

TR-1-0217

日本語形態素解析の細則
Detailed Rules for Japanese
Morphological Analysis

| | |
|----------------|-----------------|
| 江原 暉将 | 立川 映子* |
| Terumasa EHARA | Eiko TACHIKAWA* |
| 山田 久子* | 田中 麻也子* |
| Hisako YAMADA* | Mayako TANAKA* |

1991.05.28

内容梗概

ATR対話データベース (ADD) のうち、日本語形態素解析の方法の細則を記述する。

Abstract

Detailed rules for Japanese morphological analysis are described. This analysis is used in the construction of ATR Dialogue Database (ADD).

© ATR Interpreting Telephony Research Laboratories
© ATR 自動翻訳電話研究所

-
- * (株) 東洋情報システム
 - * Toyo Information Systems co., ltd.

A T R 対話データベースを作成するために、言語解析の1つとして、日本語形態素解析を実施している。これは、文献〔篠崎〕に基本的には従っているが、作業途中で種々の問題点・疑問点が発生し、その都度、解析のための細則を決定し、対処することによって現在に至っている。それらの内容は、これまで逐次記録がされているので、これらをまとめてテクニカルレポートとすることにする。次ページからの内容は以下の通りである。

| | ページ |
|-----------------------------------|-----|
| 言語情報データファイル作成問題点報告（1988年9月）・・・ | 1 |
| 言語情報データファイル検査結果報告（1988年9月）・・・ | 20 |
| 形態素情報データファイル（2）品質管理報告（1989年2月）・・・ | 34 |
| 形態素情報データファイル（3）品質管理報告（1989年8月）・・・ | 63 |
| 形態素情報データファイル（4）品質管理報告（1990年2月）・・・ | 100 |
| 形態素情報データファイル（5）品質管理報告（1990年8月）・・・ | 112 |
| 形態素情報データファイル（6）品質管理報告（1991年2月）・・・ | 160 |

参考文献

- 〔篠崎〕 篠崎直子ほか：形態素情報利用解説書（兼作業マニュアル）、
T R - 0 0 7 7

言語情報データファイル

— 作成問題点報告書 —

1988年 9月

[目次]

作業問題点報告書

| | ページ |
|--------------------------|-----|
| ・ はじめに | 1 |
| ・ 形態素品詞切りについての打ち合せ | 2 |
| [1] 形態素品詞切りについて | 2 |
| [2] 形態素解析システムについて | 2 |
| [3] 資料に基づく詳細 | 2 |
| 1) 助詞について | 2 |
| 2) 固有名詞について | 5 |
| 3) 方言について | 6 |
| 4) 数詞について | 6 |
| 5) その他 | 7 |
| ・ 言い直し文について | 10 |
| ・ 時間について | 10 |
| ・ 日付について | 11 |
| ・ 助詞について | 12 |
| ・ ”御座います”について | 12 |
| ・ 副詞について | 13 |
| ・ 造語について | 13 |
| ・ 外来語の派生語の場合 | 13 |
| ・ その他 | 14 |
| ・ 現時点での未解決問題点について | 15 |
| 1) コールフォーペーパー | 15 |
| 2) 造語の漢語成分について | 15 |
| 3) 固有名詞について | 15 |

はじめに

1. 2ページ以降の問題点は、すでに解決されており、形態素を収集していく過程で実行されています。ただし、問題点解決日以後の作業には、この点が反映されていますが、それ以前の形態素収集ファイルには手を加えていません。
2. 10ページからの問題点につきましては、現在、水野さんに問題点を提出し統一をしてもらっています。従って、現在の形態素解析の修正作業についての問題点は、すべて、水野さんに提出し、新しいマニュアルが出来た時点で、過去に蓄積してきたすべての形態素解析情報付きファイルと共に、一括して修正して行く予定です。

日付け : 88 / 3 / 8

出席者 : 吉本氏 小倉氏 前田氏 篠崎さん
松尾 三木

当社提出の問題点を解決

[1] 形態素切りについて

- ・ 助詞決定 [逆説 → 接続助詞
以外 → 細分
- ・ 固有名詞 [長単位分割
ルール作成 (三木さんに依存)
- ・ 方言 標準語に直し、その分割と同じにする。
↓
コメント欄
- ・ 数詞 分野別ルール設定
- ・ 現状、造語を名詞入力にしているが接尾辞の場合有り。
↓
篠崎さんに調査

[2] 形態素解析システムについて将来的に以下の項目に関してサポートを行いたい。

- ・ 新明解の派生語の追加
- ・ データ項目にコメント欄を設ける。

[3] 資料に基づく詳細

1) 助詞について

ex. 一度は企業に就職しようとも考えました。

と / も or 接続助詞

ex. . . . という言葉の用法とも合致していて興味深い。

と / も ?

(答え) と / も
格助 係助

* 判断の仕方

- ・ "も" を省いて意味が通じるかどうかを考える。
- ・ "も" が省けない場合、意味が逆説の場合の"とも"は接続助詞とする。

ex. 人生に対しては意欲を出そう。

て／は or 接続助詞

(答え) て／は
接助 係助

* 判断の仕方

- ・ ”は” を省いても通じるかどうかを考える。

ex. 他人となったといってもよいであろう。

て／も or 接続助詞

(答え) て／も
接助 係助

ex. 無縁の存在であるとしても

て／も or 接続助詞

(答え) 接続助詞

* 判断の仕方

- ・ ”も” を省いても通じるかどうかを考える。
- ・ 逆説を考える。

ex. その挫折の結果として起こる特殊な敵意を現わす。

と／し／て or 格助詞

(答え) 格助詞

について
において } → これらの格助詞と同じように
における } ”として” と言う格助詞を追加する。

ex. 困難な事態を打開できるとの定評...

格助詞 or 接続助詞

(答え) 格助詞

* 判断の仕方

- ・ 後に ”が” ”を” 以外の格助詞の場合は格助詞とする。

ex. それ自体最も由緒ある関なのだから。

接続詞 or だ／から

(答え) だ／から
助”だ” 接助

* 判断の仕方

- ・ ”だ”で終わることの出来るもの。
- ・ 接続詞は必ず文と文の間にある。

ex. 決心するのに半年かかりました。

接続助詞 or の／に

(答え) の／に
準体 格助

* 判断の仕方

- ・ 逆説を考える。

2) 固有名詞について

- ・ビルや会館などの名前

／ 高知県中村市立中央公民館 ／
固有名詞

ex. その国際会議場までは....

- ・川の名前

／ 四万十川 ／
固有名詞

- ・駅の名前

／ 大阪 ／ 駅 ／ → ／ 大阪駅 ／ とする。
固有名詞 固有名詞

- ・社説または章の名前

ex. 社説「部活の功罪を考える」

六日付の「勇気を出して投稿で意見を」を読み

(答え) 社説／「部活の功罪を考える」／

六日／付／の／「勇気を出して投稿で／
意見／を／」／を／読み／

- ・中国人の名前の読み方

ex. 陳雲

日本語読みで入れるのか？
字によっては変換出来ないものもあるので
どうすれば良いか。
今の所、カタカナで入力している。

(答え) 中国人の名前は日本語読みで入力する。
変換の出来ないものはカタカナ入力する。

- ・その他

趙紫陽党総書記代行は...
英国国教会
イランゲート事件
チェシャイヤー州
ホメイニ政権
M16自動小銃
チャンネル7
寺井精英電気通信大学教授
東京湾横断道路
ジブラルタル海峡トンネル
etc.

- * 判断の仕方

他の人が見ても同じようにカット出来るように、
統一をとって長単位でカットする。

3) 方音について

ex. ほんとに きれい やった で

「勘弁しておくんなせえ。 勘弁して おくんなせえ。

坊ちゃん、あっしゃア、なにも怨み憎みがあって やったんじゃねえんで ご ざん す。

ただもう逃げてえ ば っ か り に、ついこ わ く な っ て、無我 夢中 で や っ た 仕 事 な ん で。

... あっしゃア坊ちゃんに、なんとも 申 訳 の ね え、大それ た こ と を し ち め え ま し た。

ですが、こ う や っ て 今、う ぬ の 犯 し た 罪 の か ど で、こ れ か ら 死 に に 行 く と こ ろ で ご ざ ん す。 あ っ し ゃ ア 死 に て え ん で す。 よ ろ こ ん で 死 に ま す。

だから坊ちゃん、ど う か 可 哀 想 な 野 郎 だ と 思 い な す っ て、あ っ し の こ た ア、勘 弁 し て や っ て お く ん な せ ま し。 お 願 え で ご ざ ん す。」

(答え) 標準語になおして辞書にあるものは、“漢字”のところで原形になおして解析をする。

※ コメント欄が必要である。

4) 数詞について

ex. 郵便 / 番号 / 530 /

ご う さ ん れ い or ご さ ん れ い

(答え) ごさんれい

読みは、発音通りに入力する。
ただし、“ひらがな”のところには辞書に表記されている通りに入力する。

(例) データ 2
 ひらがな にい → に
 漢字 2

ex. 百万 / 平方 / キロ /

接尾語 接尾語 になるのか?

(答え) 接尾語 + 接尾語 の二つにする。

5) その他

ex. 仕事先の方へ送って下さいませんか。

補助動詞命令形 or 補助動詞連用形のイ音便

(答え) 補助動詞連用形 (イ音便ではない)

* 判断の仕方

- 1. おっしゃい
 - 2. 下さい
 - 3. いらっしゃい
 - 4. 御座い
- } → + ます

これら4つは不規則な変化をする。

- 3. の”下さい”が補助動詞命令形となるのは

(例) 下さい。

のように最後に来ている場合だけである。

- 1. 3. 4. の場合は、動詞か補助動詞かを判断する。

ex. スリムが一番と信じ...

名詞 or 副詞

(答え) 名詞 (格助詞 ”と”にかかる)

* 判断の仕方

- 後にかかる品詞による。

ex. 言葉足らずの...

足ら/ず or 名詞

(答え) 接尾辞

* 判断の仕方

- ”足らず” (造語) である。
従って、造語は接頭辞、接尾辞になりうる。
- ／舌たらず／ 辞書にあるので一つで名詞とする。
／言葉／たらず／ 辞書に無いので名詞+接尾辞とする。

ex. トンボ撮りを楽しみました。

撮り or 取り

(答え) 取り

- ・ 辞書の表記に従う

ex. 命令したにも拘らず....

名詞?

(答え) に/も/拘らず/...

- ・ 連語扱いで、データ上はカットする。
このような場合は必ずその記録を残すようにする。

ex. 親子は他人ではない。

格助詞 or 助動詞(だ)

(答え) 助動詞(だ)

ex. 遠慮がないばかりでなく

格助詞 or 接続助詞 or 助動詞(だ)

(答え) 助動詞(だ)

ex. 凡ての人
凡ての人間
状況は凡て] 名詞 or 副詞

(答え) 名詞

ex. およそ、おおよそ の品詞

* 判断の仕方

- ・ 後が格助詞の場合は名詞になる。
- ・ 副詞は用言を修飾するので後に格助詞はあまり来ない。

ex. しかし、この無批判に摂取採用できるものは、

者 or 物 or もの

(答え) 物

* 判断の仕方

- ・ 者 → 人間の場合
- ・ 物 → それ以外

ex. 圧力が高まるにつれ、

"連れて" (名詞) の変形?
または、"連れる" の連用形?

(答え) "連れる" の連用形

ex. 非人称的な...

接頭語 or 名詞

(答え) 接頭語

ex. 息を吹き返しつつある

補助動詞 or 動詞

(答え) 補助動詞

ex. 京都で一泊する

数詞+接尾語 or サ名詞

(答え) 数詞+接尾語+補助動詞

・この場合だけ例外で

／ 一 ／ 泊 ／ する ／
数 接尾 補動

ex. 雨が降りそうである

助動 補動 or 動詞

(答え) 助動詞+補助動詞

・補助動詞には

1. 「動詞+テ」に続くもの。
2. 動詞の連用形に続くもの。
3. 否定の助動詞などを伴って代動詞として用いられるもの。
4. で/ある/
助 補動 の場合

従ってこれは 4. の場合に相当する。

ex. 恐れ入ります

感動詞 or 名詞 or 切る

(答え) 感動詞

・恐れ入ります → 済みません (感動詞) の老人語

従って、恐れ入ります も感動詞とする。

1. 言い直し文について

- 括弧のある文は、どもり又は 言い直し文である。

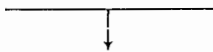
ex. あのー、(クレ)クレジットカードは …



| データ | ひらがな | 漢字 | 品詞 | 活用 | 活用形 | 音便 |
|-----|------|----|----|----|-----|----|
| クレ | クレ | クレ | 無 | -- | -- | -- |

- 意味の分かる範囲であれば、品詞付けをする。
- 活用形のあやふやなものは、後の文で判断する。

ex. (登録の方で)



(/ 登録 / の / 方 / で /) /

ex. 記入して、いただ(い)ければ …



| データ | ひらがな | 漢字 | 品詞 |
|----------|-------|-----|------|
| いただ(い)けれ | いただけれ | 戴ける | 補助動詞 |

2. 時間について

ex. 10 : 10 → / 10 / : / 10 /

| データ | ひらがな | 漢字 | 品詞 |
|-----|-------|----|----|
| 10 | じゅう | 10 | 数詞 |
| : | じ | 無 | 記号 |
| 10 | じゅっぷん | 10 | 数詞 |



| データ | ひらがな | 漢字 | 品詞 |
|-----|-------|----|----|
| 10 | じゅうじ | 10 | 数詞 |
| : | 無 | 無 | 記号 |
| 10 | じゅっぷん | 10 | 数詞 |

3. 日付について

ex. 10 / 10 → / 10 / / / 10 /

2. の時間にあわせて



| データ | ひらがな | 漢字 | 品詞 |
|-----|-------|----|----|
| 10 | じゅうがつ | 10 | 名詞 |
| / | 無 | 無 | 記号 |
| 10 | とおか | 10 | 名詞 |

| 変更前 | 変更後 |
|--|--|
| <p>4. 助詞について</p> <p>1. "まで" ex. 大阪から東京まで行く</p> <p style="text-align: center;">↓</p> <p style="text-align: center;">副助詞 ?</p> <p>2. "とか" ex. 奈良市内であるとか、 そういう所に有ります。</p> <p style="text-align: center;">↓</p> <p style="text-align: center;">並助 or 副助</p> | <p>新明解には、副助詞しかない。 吉本文法には、格助詞がある。</p> <p>従って</p> <ul style="list-style-type: none"> ・ 副助詞 ... 重いものから他を類推する。 ・ 格助詞 ... 範囲の一方の限界を表わす。 <p>(答え) 格助詞</p> <p>新明解には、副助詞しかない。 吉本文法には、副助詞には認められておらず、並列助詞しかない。</p> <p>(答え) 並列助詞</p> |
| <p>5. "御座います"について</p> <p>ex. ありがとう / ございます</p> <p style="text-align: center;">↓</p> <p style="text-align: center;">動詞 ?</p> | <p>ex. そうで、ございます。</p> <p style="text-align: center;">↓</p> <p style="text-align: center;">助動詞 補動 特殊型</p> <ul style="list-style-type: none"> ・ " — である" の意味がうすれている。 <p>ex. はい、ございます。</p> <p style="text-align: center;">↓</p> <p style="text-align: center;">本動詞 特殊型</p> <ul style="list-style-type: none"> ・ " — である" の意味である。 " はい、あります" の尊敬語 <p>(答え) 補助動詞</p> <p>* — でございます。</p> <p style="text-align: center;">↓</p> <p style="text-align: center;">助動詞</p> <p>" ございます" の前にある " で" は助動詞とする。</p> |

| 変更前 | 変更後 |
|---|---|
| <p>6. 副詞について</p> <p>1. / 直ぐ / に / 形名 助動</p> <p>2. / 即座 / に /</p> <p>3. / 同時に / 副詞</p> | <p>→ / 直ぐに / 副詞</p> <p>→ / 即座 / に / 名詞 格助</p> <p>→ / 同時 / に / 名詞 格助</p> <p>1. の場合 辞書には 直ぐ 〔副〕 一 に とある。 従って、品詞が”副詞”で 〔一に 一と がある場合には、 国語大辞典を調べて、その品詞 付けに従う。</p> |
| <p>7. 何（造語）について</p> <p>ex. / 何 / 人 / ↓ 接頭語 or 名詞？</p> | <p>・ 数詞の場合</p> <p>ex. / 3 / 人 / 数詞 接尾語</p> <p>↓</p> <p>統一をとって、代名詞+接尾語とする。</p> <p>(答え) / 何 / 人 / 代名詞 接尾語</p> <p>/ 何 / 番 / 代名詞 接尾語</p> |
| <p>8. 外来語の派生語の場合</p> <p>ex. パネルディスカッション</p> <p>辞書に パネル(名詞) —— ディスカッション</p> <p>とある場合、日本語と同じように、 派生語を一つで名詞と取るか？</p> | <p>(答え) パネルディスカッション</p> <p>従って、外来語の場合も、 派生語は一つで名詞にする。</p> |

| 変更前 | 変更後 |
|--|---|
| <p>9. その他</p> <p>ex. / 悪く / は / ない /</p> <p style="text-align: center;">└───┘ ↓ 助動詞 or 形容詞 ?</p> | <ul style="list-style-type: none"> • 助動詞の”ない”は <ul style="list-style-type: none"> 1. 動詞 未然形のあと 2. 動詞 連用形のあと+”て” 3. 形容詞、形容動詞 一部の、助動詞の連用形に付く。 • 吉本文法による。 (学校文法は使わない) • 見分け方 <ol style="list-style-type: none"> 1. ”ある” (存在を意味する)の反意語が形容詞である。 2. ”ない”を”ある”に置き換えて通じたら、形容詞である。 |

| 変更前 | 変更後 |
|---|---|
| <p>1. コールフォーペーパー</p> <p>／コール／フォー／ペーパー／ 名詞 名詞 名詞</p> <p>／call／for／paper／</p> | <p>→ ／コールフォーペーパー／ 名詞</p> <p>→ ／call for paper／ 名詞</p> <p>・一つで名詞とする。</p> |
| <p>2. 造語の漢語成分について</p> <p>ex. ／同時通訳／用／</p> <p style="text-align: center;">↓</p> <p style="text-align: center;">接尾語 or 名詞</p> <p>ex. ／講演／者／</p> <p style="text-align: center;">↓</p> <p style="text-align: center;">接尾語 or 名詞</p> <p>ex. ／会場／内／</p> <p style="text-align: center;">↓</p> <p style="text-align: center;">接尾語 or 名詞</p> <p>※ 漢語の造語成分は、 接頭語・接尾語に成りうる時 と成らない時の判断の仕方は？</p> | <p>→ 今のところ名詞にしている。</p> |
| <p>3. 固有名詞について</p> <p>ex. データ通信部長</p> <p>／データ通信部／ 固有名詞</p> <p>／データ通信／部長／？ 名 or 固 名</p> <p>ex. (株)日本交通公社</p> <p>1. ひらがなを</p> <p>[・ かぶ にほんこうつうこうしゃ ・ かぶしきがいしゃにほん… するか？</p> <p>2. (株)と日本交通公社を 分けるべきか？</p> <p>ex. 日本電気会長様 日立製作所中央研究室第三… ノース[うー]ニュース社…</p> | <p>→ 今のところ ／データ通信部長／ 固有名詞 にしている。</p> <p>→ 今のところ一つで固有名詞 にしている。</p> <p>→ 固有名詞の間に間投詞が、</p> |

入っている場合。

言語情報データファイル

— 検査結果報告書 —

1988年 9月

[目次]

検査結果報告書

ページ

| | |
|-------------------|----|
| ・ 形態素解析進行状況 | 1 |
| ・ 統計結果について | 2 |
| ・ 形態素解析修正例 | 3 |
| ・ 時間測定について | 5 |
| ・ 資料 1 | 7 |
| ・ 資料 2 | 11 |
| ・ 資料 3 | 23 |
| ・ 資料 4 | 31 |
| ・ 資料 5 | 33 |

[統計結果について]

- 資料1は1988年6月、資料2は1988年7月に統計をとったものである。

3回目の修正例 (表-1)

| | 分野 | 総単語数 | 間違い件数 | 割合(%) |
|-----|-------------|--------|-------|-------|
| 資料1 | 電話会話 | 4,355 | 19 | 0.4 |
| 資料2 | 電話会話 | 10,565 | 51 | 0.5 |
| 資料2 | KEYBOARD 会話 | 2,611 | 15 | 0.6 |

- (表-1)の通り、3回目の修正で2回目の作業者の見逃しが平均0.5%となっている。
従って、いずれも2回目の修正で、ATRから要求されている品質を、十分満たしていると思われます。
- 今期は、主に電話会話、KEYBOARD会話の形態素解析修正作業を行ったが、どちらの分野の誤りの発見率も、ばらつきは無いようである。

3回目の修正の原因 (表-2)

| | 分野 | 参照番号 | 内容 | 件数 | 割合(%) |
|-----|----------------|------|------------|----|-------|
| 資料1 | 電話会話 | 29 | その他(助詞) | 7 | 0.2 |
| | | 72 | 品詞の所の表記間違い | 5 | 0.1 |
| | | 11 | 接頭語のついた語 | 1 | 0.02 |
| 資料2 | 電話会話 | 72 | 品詞の所の表記間違い | 16 | 0.2 |
| | | 29 | その他(助詞) | 8 | 0.08 |
| | | 11 | 接頭語のついた語 | 4 | 0.04 |
| 資料3 | KEYBOARD 会話 | 71 | 漢字の所の表記間違い | 5 | 0.2 |
| | | 90 | カットのおかしい語 | 3 | 0.1 |
| | | 13 | 補助動詞 | 3 | 0.1 |

- (注) 1. (表-2)の内容は、統計結果の上位3つを選んだものである。
2. 割合は、(表-1)の総単語数をもとに計算したものである。

- 参照番号による3回目の修正の原因は(表-2)の通りである。
3回目での修正の割合は、0.2%以下になっており、分野別にみてもあまりばらつきは無いようである。
しかし、修正の原因は少し異なっています。

| 分野 | 参照番号 | 内容 | 例 |
|------|------|-------------|---|
| 電話会話 | 72 | 品詞の所の表記間違い | <ul style="list-style-type: none"> • お宅 / 様 / は / ... <div style="margin-left: 40px;">↓</div> 名詞 → 代名詞 • あるいは / 、 / ... <div style="margin-left: 40px;">↓</div> 副詞 → 接続詞 • 明日 / で / よろしい / ... <div style="margin-left: 40px;">↓</div> 助動詞 → 格助詞 ※ • セッション / は / 何 / 名 / ... <div style="margin-left: 80px;">↓ ↓</div> 名詞 → 代名詞 名詞 → 接尾語 |
| 電話会話 | 29 | その他 (助詞) | <ul style="list-style-type: none"> ※ • 其処 / から / 奈良 / まで / ... <div style="margin-left: 80px;">↓</div> 副助詞 → 格助詞 (距離を表すものは格助詞である。) ※ • キャンプ / 位置 / とか / ... <div style="margin-left: 40px;">↓</div> 副助詞 → 並列助詞 |
| 電話会話 | 11 | 接頭語のついた語 | <ul style="list-style-type: none"> • お / 決り / です / か / 。 / <div style="margin-left: 40px;">↓</div> 本動詞 → 名詞 (" です" の前は名詞である。) |

(注) 1. ※ は問題点の検討の結果、変更されたものである。

| 分野 | 参照番号 | 内容 | 例 |
|----------------|------|----------------|--|
| KEYBOARD 会話 | 71 | 漢字の所の 表記間違い | <ul style="list-style-type: none"> ・それじゃあ → それじゃ ・じゃあ → じゃ <p>(辞書の表記にあわせる。)</p> |
| KEYBOARD 会話 | 90 | カットの おかしい語 | <ul style="list-style-type: none"> ・ 非常 / に / → / 非常に / ↓ ↓ ↓ 形容名詞 助動詞 副詞 ・ お / 送り / させ / て / ... "させる" 助動詞 ↓ お / 送り / さ / せ / て / ... 補助動詞 助動詞 <p>(助動詞"させる"は 本動詞 五段 連用 とは接続しない。)</p> <ul style="list-style-type: none"> ・ もう / 一度 / ... 名詞 ↓ もう / 一 / 度 / ... 数詞 接尾語 <p>(数を表わすものは、数詞である。)</p> |
| KEYBOARD 会話 | 13 | 補助動詞 | <p>※・ どうも / ありがとう / ございます /。 本動詞 → 補助動詞</p> <p>(" — である" の意味を持つもの : 本動詞 " — である" の意味を持たないもの : 補助動詞)</p> <ul style="list-style-type: none"> ・ 考え / て / い / ます / 。 / ↓ 本動詞 → 補助動詞 <p>(「動詞 + テ」に続くものは 補助動詞である。)</p> |

(注) 1. ※ は問題点の検討の結果、変更されたものである。

・ この結果

1. 辞書の見落とし
2. 変更点
 に関するものが多いようである。

[時間測定について]

- 資料3は1988年3月、資料4、5は1988年7月に調査したものである。

参照番号を含めた場合 (表-1) イディオム可

| 修正回数 | 資料 | 1ファイル平均 所要時間 | ファイル平均 形態素数 | 1時間平均 形態素数 |
|------|----|-----------------|----------------|---------------|
| 1回目 | 3 | 1時間45分 | 291 | 167 |
| | 4 | 1時間7分 | 307 | 275 |
| 2回目 | 3 | 22分 | 291 | 791 |
| | 4 | 22分 | 306 | 853 |
| 3回目 | 3 | -- | -- | -- |
| | 4 | 9分 | 306 | 2081 |

参照番号を含まない場合 (表-2) イディオム不可

| 修正回数 | 資料 | 1ファイル平均 所要時間 | ファイル平均 形態素数 | 1時間平均 形態素数 |
|------|----|-----------------|----------------|---------------|
| 1回目 | 3 | 1時間48分 | 304 | 162 |
| | 5 | 47分 | 271 | 345 |
| 2回目 | 3 | 27分 | 304 | 680 |
| | 5 | 24分 | 271 | 680 |
| 3回目 | 3 | 15分 | 304 | 1256 |
| | 5 | 6分 | 271 | 2611 |

- 現時点での形態素解析修正作業の平均所要時間は、

1回目 平均 1時間20分 前後 / 1ファイル (= 約300形態素)
 2回目 平均 24分 前後 / 1ファイル (= 約300形態素)
 3回目 平均 10分 前後 / 1ファイル (= 約300形態素)

となっている。

1回目の修正の所要時間は、これから形態素解析システムが改善されるにつれて、ますます短縮される可能性はあるが、2回目、3回目の修正は1ファイル約300前後の形態素をチェックしていかなければならないので、2回目、3回目の修正の所要時間は、これから短縮される可能性はあまりないように思われます。

従って、2ページでも記したように、3回目の修正で2回目の作業者の見逃しが全数検査の結果0.5%になっていることから、2回目の時点で、時間的にみてもかなり高品質なデータが作成されているようです。

- ・ (表-2)の1回目の修正時間は、資料3ではイディオム登録がされておらず資料5ではイディオム登録がされている場合の計測結果です。この場合、イディオム登録することによって形態素解析システムが、かなり改善され、修正の所要時間も約1時間も短縮され、能率よく、スピーディに作業が行われているように思われます。
- ・ 全体的にみて、3月時点と7月時点の状況を比べてみると、ファイルの形態素数は少し異なっているが、1ファイル平均の所要時間、1時間平均にこなす形態素数をみると、かなり能率よく、スピーディに作業が行われているように思われます。

修正件数

1回目 628件
2回目 65件

合計 712件

$$\frac{\text{(1回目+2回目+3回目)修正合計}}{\text{形態素合計}} = \frac{712}{4355} = 0.16$$

※ 従って、ヒット率は、およそ84%ぐらいになっていると想定できる。

(6月3日現在)

[資料 2]

—— 3 回目修正状況 (参照番号可) ——

1988/7/15 三木

[電話会話 5 月分]

| ファイル名 | 総単語数 | 間違い件数 | 割合 (%) |
|---------------|--------|-------|--------|
| ED_tel-042-05 | 4,350 | 19 | 0.4 |
| ED_tel-042-06 | 2,341 | 23 | 1.0 |
| ED_tel-042-07 | 1,325 | 3 | 0.2 |
| ED_tel-052-03 | 2,549 | 6 | 0.2 |
| total | 10,565 | 51 | 0.5 |

[KEYBOARD 会話 6 月分]

| ファイル名 | 総単語数 | 間違い件数 | 割合 (%) |
|---------------|-------|-------|--------|
| ED_key-102-01 | 326 | 2 | 0.6 |
| ED_key-102-02 | 273 | 3 | 1.1 |
| ED_key-122-01 | 305 | 3 | 1.0 |
| ED_key-122-02 | 265 | 1 | 0.4 |
| ED_key-122-03 | 581 | 4 | 0.7 |
| ED_key-132-01 | 342 | 1 | 0.3 |
| ED_key-132-02 | 519 | 1 | 0.2 |
| total | 2,611 | 15 | 0.6 |

[資料 3]

— 統計結果 — (電話会話)

1988/3/11 三木

[割合]

| 修正回数 分野 | 1 回目 | 2 回目 |
|--|-------------------------|-----------------------|
| 電話会話 (前回分) ・イディオム登録不可 ・参照番号を含まない | 2162 7894 (27.4%) | 118 7891 (1.5%) |
| 電話会話 ・イディオム登録可 ・参照番号を含む | 1077 5529 (19.5%) | 68 5529 (1.2%) |

[時間]

*参照番号を含めた場合 イディオム登録可

| 修正回数 内容 | 1 回目 | 2 回目 |
|-----------------------------------|-----------|---------|
| 全体の所要時間 | 33 時間 10分 | 7 時間 |
| 1 ファイル平均の 所要時間 (平均 291 形態素) | 1 時間 45分 | 22 分 |
| 1 時間平均 | 167 形態素 | 791 形態素 |

*参照番号を含まない場合 イディオム登録不可

| 修正回数 内容 | 1 回目 | 2 回目 |
|-----------------------------------|-----------|----------|
| 全体の所要時間 | 35 時間 45分 | 8 時間 30分 |
| 1 ファイル平均の 所要時間 (平均 304 形態素) | 1 時間 48分 | 27 分 |
| 1 時間平均 | 162 形態素 | 680 形態素 |

総単語数 5529

| 内容 | 参照番号 | 件数 | 割合 (%) |
|---------------|------|------|--------|
| カットのおかしい語 | 90 | 306 | 5.5 |
| 品詞の表記間違い | 72 | 179 | 3.2 |
| 補助動詞 | 13 | 92 | 1.7 |
| その他(助詞) | 29 | 84 | 1.5 |
| 接尾語を伴うもの(数詞) | 40 | 58 | 1.0 |
| 感動詞・間投詞に関するもの | 30 | 57 | 1.0 |
| 解析できないセンテンス | 92 | 46 | 0.8 |
| 日付に関するもの(数詞) | 41 | 41 | 0.7 |
| サ変名詞 | 1 | 34 | 0.6 |
| 漢字の表記間違い | 71 | 33 | 0.6 |
| ひらがなの表記間違い | 70 | 32 | 0.6 |
| 活用形の表記間違い | 74 | 28 | 0.5 |
| 未知語の修正 | 80 | 16 | 0.3 |
| 漢字が一つ以上出ている語 | 50 | 14 | 0.3 |
| 固有名詞 | 101 | 13 | 0.2 |
| 派生語 | 100 | 11 | 0.2 |
| 形容動詞(ダ活用) | 2 | 9 | 0.2 |
| データ上、品詞を変える | 110 | 8 | 0.1 |
| 下さい(特殊型) | 14 | 5 | 0.1 |
| 御座います(特殊型) | 15 | 4 | 0.1 |
| 口語的表現の違う語 | 20 | 2 | 0.04 |
| 漢字の統一 | 51 | 2 | 0.04 |
| 形容動詞(タルト活用) | 3 | 1 | 0.02 |
| 可能動詞 | 12 | 1 | 0.02 |
| 外来語・英語の表記 | 60 | 1 | 0.02 |
| total | | 1077 | 19.5 |

電話会話 (2回目修正) イディオム登録可

総単語数 5529

| 内容 | 参照番号 | 件数 | 割合 (%) |
|--------------|------|----|--------|
| 品詞の表記間違い | 72 | 16 | 0.3 |
| カットのおかしい語 | 90 | 11 | 0.2 |
| 補助動詞 | 13 | 6 | 0.1 |
| 漢字の表記間違い | 71 | 5 | 0.1 |
| サ変名詞 | 1 | 4 | 0.07 |
| その他(助詞) | 29 | 4 | 0.07 |
| 接尾語の伴うもの(数詞) | 40 | 3 | 0.05 |
| 活用形の表記間違い | 74 | 3 | 0.05 |
| 音便の表記間違い | 75 | 3 | 0.05 |
| 下さい(特殊型) | 14 | 2 | 0.04 |
| ひらがなの表記間違い | 70 | 2 | 0.04 |
| データ上、品詞を変える | 110 | 2 | 0.04 |
| 修正ミスの訂正 | 190 | 2 | 0.04 |
| 漢字の統一 | 51 | 1 | 0.02 |
| 活用形の表記間違い | 73 | 1 | 0.02 |
| 派生語 | 100 | 1 | 0.02 |
| 固有名詞 | 101 | 1 | 0.02 |
| 個人の判断の相違 | 120 | 1 | 0.02 |
| total | | 68 | 1.2 |

電話会話 (1回目修正)

前回分 イディオム登録不可

総単語数 7894

| 内容 | 参照番号 | 件数 | 割合 (%) |
|---------------|------|------|--------|
| カットのおかしい語 | 90 | 410 | 5.2 |
| 感動詞・間投詞に関するもの | 30 | 399 | 5.1 |
| 解析できないセンテンス | 92 | 311 | 3.9 |
| 品詞の表記間違い | 72 | 283 | 3.6 |
| 補助動詞 | 13 | 109 | 1.4 |
| その他(助詞) | 29 | 98 | 1.3 |
| 活用形の表記間違い | 74 | 91 | 1.2 |
| ひらがなの表記間違い | 70 | 64 | 0.8 |
| 漢字が一つ以上出ている語 | 50 | 54 | 0.7 |
| 派生語 | 100 | 51 | 0.6 |
| サ変名詞 | 1 | 50 | 0.6 |
| 漢字の表記間違い | 71 | 47 | 0.6 |
| 固有名詞 | 101 | 42 | 0.5 |
| 接尾語を伴うもの(数詞) | 40 | 34 | 0.4 |
| 日付に関するもの(数詞) | 41 | 23 | 0.3 |
| 漢字の統一 | 51 | 19 | 0.2 |
| データ上、品詞を変える | 110 | 18 | 0.2 |
| 形容動詞(ダ活用) | 2 | 16 | 0.2 |
| 未知語の修正 | 80 | 12 | 0.2 |
| 御座います(特殊型) | 15 | 10 | 0.1 |
| 活用の表記間違い | 73 | 7 | 0.1 |
| 可能動詞 | 12 | 4 | 0.1 |
| 下さい(特殊型) | 14 | 4 | 0.1 |
| 音便の表記間違い | 75 | 3 | 0.04 |
| 形容動詞(タルト活用) | 3 | 2 | 0.02 |
| 口語的表現の違う語 | 20 | 1 | 0.01 |
| total | | 2162 | 27.4 |

電話会話 (2回目修正) 前回分 イディオム登録不可

総単語数 7891

| 内容 | 参照番号 | 件数 | 割合 (%) |
|---------------|------|-----|--------|
| 品詞の表記間違い | 72 | 22 | 0.3 |
| 漢字の表記間違い | 71 | 14 | 0.2 |
| その他(助詞) | 29 | 13 | 0.2 |
| 接尾語を伴うもの(数詞) | 40 | 12 | 0.2 |
| 補助動詞 | 13 | 10 | 0.1 |
| カットのおかしい語 | 90 | 10 | 0.1 |
| 活用形の表記間違い | 74 | 8 | 0.1 |
| ひらがなの表記間違い | 70 | 6 | 0.1 |
| 感動詞・間投詞に関するもの | 30 | 4 | 0.1 |
| 音便の表記間違い | 75 | 3 | 0.04 |
| 漢字の統一 | 51 | 3 | 0.04 |
| その他(動詞) | 19 | 3 | 0.04 |
| サ変名詞 | 1 | 2 | 0.03 |
| 修正ミスの訂正 | 190 | 2 | 0.03 |
| 未知語の修正 | 80 | 2 | 0.03 |
| 可能動詞 | 12 | 1 | 0.01 |
| データ上、品詞を変える | 110 | 1 | 0.01 |
| 活用の表記間違い | 73 | 1 | 0.01 |
| 記号が消えている場合 | 91 | 1 | 0.01 |
| total | | 118 | 1.5 |

形態素情報データファイル(2)

－ 品質管理報告 －

1989年 2月

[目 次]

品質管理報告書

ページ

1. 抜き取り検査

2. 形態素管理情報報告書

・ 言語データベース一覧表

・ 形態素解析進行状況表

・ 時間測定表

言語データベース管理表

・ 電話会話

・ キーボード会話

・ I F G モデル会話

・ S D B サンプル

— 抜き取り検査 —

1. 抜き取り検査

・抽出対象

作成期間 1988年10月7日 ~ 1989年2月22日

会話分野 電話会話
キーボード会話
IFGモデル会話
SDBサンプル会話

形態素総数

| | |
|-----------|--------|
| 電話会話 | 60,091 |
| キーボード会話 | 19,225 |
| IFGモデル会話 | 2,240 |
| SDBサンプル会話 | 8,461 |

総数 90,091 形態素

対象ファイル

電話会話

| | |
|---------------|---------------|
| ED_tel-122-01 | ED_tel-172-01 |
| ED_tel-122-02 | ED_tel-172-02 |
| ED_tel-122-03 | ED_tel-172-03 |
| ED_tel-122-04 | ED_tel-202-02 |
| ED_tel-122-05 | ED_tel-202-03 |
| ED_tel-132-01 | ED_tel-202-05 |
| ED_tel-132-02 | ED_tel-212-01 |
| ED_tel-132-03 | ED_tel-212-02 |
| ED_tel-132-04 | ED_tel-212-04 |
| ED_tel-132-05 | ED_tel-212-05 |
| ED_tel-142-01 | ED_tel-212-06 |
| ED_tel-142-02 | ED_tel-222-01 |
| ED_tel-142-03 | ED_tel-222-02 |
| ED_tel-142-04 | ED_tel-222-03 |
| ED_tel-142-05 | ED_tel-222-04 |
| ED_tel-162-01 | ED_tel-232-01 |
| ED_tel-162-02 | ED_tel-232-02 |
| ED_tel-162-03 | ED_tel-232-04 |
| ED_tel-162-04 | |
| ED_tel-162-05 | |

キーボード会話

| | |
|---------------|---------------|
| ED_key-152-01 | ED_key-232-01 |
| ED_key-152-02 | ED_key-232-02 |
| ED_key-182-01 | ED_key-232-03 |
| ED_key-192-01 | ED_key-232-04 |
| ED_key-192-02 | ED_key-232-05 |
| ED_key-202-01 | ED_key-242-01 |
| ED_key-202-02 | ED_key-242-02 |
| ED_key-212-01 | ED_key-242-03 |
| ED_key-212-02 | ED_key-242-04 |
| ED_key-212-03 | ED_key-252-01 |
| ED_key-212-04 | ED_key-252-02 |
| ED_key-212-05 | ED_key-252-03 |
| ED_key-212-06 | ED_key-252-04 |
| ED_key-212-07 | ED_key-252-05 |
| ED_key-222-01 | ED_key-252-06 |
| ED_key-222-02 | ED_key-252-07 |
| ED_key-222-03 | ED_key-262-01 |
| ED_key-222-04 | ED_key-262-02 |

ED_key-222-05
ED_key-222-06

ED_key-262-03
ED_key-262-04
ED_key-262-05

I F G モデル会話

| | |
|---------------|---------------|
| ED_ifg-012-01 | ED_ifg-012-07 |
| ED_ifg-012-02 | ED_ifg-012-08 |
| ED_ifg-012-03 | ED_ifg-012-09 |
| ED_ifg-012-04 | ED_ifg-012-10 |
| ED_ifg-012-05 | ED_ifg-012-11 |
| ED_ifg-012-06 | ED_ifg-012-12 |

S D B サンプル会話

| | |
|---------------|---------------|
| ED_sdb-012-01 | ED_sdb-012-06 |
| ED_sdb-012-02 | ED_sdb-012-07 |
| ED_sdb-012-03 | ED_sdb-012-08 |
| ED_sdb-012-04 | ED_sdb-012-09 |
| ED_sdb-012-05 | ED_sdb-012-10 |

・抽出方法

乱数を発生させランダムにファイルを抜き取る。
次に、抜き取ったファイルの形態素総数を見て、その総数より乱数を発生させてランダムに形態素データを1個を取り出し、その形態素データの入った文をファイルに出力する。
このような処理を1サイクルとし、全体形態素数の約0.8%にあたる756形態素データを抽出した。

- ・データ件数 90,017 形態素
- ・抜き取り形態素数 756 形態素
- ・不良個数 7 形態素
- ・不良率 0.93%
- ・抽出日 1989年2月8日
- ・検査日 1989年2月10日
- ・検査者 三木

2. 不良に対する考察

(1) 不良内容

1. /data3/MORPH/edout/ifg/ED_ifg-012-07

／こちら／で／は／専門的／な／質問／に／お／答え／でき／ませ／ん／。

補・変則・連用

↓
上一

2. /data3/MORPH/edout/key/ED_key-152-01

／自己／紹介／し／ます／。

名 + サ名

↓
サ名

3. /data3/MORPH/edout/tel/ED_tel-232-04

／お／知らせ／し／て／おき／ます／と／、

本・下一・連用

↓
本・五・未然 + 助・連用

4. /data3/MORPH/edout/key/ED_key-152-01

／普通／に／参加／する／方／は／どなた／でも／結構／です／。

格助詞

↓
助・連用

5. /data3/MORPH/edout/ifg/ED_ifg-012-09

／京都駅／から／タクシー／で／会議／場／まで／行く／に／は／…

いく

↓
ゆく

6. /data3/MORPH/edout/key/ED_key-212-05

／それは／助かり／ます／。

副

↓
代 + 係助詞

7. /data3/MORPH/edout/ifg/ED_ifg-012-05

／こちら／まで／お／知らせ／下さい／。

本・下一・連用

↓

・不良原因 及び 対策

1. 新品詞体系によるもの 4件

1) 本動詞 + 受身・使役の助動詞 (可能動詞以外)

ex. 含ま／れる
行わ／れる

(マニュアル 2-2 参照)

2) サ名詞 + 補助動詞

- ・副詞、形容名詞以外の下に「する」が来ているものは、前はすべてサ名詞となる。
- ・固有名詞、サ変名詞は、複合名詞なので、切るとか考えずに長単位でとる。

ex. 打合せ／する
自己紹介／する

(マニュアル 1-4 参照)

3) 代名詞 + 係助詞

- ・一語で副詞か、代名詞+係助詞か判断に迷うので、代名詞 + 係助詞に統一する。

ex. それ／は／どうも／ありがとう

- ・副詞、感動詞の「それは」は、認められない。

対策 : 新品詞体系に代わって以来、作業者が理解出来ない点や不慣れな点がたくさんあり、その度、ATR側と問題点を協議し、形態素情報データに対する品詞付けや、単語の切り方等を決定してきた。従って、マニュアルも完全でなく、作業者が新品詞体系と旧品詞体系が混同した状態でデータを作成した為、本件のようなミスが発生した。しかし、電話会話に関しては、多くの問題点を抽出・検討しマニュアルも整備されてきているので、今後、この原因で起こるミスは、確実に少なくなる。

2. 作業者の見逃しによるもの 3件

1) 活用型の間違い

ex. 「出来る」について
システムでは、変則型と出てくるので、上一型と修正するのを忘れてしまう。

2) 品詞の間違い

ex. 形容名詞について
形容名詞の後は助動詞である。

(旧品詞体系) 一に 形名詞 + 助動詞
一な の後は 副詞 + 助動詞

(新品詞体系) 一に 形名詞 + 助動詞
一な の後は 副詞 + 格助詞

3) 読みの間違い

・データに「行く」と出てきた場合、「ゆく」と読むか、「いく」と読むか、ばらつきが生じるので、作業者間で統一をとっている。

行く → 「ゆく」
行って → つ音便の時だけ「いって」

対策 : 作業者が、新品詞体系に対して不慣れであり、それに気を付けるあまり、却って、見逃しが多くなっているようである。
以後、マニュアルに、随時、見逃しの内容を追加していくことにより、このようなミスは減少すると思われる。
また、そうすることによって、作業者全員がマニュアルを完備し、均一な国文法の知識を身に付け、より良い形態素解析のデータを作成することが出来る。

— 言語データベース一覧表 —

習語データベース一覧表
 [電話会話]

ファイル名 : TEL.1

T I S 作成者 : 三木
 A T R 確認者 : 篠崎
 予想合計形態素数 : 12,919
 予定開始日 : 1989/01/12
 予定終了日 : 1989/01/20
 作業開始日 : 1989/01/12
 作業終了日 : 1989/01/20

受渡者(TIS) : 三木
 受取者(ATR) : 篠崎
 提出日 : 1989/01/20
 検収確認者 : 篠崎
 検収確認日 : 1989/01/20
 合計形態素数 : 12,904

| I D | ファイル名 | 形態素数 | 形態素 文節情報 登録更新 |
|-------|---------------------|---------|---------------------|
| T0091 | t e l - 1 4 2 - 0 1 | 3 5 4 | 8 9 . 0 1 . 2 0 |
| T0092 | t e l - 1 4 2 - 0 2 | 5 9 3 | 8 9 . 0 1 . 2 0 |
| T0093 | t e l - 1 4 2 - 0 3 | 5 3 3 | 8 9 . 0 1 . 2 0 |
| T0094 | t e l - 1 4 2 - 0 4 | 8 9 4 | 8 9 . 0 1 . 2 0 |
| T0095 | t e l - 1 4 2 - 0 5 | 5 6 0 | 8 9 . 0 1 . 2 0 |
| T0096 | t e l - 1 6 2 - 0 1 | 6 6 8 | 8 9 . 0 1 . 2 0 |
| T0097 | t e l - 1 6 2 - 0 2 | 1 3 1 0 | 8 9 . 0 1 . 2 0 |
| T0123 | t e l - 2 2 2 - 0 1 | 7 7 0 | 8 9 . 0 1 . 2 0 |
| T0124 | t e l - 2 2 2 - 0 2 | 1 1 4 9 | 8 9 . 0 1 . 2 0 |
| T0125 | t e l - 2 2 2 - 0 3 | 1 3 8 7 | 8 9 . 0 1 . 2 0 |
| T0126 | t e l - 2 2 2 - 0 4 | 1 4 0 0 | 8 9 . 0 1 . 2 0 |
| T0127 | t e l - 2 3 2 - 0 1 | 1 3 5 0 | 8 9 . 0 1 . 2 0 |
| T0128 | t e l - 2 3 2 - 0 2 | 8 3 7 | 8 9 . 0 1 . 2 0 |
| T0129 | t e l - 2 3 2 - 0 4 | 1 0 9 9 | 8 9 . 0 1 . 2 0 |

暫照データベース一覧表
 [電話会話]

ファイル名 : TEL.2

TIS作成者 : 三木
 ATR確認者 : 篠崎
 予想合計形態素数 : 16,812
 予定開始日 : 1989/01/23
 予定終了日 : 1989/02/10
 作業開始日 : 1989/01/23
 作業終了日 : 1989/02/10

受渡者(TIS) : 三木
 受取者(ATR) : 篠崎
 提出日 : 1989/02/10
 検収確認者 : 篠崎
 検収確認日 : 1989/02/10
 合計形態素数 : 17,181

| ID | ファイル名 | 形態素数 | 形態素 文節情報 登録更新 |
|-------|------------|------|---------------------|
| T0087 | tel-132-02 | 3706 | 89.02.10 |
| T0088 | tel-132-03 | 2454 | 89.02.10 |
| T0089 | tel-132-04 | 2175 | 89.02.10 |
| T0090 | tel-132-05 | 765 | 89.02.10 |
| T0098 | tel-162-03 | 913 | 89.02.10 |
| T0099 | tel-162-04 | 423 | 89.02.10 |
| T0100 | tel-162-05 | 548 | 89.02.10 |
| T0101 | tel-172-01 | 1835 | 89.02.10 |
| T0102 | tel-172-02 | 1954 | 89.02.10 |
| T0103 | tel-172-03 | 2408 | 89.02.10 |

電話データベース一覧表
 [電話会話]

ファイル名 : TEL.3

TIS作成者 : 三木
 ATR確認者 : 篠崎
 予想合計形態素数 : 12,148
 予定開始日 : 1989/03/02
 予定終了日 : 1989/03/13
 作業開始日 : 1989/03/02
 作業終了日 : 1989/03/13

受渡者(TIS) : 三木
 受取者(ATR) : 篠崎
 提出日 : 1989/03/13
 検取確認者 : 篠崎
 検取確認日 : 1989/03/13
 合計形態素数 : 13,048

| ID | ファイル名 | 形態素数 | 形態素 文節情報 登録更新 |
|--------|------------|------|---------------------|
| T0104 | tel-172-04 | 1086 | 89.03.13 |
| T0105 | tel-172-05 | 1940 | 89.03.13 |
| T0106 | tel-182-01 | 986 | 89.03.13 |
| T0107 | tel-182-02 | 1728 | 89.03.13 |
| T0108 | tel-182-03 | 649 | 89.03.13 |
| T0109 | tel-182-04 | 727 | 89.03.13 |
| T0110 | tel-182-05 | 1823 | 89.03.13 |
| T0111 | tel-182-06 | 1541 | 89.03.13 |
| T0112 | tel-192-01 | 1052 | 89.03.13 |
| ※T0113 | tel-192-02 | 1516 | 89.03.13 |

※ … 追加ファイル

言語データベース一覧表
 [SDBサンプル]

ファイル名 : SDB.1

TIS作成者 : 三木
 ATR確認者 :
 予想合計形態素数 :
 予定開始日 : 1988/12/13
 予定終了日 : 1989/01/10

作業開始日 : 1988/12/13
 作業終了日 : 1989/01/11
 納品確認者 : 篠崎
 納品確認日 : 1989/01/11

| ID | ファイル名 | 形態素数 | 形態素 文節情報 登録更新 |
|---------|------------|------|---------------------|
| S8501 | sdb-012-01 | 829 | 89.01.10 |
| S8502 | sdb-012-02 | 854 | 89.01.10 |
| 登 S8503 | sdb-012-03 | 865 | 89.01.10 |
| S8504 | sdb-012-04 | 933 | 89.01.10 |
| S8505 | sdb-012-05 | 867 | 89.01.10 |
| S8506 | sdb-012-06 | 934 | 89.01.10 |
| S8507 | sdb-012-07 | 924 | 89.01.10 |
| S8508 | sdb-012-08 | 809 | 89.01.10 |
| S8509 | sdb-012-09 | 773 | 89.01.10 |
| S8510 | sdb-012-10 | 673 | 89.01.10 |

言語データベース一覧表
 [SDBサンプル]

ファイル名 : SDB.2

T I S 作成者 : 三木
 A T R 確認者 : 篠崎
 予想合計形態素数 : 5,541
 予定開始日 : 1989/02/13
 予定終了日 : 1989/03/01
 作業開始日 : 1989/02/13
 作業終了日 : 1989/03/01

受渡者 (T I S) : 三木
 受取者 (A T R) : 篠崎
 提出日 : 1989/03/13
 検収確認者 : 篠崎
 検収確認日 : 1989/03/13
 合計形態素数 : 5,405

| I D | ファイル名 | 形態素数 | 形態素 文節情報 登録更新 |
|-----------|---------------------|---------|---------------------|
| S 8 5 1 1 | s d b - 0 2 2 - 0 1 | 1 1 8 5 | 8 9 . 0 3 . 0 1 |
| S 8 5 1 2 | s d b - 0 2 2 - 0 2 | 6 2 4 | 8 9 . 0 3 . 0 1 |
| S 8 5 1 3 | s d b - 0 2 2 - 0 3 | 7 2 8 | 8 9 . 0 3 . 0 1 |
| S 8 5 1 4 | s d b - 0 2 2 - 0 4 | 7 1 2 | 8 9 . 0 3 . 0 1 |
| S 8 5 1 5 | s d b - 0 2 2 - 0 5 | 5 0 4 | 8 9 . 0 3 . 0 1 |
| S 8 5 1 6 | s d b - 0 2 2 - 0 6 | 3 6 3 | 8 9 . 0 3 . 0 1 |
| S 8 5 1 7 | s d b - 0 2 2 - 0 7 | 4 1 2 | 8 9 . 0 3 . 0 1 |
| S 8 5 1 8 | s d b - 0 2 2 - 0 8 | 8 7 7 | 8 9 . 0 3 . 0 1 |

— 形態素解析進行状況表 —

| | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | (x1000) |
|------------------------------|-------|----|----|----|----|----|----|----|----|----|----|---------|
| 1. 電話会話 (phone) | ***** | | | | | | | | | | | 58,139 |
| ----- | | | | | | | | | | | | |
| 2. キーボード 会話 (keyboard) | ***** | | | | | | | | | | | 19,225 |
| ----- | | | | | | | | | | | | |
| 3. 雑誌 (newsweek) | ----- | | | | | | | | | | | |
| ----- | | | | | | | | | | | | |
| 4. 新聞記事 (paper) | ----- | | | | | | | | | | | |
| ----- | | | | | | | | | | | | |
| 5. 論説文 (article) | ----- | | | | | | | | | | | |
| ----- | | | | | | | | | | | | |
| 7. i f g モデル会話 (ifg) | ** | | | | | | | | | | | 2,240 |
| ----- | | | | | | | | | | | | |
| 8. s d b サンプル (sdb) | ***** | | | | | | | | | | | 13,866 |
| ----- | | | | | | | | | | | | |

— 時間測定表 —

— 参照番号を含まない場合の形態素データ作成結果（電話会話） —

1989/01/20 三木

[1回目修正]

| ファイル名 | 形態素数 | 画面数 | 時間(分) |
|--------------------|------|------|-------|
| ED_tel-142-01-01.1 | 354 | 96 | 120 |
| ED_tel-142-02-01.1 | 387 | 99 | 105 |
| ED_tel-142-02-02.1 | 206 | 59 | 60 |
| ED_tel-142-03-01.1 | 418 | 99 | 60 |
| ED_tel-142-03-02.1 | 115 | 33 | 50 |
| ED_tel-142-04-01.1 | 400 | 100 | 95 |
| ED_tel-142-04-02.1 | 402 | 100 | 125 |
| ED_tel-142-04-03.1 | 94 | 28 | 35 |
| ED_tel-142-05-01.1 | 397 | 100 | 120 |
| ED_tel-142-05-02.1 | 164 | 40 | 75 |
| ED_tel-162-01-01.1 | 535 | 98 | 105 |
| ED_tel-162-01-02.1 | 132 | 25 | 30 |
| ED_tel-162-02-01.1 | 626 | 98 | 120 |
| ED_tel-162-02-02.1 | 567 | 99 | 150 |
| ED_tel-162-02-03.1 | 116 | 17 | 35 |
| total | 4913 | 1091 | 1285 |

(2 1 時間 2 5 分)

↓

| |
|---|
| ・ 1ファイル平均 52分 (1ファイル平均328形態素とした場合) |
| ・ 1時間平均 229形態素 |

[2回目修正]

| ファイル名 | 形態素数 | 画面数 | 時間(分) |
|--------------------|------|------|-------|
| ED_tel-142-01-01.2 | 354 | 96 | 30 |
| ED_tel-142-02-01.2 | 387 | 99 | 30 |
| ED_tel-142-02-02.2 | 206 | 59 | 15 |
| ED_tel-142-03-01.2 | 418 | 99 | 15 |
| ED_tel-142-03-02.2 | 115 | 33 | 10 |
| ED_tel-142-04-01.2 | 400 | 100 | 20 |
| ED_tel-142-04-02.2 | 400 | 100 | 45 |
| ED_tel-142-04-03.2 | 93 | 28 | 10 |
| ED_tel-142-05-01.2 | 397 | 100 | 25 |
| ED_tel-142-05-02.2 | 164 | 40 | 85 |
| ED_tel-162-01-01.2 | 536 | 98 | 30 |
| ED_tel-162-01-02.2 | 132 | 25 | 10 |
| ED_tel-162-02-01.2 | 626 | 98 | 30 |
| ED_tel-162-02-02.2 | 567 | 99 | 55 |
| ED_tel-162-02-03.2 | 116 | 17 | 5 |
| total | 4911 | 1091 | 415 |

(6 時間 5 5 分)

↓

| |
|---|
| ・ 1ファイル平均 28分 (1ファイル平均327形態素とした場合) |
| ・ 1時間平均 710形態素 |

[3 回目修正]

| ファイル名 | 形態素数 | 画面数 | 時間 (分) |
|--------------------|------|------|--------|
| ED_tel-142-01-01.3 | 354 | 96 | 15 |
| ED_tel-142-02-01.3 | 387 | 99 | 15 |
| ED_tel-142-02-02.3 | 206 | 59 | 10 |
| ED_tel-142-03-01.3 | 418 | 99 | 17 |
| ED_tel-142-03-02.3 | 115 | 33 | 5 |
| ED_tel-142-04-01.3 | 400 | 100 | 15 |
| ED_tel-142-04-02.3 | 400 | 100 | 15 |
| ED_tel-142-04-03.3 | 94 | 28 | 5 |
| ED_tel-142-05-01.3 | 396 | 99 | 15 |
| ED_tel-142-05-02.3 | 164 | 40 | 12 |
| ED_tel-162-01-01.3 | 536 | 98 | 20 |
| ED_tel-162-01-02.3 | 132 | 25 | 2 |
| ED_tel-162-02-01.3 | 626 | 98 | 20 |
| ED_tel-162-02-02.3 | 567 | 99 | 20 |
| ED_tel-162-02-03.3 | 117 | 17 | 10 |
| total | 4912 | 1090 | 196 |

(3 時間 1 6 分)

↓

| |
|-----------------------------|
| ・ 1 ファイル平均 1 3 分 |
| (1 ファイル平均 3 2 7 形態素とした場合) |
| ・ 1 時間平均 1 5 0 4 形態素 |

1989/01/20 三木

[4 回目修正]

| ファイル名 | 形態素数 | 画面数 | 時間 (分) |
|-----------------|------|------|--------|
| ED_tel-222-01.4 | 770 | 155 | 60 |
| ED_tel-222-02.4 | 1149 | 245 | 150 |
| ED_tel-222-03.4 | 1398 | 331 | 300 |
| ED_tel-222-04.4 | 1399 | 287 | 300 |
| ED_tel-232-01.4 | 1350 | 386 | 70 |
| ED_tel-232-02.4 | 837 | 230 | 115 |
| ED_tel-232-04.4 | 1100 | 308 | 130 |
| total | 8003 | 1942 | 1125 |

(1 8 時間 4 5 分)

↓

| |
|--|
| <ul style="list-style-type: none"> 1 ファイル平均 2 時間 4 1 分 (1 ファイル平均 1 1 4 3 形態素とした場合) 1 時間平均 4 2 7 形態素 |
|--|

[5 回目修正]

| ファイル名 | 形態素数 | 画面数 | 時間 (分) |
|-----------------|------|------|--------|
| ED_tel-222-01.5 | 770 | 155 | 30 |
| ED_tel-222-02.5 | 1149 | 245 | 45 |
| ED_tel-222-03.5 | 1387 | 331 | 60 |
| ED_tel-222-04.5 | 1400 | 287 | 60 |
| ED_tel-232-01.5 | 1350 | 386 | 60 |
| ED_tel-232-02.5 | 837 | 230 | 25 |
| ED_tel-232-04.5 | 1099 | 308 | 30 |
| total | 7992 | 1942 | 310 |

(5 時間 1 0 分)

↓

| |
|---|
| <ul style="list-style-type: none"> 1 ファイル平均 4 5 分 (1 ファイル平均 1 1 4 1 形態素とした場合) 1 時間平均 1 5 4 7 形態素 |
|---|

— 参照番号を含まない場合の形態素データ作成結果（電話会話） —

1989/02/10 三木

[1回目修正]

| ファイル名 | 形態素数 | 画面数 | 時間(分) |
|--------------------|-------|------|-------|
| ED_tel-132-02-01.1 | 391 | 98 | 130 |
| ED_tel-132-02-02.1 | 412 | 101 | 130 |
| ED_tel-132-02-03.1 | 424 | 100 | 130 |
| ED_tel-132-02-04.1 | 351 | 95 | 95 |
| ED_tel-132-02-05.1 | 383 | 99 | 130 |
| ED_tel-132-02-06.1 | 446 | 99 | 130 |
| ED_tel-132-02-07.1 | 386 | 100 | 160 |
| ED_tel-132-02-08.1 | 406 | 97 | 120 |
| ED_tel-132-02-09.1 | 326 | 99 | 70 |
| ED_tel-132-02-10.1 | 183 | 55 | 25 |
| ED_tel-132-03-01.1 | 367 | 99 | 70 |
| ED_tel-132-03-02.1 | 367 | 96 | 38 |
| ED_tel-132-03-03.1 | 491 | 98 | 120 |
| ED_tel-132-03-04.1 | 337 | 99 | 70 |
| ED_tel-132-03-05.1 | 347 | 98 | 90 |
| ED_tel-132-03-06.1 | 347 | 98 | 100 |
| ED_tel-132-03-07.1 | 196 | 47 | 60 |
| ED_tel-132-04-01.1 | 342 | 99 | 100 |
| ED_tel-132-04-02.1 | 381 | 100 | 110 |
| ED_tel-132-04-03.1 | 391 | 99 | 115 |
| ED_tel-132-04-04.1 | 334 | 98 | 85 |
| ED_tel-132-04-05.1 | 358 | 97 | 75 |
| ED_tel-132-04-06.1 | 363 | 95 | 90 |
| ED_tel-132-05-01.1 | 308 | 97 | 60 |
| ED_tel-132-05-02.1 | 323 | 99 | 80 |
| ED_tel-132-05-03.1 | 141 | 36 | 25 |
| ED_tel-162-03-01.1 | 558 | 98 | 160 |
| ED_tel-162-03-02.1 | 355 | 67 | 165 |
| ED_tel-162-04-01.1 | 424 | 74 | 145 |
| ED_tel-162-05-01.1 | 526 | 99 | 105 |
| ED_tel-162-05-02.1 | 22 | 5 | 10 |
| ED_tel-172-01-01.1 | 455 | 100 | 195 |
| ED_tel-172-01-02.1 | 452 | 98 | 125 |
| ED_tel-172-01-03.1 | 474 | 100 | 100 |
| ED_tel-172-01-04.1 | 415 | 98 | 135 |
| ED_tel-172-01-05.1 | 38 | 7 | 10 |
| ED_tel-172-02-01.1 | 394 | 98 | 95 |
| ED_tel-172-02-02.1 | 489 | 99 | 190 |
| ED_tel-172-02-03.1 | 383 | 99 | 90 |
| ED_tel-172-02-04.1 | 363 | 96 | 100 |
| ED_tel-172-02-05.1 | 323 | 72 | 75 |
| ED_tel-172-03-01.1 | 383 | 93 | 120 |
| ED_tel-172-03-02.1 | 506 | 99 | 120 |
| ED_tel-172-03-03.1 | 493 | 99 | 135 |
| ED_tel-172-03-04.1 | 458 | 100 | 100 |
| ED_tel-172-03-05.1 | 437 | 100 | 90 |
| ED_tel-172-03-06.1 | 127 | 38 | 25 |
| total | 17176 | 4137 | 4698 |

(78時間18分)

↓

| |
|------------------------|
| ・ 1ファイル平均 1時間 40分 |
| (1ファイル平均365形態素とした場合) |
| ・ 1時間平均 219形態素 |

[2 回目修正]

| ファイル名 | 形態素数 | 画面数 | 時間 (分) |
|--------------------|-------|------|--------|
| ED_tel-132-02-01.2 | 391 | 98 | 60 |
| ED_tel-132-02-02.2 | 412 | 101 | 35 |
| ED_tel-132-02-03.2 | 423 | 100 | 60 |
| ED_tel-132-02-04.2 | 351 | 95 | 25 |
| ED_tel-132-02-05.2 | 382 | 99 | 20 |
| ED_tel-132-02-06.2 | 446 | 99 | 40 |
| ED_tel-132-02-07.2 | 387 | 100 | 60 |
| ED_tel-132-02-08.2 | 406 | 97 | 60 |
| ED_tel-132-02-09.2 | 325 | 99 | 20 |
| ED_tel-132-02-10.2 | 183 | 55 | 25 |
| ED_tel-132-03-01.2 | 367 | 99 | 15 |
| ED_tel-132-03-02.2 | 367 | 96 | 30 |
| ED_tel-132-03-03.2 | 491 | 98 | 25 |
| ED_tel-132-03-04.2 | 338 | 99 | 25 |
| ED_tel-132-03-05.2 | 348 | 98 | 15 |
| ED_tel-132-03-06.2 | 347 | 98 | 45 |
| ED_tel-132-03-07.2 | 196 | 47 | 10 |
| ED_tel-132-04-01.2 | 344 | 99 | 60 |
| ED_tel-132-04-02.2 | 381 | 100 | 20 |
| ED_tel-132-04-03.2 | 393 | 99 | 20 |
| ED_tel-132-04-04.2 | 334 | 98 | 25 |
| ED_tel-132-04-05.2 | 359 | 97 | 25 |
| ED_tel-132-04-06.2 | 363 | 95 | 40 |
| ED_tel-132-05-01.2 | 309 | 97 | 15 |
| ED_tel-132-05-02.2 | 320 | 99 | 35 |
| ED_tel-132-05-03.2 | 141 | 36 | 10 |
| ED_tel-162-03-01.2 | 557 | 98 | 55 |
| ED_tel-162-03-02.2 | 354 | 67 | 55 |
| ED_tel-162-04-01.2 | 423 | 74 | 50 |
| ED_tel-162-05-01.2 | 526 | 99 | 35 |
| ED_tel-162-05-02.2 | 22 | 5 | 10 |
| ED_tel-172-01-01.2 | 455 | 100 | 30 |
| ED_tel-172-01-02.2 | 452 | 98 | 45 |
| ED_tel-172-01-03.2 | 474 | 100 | 45 |
| ED_tel-172-01-04.2 | 415 | 98 | 25 |
| ED_tel-172-01-05.2 | 38 | 7 | 1 |
| ED_tel-172-02-01.2 | 394 | 98 | 20 |
| ED_tel-172-02-02.2 | 489 | 99 | 30 |
| ED_tel-172-02-03.2 | 385 | 99 | 30 |
| ED_tel-172-02-04.2 | 363 | 96 | 30 |
| ED_tel-172-02-05.2 | 324 | 72 | 30 |
| ED_tel-172-03-01.2 | 385 | 93 | 20 |
| ED_tel-172-03-02.2 | 506 | 99 | 25 |
| ED_tel-172-03-03.2 | 493 | 99 | 30 |
| ED_tel-172-03-04.2 | 458 | 100 | 35 |
| ED_tel-172-03-05.2 | 437 | 100 | 10 |
| ED_tel-172-03-06.2 | 127 | 38 | 5 |
| total | 16800 | 4037 | 1436 |

(2 3 時間 5 6 分)

↓

| | |
|-----------------------------|-----------|
| ・ 1 ファイル平均 | 3 1 分 |
| (1 ファイル平均 3 5 7 形態素とした場合) | |
| ・ 1 時間平均 | 7 0 2 形態素 |

[3 回目修正]

| ファイル名 | 形態素数 | 画面数 | 時間(分) |
|--------------------|-------|------|-------|
| ED_tel-132-02-01.3 | 391 | 98 | 40 |
| ED_tel-132-02-02.3 | 412 | 100 | 20 |
| ED_tel-132-02-03.3 | 423 | 100 | 25 |
| ED_tel-132-02-04.3 | 349 | 95 | 11 |
| ED_tel-132-02-05.3 | 382 | 99 | 25 |
| ED_tel-132-02-06.3 | 446 | 99 | 15 |
| ED_tel-132-02-07.3 | 387 | 100 | 20 |
| ED_tel-132-02-08.3 | 405 | 97 | 15 |
| ED_tel-132-02-09.3 | 325 | 99 | 20 |
| ED_tel-132-02-10.3 | 186 | 55 | 15 |
| ED_tel-132-03-01.3 | 367 | 99 | 30 |
| ED_tel-132-03-02.3 | 367 | 96 | 20 |
| ED_tel-132-03-03.3 | 491 | 98 | 15 |
| ED_tel-132-03-04.3 | 338 | 99 | 15 |
| ED_tel-132-03-05.3 | 348 | 98 | 15 |
| ED_tel-132-03-06.3 | 347 | 98 | 10 |
| ED_tel-132-03-07.3 | 196 | 47 | 5 |
| ED_tel-132-04-01.3 | 344 | 99 | 10 |
| ED_tel-132-04-02.3 | 381 | 100 | 15 |
| ED_tel-132-04-03.3 | 394 | 99 | 15 |
| ED_tel-132-04-04.3 | 334 | 98 | 20 |
| ED_tel-132-04-05.3 | 359 | 97 | 15 |
| ED_tel-132-04-06.3 | 363 | 95 | 25 |
| ED_tel-132-05-01.3 | 304 | 95 | 15 |
| ED_tel-132-05-02.3 | 320 | 99 | 15 |
| ED_tel-132-05-03.3 | 141 | 36 | 10 |
| ED_tel-162-03-01.3 | 558 | 98 | 30 |
| ED_tel-162-03-02.3 | 355 | 67 | 25 |
| ED_tel-162-04-01.3 | 423 | 74 | 15 |
| ED_tel-162-05-01.3 | 526 | 99 | 20 |
| ED_tel-162-05-02.3 | 22 | 5 | 1 |
| ED_tel-172-01-01.3 | 455 | 100 | 15 |
| ED_tel-172-01-02.3 | 452 | 98 | 15 |
| ED_tel-172-01-03.3 | 474 | 100 | 20 |
| ED_tel-172-01-04.3 | 416 | 98 | 20 |
| ED_tel-172-01-05.3 | 38 | 7 | 1 |
| ED_tel-172-02-01.3 | 394 | 98 | 10 |
| ED_tel-172-02-02.3 | 489 | 99 | 20 |
| ED_tel-172-02-03.3 | 384 | 99 | 15 |
| ED_tel-172-02-04.3 | 363 | 96 | 20 |
| ED_tel-172-02-05.3 | 324 | 72 | 12 |
| ED_tel-172-03-01.3 | 386 | 93 | 20 |
| ED_tel-172-03-02.3 | 506 | 99 | 20 |
| ED_tel-172-03-03.3 | 493 | 99 | 20 |
| ED_tel-172-03-04.3 | 458 | 100 | 20 |
| ED_tel-172-03-05.3 | 438 | 100 | 20 |
| ED_tel-172-03-06.3 | 127 | 38 | 10 |
| total | 17181 | 4134 | 805 |

(1 3 時間 2 5 分)

↓

| | |
|-----------------------------|-------------|
| ・ 1 ファイル平均 | 1 7 分 |
| (1 ファイル平均 3 6 6 形態素とした場合) | |
| ・ 1 時間平均 | 1 2 8 1 形態素 |

— 参照番号を含まない場合の形態素データ作成結果（電話会話） —

1989/03/13 三木

[1回目修正]

| ファイル名 | 形態素数 | 画面数 | 時間 (分) |
|--------------------|-------|------|--------|
| ED_tel-172-04-01.1 | 316 | 96 | 90 |
| ED_tel-172-04-02.1 | 146 | 32 | 175 |
| ED_tel-172-04-03.1 | 352 | 79 | 75 |
| ED_tel-172-05-01.1 | 415 | 99 | 150 |
| ED_tel-172-05-02.1 | 500 | 98 | 90 |
| ED_tel-172-05-03.1 | 461 | 98 | 60 |
| ED_tel-172-05-04.1 | 470 | 99 | 120 |
| ED_tel-172-05-05.1 | 91 | 28 | 15 |
| ED_tel-182-01-01.1 | 425 | 100 | 95 |
| ED_tel-182-01-02.1 | 562 | 100 | 180 |
| ED_tel-182-01-03.1 | 2 | 1 | 1 |
| ED_tel-182-02-01.1 | 408 | 100 | 40 |
| ED_tel-182-02-02.1 | 414 | 100 | 130 |
| ED_tel-182-02-03.1 | 426 | 98 | 150 |
| ED_tel-182-02-04.1 | 471 | 96 | 125 |
| ED_tel-182-03-01.1 | 517 | 100 | 70 |
| ED_tel-182-03-02.1 | 132 | 26 | 30 |
| ED_tel-182-04-01.1 | 398 | 100 | 105 |
| ED_tel-182-04-02.1 | 328 | 69 | 80 |
| ED_tel-182-05-01.1 | 461 | 100 | 60 |
| ED_tel-182-05-02.1 | 571 | 99 | 175 |
| ED_tel-182-05-03.1 | 596 | 99 | 135 |
| ED_tel-182-05-04.1 | 195 | 36 | 35 |
| ED_tel-182-06-01.1 | 424 | 99 | 80 |
| ED_tel-182-06-02.1 | 424 | 100 | 65 |
| ED_tel-182-06-03.1 | 359 | 100 | 100 |
| ED_tel-182-06-04.1 | 337 | 81 | 90 |
| ED_tel-192-01-01.1 | 337 | 98 | 70 |
| ED_tel-192-01-02.1 | 401 | 100 | 80 |
| ED_tel-192-01-03.1 | 312 | 74 | 85 |
| ED_tel-192-02-01.1 | 453 | 98 | 120 |
| ED_tel-192-02-02.1 | 407 | 99 | 60 |
| ED_tel-192-02-03.1 | 464 | 99 | 190 |
| ED_tel-192-02-04.1 | 196 | 44 | 40 |
| total | 12771 | 2845 | 3166 |

(5 2 時間 4 6 分)

↓

| |
|-----------------------------|
| ・ 1 ファイル平均 1 時間 3 3 分 |
| (1 ファイル平均 3 7 6 形態素とした場合) |
| ・ 1 時間平均 2 4 2 形態素 |

*1 --- このファイルは 4 0 0 形態素ぐらい、予想されるが
マシントラブルのため途中で切れている

[2 回目修正]

| ファイル名 | 形態素数 | 画面数 | 時間 (分) |
|--------------------|-------|------|--------|
| ED_tel-172-04-01.2 | 316 | 96 | 30 |
| ED_tel-172-04-02.2 | 418 | 98 | 30 |
| ED_tel-172-04-03.2 | 352 | 79 | 20 |
| ED_tel-172-05-01.2 | 415 | 99 | 20 |
| ED_tel-172-05-02.2 | 500 | 98 | 40 |
| ED_tel-172-05-03.2 | 464 | 98 | 80 |
| ED_tel-172-05-04.2 | 470 | 99 | 15 |
| ED_tel-172-05-05.2 | 91 | 28 | 2 |
| ED_tel-182-01-01.2 | 423 | 100 | 35 |
| ED_tel-182-01-02.2 | 561 | 100 | 30 |
| ED_tel-182-01-03.2 | 2 | 1 | 1 |
| ED_tel-182-02-01.2 | 409 | 100 | 30 |
| ED_tel-182-02-02.2 | 414 | 100 | 15 |
| ED_tel-182-02-03.2 | 434 | 98 | 30 |
| ED_tel-182-02-04.2 | 471 | 96 | 30 |
| ED_tel-182-03-01.2 | 517 | 100 | 55 |
| ED_tel-182-03-02.2 | 132 | 26 | 5 |
| ED_tel-182-04-01.2 | 398 | 100 | 20 |
| ED_tel-182-04-02.2 | 328 | 69 | 30 |
| ED_tel-182-05-01.2 | 461 | 100 | 40 |
| ED_tel-182-05-02.2 | 570 | 99 | 60 |
| ED_tel-182-05-03.2 | 596 | 99 | 35 |
| ED_tel-182-05-04.2 | 195 | 36 | 15 |
| ED_tel-182-06-01.2 | 423 | 99 | 20 |
| ED_tel-182-06-02.2 | 424 | 100 | 30 |
| ED_tel-182-06-03.2 | 359 | 100 | 20 |
| ED_tel-182-06-04.2 | 337 | 81 | 20 |
| ED_tel-192-01-01.2 | 337 | 98 | 20 |
| ED_tel-192-01-02.2 | 402 | 100 | 20 |
| ED_tel-192-01-03.2 | 313 | 74 | 25 |
| ED_tel-192-02-01.2 | 453 | 98 | 15 |
| ED_tel-192-02-02.2 | 406 | 99 | 20 |
| ED_tel-192-02-03.2 | 463 | 99 | 25 |
| ED_tel-192-02-04.2 | 196 | 44 | 10 |
| total | 13050 | 2911 | 893 |

(1 4 時間 5 3 分)

↓

| |
|---------------------------|
| ・ 1 ファイル平均 26分 |
| (1 ファイル平均 384 形態素とした場合) |
| ・ 1 時間平均 877 形態素 |

[3 回目修正]

| ファイル名 | 形態素数 | 画面数 | 時間 (分) |
|--------------------|-------|------|--------|
| ED_tel-172-04-01.3 | 316 | 96 | 9 |
| ED_tel-172-04-02.3 | 418 | 98 | 20 |
| ED_tel-172-04-03.3 | 352 | 79 | 20 |
| ED_tel-172-05-01.3 | 415 | 99 | 15 |
| ED_tel-172-05-02.3 | 500 | 98 | 30 |
| ED_tel-172-05-03.3 | 464 | 98 | 20 |
| ED_tel-172-05-04.3 | 470 | 99 | 45 |
| ED_tel-172-05-05.3 | 91 | 28 | 10 |
| ED_tel-182-01-01.3 | 423 | 100 | 10 |
| ED_tel-182-01-02.3 | 561 | 100 | 20 |
| ED_tel-182-01-03.3 | 2 | 1 | 1 |
| ED_tel-182-02-01.3 | 409 | 100 | 30 |
| ED_tel-182-02-02.3 | 414 | 100 | 25 |
| ED_tel-182-02-03.3 | 434 | 98 | 25 |
| ED_tel-182-02-04.3 | 471 | 96 | 35 |
| ED_tel-182-03-01.3 | 517 | 100 | 35 |
| ED_tel-182-03-02.3 | 132 | 26 | 5 |
| ED_tel-182-04-01.3 | 398 | 100 | 20 |
| ED_tel-182-04-02.3 | 329 | 69 | 10 |
| ED_tel-182-05-01.3 | 461 | 100 | 25 |
| ED_tel-182-05-02.3 | 571 | 99 | 20 |
| ED_tel-182-05-03.3 | 596 | 99 | 20 |
| ED_tel-182-05-04.3 | 195 | 36 | 15 |
| ED_tel-182-06-01.3 | 424 | 99 | 15 |
| ED_tel-182-06-02.3 | 421 | 99 | 5 |
| ED_tel-182-06-03.3 | 359 | 100 | 20 |
| ED_tel-182-06-04.3 | 337 | 81 | 15 |
| ED_tel-192-01-01.3 | 337 | 98 | 10 |
| ED_tel-192-01-02.3 | 402 | 100 | 10 |
| ED_tel-192-01-03.3 | 313 | 74 | 18 |
| ED_tel-192-02-01.3 | 451 | 98 | 25 |
| ED_tel-192-02-02.3 | 407 | 99 | 10 |
| ED_tel-192-02-03.3 | 462 | 99 | 10 |
| ED_tel-192-02-04.3 | 196 | 44 | 10 |
| total | 13048 | 2910 | 613 |

(1 0 時間 1 3 分)

↓

| |
|---|
| ・ 1 ファイル平均 1 8 分 (1 ファイル平均 3 8 4 形態素とした場合) |
| ・ 1 時間平均 1 2 7 7 形態素 |

— 時間測定表 (SDB サンプル会話) —

※机上の修正分について

1989/03/01 三木

[1 回目修正]

| ファイル名 | 形態素数 | 画面数 | 時間 (分) |
|--------------------|------|-----|--------|
| ED_sdb-022-01-01.1 | 791 | 100 | 125 |
| ED_sdb-022-01-02.1 | 399 | 37 | 70 |
| ED_sdb-022-02-01.1 | 624 | 72 | 90 |
| ED_sdb-022-03-01.1 | 727 | 85 | 160 |
| ED_sdb-022-04-01.1 | 712 | 67 | 115 |
| ED_sdb-022-05-01.1 | 504 | 57 | 100 |
| ED_sdb-022-06-01.1 | 363 | 55 | 90 |
| ED_sdb-022-07-01.1 | 412 | 51 | 80 |
| ED_sdb-022-08-01.1 | 699 | 100 | 115 |
| ED_sdb-022-08-02.1 | 180 | 22 | 25 |
| total | 5411 | 646 | 970 |

(1 6 時間 1 0 分)

↓

| |
|-----------------------------|
| ・ 1 ファイル平均 1 時間 3 7 分 |
| (1 ファイル平均 5 4 1 形態素とした場合) |
| ・ 1 時間平均 3 3 5 形態素 |

[2 回目修正]

| ファイル名 | 形態素数 | 画面数 | 時間 (分) |
|--------------------|------|-----|--------|
| ED_sdb-022-01-01.2 | 787 | 100 | 45 |
| ED_sdb-022-01-02.2 | 398 | 37 | 20 |
| ED_sdb-022-02-01.2 | 624 | 72 | 50 |
| ED_sdb-022-03-01.2 | 728 | 85 | 40 |
| ED_sdb-022-04-01.2 | 712 | 67 | 30 |
| ED_sdb-022-05-01.2 | 504 | 57 | 15 |
| ED_sdb-022-06-01.2 | 363 | 55 | 25 |
| ED_sdb-022-07-01.2 | 412 | 51 | 10 |
| ED_sdb-022-08-01.2 | 697 | 100 | 45 |
| ED_sdb-022-08-02.2 | 180 | 22 | 10 |
| total | 5405 | 646 | 290 |

(4 時間 5 0 分)

↓

| |
|-----------------------------|
| ・ 1 ファイル平均 2 9 分 |
| (1 ファイル平均 5 4 0 形態素とした場合) |
| ・ 1 時間平均 1 1 1 8 形態素 |

形態素情報データファイル(3)

－ 品質管理報告 －

1989年 8月

< 目次 >

| | |
|--------------------------|----|
| 1. 形態素情報の抜き取り検査 | 1 |
| 1 - 1. 検査対象 | 2 |
| 1 - 2. 検査方法と検査結果 | 7 |
| 1 - 3. 不良に対する考察 | 8 |
| 1 - 4. 品質向上に対する考察 | 9 |
| 1 - 4 - 1. 考察の要点 | 9 |
| 1 - 4 - 2. 新規事項（取り決め） | 10 |
| 1 - 4 - 3. 形態素抽出プログラム出力例 | 14 |
| 2. 「で」に対する考察と形態素情報の付け方 | 17 |
| 2 - 1. 概要 | 18 |
| 2 - 2. 「で」の品詞と具体例 | 18 |
| 3. 不良テキストに対する取り扱い | 31 |
| 3 - 1. ゆれに対する問題点 | 32 |
| 3 - 2. 不良対象報告 | 34 |

1. 形態素情報の抜き取り検査

1-1. 検査対象

■ 作業期間 1989年3月1日 ~ 1989年7月31日

■ 会話分野と形態素数

| 会話分野 | 形態素数 | ファイル数 | 1ファイル平均 (単位:形態素) |
|---------|---------|-------|---------------------|
| 電話会話 | 86,978 | 86 | 1011 |
| キーボード会話 | 53,782 | 132 | 407 |
| 手紙文 | 12,252 | 25 | 490 |
| ※※※※※※※ | 153,012 | 243 | ※※※※※※ |

(注)・1989年前期分の目標単語数は18万語であるが、残り3万語は8月31日に納品予定である。これは、次回の検査対象とする。

・ファイル数は会話数を表す

■ 対象ファイル

<電話会話>

ファイル数 : 86ファイル

ファイル名 :

| | |
|---------------|---------------|
| ED_tel-172-04 | ED_tel-312-01 |
| ED_tel-172-05 | ED_tel-312-02 |
| ED_tel-182-01 | ED_tel-312-03 |
| ED_tel-182-02 | ED_tel-312-04 |
| ED_tel-182-03 | ED_tel-322-01 |
| ED_tel-182-04 | ED_tel-322-02 |
| ED_tel-182-05 | ED_tel-322-03 |
| ED_tel-182-06 | ED_tel-322-04 |
| ED_tel-192-01 | ED_tel-322-05 |
| ED_tel-192-02 | ED_tel-332-01 |
| ED_tel-192-03 | ED_tel-332-02 |
| ED_tel-242-04 | ED_tel-332-03 |
| ED_tel-252-01 | ED_tel-332-04 |
| ED_tel-252-02 | ED_tel-342-01 |
| ED_tel-252-03 | ED_tel-342-02 |
| ED_tel-252-04 | ED_tel-342-03 |
| ED_tel-252-05 | ED_tel-342-04 |
| ED_tel-262-01 | ED_tel-342-05 |
| ED_tel-262-02 | ED_tel-342-06 |
| ED_tel-262-03 | ED_tel-342-07 |
| ED_tel-272-01 | ED_tel-082-01 |
| ED_tel-272-03 | ED_tel-082-02 |
| ED_tel-272-04 | ED_tel-082-03 |
| ED_tel-282-01 | ED_tel-082-04 |
| ED_tel-282-02 | ED_tel-022-01 |
| ED_tel-282-03 | ED_tel-022-02 |
| ED_tel-282-04 | ED_tel-022-03 |
| ED_tel-282-05 | ED_tel-022-04 |
| ED_tel-282-06 | ED_tel-022-05 |
| ED_tel-282-07 | ED_tel-022-06 |
| ED_tel-282-08 | ED_tel-022-07 |
| ED_tel-282-09 | ED_tel-022-08 |
| ED_tel-292-01 | ED_tel-022-09 |

ED_tel-292-02
ED_tel-292-03
ED_tel-292-04
ED_tel-292-05
ED_tel-302-01
ED_tel-302-02
ED_tel-302-03
ED_tel-302-04
ED_tel-302-05
ED_tel-302-06

ED_tel-022-10
ED_tel-022-11
ED_tel-022-12
ED_tel-022-13
ED_tel-022-14
ED_tel-032-01
ED_tel-032-02
ED_tel-032-03
ED_tel-032-04
ED_tel-032-05

<キーボード会話>

ファイル数 : 132ファイル

ファイル名 : ED_key-282-01 ED_key-552-01
ED_key-292-01 ED_key-552-02
ED_key-292-02 ED_key-552-03
ED_key-312-02 ED_key-552-04
ED_key-312-03 ED_key-562-01
ED_key-322-01 ED_key-572-01
ED_key-322-02 ED_key-572-02
ED_key-322-03 ED_key-572-03
ED_key-322-04 ED_key-572-04
ED_key-332-05 ED_key-582-01
ED_key-342-01 ED_key-582-02
ED_key-342-03 ED_key-582-03
ED_key-342-05 ED_key-592-01
ED_key-362-02 ED_key-592-02
ED_key-362-03 ED_key-592-03
ED_key-362-04 ED_key-602-01
ED_key-372-02 ED_key-602-02
ED_key-372-04 ED_key-612-01
ED_key-382-02 ED_key-612-02
ED_key-382-03 ED_key-622-01
ED_key-392-01 ED_key-622-02
ED_key-392-02 ED_key-622-03
ED_key-392-03 ED_key-012-01
ED_key-402-01 ED_key-012-02
ED_key-402-02 ED_key-012-03
ED_key-412-01 ED_key-012-04
ED_key-412-02 ED_key-012-05
ED_key-412-03 ED_key-012-06
ED_key-422-01 ED_key-022-01
ED_key-422-02 ED_key-022-02
ED_key-422-03 ED_jey-022-03
ED_key-422-04 ED_key-022-04
ED_key-432-01 ED_key-032-01
ED_key-442-01 ED_key-032-02
ED_key-442-02 ED_key-032-03
ED_key-442-03 ED_key-032-04
ED_key-442-04 ED_key-042-01
ED_key-462-01 ED_key-042-02
ED_key-462-02 ED_key-042-03
ED_key-472-01 ED_key-042-04
ED_key-472-02 ED_key-042-05
ED_key-472-03 ED_key-052-01
ED_key-472-04 ED_key-052-02
ED_key-472-05 ED_key-052-03
ED_key-472-07 ED_key-052-04
ED_key-482-01 ED_key-062-01
ED_key-482-02 ED_key-062-02
ED_key-492-01 ED_key-062-03
ED_key-492-05 ED_key-072-01
ED_key-492-06 ED_key-072-02
ED_key-502-01 ED_key-082-01
ED_key-502-02 ED_key-082-02
ED_key-502-03 ED_key-092-01
ED_key-512-01 ED_key-092-02
ED_key-512-02 ED_key-102-01
ED_key-512-03 ED_key-102-02
ED_key-512-04 ED_key-112-01
ED_key-522-01 ED_key-112-02
ED_key-522-02 ED_key-112-03
ED_key-532-01 ED_key-112-04
ED_key-532-02 ED_key-112-05

ED key-532-03
ED key-532-04
ED key-532-05
ED key-542-01
ED key-542-02

ED key-122-01
ED key-122-02
ED key-122-03
ED key-132-01
ED key-132-02

<手紙文>

ファイル数 : 25ファイル

| | | |
|---------|---------------|---------------|
| ファイル名 : | ED_ltr-012-01 | ED_ltr-012-14 |
| | ED_ltr-012-02 | ED_ltr-012-15 |
| | ED_ltr-012-03 | ED_ltr-012-16 |
| | ED_ltr-012-04 | ED_ltr-012-17 |
| | ED_ltr-012-05 | ED_ltr-012-18 |
| | ED_ltr-012-06 | ED_ltr-012-19 |
| | ED_ltr-012-07 | ED_ltr-012-20 |
| | ED_ltr-012-08 | ED_ltr-012-21 |
| | ED_ltr-012-09 | ED_ltr-012-22 |
| | ED_ltr-012-10 | ED_ltr-012-23 |
| | ED_ltr-012-11 | ED_ltr-012-24 |
| | ED_ltr-012-12 | ED_ltr-012-25 |
| | ED_ltr-012-13 | |

1 - 2. 検査方法と検査結果

乱数を発生させランダムにファイルを抜き取る。
次に、抜き取ったファイルの形態素総数を見て、その総数より乱数を発生させてランダムに形態素データを1個取り出し、その形態素データの入った文をファイルに出力する。
このような処理を1サイクルとし、全体形態素数の約1%にあたる1530形態素データを抽出した。

- ・ データ件数 153,012 形態素
- ・ 抜き取り形態素数 1,530 形態素
- ・ 不良個数 1 形態素
- ・ 不良率 0.065%
- ・ 抽出日 1989年8月1日
- ・ 検査日 1989年8月2日
- ・ 検査者 立川

1-3. 不良に対する考察

(1) 不良内容

/data3/MORPH/edout/key/ED_key-442-04

／この／ような／高名／な／会議／に／スピーカー／として／、／

助動詞 連用形 → 助動詞 連体形

(2) 不良原因

助動詞 「ようだ」は次のように活用する

| 基本形 | 未然形 | 連用形 | 終止形 | 連体形 | 假定形 | 命令形 |
|-----|------|--------------------|-----|-----|------|-----|
| ようだ | ようだろ | ようだっ ようだ ように | ようだ | ような | ようなら | ○ |

「ような」は連体形にしかないのでシステムは正しく出力してくる。
ゆえに、これは人為的ミスだと言える。

(3) 対策

こういう、人為的誤りは見つけた時に、作業者全員に報告し、その語について、特に注意する。

1 - 4. 品質向上に対する考察

1 - 4 - 1. 考察の要点

前回（1989年2月納品、不良率0.93%）に比べて品質がかなり向上した。
以下にその要因と見られる点を記す

（1）新品詞体系に慣れてきた

- ・マニュアルに書かれてある矛盾や追加は納品ごとに新規事項（取り決め）として統一する事にした為、矛盾しているまま次回に進む事がなくなった。（1-4-2を参照）
- ・作業者が新品詞体系のマニュアルをより深く理解してきた。

（2）形態素抽出プログラム（extract）を形態素のチェック用として使用

- ・これは品詞別に形態素情報が出てくる。この為、違った角度から形態素情報を観察でき、誤りを、発見しやすい。
- ・紙上でチェックする（1-4-3を参照）
（ひらがなの入力ミス、正規表現のゆれなどを特に見つけやすい）
- ・全く同じ形態素情報はまとまって出て来るので、チェックする情報がしぼられる。

（3）システムのバージョンアップ

- ・ヒット率が向上した為、単純な切り方については修正しなくても良くなり、その分ミスが少なくなった。

1-4-2. 新規事項（取り決め）

★ 納品日 …………… 1989.03.31 作成者 …………… 柴田

< 分野 : 手紙文 >

取り決め内容

- ・ データ上に " _ " (under ber) が入っている形態素データの扱い

固有名詞の間に _ が入っているものについて、ひらがな、正規表現では " _ " を削除する

清水_貢

| データ | ひらがな | 漢字 | 品詞 | 活種 | 活形 | 音便 |
|------|--------|-----|-----|----|----|----|
| 清水_貢 | しみずみつぐ | 清水貢 | 固名詞 | 無 | 無 | 無 |

★ 納品日 …………… 1989.04.28 作成者 …………… 柴田

< 分野 : キーボード会話 >

取り決め内容

- ・ " 1 s t アナウンスメント " " 2 s t アナウンスメント " のひらがなはファーストアナウンスメント、セカンドアナウンスメントとする

| データ | ひらがな | 漢字 | 品詞 | 活種 | 活形 | 音便 |
|---|------|----|----|----|----|----|
| 1 s t アナウンスメント ファーストアナウンスメント 1 s t アナウンスメント | | | 名詞 | 無 | 無 | 無 |

取り決め内容

- ・ 明らかに打ち間違えと思われるものについては、品詞は " その他 " にしてコメントを入れる

議事録を買買うことが出来るでしょうか。

**

| データ | ひらがな | 漢字 | 品詞 | 活種 | 活形 | 音便 |
|-----|------|----|-----|----|----|----|
| 買 | かい | 買 | その他 | 無 | 無 | 無 |

※ コメントで others (打ち間違え)

★ 納品日 …………… 1989.05.17 作成者 …………… 山田

< 分野 : 電話会話 >

取り決め内容

- ・ 漢字で " 毎年 " と出てきたら、ひらがなは " まいねん " とする。

品詞は次の通りである。

ひらがなで”まいとし”ならば、”とし”は、名詞とする。

毎年

| データ | ひらがな | 漢字 | 品詞 | 活種 | 活形 | 音便 |
|----------|----------|----|----------|--------|--------|--------|
| 毎年 | まい ねん | 毎年 | 接頭 接尾 | 無 無 | 無 無 | 無 無 |
| まい とし | まい とし | 毎年 | 接頭 名詞 | 無 無 | 無 無 | 無 無 |

・漢字で”毎月”と出てきたら、ひらがなは、”まいつき”とする。

品詞は次の通りである。

毎月

| データ | ひらがな | 漢字 | 品詞 | 活種 | 活形 | 音便 |
|-----|----------|----|----------|--------|--------|--------|
| 毎月 | まい つき | 毎月 | 接頭 接尾 | 無 無 | 無 無 | 無 無 |

★ 納品日 …………… 1989.06.02

作成者 …………… 田中

< 分野 : 電話会話 >

取り決め内容

・”おいで”について

お / いで
接頭 ー

いで (正規表現は いづ) は下二段活用であるが、システムに下二段がエントリーされていないため下一段とし、コメントで次のように表記する。

kassyu (下二段)

いで

| データ | ひらがな | 漢字 | 品詞 | 活種 | 活形 | 音便 |
|-----|------|----|-----|----|----|----|
| いで | いで | いづ | 本動詞 | 下一 | 連用 | 無 |

・”ウ音便”について

よろしゅう / ございます
本動

よろしゅうは、よろしいの文語形であるため、コメントで次のように表記する。

文語 (ウ音便)

よろしゅう

| データ | ひらがな | 漢字 | 品詞 | 活種 | 活形 | 音便 |
|-------|-------|-----|-----|----|----|----|
| よろしゅう | よろしゅう | 宜しい | 形容詞 | 無 | 連用 | ウ音 |

< 分野 : 電話会話 >

取り決め内容

- ・今回から、活用の種類に四段活用、上二段活用、下二段活用、ラ行変格活用、ナ行変格活用、音便にウ音便が加わり、エディターで入力できるように成ったので、此を使用する事にし、コメントで 文語 と表記する。

よって、6月2日の取り決め内容も今回から以下のように変更する。

お / いで
接頭 ———

いで

| データ | ひらがな | 漢字 | 品詞 | 活種 | 活形 | 音便 |
|------|------|----|-----|----|----|----|
| いで | いで | いつ | 本動詞 | 下二 | 連用 | 無 |
| コメント | 文語 | | | | | |

よろしゅう / ございます
————— 本動

よろしゅう

| データ | ひらがな | 漢字 | 品詞 | 活種 | 活形 | 音便 |
|-------|-------|-----|-----|----|----|----|
| よろしゅう | よろしゅう | 宜しい | 形容詞 | 無 | 連用 | ウ音 |
| コメント | 文語 | | | | | |

< 分野 : 電話会話 >

取り決め内容

- ・以前は /とんでも/ない/ としていたものをマニュアル、新明解に
副詞
基づいて 現在、/とんでもない/ としている。
形容詞

”とんでも御座いません” の場合は マニュアル、新明解、学研に
エントリーされていない為、以前の /とんでも/ない/ に基づいて、
副詞
/とんでも/御座いませ/ん/ とする。

ただし、コメントで次のように表記する。

others (副詞を作りました)

とんでも / 御座いませ / ん
本動 助動

とんでも

| データ | ひらがな | 漢字 | 品詞 | 活種 | 活形 | 音便 |
|------|-------------------|------|----|----|----|----|
| とんでも | とんでも | とんでも | 副詞 | 無 | 無 | 無 |
| コメント | others (副詞を作りました) | | | | | |

★ 納品日 …………… 1989.07.31

作成者 …………… 田中

< 分野 : キーボード会話、電話会話 >

取り決め内容

・ ” すべき ” について

マニュアル 13-5 (その他) の (8) 文語の扱いに、

廃止 すべき だ

→ す (文語) 助動「す」の終止

とエントリーされているが、マニュアル 9-5 (助動詞) には、

落胆 すべき

→ す (文語) 補動「す」の終止

とエントリーされている。

この2つを同一と見做し、今回からマニュアル 9-5 に統一する。

す / べき

| データ | ひらがな | 漢字 | 品詞 | 活種 | 活形 | 音便 |
|------|------|----|----|----|----|----|
| す | す | す | 補動 | サ変 | 終止 | 無 |
| コメント | 文語 | | | | | |

・ 今回 キーボード会話に

通訳より訂正 : ” 記号言語 ” は ” 手話 ” の間違いです。

と、会話以外の言葉が出てきた為、通訳より訂正 : を発話マーク扱いとし、以下 すべて品詞単位に切る事にし、それぞれの形態素に

コメント others (会話外) と入れる事にする。

(詳しくは、別紙 - コメント欄の記入方法について - を参照の事)

1-4-3. 形態素抽出プログラム (extract) 出力例

| データ | 正 規 | 正 規 の 読 み | 活用の種類 | 件 数 |
|--------|------|-----------|-------|-----|
| [記号] | | | | |
| 。 | | | | 35 |
| 、 | | | | 16 |
| 質問者： | | | | 14 |
| 通訳者： | | | | 14 |
| ？ | | | | 1 |
| [形容詞] | | | | |
| 宜しい | 宜しい | よろしい | | 2 |
| 詳しい | 詳しい | くわしい | | 1 |
| いい | 良い | いい | | 1 |
| [普通名詞] | | | | |
| 登録 | 登録 | とうろく | | 4 |
| 会議 | 会議 | かいぎ | | 3 |
| 申込 | 申込み | もうしこみ | | 2 |
| 料 | 料 | りょう | | 2 |
| ドル | ドル | ドル | | 2 |
| 申し訳 | 申訳 | もうしわけ | | 2 |
| 事 | 事 | こと | | 2 |
| こと | 事 | こと | | 1 |
| 8月 | 8月 | はちがつ | | 1 |
| とき | 時 | とき | | 1 |
| 円 | 円 | えん | | 1 |
| 通し | 通し | とおし | | 1 |
| ホテル | ホテル | ホテル | | 1 |
| 払い戻し | 払戻し | はらいもどし | | 1 |
| 同時 | 同時 | どうじ | | 1 |
| サービス | サービス | サービス | | 1 |
| 方 | 方 | かた | | 1 |
| 会場 | 会場 | かいじょう | | 1 |
| 他 | 外 | ほか | | 1 |
| [サ変名詞] | | | | |
| 参加 | 参加 | さんか | | 3 |
| 運営 | 運営 | うんえい | | 1 |
| [代名詞] | | | | |
| 何 | 何 | なに | | 1 |
| 私 | 私 | わたし | | 1 |
| [数詞] | | | | |
| 100 | 100 | ひゃく | | 2 |
| 三 | 三 | みっ | | 1 |
| 二 | 二 | ふつ | | 1 |
| 三 | 三 | さん | | 1 |
| 二 | 二 | に | | 1 |
| [副詞] | | | | |
| まず | 先 | まず | | 1 |
| 幾ら | 幾等 | いくら | | 1 |
| もし | 若し | もし | | 1 |
| そう | そう | そう | | 1 |
| 全然 | 全然 | ぜんぜん | | 1 |
| どうも | どうも | どうも | | 1 |
| 特に | 特に | とくに | | 1 |

| | | | | |
|--------|-------|-------|--|----|
| [連体詞] | | | | |
| その | 其 | その | | 1 |
| [接続詞] | | | | |
| それでも | それでも | それでも | | 1 |
| では | では | では | | 1 |
| [感動詞] | | | | |
| はい | はい | はい | | 4 |
| もしもし | もしもし | もしもし | | 2 |
| ええ | ええ | ええ | | 1 |
| ありがとう | ありがとう | ありがとう | | 1 |
| [助動詞] | | | | |
| ます | ます | ます | | 17 |
| です | です | です | | 16 |
| ん | ん | ん | | 6 |
| う | う | う | | 5 |
| た | た | た | | 5 |
| ない | ない | ない | | 4 |
| たい | たい | たい | | 3 |
| だ | だ | だ | | 2 |
| [助詞] | | | | |
| だけ | だけ | だけ | | 1 |
| [接続助詞] | | | | |
| が | が | が | | 8 |
| て | て | て | | 8 |
| ば | ば | ば | | 1 |
| [格助詞] | | | | |
| を | を | を | | 8 |
| で | で | で | | 8 |
| に | に | に | | 5 |
| の | の | の | | 4 |
| と | と | と | | 2 |
| が | が | が | | 2 |
| から | から | から | | 1 |
| [終助詞] | | | | |
| か | か | か | | 10 |
| ね | ね | ね | | 1 |
| [接尾辞] | | | | |
| 語 | 語 | ご | | 4 |
| ドル | ドル | ドル | | 2 |
| 日間 | 日間 | か | | 2 |
| 代目 | 代目 | かん | | 1 |
| 費分 | 費分 | だい | | 1 |
| 者 | 者 | め | | 1 |
| | | ひ | | 1 |
| | | ぶん | | 1 |
| | | しゃ | | 1 |
| [接頭辞] | | | | |
| 御 | 御 | お | | 7 |
| お | 御 | お | | 5 |

| | | | |
|--------|-------|-------|----|
| [補助動詞] | | | |
| する | する | する | 10 |
| くださる | くださる | くださる | 3 |
| 致す | いたす | いたす | 2 |
| いる | いる | いる | 1 |
| 居る | おる | おる | 1 |
| できる | できる | できる | 1 |
| 戴く | いただく | いただく | 1 |
| みる | みる | みる | 1 |
| ございます | ございます | ございます | 1 |
| おる | おる | おる | 1 |
| [固有名詞] | | | |
| 英 | 英 | えい | 2 |
| アメリカ | アメリカ | アメリカ | 1 |
| 日本 | 日本 | にほん | 1 |
| [本動詞] | | | |
| 分る | 分る | わかる | 5 |
| 有る | 有る | ある | 4 |
| 出来る | 出来る | できる | 4 |
| 願う | 願う | ねがう | 3 |
| 支払う | 支払う | しはらう | 2 |
| 返す | 返す | かえす | 2 |
| 御座います | 御座います | ございます | 2 |
| 知る | 知る | しる | 1 |
| 教える | 教える | おしえる | 1 |
| 送る | 送る | おく | 1 |
| 聞く | 聞く | きく | 1 |
| 構う | 構う | かまう | 1 |
| 為る | 為る | なす | 1 |
| 含む | 含む | ふくむ | 1 |
| なる | なる | なる | 1 |
| 言う | 言う | いう | 1 |
| す | 話す | はなす | 1 |
| 頼める | 頼める | たのめる | 1 |
| [準体助詞] | | | |
| の | の | の | 8 |
| [並列助詞] | | | |
| と | と | と | 2 |
| [係助詞] | | | |
| は | は | は | 8 |
| も | も | も | 1 |

2. 「で」に対する考察と形態素情報の付け方

2 - 1. 概要

形態素解析作業において判断が明確でない為に統一が取りにくかった「で」についてデータを集め基準を明確にした。

「で」には次の4種類がある。

- [1] 接続詞の「で」
- [2] 接続助詞の「で」
- [3] 格助詞の「で」
- [4] 助動詞「だ」の連用形である「で」

この中で[1]と[2]は判断がつき易い。しかし、[3]と[4]の区別は非常に難しい。そこで、これらを明確にする為に新明解第三版・学研の国語大辞典第二版及び形態素マニュアルをもとに過去のデータすべてを分類分けし、統一をとる事にした。

[3]と[4]の分類分けの例を以下に記すと共に具体例を添付する。

2 - 2. 「で」の品詞と具体例

| 品詞 | 基準 | 例文 |
|-----|---|---|
| [1] | 接続詞 [それで]と同様 | 会議は九時から始まりました。で、どんな事が決まりましたか。 |
| [2] | 用言の連用形に付く「で」の前が原因・理由となり、「で」の後ろが結果となる。 | 皆が騒い [で] うるさい。 ころん [で] 痛い。 読ん [で] 感動した。 |
| [3] | <ul style="list-style-type: none"> ・状態 ・原因・理由 ・時点 ・方法・手段・材料 ・動作の主体 ・場所・場面 ・限られた語につけて動作の主体を間接的に表す | (1) で記す (2) で記す (3) で記す (4) で記す (5) で記す (6) で記す (7) で記す |
| [4] | <ul style="list-style-type: none"> ・並列しているもの ・中止法の「で」 ・後ろに "ある" 及び "ない" が続くもの ・形容名詞に連なるもの | (8) で記す (9) で記す (10) で記す (11) で記す |

(1) 格助詞 どういう状態で行われるか

< 新明解・学研の用例 >

みんな [で] やろう

ふたり [で] でかけた

フルスピード [で] 走った

三つ [で] 百円

千円 [で] つりをくれ

申込みは明日 [で] 締め切る

三日 [で] 仕上げる

明かりなし [で] は探せない

十日に一編の割 [で] けんかをしていた

< 言語データ (形態素) から抽出したもの >

それから、私の手書 [で] 読めないところがあり

独りぼっち [で] 住んでいた

先生がおわかりになる限り [で] 結構ですので早々に事務局まで

会議の参加申込み用紙は先生のお名前 [で] 届いておりますが

どんなサイズ [で] 掲載されるのか

配偶者のみこの名称 [で] 同行できるとも聞きました

こちらから寄附金なし [で] 参加登録のみで

午後7時のフライト [で] ファーストクラスのウェイトングリストに

国際コンピュータ会議の件 [で] お手紙差し上げます

東京都豊島区東池袋3丁目2番5号 [で] よろしいですね

コンピュータ業界の専門紙でありまして普通 [で] は得られない情報や

おそらく15分以内 [で] ご連絡できると思います

5泊6日 [で] 予約させて頂きます

そうですね、2枚くらい [で] 結構かと思います

貴局はただ許可して下さるだけ [で] 結構であります

外務省の方に申請をするつもり [で] おります

そのつもり [で] はいるんです

郵便物を処理するだけ [で] 優に半日はかかる

私の発表が、完全な形 [で] お見せ出来るよう正式な

40分程度 [で] お願いしているのですが

以上 [で] ご質問にはすべてお答えさせて

ジーパンにティーシャツといった格好 [で] 駅まで自転車を走らせる

こちらの時間 [で] 木曜日の朝までに

知的な意味 [で] 非常に刺激的な会議になることと思います

早春の味と香を、素朴な風味 [で] 味あわせてくれるものです

横目 [で] ちらっと見れば主人の立ち上がる気配がする

一般参加ということ [で] お願いしているんですけども

内容が重複しないような形 [で]、私の専門の方を

会議がもう2ヶ月後ということ [で] 伺ってますが、どうでしょうか

一律料金ということ [で] ご理解頂ければと存じます

十五時から十五時半ということ [で]、休憩の方入れさせていただきます

処理学会の会員ということ [で] 特別料金は設定してない

講演依頼という形 [で] レターが来てるんです

登録料をカードでお支払いということ [で] 手続きさせていただきます

当然申込という形 [で] 受け付ける必要があるかどうか

最終の打ち合せをすること [で] いかがでしょうか

一般は八万五千円ということ [で] 決めさしていただいております

もう一台スクリーンということ [で] お使いいただくこと可能

送らせて頂くということ [で] 今回の御連絡を終らせて頂き

官産学の一体ということ [で] 行われるもので

(2) 格助詞 どんな原因・理由で行われるか

< 新明解・学研の用例 >

病氣 [で] 休む

受験勉強 [で] 忙しい

おかげ [で] 助った

そういう事 [で] 今は手一杯だ

流行の疱疹 [で] 死んだ

私たちを見つけた喜び [で] 真裸のまま

< 言語データ (形態素) から抽出したもの >

家族の都合 [で] 会議終日の30日に日本を発ちたい

健康あるいは身体障害者等の都合 [で] お一人でお越しになれない

写真は製版の都合 [で] 一番急いでおります

熱 [で] うるんだ青い空に積乱雲がある

不眠不休の徹夜続き [で] 目が腫れあがっている

軍の依頼 [で] 中国人運転手の通訳をしながら

お陰様 [で] シカゴで行われましたIEEE主催の

おかげさま [で] 完成することができ

運営を進める、という方針 [で] この参加費の設定がなされた

こうした意味 [で] 機械文明との新しい調和を考えなければ

これでは、母親がなげくのも無理はありません

会議に出席するつもり [で] リプライカードをお送り致します

主人の仕事は建築業 [で] 作業服を着て出勤します

露と霧 [で] ぐっしょり砂がぬれて

会議の準備 [で] お忙しいところ誠に恐縮ですが

今年の会議そのもの [で] 大分忙しいと思いますので

中学は弁当制 [で] 持って行けない場合は

カルシウムが含まれていること [で]、現代の食生活の中で

病状が重くないという事 [で] 少し安心しました

そういうこと [で]、そちらの方からお送りいただく

そういうこと [で]、バイバイ、どうもありがとうございました

変更ということ [で] 実は非常に困るんです

(3) 格助詞 どんな時点で問題にされるか

< 新明解・学研の用例 >

今日 [で] は月旅行は夢ではなくなった

現在 [で] はなんらの疑惑もいだいていない

八十三歳 [で] 死んだ

< 言語データ (形態素) から抽出したもの >

スライドの発表 [で] は自作のカラービデオフィルムを使用

現在の時点 [で] 貴社用にと予定しております

29日の夕刻 [で] はいかがでしょうか

その他の日時 [で] は講師の方々の出席が難しい

会議が始まってから [で] は遅いですか

(4) 格助詞 どんな方法・手段・材料で行われるか

< 新明解・学研の用例 >

鉛筆 [で] 書く

手料理 [で] ごちそうする

米 [で] 酒をつくる

ラジオ [で] 聞いた話

濡れ手 [で] 粟をつかむ

私から [で] いけませんか

貨物船 [で] 南米に行く

< 言語データ (形態素) から抽出したもの >

我慢しきれず大声 [で] 怒鳴りつけ、その日の撮影は中止に

その反射光を半透明なフィルム [で] 受ける

鉢植え [で] 育てているツツジ類、ツバキ、アセビなど

指紋 [で] よごれ、ぐにゃぐにゃになってしまうようだ

午後四時の電車 [で] 帰りますから牛乳をあたためて

その旨文書 [で] 御通知下さいまして深くお礼申し上げます

少年は指をズボン [で] ふくと、素直にいそいそと立ち上がった

発表が日本語 [で] 行われる場合、英語への同時通訳は

20部別便 [で] お送りさせて頂きました

どのような書式 [で] 書けばいいんですか

登録用紙 [で] 手続をして下さい

参加料と出展料だけ [で] 運営しております

収入をやりくりして、現金 [で] サービスを利用している

フロアプラン [で] 問題ございませんでしょうか

シュロ縄など [で] かたく結束し、傷口を

その分はその分で別便 [で] ですね

電信振込 [で]、下記口座へ送金いただきたいのです

朝鮮ニンジンの細胞を、タンク培養 [で] 迅速、かつ大量に増やす

救急車 [で] 病院に運ばれる途中に息をひきとった

(5) 格助詞 動作・作用の主体を表わす

< 新明解・学研の用例 >

彼の説 [で] はこうなっている

野党側 [で] は強い反対を示した

気象庁 [で] は颱風警報を出した

これは私たち [で] 考え出したやりかたです

< 言語データ (形態素) から抽出したもの >

これらリモコンは当方 [で] 用意致します

当方 [で] は「同伴者」はあくまでも配偶者

貴局 [で] は「アカンパニング・パーソン」をどう既定して

手配は弊社の方 [で] させていただきますので

こちら [で] は専門的な質問にお答えできません

ほかの企業の方 [で] 御参加される方

1企業 [で] 1部屋まるまる予約するということは控え

営利団体と非営利団体の間 [で] 区別がなされるべきだ

先生個人 [で] 支払って頂く費用との区別ですが

しかしながら、植物 [で] は、受精した卵細胞からの固体

高等動物 [で] は植物に見られるような体細胞からの

(6) 格助詞 どんな場所・場面で行われるか

< 新明解・学研の用例 >

日本〔で〕万博が行われた

教室〔で〕の学習態度

最終審査〔で〕はねられた

法廷〔で〕白黒を争う

団子坂〔で〕お目に懸ったの

< 言語データ(形態素)から抽出したもの >

ラスベガス〔で〕開かれることになっております

雪解けの山の日だまり〔で〕は、フクジュソウやミスミソウなど

大阪〔で〕の会議が開催されますホテルにて

さんご礁の海辺〔で〕は天然のものも多く取れますが

肩の汗をぬぐって支度部屋〔で〕呼吸を整える

そこ〔で〕皆様方とも話し合いました結果

貴社〔で〕の出展内容が決定次第

植物の細胞をタンク内〔で〕大量に培養して、化学合成するのが

当日集合場所〔で〕お支払い下さい

当事務局内〔で〕決定出来ることではございません

すぐそば〔で〕、はあはあ弾ませている呼吸だけが

このあたり一帯〔で〕は、古くから、結婚式とか葬式

ひびのはいった直線コース〔で〕、子供がキャッチボールをやっている

4～5月は、各地〔で〕演芸即売会が開催されます

人かげのたえた所〔で〕は、スズメもやがてすがたを消すだろう

秋にはもう少し遠く〔で〕見られることはあるが

地下一階の中小会議室〔で〕行なう予定です

自動車や精密機械など〔で〕技術系の採用を抑えるところが目立ち

電話番号は大学校内〔で〕六三四、八九、六三四三です

I C C S会議〔で〕のサテライト・シンポジウムの開催を

会議〔で〕扱う話題に関して

会議の晩餐会〔で〕スピーチをするようにとの依頼を受け

それらの分野〔で〕、特に威力を発揮するということが

(7) 格助詞 組織や団体名、あるいは「お隣」「お宅」など限られた語につけて
副詞句の形で述部の表わす動作の主体を間接的に表わす

< 新明解・学研の用例 >

お隣 [で] おはぎを持って来て下さった

< 言語データ (形態素) から抽出したもの >

該当する形態素なし

(8) 助動詞 並列しているもの

< 新明解・学研の用例 >

今日は十日 [で] 金曜日だ

外国人 [で] 日本に居る人

兄は政治家 [で] 弟は芸術家だ

< 言語データ (形態素) から抽出したもの >

半農半漁の静かな町 [で]、かつお漁の母港としても知られている

色白で透けるような肌 [で]、やや薄い色の瞳であった

貯蔵してあるダリア [で]、すでに芽の出ているものは

もずくは本州のものにくらべて、太目 [で] 柔らかいのが特徴ですが

エージェントは JTB [で] 電話番号は 353-1234 です

テーブルのサイズは 2 平方メートル [で] 料金は 100,000 円です

講師および座長の方 [で] 配偶者がいっらしやらない

個人のもの以外 [で] 会社の小切手ではだめでしょうか

背の高さは 170 センチほど [で] 目が大きくやや太っている

旗多郡佐賀町の中心となるところ [で] 南に土佐湾を見、

- (9) 助動詞 中止法 「です・だ」に言い換えておかしくないもので あとに対応する「です・だ」が来ている

< 新明解・学研の用例 >

各々ダブル [で] 期間は 8 / 5 ~ 8 / 8 までです

この国際会議は第一回目のもの [で] 同時通訳に関するもので

< 言語データ (形態素) から抽出したもの >

人為的には起こし得る訳 [で] これは動物細胞とは非常に異なる
暖房は暑すぎるくらい [で] 部屋を出る時に毛布をかけて
娘のフィアンセ [で] こいつだけにはどうしても負けられない
誰もが持っている潜在的能力の一つ [で] 一つの細胞から
クシャミをされるといわれる程 [で] 日本経済も不況です
品種改良とか育種といわれる領域 [で] もう一つは大量クローン増殖と
時間はおよそ一時間から一時間半程 [で] 出来れば会議のメイン会場の近く
婚家は実家の近く [で] ハイヤーを使うほどの距離ではなかった
来た年賀状は三百枚程度 [で] 丁度出した分と同じぐらいだ
セッション会場はホテル 2 階の鳳凰の間 [で] そしてセブションは
10月27日より30日まで [で] 会場はホテルニューオータニで
内容も最先端に位置するもの [で] 自分の携わる仕事にも関係
午後6時よりはじまりますウェルカムレセプションのみ [で] 翌日28日

(10) 助動詞 後ろに“ある”及び“ない”が続くもの

< 新明解・学研の用例 >

会議に出席する訳 [で] はありません

それは鳥 [で] もない

バルバーン先生 [で] いらっしゃいますね

< 言語データ (形態素) から抽出したもの >

実際そう [で] もなければ五年間もこの個室で

選択方法の改善で収まるもの [で] はない

われら凡人はスキャンダルが嫌い [で] はない

過大評価することは適切 [で] はあるまい

それぞれ出世をしている様子 [で] あった

どちらの団体の会員 [で] いらっしゃいますか

知識水準に問題があるの [で] ございます

パソ・ローブルズの学会員 [で] ございます

あいにく [で] はございますが、即答しかねます

(11) 助動詞 形容名詞+助動詞「だ」の連用形

< 言語データ(形態素)から抽出したもの >

これほど躍動感がなくて、静か[で]沈痛な胸上げというもの
赤子のように無欲[で]愛らしい表情をする父に
女性とは逆[で]、何とか常識を破って
乱暴[で]乱暴でゆく先が案じられると母が
十歳年上のご主人がお元気[で]家事一切をなさっていました
色白[で]、透けるような肌で、やや薄い色の
あけっぴろげ[で]こだわりのないおおらかな気分が
パリから大阪行きのチケットは変更不要[で]よろしいのでしょうか
以前はパンより御飯が好き[で]、喜んでお代わりをした
グラフのことで不行き届き[で]御迷惑をかけて申し訳ないんです
遊びも勉強もマイペース[で]、帰るべき時間にはきちんと
人間とは微妙[で]複雑な生き物である

3. 不良テキストに対する取り扱い

3 - 1. ゆれに対する問題点

形態素マニュアル 2-7 には

{ ご、お } + < 1 > + 補助動詞

< 1 > 尊敬表現では接頭語のあとにくるのは名詞、サ変名詞、動詞の連用形、などがあり、この順に名詞的色彩が薄れる。このタイプ（うしろに補助動詞がくるタイプ）は名詞的色彩のうすいもの（サ変名詞、動詞・連用）を好む。「ご」のあとには漢語すなわちサ変名詞が、「お」のあとには動詞の連用形がくるというのが原則である。

このときの動詞の連用形のなかには名詞としての用法が辞書に記載されているものがある。しかし、このタイプの尊敬表現においては名詞としては取り扱わない。

ex.1 経理担当の方にお話しできますでしょうか

動詞

ex.2 喜んでお引受け致しました

動詞

ex.3 どのようにお支払いしたらよろしいんでしょう

動詞

ex.4 よろしくお願ひします

動詞

ex.5 お電話下さい

サ名

ex.6 お返事下さい

サ名

と書いてある。従ってこのマニュアルの例に従えば次の [] 内は動詞になり、語尾変化が漢字表記に含まれる動詞が存在する事になる。

ex.1 お [話] したいんですが、お [話] できますでしょうか。

ex.2 本国の税務省にお [支払] して頂く事になっておりますが

ex.3 セカンドアナウンスメントを御 [受取] 致しました

こういう語が出現した場合の対処について

案 1 . 活用語尾のない動詞にする

案 2 . 名詞にする

案 3 . 原文を修正する

以下にその他の例を記す

●動詞形 → 話す 名詞形 → 話

この件で、先生ともお〔話〕しました結果、

では、その時にまたお〔話〕いたしましょう

講演をしておりますジェンファー先生とお〔話〕できないかと思ひまして

●動詞形 → 引(き)受ける 名詞形 → 引受け

会議の企画委員長をお〔引受〕頂いておひまして、

●動詞形 → 支払う 名詞形 → 支払(い)

日本円では幾らお〔支払〕したらいいのですか？

●動詞形 → 願う 名詞形 → 願(い)

是非、お〔願〕します

宜しくお〔願〕致します

ご住所とお名前をお〔願〕致します

●動詞形 → 受取る 名詞形 → 受取

小切手を〔受取〕次第参加登録証をお送り致します

登録料はドルでなければ御〔受取〕できません

●動詞形 → 振込む 名詞形 → 辞書にナシ

御〔振込〕されてから私どもの口座への

お〔振込〕頂ければ非常に嬉しく思ひます

不足分の400円を至急、お〔振込〕下さい

ブローシーディング代を差し引いた登録料を〔振込〕ます。

●動詞形 → 待つ 名詞形 → 辞書にナシ

ちょっと書き取りますのでお〔待〕いただけますか

はい、お〔待〕致しております

当日、お〔待〕しております

3 - 2. 不良対象報告

- ・ 発話者マークが間違っている時がある
- ・ 文章が句点で終わっていないため、次に来る発話マークが、発話マーク扱いにならない
- ・ ブランクが入っている
- ・ 固有名詞（地名・人名）に読みがなが無いため、品質が落ちる
- ・ ミスタイプと思われるテキストの品詞付けに困る
- ・ 日本語テキスト内に英字、ブランクの半角が存在している

形態素情報データファイル(4)

－ 品質管理報告 －

1990年 2月

< 目次 >

1. 形態素情報の抜き取り検査

1 - 1. 検査対象 1

1 - 2. 検査方法と検査結果 6

1 - 3. 不良に対する考察 7

2. 問題点

2 - 1. 固有名詞の読みについて 9

2 - 2. 固有名詞の品詞付けについて 9

3. 相槌について 10

1. 形態素情報の抜き取り検査

1-1. 検査対象

■ 作業期間 1989年8月1日 ~ 1990年1月31日

■ 会話分野と形態素数

| 会話分野 | 形態素数 | ファイル数 | 1ファイル数平均 (単位:形態素数) |
|---------|---------|-------|-----------------------|
| 電話会話 | 162,170 | 98 | 1,655 |
| キーボード会話 | 37,898 | 41 | 924 |
| SDB会話 | 2,070 | 4 | 5,175 |
| ※※※※※※※ | 202,138 | 143 | ※※※※※※※ |

(注) ファイル数は会話数を表す

■ 対象ファイル

< 電話会話 >

ファイル数 : 98ファイル

| 形態素数 | : 画面数 | : ファイル名 |
|------|-------|---------------|
| 967 | 271 | ED_tel-032-06 |
| 736 | 193 | ED_tel-032-07 |
| 1184 | 316 | ED_tel-032-08 |
| 1132 | 317 | ED_tel-032-09 |
| 1179 | 322 | ED_tel-032-10 |
| 1407 | 346 | ED_tel-032-11 |
| 897 | 248 | ED_tel-032-12 |
| 708 | 194 | ED_tel-032-13 |
| 1382 | 331 | ED_tel-032-14 |
| 1163 | 323 | ED_tel-032-15 |
| 1429 | 388 | ED_tel-032-16 |
| 1609 | 391 | ED_tel-032-17 |
| 1558 | 425 | ED_tel-032-18 |
| 1018 | 255 | ED_tel-032-19 |
| 2083 | 830 | ED_tel-042-01 |
| 1590 | 590 | ED_tel-042-02 |
| 2587 | 995 | ED_tel-042-03 |
| 2102 | 815 | ED_tel-042-04 |
| 4346 | 1613 | ED_tel-042-05 |
| 2331 | 783 | ED_tel-042-06 |
| 1337 | 515 | ED_tel-042-07 |
| 1096 | 392 | ED_tel-052-01 |
| 1380 | 500 | ED_tel-052-02 |
| 2560 | 814 | ED_tel-052-03 |
| 2165 | 747 | ED_tel-052-04 |
| 2579 | 854 | ED_tel-052-05 |
| 2702 | 878 | ED_tel-052-06 |
| 2725 | 835 | ED_tel-052-07 |
| 1871 | 640 | ED_tel-052-08 |
| 721 | 234 | ED_tel-062-01 |
| 313 | 118 | ED_tel-062-02 |
| 531 | 207 | ED_tel-062-03 |
| 419 | 148 | ED_tel-062-04 |
| 255 | 82 | ED_tel-062-05 |
| 134 | 52 | ED_tel-062-06 |
| 525 | 192 | ED_tel-062-07 |
| 1275 | 446 | ED_tel-062-08 |
| 515 | 171 | ED_tel-062-09 |
| 636 | 130 | ED_tel-082-05 |
| 694 | 173 | ED_tel-082-06 |
| 1079 | 216 | ED_tel-082-07 |
| 757 | 205 | ED_tel-092-01 |
| 803 | 223 | ED_tel-092-02 |
| 794 | 223 | ED_tel-092-03 |
| 578 | 145 | ED_tel-092-04 |
| 609 | 162 | ED_tel-092-05 |
| 358 | 67 | ED_tel-102-01 |
| 609 | 110 | ED_tel-102-02 |
| 636 | 139 | ED_tel-102-03 |
| 584 | 113 | ED_tel-102-04 |
| 314 | 72 | ED_tel-102-05 |
| 778 | 153 | ED_tel-102-06 |
| 657 | 166 | ED_tel-112-01 |
| 655 | 176 | ED_tel-112-02 |
| 668 | 192 | ED_tel-112-03 |

| | | |
|--------|-------|---------------|
| 842 | 220 | ED_tel-112-04 |
| 989 | 244 | ED_tel-112-05 |
| 2668 | 312 | ED_tel-352-01 |
| 3415 | 457 | ED_tel-352-02 |
| 2743 | 324 | ED_tel-352-03 |
| 1135 | 155 | ED_tel-352-04 |
| 1748 | 219 | ED_tel-352-05 |
| 4375 | 633 | ED_tel-352-06 |
| 2350 | 348 | ED_tel-352-07 |
| 2099 | 321 | ED_tel-352-08 |
| 2560 | 430 | ED_tel-352-09 |
| 2541 | 346 | ED_tel-352-10 |
| 4632 | 776 | ED_tel-352-11 |
| 2700 | 400 | ED_tel-352-12 |
| 1246 | 194 | ED_tel-362-01 |
| 2187 | 335 | ED_tel-362-02 |
| 945 | 110 | ED_tel-362-03 |
| 3267 | 318 | ED_tel-362-04 |
| 4115 | 376 | ED_tel-362-05 |
| 2394 | 252 | ED_tel-362-06 |
| 1657 | 158 | ED_tel-362-07 |
| 2904 | 319 | ED_tel-362-08 |
| 4433 | 435 | ED_tel-362-09 |
| 2924 | 316 | ED_tel-362-10 |
| 1587 | 185 | ED_tel-372-01 |
| 1517 | 160 | ED_tel-372-02 |
| 1561 | 162 | ED_tel-372-03 |
| 1672 | 188 | ED_tel-372-04 |
| 2714 | 278 | ED_tel-382-01 |
| 1573 | 184 | ED_tel-382-02 |
| 1385 | 147 | ED_tel-382-03 |
| 1893 | 187 | ED_tel-382-04 |
| 1510 | 200 | ED_tel-392-01 |
| 1566 | 198 | ED_tel-392-02 |
| 1123 | 114 | ED_tel-392-03 |
| 3739 | 442 | ED_tel-392-04 |
| 2196 | 253 | ED_tel-392-05 |
| 1933 | 228 | ED_tel-402-01 |
| 1831 | 231 | ED_tel-402-02 |
| 2355 | 267 | ED_tel-402-03 |
| 1447 | 186 | ED_tel-402-04 |
| 983 | 102 | ED_tel-402-05 |
| 2013 | 237 | ED_tel-412-01 |
| 162170 | 32383 | total |

■ 対象ファイル

< キーボード会話 >

ファイル数 : 41ファイル

| 形態素数 | 画面数 | ファイル名 |
|-------|------|---------------|
| 1015 | 106 | ED_key-692-01 |
| 1256 | 143 | ED_key-692-02 |
| 860 | 104 | ED_key-702-01 |
| 532 | 68 | ED_key-702-02 |
| 606 | 68 | ED_key-702-03 |
| 1051 | 128 | ED_key-702-04 |
| 1260 | 150 | ED_key-702-05 |
| 1152 | 133 | ED_key-702-06 |
| 703 | 88 | ED_key-702-07 |
| 766 | 87 | ED_key-712-01 |
| 868 | 100 | ED_key-712-02 |
| 776 | 98 | ED_key-712-03 |
| 972 | 105 | ED_key-722-01 |
| 1010 | 128 | ED_key-722-02 |
| 1208 | 139 | ED_key-722-03 |
| 808 | 88 | ED_key-732-01 |
| 612 | 86 | ED_key-732-02 |
| 1041 | 118 | ED_key-742-01 |
| 1133 | 137 | ED_key-742-02 |
| 857 | 91 | ED_key-742-03 |
| 911 | 118 | ED_key-752-01 |
| 680 | 88 | ED_key-752-02 |
| 1053 | 118 | ED_key-752-03 |
| 959 | 110 | ED_key-752-04 |
| 1059 | 127 | ED_key-762-01 |
| 719 | 84 | ED_key-762-02 |
| 924 | 114 | ED_key-762-03 |
| 899 | 99 | ED_key-772-01 |
| 980 | 104 | ED_key-772-02 |
| 1257 | 152 | ED_key-772-03 |
| 965 | 108 | ED_key-772-04 |
| 926 | 124 | ED_key-772-05 |
| 920 | 108 | ED_key-782-01 |
| 572 | 64 | ED_key-782-02 |
| 983 | 120 | ED_key-782-03 |
| 1098 | 151 | ED_key-782-04 |
| 920 | 126 | ED_key-792-01 |
| 891 | 106 | ED_key-792-02 |
| 940 | 116 | ED_key-802-01 |
| 731 | 85 | ED_key-802-02 |
| 1025 | 127 | ED_key-802-03 |
| 37898 | 4514 | total |

■ 対象ファイル

< SDB会話 >

ファイル数 : 4ファイル

| 形態素数 | 画面数 | ファイル名 |
|------|-----|---------------|
| 588 | 47 | ED_sdb-032-01 |
| 393 | 40 | ED_sdb-032-02 |
| 602 | 60 | ED_sdb-032-03 |
| 487 | 66 | ED_sdb-032-04 |
| 2070 | 213 | total |

1 - 2. 検査方法と検査結果

乱数を発生させランダムにファイルを抜き取る。
次に、抜き取ったファイルの形態素総数を見て、その総数より乱数を発生させてランダムに形態素データを1個取り出し、その形態素データの入った文をファイルに出力する。
このような処理を1サイクルとし、全体形態素数の約1%にあたる2,021形態素データを抽出した。

- ・ データ件数 202,138形態素
- ・ 抜き取り形態素数 2,021形態素
- ・ 不良個数 2形態素
- ・ 不良率 0.099%
- ・ 抽出日 1990年2月1日
- ・ 検査日 1990年2月2日
- ・ 検査者 立川

1-3. 不良に対する考察

不良形態素は「見れる」と「遣る」の2つであった。
以下に原因を説明する。

■ 「見れる」について

(1) 不良内容

/data3/MORPH/product/tel/ED_tel-352-12

(誤) 妻籠とか、 / 見 / れる / コースできますか。

(正) 妻籠とか、 / 見れる / コースできますか。

(2) 不良原因

形態素作業マニュアル2-3に、可能動詞について下記のような記述があるため「見 / れる」は誤りである。

可能動詞について

可能動詞は一まとめで動詞とみる。

eg. よめる → /よめる/
動詞

みれる → /みれる/
動詞

ただし、

みられる → み/られる
動詞 助動詞

今回のミスはこれらを取り違えた人為的ミスだと言える

(3) 対策

こういう人為的ミスはすべてを解消する事は不可能に近いが、これまでの検査報告書に記述してきた誤りについては重点的にチェックしたい。

(注) 今回の誤りと同じものに使役動詞がある。

■ 「遣る」について

(1) 内容

/data3/MORPH/product/key/ED_key-772-05

夜はバーベキューなんかも / やり / たいんで。

| | データ | ひらがな | 漢字 | 品詞 | 活種 | 活用形 | 音便 |
|-----|-----|------|----|-----|----|-----|----|
| (誤) | やり | やり | 遣る | 本動詞 | 五段 | 連用形 | 無 |
| (正) | やり | やり | 遣る | 本動詞 | 五段 | 連用形 | 無 |

漢字（正規表現）が間違っている

(2) 不良原因

1. 新明解の電子化辞書（ /data1/MEIKAI ）の入力ミス箇所が形態素用辞書にそのままエントリーされていたため
2. よく似た漢字の為、見落としていた

(3) 対策

1. 形態素解析辞書については、誤り箇所を訂正した
2. よく似ている漢字は特に注意したい
たとえば

覧 / 監

膳 / 騰

荻 / 萩

送 / 迭

幣 / 弊

緑 / 縁

博 / 搏 / 搏

虜 / 慮

項 / 頃

租 / 祖

2. 問題点

2 - 1. 固有名詞の読みについて

/data3/MORPH/product/tel/ED_tel-392-05

三島から沼津に出まして、

- ・ 固有名詞の読みについては原文に入れている半角文字及び英文を参照する事になっている。
「沼津」については原文に半角のよみがなく、英文には下記のようなった。

< 英文 >

from Mishima to Numazu,

このため、現在は「ぬまつ」とせずに、「ぬまず」としているが
いいのでしょうか？

2 - 2. 固有名詞の品詞付けについて

/data3/MORPH/product/tel/ED_tel-362-01

添乗員が付きまして、{はい} [あの一] 西海岸をご案内させていただく

- ・ これは話の流れから推察すると、アメリカの西海岸と思われる。
..... 固有名詞

西海岸という単語についてだけみると、日本にも西海岸はある。
..... 普通名詞 + 普通名詞

現在は、「西 / 海岸」にしている。いいのでしょうか。

普通名詞 + 普通名詞

3. 相槌について

11月30日納品分より、相槌を規定するようになり、ATR側にストックしてある原文については下記のように相槌を取り入れた

原文
|
ATR 文の認定をする
|
TIS

- ・相槌を中括弧にいれる
- ・ATR側で1文とみなされたものの中に、句点があるときは、句点の前が感動詞、接続助詞なら読点に変え、それ以外は句点を取り除く。 **** (1)
- ・ATR側で1文とみなされたものの中に、発話マークがあるときは取り除く。

tel-402-02

< 原文 >

担当者：はい。
では[えーと]人数は何名様っておっしゃってましたでしょう。
申込者：[あの]3人で。
担当者：はい。
申込者：[あの]子供が小学校6年生なんですけれども。
担当者：3名様で6年生。
申込者：はい。
担当者：そうすると、ベッドがみつつあるお部屋をひと部屋、ご用意させていただきます。
申込者：はい。
担当者：よろしいわけですね。

< ATR >

担当者：はい。
では[えーと]人数は何名様っておっしゃってましたでしょう。/
申込者：[あの]3人で。
担当者：はい。
申込者：[あの]子供が小学校6年生なんですけれども。/
担当者：3名様で6年生。/
申込者：はい。/
担当者：そうすると、ベッドがみつつあるお部屋をひと部屋、ご用意させていただきます。
申込者：はい。
担当者：よろしいわけですね。/

< TIS >

担当者：はい、では[えーと]人数は何名様っておっしゃってましたでしょう。
申込者：[あの]3人で{はい}[あの]子供が小学校6年生なんですけれども。
担当者：3名様で6年生。
申込者：はい。
担当者：そうすると、ベッドがみつつあるお部屋をひと部屋、ご用意させていただきます。
申込者：{はい}よろしいわけですね。

(1)

(注) いただければ、{はい}よろしいわけですね。

—
接続助詞

形態素情報データファイル（５）

－ 品質管理報告 －

1990年 8月

< 目次 >

| | |
|-------------------------|----|
| 1. 形態素情報の抜き取り検査 | |
| 1 - 1. 検査対象 | 1 |
| 1 - 2. 検査方法と検査結果 | 7 |
| 1 - 3. 不良に対する考察 | 8 |
| 2. 問題点解消の為の経緯 | 11 |
| 2 - 1. パンフレット名・ツアー名について | 12 |
| 2 - 2. 文節の切り方について | 14 |
| 2 - 3. お+動词语幹の扱いについて | 16 |
| 2 - 4. 複合格助詞について | 19 |
| 2 - 5. 補助動詞について | 20 |
| 3. 文節マニュアル第一版作成 | 22 |

1. 形態素情報の抜き取り検査

1-1. 検査対象

■ 対象期間 1990年2月1日 ~ 1990年7月31日

■ 会話分野と形態素数

| 会話分野 | 形態素数 | ファイル数 | 1ファイル数平均 (単位:形態素数) |
|----------|---------|-------|-----------------------|
| 電話会話 | 105,249 | 42 | 2,506 |
| キーボード会話 | 73,873 | 87 | 849 |
| SDB会話 | 41,501 | 75 | 553 |
| ※※※※※※※※ | 220,623 | 204 | ※※※※※※※※ |

(注) ファイル数は会話数を表す

■ 対象ファイル

< 電話会話 >

ファイル数 : 42ファイル

| 形態素 | : 画面数 | : | ファイル名 |
|--------|-------|---|---------------|
| 1681 | 205 | | ED_tel-412-02 |
| 2515 | 325 | | ED_tel-412-03 |
| 2772 | 283 | | ED_tel-412-04 |
| 1937 | 155 | | ED_tel-422-01 |
| 2879 | 238 | | ED_tel-422-02 |
| 2886 | 257 | | ED_tel-422-03 |
| 2575 | 232 | | ED_tel-422-04 |
| 4654 | 386 | | ED_tel-422-05 |
| 1536 | 130 | | ED_tel-422-06 |
| 1330 | 160 | | ED_tel-422-07 |
| 2119 | 246 | | ED_tel-422-08 |
| 2422 | 218 | | ED_tel-432-01 |
| 2466 | 233 | | ED_tel-432-02 |
| 1824 | 150 | | ED_tel-432-03 |
| 2535 | 234 | | ED_tel-432-04 |
| 1296 | 131 | | ED_tel-432-05 |
| 3144 | 330 | | ED_tel-442-01 |
| 1570 | 168 | | ED_tel-442-02 |
| 2474 | 253 | | ED_tel-442-03 |
| 1227 | 133 | | ED_tel-442-04 |
| 1642 | 156 | | ED_tel-442-05 |
| 2735 | 281 | | ED_tel-442-06 |
| 1477 | 143 | | ED_tel-452-01 |
| 2015 | 222 | | ED_tel-452-02 |
| 1915 | 256 | | ED_tel-452-03 |
| 2202 | 243 | | ED_tel-452-04 |
| 2701 | 343 | | ED_tel-452-05 |
| 2299 | 253 | | ED_tel-452-06 |
| 2500 | 298 | | ED_tel-452-07 |
| 2910 | 261 | | ED_tel-462-01 |
| 1638 | 164 | | ED_tel-462-02 |
| 4377 | 428 | | ED_tel-462-03 |
| 2204 | 207 | | ED_tel-472-01 |
| 2709 | 255 | | ED_tel-472-02 |
| 2370 | 241 | | ED_tel-472-03 |
| 2129 | 192 | | ED_tel-482-01 |
| 1558 | 157 | | ED_tel-482-02 |
| 2619 | 261 | | ED_tel-482-03 |
| 3767 | 306 | | ED_tel-482-04 |
| 2224 | 199 | | ED_tel-482-05 |
| 7109 | 667 | | ED_tel-482-06 |
| 4307 | 395 | | ED_tel-482-07 |
| 105249 | 10395 | | total |

■ 対象ファイル

< キーボード会話 >

ファイル数 : 87 ファイル

| 形態素 | : 画面数 | : ファイル名 |
|------|-------|---------------|
| 730 | 91 | ED_key-812-01 |
| 686 | 75 | ED_key-812-02 |
| 1310 | 133 | ED_key-812-03 |
| 1104 | 120 | ED_key-812-04 |
| 823 | 78 | ED_key-812-05 |
| 899 | 106 | ED_key-822-01 |
| 640 | 78 | ED_key-822-02 |
| 896 | 88 | ED_key-822-03 |
| 1121 | 114 | ED_key-822-04 |
| 606 | 67 | ED_key-822-05 |
| 894 | 99 | ED_key-832-01 |
| 956 | 128 | ED_key-832-02 |
| 858 | 117 | ED_key-832-03 |
| 974 | 125 | ED_key-832-04 |
| 962 | 116 | ED_key-832-05 |
| 943 | 101 | ED_key-842-01 |
| 581 | 84 | ED_key-842-02 |
| 876 | 101 | ED_key-842-03 |
| 877 | 122 | ED_key-842-04 |
| 870 | 107 | ED_key-842-05 |
| 680 | 93 | ED_key-852-01 |
| 757 | 96 | ED_key-852-02 |
| 789 | 107 | ED_key-852-03 |
| 772 | 95 | ED_key-852-04 |
| 785 | 96 | ED_key-852-05 |
| 1021 | 137 | ED_key-862-01 |
| 867 | 122 | ED_key-862-02 |
| 787 | 111 | ED_key-862-03 |
| 995 | 123 | ED_key-862-04 |
| 788 | 105 | ED_key-862-05 |
| 1000 | 140 | ED_key-872-01 |
| 1061 | 121 | ED_key-872-02 |
| 822 | 96 | ED_key-872-03 |
| 972 | 105 | ED_key-872-04 |
| 924 | 125 | ED_key-882-01 |
| 604 | 71 | ED_key-882-02 |
| 639 | 76 | ED_key-882-03 |
| 681 | 83 | ED_key-882-04 |
| 846 | 113 | ED_key-882-05 |
| 815 | 100 | ED_key-892-01 |
| 607 | 76 | ED_key-892-02 |
| 1120 | 119 | ED_key-892-03 |
| 1091 | 120 | ED_key-892-04 |
| 790 | 99 | ED_key-892-05 |
| 915 | 91 | ED_key-892-06 |
| 816 | 79 | ED_key-892-07 |
| 817 | 92 | ED_key-892-08 |
| 837 | 86 | ED_key-902-01 |
| 691 | 83 | ED_key-902-02 |
| 769 | 97 | ED_key-902-03 |
| 958 | 128 | ED_key-902-04 |
| 990 | 108 | ED_key-902-05 |
| 731 | 85 | ED_key-902-06 |
| 883 | 103 | ED_key-902-07 |

| | | |
|-------|------|---------------|
| 770 | 84 | ED_key-912-01 |
| 705 | 85 | ED_key-912-02 |
| 594 | 79 | ED_key-912-03 |
| 616 | 71 | ED_key-912-04 |
| 795 | 110 | ED_key-912-05 |
| 921 | 117 | ED_key-912-06 |
| 849 | 97 | ED_key-922-01 |
| 827 | 87 | ED_key-922-02 |
| 585 | 72 | ED_key-922-03 |
| 849 | 110 | ED_key-922-04 |
| 938 | 109 | ED_key-922-05 |
| 944 | 112 | ED_key-922-06 |
| 1074 | 121 | ED_key-922-07 |
| 838 | 90 | ED_key-922-08 |
| 878 | 96 | ED_key-932-01 |
| 993 | 90 | ED_key-932-02 |
| 708 | 96 | ED_key-932-03 |
| 814 | 85 | ED_key-932-04 |
| 761 | 88 | ED_key-932-05 |
| 909 | 111 | ED_key-932-06 |
| 836 | 103 | ED_key-932-07 |
| 953 | 116 | ED_key-932-08 |
| 962 | 95 | ED_key-942-01 |
| 923 | 92 | ED_key-942-02 |
| 878 | 115 | ED_key-942-03 |
| 995 | 125 | ED_key-942-04 |
| 795 | 108 | ED_key-942-05 |
| 791 | 88 | ED_key-952-01 |
| 731 | 85 | ED_key-952-02 |
| 829 | 91 | ED_key-952-03 |
| 682 | 76 | ED_key-952-04 |
| 1025 | 122 | ED_key-952-05 |
| 879 | 127 | ED_key-952-06 |
| 73873 | 8784 | total |

■ 対象ファイル

< SDB会話 >

ファイル数 : 75ファイル

| 形態素 | 画面数 | ファイル名 |
|------|-----|---------------|
| 124 | 21 | ED_sdb-042-01 |
| 140 | 19 | ED_sdb-042-02 |
| 467 | 37 | ED_sdb-052-01 |
| 553 | 39 | ED_sdb-052-02 |
| 759 | 64 | ED_sdb-052-03 |
| 721 | 53 | ED_sdb-052-04 |
| 673 | 63 | ED_sdb-052-05 |
| 105 | 19 | ED_sdb-062-01 |
| 86 | 12 | ED_sdb-062-02 |
| 136 | 23 | ED_sdb-062-03 |
| 879 | 201 | ED_sdb-072-01 |
| 1012 | 204 | ED_sdb-072-02 |
| 917 | 204 | ED_sdb-072-03 |
| 1143 | 216 | ED_sdb-072-04 |
| 2259 | 346 | ED_sdb-082-01 |
| 567 | 62 | ED_sdb-082-02 |
| 1001 | 119 | ED_sdb-082-03 |
| 583 | 83 | ED_sdb-082-04 |
| 584 | 84 | ED_sdb-082-05 |
| 322 | 51 | ED_sdb-092-01 |
| 348 | 46 | ED_sdb-092-02 |
| 260 | 45 | ED_sdb-092-03 |
| 355 | 45 | ED_sdb-092-04 |
| 326 | 40 | ED_sdb-092-05 |
| 284 | 38 | ED_sdb-092-06 |
| 414 | 49 | ED_sdb-092-07 |
| 341 | 51 | ED_sdb-092-08 |
| 358 | 49 | ED_sdb-092-09 |
| 289 | 40 | ED_sdb-092-10 |
| 308 | 36 | ED_sdb-092-11 |
| 287 | 38 | ED_sdb-092-12 |
| 236 | 32 | ED_sdb-092-13 |
| 341 | 45 | ED_sdb-092-14 |
| 336 | 47 | ED_sdb-092-15 |
| 305 | 38 | ED_sdb-092-16 |
| 284 | 39 | ED_sdb-092-17 |
| 314 | 42 | ED_sdb-092-18 |
| 295 | 40 | ED_sdb-092-19 |
| 288 | 31 | ED_sdb-092-20 |
| 414 | 78 | ED_sdb-102-01 |
| 460 | 60 | ED_sdb-102-02 |
| 278 | 57 | ED_sdb-102-03 |
| 330 | 53 | ED_sdb-102-04 |
| 293 | 36 | ED_sdb-102-05 |
| 285 | 57 | ED_sdb-102-06 |
| 379 | 59 | ED_sdb-102-08 |
| 243 | 42 | ED_sdb-102-09 |
| 447 | 80 | ED_sdb-102-10 |
| 322 | 49 | ED_sdb-102-11 |
| 673 | 129 | ED_sdb-102-12 |
| 710 | 135 | ED_sdb-102-13 |
| 759 | 130 | ED_sdb-102-14 |
| 373 | 56 | ED_sdb-102-15 |
| 688 | 126 | ED_sdb-102-16 |

| | | |
|-------|------|---------------|
| 726 | 143 | ED_sdb-102-17 |
| 896 | 170 | ED_sdb-102-18 |
| 699 | 124 | ED_sdb-102-19 |
| 605 | 109 | ED_sdb-102-20 |
| 576 | 103 | ED_sdb-102-21 |
| 820 | 145 | ED_sdb-102-22 |
| 631 | 101 | ED_sdb-102-23 |
| 255 | 35 | ED_sdb-102-24 |
| 685 | 125 | ED_sdb-102-25 |
| 732 | 151 | ED_sdb-102-26 |
| 671 | 144 | ED_sdb-102-27 |
| 825 | 171 | ED_sdb-102-28 |
| 781 | 155 | ED_sdb-102-29 |
| 732 | 147 | ED_sdb-102-30 |
| 304 | 46 | ED_sdb-112-01 |
| 512 | 69 | ED_sdb-112-02 |
| 1063 | 130 | ED_sdb-112-03 |
| 179 | 20 | ED_sdb-112-04 |
| 517 | 62 | ED_sdb-112-05 |
| 666 | 59 | ED_sdb-112-06 |
| 2972 | 573 | ED_sdb-122-01 |
| 41501 | 6640 | total |

1 - 2. 検査方法と検査結果

乱数を発生させランダムにファイルを抜き取る。
次に、抜き取ったファイルの形態素総数を見て、その総数より乱数を発生させてランダムに形態素データを1個取り出し、その形態素データの入った文をファイルに出力する。
このような処理を1サイクルとし、全体形態素数の約1%にあたる2,206形態素データを抽出した。

- ・ データ件数 220,623形態素
- ・ 抜き取り形態素数 2,206形態素
- ・ 不良個数 2形態素
- ・ 不良率 0.090%
- ・ 抽出日 1990年8月1日
- ・ 検査日 1990年8月2日
- ・ 検査者 立川

1-3. 不良に対する考察

(1) 内容

/data3/MORPH/product/sdb/ED_sdb-092-08

AAA:いらっしゃいませ。

??
? データ ? ひらがな ? 漢字 ? 品詞 ? 活種 ? 活形 ? 音便 ?
??
?(誤) ? ませ ? ませ ? ます ? 助動詞 ? 無 ? 未然形 ? 無 ?
??
?(正) ? ませ ? ませ ? ます ? 助動詞 ? 無 ? 命令形 ? 無 ?
??

(2) 不良原因

システムが未然形で出力してきたものを見落とした、人為的ミスである。

(3) 対策

書き換え規則システムで自動的に修正する。

(1) 内容

/data3/MORPH/product/tel/ED_tel-452-03

なにか行き違いがあったんだと思うんですけど、

??
? データ ? ひらがな ? 漢字 ? 品詞 ? 活種 ? 活形 ? 音便 ?
??
? (誤) ? 思う ? おもう ? 思う ? 本動詞 ? 無 ? 連体形 ? 無 ?
??
? (正) ? 思う ? おもう ? 思う ? 本動詞 ? 五段 ? 連体形 ? 無 ?
??

(2) 不良原因

システムは動詞の活用の種類を「無」にして出力してくることはない。
故に、これは人為的ミスであると思われる。このような活用の種類、
活用形のチェックとして、形態素抽出プログラム (extract) を
利用しているが見落としていた。 (1-3-1を参照)

(3) 対策

形態素修正時に見逃さないように重点的にチェックすると共に、
extractの出力は2人以上がチェックする事とする。

1-3-1. 形態素抽出プログラム (extract)

tel-452-03

[本動詞]

| | | | | |
|--------|--------|--------|-----|----|
| いう | 言う | いう | 五段 | 14 |
| なる | 成る | なる | 五段 | 7 |
| 待つ | 待つ | まつ | 五段 | 6 |
| わかる | 分る | わかる | 五段 | 6 |
| とる | 取る | とる | 五段 | 5 |
| 願う | 願う | ねがう | 五段 | 5 |
| 調べる | 調べる | しらべる | 下一段 | 3 |
| いただく | 戴く | いただく | 五段 | 3 |
| 越す | 越す | こす | 五段 | 3 |
| かしまる | 畏まる | かしまる | 五段 | 3 |
| 渡す | 渡す | わたす | 五段 | 3 |
| 入る | 入る | はいる | 五段 | 2 |
| 付く | 付く | つく | 五段 | 2 |
| 思う | 思う | おもう | 五段 | 2 |
| 持つ | 持つ | もつ | 五段 | 2 |
| 預かる | 預る | あずかる | 五段 | 2 |
| 伝わる | 伝わる | つたわる | 五段 | 1 |
| 聞き忘れる | 聞忘れる | ききわすれる | 下一段 | 1 |
| ある | 有る | ある | 五段 | 1 |
| 思う | 思う | | | 1 |
| ございます | 御座います | ございます | 特殊 | 1 |
| つかる | 漬る | つかる | 五段 | 1 |
| できる | 出来る | できる | 上一段 | 1 |
| やる | 遣る | やる | 五段 | 1 |
| 出る | 出る | でる | 変則 | 1 |
| 問い合わせる | 問合せる | といあわせる | 下一段 | 1 |
| 含む | 含む | ふくむ | 下一段 | 1 |
| かける | 掛ける | かける | 下一段 | 1 |
| 求める | 求める | もとめる | 下一段 | 1 |
| 伺う | 伺う | うかがう | 五段 | 1 |
| いらっしゃる | いらっしゃる | いらっしゃる | 特殊 | 1 |
| 伝える | 伝える | つたえる | 下一段 | 1 |
| 教える | 教える | おしえる | 下一段 | 1 |
| 申す | 申す | もうす | 五段 | 1 |
| かかる | 掛る | かかる | 五段 | 1 |
| 行ける | 行ける | ゆける | 下一段 | 1 |
| 行く | 行く | ゆく | 五段 | 1 |
| ゆう | 言う | ゆう | 五段 | 1 |
| 支払う | 支払う | しはらう | 五段 | 1 |
| つく | 付く | つく | 五段 | 1 |
| 為る | 為る | する | サ変 | 1 |
| おっしゃる | おっしゃる | おっしゃる | 特殊 | 1 |

形態素抽出プログラムにおいて正規表現が抜けているのはその形態素の活用の種類が抜けているため、正規表現が作れなかったからである。

2. 問題点解消のための経緯

弊社で形態素切り、文節切りを行なっている際に現在のマニュアルでは判断できない点が数件でてきた。この判断できない点に関してATR側（江原氏）と弊社で協議した経緯を次ページより記す。

【問題点】

パンフレット名、ツアー名の品詞付け、文節の切り方をどうすればよいか。

【弊社の質問】

現在、形態素作業マニュアル 1-2 (固有名詞) の表記に従い、各旅行会社が独自に

*1

創造したブランド(商品名)の「ルック」「パレット」「マッハ」は本の題名と同じレベルとして固有名詞としています。

パンフレット名、ツアー名は本で言えば章にあたるを考え、短単位分割で名詞として扱っています。いいのでしょうか？ また、文節切りはどのようなのでしょうか？

*1 本の名前は一つで固有名詞、章・社説名は普通名詞扱いで短単位分割

(ex-1) 当社の「ルック」という商品の中の「ヨーロッパベストルート12」という中の「フリープランロンドン・ローマ・パリ」というツアーなんです。

／ ルック ／
固名

／ ヨーロッパ ／ ベストルート ／ 12 ／
固名 普名 数詞

／ フリープラン ／ ロンドン ／ ・ ／ ローマ ／ ・ ／ パリ ／
普名 固名 記号 固名 記号 固名

(ex-2)パンフレット名

「リリカルヨーロッパ」というパンフレットなんです

(ex-3)ツアー名

「アルハムブラ宮殿とバルセロナ、マドリード(よう)8日間」というツアー

「ハイライト山水(し)詩情、桂林と北京、上海7」というツアーに参加

「倉敷・道後温泉」にしたいと思います

(ex-4) オブショナルツアー名

「古代遺跡の音と光のショー」というオブショナルツアーが夜に御座います

6時間コースの「市内洛北コース」と「嵐山コース」がよろしいのでは
ないでしょうか。

(ex-5) ブランド名ではないがツアー名であるかどうか判断しにくいもの

・ バンクーバーのスキーツアーをお願いしたいのですが

1. 「バンクーバーのスキーツアー」というツアー名なのか？

2. バンクーバーに行くためのスキーツアーなのか？

・ ハワイ 6 日間のツアーをお願いします

・ JAL で行く香港・シンガポール 6 日間を希望なんです

【ATRの回答】

形態素情報は今まで通り、短単位分割にする。しかし、コメントをそれぞれに入れる

ex. 「ヨーロッパベストルート12」という

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |
| ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |

文節情報はパンフレット・ツアー名及びそのすぐ後に続く付属語を含めて一文節とする

ex. / 「リリカルヨーロッパ」と / いう / パンフレットなんです

2 - 2. 文節の切り方について

1990.04.17 T I S 立川

【問題点】

「世界一」の文節の切り方はどのようにすればよいか

【弊社の質問】

< 「世界一」の文節の切り方について >

以前、文節についての問題点会議で自立語の後ろに数詞が来ると、文節は切るというルールが決まりました。そして、それに従って以下のように対応しました。

(ex-1) 約 5 0 0 名の参加者がありました。

約
(副詞)

5 0 0 名 の
(数詞) (接尾語) (格助詞)

(ex-2) 先生の発表は会議 3 日目の 2 9 日第一分科会の 1 0 時から 1 1 時まで

会議
(普名詞)

3 日 目 の
(数詞) (接尾語) (接尾語) (格助詞)

(ex-3) / 気温 / 1 0 度 /

(ex-4) / 標高 / 5 0 0 メートル /

(ex-5) / 摂氏 / 1 0 度 /

(ex-6) / 身長 / 1 5 0 C M /

しかし、世界一を「世界 / 一」とするには意味から考えて、おかしいと思うのですが、どうすれば良いですか？

T0230 : tel-422-05

ただ、[あの]シカゴのやはり有名なシカゴ美術館がございませし、あるいは世界一ののっぽビルのサースタワーというのがあるんですね。

【ATRの回答】

1990.4.18 庄山

問い合わせ：東洋情報システム

☆ 「世界一」を1つの文節として扱うかどうか。

返答：ATR自動翻訳電話研究所

短単位の切り方に基づき、世界／一と2つの文節とする。

【問題点】

「ごらん」の品詞付けをどうするか

【弊社の質問】

下記の「ごらん」の品詞付けについて

・テレビをつけてごらん。

・トランジスタラジオを持ってきてごらん。

・こっちへ来てみてごらん。

・きょうの新聞読んでごらん。

「ごらん」は新明解第三版には次のように書いてある

ごらん【御覧】

- [一] 他人が見ることの敬語。
「～ください・あっちを～ {ご覧なさいの、ぞんざいな言い方}」
- [二] {補助動詞的に} 「～して見よ」の丁寧語。
「～て御覧なさい {＝～てみなさい}」の、ぞんざいな表現。
「書いて～」

今回の「ごらん」は新明解の [二] になる。品詞付け及び文節は次の3通りが考えられるのですが、どうしたらいいのでしょうか？

1. ご / らん

接頭語 + サ変名詞 にする

- ・「ご覧なさい」の省略形と考えるとこの形にする

<問題点>

補助動詞的である（付属語）にもかかわらず自立語扱いされ、「ごらん」で一文節になってしまう

2. 補助動詞にする

- ・付属語である事を重視してこの形にする

<問題点>

「ごらんなさる」という補助動詞ができてしまう
活用語尾のない補助動詞になってしまう

??
 ? データ ? ひらがな ? 漢 字 ? 品 詞 ? 活 種 ? 活 形 ? 音 便 ?
 ???
 ? ごらん ? ごらん ? ごらん なさる ? 補 助 動 詞 ? 特 殊 ? 命 令 形 ? 無 ?
 ???

3. 普通名詞にする

自立語になってしまうので「ごらん」で一文節になってしまう

その外同じような例として

・買っておいで

・走ってお行き

【ATRの回答】

1990.6.12 江原氏

お+動詞語幹の扱いについて
 ?????????????????????

(1) 形態素の切り方

動詞語幹部分だけで(現代語として)意味を成すときは2形態素、
 そうでないときは1形態素

- ex. お行き 「行き」が意味を成すので2形態素
 ごらん 「らん」が意味を成さないので1形態素
 おいで 「いで」が(現代語として)意味を
 成さないので1形態素

(2) 品詞の決め方

本動詞のていねい形の場合は本動詞。補助動詞のていねい形の場合は
 補助動詞とする。

| | | | |
|-----------------|---|----------------|------|
| あっちを見よ ?? | → | あっちをごらん ??? | 本動詞 |
| テレビをつけてみよ ?? | → | つけてごらん ??? | 補助動詞 |
| 買ってこい ?? | → | 買っておいで ??? | 補助動詞 |
| 走って行け ?? | → | 走ってお行き **?? | 補助動詞 |

** 接頭語

「お行き」で1文節

| | | | |
|-------------|---|--------------|-----|
| 読んで返せ ?? | → | 読んでお返し ?? | 本動詞 |
|-------------|---|--------------|-----|

(3) 標準表現

「ごらんなさる」の省略形とはせず「ごらん」又は「御覧」（用字は新明解による）とする

(4) 活用型・活用形

「ごらん」「おいで」 → 活用型は特殊
活用形は語幹

(5) 文節の切り方

本動詞のときは新たな文節とする

「買っておいで。」 → 1文節

「買って／お返し。」 → 2文節

(6) その他1.

さらに補助動詞がつくこともある。

「こっち／へ／来／て／み／て／ごらん／なさい／。」
*** ***** *****

**** 補助動詞

(7) その他2.

本動詞と補助動詞の区別がむづかしい場合

「新聞をもっと目から離してごらん」

- ・「試しに離してみなさい」の意味のときは「見る」の本来の意味でなく、アスペクトであるから補助動詞
- ・「健康に悪いからもっと離して見なさい」の意味のときは「見る」の本来の意味であるから本動詞

【問題点】

複合格助詞であるかどうかの判断をどうすればよいか。

【回答】

< 「をもって」 について >

「をもって」 を 「を」 + 本動詞「持つ」 + 接続助詞「て」にするか、複合格助詞にするかの基準は次による。

「持つ」の部分の語義が「所有する」の場合は前者、「使用する」の場合は後者とする。

情熱をもって勉強する

| | | |
|-------------|-------|----|
| 情熱を所有して勉強する | | 前者 |
| 情熱を使用して勉強する | | 後者 |

書面をもって通知する

| | | |
|-------------|-------|----|
| 書面を所有して通知する | | 前者 |
| 書面を使用して通知する | | 後者 |

鉛筆をもって勉強する

| | | |
|-------------|-------|-----|
| 鉛筆を所有して勉強する | | 前者 |
| 鉛筆を使用して勉強する | | 前者* |

* 所有しかつ使用して

| | | |
|------------|-------|----------|
| 恨みをもって首を切る | | 前者 |
| 刀をもって首を切る | | 前者 or 後者 |

【問題点】

抽象性の高い動詞の品詞付けはどのようにすればよいか。

【弊社の質問】

抽象性の高い動詞については本動詞か補助動詞か文字の上からでは判断しにくいいため、基準をどうすればよいのでしょうか。

ex. 何かよく保険に入ってから飛行機に乗って^{~~~~~}いくって

聞きましたんで。

ex. ヨーロッパの方に持って行かれます外貨のご案内な^{~~~~~}

どがございます。

ex. 保険って入って行かれるのかしら。ヨーロッパって^{~~~~~}

言うと。

形態素作業マニュアル2-11では「本来の意味を失っているかどうか」で判断するようになっている。しかしこの判断では迷うものがある。下記にマニュアルに載っている例をあげる。

・ 生きていく (2-11)
^{~~~~}

補動

・ 彼は自転車に乗ってきた (2-11)
^{^^}

本動

・ 外へ出て行く (2-10)
^{~~~~}

補動

再び補助動詞について
????????????

原則

- ・動詞が本来の意味を持つときは本動詞
- ・本来の意味でなく付属的な意味のときは補助動詞

本来の意味とは

- ・辞書による

例として、「行く」「来る」のときは
空間的に移動するときに本来の中心的な意味

例. 「生きて行く」
生きてそして、どこかへ空間的に移動した場合は本動詞。
生きる事が継続した場合は補助動詞。

「外へ出て行く」
微妙な例。出てどこかへ行った場合は本動詞。出ることが中心で、
行き先はどうでも良いときは補助動詞。

問題の例

「飛行機に乗っていく」
多分補助動詞。行き先が問題ではないため。

「持って行かれます外貨」
補助動詞。行き先が問題でない。

「入って行かれるのかしら」
本動詞。保険に入ってそれから、空間的に移動する。尊敬の意味の
ときは補助動詞。「お風呂に入って行かれるのかしら、お客様」の
時は補助動詞。

3. 文節マニュアル第一版作成

文節について

1. 文節とは何か？

「文節とは何か」について『中学生の国文法の実力』によると以下のように説明されている。

??
? 文をくみだてていることばを、意味がわかる範囲内で、できるだけ ?
? こまかくくぎった一つ一つ。… (途中略) … 文を文節に分ける ?
? ときは、できるだけたくさん「ね」をいれてみるとよい。 ?
??

しかし、「ね」を挿入できる箇所というのは曖昧であり、判断に個人差が生じる可能性がある。そこで、ここでは形態素解析の結果を利用して、「自立語1個を中心に文節は構成される」と考えて作業を行う。

<基本ルール> 接頭辞*+自立語+接尾辞*+付属語*

ここで、
[1]「*」は0個以上複数の連続を許すことを表している。

[2]自立語とは以下の品詞を指す。
名詞 (固名詞、普名詞、サ名詞、形容名詞、数詞、代名詞)
本動詞
形容詞
副詞
連体詞
接続詞
感動詞
間投詞

[3]付属語とは以下の品詞を指す。
助動詞
助詞 (格助詞、係助詞、副助詞、並助詞、接助詞、終助詞、準助詞)
補助動詞
記号
その他

<基本ルールに対する例外>

- 1.自立語が0個という例外
(1)接頭語+接尾語で一つの自立語扱いとなるもの。
(例) 幾日
何人
- 2.自立語が2個以上という例外
(2)複数の単語 (特に普名詞) の連続が複合名詞として1自立語扱いとなるもの。
(例) 口座 番号
計算 機 工学

2. 文節プログラムの出力と修正

さて、文節プログラムでは、1. の要求をなるべく機械的に処理し、人間の負担を減らすべく作られており、ルールを満たすべく以下の3通りの組み合わせを捜し出して、文節単位とみなすのである。

1. 接頭辞*+自立語+接尾辞*+付属語*
 2. 接頭辞*+接尾辞*+付属語*
 3. 接頭辞*+普通名詞*+接尾辞*+付属語*
- ex:何人でも

しかし、これでは完全とはいえない。何故なら、

- (1) 複合名詞相当語は普通名詞連続のみではない。
→ 普通名詞連続以外の複合名詞相当語は、(複数の文節になっているものを)1つの文節にまとめなくてはならない。
- (2) 複合名詞相当語でない(偶然の)普通名詞連続がある。
→ このときには、1つの文節になっているものを2つ以上の文節に分割しなくてはならない。

そこで、人手による修正が必要となる。修正内容については次節(3.)に例をあげて詳しく説明する。

3. 文節プログラムの出力を修正するもの

3. 1 複数の文節になっているものを一つの文節にまとめるもの。

(1) 数詞を含む複合名詞相当語

典型的な例(「数詞+接尾語」+名詞など)

- お値段のほうはだいたい5万円前後になろうかと存じます。

5万 円 前後 に
数 詞 接尾辞 普名詞 格助詞

- はい、プログラムは厚さ7mm程度のパンフレット状のものです。

7 mm 程度 の
数 詞 記号 普名詞 格助詞

- 3部屋予約しており、それぞれ10小間分けて使用する予定です。

3 部屋
数 詞 普名詞

10 小 間 に
数 詞 接頭辞 普名詞 格助詞

- バランスよく配置したいと考えておりますので、1企業で1部屋
まるまる予約するということは控えていただきたい意向であります。

1 企業 で
数 詞 普名詞 格助詞

1 部屋
数 詞 普名詞

- 3ヶ月後のそちらからの次のサーキュラーが来るのを待つことにし
ます。

3 ヶ 月 後 の
数 詞 接尾辞 接尾辞 普名詞 格助詞

- 300人収容の松の間という部屋が最大だったのですが、まだ
450人収容の鳳凰の間という部屋があいておりますので、

300 人 収容 の
数 詞 接尾辞 普名詞 格助詞

450 人 収容 の
数 詞 接尾辞 普名詞 格助詞

- 四色刷り全面広告で1週間につき1350ドルです。
^^^^^^^^^^^^^^

四色刷り全面広告で
数詞 接尾辞 普名詞 普名詞 普名詞 格助詞

- その他3会場同時進行のセッションが行われます。
^^^^^^^^^^^^^^

3会場同時進行の
数詞 普名詞 普名詞 普名詞 格助詞

典型的な例（「接頭辞＋数詞」＋名詞など）

- 先生の発表は会議3日目の29日第一分科会の10時から11時までです。
^^^^^^^^^^

第 一 分 科 会 の
接頭辞 数 詞 普名詞 普名詞 格助詞

- 実は、会議の第一アナウンスメントを見ていたのですが、学会員としての
^^^^^^^^^^^^^^^^^^^^
登録料の割引に関する条項がないようなのですが。

第 一 アナウンスメント を
接頭辞 数 詞 普名詞 格助詞

- 第一セッションの知能ロボットに関しては、最先端技術に関する発表
^^^^^^^^^^^^^^
が研究者より行われます。

第 一 セッション の
接頭辞 数 詞 普名詞 格助詞

ちょっと紛らわしい例

- 第二、第三セッションのエキスパートシステム、画像理解については
^^^^^^^^^^^^^^^^^^
企業の開発担当者を中心に発表されます。

第 二 、
接頭辞 数 詞 記 号

第 三 セッション の
接頭辞 数 詞 普名詞 格助詞

（次の例のように「4、50」が「4」と「50」ではなく、全体で「40」
から「50」くらいの数を表していると考えられるときは1文節にする。「二、三」も同じ）

- 4、50名収容の部屋でしたら、いくつか用意できると思います。
^^^^^^^^^^^^^^

4 、 50 名 収容 の
数 詞 記 号 数 詞 接尾辞 普名詞 格助詞

●ホテルについて二、三お伺いしてよろしいでしょうか？
^^^^

二、三
数詞 記号 数詞

●お値段は10,000、2、3,000円ぐらいです。
^^

10,000、2、3,000 円
数詞 記号 数詞 記号 数詞 接尾辞

その他（数詞関連の修正しなくてもよい例）

- 約500名の参加者がありました。
^^^^^^^^

約
副 詞

500名 の
数 詞 接尾辞 格助詞

- 先生の発表は会議3日目の29日第一分科会の10時から11時までです。
^^^^^^^^

会議
普名詞

3 日 目 の
数 詞 接尾辞 接尾辞 格助詞

- 登録料は、おひとり¥50,000となっております。
^^^^^^^^

お ひと り
接頭辞 数 詞 接尾辞

¥ 50,000 と
記 号 数 詞 格助詞

- 最低一週間はかかってしまいますので、
^^^^^^^^

最低
普名詞

一 週 間 は
数 詞 接尾辞 接尾辞 係助詞

- ファックスはCT-355です。
^^^^^^^^

CT -
普名詞 記号

355 です。
数 詞 助動詞 記号

- 2x2ということで40,000円ですね。
^^^^^^

2 x
数 詞 記 号

2
数 詞

● 10 / 27
~~~~~

10 /  
数詞 記号

27  
数詞

数詞が後にでてくるものは、切ったままにしておく。

\* + 数詞 → \* / 数詞 (切る)

数詞 + \* → 数詞 \* (ひっつける)

- (例) ・ 気温 / 10 度                      ・ 3 分の / 1 程度の  
・ 標高 / 500 メートル                  ・ 一泊 / 2 食付  
・ 摂氏 / 10 度                          ・ 世界 / 一  
・ 身長 / 150 c m                      ・ 1 時間 / 30 分  
・ 7 時 / 30 分                          ・ 7 : / 00  
・ 会議 / 三日目                        ・ 一泊 / 5000 円  
・ 昭和 / 42 年  
・ 午後 / 7 時

(2) 数詞を含まない複合名詞相当語

典型的な例

- 予定参加者数は275人から325人です。  
^^^^^^^^

予定 参加 者 数 は  
普名詞 普名詞 接尾辞 普名詞 係助詞

- 皆さん一般参加者と同じ登録料となっております。  
^^^^^^^^

一般 参加 者 と  
普名詞 普名詞 接尾辞 格助詞

- それに出国申請書作成に際し、私がこの会議の正式な参加者であるとい  
^^^^^^^^^^^^^^^^  
う事務局発行の証明書が必要となります。

出国 申請 書 作成 に  
普名詞 普名詞 接尾辞 普名詞 格助詞

- 開催時期は会期中半ばを考えているのですが、それでいかがでしょうか。  
^^^^^^^^

会期 中 半ば を  
普名詞 接尾辞 普名詞 格助詞

- コピー機以外に何か準備するものではありませんでしょうか。  
^^^^^^^^

コピー 機 以外 に  
普名詞 接尾辞 普名詞 格助詞

- 海外のスピーカーと企画委員会スタッフの親睦をはかろうというこちらの  
^^^^^^^^^^^^^^^^  
希望はご存知ですね。

企画 委員 会 スタッフ の  
普名詞 普名詞 普名詞 普名詞 格助詞

- 会議当日お会いできることを楽しみにしております。  
^^^^^^^^

会議 当日  
普名詞 普名詞

- 締切日以降にキャンセルされた方への払い戻しはありません。

~~~~~

締切 日 以降 に
普名詞 普名詞 普名詞 格助詞

- その両学会のいずれかに会員となっておられる方しか後援発表はできません。

~~~~~

両 学会 の  
普名詞 普名詞 格助詞

- 一応今のところは各分科会のテーマしか決定しておりません。

~~~~~

各 分科 会 の
接頭辞 普名詞 普名詞 格助詞

- 選ばれた方には講師謝礼費、往復渡航運賃を支払います。

~~~~~

講師 謝礼 費  
普名詞 普名詞 接尾辞

- 会議開催も近づいてきた

~~~~~

会議 開催 も
普名詞 普名詞 係助詞

- おたくさまの講演名義使用の許可を頂く。

~~~~~

講演 名義 使用 の  
普名詞 普名詞 普名詞 格助詞

- 事務局発行の証明書が必要となります。

~~~~~

事務 局 発行 の
普名詞 普名詞 普名詞 格助詞

- スピーカーの論文要旨をまとめたものを発行するようになっております。

~~~~~

論文 要旨 を  
普名詞 普名詞 格助詞

- 材料機器など事務局で準備する

材料 機器 など  
普名詞 普名詞 副助詞

- 講演招待依頼の手紙には、

講演 招待 依頼 の  
普名詞 普名詞 普名詞 格助詞

- パソコンの家庭ユーザー向けオンラインサービスを提供しているのですが、

家庭 ユーザー 向け オンラインサービス を  
普名詞 普名詞 普名詞 普名詞 格助詞

- マサチューセッツ工科大学教授G. T. Williamsです。

マサチューセッツ工科大学 教授 G. T. Williams です。  
固名詞 普名詞 固名詞 助動詞

- 中世イギリスの風物を楽しむ

中世 イギリス の  
普名詞 固名詞 格助詞

- 中世イギリスから航海に出た人々は

中世  
普名詞  
イギリス から  
固名詞 格助詞

- 会議すべてまとめて皆さん聞いて頂きますので

会議 すべて  
普名詞 普名詞

- 白黒印刷全面広告で一週間につき700ドルです。

白黒 印刷 全面 広告 で  
普名詞 普名詞 普名詞 普名詞 格助詞

- 今回の旅行は買物優先なんで  
^^^^^^

買物      優先      な      んで  
普名詞   普名詞   助動詞   接助詞

- 同じ研究をしています  
^^^^^^

同じ  
形名詞

研究                  を  
普名詞              格助詞

- 日本は衣食住のバランスがとれている。  
^^^^^^

衣      食      住      の  
普名詞   普名詞   普名詞   格助詞

- 日英の同時通訳  
^^^^

日      英      の  
固名詞   固名詞   格助詞

(3) 複合動詞

本動詞の連続を複合名詞と同様、複合動詞として 1 自立語扱いとする。

- パルバーン先生の来日を当局としても待ち望んでおります。  
~~~~~

待ち 望ん で おり ます 。
本動詞 本動詞 接助詞 補動詞 助動詞 記号

- キャンセル料を払い戻すことはできません。
~~~~~

払い 戻す  
本動詞 本動詞

(4) プログラムの都合による特殊例

数字で始まる普通名詞

(下の例の「一級」また「一律」など、数字で始まる普通名詞はプログラムでは普通名詞連続からはずされている。これは月日について／一月／一日／のように文節を切りたいため「数字で始まる普通名詞は連続していても切る」という規則を導入したためである。よって月日以外の数字で始まる普通名詞連続は手修正でまとめる必要がある。)

- 今日は5月25日です。

5月  
普名詞

25日 　　です 　　。  
普名詞 　助動詞 　記号

- これは一級品のビデオなんですよ。

一級 　品 　　の  
普名詞 普名詞 格助詞

- ですから、一律料金ということでご理解いただければと存じます。

一律 　料金 　と  
普名詞 普名詞 格助詞

(5) 特殊な副詞

名詞的に前後の名詞と直接つながっている例

- なるべく九月一杯にお送りしたいと思います。

九月 一杯 に  
普名詞 副詞 格助詞

- もちろん、料金は、その場合、別途料金になります。

別途 料金 に  
副詞 普名詞 格助詞

- \*別途追加料金をうけたまわります。

別途  
副詞

追加 料金を  
普名詞 普名詞 格助詞

? 別途料金を / になります

?

? 別途 / 資料を / お送りします /

文のつながりを見て「別途、資料を送ります」という意味なら切る。  
別途資料というものを送るといふのなら一つにする。

- 参加申込書は、しばらく前に送ったのですが、

しばらく 前 に  
副詞 普名詞 格助詞

文のつながりを見て、その副詞が前後の名詞にかかるかどうかを見る。

- それでは 1時ちょうど発の便をお願いします。

1 時 ちょうど 発  
数詞 接尾辞 副詞 普名詞



- 1時間ちょっとで行けます

^^^^^^^^^^

|    |     |      |     |
|----|-----|------|-----|
| 1  | 時間  | ちょっと | で   |
| 数詞 | 接尾辞 | 副詞   | 格助詞 |

- もう一度お願い致します

^^^^^^^^^^

|    |    |     |
|----|----|-----|
| もう | 一  | 度   |
| 副詞 | 数詞 | 接尾辞 |

- ぜひ一度お願いします

^^^^^^^^^^

|    |
|----|
| ぜひ |
| 副詞 |

|    |     |
|----|-----|
| 一  | 度   |
| 数詞 | 接尾辞 |

- 彼が来てもう1週間になります

^^^^^^^^^^

|         |
|---------|
| ?<br>もう |
| ?<br>副詞 |

|         |          |          |          |
|---------|----------|----------|----------|
| ?<br>一  | ?<br>週   | ?<br>間   | ?<br>に   |
| ?<br>数詞 | ?<br>接尾辞 | ?<br>接尾辞 | ?<br>格助詞 |

- もう1週間滞在したい

^^^^^^^^^^

|    |    |     |     |
|----|----|-----|-----|
| もう | 1  | 週   | 間   |
| 副詞 | 数詞 | 接尾辞 | 接尾辞 |

- それはつい最近のことです

^^^^^^^^^^

|    |     |     |
|----|-----|-----|
| つい | 最近  | の   |
| 副詞 | 普名詞 | 格助詞 |

(6) 形容詞と名詞

- 移動の距離も長いものですから、なるべく荷物を軽くしたいと思ひまして、  
~~~~~

長い
形容詞

もの です から 、
普名詞 助動詞 接助詞 記号

- どうかご懸念なく身軽なご旅行をお楽しみ下さいませ。
~~~~~

ご            懸念  
接頭辞      普名詞

なく  
形容詞

??????  
? 注意 ?  
??????

代名詞と連体詞は少々気持ち悪くてもまとめず、切ったままにしておく。

- また、私自身でコピーを<sup>~~~~~</sup>持っておく必要がありますので、こちらにも一部  
お送り願えませんでしょうか。

私  
代名詞

自身 で  
普名詞 格助詞

- はい、当事務局と<sup>~~~~~</sup>しましてできる限りお手伝いしたいと思っております。

当  
連体詞

事務局 と  
普名詞 普名詞 格助詞

- それ以外の日時は少しむずかしいかと思えます。  
~~~~~

それ
代名詞

以外 の
普名詞 格助詞

4. 1つの文節になっているものを2つ以上に分割するもの。

(1) 複合名詞相当語でない(偶然の)普通名詞連続

典型的な例 その1

- アレンジを、今事務局のもとで活動されているATRにお願いしたいと
思います。

今
普名詞

事務 局 の
普名詞 普名詞 格助詞

- 先生には当初招待講演の方引き受けて頂いておりましたが残念ながら
出席して頂けないと連絡をうけました。

当初
普名詞

招待 講演 の
普名詞 普名詞 格助詞

- 他に問題がなければ、再度最終フロアプランを検討致しまして先生の方に
お送りさせていただきます。

再度
普名詞

最終 フロアプラン を
普名詞 普名詞 格助詞

- はい、実は先日書類をお送りしたところ、宛先不明で戻って参りました。

先日
普名詞

書類 を
普名詞 格助詞

- 今回ビデオということですので、どれぐらいの規模の機器が必要で、
搬入及び経費等詳しくわからないと決定できないとのことでした。

今回
普名詞

ビデオ と
普名詞 格助詞

- ということは、協賛団体の会員であっても今回特典はないわけですね。
^^^^^^

今回
普名詞

特典 は
普名詞 係助詞 ("今回"など時をあらわすものは切る)

典型的な例 その2

- お忙しいところ本当にありがとうございました。
^^^^^^

お 忙しい
接頭辞 形容詞

ところ
普名詞

本当 に
普名詞 格助詞

- ファーストアナウンスメントについているリプライカードに必要事項を記入の上事務局までお送り下さい。
^^^^^^

記入 の
普名詞 格助詞

上
普名詞

事務 局 まで
普名詞 普名詞 格助詞

- 宣伝のためホテル内に特別の展示を行なってもよいでしょうか。
^^^^^^

宣伝 の
普名詞 格助詞

ため
普名詞

ホテル 内 に
普名詞 接尾辞 格助詞

- 私としましてはできる限り多くの方にこの会議のことを知って頂きたいので、
どのようなところに配布すれば一番効率が良いのでしょうか。

できる
本動詞

限り
普名詞

多く の
普名詞 格助詞

- 会場の確保、その他備品、運営などについて詳しくお知らせ頂きたいのですが。

その
連体詞

他
普名詞

備品 、
普名詞 記号

運営 など について
普名詞 副助詞 格助詞

- アメリカ企業、ヨーロッパの企業、アジアの企業、日本の企業バランスよく
配置したいと考えておりますので、

アジア の
固名詞 格助詞

企業 、
普名詞 記号

日本 の
固名詞 格助詞

企業
普名詞

バランス
普名詞

よく
形容詞

- お互い話も合うのではないかと思い、同会議開催の折り、非公式な懇親会を
開くことになりました。

お 互い
接頭辞 普名詞

話 も
普名詞 係助詞

合 う の で は な い か と
本動詞 準助詞 助動詞 係助詞 助動詞 終助詞 格助詞

◎ その他

- どういう機能があるか、という点です。
^^^^^^^^^^

ある か 、 と
本動詞 終助詞 記号 格助詞

- はい、でしたら、それで お願いします。
^^^^^^^^^^

はい 、 でしたら
感動詞 記号 助動詞 接助詞

- スペルは T A N A K A です。
^^^^^^^^^^

T A
普名詞 普名詞

N A
普名詞 普名詞

K A です 。
普名詞 普名詞 助動詞 記号

*カタカナ (ティーエーエヌエーケーエー) の場合も同じ

- スペルは T O M です。(外国人の場合)

^^^^^^^^^^
(ティーオーエム)

T O M です 。
普名詞 普名詞 普名詞 助動詞 記号

- 電話番号は・・・

* 0 3 - / 6 7 2 - / 5 7 3 9 です /

* 0 3 、 / 6 7 2 の / 5 7 3 9 です /

* 0 3 / (6 7 2) / 5 7 3 9 です /

* 電話番号 / 0 3 の / 6 7 2 / 5 7 3 9 /

* 普通貯金 / 8 5 7 の / 8 4 3 5 /

◎会話文を、文節の基本ルールにすべてあてはめる事には無理がある。
よって その場に応じての特別のルールをもうけ、対応する。

旅行会話におけるパンフレット名、ツアー名の扱いについては文節は一つにする

- 「ヨーロッパハイライト・ルート10」というツアーに

「ヨーロッパハイライト・ルート10」
記号 固名詞 普名詞 記号 普名詞 数詞 記号

- 「サッポロフリープランデラックス5DAY」に予約してた、

「サッポロフリープランデラックス5DAY」
記号 固名詞 普名詞 普名詞 数詞 接尾辞 記号

- 「沖縄本島・4日間[えー]5日間」というのが

「沖縄本島・4日間[えー]5日間」
記号 固名詞 普名詞 記号 数詞 接尾 接尾 記号 間投詞 記号 数詞 接尾 接尾 記号

- 会話文における ”相槌” の扱いについて

相槌を表す { } は前の文節に付ける

*あるかも／しれないんで、{ [あー] そうですか } /

*500,000ぐらいという予算で／言ってみましたけど、{はい} / どうでしょうかね、

- 間投詞、言い間違いの扱いについて

* [あー]、あのですね

[あー] / 、
記号 間投詞 記号 記号

* (え、おい) おいしい食べ物は・・・

/ (え 、 おい) /
記号 その他 記号 その他 記号

※その他は付属語扱い

形態素情報データファイル(6)

－ 品質管理報告 －

1991年 2月

< 目次 >

1. 形態素情報の抜き取り検査

| | | |
|------------------|-------|---|
| 1 - 1. 検査対象 | | 1 |
| 1 - 2. 検査方法と検査結果 | | 8 |
| 1 - 3. 不良に対する考察 | | 9 |

1. 形態素情報の抜き取り検査

1-1. 検査対象

■ 対象期間 1990年8月1日 ~ 1991年2月12日

■ 会話分野と形態素数

| 会話分野 | 形態素数 | ファイル数 | 1ファイル数平均 (単位:形態素数) |
|---------|---------|-------|-----------------------|
| キーボード会話 | 184,376 | 265 | 696 |
| SDB会話 | 15,108 | 11 | 1,373 |
| 合計 | 199,484 | 276 | ※※※※※※※※ |

(注) ファイル数は会話数を表す

■ 対象ファイル

< キーボード会話 >

ファイル数 : 265ファイル

| 形態素 | 画面数 | ファイル名 |
|------|-----|----------------------------------|
| 1012 | 134 | /morph/product/key/ED_key-a12-01 |
| 842 | 86 | /morph/product/key/ED_key-a12-02 |
| 1118 | 120 | /morph/product/key/ED_key-a12-03 |
| 1499 | 184 | /morph/product/key/ED_key-a12-04 |
| 1120 | 107 | /morph/product/key/ED_key-a12-06 |
| 802 | 94 | /morph/product/key/ED_key-a12-07 |
| 1261 | 131 | /morph/product/key/ED_key-a22-01 |
| 921 | 102 | /morph/product/key/ED_key-a22-02 |
| 815 | 92 | /morph/product/key/ED_key-a22-03 |
| 864 | 110 | /morph/product/key/ED_key-a22-04 |
| 893 | 110 | /morph/product/key/ED_key-a22-05 |
| 966 | 114 | /morph/product/key/ED_key-a32-01 |
| 1071 | 137 | /morph/product/key/ED_key-a32-02 |
| 1046 | 142 | /morph/product/key/ED_key-a32-03 |
| 737 | 108 | /morph/product/key/ED_key-a32-04 |
| 919 | 121 | /morph/product/key/ED_key-a32-05 |
| 894 | 120 | /morph/product/key/ED_key-a32-06 |
| 1338 | 174 | /morph/product/key/ED_key-a42-01 |
| 852 | 106 | /morph/product/key/ED_key-a42-02 |
| 1372 | 164 | /morph/product/key/ED_key-a42-03 |
| 571 | 64 | /morph/product/key/ED_key-a42-04 |
| 987 | 134 | /morph/product/key/ED_key-a52-01 |
| 1082 | 130 | /morph/product/key/ED_key-a52-02 |
| 772 | 104 | /morph/product/key/ED_key-a52-03 |
| 776 | 99 | /morph/product/key/ED_key-a52-04 |
| 1050 | 135 | /morph/product/key/ED_key-a52-05 |
| 934 | 118 | /morph/product/key/ED_key-a52-06 |
| 686 | 108 | /morph/product/key/ED_key-a62-01 |
| 973 | 137 | /morph/product/key/ED_key-a62-02 |
| 698 | 96 | /morph/product/key/ED_key-a62-03 |
| 765 | 108 | /morph/product/key/ED_key-a62-04 |
| 546 | 72 | /morph/product/key/ED_key-a62-05 |
| 686 | 106 | /morph/product/key/ED_key-a72-01 |
| 585 | 81 | /morph/product/key/ED_key-a72-02 |
| 728 | 109 | /morph/product/key/ED_key-a72-03 |
| 583 | 85 | /morph/product/key/ED_key-a72-04 |
| 835 | 115 | /morph/product/key/ED_key-a82-01 |
| 1178 | 150 | /morph/product/key/ED_key-a82-02 |
| 742 | 97 | /morph/product/key/ED_key-a82-03 |
| 791 | 100 | /morph/product/key/ED_key-a82-04 |
| 846 | 111 | /morph/product/key/ED_key-a82-05 |
| 1148 | 149 | /morph/product/key/ED_key-a82-06 |
| 1291 | 171 | /morph/product/key/ED_key-a82-07 |
| 766 | 122 | /morph/product/key/ED_key-a92-01 |
| 971 | 104 | /morph/product/key/ED_key-a92-02 |
| 892 | 128 | /morph/product/key/ED_key-a92-03 |
| 913 | 118 | /morph/product/key/ED_key-a92-04 |
| 800 | 111 | /morph/product/key/ED_key-a92-05 |
| 877 | 115 | /morph/product/key/ED_key-b12-01 |
| 864 | 115 | /morph/product/key/ED_key-b12-02 |
| 944 | 121 | /morph/product/key/ED_key-b12-03 |
| 545 | 83 | /morph/product/key/ED_key-b12-04 |
| 978 | 140 | /morph/product/key/ED_key-b12-05 |
| 1015 | 129 | /morph/product/key/ED_key-b12-06 |

| | | |
|------|-----|----------------------------------|
| 510 | 82 | /morph/product/key/ED_key-b22-01 |
| 946 | 138 | /morph/product/key/ED_key-b22-02 |
| 1160 | 157 | /morph/product/key/ED_key-b22-03 |
| 792 | 108 | /morph/product/key/ED_key-b22-04 |
| 974 | 122 | /morph/product/key/ED_key-b22-05 |
| 781 | 109 | /morph/product/key/ED_key-b32-01 |
| 1138 | 141 | /morph/product/key/ED_key-b32-02 |
| 1230 | 180 | /morph/product/key/ED_key-b32-03 |
| 851 | 122 | /morph/product/key/ED_key-b32-04 |
| 537 | 64 | /morph/product/key/ED_key-b32-05 |
| 623 | 85 | /morph/product/key/ED_key-b32-06 |
| 731 | 88 | /morph/product/key/ED_key-b42-01 |
| 774 | 105 | /morph/product/key/ED_key-b42-02 |
| 636 | 90 | /morph/product/key/ED_key-b42-03 |
| 885 | 111 | /morph/product/key/ED_key-b42-04 |
| 1095 | 109 | /morph/product/key/ED_key-b52-01 |
| 1451 | 169 | /morph/product/key/ED_key-b52-02 |
| 1107 | 126 | /morph/product/key/ED_key-b52-03 |
| 540 | 68 | /morph/product/key/ED_key-b52-04 |
| 1337 | 120 | /morph/product/key/ED_key-b62-01 |
| 896 | 103 | /morph/product/key/ED_key-b62-02 |
| 1355 | 158 | /morph/product/key/ED_key-b62-03 |
| 1228 | 119 | /morph/product/key/ED_key-b62-04 |
| 1309 | 133 | /morph/product/key/ED_key-b72-01 |
| 844 | 83 | /morph/product/key/ED_key-b72-02 |
| 992 | 100 | /morph/product/key/ED_key-b72-03 |
| 669 | 91 | /morph/product/key/ED_key-b82-01 |
| 607 | 72 | /morph/product/key/ED_key-b82-02 |
| 750 | 97 | /morph/product/key/ED_key-b82-03 |
| 797 | 85 | /morph/product/key/ED_key-b82-04 |
| 963 | 115 | /morph/product/key/ED_key-b92-01 |
| 1231 | 141 | /morph/product/key/ED_key-b92-02 |
| 1184 | 148 | /morph/product/key/ED_key-b92-03 |
| 710 | 91 | /morph/product/key/ED_key-b92-04 |
| 812 | 90 | /morph/product/key/ED_key-c12-01 |
| 823 | 83 | /morph/product/key/ED_key-c12-02 |
| 974 | 124 | /morph/product/key/ED_key-c12-03 |
| 597 | 68 | /morph/product/key/ED_key-c22-01 |
| 1041 | 107 | /morph/product/key/ED_key-c22-02 |
| 1011 | 112 | /morph/product/key/ED_key-c22-03 |
| 754 | 92 | /morph/product/key/ED_key-c22-04 |
| 830 | 91 | /morph/product/key/ED_key-c22-05 |
| 648 | 77 | /morph/product/key/ED_key-c32-01 |
| 934 | 103 | /morph/product/key/ED_key-c32-02 |
| 517 | 70 | /morph/product/key/ED_key-c32-03 |
| 683 | 80 | /morph/product/key/ED_key-c32-04 |
| 400 | 57 | /morph/product/key/ED_key-c32-05 |
| 964 | 107 | /morph/product/key/ED_key-c32-06 |
| 919 | 92 | /morph/product/key/ED_key-c42-01 |
| 681 | 80 | /morph/product/key/ED_key-c42-02 |
| 1106 | 126 | /morph/product/key/ED_key-c42-03 |
| 907 | 99 | /morph/product/key/ED_key-c42-04 |
| 929 | 117 | /morph/product/key/ED_key-c52-01 |
| 1128 | 111 | /morph/product/key/ED_key-c52-02 |
| 1028 | 124 | /morph/product/key/ED_key-c52-03 |
| 851 | 101 | /morph/product/key/ED_key-c52-04 |
| 695 | 69 | /morph/product/key/ED_key-c62-01 |
| 835 | 95 | /morph/product/key/ED_key-c62-02 |
| 749 | 82 | /morph/product/key/ED_key-c62-03 |
| 666 | 69 | /morph/product/key/ED_key-c62-04 |
| 761 | 92 | /morph/product/key/ED_key-c72-01 |
| 599 | 73 | /morph/product/key/ED_key-c72-02 |
| 562 | 69 | /morph/product/key/ED_key-c72-03 |
| 521 | 73 | /morph/product/key/ED_key-c72-04 |
| 506 | 54 | /morph/product/key/ED_key-c72-05 |
| 338 | 44 | /morph/product/key/ED_key-c82-01 |

| | | |
|------|-----|----------------------------------|
| 952 | 108 | /morph/product/key/ED_key-c82-02 |
| 777 | 85 | /morph/product/key/ED_key-c82-03 |
| 291 | 35 | /morph/product/key/ED_key-c82-04 |
| 628 | 86 | /morph/product/key/ED_key-c82-05 |
| 415 | 71 | /morph/product/key/ED_key-c82-06 |
| 859 | 94 | /morph/product/key/ED_key-c92-01 |
| 525 | 67 | /morph/product/key/ED_key-c92-02 |
| 695 | 104 | /morph/product/key/ED_key-c92-03 |
| 916 | 132 | /morph/product/key/ED_key-c92-04 |
| 648 | 84 | /morph/product/key/ED_key-c92-05 |
| 849 | 111 | /morph/product/key/ED_key-c92-06 |
| 418 | 67 | /morph/product/key/ED_key-d12-01 |
| 1090 | 141 | /morph/product/key/ED_key-d12-02 |
| 908 | 111 | /morph/product/key/ED_key-d12-03 |
| 778 | 97 | /morph/product/key/ED_key-d12-04 |
| 301 | 52 | /morph/product/key/ED_key-d12-05 |
| 362 | 55 | /morph/product/key/ED_key-d12-06 |
| 355 | 55 | /morph/product/key/ED_key-d12-07 |
| 667 | 97 | /morph/product/key/ED_key-d22-01 |
| 526 | 69 | /morph/product/key/ED_key-d22-02 |
| 376 | 55 | /morph/product/key/ED_key-d22-03 |
| 414 | 63 | /morph/product/key/ED_key-d22-04 |
| 656 | 95 | /morph/product/key/ED_key-d22-05 |
| 531 | 71 | /morph/product/key/ED_key-d22-06 |
| 548 | 72 | /morph/product/key/ED_key-d22-07 |
| 622 | 89 | /morph/product/key/ED_key-d22-08 |
| 682 | 96 | /morph/product/key/ED_key-d32-01 |
| 492 | 62 | /morph/product/key/ED_key-d32-02 |
| 439 | 70 | /morph/product/key/ED_key-d32-03 |
| 609 | 91 | /morph/product/key/ED_key-d32-04 |
| 747 | 85 | /morph/product/key/ED_key-d32-05 |
| 503 | 73 | /morph/product/key/ED_key-d32-06 |
| 543 | 87 | /morph/product/key/ED_key-d32-07 |
| 807 | 111 | /morph/product/key/ED_key-d42-01 |
| 559 | 82 | /morph/product/key/ED_key-d42-02 |
| 523 | 75 | /morph/product/key/ED_key-d42-03 |
| 578 | 75 | /morph/product/key/ED_key-d42-04 |
| 477 | 59 | /morph/product/key/ED_key-d42-05 |
| 701 | 90 | /morph/product/key/ED_key-d42-06 |
| 723 | 99 | /morph/product/key/ED_key-d42-07 |
| 457 | 61 | /morph/product/key/ED_key-d42-08 |
| 574 | 72 | /morph/product/key/ED_key-d52-01 |
| 709 | 80 | /morph/product/key/ED_key-d52-02 |
| 1071 | 134 | /morph/product/key/ED_key-d52-03 |
| 394 | 62 | /morph/product/key/ED_key-d52-04 |
| 345 | 53 | /morph/product/key/ED_key-d52-05 |
| 334 | 48 | /morph/product/key/ED_key-d52-06 |
| 676 | 91 | /morph/product/key/ED_key-d52-07 |
| 551 | 91 | /morph/product/key/ED_key-d52-08 |
| 601 | 85 | /morph/product/key/ED_key-d62-01 |
| 505 | 74 | /morph/product/key/ED_key-d62-02 |
| 502 | 69 | /morph/product/key/ED_key-d62-03 |
| 501 | 65 | /morph/product/key/ED_key-d62-04 |
| 482 | 73 | /morph/product/key/ED_key-d62-05 |
| 499 | 72 | /morph/product/key/ED_key-d62-06 |
| 359 | 50 | /morph/product/key/ED_key-d62-07 |
| 385 | 45 | /morph/product/key/ED_key-d72-01 |
| 400 | 59 | /morph/product/key/ED_key-d72-02 |
| 508 | 67 | /morph/product/key/ED_key-d72-03 |
| 278 | 44 | /morph/product/key/ED_key-d72-04 |
| 486 | 68 | /morph/product/key/ED_key-d72-05 |
| 531 | 71 | /morph/product/key/ED_key-d72-06 |
| 574 | 87 | /morph/product/key/ED_key-d72-07 |
| 594 | 93 | /morph/product/key/ED_key-d82-01 |
| 370 | 57 | /morph/product/key/ED_key-d82-02 |
| 764 | 90 | /morph/product/key/ED_key-d82-03 |

| | | |
|-----|-----|----------------------------------|
| 438 | 66 | /morph/product/key/ED_key-d82-04 |
| 355 | 51 | /morph/product/key/ED_key-d82-05 |
| 706 | 88 | /morph/product/key/ED_key-d82-06 |
| 596 | 70 | /morph/product/key/ED_key-d82-07 |
| 580 | 78 | /morph/product/key/ED_key-d82-08 |
| 505 | 74 | /morph/product/key/ED_key-d92-01 |
| 670 | 78 | /morph/product/key/ED_key-d92-02 |
| 610 | 93 | /morph/product/key/ED_key-d92-03 |
| 787 | 94 | /morph/product/key/ED_key-d92-04 |
| 667 | 91 | /morph/product/key/ED_key-d92-05 |
| 584 | 95 | /morph/product/key/ED_key-d92-06 |
| 559 | 85 | /morph/product/key/ED_key-d92-07 |
| 805 | 103 | /morph/product/key/ED_key-d92-08 |
| 515 | 74 | /morph/product/key/ED_key-d92-09 |
| 721 | 114 | /morph/product/key/ED_key-e12-01 |
| 693 | 93 | /morph/product/key/ED_key-e12-02 |
| 562 | 77 | /morph/product/key/ED_key-e12-03 |
| 538 | 72 | /morph/product/key/ED_key-e12-04 |
| 483 | 67 | /morph/product/key/ED_key-e12-05 |
| 759 | 98 | /morph/product/key/ED_key-e12-06 |
| 814 | 98 | /morph/product/key/ED_key-e12-07 |
| 705 | 86 | /morph/product/key/ED_key-e12-08 |
| 619 | 96 | /morph/product/key/ED_key-e22-01 |
| 437 | 67 | /morph/product/key/ED_key-e22-02 |
| 642 | 90 | /morph/product/key/ED_key-e22-03 |
| 470 | 67 | /morph/product/key/ED_key-e22-04 |
| 660 | 95 | /morph/product/key/ED_key-e22-05 |
| 595 | 83 | /morph/product/key/ED_key-e22-06 |
| 852 | 116 | /morph/product/key/ED_key-e22-07 |
| 500 | 72 | /morph/product/key/ED_key-e22-08 |
| 830 | 106 | /morph/product/key/ED_key-e32-01 |
| 648 | 99 | /morph/product/key/ED_key-e32-02 |
| 597 | 101 | /morph/product/key/ED_key-e32-03 |
| 721 | 86 | /morph/product/key/ED_key-e32-04 |
| 522 | 78 | /morph/product/key/ED_key-e32-05 |
| 754 | 102 | /morph/product/key/ED_key-e32-06 |
| 634 | 98 | /morph/product/key/ED_key-e32-07 |
| 753 | 94 | /morph/product/key/ED_key-e32-08 |
| 760 | 88 | /morph/product/key/ED_key-e42-01 |
| 769 | 91 | /morph/product/key/ED_key-e42-02 |
| 743 | 122 | /morph/product/key/ED_key-e42-03 |
| 960 | 127 | /morph/product/key/ED_key-e42-04 |
| 328 | 44 | /morph/product/key/ED_key-e42-05 |
| 548 | 56 | /morph/product/key/ED_key-e42-06 |
| 572 | 81 | /morph/product/key/ED_key-e42-07 |
| 567 | 78 | /morph/product/key/ED_key-e42-08 |
| 216 | 31 | /morph/product/key/ED_key-e42-09 |
| 227 | 33 | /morph/product/key/ED_key-e52-01 |
| 650 | 65 | /morph/product/key/ED_key-e52-02 |
| 331 | 46 | /morph/product/key/ED_key-e52-03 |
| 500 | 70 | /morph/product/key/ED_key-e52-04 |
| 252 | 34 | /morph/product/key/ED_key-e62-01 |
| 324 | 35 | /morph/product/key/ED_key-e62-02 |
| 310 | 43 | /morph/product/key/ED_key-e62-03 |
| 300 | 38 | /morph/product/key/ED_key-e62-04 |
| 584 | 81 | /morph/product/key/ED_key-e62-05 |
| 684 | 97 | /morph/product/key/ED_key-e72-01 |
| 211 | 34 | /morph/product/key/ED_key-e72-02 |
| 583 | 75 | /morph/product/key/ED_key-e72-03 |
| 272 | 35 | /morph/product/key/ED_key-e82-01 |
| 312 | 42 | /morph/product/key/ED_key-e82-02 |
| 375 | 68 | /morph/product/key/ED_key-e82-03 |
| 428 | 52 | /morph/product/key/ED_key-e82-04 |
| 301 | 46 | /morph/product/key/ED_key-e82-05 |
| 291 | 41 | /morph/product/key/ED_key-e82-06 |
| 403 | 61 | /morph/product/key/ED_key-e92-01 |

| | | |
|--------|-------|----------------------------------|
| 454 | 66 | /morph/product/key/ED_key-e92-02 |
| 455 | 61 | /morph/product/key/ED_key-e92-03 |
| 490 | 68 | /morph/product/key/ED_key-e92-04 |
| 230 | 38 | /morph/product/key/ED_key-e92-05 |
| 206 | 33 | /morph/product/key/ED_key-f12-01 |
| 473 | 70 | /morph/product/key/ED_key-f12-02 |
| 357 | 42 | /morph/product/key/ED_key-f12-03 |
| 311 | 39 | /morph/product/key/ED_key-f12-04 |
| 196 | 29 | /morph/product/key/ED_key-f12-05 |
| 298 | 39 | /morph/product/key/ED_key-f12-06 |
| 244 | 34 | /morph/product/key/ED_key-f22-01 |
| 229 | 33 | /morph/product/key/ED_key-f22-02 |
| 244 | 35 | /morph/product/key/ED_key-f22-03 |
| 184376 | 23760 | total |

■ 対象ファイル

< SDB会話 >

ファイル数 : 11ファイル

| 形態素 | 画面数 | ファイル名 |
|-------|------|----------------------------------|
| 3204 | 505 | /morph/product/sdb/ED_sdb-122-02 |
| 536 | 50 | /morph/product/sdb/ED_sdb-132-01 |
| 1106 | 129 | /morph/product/sdb/ED_sdb-132-02 |
| 266 | 41 | /morph/product/sdb/ED_sdb-132-03 |
| 1279 | 121 | /morph/product/sdb/ED_sdb-142-01 |
| 699 | 70 | /morph/product/sdb/ED_sdb-142-02 |
| 3587 | 483 | /morph/product/sdb/ED_sdb-152-01 |
| 552 | 80 | /morph/product/sdb/ED_sdb-162-01 |
| 277 | 41 | /morph/product/sdb/ED_sdb-162-02 |
| 520 | 88 | /morph/product/sdb/ED_sdb-162-03 |
| 3082 | 416 | /morph/product/sdb/ED_sdb-172-01 |
| 15108 | 2024 | total |

1 - 2. 検査方法と検査結果

乱数を発生させランダムにファイルを抜き取る。
次に、抜き取ったファイルの形態素総数を見て、その総数より乱数を発生させてランダムに形態素データを1個取り出し、その形態素データの入った文をファイルに出力する。
このような処理を1サイクルとし、全体形態素数の約1%にあたる1,994形態素データを抽出した。

- ・ データ件数 199,484形態素
- ・ 抜き取り形態素数 1,994形態素
- ・ 不良個数 2形態素
- ・ 不良率 0.1%
- ・ 抽出日 1991年 2月8日
- ・ 検査日 1991年 2月12日
- ・ 検査者 立川

(別紙)

動詞・連用形と名詞形が同形の形態の品詞付けについて

抜き取り検査の結果過去のデータと現在のデータとに相違が出ました。

過去の品詞付け方法 (1990.09.28納品時まで)

????????

今までは第一に新明解を準拠とする考え方で、名詞としてエントリーされていれば名詞として品詞付けする

..... 申し込み書
***** **
名詞 接尾辞

また、"接頭辞「お」の後には動詞・連用形がくるのが原則"などがマニュアルに記述されていれば採用するとう考え方で

..... お申し込み書
***** **
動詞 接尾辞

としていました。

新しい品詞付け方法 (1990.10.29納品時から)

????????

作業を進めて行く上でこれらの矛盾が大きくなってきたため、マニュアルの<1-15> <2-7>を採用し**動詞は接尾辞に接続し、(接尾辞の前は動詞)名詞は名詞(名詞の前は名詞)に接続するという考え方に今回から一本化しました。**

eg. お申し込み方は明日ご説明します
***** **
動詞+接尾辞

eg. 申し込み方は明日ご説明します
***** **
動詞+接尾辞

eg. お申し込み金は今日中に戴けますか
***** **
普通名詞+普通名詞

< マニュアル1-15 (転成名詞) >

c f. 使い方
*** ** 動詞・連用形+接尾辞

< マニュアル2-7 (動詞) >

お送り先
*** ** 名詞+名詞
(動詞なら連体形にならんとあかん)

申し込み用紙
***** **** 名詞+名詞
(上に同じ)

※ 本来、転成名詞に接尾辞が付く場合はその転成名詞に格助詞「が」を付けて意味が通じる場合は名詞とするところだが、過去のデータと大きく異なる為、上記の「新しい品詞付け方法」をもとに作業をする

(添付1)

名詞としていたが現在は動詞とするデータ

FILE:ED key-a92-05-01

それとですね、他にも、**申込書**の方でいくつか不明点がある

普名詞 --- --- 申込
接尾辞 --- --- 書

FILE:ED key-a92-03-02

それから、お手数なんですけど、**申込書**を書き直ししていただかないと

普名詞 --- --- 申込
接尾辞 --- --- 書

FILE:ED key-a92-03-02

行けるかどうかわからないから、こちらに**申込書**を送っていただけないかしら

普名詞 --- --- 申込
接尾辞 --- --- 書

FILE:ED key-a92-01-02

まあ、全部わかったと思いますので**申込書**と振込用紙を送ってください。

普名詞 --- --- 申込
接尾辞 --- --- 書

FILE:ED key-a92-01-01

担当者：それでは、**申込書**と、申込金を納めていただくと、

普名詞 --- --- 申込
接尾辞 --- --- 書

FILE:ED key-a92-01-01

ますので、こちらにいらっしゃれますか。それとも**申込書**と振込用紙をお送りしましょうか

普名詞 --- --- 申込
接尾辞 --- --- 書

FILE:ED key-a92-01-01

。通訳者：おうかがいする時間がないもので、**申込書**を送ってもらえますか。あとは電話

普名詞 --- --- 申込
接尾辞 --- --- 書

FILE:ED key-a82-02-02

今日打ち合せした内容を加えた**見積書**を送っていただけますか。

普名詞 --- --- 見積
接尾辞 --- --- 書

FILE:ED key-a82-02-02

至急**見積書**とコースの日程表を送らせていただきます

普名詞 --- --- 見積
接尾辞 --- --- 書

FILE:ED key-a62-02-02

ただ最初の予定を変えるとしましたら**申込書**を書き直さなくてはなりませんので、

普名詞 --- --- 申込
接尾辞 --- --- 書

FILE:ED key-a52-05-02

料金の方が50,000円になります。**申込書**と申込金を3日間以内におさめ

普名詞 --- --- 申込
接尾辞 --- --- 書

FILE:ED key-a52-05-02

申込書の方はご自宅にお送りしますので。

普名詞 --- --- 申込
接尾辞 --- --- 書

FILE:ED key-a52-01-01

担当者：先日、お送りいただいた英文**申込書**のことでちょっと確認したいことがある

普名詞 --- --- 申込

接尾辞 --- --- 書

FILE:ED key-a52-01-01

担当者：今、お手元の方に**申込書**の控えがございますでしょうか。

普名詞 --- --- 申込

接尾辞 --- --- 書

FILE:ED key-a42-02-01

担当者：では、**申込書**を書いていただきたいのですが

普名詞 --- --- 申込

接尾辞 --- --- 書

FILE:ED key-a22-01-02

予約をお取りします。**申込書**に記入していただきたいので

普名詞 --- --- 申込

接尾辞 --- --- 書

FILE:ED key-a12-03-02

ナイロビヒルトンにシングルで3泊予約いたします。**申込書**をお送りしますので、

普名詞 --- --- 申込

接尾辞 --- --- 書

FILE:ED key-b12-03-02

担当者：はい。それから**申込書**を書き直していただかなければなりませんので

普名詞 --- --- 申込

接尾辞 --- --- 書

(注)

本資料はATR自動翻訳電話研究所が(株)東洋情報システムに発注した言語データベースについての品質管理報告書をまとめたものである。