

TR-IT-0290

言語フィルターの改善

酒井 慎介 藤澤 謙 Nick Campbell

1999.2.26

概要

TDMT の辞書の読み、活用語尾の区切り方、アクセント情報を kan2rom と同じ正しいものに修正し、それに伴い、gfilter を正しい区切りに対応でき、TDMT での形態素結果を kan2rom の形態素解析結果と同様に変換できるように修正した。

また、幾つかのバグにより、実装できていなかった付属語検出エラー時のリカバリ機能のバグを修正し、実装した。

©ATR Interpreting Telecommunications
Research Laboratory.

©ATR 音声翻訳通信研究所

もくじ

1	はじめに	1
2	TDMT 日本語辞書の修正	2
2.1	TDMT 日本語辞書	2
2.2	辞書の形式	3
2.3	修正内容	4
2.4	差分辞書 jg.dic の修正	4
2.4.1	差分辞書 Jg.dic	4
2.4.2	修正結果	5
2.5	活用語の正しい区切りへの修正	8
2.5.1	活用語尾	8
2.5.2	修正結果	9
2.6	考察、まとめ	11
3	言語フィルター gfilter の改良	12
3.1	gfilter	12
3.2	これまでの gfilter の問題点	15
3.3	gfilter の修正方法	17
3.4	不具合の要因	18
3.5	対策	23
3.6	結果	24
4	修正した gfilter の評価	25

4.1	gfilter の評価	25
4.2	考察、まとめ	27
5	gfilter 出力での BI の変更	28
5.1	概要	28
5.2	修正ファイル	28
5.3	修正内容	28
6	gfilter の無声化処理の C 版への切り替え	29
6.1	概要	29
6.2	修正ファイル	29
6.3	修正内容	29
7	付属語検出エラーのリカバリ機能の実装	30
7.1	リカバリ機能	30
7.2	リカバリ機能のバグ	30
7.3	サ変名詞の付属語がエラーになる要因	32
7.4	結果	34
7.5	考察、まとめ	35
8	修正したファイルのリスト	37
9	改良後の gfilter の出力結果	38

第 1 章

はじめに

TDMT の形態素解析結果を利用するために、TDMT 辞書に読み、アクセント情報を与えているが、それは kan2rom と動詞、形容詞などの活用語尾の区切り方が異なっている。これは、gfilter を通して mor2bou にて韻律処理を行う際に、正しい付属語、アクセント処理ができないため、出力として、アクセントがおかしくなってしまう。

そこで、TDMT の辞書の読み、区切り方、アクセント情報を kan2rom と同じ正しいものに修正する。それに伴い、gfilter も正しい区切りに対応できるように修正が必要となる。TDMT で形態素結果を kan2rom の出力と同様に変換できるよう、gfilter を修正する。

また、mor2bou に、文法的におかしな文章が入力されると、付属語検出エラーとなる場合があるが、これまでは、エラーとなった部分以降の読みが出力されない問題があった。それを回避するために、付属語検出エラー時に、表層の読みをそのまま出力する付属語リカバリ機能が用意されていたが、幾つかのバグのため、gfilter にリンクするときのみ用いられていた。

そこで、それらのバグとなる要因を解消し、付属語リカバリ機能を実装した。

尚、このレポートの内容の中でいくつかのファイルを修正し、commit しているが、私の修正前の gfilter 及び morph は

```
$ cvs checkout -r stable_990114 gfilter (or morph)
```

で CVS から checkout できる。

第 2 章

TDMT 日本語辞書の修正

2.1 TDMT 日本語辞書

TDMT の辞書には形態素解析辞書と生成辞書の二種類ある。形態素解析辞書のエントリで生成辞書に含まれていないものは、差分辞書に記述されている。

本研究で修正対象となるのは生成辞書と差分辞書である。

TDMT 生成辞書

- dohsi.dic : TDMT 側の品詞で本動詞のもの
- keiyou.dic : TDMT 側の品詞で形容詞、形容名詞のもの
- fukushi.dic : TDMT 側の品詞で副詞、感動詞、接続詞のもの
- jodoshi.dic : TDMT 側の品詞で助動詞のもの
- meishi.dic : TDMT 側の品詞で固有名詞、普通名詞、日時のもの
- rentaishi.dic: TDMT 側の品詞で連体詞のもの
- setsuji.dic : TDMT 側の品詞で接頭辞、接尾辞のもの
- joshi.dic : TDMT 側の品詞で助詞一般のもの

TDMT 差分辞書

- jg.dic : 形態素解析辞書に登録されていて、生成辞書に登録されていないもの

2.2 辞書の形式

辞書の形式は TDMT 品詞、表層、読み、アクセント情報からなる。

辞書のフォーマット

TDMT 品詞 "表層" 読み "表層の読み" アクセント "アクセント情報"

例

普通名詞 "大仏" 読み "だいぶつ" アクセント "1000001000000000"

副詞 "残念ながら" 読み "ざんねんながら" アクセント "5000001000000051"

また、活用語尾をもつ形容詞、動詞、アクセントが2つ以上の名詞などは、読みをわけて、以下のような形式にする。

読みをわける場合

TDMT 品詞 "表層" 読み "表層の読み" アクセント "アクセント情報" 読み "表層の読み" アクセント "アクセント情報" . . .

例

本動詞 "送る" 読み "おく" アクセント "2750001000000001" 読み "る" アクセント "0"

形容詞 "乾いてない" 読み "かわ" アクセント "20000010000000021" 読み "いて" アクセント "0" 読み "な" アクセント "30000010000000011" 読み "い" アクセント "0"

普通名詞 "京都南インターチェンジ" 読み "きょうと" アクセント "1130001000000010" 読み "みなみいんたーちえんじ" アクセント "1000001000000080"

2.3 修正内容

辞書の修正内容は以下の2つに大別される。

- 差分辞書 `kg.dic` に関しては、読み、アクセント情報は `KDD kan2rom` の出力をそのまま使用したものとなっている。そのため、固有名詞や、区切り方が適切でない語などに関しては、読みや形態素の区切りを誤ったり、アクセントの位置が正しくなかったり、未知語として処理されているものなどが多く存在する。

従って、そのようなエントリは、表層に対して正しい読み、区切り、アクセント情報を与える。

- `gfilter` では付属語検出エラーを減らすために、名詞以外の付属語処理を行っていない。そのため、`TDMT` 辞書には動詞、形容詞など活用語をもつものに対して、形態素区切りを `kan2rom` とは異なる与え方をしている。このままでは、`mor2bou` でのアクセント処理にて、正しい活用語の処理を行うことができず、不十分なアクセント処理となってしまう。

従って、後述する `gfilter` の改良に伴い、活用語の与え方を `kan2rom` と同じ正しいものに修正する。全ての辞書が修正対象となる。

2.4 差分辞書 `kg.dic` の修正

2.4.1 差分辞書 `Jg.dic`

差分辞書 `kg.dic` では `kan2rom` の出力がそのまま使用されており、いろいろな誤りが存在している。

- 読みが誤っているもの 表層の読みが正しくないため、アクセント情報が正しいものではない。
- 区切りが誤っているもの 表層の区切り方が正しくないため、正しいアクセント情報ではないものがある
- アクセント位置が誤っているもの アクセント位置だけが正しくないものは、`kan2rom` の辞書 `base.dic` に誤りがある場合に起きる。まれに存在する。
- 未知語のもの 未知語として処理され、アクセント情報は正しくない。

また、誤りではないが、複合名詞が複数の名詞に区切られていたら、アクセントが一つに収まる範囲で、一つに結合し、長い単語で登録した。これは、`mor2bou` の複合処理におけるアクセント位置計算での、誤りの危険性を避けることができるからである。

2.4.2 修正結果

以下に修正した箇所の一例を挙げる。

- 読みが誤っているもの

読みが誤っているもの

(修正前) サ変名詞 "直行" 読み "なおゆき" アクセント "11200010000000021"
(修正後) サ変名詞 "直行" 読み "ちょっこう" アクセント "10024210000000000"

(修正前) 形容名詞 "高尚" 読み "たかひさ" アクセント "11200010000000021"
(修正後) 形容名詞 "高尚" 読み "こうしょう" アクセント "40000010000000000"

(修正前) 普通名詞 "頭皮" 読み "あたまかわ" アクセント "a0000010000000031"
(修正後) 普通名詞 "頭皮" 読み "とうひ" アクセント "10000010000000010"

(修正前) 副詞 "殆んど" 読み "たいんど" アクセント "a0000010000000010"
(修正後) 副詞 "殆んど" 読み "ほとんど" アクセント "50000010000000020"

- 区切りが誤っているもの

区切りが誤っているもの

(修正前) 普通名詞 "アンチェインドメロディー" 読み "あん" アクセント "11200010000000010" 読み "ちえ" アクセント "a0000010000000010" 読み "いんど" アクセント "11300010000000010" 読み "めろでいー" アクセント "10000010000000010"
(修正後) 普通名詞 "アンチェインドメロディー" 読み "あんちえいんどめろでいー" アクセント "11000010000000070"

(修正前) 感動詞 "ごちそうさま" 読み "ごち" アクセント "a0000010000000010" 読み "そう" アクセント "50000010000000011" 読み "さま" アクセント "100000100010001"
(修正後) 感動詞 "ごちそうさま" 読み "ごちそうさま" アクセント "710000100010001"

● アクセント位置が正しくないもの

———— アクセント位置が正しくないもの ————

(修正前) 普通名詞 "てんどん" 読み "てんどん" アクセント "1000001000000031"

(修正後) 普通名詞 "てんどん" 読み "てんどん" アクセント "1000001000000001"

(修正前) サ変名詞 "案内" 読み "あんない" アクセント "1002411000000040"

(修正後) サ変名詞 "案内" 読み "あんない" アクセント "1002411000000030"

● 未知語のもの

———— 未知語のもの ————

(修正前) 普通名詞 "にぎり寿司" 読み "にぎり" アクセント "a000001000000010"
読み "すし" アクセント "1000101000110020"

(修正後) 普通名詞 "にぎり寿司" 読み "にぎりずし" アクセント "10000010000000
30"

(修正前) 普通名詞 "にんじん" 読み "にんじん" アクセント "a000001000000010"

(修正後) 普通名詞 "にんじん" 読み "にんじん" アクセント "1000001000000000"

(修正前) 普通名詞 "彩子" 読み "あやこ" アクセント "a000001000000011"

(修正後) 普通名詞 "彩子" 読み "あやこ" アクセント "1120001000000011"

● 結合し、長い単語にしたもの

結合し、長い単語にしたもの

(修正前) 普通名詞 "クリアランスセール" 読み "くりあ" アクセント "4002431000000020" 読み "らんす" アクセント "a000001000000010" 読み "せーる" アクセント "1000001000000010"

(修正後) 普通名詞 "クリアランスセール" 読み "くりあらんすせーる" アクセント "10000010000000070"

(修正前) 普通名詞 "近鉄バッファローズ" 読み "きんてつ" アクセント "1000001000000000" 読み "ぼっふあ" アクセント "a000001000000010" 読み "ろーず" アクセント "1000001000000010"

(修正後) 普通名詞 "近鉄バッファローズ" 読み "きんてつばふあろーず" アクセント "11400010000000050"

(修正前) 普通名詞 "京都国際観光ホテル" 読み "きょうと" アクセント "1130001000000010" 読み "こくさい" アクセント "1000201000000000" 読み "かんこう" アクセント "1002401000000000" 読み "ほてる" アクセント "1000001000000010"

(修正後) 普通名詞 "京都国際観光ホテル" 読み "きょうと" アクセント "1130001000000010" 読み "こくさいかんこうほてる" アクセント "10000010000000090"

2.5 活用語の正しい区切りへの修正

2.5.1 活用語尾

活用語尾とは、動詞、形容詞などで見られる活用の種類によって、変化する部分である。例えば、動詞「動く」であれば、「うご」までは活用に関係なく、変化しないが、それ以降は「動かない」、「動け」などのように変化する。従って、「動く」の場合は「く」が活用語尾となる。

前述したように、これまでの gfilter では、活用語尾を区切るようにすると、付属語検出エラーが発生する危険性があることから、TDMT の辞書には、

```
本動詞 "動く" 読み "うごく" アクセント "2050001000000021"
```

のように登録されている。

しかし、このままでは mor2bou で正しいアクセント処理が行えない。従って、下のように、活用語尾を区切り、kan2rom の形態素解析と同様の出力になるようにする。

```
本動詞 "動く" 読み "うご" アクセント "2050001000000021" 読み "く" アクセント "0"
```

また、難解なエントリに対しては、/DB/PI/SunOS/bin/awss (以下、awss) を利用した。awss は kan2rom の形態素解析結果を出力するものである。その結果と同じになるように区切り、読み、アクセント情報を与えた。

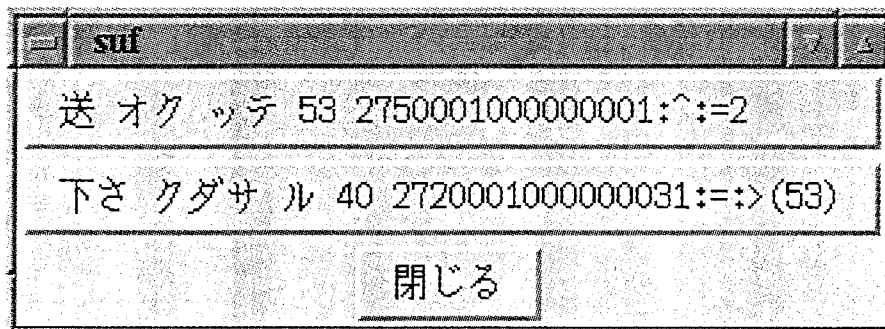


図 2.1: awss の出力結果 入力「送って下さる」

2.5.2 修正結果

全ての辞書を調べた結果、活用語尾の区切りの修正が必要となった辞書は、以下のものである。

doshi.dic keiyou.dic jg.dic jodoshi.dic

以下には、各辞書において修正を行った箇所の一例を示す。

● doshi.dic の修正例

— doshi.dic —

(修正前) 本動詞 "見せる" 読み "みせる" アクセント "2900001000000011"

(修正後) 本動詞 "見せる" 読み "みせ" アクセント "2900001000000011" 読み "る" アクセント "0"

(修正前) 本動詞 "保険を掛ける" 読み "ほけん" アクセント "1000001000000000" 読み "を" アクセント "0" 読み "かける" アクセント "2900511000100021"

(修正後) 本動詞 "保険を掛ける" 読み "ほけん" アクセント "1000001000000000" 読み "を" アクセント "0" 読み "かけ" アクセント "2900511000100021" 読み "る" アクセント "0"

(修正前) 本動詞 "送って下さる" 読み "おくってくださる" アクセント "2700001000000001"

(修正後) 本動詞 "送って下さる" 読み "おく" アクセント "2750001000000001" 読み "って" アクセント "0" 読み "くださ" アクセント "2720001000000031" 読み "る" アクセント "0"

(修正前) 本動詞 "薬味のある" 読み "やくみのある" アクセント "27100020000000051"

(修正後) 本動詞 "薬味のある" 読み "やくみ" アクセント "1000001000000000" 読み "の" アクセント "0" 読み "あ" アクセント "27100010000000011" 読み "る" アクセント "0"

• keiyou.dic の修正例

keiyou.dic

(修正前) 形容詞 "美味しい" 読み "おいしい" アクセント "3000001000000001"

(修正後) 形容詞 "美味しい" 読み "おいし" アクセント "3000001000000031" 読み "い" アクセント "0"

(修正前) 形容詞 "分かりやすい" 読み "わかりやすい" アクセント "30000010000000051"

(修正後) 形容詞 "分かりやすい" 読み "わかりやす" アクセント "3000001000000001" 読み "い" アクセント "0"

• jg.dic の修正例

jg.dic

(修正前) 形容詞 "なまぬるい" 読み "なまぬるい" アクセント "30000010000000041"

(修正後) 形容詞 "なまぬるい" 読み "なまぬる" アクセント "30000010000000041" 読み "い" アクセント "0"

(修正前) 形容詞 "頑固だ" 読み "がんこだ" アクセント "40000010000000010"

(修正後) 形容詞 "頑固だ" 読み "がんこ" アクセント "40000010000000010" 読み "だ" アクセント "0"

(修正前) 副詞 "いっぺんに" 読み "いっぺんに" アクセント "51000010000000001"

(修正後) 副詞 "いっぺんに" 読み "いっぺん" アクセント "51000010000000001" 読み "に" アクセント "0"

(修正前) 本動詞 "返してもらおう" 読み "かえしてもらおう" アクセント "2250511000100011"

(修正後) 本動詞 "返してもらおう" 読み "かえ" アクセント "2250511000100011" 読み "してもらおう" アクセント "0"

(修正前) 連体詞 "何人かの" 読み "なんにんかの" アクセント "10000010000000011"

(修正後) 連体詞 "何人かの" 読み "なんにんか" アクセント "10000010000000011" 読み "の" アクセント "0"

• jodoshi.dic の修正例

— jodoshi.dic —

(修正前) 助動詞 "願う" 読み "ねがう" アクセント "2800001000000021" 読み "う" アクセント "0"

(修正後) 助動詞 "願う" 読み "ねが" アクセント "2800001000000021" 読み "う" アクセント "0"

2.6 考察、まとめ

ここで示した辞書以外でも、gfilter のテストなどで発見された辞書の誤りはその都度修正している。修正した辞書については巻末の修正ファイル一覧に記す。また、辞書の修正には、細心の注意を払って行ったつもりであるが、莫大な量であること、また、全て手作業で行ったということから、多少の誤りはあるかと思われる。よって、これはおかしいというものがあれば、修正して頂きたいと思う。

第 3 章

言語フィルター - gfilter の改良

3.1 gfilter

gfilter は、TDMT で形態素解析された結果を、入力とし、CHATR の Uttrance Phoneme 形式のアクセント付きローマ字表記で出力するものである。

アクセント処理は、mor2bou で行うため、gfilter の基本的な役割は、TDMT の 出力結果を、mor2bou への入力データに変換することである。

— TDMT の出力 —

```
JF できれば | 接続副詞 | できれば | 7000001000000021
、 | 特殊 | |
ホテルに | 格助詞 | ほてる に | 1000001000000010 0
もう | 接頭辞 | もう | 50000010000000011
二 | 数詞 | 2 | 12000010000000011
泊を | 格助詞 | はく を | 9100001000002020 0
滞在する | 本動詞 | たいざい する | 1002401000000000 0
ので | 接続助詞 | ので | 0
、 | 特殊 | |
部屋を | 格助詞 | へや を | 1000101000100021 0
変え | 本動詞 | かえ | 2900511000100001
たい | 助動詞 | たい | 0
です | 助動詞 | です | 0
。 | 特殊 | |
```

mor2bou への入力

できれば デキレバ 0 0 7000001000000021
、 0 0 0 d000001000000001
ホテルに ホテル ニ 0 1000001000000010
もう モウ 0 0 50000010000000011
二 2 0 0 12000010000000011
泊を ハク ラ 0 91000010000002020
滞在するので タイザイ スルノデ 0 1002401000000000
、 0 0 0 d0000010000000001
部屋を ヘヤ ラ 0 1000101000100021
変えたいです カエ タイデス 0 2900511000100001
。 0 0 0 0000001000000001

mor2bou の出力

deki'reba 3 ho'teruni 1 mo'0 1 nihaku'o 1 taizai surunode 2 heya'o kaetai'desU 5

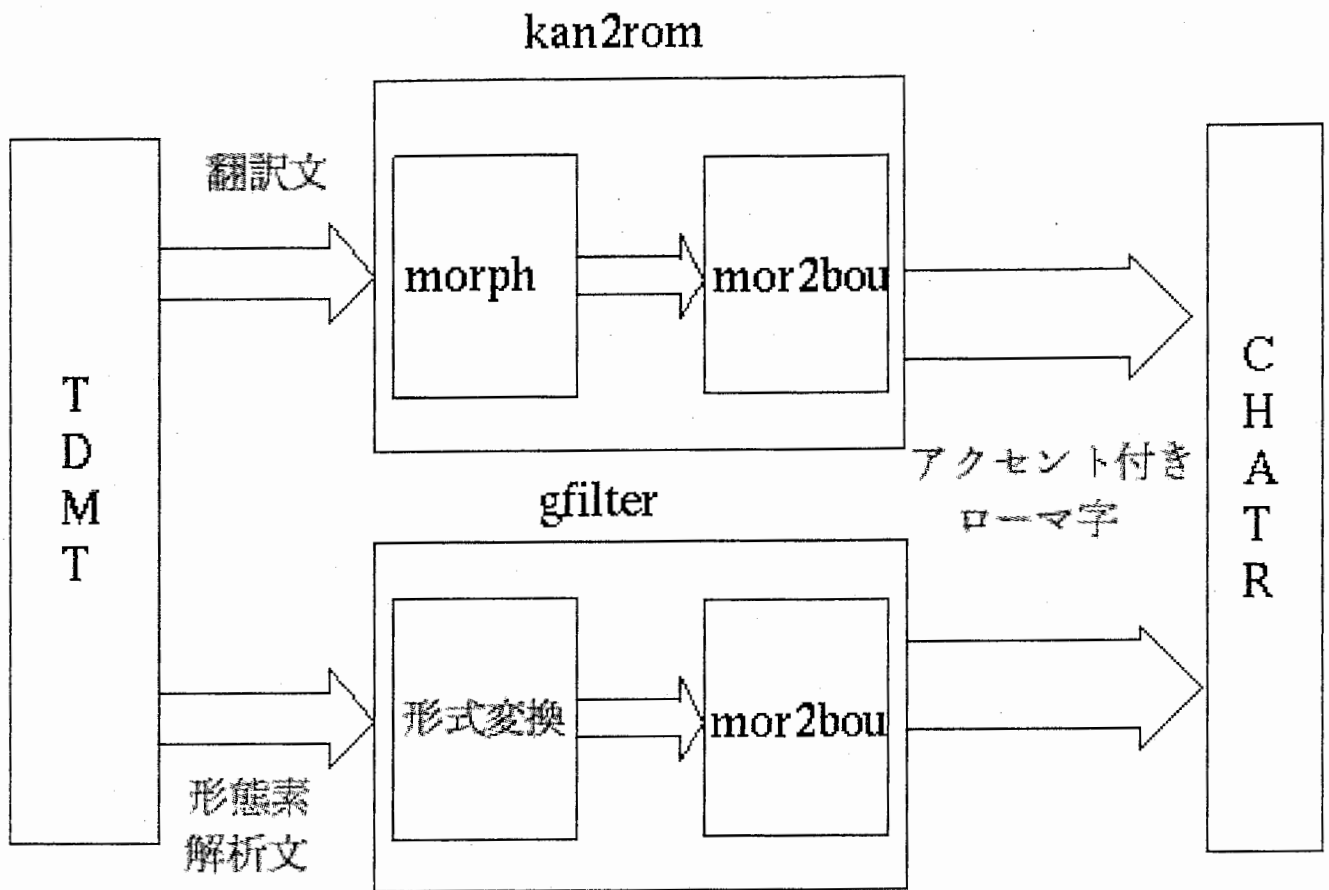


図 3.1: gfilter と kan2rom の比較

3.2 これまでの gfilter の問題点

これまでの gfilter は、活用語尾を区切っていなかったため、正しい付属語処理、アクセント処理が行われず、アクセント位置や語句の区切りが正しくない。

- これまでの gfilter の出力

TDMT 出力 (辞書修正前)

```
行け | 本動詞 | いけ | 2060001000000001
ば | 接続助詞 | ば | 0
良い | 形容詞 | いい | 3000001000000011
か | 終助詞 | か | 0
教え | 本動詞 | おしえ | 2900001000000001
て頂け | 助動詞 | ていただけ | 0
ます | 助動詞 | ます | 0
か | 終助詞 | か | 0

十 | 数詞 | 10 | 12000010000000010
時まで | 格助詞 | じ まで | 9000001000000a000 0
開い | 本動詞 | あい | 2000001000000001
てい | 助動詞 | てい | 0
ます | 助動詞 | ます | 0
```

mor2bou 出力

```
ike 1 ba 1 iika 3 oshie 2 teetadakemasUka 5

juuji'made ai 1 teemasU 5
```

このように、動詞、形容詞の活用語尾を区切っていないため、動詞とその後に続く助詞などの付属語処理、アクセント処理が正しくできていない。

前章で TDMT 辞書の修正をしたので、TDMT の出力は次頁のように変わる。しかし、これまでの gfilter はこの正しい区切りに対応していないので、出力はまだ不完全であり、付属語エラーも起きてしまう。

- これまでの gfilter に活用語尾を区切りを正しいものに修正した TDMT の出力結果を入力した場合。

TDMT 出力 (辞書修正後)

行け | 本動詞 | い け | 2060001000000001 0
 ば | 接続助詞 | ば | 0
 良い | 形容詞 | い い | 3000001000000011 0
 か | 終助詞 | か | 0
 教え | 本動詞 | おしえ | 2900001000000001
 て頂け | 助動詞 | ていただけ | 0
 ます | 助動詞 | ます | 0
 か | 終助詞 | か | 0

 十 | 数詞 | 10 | 1200001000000010
 時まで | 格助詞 | じ まで | 900000100000a000 0
 開い | 本動詞 | あ い | 2000001000000001 0
 てい | 助動詞 | てい | 0
 ます | 助動詞 | ます | 0

mor2bou 出力

mor2bou 出力
 ike 1'ba 1 i'i ka 3 oshie 2 teetadakemasUka 1

 FUZOKUGO error ... recovered
 juuji'made ai 1 teemasU 5

上の入力の場合正しくは、

ike'ba 1 i'ika 1 shi'zukana heya'o oshiete itadakemasUka 1

このように出力されなければならない。従って、正しい mor2bou の出力が得られるように、gfilter を修正する必要がある。

3.3 gfilter の修正方法

gfilter の役割は TDMT の出力を、kan2rom の形態素解析結果に相当する mor2bou の入力へと変換することである。gfilter では、TDMT の出力を、awss の形態素解析の出力と同じ形に変換できれば良い。

gfilter の修正手順は次のようにした。

- 1 TDMT の出力サンプル（辞書修正後のもの）をテストし、付属語エラーになるものを調べる。
- 2 付属語エラーにならなくとも、アクセントがおかしいものを調べる。
- 3 gfilter から mor2bou へ入力されている情報を表示させるようにして、上記二項目で、対象となったサンプルが、awss の形態素解析と、どのように異なるのかを調べ、その要因を探索する。
- 4 要因が確認できたら、gfilter を修正する。
- 5 問題のサンプルおよび、できるだけ多くのサンプルを入力し、不具合が解消されているか、生じないかテストをする。

3.4 不具合の要因

修正手順 1、2、3 で判明した不具合の要因は以下のようなものである。ただし、以下では、awss の出力を比較しやすいように gfilter と同じ形式にしている。

- 要因 1

awss (kan2rom) の形態素解析では、動詞 (アクセント情報の 1 文字目が 2 のもの) の後に助詞 (アクセント情報が "0000000000000000" のもの) を、動詞の付属語に結合している。

————— TDMT 出力 —————

```
開い | 本動詞 | あ い | 2000001000000001 0  
てい | 助動詞 | てい | 0  
ます | 助動詞 | ます | 0
```

————— gfilter、awss からの mor2bou の入力 —————

```
(gfilter)  
開い ア イ 0 2000001000000001  
ています テイマス 0 0 0000000000000000  
  
(awss)  
開いています ア イ テイマス 0 2000001000000001
```

————— 両者の mor2bou からの出力 —————

```
(gfilter)  
ai 1 teemasU  
  
(awss)  
aite imasu
```

● 要因 2

要因 1 で、動詞が付属語を持たなければ、後ろの助詞を付属語にしている。

— TDMT の出力 —

見 | 本動詞 | み | 2900001000000011
まし | 助動詞 | まし | 0
た | 助動詞 | た | 0
か | 終助詞 | か | 0

— gfilter、awss からの mor2bou の入力 —

(gfilter)
見 ミ 0 0 2900001000000011
ましたか マシタカ 0 0 0000000000000000

(awss)
見ましたか ミ マシタカ 0 2900001000000011

— 両者の mor2bou からの出力 —

(gfilter)
mi' 2 mashItaka

(awss)
mima'shitaka

• 要因 3

形容詞、形容動詞の場合も動詞と同様の処理が行われている。

— TDMT の出力 —

嬉しい | 形容詞 | うれし い | 3000001000000021 0
です | 助動詞 | です | 0
確か | 形容名詞 | たしか | 4000001000000011
です | 助動詞 | です | 0

— gfilter、awss からの mor2bou の入力 —

(gfilter)
嬉しい ウレシ イ 0 3000001000000021
です デス 0 0 0000000000000000
確か タシカ 0 0 4000001000000011
です デス 0 0 0000000000000000

(awss)
嬉しいです ウレシ イデス 0 3000001000000021
確かです タシカ デス 0 4000001000000011

— 両者の mor2bou からの出力 —

(gfilter)
ureshi'i desu
tashIkadesU

(awss)
ureshi'idesu
ta'shikadesu 5

• 要因 4

語幹のないカ変、サ変動詞（来る、する）の場合は、後ろの助詞を結合したものを付属語にしている。

TDMT の出力

```
し | 本動詞 | し | 2b300010000000001
ます | 助動詞 | ます | 0
来 | 本動詞 | き | 2a000010000000011
てい | 助動詞 | てい | 0
まし | 助動詞 | まし | 0
た | 助動詞 | た | 0
```

gfilter、awss からの mor2bou の入力

```
(gfilter)
し し 0 0 2b300010000000001
ます マス 0 0 0000000000000000
来 キ 0 0 2a000010000000011
ていました テイマシタ 0 0 0000000000000000
```

```
(awss)
します シ シマス 0 2b300010000000001
来ていました キ キテイマシタ 0 2a000010000000011
```

両者の mor2bou からの出力

```
(gfilter)
shi 1 masU 5
ki' 1 teemashIta
```

```
(awss)
shima'su 5
ki'te imashita
```


• 要因 5

TDMT の形態素解析結果で 1 行に名詞、動詞など、助詞以外の品詞が複数混在している場合に、gfilter ではそれを統合し、先頭の品詞にしてしまっている。

TDMT の出力

```
レストランに関する | 格助詞 | れすとらん に かん する | 10000010000000010 0  
2b100010000000010 0
```

gfilter、awss からの mor2bou の入力

(gfilter)

```
レストランに関する レストランニカンスル 0 0 10000010000000010
```

(awss)

```
レストランに レストラン ニ 0 10000010000000010  
関する カン スル 0 2b100010000000010
```

両者の mor2bou からの出力

(gfilte)

```
resUtoraNnikaNsuru
```

(awss)

```
re'sutoraNni kaNsu'ru
```

3.5 対策

前節の要因に起因し、修正対象となるファイルは /gfilter/src/tdmt_in.c である。tdmt_in.c の中で TDMT の出力を 1 行ごとに構造体へセットしている関数 set_tdmtdata() に以下の処理を加える。

このプログラムの修正は次の通りである。

gfilter/src/tdmt_in.c [1.3] から /tdmt_in.c.[1.4]

- 要因 1、2 の対策処理

現在の入力エントリの品詞が助詞（アクセント情報"0"）であり、前エントリの品詞が、動詞（アクセント情報の先頭が"2"）であれば、助詞の読みを前エントリの動詞の付属語に結合させる。動詞に付属語がなければ、その助詞を付属語とする。

- 要因 3 の対策処理

現在の入力エントリの品詞が助詞であり、前エントリの品詞が形容詞、形容動詞（アクセント情報の先頭が"3"、"4"）であれば、動詞の場合と同様に付属語化させる。

- 要因 4 の対策処理

入力が語幹のないカ変、サ変動詞「来る、する」（アクセント情報の先頭が"2a"、"2b3"）であれば、自立語と付属語を結合させ、それを付属語とする。付属語がなければ、自立語を付属語にする。

- 要因 5 の対策処理

入力エントリ 1 行に、助詞を除いた品詞が複数混在している場合は、その 2 個目以降を分解し、2 行にわける関数 bunkai() を新たに加える。ただし、読み、アクセント情報などは、分解が可能だが、表層は分解困難であり、mor2bou での韻律処理に影響がないことから、分解後も同じ表層を用いる。

——— 分解例 ———

(TDMT 出力)

迷惑に対し | 格助詞 | めいわく に たいし | 1002401000000010 0 5000001000000011

(これまでの gfilter)

迷惑に対し メイワクニタイシ 0 0 1002401000000010

(修正後の gfilter)

迷惑に対し メイワク ニ 0 1002401000000010

迷惑に対し タイシ 0 0 5000001000000011

3.6 結果

サンプルをテストすることで、発見された gfilter と kan2rom の出力の違いは、このプログラムの修正によって、解決されている。また、修正のため新たに発生したバグは発見されていない。

以下に、要因で挙げたサンプルを修正後の gfilter に入力したものを示す。他のサンプルの出力結果は、巻末に掲載しているなので、そちらを参照。また gfilter の評価については、次章で示す。

修正後 gfilter からの mor2bou の入力

(要因 1)

開いています ア イテイマス 0 2000001000000001

(要因 2)

見ましたか ミ マシタカ 0 2900001000000011

(要因 3)

嬉しいです ウレシ イデス 0 30000010000000021

確かです たしか デス 0 40000010000000011

(要因 4)

します シ シマス 0 2b300010000000001

来ていました キ キテイマシタ 0 2a000010000000011

(要因 5)

レストランに関する レストラン ニ 0 10000010000000010

レストランに関する カン スル 0 2b100010000000010

第 4 章

修正した gfilter の評価

4.1 gfilter の評価

巻末に示してある、修正前と修正後の gfilter の出力を比較すると、以下のものであった。TDMT の出力は 99/02/19 に三研の土居氏により作成された

~doi/TMP/fujisawa/G-FILTER/J/ej*

を用いた。

- 読みがよくなったもの (ej は省略、サンプル番号のみ)
2,11,34,73,102
- アクセントがよくなったもの
4,5,6,10,11,12,13,32,38,50,52,57,64,80,81,86,90,92,97,99,103,104,106,
110,112,116,122,136,137,141,144,145,148,149,150,153,154,155,157,161
- アクセントが悪くなったもの
55,68,147
悪くなったのではなく修正前のものが、1 行中のサ変名詞と動詞を結合し、結果的に偶然アクセントがあっただけである。つまり、この問題は gfilter の問題ではなく、mor2bou のサ変名詞にまつわる問題の一つである。以降でも述べるが、サ変名詞周りのアクセント処理は問題が多い
- 付属語エラーがなくなったもの
7,57,108,114,120,121

- 新たに付属語エラーになったもの

9,24,26,37,44,46,49,62,67,75,77,87,89,113,118,124,131,133,146,152

これらは、決して gfilter に問題があるものではない。どれもエラーになるべきもので、修正前の gfilter でエラーになっていないのは、付属語処理ができていないからである。

以下にエラーになる理由を示す。

24,37,49,89,133,146,152 :

TDMT の出力がおかしい。「し」が「する」になっている。TDMT 側で修正予定

26,44,46,75,77,87,131 :

「くです」が 形容詞の付属語としておかしいと判定される。これは、付属語検索テーブルに「形容詞 -> く -> です」の遷移を登録すればエラーはなくなるが未対応。

67 :

TDMT の出力がおかしい。「必要ではあり」が自立語となっている。

TDMT 側で修正予定

113:

TDMT の出力がおかしい。「いる」「いない」の「い」の品詞 290 (上一段活用)が、「いる」「いらぬ」の「い」の 270 (ラ行五段活用)になっている。

これは TDMT 辞書で 270 で登録されているからで、290 に修正した。

- エラーのままのもの

9,62,118,124

以下にエラー理由を示す。

9 :

TDMT の出力が文法的におかしい。「予約の をします」 TDMT 側で修正予定。

62 :

TDMT の出力が文法的におかしい。「十二時間たら」

118 :

「にです」が 名詞の付属語としておかしいと判定される。これは、付属語検索

テーブルに「名詞 -> に -> です」の遷移を登録すればエラーにはならなくなるが未対応。

124 :

「持って行って」が「も って」「い って」になれば良い。辞書を修正したので解決するはず。

- その他

133 :

感動詞「それでは」は全て自立語で、付属語なしだが、mor2bou では、付属語の「は」「へ」「を」は「わ」「え」「お」に変換するが、自立語内の「は」などは変換されない、従って、「それでは」の読みを「それでわ」に辞書を修正。

4.2 考察、まとめ

今回、三研のほうから頂いた TDMT の出力サンプルでの、出力結果を見ると、アクセント処理は明らかに向上されている。正しい区切り対応させたことによるエラーなどは起こっていない。今回修正したのは、TDMT 出力を mor2bou 入力に形式変換させる部分であるが、ここでの役割は十分に満たしており、これ以上の変換はできないものと思われる。これ以上のアクセントの向上は、TDMT の形態素解析結果次第であると思われる。

第 5 章

gfilter 出力での BI の変更

この章と、次章は佐渡詩郎氏のテクニカルレポート (TR-IT-0291) に掲載されているものであるが、内容が gfilter に関わるものなので、こちらにも引用している。

5.1 概要

CHATR は、スペース区切り (BI 無し) のローマ字列を合成すると正しく target F0 を予測しない。mor2bou を修正し、この問題に対処したが (TR-IT-0291 参照)、出力修正箇所が mor2bou.c であるため、gfilter が出力する BI を変更するためには、gfilter のソースコードを変更する必要がある。

5.2 修正ファイル

修正したファイルは次の通りである。

```
gfilter/src/tdmt_mor2bou.c[1.3]
```

5.3 修正内容

tdmt_mor2bou.c に記述されている mor2bou_text_out() を変更し、BI を伴わない空白区切りが存在した場合、強制的に BI(1) を挿入するようにした。ただし、gfilter は、誤って BI を多く入れる場合があるようなので、BI が余分に出力される場合は、最も深い統語境界を持つ BI をひとつだけ出力するようにした。

第 6 章

gfilter の無声化処理の C 版への切り替え

6.1 概要

gfilter からは Perl 版の無声化フィルタが呼び出されていたが、C 版で実装した無声化フィルタを呼び出すようにした。

6.2 修正ファイル

修正したファイルは次の通りである。

```
gfilter/src/makefile[1.4]
gfilter/src/gfilter.c[1.5]
morph/PROSODY/cal_accbou/unvoice.c[1.3]
```

6.3 修正内容

gfilter.c 中に記述されている、Perl 版無声化フィルタを呼び出す関数 `unvoice()` を削除した。プロトタイプ宣言

```
static void unvoice( char * )
```

を削除し、新たにプロトタイプ宣言

```
extern void unvoice( char * )
```

を記述した。これは `unvoice.c` に記述されている `unvoice()` を呼び出すためのものである。

`unvoice.c` に記述されている `unvoice()` を修正した。 `mor2bou` はアクセント句ごとに `unvoice()` を呼び出すが、 `gfilter` は文ごとに `unvoice()` を呼び出す。このため、文ごと、アクセント句ごとのどちらで呼び出されても、文末の `/desu/`、 `/masu/` の無声化を行うようにした。

第 7 章

付属語検出エラーのリカバリ機能の実装

7.1 リカバリ機能

これまで、mor2bou に、文法的におかしな文章が入力されると、付属語検出エラーを発生することがあり、エラーとなった部分以降の読みが出力されないといった、問題があった。

たとえ、おかしな文章であっても、最低でも表層の読みだけは出力することが望まれる。

そのために、付属語検出エラー時に、表層の読みをそのまま出力とするリカバリ機能を用意したが、バグが存在することから、gfilter にリンクするときのみ用いられていた。

この機能を実装すべく、バグの修正を行う。

7.2 リカバリ機能のバグ

リカバリ機能のバグで明らかになっているのは、サ変名詞が、動詞として使用される場合、つまり付属語として「する」「される」などを持つ場合である。

<http://www.itl.atr.co.jp/~fujisawa/morph/bugs/aclose.html>

サ変名詞で付属語検出エラー を参照

このサ変名詞の場合、文法的に正しいのだが、リカバリ機能では付属語検出エラーになると判断し、表層の読みをアクセント処理など通さず、そのまま出力してしまうということである。

TDMT の出力

予約され | 助動詞 | よやく され | 1002401000000000 0 0
てい | 助動詞 | てい | 0
ます | 助動詞 | ます | 0

滞在する | 本動詞 | たいざい する | 1002401000000000 0
ので | 接続助詞 | ので | 0

リカバリ機能 ON 時の mor2bou 出力

FUZOKUGO error ... recovered
yoyakUsareteemasu

FUZOKUGO error ... recovered
taizai surunode

リカバリ機能 OFF 時の mor2bou 出力

yoyaku sarete imasu

taizai surunode

リカバリ機能を ON にすると、上の例などは付属語としての処理が誤っている。下の例は、OFF 時と同じ出力だが、FUZOKUGO error ... recoverd の文字が出力されている。サ変名詞の付属語をエラーと判断しないように修正しなければならない。

7.3 サ変名詞の付属語がエラーになる要因

リカバリ機能は morph/PROSODY/srh_htbl/srh_ftbl.c の searchftbl() 内にあるが、この関数が呼び出されるのは、morph/PROSODY/cal_accbou/inritu.c の ktstoirt() 内からである。このソースを確認したところ、リカバリ機能が付属語検出エラーと判断するのは、searchftbl() を抜ける時に、付属語が検出されていない時である。

しかし、サ変名詞などのように、他の派生情報があるもの（サ変名詞は名詞としてと、動詞としての2通りある）は、一度この searchftbl() を抜け ktstoirt() に戻り、品詞のエントリ番号を変えてから、再び searchftbl() で付属語を検出するようになっている。

従って、これまで、派生情報（アクセント情報の4文字目）が1以上のものは、そのときの活用の上の方次第では、サ変名詞に関わらず、リカバリ機能で誤って、エラーと判断されていたということになる。

これを修正するため、まず presearchftbl() という関数を作成する。そして、ktstoirt() から searchftbl() を呼び出す前に presearchftbl() を呼び、その中で searchftbl() を呼び出すようにする。presearchftbl() 内で、派生情報の有無を考慮して、付属語の検出エラーの判定を行うようにした。

この時派生情報の有無として、ktstoirt() から presearchftbl() には hsg_fg（他の派生情報を持つ時1、派生情報を持たない時0）が渡される。

次頁にソースの修正の様子を図でしめす。

このプログラムの修正は

```
morph/PROSODY/cal_accbou/inritu.c [1.7] から /inritu.c [1.8]  
morph/PROSODY/srh_ftbl/srh_ftbl.c [1.7] から /srh_ftbl.c [1.8]
```

へのものである。

inritu.c

```
ktstoirt(){  
  
    for(;;){  
  
        searchftbl()  
  
        他に派生情報があれば  
        entry 変更  
    }  
}
```

srh ftbl.c

```
searchftbl(){  
  
    付属語検索  
  
    付属語がなければ  
    リカバリ機能  
}
```

図 7.1: 修正前の inritu.c と srh_ftbl.c

inritu.c

```
ktstoirt(){  
  
    for(;;){  
  
        hsg_fg= 0 or 1  
        presearchftbl()  
        他に派生情報があれば  
        entry 変更  
    }  
}
```

srh ftbl.c

```
searchftbl(){  
    付属語検索  
}  
presearchftbl()  
{  
    searchftbl()  
    付属語がなく hsg_fg=0 なら  
    リカバリ機能  
}
```

図 7.2: 修正後の inritu.c と srh_ftbl.c

7.4 結果

結果は以下のようにリカバリ機能 OFF 時と同じようになり、サ変名詞が付属語検出エラーと判断されることはなくなった。また、本来エラーになるべきものは、そのままエラーが検出されるようになっている。

TDMT 出力

予約され | 助動詞 | よやく され | 1002401000000000 0 0

てい | 助動詞 | てい | 0

ます | 助動詞 | ます | 0

滞在する | 本動詞 | たいざい する | 1002401000000000 0

ので | 接続助詞 | ので | 0

(上のを誤りにしたもの)

滞在する | 本動詞 | たいざい する | 1002401000000000 0

で | 接続助詞 | で | 0

mor2bou 出力

yoyaku sarete imasu

taizai surunode

FUZOKUGO error ... recovered

taizaisurude

7.5 考察、まとめ

付属語検出エラー時のリカバリ機能は修正に起因した新しいバグもなく問題なく実装できているようである。

しかし、mor2bou には、サ変名詞の付属語が離れてしまい、付属語のアクセント処理が行われていない、というバグが存在する。

<http://www.itl.atr.co.jp/~fujisawa/morph/bugs/aopen.html>

サ変名詞につく付属語が分離する

前頁の「滞在するので」は本来「taizaisuru'node」が正しいアクセント処理である。この直接の原因は srh_ftbl.c serchftbl() 内の、アクセント句の区切りという部分であることは判明している。

サ変名詞の付属語の「する」「される」などは、付属語検索の結果、function = F_JCONTQ というのが与えられ、この結果、アクセント句の区切りというところで、区切られているのである。

サ変名詞の付属語は、エン트리番号として 241 というのもっているもので、これを条件文に利用すれば、アクセント句を区切らずには済み、付属語のアクセント処理も行われるようになる。

しかし、「する」に関しては、「する」自体がアクセントを持たないため、問題ないのであるが、「します」などは「しま'す」とアクセントを持つ。サ変名詞自体が、アクセントを持たない場合は、問題ないのだが、アクセントを持つもの「キャ'ンセル」などは、付属語「します」を持つと、アクセントが「キャンセルしま'す」と、一つにされてしまう。

すなわち、サ変名詞は、付属語として、「しま'す」「した'い」「すれ'ば」などアクセントを持つものを持ってはいけないといえる。

従って、サ変名詞の付属語が離れてしまうというバグは、今のままでは、付属語を結合するわけにもいかず、今後の大きな課題になるであろう。

謝辞

本研究を進めるにあたって、暖かく見守りつつ、多くの御指導を頂いた Nick Campbell 第二研究室室長、藤澤 謙研究員に心から感謝致します。

さらに、本研究に対する適切な助言を頂き、様々な相談にのって下さった ATR 音声翻訳研究所第二研究室の皆様へ深く感謝致します。

1999 年 2 月 26 日

酒井 慎介

第 8 章

修正したファイルのリスト

- TDMT 辞書の修正

gfilter/dic/japanese/dic/jg.dic [1.10]
gfilter/dic/japanese/dic/doshi.dic [1.5]
gfilter/dic/japanese/dic/keiyou.dic [1.2]
gfilter/dic/japanese/dic/jodoshi.dic [1.2]
gfilter/dic/japanese/dic/setsuji.dic [1.4]
gfilter/dic/japanese/dic/fukushi.dic [1.4]
gfilter/dic/japanese/dic/meishi.dic [1.2]

- gfilter の改良

gfilter/src/tdmt_in.c [1.4]

- 佐渡氏のテクニカルレポート内の内容 5 章 6 章で修正されたもの

- gfilter 出力での BI の変更

gfilter/src/tdmt_mor2bou.c [1.3]

- gfilter の無声化処理の C 版への切り替え

gfilter/src/makefile [1.4]

gfilter/src/gfilter.c [1.5]

morph/PROSODY/cal_accbou/unvoice.c [1.3]

- リカバリ機能の実装

morph/PROSODY/cal_accbou/inritu.c [1.8]

morph/PROSODY/srh_ftbl/srh_ftbl.c [1.8]

第 9 章

改良後の gfilter の出力結果

以下には、正しい区切りに対応できるように改良した gfilter に 3 研からのサンプルテキストを入力し、出力された結果を示す。比較できるように、修正前の gfilter の出力も示す。どちらの場合も付属語検索エラーのリカバリ機能も実装している。

サンプルテキストは ~xssakai/simura2/Ssample/ej*** である。

上が修正前 gfilter の出力、下が修正後 gfilter の出力である。

ej001:

```
sui'itowa 1 i'ppaku 1 nanahyakudo'rudesu 1 tsuiNru'umuwa 2 i'ppaku
1 saNbyakudo'rudesU 5
sui'itowa 1 i'ppaku 1 nanahyakudo'rudesu 1 tsuiNru'umuwa 2 i'ppaku
1 saNbyakudo'rudesU 5
```

ej002:

```
ha'i 1 jeea'aruno oosaka'ekIkara 1 sUta'jiamuni iku hoohoowooshiete'etadakemasUka 5
ha'i 1 jeea'aruno oosaka'ekIkara 1 sUta'jiamuni iku hoohoowo 1 oshiete
itadakemasUka 5
```

ej003:

```
ha'i 1 su'uzaN 3 bura'uN 3 e'su 3 yu'u 3 e'su 3 e'e 3 e'nu 3 bi'i 3 a'aru
3 o'o 3 da'buryuu 3 e'nudesU 5
ha'i 1 su'uzaN 3 bura'uN 3 e'su 3 yu'u 3 e'su 3 e'e 3 e'nu 3 bi'i 3 a'aru
3 o'o 3 da'buryuu 3 e'nudesU 5
```

ej004:

```
deki'reba 3 ho'teruni 1 mo'o 1 nihaku'o 1 taizai surunode 2 heya'o kae 2 taidesU 5
```

deki'reba 3 ho'teruni 1 mo'0 1 nihaku'o 1 taizai surunode 2 heya'o kaetai'desU 5

ej005:

jidaima'tsurino fUtsUkama'eto 1 chekkui'N shItaidesu 1 do'chiraga 2 juugatsu'
1 ni'juu ni' 1 hiniari 1 masUka 5

jidaima'tsurino fUtsUkama'eto 1 chekkui'N shItaidesu 1 do'chiraga 2 juugatsu'
1 ni'juu ni' 1 hini arima'sUka 5

ej006:

ka'igino kamimaeo 1 te'hai suru 1 koto'ga deki'ru yooni 3 shizukanaheyaga
i'i desU 5

ka'igino kamimaeo 1 te'hai suru 1 koto'ga deki'ru yooni 1 shizukanaheyaga
i'idesU 5

ej007:

FUZOKUGO error ... recovered

ha'i 1 ho'teruno 1 nikaino 1 shoteNwa 2 go'go 1 juuji'made ai 1 teemasU 5

ha'i 1 ho'teruno 1 nikaino 1 shoteNwa 2 go'go 1 juuji'made aite imasU 5

ej008:

sorekara 3 kokono deNwaba'Ngoowa 2 re'e 3 saN 3 saN 3 ni' 3 yo'N 3 go' 3 ni'
3 ni' 3 saN 3 saNde'su 5

sorekara 3 kokono deNwaba'Ngoowa 2 re'e 3 saN 3 saN 3 ni' 3 yo'N 3 go' 3 ni'
3 ni' 3 saN 3 saNde'su 5

ej009:

FUZOKUGO error ... recovered

yutta jikaNno aidani 1 ha'i 1 chokUsetsu 1 reNraku shinakereba narimaseN
1 yoyakunowoshimasu 3 3

FUZOKUGO error ... recovered

yutta jikaNno aidani 1 ha'i 1 chokUsetsu 1 reNraku shinakereba narimaseN
1 yoyakunowoshimasu 3 3

ej010:

namaeo shiri 2 taidesUtsu'aani 1 nanimyooowa saNka shimasU 5

namaeo shirita'idesu 1 tsu'aani 1 nanimyooowa saNka shimasU 5

ej011:

sorekara 3 i'kurani na'ru ka 3 oshie 2 teetadakemasUka 3 do'noyooni 1 soreno
da'ikiNo shIharau ho'oga i'i noka 5

sorekara 3 i'kurani na'ruka 1 oshiete itadakemasUka 1 do'noyooni 1 soreno
da'ikiNo shIharau ho'oga i'inoka 5

ej012:

resUtoraNnikaNsurujo'ohooga 2 ari 2 masUka ko'obeno 1 i'i yo'runo nagamede'su
5

re'sUtoraNni kaNsu'ru jooohooga arima'sUka 1 ko'obeno 1 i'i yo'runo nagamede'su
5

ej013:

yoroshi'i desu 1 sorede'ha 3 kakuniNba'Ngoowa 2 go' 3 ichi' 3 saN 3 yo'N
3 yo'Ndesu 1 kooeNwa 1 juubaNmeno migino go'go 1 saNji'ni hajime 2 masU 5
yoroshi'idesu 1 sorede'ha 3 kakuniNba'Ngoowa 2 go' 3 ichi' 3 saN 3 yo'N
3 yo'Ndesu 1 kooeNwa 1 juubaNmeno migino go'go 1 saNji'ni hajimema'su 5

ej014:

taiheN 1 sumimase'N 1 yoyaku sareru 1 kiNyo'obiwa aru to 1 mooshiwakena'i
1 nichiyobidesu 1 shima'tte imasU 5

taiheN 1 sumimase'N 1 yoyaku sareru 1 kiNyo'obiwa aru'to 1 mooshiwakena'i
1 nichiyobidesu 1 shima'tte imasU 5

ej015:

ha'i 1 kyo'o kokunaifu'raitode 2 tookyoo shUsshiN 3 ja'runo ichi' 3 re'e
3 kyuubi'NdesU 5

ha'i 1 kyo'o kokunaifu'raitode 2 tookyoo shUsshiN 3 ja'runo ichi' 3 re'e
3 kyuubi'NdesU 5

ej016:

koNnichIha 5

koNnichIha 5

ej017:

peniNshuraho'terudesU 5

peniNshuraho'terudesU 5

ej018:

do'oshimashItaka 5

do'oshimashItaka 5

ej019:

ha'i 3 3 3

ha'i 3 3 3

ej020:

do'oshimashItaka 5

do'oshimashItaka 5

ej021:

mochi'roN 1 desU 5

mochi'roN 1 desU 5

ej022:

na'Njini 1 chekkui'N suru 1 yoteede'sUka 5

na'Njini 1 chekkui'N suru 1 yoteede'sUka 5

ej023:

mochi'roN 1 desU 5

mochi'roN 1 desU 5

ej024:

de'suga 3 maemo'tte 1 soreno da'ikiNo shIharau yooni

1 onegaisuranakerebanarimaseN 5

FUZOKUGO error ... recovered

de'suga 3 maemo'tte 1 soreno da'ikiNo 1 shIharau yooni

1 o negaisu'ranakerebanarimaseN 5

ej025:

de'ha 3 kurejittoka'adono baNgo'oo ukagatte'mo 1 yoroshi'i desUka 5

de'ha 3 kurejittoka'adono baNgo'oo ukagatte'mo 1 yoroshi'idesUka 5

ej026:

machigaina'ku 2 desU 5
FUZOKUGO error ... recovered
machigaina'kudesU 5

ej027:

saika'kuniN sasete kudasai 5
saika'kuniN sasete kudasai 5

ej028:

masUtaaka'adodesU 5
masUtaaka'adodesU 5

ej029:

baNgo'o 1 go' 3 ni' 3 na'na 3 kyu'u 3 saN 3 kyu'u 3 ni' 3 re'e 3 ni' 3 yo'N
3 roku' 3 kyu'u 3 re'e 3 re'e 3 kyu'u 3 hachi'desU 5
baNgo'o 1 go' 3 ni' 3 na'na 3 kyu'u 3 saN 3 kyu'u 3 ni' 3 re'e 3 ni' 3 yo'N
3 roku' 3 kyu'u 3 re'e 3 re'e 3 kyu'u 3 hachi'desU 5

ej030:

tekudasai 5
tekudasai 5

ej031:

watashino namaewa 2 jo'N 3 fi'rippUsudesU 5
watashino namaewa 2 jo'N 3 fi'rippUsudesU 5

ej032:

sorekara 3 sa'rani 1 shItsumoNga aru baaini'wa 2 eNryo shinai 1 dekudasai
3 deNwade'su 5
sorekara 3 sa'rani 1 shItsumoNga aru baaini'wa 2 eNryo shinaide 1 kudasa'i
1 deNwade'su 5

ej033:

dooitashima'shIte 5
dooitashima'shIte 5

ej034:

de'ha 3 peniNshuraho'terude 2 ana'tani 1 a'u nowo tanoshi'mini shi 1 teemasU 5

de'ha 3 peniNshuraho'terude 2 ana'tani 1 a'u nowo tanoshi'mini shIte imasU 5

ej035:

mo'shimoshi 1 furoNtode'sUkudesu 1 do'oshimashItaka 5

mo'shimoshi 1 furoNtode'sUkudesu 1 do'oshimashItaka 5

ej036:

mochi'roN 1 desU 5

mochi'roN 1 desU 5

ej037:

deki'ru 1 kagi'ri 2 nokotoo su 1 masU 5

FUZOKUGO error ... recovered

deki'ru 1 kagi'rino 1 koto'o surumasU 5

ej038:

do'kode 2 ni'motsuo 1 sa'igoni mi' 2 mashItaka 5

do'kode 2 ni'motsuo 1 sa'igoni mima'shItaka 5

ej039:

wakarima'shIta 5

wakarima'shIta 5

ej040:

taiheN 1 sumimase'N 3 3

taiheN 1 sumimase'N 3 3

ej041:

shinamonoo 1 setsumee shIte itadakemasUka 5

shinamonoo 1 setsumee shIte itadakemasUka 5

ej042:

wakarima'shIta 5

wakarima'shIta 5

ej043:

sorede'ha 3 tsu'aade 2 kokoni iru nodesune 5

sorede'ha 3 tsu'aade 2 kokoni iru'nodesune 5

ej044:

machigaina'ku 2 desU 5

FUZOKUGO error ... recovered

machigaina'kudesU 5

ej045:

sorekara 3 uketotta reshi'itono baNgo'oo ukagatte'mo 1 i'i desUka 5

sorekara 3 uketotta reshi'itono baNgo'oo ukagatte'mo 1 i'idesUka 5

ej046:

machigaina'ku 2 desU 5

FUZOKUGO error ... recovered

machigaina'kudesU 5

ej047:

sorekara 3 su'guni 1 shinamonono na'kani atta na'nikaga 2 hItsuyoode'sUka 5

sorekara 3 su'guni 1 shinamonono na'kani atta na'nikaga hItsuyoode'sUka 5

ej048:

wakarima'shIta 5

wakarima'shIta 5

ej049:

ha'i 1 soo su 1 masU 5

FUZOKUGO error ... recovered

ha'i 1 soo surumasU 5

ej050:

mitsUkeru made 3 da'rekani 1 soodaN shi 1 kanarazu sagashi 2 masU 5

mitsUkeruma'de 1 da'rekani 1 soodaN shi 1 kanarazu sagashima'su 5

ej051:

mitsUkeshidaide'Nwa shimasU 5

mitsUkeshidaide'Nwa shimasU 5

ej052:

ana'tani 1 okake shIte iru 1 me'ewakunitaishi 1 taiheN 1 sumimase'N 5

ana'tani 1 okake shIte iru 1 me'ewakuni 1 ta'ishi 1 taiheN 1 sumimase'N 5

ej053:

sayoonara 5

sayoonara 5

ej054:

mo'shimoshi 5

mo'shimoshi 5

ej055:

gureehaNdo'shani 1 ode'Nwaitadaki 3 ari'gatoo 5

gureehaNdo'shani 1 odeNwai'tadaki 1 ari'gatoo 5

ej056:

do'oshimashItaka 5

do'oshimashItaka 5

ej057:

FUZOKUGO error ... recovered

sorutoreekUshI'tiikara 1 rasube'gasumade ika 1 nakerebanarimaseN 5

sorutoreekUshI'tiikara 1 rasube'gasumade ikana'kereba narimaseN 5

ej058:

so'odesune 5

so'odesune 5

ej060:

shUppatsuji'kaNwa 2 re'ej1 1 juugo'fuN 1 go'zeN 1 kyu'uji 1 juugo'fuN

1 sho'ogodesU 5

shuppatsuji'kaNwa 2 re'ej1 1 juugo'fuN 1 go'zeN 1 kyu'uji 1 juugo'fuN

1 sho'ogodesu 5

ej061:

eeto 1 go'zeN 1 kyu'uji 1 juugofu'Nno 1 ba'suwa 2 go'go 1 hachiji'madeni
1 toochaku shimasU 5

eeto 1 go'zeN 1 kyu'uji 1 juugofu'Nno 1 ba'suwa 2 go'go 1 hachiji'madeni
1 toochaku shimasu 5

ej062:

FUZOKUGO error ... recovered

iie 1 jikaNno tameni 1 juunijikaNtara' 1 heNkoo shIte kudasai 5

FUZOKUGO error ... recovered

iie 1 jikaNno tameni 1 juunijikaNtara' 1 heNkoo shIte kudasai 5

ej063:

otonano katamichiu'NchiNwa 2 hachijuu nido'ru 1 nana'juu 1 seNtode'su 5

otonano katamichiu'NchiNwa 2 hachijuu nido'ru 1 nana'juu 1 seNtode'su 5

ej064:

ta'no shItsumoNga ari 2 masUka 5

ta'no shItsumoNga arima'sUka 5

ej065:

rasube'gasUkara 1 rosaNje'rusumadeno otonano katamichiu'NchiNwa yo'Njuu
yoNdo'rudesU 5

rasube'gasUkara 1 rosaNje'rusumadeno otonano katamichiu'NchiNwa yo'Njuu
yoNdo'rudesU 5

ej066:

ryokoo 3 sore 3 rokuji'kaN 1 kaka'tte kudasai 5

ryokoo 3 sore 3 rokuji'kaN 1 kaka'tte kudasai 5

ej067:

iie 1 yoyakuwa 2 hItsuyooodehaarima'seN 5

FUZOKUGO error ... recovered

iie 1 yoyakuwa hItsuyooodehaarimaseN 5

ej068:

gureehaNdo'shani 1 ode'Nwaitadaki 3 ari'gatoo 5
gureehaNdo'shani 1 odeNwai'tadaki 1 ari'gatoo 5

ej070:

koNnichIha 5
koNnichIha 5

ej071:

kaNko'okyokudesU 5
kaNko'okyokudesU 5

ej072:

do'oshimashItaka 5
do'oshimashItaka 5

ej073:

do'noyoona ho'teruo sagashi 2 teemasUka 5
do'noyoona ho'teruo sagashIte imasUka 5

ej074:

sorekara 3 yosaNno habade'su 5
sorekara 3 yosaNno habade'su 5

ej075:

machigaina'ku 2 desU 5
FUZOKUGO error ... recovered
machigaina'kudesU 5

ej076:

shokujitsUkiga i'i desUka 5
shokujitsUkiga i'idesUka 5

ej077:

machigaina'ku 2 desU 5
FUZOKUGO error ... recovered
machigaina'kudesU 5

ej078:

cho'tto ma'tte kudasai 5

cho'tto ma'tte kudasai 5

ej079:

omatasesUtteete 1 mooshiwakearimase'N 5

omatase sUsUssUteete 1 mooshiwakearimase'N 5

ej080:

ekUserushio'oruniwa 2 i'ppaku 1 rokujuu godo'runo heya'ga ari 2 masU 5

ekUserushio'oruniwa 2 i'ppaku 1 rokujuu godo'runo heya'ga arima'su 5

ej081:

sorekara 3 hoterumaNsUfi'irudoniwa 2 rokujuu godoru 1 kara 1 i'ppaku

1 nana'juu godo'runo heya'ga ari 2 masU 5

sorekara 3 hoterumaNsUfi'irudoniwa 2 rokujuu godoru 1 kara 1 i'ppaku

1 nana'juu godo'runo heya'ga arima'su 5

ej082:

hoterumaNsUfi'irudowa 1 buroodooe'eni 1 totemo chIka'i desU 5

hoterumaNsUfi'irudowa 1 buroodooe'eni 1 totemo chIka'idesU 5

ej083:

mochi'roN 1 desU 5

mochi'roN 1 desU 5

ej084:

namaeto 1 donokuraino jikaNo 1 onegai dekimasUka 1 taizai shimasU 5

namaeto 1 donokuraino jikaNo 1 onegai dekimasUka 1 taizai shimasU 5

ej086:

na'Njini 1 ho'teruni tsU'ku tomoi 1 masUka 5

na'Njini 1 ho'teruni tsU'kUto omoima'sUka 5

ej087:

machigaina'ku 2 desU 5

FUZOKUGO error ... recovered

machigaina'kudesU 5

ej088:

yoyakuo 1 kakuniN suru 1 aida 3 sUko'shi ma'tte itadakemasUka 5

yoyakuo 1 kakuniN suru 1 aida 3 sUko'shi ma'tte itadakemasUka 5

ej089:

omatasesuri 1 mooshiwakearimase'N 3 3 3

FUZOKUGO error ... recovered

omatase suri 1 mooshiwakearimase'N 3 3 3

ej090:

koNbaNyo'yakuga ari 2 masU 5

koNbaNyo'yakuga arima'su 5

ej091:

mochi'roN 1 desU 5

mochi'roN 1 desU 5

ej092:

gobaNgaito' 1 rokuba'Ngaino aidano 1 yo'Njuu yoNchoomeni 1 ari 2 masU 5

gobaNgaito' 1 rokuba'Ngaino aidano 1 yo'Njuu yoNchoomeni 1 arima'su 5

ej093:

sorekara 3 deNwaba'Ngoowa 2 kyu'u 3 yo'N 3 yo'N 3 roku' 3 re'e 3 go' 3 re'edesU 5

sorekara 3 deNwaba'Ngoowa 2 kyu'u 3 yo'N 3 yo'N 3 roku' 3 re'e 3 go' 3 re'edesU 5

ej094:

dooitashima'shIte 5

dooitashima'shIte 5

ej095:

sayoonara 5

sayoonara 5

ej096:

ha'i 1 juuni' 3 gojuugo'oshItsuno fi'rippUsudesU 5

ha'i 1 juuni' 3 gojuugo'oshItsuno fi'rippUsudesU 5

ej097:

eakoNwa koware' 2 teerutoomoi 1 masU 5

eakoNwa koware'te iruto omoima'su 5

ej098:

suzushi'i ku'ukiga de' 2 maseN 5

suzushi'i ku'ukiga demase'N 5

ej099:

ha'i 1 deNgeNwa ha'itte imasu 1 ku'ukiwa de' 2 teemasu 1 ta'da 1

o'NdogasagatteemaseN 5

ha'i 1 deNgeNwa ha'itte imasu 1 ku'ukiwa de'te imasu 1 ta'da 1 o'Ndogo

saga'tte imaseN 5

ej100:

ari'gato 5

ari'gato 5

ej101:

gozo'Njino yooni 2 kokowa 2 hoNtooni atsu'i desU 5

gozo'Njino yooni 2 kokowa 2 hoNtooni atsu'idesU 5

ej102:

su'guni 1 da'rekao yoko'shi 2 teetadakemasUka 5

su'guni 1 da'rekao yoko'shIte itadakemasUka 5

ej103:

ma'doo akeru koto'ga deki'ru nara 3 tasUka'ri 2 masU 5

ma'doo akeru koto'ga deki'runara 1 tasUkarima'su 5

ej104:

donokuraino kIkaNe'akoNo 1 shu'uri suru 1 nonikakari 1 masUka 5

donokuraino kIkaNe'akoNo 1 shu'uri suru 1 noni 1 kakarima'sUka 5

ej105:

kii 1 te 1 ureshi'i desu 3 3

kiite 1 ureshi'idesu 3 3

ej106:

shu'uriga ha'yaku okonawareru koto'ga deki'ru nara 3 iie 1 jitsU'ha 1 hoNtooni

1 wa'zawaza 1 ro'biini iki 2 takuarimaseN 5

shu'uriga ha'yaku okonawareru koto'ga deki'runara 1 iie 1 jitsU'ha 1 hoNtooni

1 wa'zawaza 1 ro'biini ikIta'ku arimaseN 5

ej107:

ha'i 5

ha'i 5

ej108:

FUZOKUGO error ... recovered

yoroshi'i 1 de'mo 1 iso'i 1 dekudasai 5

yoroshi'i 1 de'mo 1 iso'ide kudasa'i 5

ej109:

mo'shimoshi 5

mo'shimoshi 5

ej110:

nyuuyo'okUkaNkookyokudesU 5

nyuuyo'oku kaNko'okyokudesU 5

ej111:

do'oshimashItaka 5

do'oshimashItaka 5

ej112:

tashIkadesU 5

ta'shIkadesU 5

ej113:

imashi'naini i 1 masune 5

FUZOKUGO error ... recovered
imashi'naini imasune 5

ej114:

FUZOKUGO error ... recovered
amerika'gaito 1 gobaNgaio' shIttee 1 masUka 5
amerika'gaito 1 gobaNgaio' shItte imasUka 5

ej116:

bijutsU'kaNwa 2 rokkUferaase'Ntaano ichiburo'kku kItani ari 2 masU 5
bijutsU'kaNwa 2 rokkUferaase'Ntaano ichiburo'kku kItani arima'su 5

ej117:

de'suga 3 bijutsU'kaNwa 2 kyo'o 1 shima'tte imasU 5
de'suga 3 bijutsU'kaNwa 2 kyo'o 1 shima'tte imasU 5

ej118:

FUZOKUGO error ... recovered
maishuusuiyo'obito 1 juunigatsu' 1 ni'juu go' 1 hinokurisumasu'nidesU 5
FUZOKUGO error ... recovered
maishuusuiyo'obito 1 juunigatsu' 1 ni'juu go' 1 hino kurisumasu'nidesU 5

ej119:

ha'i 1 sumimase'N 5
ha'i 1 sumimase'N 5

ej120:

FUZOKUGO error ... recovered
kiNyo'obikara 1 kayo'obimade 1 go'zeN 1 juuichiji'kara 1 go'go 1 rokuji'made
ai 1 teemasU 5
kiNyo'obikara 1 kayo'obimade 1 go'zeN 1 juuichiji'kara 1 go'go 1 rokuji'made
aite imasU 5

ej121:

FUZOKUGO error ... recovered
sorekara 3 mokuyo'obiniwa 1 kyuuji'made ai 1 teemasU 5
sorekara 3 mokuyo'obiniwa 1 kyuuji'made aite imasU 5

ej122:

ze'eo fUkume' 2 teo'tonanitaishi 1 yoNdo'ru 1 goju'u 1 seNtode'su 5
ze'eo fUkume'te 1 otonani 1 ta'ishi 1 yoNdo'ru 1 goju'u 1 seNtode'su 5

ej123:

dooitashima'shIte 5
dooitashima'shIte 5

ej124:

FUZOKUGO error ... recovered
kigaruni 1 ha'itte kudasai 1 nyuuyo'okuno muryoono gaidobu'kkuo motteittekudasai 5
FUZOKUGO error ... recovered
kigaruni ha'itte kudasai 1 nyuuyo'okuno muryoono gaidobu'kkuo motteittekudasai
5

ej126:

koNbaNha 1 eaapashIfI'kkudesU 5
koNbaNha 1 eaapashIfI'kkudesU 5

ej127:

me'arii 3 fi'rippUsudesU 5
me'arii 3 fi'rippUsudesU 5

ej128:

do'oshimashItaka 5
do'oshimashItaka 5

ej129:

ma'sanisono 1 eeto 5
ma'sanisono 1 eeto 5

ej130:

namaeto 1 ge'Nzai 1 yoyaku sarete iru 1 furaitoo ukagatte'mo 1 yoroshi'i desUka 5
namaeto 1 ge'Nzai 1 yoyaku sarete iru 1 furaitoo ukagatte'mo 1 yoroshi'idesUka 5

ej131:

machigaina'ku 2 desU 5
FUZOKUGO error ... recovered
machigaina'kudesU 5

ej132:

sorede'ha 3 nibi'No 1 kya'Nseru shItaidesu 3 a 1 hachigatsu' 1 ju'u 1 hino
1 ichi'desU 5
sorede'ha 3 nibi'No 1 kya'Nseru shItaidesu 3 a 1 hachigatsu' 1 ju'u 1 hino
ichi'desU 5

ej133:

dochirano furaitoni su 1 masUka 5
FUZOKUGO error ... recovered
dochirano furaitoni surumasUka 5

ej134:

sUko'shi ma'tte kudasai 5
sUko'shi ma'tte kudasai 5

ej135:

taiheN 1 sumimase'N 1 ga 3 sono furaitowa yoyaku sarete imasU 5
taiheN 1 sumimase'N 1 ga 3 sono furaitowa yoyaku sarete imasU 5

ej136:

cho'ttomattekudasai shira'beru 2 masU 5
cho'tto ma'tte kudasa'i 1 shiraberuma'su 5

ej137:

nibi'Nni 1 akiga ari 2 masu 3 a 1 go'go 1 saNji'ni 1 nyuuyo'okuno
jeeefUke'eo de' 2 te 3 go'ji 1 saNjuppu'Nni 1 rosaNze'rusUkokUsaikuukooni
1 toochaku suru 1 hachigatsu' 1 juuichi' 1 hino 1 na'nadesU 5
nibi'Nni 1 akiga arima'su 3 a 1 go'go 1 saNji'ni 1 nyuuyo'okuno
jeeefUke'eo de'te 1 go'ji 1 saNjuppu'Nni 1 rosaNze'rusu kokUsaiku'ukooni
1 toochaku suru 1 hachigatsu' 1 juuichi' 1 hino na'nadesU 5

ej138:

ma'sanisono 1 eeto 5

ma'sanisono 1 eeto 5

ej139:

yoyakuo 1 kakuniN sasete kudasai 5

yoyakuo 1 kakuniN sasete kudasai 5

ej140:

ka'zukodesu 3 3

ka'zukodesu 3 3

ej141:

atarashi'i furaitowa niba'Ndesu 3 a 1 go'go 1 saNji'ni 1 nyuuyo'okuno

jeeefUke'eo de' 3 go'go 1 go'ji 1 saNjuppu'Nni 1 rosaNze'rusUkokUsaikuukooni

1 toochaku suru 1 hachigatsu' 1 juuichi' 1 hinomonode'su 5

atarashi'i furaitowa niba'Ndesu 3 a 1 go'go 1 saNji'ni 1 nyuuyo'okuno

jeeefUke'eo de' 3 go'go 1 go'ji 1 saNjuppu'Nni 1 rosaNze'rusu kokUsaiku'ukooni

1 toochaku suru 1 hachigatsu' 1 juuichi' 1 hino mono'desU 5

ej142:

sorekara 3 nibi'Nno 1 furu'i yoyakuo 1 kya'Nseru shimasu 3 a 1 hachigatsu'

1 ju'u 1 hino 1 ichi'desU 5

sorekara 3 nibi'Nno 1 furu'i yoyakuo 1 kya'Nseru shimasu 3 a 1 hachigatsu'

1 ju'u 1 hino ichi'desU 5

ej143:

mattaku' tadashi'i desUka 5

mattaku' tadashi'idesUka 5

ej144:

kaNtaNdesU 5

kaNtaNde'su 5

ej145:

atarashi'i chIke'ttoo to'ri 2 masU 5

atarashi'i chIke'ttoo torima'su 5

ej146:

i'i ryokooo su 1 tekudasai 5
FUZOKUGO error ... recovered
i'i ryokooo surutekudasai 5

ej147:

eaapashIfI'kkuo 1 goriyooitadaki 3 ari'gato 5
eaapashIfI'kkuo 1 goriyooi'tadaki 1 ari'gato 5

ej148:

mo'shimoshi 1 kootsuushudaNnikaNsurujo'ohooo e' 2 taidesU 5
mo'shimoshi 1 kootsu'ushudaNni kaNsu'ru jooohoo eta'idesU 5

ej149:

e'Nryakujini iki 2 taidesU 5
e'Nryakujini ikIta'idesU 5

ej150:

do'oyatte teniire'ru koto'ga deki' 2 masUka 3 3
do'oyatte teniire'ru koto'ga dekima'sUka 3 3

ej151:

do'chiraga 2 ichibaN haya'i desUka 5
do'chiraga 2 ichibaN haya'idesUka 5

ej152:

norikaeo su 1 nakerebanarimaseNka 5
FUZOKUGO error ... recovered
norikaeo surunakerebanarimaseNka 5

ej153:

yoNsai'no 1 musumetoisshoniryo'koo shIte imasU 5
yoNsai'no 1 musume'to 1 isshoni 1 ryokoo shIte imasU 5

ej154:

do'chiraga 2 ichibaN 1 beNride'sUka 5
do'chiraga 2 ichibaN be'NridesUka 5

ej155:

sorekara 3 do'kokara 1 ba'suwa de' 2 masUka 5

sorekara 3 do'kokara 1 ba'suwa dema'sUka 5

ej157:

ba'suga de'ru 1 i'kUtsUkano jikaNo oshieru koto'ga deki're 1 ba 3 i'i desU 5

ba'suga de'ru 1 i'kUtsUkano jikaNo oshieru koto'ga deki'reba 1 i'idesU 5

ej158:

ha'i 1 soo 1 desU 5

ha'i 1 soo 1 desU 5

ej159:

sono jikaNde i'i desU 5

sono jikaNde i'idesU 5

ej160:

sa'Njookara 1 e'Nryakujimadewa 1 donokuraino jikaNde'sUka 5

sa'Njookara 1 e'Nryakujimadewa 1 donokuraino jikaNde'sUka 5

ej161:

su'beteno shItsumoNni kota'e 2 mashIta 5

su'beteno shItsumoNni kotaema'shIta 5