

TR-IT-0168

対話音声認識用日本語文法の 大語彙化および拡張の検討

Study on Enhancement of Vocabulary Size and Coverage of
the Japanese Grammar for Dialogue Speech Recognition

竹澤 寿幸

Toshiyuki TAKEZAWA

1996. 5

内容梗概

音声言語データベースから新たに50対話を選び、音声認識用日本語文法による正解構文木を作成した。「音声認識用日本語文法 Ver. 2」をあまり越えない範囲でCFG規則を加えたので、Ver. 2の文法で部分木として処理する方針とした文(主として不適格文)は依然として部分木のままで処理している。以前と比べて具体的にどのような興味深い言語現象が現れたかを列挙し、将来の語彙・文法の拡張作業や部分木併合処理の研究に備えることにした。本報告書ではその概要を述べる。

ATR 音声翻訳通信研究所

ATR Interpreting Telecommunications Research Laboratories

© 株式会社 エイ・ティ・アール音声翻訳通信研究所

© 1996 by ATR Interpreting Telecommunications Research Laboratories

目次

1	まえがき	1
2	対話データの選定	1
3	語彙および文法の拡張	3
4	部分木として処理する言語現象	4
4.1	格助詞の省略	4
4.2	助動詞の省略	5
4.3	述語の省略	5
4.4	主節の省略	5
4.5	断片文	5
4.6	箇条発話	6
4.7	主格助詞の「の」	6
4.8	連体助詞の「の」	6
4.9	連用中止	6
4.10	会話文に特有な終助詞	6
4.11	複雑な並列句	7
4.12	同格	7
4.13	節に接続する接尾辞	7
4.14	挿入文	7
5	むすび	7
	謝辞	8
	参考文献	8
A	付録 1: MSLR 法による音声認識用日本語文法の検討	9
A.1	はじめに	9
A.2	音声認識用日本語文法の改良案	9
A.3	おわりに	13
	参考文献	14
B	付録 2: 新規登録単語一覧	15

表目次

1	新規登録単語の内訳	3
2	新規登録規則の内訳	4
3	文法の諸元	7

1 まえがき

1995年度上半期に「部分木を単位とする音声認識用日本語文法」[竹沢 95a]を改良した[竹沢 95b]。これは、動詞と後置詞句との共起制約の規則化と複合語の部分文法化を施したもので、その結果、音素パープレキシティで20～30%、単語パープレキシティで40～60%の改善を達成した[竹沢 95b]。

一方、ATR音声言語データベースから選んだ12対話について、*N*グラム言語モデルの実験を行い、その結果、前終端記号(プレターミナル)のバイグラムが有効であることが分かった[竹沢 95c]。

以上の成果をふまえ、今回、オープンデータによるバイグラムデータを収集するための資料として、50対話例の正解構文木を作成した。その過程で現れた興味深い言語現象を整理し、将来の語彙・文法の拡張作業や部分木併合処理の研究に備えることにした。本報告書ではそれらの概要を述べる。

2 対話データの選定

「音声認識用日本語文法」は、ATR音声言語データベースから選んだ12の対話データを扱えるように規則および語彙を設定している。これらの対話例は次のような話題に関するものである。

対話番号	対話例番号	話題(内容説明)
TAS22001	対話 A	ホテルの部屋の予約(ニューワシントンホテル)
TAS12009	対話 B	ホテルの部屋の予約(キャンセル)
TAS33002	対話 C	ホテルの部屋の予約(京都観光ホテル)
TAC23034	対話 D	ホテルの部屋の予約(滞在延長)
TCC22074	対話 E	ホテルでのサービス(友達の病気)
TCC22094	対話 F	ホテルでのサービス(ルームサービス)
TCC23034	対話 G	ホテルでのトラブル(ブリーフケースの紛失)
TDS32007	対話 H	ホテルの会議室の相談・予約(パーティ予約)
TGS12001	対話 I	飛行機のフライトの予約・変更・解約(変更)
THS12002	対話 J	バス・列車の切符の問い合わせ(列車の予約)
TRS12001	対話 K	演劇・コンサートのチケットの予約・購入
TSS12002	対話 L	レストランの予約

今回、音声言語データベース[浦谷 94]の「音声対話ファイル」の中から50の対話例を選定するにあたっては、話題の分布がこれとだいたい似たものになること、なるべく多様なものになるように留意した。

対話番号	話題 (内容説明)
TAS12002	ホテルの部屋の予約 (ニューワシントンホテル)
TAS12005	ホテルの部屋の予約 (ホリデイインプラザ)
TAS12007	ホテルの部屋の予約 (ギャランティ)
TAS12010	ホテルの部屋の予約 (キャンセル)
TAS12011	ホテルの部屋の予約 (リゾートホテル)
TAS12015	ホテルの部屋の予約 (宿泊者の変更)
TAS12020	ホテルの部屋の予約 (リゾートホテル)
TAS13001	ホテルの部屋の予約 (京都観光ホテル)
TAS13002	ホテルの部屋の予約 (京都観光ホテル)
TAS13007	ホテルの部屋の予約 (二人二泊)
TAS13009	ホテルの部屋の予約 (一人一泊)
TAS13011	ホテルの部屋の予約 (人数の変更)
TAS13018	ホテルの部屋の予約 (ウィークエンドレディースパック)
TAS13022	ホテルの部屋の予約 (鹿王院)
TBS12001	ホテルの紹介 (ダウンタウンのホテル)
TBS12003	ホテルの紹介 (夫婦での旅行)
TBS13001	ホテルの紹介 (京都ツーリスト)
TCS12001	ホテルでのサービス (クリーニング)
TCS12003	ホテルでのサービス (テレビの故障)
TCS12005	ホテルでのサービス (子供の発熱)
TCS12021	ホテルでのサービス (ハーフムーンクルーズ)
TCS12023	ホテルでのサービス (バースデイパーティー)
TCS13001	ホテルでのサービス (クリーニング)
TCS13018	ホテルでのサービス (カメラの忘れ物)
TCS13020	ホテルでのサービス (宿坊の問い合わせ)
TCS13022	ホテルでのサービス (京都観光の問い合わせ)
TDS12003	ホテルの施設の利用 (プール)
TDS12004	ホテルの施設の利用 (プール)
TDS12005	ホテルの施設の利用 (ランチミーティング)
TDS13005	ホテルの施設の利用 (出張パーティー)
TGS12002	飛行機のフライトの予約・変更・解約 (便の変更)
TGS13001	飛行機のフライトの予約・変更・解約 (日にちの変更)
TGS13002	飛行機のフライトの予約・変更・解約 (日にちの変更)
THS12001	バス・列車の切符の問い合わせ・購入 (列車の問い合わせ)
THS12003	バス・列車の切符の問い合わせ・購入 (バスの問い合わせ)
THS13001	バス・列車の切符の問い合わせ・購入 (新幹線の問い合わせ)
TIS12001	交通手段の問い合わせ (自由の女神に行くには)
TIS13001	交通手段の問い合わせ (延暦寺に行くには)
TKS12001	レンタカーの問い合わせ・申し込み (アムステルダム)
TKS13001	レンタカーの問い合わせ・申し込み (京都)
TOS12001	旅行パックの問い合わせ・申し込み (ショッピング)
TOS12003	旅行パックの問い合わせ・申し込み (美術館)
TOS13001	旅行パックの問い合わせ・申し込み (京都バスツアー)
TOS13003	旅行パックの問い合わせ・申し込み (京都バスツアー)
TRS12002	演劇・コンサートのチケットの予約・購入 (オペラ座の怪人)
TRS13001	演劇・コンサートのチケットの予約・購入 (歌舞伎)
TRS13002	演劇・コンサートのチケットの予約・購入 (歌舞伎)
TSS12001	レストランの予約 (レインボールーム)
TSS13001	レストランの予約 (和食)
TSS13002	レストランの予約 (和食)

3 語彙および文法の拡張

新しく登録した単語は、962語(付録2参照)であった。品詞別に語数を挙げると次のようになる。

表 1: 新規登録単語の内訳

品詞	語数	パーセント
固有名詞	134	13.9
人名	2	0.2
サ変名詞	82	8.5
形容名詞	22	2.3
普通名詞	357	37.1
数詞	1	0.1
日時	1	0.1
代名詞	7	0.7
本動詞	118	12.3
形容詞	23	2.4
副詞	55	5.7
連体詞	5	0.5
接続詞	15	1.6
感動詞	18	1.9
間投詞	18	1.9
助動詞・補助動詞	28	2.9
格助詞	5	0.5
引用助詞	2	0.2
副助詞	8	0.8
連体助詞	3	0.3
接続助詞	11	1.1
並立助詞	1	0.1
終助詞	4	0.4
接頭辞	4	0.4
接尾辞	37	3.9
活用語尾	2	0.2
合計	962	100.0

また、正解木をできるだけ多くするために、「音声認識用日本語文法 Ver. 2」[竹沢 95b]をあまり越えない範囲で句構造規則を加えた。新しく登録した規則数は405規則、その内訳は次の通りである。

表 2: 新規登録規則の内訳

規則のタイプ	規則数	パーセント
述語	110	27.2
活用	51	12.6
後置詞句	75	18.5
副詞句	4	1.0
係り受け	71	17.5
名詞句	8	2.0
住所	16	4.0
数詞	12	3.0
連体句	7	1.7
並列句	6	1.5
複合語	45	11.0
合計	405	100.0

これを見ると、単語では固有名詞・普通名詞・本動詞が多く、規則では述語関連の規則と後置詞句・係り受けの規則が多いことが分かる。特に、述語や後置詞句・係り受け関連の規則は、品詞を細分化して制約を強くしているものが多い。ということは、これらの規則のひとつひとつがカバーできる言語表現が限られているということで、扱う文を増やせばそれに応じてこの種の規則をそのつど新たに加えなければならなくなる。

これを避けるためには、句構造規則はできるだけ一般的なものにし、その代わりに、前終端記号(プレターミナル)の接続情報を使って制約を加えるという方法が考えられる。これについては、付録1を参照していただきたい。

4 部分木として処理する言語現象

今回は「音声認識用日本語文法 Ver. 2」[竹沢 95b]をあまり越えない範囲で CFG 規則を加えたので、Ver. 2の文法[竹沢 95b]で部分木として処理する方針とした文(主として不適格文)は依然として部分木のままで処理している。そのような文は部分木として出力されることになるから、あとの言語処理部で併合処理を施す必要がある。ここでは、調査対象を50対話に広げたことによって、以前と比べて具体的にどのような興味深い言語現象が現れたかを列挙し、将来の語彙・文法の拡張作業や部分木併合処理の研究に備えることにしたい。

4.1 格助詞の省略

主格助詞の「が」や目的格助詞の「を」や連体助詞の「の」がしばしば省略される。

例文1 女性のお客様(が)三名様までで、土曜日の夜にご利用いただけるパジャマパーティのプランです。

例文2 えーと、じゃあ、二名で、二日とも朝食の予約(を)お願いします。

例文3 カードと現金(の)どちらになさいますか。

ただし、文法的に言えば、助詞が本当に省略されているのか、それとも、もともと助詞が存在しなかったのかは、微妙な問題である。

4.2 助動詞の省略

重文において、前の節の助動詞「で」が省略されることがしばしば生じる。

例文 4 スタンダードツインの方は一万八千円(で)、デラックスツインの方は二万一千円となっております。

この文は、「スタンダードツインの方は一万八千円」と「デラックスツインの方は二万一千円」の並列構造で、「となっております」が両方を受けていると考えることもできる。また、サ変名詞につく助動詞もしばしば省略される。

例文 5 例えば、自動車で旅行(する)ですとか、長期滞在(する)ですとか。

4.3 述語の省略

述語は文の根幹とも言うべき重要な部分であるから、それが省略されるのは、省略されてもコミュニケーションにさしつかえない場合だけである。ほとんど自明であるとか、慣用的な言い回しであるとか、といったことが多い。

例文 6 それはちょっと。

例文 7 特典とおっしゃいますと。

例文 8 では、楽しい旅を。

例文 9 そうですねー、歩いて三十分っていうところでしょうか。

例文 6 は、「ロビーにテラスがございますので、そちらに行ってみていただけませんかでしょうか」という要望に対して、客が渋っているときの発話である。

4.4 主節の省略

次は、ホテル側が用件を尋ねたのに対して、客が答えた発話である。

例文 10 ちょっと困ったことが起こりました。

例文 11 チケット持ってるんですけども、都合が悪くて。

接続助詞で終わっているのも、文法的には主節が省略されていると考えられるが、対話の当事者には後に何かが続くとは意識されていないだろう。前の「特典とおっしゃいますと」と同じく、このような表現をそれだけで完結した発話として対話の流れの中に位置づけるべきだろう。

4.5 断片文

格助詞と動詞・助動詞が同時に省略されることもある。そうになると、ほとんど断片の連なりと変わらなくなる。

例文 12 カード番号、五二七九、三九二零、二四六九、零零九八。

例文 13 東京シティーホテル、ご滞在。

4.6 箇条発話

システムが確認のために予約事項を列挙する発話は、旅行ガイドのような対話では必ず現れる。それも、単に名詞句を列挙するだけでなく、文の断片のようなものが挟まれたりして、非常に複雑な構文になることがある。

例文 14 ジョン・フィリップス様、二名様、ご宿泊予定日は十月二十七日木曜日、二十八日金曜日です。

例文 15 デラックスツインのお部屋、そして、朝食を二名様、二回ということでしょうか。

4.7 主格助詞の「の」

「の」が主格助詞として使われることがしばしばあるが、それを規則化すると曖昧性 (ambiguity) がかなり増えることになるだろう。

例文 16 あなたの言ったホテルの電話番号を教えてください。

例文 17 交通の便利なところがいいですね。

4.8 連体助詞の「の」

連体助詞の「の」は典型的には名詞句に接続して連体修飾句を作るが、しばしば、名詞句以外の句に接続することがある。

例文 18 では、続いて、JRを使っての行き方をご説明いたします。

4.9 連用中止

経験によれば、連用中止を規則化すると一挙に解析の曖昧性 (ambiguity) が増えてしまう。しかし、対話では頻繁に現れる。

例文 19 シングルルームで一名の予約から、ツインルームで二名の予約に変えたいのです。

例文 20 こちらも 祇園にあり、観光の方に大変人気のあるホテルです。

あとの例文は、主格助詞の「の」が使われている文でもある。

4.10 会話文に特有な終助詞

例文 21 値段はいくらなの。

丁寧な会話文では「値段はいくらなんですか」となるところである。これならば現在の文法でも解析可能である。ところが、くだけた会話文では例文のような表現がいくらでも現れる。

4.11 複雑な並列句

並列句が重なり入れ子になって、非常に複雑な構造になっている文がある。

例文 22 その他に、ゲームが二種類、キャンドル、花束、それから、当ホテルからのプレゼントがございます。

4.12 同格

同格、または、言い換えといってもいい例である。今回の調査範囲ではこの例はあまり多くないが、会話ではわりに頻出する可能性がある。

例文 23 市街地、交通の便利なところがいいですね。

例文 24 えーと、あとに出るやつ、「のぞみ」をお願いします。

4.13 節に接続する接尾辞

補語と述語を備えた節に接尾辞が接続することがある。

例文 25 今そちらで公演中の芝居なんですけど、「オペラ座の怪人」でしたよね。

4.14 挿入文

ひとつの文の中に、挿入文と、それを受けての照応が存在するといった、複雑な構造の文がある。

例文 26 でも、一人あたり八千円の予算と思ってたんで、えーと、竹ですか、それをお願いします。

「竹ですか」という問いかけの文が挿入されているが、このようにひとつの文の中で、情報提供、質問、要求といった発話タイプの複雑な転換が行われている。

5 むすび

語彙・文法を拡張した結果、文法の諸元は次の通りになった。

表 3: 文法の諸元

バージョン	対象対話	語彙数	規則数	パープレキシティ	
				音素	単語
2-A	50 対話	1972	4186	4.47	110.44

なお、対象対話 50 対話の総文数は 1959 文で、それに対して解析に成功した文の数は 1434 文 (73.2%) であった。今回は、前終端記号のバイグラムデータを得ることが主な目的であったので、解析できない文は部分木として出力した。もしも、対象対話についてできるだけ正解木が出るようにするとすると、新たに加えなければならない CFG 規則の数はずっと増えることになるだろう。また、それに伴って、パープレキシティの値も大きくなるはずである。

謝辞

本研究を進めるにあたり、協力をいただいた衛藤純司氏に感謝します。また、定期的に会合に参加していただき、助言をいただいた 荒川直哉、加藤直人、巖寺俊哲、田代敏久 各氏に感謝します。

参考文献

- [浦谷 94] 浦谷則好, 竹沢寿幸, 松尾秀彦, 森田千帆: “音声言語データベースの構成”, *ATR* テクニカルレポート, TR-IT-0056 (1994-05).
- [竹沢 94] 竹沢寿幸, 田代敏久, 衛藤純司: “部分木を単位とする音声言語解析のための言語現象の調査”, *ATR* テクニカルレポート, TR-IT-0080 (1994-11).
- [竹沢 95a] 竹沢寿幸, 田代敏久, 衛藤純司: “部分木を単位とする音声認識用日本語文法”, *ATR* テクニカルレポート. TR-IT-0110 (1995-04).
- [竹沢 95b] 竹沢寿幸, 田代敏久, 衛藤純司: “部分木を単位とする音声認識用日本語文法の改良”, *ATR Technical Report*, TR-IT-0139 (1995-10).
- [竹沢 95c] 竹沢寿幸, 森元暎: “部分木を単位とする構文規則と前終端記号のバイグラムを利用した連続音声認識”, 信学技報, NLC95-54, SP95-89, 情処研報, 95-SLP-9-8 (1995-12).

A 付録 1: MSLR 法による音声認識用日本語文法の検討

A.1 はじめに

音声認識用の文法に課せられる役割は次の 2 つである。

- (1) 文の構造記述を与えること。
- (2) 単語や句などの接続に制約を加えて探索空間を絞ること。

ところが、この 2 つの要請は互いに相反することが多い。その理由は、音声認識と言語解析を統合的に行うためには構造記述はできるだけ一般的であることが望ましいが、音声認識を正確に行うためには特殊ではあってもできるだけ詳細な制約を加えなければならないからである。

現在は、文脈自由文法 (CFG) がこれら 2 つの役割を負わされている。その結果どういふことが起こるかという、単語や句の接続制約を強めようとするれば非終端記号を細分化しなければならず、それはただちに構造記述に反映して特殊な規則がむやみに増えるということになる。このことは、5000 ~ 10000 語と語彙サイズを拡大していくにしたがって、文法開発にさまざまな困難をもたらすことになるだろう。

文献 [田中 95] で提案されている方法 (MSLR 法 = Morphological & Syntactic LR 法とよぶ) は、このような問題を解決するために構造記述と接続制約記述を分離してそれぞれの要請を互いに独立に満足させようとするものである。これによると、

- (1) 文法記述者は、接続可能性を考慮した新たな非終端記号の導入を行うことなしに、CFG の記述が可能である。
- (2) 接続可能性を表す接続表の記述者は、CFG 規則とは無関係に接続表の記述が可能である。

この方法にしたがって音声認識用文法を記述してみると、実際に CFG 規則は少数のごく一般的な規則ですむようになり、接続制約はより一層精密で柔軟な記述が可能になる。

A.2 音声認識用日本語文法の改良案

MSLR 法は、文法的な制約のうち、大域的なものを CFG で、局所的なものを接続表で記述し、両方をひとつの LR 表に組み込もうとするものである。こうすることによって、文法記述が非常に見通しのよいものになるうえに、LR 表のアクションの数を大幅に削減することができるようになる (接続表を LR 表に組み込む方法、および、制約を伝播してアクションを削除する方法については、[田中 95, 綾部 96] を参照していただきたい)。

この方法の利点を知るには、従来の文法と新しく改良した文法を比べてみるのがよい。中でも文末述語の規則は MSLR 法の効果が最も著しく現れた部分であるので、これを例にとって説明しよう。

周知のように、日本語の文末述語は、動詞・助動詞の相互承接関係が厳密に定まっています。接続制約を記述しやすい部分である。しかし、それを CFG 規則だけで記述するとなると、規則の数が膨大なものになってしまう。おまけに、述語というのはアスペクトやモダリティや敬語表現などの要素を含み、日本語の根幹をなす「陳述形式」の多様性が集中して現

れる部分でもある。そのため、語彙を拡大し、カバーすべき文を増やすにしたがって、文末述語を形成する CFG 規則は際限なく増え続けるということにもなる。事実、これまでの文法開発の経緯を振り返ってみても、文末述語の規則を作成し調整することに大半の精力が費やされたと言ってもよい。

しかし、MSLR 法にしたがって記述すると、こうした労苦が劇的と言ってもいいほどに軽減されることになる。次の例文で説明しよう。

例文 1 ここで少し待たせていただけませんか。

「待たせていただけませんか」という文末述語を解析するために、従来の文法では次のような規則が必要であった。

```
(<vp-sfp> <--> (<vaux-sfp>))
(<vaux-sfp> <--> (<vaux-guess-syusi> <aux-sfp>))
(<vaux-guess-syusi> <--> (<vaux-masu-negt-syusi> <aux-guess-deshou>))
(<vaux-masu-negt-syusi> <--> (<vaux-masu-mizen1> <aux-negt-n>))
(<vaux-masu-mizen1> <--> (<vaux-te-renyo> <aux-masu-mizen1>))
(<vaux-te-renyo> <--> (<verb-1-renyo> <aux-te-renyo>))
(<verb-1-renyo> <--> (<vstem-1dan>))
(<aux-te-renyo> <--> (<auxstem-te-1>))
(<aux-masu-mizen1> <--> (<auxstem-masu> <vinfl-spe-se>))
(<aux-guess-deshou> <--> (<aux-cop-desu-mizen2> <aux-intn>))
(<aux-cop-desu-mizen2> <--> (<auxstem-desu> <vinfl-spe-sho>))
(<vstem-1dan> <--> (m a t a s e))
(<auxstem-te-1> <--> (t e i t a d a k e))
(<auxstem-masu> <--> (m a))
(<vinfl-spe-se> <--> (s e))
(<aux-negt-n> <--> (n))
(<auxstem-desu> <--> (d e))
(<vinfl-spe-sho> <--> (sh o))
(<aux-intn> <--> (u))
(<aux-sfp> <--> (k a))
```

これらの規則によって、次のような構造が与えられる。

```

<vp-sfp>
|--<vaux-sfp>
  |--<vaux-guess-syusi>
    | |--<vaux-masu-negt-syusi>
    | | |--<vaux-masu-mizen1>
    | | | |--<vaux-te-renyo>
    | | | | |--<verb-1-renyo>
    | | | | | |--<vstem-1dan>
    | | | | | |--matase
    | | | | | |--<aux-te-renyo>
    | | | | | |--<auxstem-te-1>
    | | | | | |--teitadake
    | | | |--<aux-masu-mizen1>
    | | | | |--<auxstem-masu>
    | | | | | |--ma
    | | | | | |--<vinfl-spe-se>
    | | | | | |--se
    | | |--<aux-negt-n>
    | | | |--n
    | |--<aux-guess-deshou>
    | | |--<aux-cop-desu-mizen2>
    | | | |--<auxstem-desu>
    | | | | |--de
    | | | | |--<vinfl-spe-sho>
    | | | | | |--sho
    | | |--<aux-intn>
    | | | |--u
  |--<aux-sfp>
    |--ka

```

これに対して、MSLR 法による文法記述では、次の5つの CFG 規則があれば足りる。

```

(<vp> <--> (<vaux>))
(<vaux> <--> (<verb> <aux>))
(<vaux> <--> (<vaux> <aux>))
(<verb> <--> (<vstem> <vinfl>))
(<aux> <--> (<auxstem> <auxinfl>))

```

もちろん、これだけでは動詞・助動詞の相互承接関係を規定することができないから、それを規定するために、例えば次のような規則をインタフェースにして、前終端記号の接続表を作成することになる。

```

(<vstem> <--> (<mata>))
(<auxstem> <--> (<se>))
(<auxstem> <--> (<teitadake>))
(<auxstem> <--> (<ma>))
(<auxinfl> <--> (<se>))
(<auxstem> <--> (<n>))
(<auxstem> <--> (<de>))
(<auxinfl> <--> (<sho>))
(<auxstem> <--> (<u>))
(<auxstem> <--> (<ka>))

```

	matase	teitadake	ma	se	n	de	sho	u	ka
matase	0	1	1	0	0	0	0	0	0
teitadake	0	0	1	0	0	0	0	0	0
ma	0	0	0	1	0	0	0	0	0
se	0	0	0	0	1	0	0	0	0
n	0	0	0	0	0	1	0	0	1
de	0	0	0	0	0	0	1	0	0
sho	0	0	0	0	0	0	0	1	0
u	0	0	0	0	0	0	0	0	1
ka	0	0	0	0	0	0	0	0	0

(1は接続可能、0は接続不可能)

以上の CFG 規則と接続表によって、次のような構造が与えられる。

```

<vp>
|--<vaux>
  |--<vaux>
    | |--<vaux>
      | | |--<vaux>
        | | | |--<vaux>
          | | | | |--<vaux>
            | | | | | |--<verb>
              | | | | | |--<vstem>
                | | | | | |--matase
                  | | | | | |--<aux>
                    | | | | | |--<auxstem>
                      | | | | | |--teitadake
                        | | | | | |--<aux>
                          | | | | | |--<auxstem>
                            | | | | | |--ma
                              | | | | | |--<auxinfl>
                                | | | | | |--se
                                  | | | | |--<aux>
                                    | | | | |--n
                                      | | | |--<aux>
                                        | | | |--<auxstem-desu>
                                          | | | |--de
                                            | | | |--<vinfl-spe-sho>
                                              | | | |--sho
                                                | |--<aux-intn>
                                                  |--u
                                                    |--<aux-sfp>
                                                      |--ka

```

両方を比べれば、MSLR法による文法記述の方が、規則数もはるかに少なく済み、また、構造も単純でずっと理解しやすいことが分かるだろう。そのうえ、上に挙げた5つのCFG規則ですべての述語をカバーすることができるので、語彙や扱う文がどれほど増えても新しい規則を加える必要がない。これらのことは、大規模な文法を開発するうえで大きなメリットとなる。

今回、文末述語のほかにも後置詞句や数詞の形成規則を改良した。それらについても、同様に規則数の大幅な削減とより一層精密な接続制約の組み込みが可能になった。参考までに、新しい文法の緒元を従来の文法のそれと比べてみる。

文法	規則数	語彙サイズ
旧文法	1809	1010
新文法	858	1010

A.3 おわりに

MSLR法の利点を、新旧の文法を比較することによって説明した。この方法の主な狙いは、一般的な構造記述と局所的な接続制約をそれぞれCFG規則と接続表に振り分けることにあり、これによって、次のようなことが可能になるだろう。

- (1) 音声認識用の CFG 規則と言語解析用の CFG 規則をほとんど同じものにすることができ、両者のインタフェースにまつわる問題を軽減することができる。
- (2) 前終端記号の設定を工夫することによって、精密で柔軟な接続制約を記述することができる。

もちろん、CFG 規則の開発に費やされる労力が軽減される代わりに、接続表を作成する労力が新たに加わるということがある。したがって、MSLR 法の利点を十分に引き出すためには、精密で正確な接続表をいかに効率よく作るかという問題を解決しなければならない。そのために、音声言語データベースのコーパスから半自動的に接続表を抽出するというようなことも試みる必要があろう。

また、接続表に確率値を与えることによって、さらに認識率を上げることができるだろう。

参考文献

- [田中 95] 田中穂積, 李輝, 徳永健伸: “自然言語処理の新しい方法 — LR 表工学の提案 (1)”, 人工知能学会研究会資料, SIG-J-9501-1, pp. 1-8 (1995-12).
- [綾部 96] 綾部寿樹: “複数の接続表の制約の LR 表への組み込み”, 東京工業大学工学部情報工学科卒業論文 (1996).
- [保坂 91] 保坂順子, 竹沢寿幸: “SL-Trans における音声認識のための構文規則の概要”, ATR テクニカルレポート, TR-I-0193 (1991).
- [竹沢 95] 竹沢寿幸, 田代敏久, 衛藤純司: “部分木を単位とする音声認識用日本語文法”, ATR テクニカルレポート. TR-IT-0110 (1995-04).
- [田代 95] 田代敏久, 森元暹: “音声言語処理のための構文解析ツールキット”, 情報処理学会自然言語処理研究会, 106-12, pp. 67-72 (1995-03).

B 付録2: 新規登録単語一覧

品詞	見出し語	標準形	備考
固有名詞	アールシーエービル	アールシーエービル	
固有名詞	アムステルダム営業所	アムステルダム営業所	
固有名詞	アムステルダム街営業所	アムステルダム街営業所	
固有名詞	アメリカ	アメリカ	
固有名詞	アメリカ街	アメリカ街	
固有名詞	アメリカンエクスプレス	アメリカンエクスプレス	
固有名詞	嵐山線	嵐山線	
固有名詞	叡山駅	叡山駅	
固有名詞	エイビス	エイビス	
固有名詞	エルエー	エルエー	
固有名詞	エルエーエックス空港	エルエーエックス空港	
固有名詞	延暦寺	延暦寺	
固有名詞	鎌倉時代	鎌倉時代	
固有名詞	鴨川ホテル	鴨川ホテル	
固有名詞	烏丸口	烏丸口	
固有名詞	烏丸線	烏丸線	
固有名詞	河原町	河原町	
固有名詞	河原町駅	河原町駅	
固有名詞	勸進帳	勸進帳	
固有名詞	祇園	祇園	
固有名詞	如月	如月	
固有名詞	北大路駅	北大路駅	
固有名詞	北大路バスターミナル	北大路バスターミナル	
固有名詞	キャノン	キャノン	
固有名詞	京都駅	京都駅	
固有名詞	京都駅前ホテル	京都駅前ホテル	
固有名詞	京都国際観光旅館	京都国際観光旅館	
固有名詞	京都新阪急ホテル	京都新阪急ホテル	
固有名詞	京都センチュリーホテル	京都センチュリーホテル	
固有名詞	京都ツーリスト	京都ツーリスト	
固有名詞	京都の三名所めぐり	京都の三名所めぐり	ツアー名
固有名詞	京都ホテル	京都ホテル	
固有名詞	清水寺	清水寺	
固有名詞	金閣	金閣	
固有名詞	金閣寺	金閣寺	
固有名詞	銀閣寺	銀閣寺	
固有名詞	近代美術館	近代美術館	
固有名詞	グレイハウンドインターナショナル	グレイハウンドインターナショナル	
固有名詞	グレイハウンドバス	グレイハウンドバス	
固有名詞	京阪	京阪	
固有名詞	京阪電車	京阪電車	
固有名詞	京阪電鉄	京阪電鉄	
固有名詞	京福	京福	
固有名詞	ケーアンドケー運輸	ケーアンドケー運輸	
固有名詞	51ストリート	51ストリート	
固有名詞	五十三番通り	五十三番通り	

品詞	見出し語	標準形
固有名詞	52ストリート	52ストリート
固有名詞	五番街	五番街
固有名詞	サウスフェリー	サウスフェリー
固有名詞	嵯峨駅	嵯峨駅
固有名詞	坂本駅	坂本駅
固有名詞	三十三間堂	三十三間堂
固有名詞	三条	三条
固有名詞	三条駅	三条駅
固有名詞	JR	JR
固有名詞	シェラトンバルハーバーリゾート	シェラトンバルハーバーリゾート
固有名詞	ジオメトロ	ジオメトロ
固有名詞	四条河原町	四条河原町
固有名詞	七条	七条
固有名詞	七条駅	七条駅
固有名詞	シビック	シビック
固有名詞	自由の女神	自由の女神
固有名詞	十番街	十番街
固有名詞	ジュエルナイトツアー	ジュエルナイトツアー
固有名詞	手工芸センター	手工芸センター
固有名詞	新大阪	新大阪
固有名詞	世界貿易センター	世界貿易センター
固有名詞	セントコロンバス駅	セントコロンバス駅
固有名詞	ソーホー	ソーホー
固有名詞	ソルトレイクシティー	ソルトレイクシティー
固有名詞	太平洋	太平洋
固有名詞	宝ヶ池プリンスホテル	宝ヶ池プリンスホテル
固有名詞	千鳥	千鳥
固有名詞	チャイナタウン	チャイナタウン
固有名詞	中京	中京
固有名詞	東寺	東寺
固有名詞	ナイアガラ	ナイアガラ
固有名詞	成田	成田
固有名詞	二条城	二条城
固有名詞	日本	日本
固有名詞	ニュー京都ホテル	ニュー京都ホテル
固有名詞	ニューヨーク観光局	ニューヨーク観光局
固有名詞	ニューヨーク店	ニューヨーク店
固有名詞	ニューヨークホテル	ニューヨークホテル
固有名詞	仁和寺	仁和寺
固有名詞	パークアベニュー	パークアベニュー
固有名詞	博多	博多
固有名詞	畑中	畑中
固有名詞	バッテリー公園	バッテリー公園
固有名詞	ハドソン川	ハドソン川
固有名詞	浜大津駅	浜大津駅
固有名詞	パワリー	パワリー
固有名詞	比叡山	比叡山
固有名詞	ひかり	ひかり

品詞	見出し語	標準形
固有名詞	VISA	VISA
固有名詞	平安神宮	平安神宮
固有名詞	ブロードウェイ	ブロードウェイ
固有名詞	ベニンシュラニューヨーク	ベニンシュラニューヨーク
固有名詞	ポートオーソリティーバスターミナル	ポートオーソリティーバスターミナル
固有名詞	ホテルニューヨーク	ホテルニューヨーク
固有名詞	ホテルプラザ	ホテルプラザ
固有名詞	ホテルプラザ通り	ホテルプラザ通り
固有名詞	マイアミ空港	マイアミ空港
固有名詞	マジソンスクエアガーデン	マジソンスクエアガーデン
固有名詞	マジソンスクエア公園	マジソンスクエア公園
固有名詞	マジソンホテル	マジソンホテル
固有名詞	マツダ	マツダ
固有名詞	マンハッタン	マンハッタン
固有名詞	南座前	南座前
固有名詞	ミルバーン	ミルバーン
固有名詞	八坂会館	八坂会館
固有名詞	八坂神社	八坂神社
固有名詞	ユニオンスクエア	ユニオンスクエア
固有名詞	ラスベガス	ラスベガス
固有名詞	リトルイタリー	リトルイタリー
固有名詞	リバティーアイランド	リバティーアイランド
固有名詞	竜安寺	竜安寺
固有名詞	竜安寺前	竜安寺前
固有名詞	ロキシーアンドネルズビル	ロキシーアンドネルズビル
固有名詞	鹿王院	鹿王院
固有名詞	ロサンジェルス支店	ロサンジェルス支店
固有名詞	ロスアンジェルス	ロスアンジェルス
固有名詞	ロックフェラーセンター	ロックフェラーセンター
固有名詞	ワトソンクリニック	ワトソンクリニック
固有名詞	ワトソンズクリニック	ワトソンズクリニック
固有名詞	京都	京都
固有名詞	精華町	精華町
固有名詞	中京	中京
固有名詞	寺町通り	寺町通り
固有名詞	二条北	二条北
サ変名詞	安心	安心
サ変名詞	位置	位置
サ変名詞	移動	移動
サ変名詞	飲食	飲食
サ変名詞	受付	受付
サ変名詞	運転	運転
サ変名詞	営業	営業
サ変名詞	お届け	お届け
サ変名詞	オン	オン
サ変名詞	開始	開始
サ変名詞	外出	外出
サ変名詞	貸し出し	貸し出し

品詞	見出し語	標準形
サ変名詞	我慢	我慢
サ変名詞	鑑賞	鑑賞
サ変名詞	記念	記念
サ変名詞	クリーニング	クリーニング
サ変名詞	見学	見学
サ変名詞	公演	公演
サ変名詞	交換	交換
サ変名詞	合計	合計
サ変名詞	交通	交通
サ変名詞	合流	合流
サ変名詞	混乱	混乱
サ変名詞	散歩	散歩
サ変名詞	持参	持参
サ変名詞	実演	実演
サ変名詞	質問	質問
サ変名詞	指定	指定
サ変名詞	集合	集合
サ変名詞	修理	修理
サ変名詞	出張	出張
サ変名詞	循環	循環
サ変名詞	紹介	紹介
サ変名詞	ショッピング	ショッピング
サ変名詞	心配	心配
サ変名詞	正座	正座
サ変名詞	精算	精算
サ変名詞	清掃	清掃
サ変名詞	説教	説教
サ変名詞	設備	設備
サ変名詞	選択	選択
サ変名詞	代表	代表
サ変名詞	旅	旅
サ変名詞	ダンス	ダンス
サ変名詞	注意	注意
サ変名詞	挑戦	挑戦
サ変名詞	調理	調理
サ変名詞	直通	直通
サ変名詞	ツアー	ツアー
サ変名詞	追加	追加
サ変名詞	都合	都合
サ変名詞	提供	提供
サ変名詞	デコレーション	デコレーション
サ変名詞	手伝い	手伝い
サ変名詞	テーブルコーディネート	テーブルコーディネート
サ変名詞	搭乗	搭乗
サ変名詞	ドライブ	ドライブ
サ変名詞	乗り換え	乗り換え
サ変名詞	入場	入場
サ変名詞	拝観	拝観

品詞	見出し語	標準形
サ変名詞	発券	発券
サ変名詞	引き渡し	引き渡し
サ変名詞	復唱	復唱
サ変名詞	フライト	フライト
サ変名詞	返却	返却
サ変名詞	放送	放送
サ変名詞	防音	防音
サ変名詞	保証	保証
サ変名詞	補償	補償
サ変名詞	待ち合わせ	待ち合わせ
サ変名詞	満喫	満喫
サ変名詞	ミーティング	ミーティング
サ変名詞	迷惑	迷惑
サ変名詞	メモ	メモ
サ変名詞	免責	免責
サ変名詞	申し込み	申し込み
サ変名詞	要望	要望
サ変名詞	了承	了承
サ変名詞	レンタル	レンタル
サ変名詞	割引	割引
サ変名詞	割り増し	割り増し
サ変名詞	引き渡し	引き渡し
形容名詞	安全	安全
形容名詞	一緒	一緒
形容名詞	嫌	嫌
形容名詞	可能	可能
形容名詞	完全	完全
形容名詞	簡単	簡単
形容名詞	急	急
形容名詞	きれい	きれい
形容名詞	豪華	豪華
形容名詞	静か	静か
形容名詞	上等	上等
形容名詞	親切	親切
形容名詞	好き	好き
形容名詞	正確	正確
形容名詞	楽しみ	楽しみ
形容名詞	適当	適当
形容名詞	苦手	苦手
形容名詞	不明	不明
形容名詞	便利	便利
形容名詞	有名	有名
形容名詞	余分	余分
形容名詞	楽	楽
普通名詞	アイランド	アイランド
普通名詞	朝	朝
普通名詞	辺り	辺り
普通名詞	後	後

品詞	見出し語	標準形
普通名詞	アンティーク	アンティーク
普通名詞	案内書	案内書
普通名詞	以下	以下
普通名詞	行き方	行き方
普通名詞	幾分	幾分
普通名詞	以上	以上
普通名詞	イタリアン	イタリアン
普通名詞	イタリアンランチ	イタリアンランチ
普通名詞	位置	位置
普通名詞	一流	一流
普通名詞	一緒	一緒
普通名詞	以内	以内
普通名詞	入り	入り
普通名詞	イングリッシュスタイル	イングリッシュスタイル
普通名詞	ウークエンドレディースパック	ウークエンドレディースパック
普通名詞	ウインドウショッピング	ウインドウショッピング
普通名詞	ウエスタンスタイル	ウエスタンスタイル
普通名詞	受け付けカウンター	受け付けカウンター
普通名詞	うち	うち
普通名詞	海	海
普通名詞	梅	梅
普通名詞	裏	裏
普通名詞	裏手	裏手
普通名詞	売り切れ	売り切れ
普通名詞	運転手	運転手
普通名詞	駅	駅
普通名詞	エコノミーカー	エコノミーカー
普通名詞	エコノミーサイズ	エコノミーサイズ
普通名詞	エレベーター	エレベーター
普通名詞	宴会部	宴会部
普通名詞	オートマチック	オートマチック
普通名詞	多く	多く
普通名詞	音	音
普通名詞	オレンジ	オレンジ
普通名詞	会議室	会議室
普通名詞	外国	外国
普通名詞	懐石	懐石
普通名詞	会席料理	会席料理
普通名詞	ガイドブック	ガイドブック
普通名詞	カウンター	カウンター
普通名詞	各種	各種
普通名詞	学生	学生
普通名詞	数	数
普通名詞	ガソリン	ガソリン
普通名詞	型	型
普通名詞	格好	格好
普通名詞	角	角
普通名詞	歌舞伎	歌舞伎

品詞	見出し語	標準形
普通名詞	紙	紙
普通名詞	かも	かも
普通名詞	代わり	代わり
普通名詞	間隔	間隔
普通名詞	観光客	観光客
普通名詞	観光局	観光局
普通名詞	感じ	感じ
普通名詞	看板	看板
普通名詞	キー	キー
普通名詞	北側	北側
普通名詞	キッチン	キッチン
普通名詞	切符	切符
普通名詞	ギャラリー	ギャラリー
普通名詞	ギャランテイ	ギャランテイ
普通名詞	キャンドル	キャンドル
普通名詞	休館日	休館日
普通名詞	救急医	救急医
普通名詞	救急車	救急車
普通名詞	狂言	狂言
普通名詞	局番	局番
普通名詞	金	金
普通名詞	金額	金額
普通名詞	キングサイズ	キングサイズ
普通名詞	近所	近所
普通名詞	空港	空港
普通名詞	クリーニングカウンター	クリーニングカウンター
普通名詞	グリーン車	グリーン車
普通名詞	クリスマス	クリスマス
普通名詞	クルーザー	クルーザー
普通名詞	グループ	グループ
普通名詞	車	車
普通名詞	クレジットカード	クレジットカード
普通名詞	ケーブル	ケーブル
普通名詞	ケーブルカー	ケーブルカー
普通名詞	ゲーム	ゲーム
普通名詞	劇場	劇場
普通名詞	景色	景色
普通名詞	化粧品	化粧品
普通名詞	化粧用	化粧用
普通名詞	現地	現地
普通名詞	子	子
普通名詞	コーヒー	コーヒー
普通名詞	高級品	高級品
普通名詞	航空便	航空便
普通名詞	紅茶	紅茶
普通名詞	国際	国際
普通名詞	五重塔	五重塔
普通名詞	琴	琴

品詞	見出し語	標準形
普通名詞	子供	子供
普通名詞	好み	好み
普通名詞	コンセント	コンセント
普通名詞	コンチネンタルスタイル	コンチネンタルスタイル
普通名詞	コンピューター	コンピューター
普通名詞	最後	最後
普通名詞	サウナ	サウナ
普通名詞	作品	作品
普通名詞	酒	酒
普通名詞	座禅	座禅
普通名詞	座敷	座敷
普通名詞	サッカー	サッカー
普通名詞	茶道	茶道
普通名詞	サラダ	サラダ
普通名詞	山地	山地
普通名詞	市	市
普通名詞	シェフ	シェフ
普通名詞	市外	市外
普通名詞	市街	市街
普通名詞	市街地	市街地
普通名詞	自信	自信
普通名詞	施設	施設
普通名詞	時代物	時代物
普通名詞	下	下
普通名詞	自宅	自宅
普通名詞	自動車	自動車
普通名詞	品	品
普通名詞	市内	市内
普通名詞	品数	品数
普通名詞	芝居	芝居
普通名詞	自分	自分
普通名詞	ジム	ジム
普通名詞	ジャケット	ジャケット
普通名詞	シャーベット	シャーベット
普通名詞	シャンパン	シャンパン
普通名詞	種	種
普通名詞	住職	住職
普通名詞	週末	週末
普通名詞	宿泊日	宿泊日
普通名詞	状況	状況
普通名詞	条件	条件
普通名詞	乗車券	乗車券
普通名詞	症状	症状
普通名詞	消費税	消費税
普通名詞	情報	情報
普通名詞	正面	正面
普通名詞	食器	食器
普通名詞	所要	所要

品詞	見出し語	標準形
普通名詞	新幹線	新幹線
普通名詞	スクランブルエッグ	スクランブルエッグ
普通名詞	薦め	薦め
普通名詞	スタンダードタイプ	スタンダードタイプ
普通名詞	スタンダードツイン	スタンダードツイン
普通名詞	ストリート	ストリート
普通名詞	すべて	すべて
普通名詞	スベル	スベル
普通名詞	請求書	請求書
普通名詞	設備	設備
普通名詞	全員	全員
普通名詞	前菜	前菜
普通名詞	前日	前日
普通名詞	専用車	専用車
普通名詞	早朝	早朝
普通名詞	外	外
普通名詞	それなり	それなり
普通名詞	代金	代金
普通名詞	滞在地	滞在地
普通名詞	対人	対人
普通名詞	対物	対物
普通名詞	ダウンタウン	ダウンタウン
普通名詞	タオル	タオル
普通名詞	宅	宅
普通名詞	タクシー	タクシー
普通名詞	竹	竹
普通名詞	出し物	出し物
普通名詞	ただいま	ただいま
普通名詞	ダブルサイズ	ダブルサイズ
普通名詞	単位	単位
普通名詞	短期	短期
普通名詞	誕生日	誕生日
普通名詞	治安	治安
普通名詞	地下鉄	地下鉄
普通名詞	チキン	チキン
普通名詞	知人	知人
普通名詞	地図	地図
普通名詞	地方	地方
普通名詞	中間	中間
普通名詞	昼食	昼食
普通名詞	中心地	中心地
普通名詞	長期	長期
普通名詞	チョコレートクリーム	チョコレートクリーム
普通名詞	付き	付き
普通名詞	勤め	勤め
普通名詞	テーブル	テーブル
普通名詞	テーブル	テーブル
普通名詞	テーブル席	テーブル席

品詞	見出し語	標準形
普通名詞	程度	程度
普通名詞	定評	定評
普通名詞	停留所	停留所
普通名詞	手元	手元
普通名詞	テラス	テラス
普通名詞	デラックス	デラックス
普通名詞	デラックスツイン	デラックスツイン
普通名詞	テレビ	テレビ
普通名詞	点	点
普通名詞	天気	天気
普通名詞	電源	電源
普通名詞	同時	同時
普通名詞	当日	当日
普通名詞	道路	道路
普通名詞	特典	特典
普通名詞	特別地方消費税	特別地方消費税
普通名詞	年	年
普通名詞	途中	途中
普通名詞	特急券	特急券
普通名詞	隣	隣
普通名詞	徒歩	徒歩
普通名詞	トマトジュース	トマトジュース
普通名詞	トラベルセット	トラベルセット
普通名詞	中身	中身
普通名詞	なし	なし
普通名詞	何事	何事
普通名詞	斜め	斜め
普通名詞	南部	南部
普通名詞	何名	何名
普通名詞	西側	西側
普通名詞	日時	日時
普通名詞	日没	日没
普通名詞	日程	日程
普通名詞	日本語	日本語
普通名詞	庭	庭
普通名詞	人気	人気
普通名詞	願い	願い
普通名詞	ネクタイ	ネクタイ
普通名詞	年	年
普通名詞	年中	年中
普通名詞	年齢	年齢
普通名詞	残り	残り
普通名詞	乗り継ぎ	乗り継ぎ
普通名詞	パーティ	パーティ
普通名詞	初め	初め
普通名詞	パジャマパーティー	パジャマパーティー
普通名詞	パジャマ	パジャマ
普通名詞	バスターミナル	バスターミナル

品詞	見出し語	標準形
普通名詞	バスツアー	バスツアー
普通名詞	話	話
普通名詞	花束	花束
普通名詞	速さ	速さ
普通名詞	晩	晩
普通名詞	半額	半額
普通名詞	パンフレット	パンフレット
普通名詞	ビール	ビール
普通名詞	東側	東側
普通名詞	飛行機	飛行機
普通名詞	ビジネス	ビジネス
普通名詞	美術館	美術館
普通名詞	左	左
普通名詞	左側	左側
普通名詞	日付	日付
普通名詞	必要	必要
普通名詞	人	人
普通名詞	標準	標準
普通名詞	昼	昼
普通名詞	昼間	昼間
普通名詞	広さ	広さ
普通名詞	便	便
普通名詞	ファーストクラス	ファーストクラス
普通名詞	夫婦	夫婦
普通名詞	フェリー	フェリー
普通名詞	付近	付近
普通名詞	服	服
普通名詞	普通車	普通車
普通名詞	ブティック	ブティック
普通名詞	船	船
普通名詞	フライトナンバー	フライトナンバー
普通名詞	フランス料理	フランス料理
普通名詞	フリーマイリッジ	フリーマイリッジ
普通名詞	フリーマイリッジ料金制	フリーマイリッジ料金制
普通名詞	フリーマイレッジ	フリーマイレッジ
普通名詞	フル	フル
普通名詞	フレンチ	フレンチ
普通名詞	フレンチランチ	フレンチランチ
普通名詞	ふろ	ふろ
普通名詞	フロア	フロア
普通名詞	ブロック	ブロック
普通名詞	フロントデスク	フロントデスク
普通名詞	雰囲気	雰囲気
普通名詞	別	別
普通名詞	ベランダ	ベランダ
普通名詞	宝石店	宝石店
普通名詞	方式	方式
普通名詞	保険	保険

品詞	見出し語	標準形
普通名詞	掘り出し物	掘り出し物
普通名詞	本日	本日
普通名詞	マーケット	マーケット
普通名詞	毎週	毎週
普通名詞	毎日	毎日
普通名詞	マイル	マイル
普通名詞	前	前
普通名詞	町	町
普通名詞	間違い	間違い
普通名詞	間違え	間違え
普通名詞	松	松
普通名詞	末日	末日
普通名詞	窓口	窓口
普通名詞	満タン	満タン
普通名詞	ミーティングルーム	ミーティングルーム
普通名詞	見ごろ	見ごろ
普通名詞	ミサ	ミサ
普通名詞	店	店
普通名詞	道	道
普通名詞	道順	道順
普通名詞	土産	土産
普通名詞	無休	無休
普通名詞	無料	無料
普通名詞	目	目
普通名詞	目玉焼き	目玉焼き
普通名詞	免許証	免許証
普通名詞	免税店	免税店
普通名詞	面倒	面倒
普通名詞	最寄り	最寄り
普通名詞	門限	門限
普通名詞	役	役
普通名詞	夜景	夜景
普通名詞	薬剤師	薬剤師
普通名詞	休み	休み
普通名詞	薬局	薬局
普通名詞	夕食	夕食
普通名詞	予定日	予定日
普通名詞	ランチ	ランチ
普通名詞	ランチメニュー	ランチメニュー
普通名詞	リムジンバス	リムジンバス
普通名詞	料理	料理
普通名詞	利用日	利用日
普通名詞	両方	両方
普通名詞	旅館	旅館
普通名詞	ルートエム	ルートエム
普通名詞	ルームナンバー	ルームナンバー
普通名詞	歴史	歴史
普通名詞	レンタカー	レンタカー

品詞	見出し語	標準形	備考
普通名詞	ロッカー	ロッカー	
普通名詞	ワイン	ワイン	
普通名詞	わき	わき	
普通名詞	和風	和風	
普通名詞	内	内	
普通名詞	あさって	あさって	
普通名詞	今週	今週	
普通名詞	今度	今度	
普通名詞	今晚	今晚	
普通名詞	先程	先程	
普通名詞	毎朝	毎朝	
普通名詞	明晩	明晩	
普通名詞	翌朝	翌朝	
普通名詞	いかほど	いかほど	
普通名詞	全員	全員	
普通名詞	夕方	夕方	
普通名詞	夜	夜	
数詞	三	三	「み」(「さん」の「三」は既にあり)
日時	一日	一日	「いちにち」(「ついたち」の「一日」は既にあり)
代名詞	あたし	あたし	
代名詞	あなた	あなた	
代名詞	それ	それ	
代名詞	僕	僕	
代名詞	そっち	そっち	
代名詞	いくつ	いくつ	
代名詞	どっち	どっち	
本動詞	書	書	
本動詞	風邪を引	風邪を引	
本動詞	続	続	
本動詞	届	届	
本動詞	寛	寛	
本動詞	いた	致	
本動詞	起こ	起こ	
本動詞	下ろ	下ろ	
本動詞	貸	貸	
本動詞	返	返	
本動詞	貸し出	貸し出	
本動詞	繰り返	繰り返	
本動詞	引き落と	引き落と	
本動詞	見下ろ	見下ろ	
本動詞	戻	戻す	
本動詞	渡	渡	
本動詞	立	立	
本動詞	並	並	
本動詞	呼	呼	
本動詞	喜	喜	
本動詞	組	組	
本動詞	済	済	

品詞	見出し語	標準形
本動詞	申し込	申し込
本動詞	読	読
本動詞	上が	上が
本動詞	預か	預か
本動詞	当た	当た
本動詞	いらっしゃ	いらっしゃ
本動詞	要	要
本動詞	受け取	受け取
本動詞	映	映
本動詞	送	送
本動詞	起こ	起こ
本動詞	おっしゃ	おっしゃ
本動詞	終わ	終わ
本動詞	帰	帰
本動詞	限	限
本動詞	決ま	決ま
本動詞	下さ	下さ
本動詞	困	困
本動詞	ご覧にな	ご覧にな
本動詞	異な	異な
本動詞	差し掛か	差し掛か
本動詞	仕上が	仕上が
本動詞	知	知
本動詞	突き当た	突き当た
本動詞	突っ切	突っ切
本動詞	詰ま	詰ま
本動詞	通	通
本動詞	なさ	なさ
本動詞	始ま	始ま
本動詞	曲が	曲が
本動詞	回	回
本動詞	戻	戻
本動詞	因	因
本動詞	渡	渡
本動詞	下さ	下さ
本動詞	行な	行な
本動詞	込み合	込み合
本動詞	支払	支払
本動詞	住ま	住ま
本動詞	沿	沿
本動詞	手伝	手伝
本動詞	整	整
本動詞	払	払
本動詞	間に合	間に合
本動詞	浴び	浴び
本動詞	合わせ	合わせ
本動詞	行け	行け
本動詞	承れ	承れ

品詞	見出し語	標準形
本動詞	受け付け	受け付け
本動詞	遅れ	遅れ
本動詞	行われ	行われ
本動詞	下り	下り
本動詞	替え	替え
本動詞	兼ね	兼ね
本動詞	感じ	感じ
本動詞	消え	消え
本動詞	着替え	着替え
本動詞	聞け	聞け
本動詞	気を付け	気を付け
本動詞	決め	決め
本動詞	加え	加え
本動詞	越え	越え
本動詞	ご覧になれ	ご覧になれ
本動詞	差し上げ	差し上げ
本動詞	過ぎ	過ぎ
本動詞	薦め	薦め
本動詞	済ませ	済ませ
本動詞	楽しめ	楽しめ
本動詞	尋ね	尋ね
本動詞	付け	付け
本動詞	連れ	連れ
本動詞	出掛け	出掛け
本動詞	届け	届け
本動詞	長すぎ	長すぎ
本動詞	寝	寝
本動詞	乗れ	乗れ
本動詞	控え	控え
本動詞	含め	含め
本動詞	間に合わせ	間に合わせ
本動詞	申し上げ	申し上げ
本動詞	申しつけ	申しつけ
本動詞	戻れ	戻れ
本動詞	分かれ	分かれ
本動詞	忘れ	忘れ
本動詞	座れ	座れ
本動詞	取れ	取れ
本動詞	見え	見え
本動詞	もらえ	もらえ
本動詞	し	する
本動詞	来	来
本動詞	来る	来る
形容詞	美し	美し
形容詞	多	多
形容詞	大き	大き
形容詞	黄色	黄色
形容詞	細か	細か

品詞	見出し語	標準形	備考
形容詞	しょうがな	しょうがな	
形容詞	狭	狭	
形容詞	樂し	樂し	
形容詞	近	近	
形容詞	遠	遠	
形容詞	長	長	
形容詞	早	早	
形容詞	速	速	
形容詞	広	広	
形容詞	ふさわし	ふさわし	
形容詞	古	古	
形容詞	欲し	欲し	
形容詞	短	短	
形容詞	珍し	珍し	
形容詞	面倒くさ	面倒くさ	
形容詞	分かりやす	分かりやす	
形容詞	い	い	
形容詞	よさ	よさ	特殊なもの(「よさそう」の「よさ」)
副詞	後	後	
副詞	一応	一応	
副詞	一番	一番	
副詞	一瞬	一瞬	
副詞	一体	一体	
副詞	いっぱい	いっぱい	
副詞	うっかり	うっかり	
副詞	必ず	必ず	
副詞	かなり	かなり	
副詞	きっと	きっと	
副詞	最近	最近	
副詞	さっき	さっき	
副詞	十分	十分	
副詞	すぐ	すぐ	
副詞	せっかく	せっかく	
副詞	ぜひ	ぜひ	
副詞	全然	全然	
副詞	そのまま	そのまま	
副詞	それほど	それほど	
副詞	そんなに	そんなに	
副詞	大変に	大変に	
副詞	たくさん	たくさん	
副詞	多少	多少	
副詞	例えば	例えば	
副詞	直接	直接	
副詞	ついでに	ついでに	
副詞	どうしても	どうしても	
副詞	どうぞ	どうぞ	
副詞	どうやって	どうやって	
副詞	どうやっても	どうやっても	

品詞	見出し語	標準形
副詞	特に	特に
副詞	とつても	とつても
副詞	とても	とても
副詞	どのくらい	どのくらい
副詞	なお	なお
副詞	なるべく	なるべく
副詞	念のため	念のため
副詞	はっきり	はっきり
副詞	ぱっと	ぱっと
副詞	ぶらぶら	ぶらぶら
副詞	誠に	誠に
副詞	また	また
副詞	まっすぐ	まっすぐ
副詞	もう少し	もう少し
副詞	もうちょっと	もうちょっと
副詞	もっと	もっと
副詞	やっぱり	やっぱり
副詞	やはり	やはり
副詞	ゆっくり	ゆっくり
副詞	ゆったり	ゆったり
副詞	ゆったりと	ゆったりと
副詞	よく	よく
副詞	おそらく	おそらく
副詞	失礼ですが	失礼ですが
副詞	申し訳ありませんが	申し訳ありませんが
連体詞	大きな	大きな
連体詞	このような	このような
連体詞	どの	どの
連体詞	どんな	どんな
連体詞	あの	あの
接続詞	そういたしますと	そういたしますと
接続詞	そうすると	そうすると
接続詞	それが	それが
接続詞	それじゃあ	それじゃあ
接続詞	それだったら	それだったら
接続詞	それとね	それとね
接続詞	それに	それに
接続詞	ただ	ただ
接続詞	つぎに	つぎに
接続詞	ですが	ですが
接続詞	ですけれども	ですけれども
接続詞	でも	でも
接続詞	なので	なので
接続詞	または	または
接続詞	それと	それと
感動詞	いや	いや
感動詞	え	え
感動詞	すいません	すいません

品詞	見出し語	標準形	備考
感動詞	どうぞ	どうぞ	
感動詞	また	また	
感動詞	ごめんなさい	ごめんなさい	
感動詞	こんにちわ	こんにちわ	
感動詞	さようなら	さようなら	
感動詞	さよなら	さよなら	
感動詞	失礼します	失礼します	
感動詞	すいませんでした	すいませんでした	
感動詞	すみませんでした	すみませんでした	
感動詞	そう	そう	
感動詞	そっか	そっか	
感動詞	そっかあ	そっかあ	
感動詞	申し訳ありません	申し訳ありません	
感動詞	申し訳ない	申し訳ない	
感動詞	よろしく	よろしく	
間投詞	いや	いや	
間投詞	うーんと	うーんと	
間投詞	うーんとー	うーんとー	
間投詞	えーあの	えーあの	
間投詞	えーっとー	えーっとー	
間投詞	えーっとね	えーっとね	
間投詞	えーとです	えーとです	
間投詞	えっ	えっ	
間投詞	えっとー	えっとー	
間投詞	おう	おう	
間投詞	そう	そう	
間投詞	そうだなー	そうだなー	
間投詞	そうですね	そうですね	
間投詞	っとー	っとー	
間投詞	ふーん	ふーん	
間投詞	ま	ま	
間投詞	まあ	まあ	
間投詞	んと	んと	
人名	トーマス	トーマス	
人名	ネルソン	ネルソン	
助動詞	じゃあ	じゃあ	「～じゃあない～」の「じゃあ」
助動詞	させ	させ	
助動詞	られ	られ	
助動詞	てくれ	てくれ	
助動詞	てい	てい	
助動詞	と	と	
助動詞	でくださ	でくださ	
助動詞	ちゃ	ちゃ	
助動詞	たらよ	たらよ	
助動詞	たらい	たらい	
助動詞	てもかまいません	てもかまいません	
助動詞	なければいけません	なければいけません	
助動詞	ばい	ばい	

品詞	見出し語	標準形
助動詞	ばよ	ばよ
助動詞	だ	だ
助動詞	かもしれな	かもしれな
助動詞	とよ	とよ
助動詞	かもしれません	かもしれません
助動詞	つもり	つもり
助動詞	つもりで	つもりで
助動詞	ようで	ようで
助動詞	そう	そう
助動詞	であれ	であれ
補助動詞	すれ	する
補助動詞	なさ	なさ
補助動詞	願	願
補助動詞	になれ	になれ
格助詞	の	の
格助詞	にかけて	にかけて
格助詞	に関して	に関して
格助詞	につき	につき
格助詞	によって	によって
副助詞	ごと	ごと
副助詞	ずつ	ずつ
副助詞	とう	とう
副助詞	とか	とか
副助詞	とも	とも
副助詞	なんて	なんて
副助詞	のみ	のみ
副助詞	ばかり	ばかり
連体助詞	っていう	っていう
連体助詞	といった	といった
連体助詞	のような	のような
並立助詞	や	や
接続助詞	けれど	けれど
接続助詞	ても	ても
接続助詞	ながら	ながら
接続助詞	し	し
接続助詞	なら	なら
接続助詞	のに	のに
接続助詞	ので	ので
接続助詞	んで	んで
接続助詞	のに	のに
接続助詞	ものですから	ものですから
終助詞	っけ	っけ
終助詞	って	って
終助詞	なあ	なあ
終助詞	の	の
引用助詞	かどうか	かどうか
引用助詞	って	って

品詞	見出し語	標準形
接頭辞	各	各
接頭辞	何	何
接頭辞	毎	毎
接頭辞	約	約
接尾辞	都	都
接尾辞	道	道
接尾辞	府	府
接尾辞	県	県
接尾辞	区	区
接尾辞	市	市
接尾辞	区	区
接尾辞	たち	たち
接尾辞	ら	ら
接尾辞	セント	セント
接尾辞	行き	行き
接尾辞	街	街
接尾辞	側	側
接尾辞	沿い	沿い
接尾辞	代	代
接尾辞	付き	付き
接尾辞	用	用
接尾辞	通り	通り
接尾辞	者	者
接尾辞	中	中
接尾辞	過ぎ	過ぎ
接尾辞	末	末
接尾辞	目	目
接尾辞	当たり	当たり
接尾辞	置き	置き
接尾辞	増し	増し
接尾辞	当たり	当たり
接尾辞	さ	さ
接尾辞	コース	コース
接尾辞	台	台
接尾辞	度	度
接尾辞	部	部
接尾辞	分	分
接尾辞	メートル	メートル
接尾辞	歳	歳
接尾辞	回	回
接尾辞	品	品
接尾辞	本	本
接尾辞	日	日
活用語尾	し	し
活用語尾	す	す