	動作主体
を	動作対象
の	連体修飾
上記以外	その他

4 情報の付与

今後、本解析システムを様々な目的へ適用を進める予定である。そこで、辞書上の情報をもとに用途に応じて解析結果に必要な情報を付与する。

4.1 文節意図を付与する

解析の精度を上げるためと、翻訳などで必要な情報を取り出しやすくするために、「係り受けの語幹まで」と、それに続く「付属語の並び」をまとめて管理している。

例 係り受け 付属語の並び
お酒を飲 んでください

さらに、付属語の並びの持つ表 3 のような性質を「文節意図」と呼ぶこととする。これは一般に命題に含まれる否定・肯定、ボイス、テンスなどを、モダリティーと一緒にしたものである。(益岡 2000)

乾健太郎「KURA」佐藤理史「醍醐プロジェクト」は、ここでいう係り受けの部分をより分かりやすくするための置き換えを説明したものであるので、本システムとは目的が異なる。

文の文節意図の把握

翻訳以外でも、例えば内容が「依頼」の記事を集めようとしたとき、記事を解析して係り受けにまでしても、結局人手で読み直して分類する必要があった。

文末の文節意図が「依頼」の文を含む記事を集めると、その目的が達成される。ほとんどの場合、文末の文節意図が文全体の文節意図を表している。

文節意図と主語の推定

省略された主語は文脈を調べないと分からない場合もあるが、文節意図を調べると推測できる場合がある。翻訳などのために省略された主語の人称を推定する。(4.3 で述べる)

同じ文節意図でも丁寧さの違いなどでいろいろな書き方がある。表 4 に「依頼」の文節意図の例を上げたが、ここに上げたのはその一部でこのほかにもいくつかの書き方がある。

ひとつの文節が複数の文節意図を持つことがある。たとえば下記の文節は、「禁止」、「否定」、「疑問」、「推量」、「丁寧」の 5 つの文節意図を持っている。

例 逢ってはいけないものでございましょうか

表 3 代表的な「文節意図」と人称の例

文節意図	人称	例
願望	1 人称	飲みたい
意志	1 人称	飲もう
許可要求	1 人称	飲んでかまいませんか
依頼	2 人称	飲んでください
勧誘	2 人称	飲みますか
禁止	2 人称	飲んではいけません
確認	2 人称	飲んだね
伝聞	3 人称	飲むそうだ
疑問		飲めますか
義務		飲むべきです
譲歩		飲んでもかまいません
推量		飲むだろう
否定		飲まない
可能		飲める
受身・尊敬	2 人称	飲まれる

表 4 「依頼」の「文節意図」の例

- 飲んでほしい
- 飲んでください
- 飲んでくださいませんか
- 飲んでくれ
- 飲んでくれないか
- 飲んでくれないですか
- 飲んでくれない
- 飲んでくれませんか
- 飲んでくれませんか

付属語の組み合わせによって，文節意図は変化する

- 例 飲むつもりです 1 人称 意志
- 飲むつもりですね 2 人称 確認

このような複雑な文節意図を持つ文節を扱うためと，解析の速度を速くするために，辞書上で

は数の付属語を「付属語の並び」としてまとめた形で管理している．筆者らのシステムでは解析辞書(5.2で述べる)に1,300,000行の「付属語の並び」を持っている．

4.2 良しあし，注目度を付与する

レコメンデーションでは，良しあしの情報を手がかりのひとつにする．

単独で良しあしの決められる辞書上の用語には，良しあしのフラグを付けてある．

例 美しい	(良い)	汚い	(悪い)
静寂	(良い)	騒音	(悪い)
さっぱりする	(良い)	さっぱりだ	(悪い)

しかし，用語単独では良しあしが決められず，係り受け関係を調べないと決められないことがある．

例 寿命	が	延びる	(良い)
寿命	が	短い	(悪い)

「寿命」，「延びる」，「短い」など用語は単独では良しあしの性質は持っていないが，組み合わせられたときに良しあしの性質が出てくる．筆者らのシソーラスでは，「係り用語」と，「受けの用語」と，「間にはさまれた格助詞」と，「良しあし」の組で管理している．シソーラスの係り受け語を調べて係り受けの良しあしを決める．

係り，受けのそれぞれの用語の同義語，狭義語をシソーラスで拡張して，係り受け語として登録されていない係り受けにも対応できるようにした．

例 ビール	が	冷えている	
		麦酒	が 冷えている (ビールの同義語)
		生ビール	が 冷えている (ビールの狭義語)

否定の文節

良しあしは否定があると逆転する．

例 ビール	が	冷えている	(良い)
ビール	が	冷えていない	(悪い)

日本語では「ない」と書いてあっても否定だとは決められない(表5参照)．否定になるかどうかも付属語の並びに記述してある(5.2で述べる)．

注目度を付与して良しあしを辞書に登録することができる

「良しあし」の判断基準は普遍的なものではない．ユーザーによって異なることがある．また自社や競合他社の商品名のようにその評判をいつも注目しておきたい用語もある．

筆者らのシステムはユーザーにより適切なレコメンデーションをするために，このような用語または係り受けに注目度，良しあしをつけて登録する仕組みが用意してある．(表6参照)

表 5 「ない」を含んでいても否定にならない例

例	文節意図
飲まないか	勧誘
飲むかも知れない	推量
飲まなければならない	義務
飲んでもらえないか	依頼

表 6 良しあし, 注目度の例

係り	受け	注目度	良しあし
メタミドホス		7	悪い
騒音	小さい	2	良い
リンゴ	甘い		良い
判断	甘い		悪い

4.3 主語の人称の推定

日本語はしばしば主語が省略されるが，そのまま翻訳すると訳文があいまいになることが少なくない．またビジネス文書でも主語を解析して認識することは重要である．実際の文章では，待遇表現や文節意図で暗に主語の人称を示している．本論文ではこの点に注目して，文節意図の情報を利用することにより主語を推定する方法を説明する．

待遇表現によって主語の人称を推定する

ビジネス文書では，待遇表現は適切に使用されているので有効な推定法である．解析辞書には文節がどの待遇表現になるかが記述してある．

謙譲語が使われている動詞の主語は1人称である．

例 申し上げたのは (1人称が) 申し上げたのは

尊敬語が使われている動詞の主語は2人称ないしは3人称である．

例 おっしゃったのは (2人称が) おっしゃったのは

文節意図によって主語の人称を推定する

待遇表現でも主語が推定できなかったときに文節意図によって主語の人称を推定する．(表 3 参照)

例えば，文節意図が「意志」のときは，主語は1人称である．

例 飲みたい「意志」 (1人称が) 飲みたい

文節意図が「依頼」「指示」のときは，受けの主語は2人称である．

例 送ってくれ「依頼」 (2人称が)送ってくれ

5 辞書

ブログやメールに代表されるような文書を扱うために、時事的な用語や省略語も積極的に登録している。送り仮名や訳語などの差異による異表記語も網羅的に収集した。よく使われる用語であれば誤った用語(例「キューピット」cupid)も積極的に採択してある。反面、古語や文学作品にしか出てこない用語は採択していない。

すべての辞書は共通の品詞に分類してある。用言は語幹と活用形で管理している。

5.1 シソーラス

自然言語処理を目的とした一般語を主とするシソーラスである。いわゆる名詞だけでなく、動詞、形容詞、形容動詞、副詞、代名詞、擬態語さらに慣用句までを登録している。

「広義語 - 狭義語」の関係は、自然言語処理で広義語に適用した規則が狭義語にも適用できるように同じ属性のものだけとした。「自動車」 - 「タイヤ」のような全体 - 部分関係は関連語とした。品詞の異なる用語、「自動詞」 - 「他動詞」の対応なども関連語とした。

用語間の意味関係として、表7のものを用意した。詳細は(国分, 岡野 2010)を参照されたい。

表7 シソーラスの用語同士の意味関係

意味関係	説明
同義語	例 犬, ドッグ 表記の揺れも含む
反義語	例 強い, 弱い
狭義語	例 犬, 秋田犬
広義語	例 犬, 哺乳類
関連語	例 犬, キツネ 犬, 犬小屋
推奨語	用語標準化のために同義語のグループの中の、言語工学研究所で推奨するひとつの用語にマークがしてある
係り受け語	係り受けになる用語の組み合わせである 係りには名詞の意味(5.2で述べる)を指定できる 係りの用語と受けの用語との間に入る格助詞および、良しあしの情報が入っている 例 ネット で 調べる 寿命 が 長い (良い)

シソーラスのその他の項目

エラーフラグ	誤った表記，差別語
品詞	動詞の活用形を含めて 24 種類
注目度	レコメンデーションのためにユーザーがマークした用語．
異なり語の数	440,000 語

(竹内 2008) は動詞の性質を分類するために動詞のそれぞれの性質で係り受けする名詞の一例を示したものである．一方筆者らのこのシソーラスでは，受けになる個々の用言を中心に，その用言の係りとなり得る名詞をなるべく網羅的に収集した．

5.2 解析辞書

ここで実現しているアルゴリズムは，辞書の情報で制御する方式をとっている．そのために必要になる情報が各用語に付与してある．

自立語

自立語の構成要素である接頭辞，接尾辞，助数詞なども含む．

語数	240,000 語
品詞	動詞の活用形も含めて 24 種類
良しあし	自立語の良しあしが記入してある

否定フラグ

名詞の意味 係り受け関係を調べるために次の 10 種類ある．

1 つの用語が複数の意味を重複して持てる．

意味	例
人	先生，山田，雅子
機関	学校，研究所
物	机，物理現象も含む
時	昨年
場所	東京，駅前
数量詞	5 本，少し
抽象名詞	芸術，甘さ
動作名詞	サ変動詞
数詞	9，二百
不定	代名詞，未知語などで意味が決定できないもの．

用言 カッコ内は活用語尾である．

品詞・活用形	例
サ変名詞形	勉強(する)

サ変非名詞形	察(する)
ザ変	信(ずる)
一段	生き(る)
カ行五段	書(く)
カ行五段例外	行(く)
ガ行五段	泳(ぐ)
サ行五段	押(す)
タ行五段	立(つ)
ナ行五段	死(ぬ)
バ行五段	遊(ぶ)
マ行五段	飲(む)
ラ行五段	走(る)
ラ行五段例外	おっしゃ(る)
ワア行5段	買(う)
ワ行五段例外	問(う)
形容詞	青(い)
形容動詞	閑静(な)
形容動詞とノたる形	鬻鑠(たる)
副詞	さっぱり
連体詞	こんな
打ち消しの動詞	年端もいか(ない:助動詞)
打ち消しの形容詞	必要(ない:形容詞)

動詞の性質

自・他動詞	限定用法から叙述用法に変換するときに挟む助詞を決定するため.
移動性の動詞	自動詞であるが「を格」をとる. 例 道を行く
待遇表現	尊敬語・謙讓語 翻訳などで主語を決定するため.
	例 おっしゃる (尊敬語)
	申し上げる (謙讓語)

付属語の並び

文節意図を付与するために助詞, 助動詞だけでなく, いわゆる機能表現とその組み合わせをまとめた形で扱っている.

付属語

助詞

助動詞とその活用語尾

形式名詞

機能表現のための動詞およびその活用語尾

例えば推量を表す「～かも知れない」というときの「知れる」という動詞．

機能表現のための形容詞およびその活用語尾

行数 1,300,000 行

エラーフラグ 間違った表記

文節意図 表 3 参照

並列フラグ 並列を構成しうるかどうか

待遇表現 尊敬語・謙譲語

否定フラグなど

6 結果

C G M (消費者生成メディア) の例として Yahoo!知恵袋データの 2004 年 4 月分の質問記事 (5,957 記事, 15,883 文) を用いて, 評価した．まず, 筆者らのシステムと Cabocha との解析精度を比較した．会話体の文章なので両者ともあまり良い結果はでなかったが正解率は Cabocha を 13.8 ポイント上回っている．

	正しい	間違い	正解率
筆者らのシステム	11,690 文	4,193 文	73.6%
Cabocha	9,498 文	6,385 文	59.8%

両システムの処理時間も比較してみた．筆者らのシステムはシソーラスを参照しているので, 処理が遅くなる恐れがあったが, 両者はほとんど差がなかった．これは, 筆者らのシステムでは付属語をまとめた形で扱うことによって, 辞書へのアクセス回数を減らしたためと考えられる．

筆者らのシステム	1 分 39 秒
Cabocha	1 分 30 秒

次に, 筆者らのシステムでシソーラスを組み込んで解析した結果と, 組み込まないで解析した結果との差を調べた．係り受け構造の違いと, 並列構造の違いとに分けて集計した．しかし, シソーラスを組み込んだ結果, シソーラスを組み込まないときには正しく解析できた結果を, かえって間違った構造にしてしまった場合もあった．

差分の生じた 172 文の内訳は改善 160 文, 悪化 12 文, 差引 148 文

全体 15,883 文に対して 0.9 % の向上が観測された．

	係り受け	並列	合計
正しい構造にした	106 文	54 文	160 文
間違った構造にしてしまった	8 文	4 文	12 文
合計	114 文	58 文	172 文

成功例

「音楽がいつまでたっても始まりません。」

音楽が
いつまで
たっても
始まりません。

「音楽 / が / 始ま / 」という係り受けが登録されているので、「音楽が」という文節が「たっても」という文節ではなく、「始まりません」という文節に係った。

間違った構造にしてしまった例

「警察の方に話がいつまでたっても始まりません。」

話が
警察の方に
いつまでたっても
始まりません

「話 / が / 分かる / 」という係り受けが登録されていたため、「話が」という文節が「いつまでたっても」という文節ではなく、「始まりません」という文節に係ってしまった。

下記のサイトから使ってみられるようにしてあるので、試用して評価していただくことを希望する。

<http://www.gengokk.co.jp/koubun/>

7 おわりに

CGMのような会話体の文章を扱うためには、より一層の誤りを含んだデータに対応できることが要求される。辞書は手作業で収集したもので、漏れも多いと思う。係り受けがシソーラスに登録されていないために、今回の評価の対象にならなかった記事もあると思われる。今後大規模コーパスを解析して、係り受けを抽出してシソーラスの係り受け語を充実させていく計画である。高速化についても現在改造中で近日中に発表する予定である。

全世界に言語の数は多い。現在までに日本語から外国語への翻訳プログラムが未着手の言語を対象に、翻訳プログラムが作れないかと思っている。今後、外部の人を含めて実用化を進めたいと思っている。自然文検索や翻訳だけでなく、いろいろな応用が考えられる。

係り受けだけで「よしあし」を決定しているが、3つ以上の文節が組み合わさって「よしあし」が決定される場合がある。これから充実させていく必要がある。

謝 辞

本稿に対して有益なご意見, ご指摘をいただきました査読者の方に感謝いたします。また国立情報学研究所が提供する「Yahoo!知恵袋-研究機関提供用データ」を利用させていただいたことを, 感謝いたします。

参考文献

- 益岡隆志 (2000). “命題とモダリティの境界を求めて” 日本語文法の諸相. くろしお出版.
- 乾健太郎. “KURA.”, http://cl.aist-nara.ac.jp/kura/doc/publication_list.html.
- 宮崎和人, 安達太郎, 野田春美, 高梨信乃 (2002). “モダリティ” 新日本基本文法選書. くろしお出版.
- 国分芳宏, 岡野弘行 (2010). 複数の観点で分類した自然言語処理用シソーラス. 自然言語処理, 17(1), pp.247-263.
- 黒崎禎夫 (1992). 長い日本語文における並列構造の推定. 情報処理学会論文誌, 33(8), pp.1022-1031.
- 佐藤理史. “醍醐プロジェクト.”, <http://sslslab.nuee.nagoya-u.ac.jp/sato/research/daigo.html>.
- 仁田義雄 (1991). 日本語のモダリティと人称. ひつじ書房.
- 竹内孔一 (2008). 動詞項構造シソーラス. 竹内孔一 HP.
- 藤尾正和 (2000). 語彙統計モデルに基づく日本語依存構造解析. 奈良先端科学技術大学院大学博士情報科学研究科博士論文.
- 南不二男 (1974). 現代日本語の構造. 大修館書店.
- 白井諭 (1995). 階層的認識構造に着目した日本語従属節間の係り受け解析の方法とその精度. 情報処理学会論文誌, 36(10), pp.2353-2360.

略歴

- 国分 芳宏 (正会員) : 1966年東京理科大学理学部応用物理学科卒。同年日本科学技術情報センター入社。1985年株式会社言語工学研究所設立代表取締役就任。自然言語処理、シソーラス作成に従事 情報処理学会会員。
- 梅北 浩二 : 1990年日本電子専門学校卒。1996年株式会社言語工学研究所入社 自然言語処理開発に従事
- 松下 栄一 : 1999年東京理科大学大学院工学研究科工業化学専攻卒。2004年株式会社言語工学研究所入社 自然言語処理開発に従事
- 末岡 隆史 : 1975年東京理科大学工学部電気工学科卒。1987年株式会社言語工学研究所入社 自然言語処理開発に従事