日本語Dictation Systemのための統計的言語モデルに関する一考察

3R－7

西村雅史* 大嶋良明* 野崎広志**
日本アイ・ビー・エム 東京基礎研究所* 大和研究所**

1．はじめに

近年，欧米では単語のN-Gramのような統計的言語モデル[1]を利用した口述筆記(Dictation)システムが，まだ離散単語発声ではあるが，実用化されはじめている．一方，日本語に関しては，語順に関する制約が弱いという知見から，N-Gramモデルの有効性について疑問視されていることや，単語の概念が明確でないため，離散発声単位として適したものがないなどの理由で，欧米と同様の構成のシステムはあまり研究されていない．

日本語においても，潜在意識的ではあるが意味のある最小の単位としての単語が存在する[2]．ただ，機械による処理を前提とする場合に用いる文法は，これとはまったく異なるものを単語として扱うのが一般的である．

今回，実際に人間の振る舞いを観察することで，この"潜在意識的な日本語の単語"を抽出した．また，機械的に自動抽出された形態素解析結果との対応関係を推定することによって，この単語単位を自動生成し，N-Gramモデルを構築した．他の単位とパープレキシティによる比較を行った結果は，日本語においても(大語彙，離散単語発声による)Dictationが実現可能であることを示唆している．むろん，この単位は連続発声による認識にも容易に適用出来る．

2．形態素表記のカバレージ

一般に，日本語は語彙数が多いと言われるが，語彙数とカバレージの関係についてはまだ十分な調査がなされているとは言えない．そこで，まず，一年分の新聞記事全紙面（日本経済新聞1992年12月-1993年11月までの約180万文）を文献[3]の形態素解析プログラムで解析し，このうち1993年11月の１ヶ月分をテストデータとして，目的とするカバレージを得るのに必用な語彙数を訓練テキストデータ量(1992年12月分から順次月単位で追加)毎に求めた．結果を図１に示す．この図から，９７%以下のカバレージなら，３ヶ月分程度のデータ量で必要語彙数はある程度飽和する傾向が読み取れる．今回，これ以降の実験では，全体の処理量を押さえるため，３ヶ月分のデータだけを用いることにする．

3．発声と言語モデルの単位

英語のDictationシステムでは言語モデルの単位(認識の最小単位)は単語であり，発声の単位は単語または単語の連鎖である．一方，日本語の言語モデルの単位としては，音素，音節，文字，形態素，単語，文節などが考えられるが，文節以上の単位では，語彙数が膨大になってしまうし，音素や音節だと，トライグラム程度の統計的言語モデルではあまり役に立たないであろう．

また，発声の単位としては，連続音声認識のアルゴリズムを用いることで言語単位の連鎖が受理可能になるが，逆に言語モデルの単位より短くてはいけない．我々は，自然言語処理の結果得られた形態素の表記をそのまま言語モデルの単位としてきたが，促音/っ/が１つの単位になっていたり，非常に長い複合語が含まれているなど，そのまま発声や認識の単位にするには無理がある．これは，形態素が，言語の解析作業によって得られた単位であり，人間の意識や，発声方法を考慮したものではないためである．

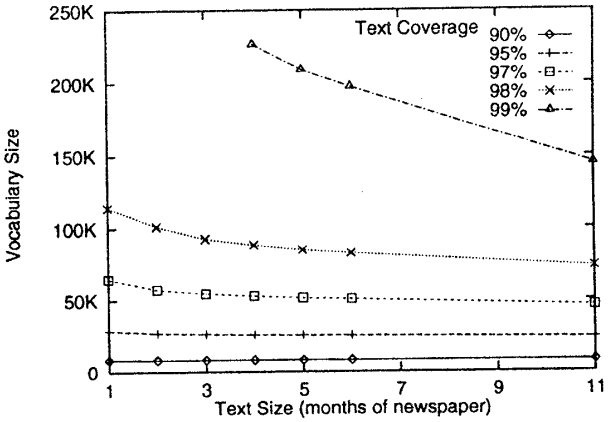

図１ 各カバレージにおける，訓練テキストデータ量と必要語彙数の関係（形態素表記）

日本語の単語については過去に多くの検討が文法学者によってなされているが，この内，文献[2]で説かれているように，[言語主体の意識によって自然に単位として認識されるもの]を単語として定義出来れば，これを発声や認識の基本単位とし，英語と同様の処理が可能になると期待される．つまり，離散発声によるDictationも可能になる．

3-1．単語と文節の切り出し実験

日本語では音節より長く，文より短い単位について明確な定義はない．どれを単語や文節と考えるかは個々の人に依存するので，その振る舞いを観察してモデル化するために，次のような実験を行った．

披験者は５名で，データは新聞から抜粋した１００文である．各披験者に対し次の２通りの作業を指示した．ただし，1. については，３文程度の切り出し例を提示してから作業を開始している．

1．意味のある最小単位に切る(以後単語と呼ぶ)
2．文節単位に切る．

なお，実際に発声を切り出すのは非常に手間がかかるので，対象文を表示して，それを読み上げつつ，ポーズの位置に，改行を挿入するという作業で模倣している．また，促音については発声を十分意識して切るように注意した．

一方，切り出し位置の一致度としては次の尺度を使う．

$$M = (n_c/n_1 + n_c/n_2)/2$$

ここで，n_c：共通に現れたセグメント数
n_1：比較対象1のセグメント数
n_2：比較対象2のセグメント数

結果を単語については表１に，文節については表２に示す．平均一致度は，それぞれ単語８６.3%，文節８６.5%で，ほ

A Study on Statistical Language Modeling for Japanese Dictation Systems
IBM Research, Tokyo Research Laboratory*, Yamato Laboratory**
Masafumi Nishimura*, Yoshiaki Ohshima*, Hiroshi Nozaki**

とんど差が無い．この値自体は，決して高いものではないが，単語単位も，従来比較的安定していると考えられていた，文節と同程度の安定性を持つことは興味深い．また，話者間で，一致度にそれほど大きなばらつきがないことから，人間が考えるところの単語を精度良く，自動抽出出来る可能性がある．

表1 単語セグメントの一致度(%)

	披験者B	披験者C	披験者D	披験者E
披験者A	86.7	89	86.4	86
披験者B	-	83.3	82.5	83.9
披験者C	-	-	88.7	86.3
披験者D	-	-	-	89.9

表2 文節セグメントの一致度(%)

	披験者B	披験者C	披験者D	披験者E
披験者A	90.4	87.3	89.2	84.9
披験者B	-	83.6	86.3	85.2
披験者C	-	-	88	85.2
披験者D	-	-	-	85.1

3-2．単語単位の自動生成

形態素解析結果[3]に対し，かな漢字変換のアルゴリズムを用いて複合語の分割処理を行なったのち，これと上記実験で観測された人間の定義した単語との対応関係をDPを使って自動推定した．このうち，比較的頻度の高い対応関係を４４の文法規則で表現することで，形態素連鎖から，単語連鎖を自動生成した．この時の平均一致度は８２.7%であった．また，各披験者について，接続助詞や助動詞の分離，接合傾向を反映させた１０程度の規則を追加適用すれば，平均一致度は９２%まで改善された．ただ，離散発声の認識を考えた場合には，この一致度ではまだ十分とは言えない．特に，３文字熟語(/四半期/，/預貸金/など)や，長い付属語連鎖を中心にして，揺らぎが残る．今後，個々の単語を調査し，複数の切り出しパターンを登録してゆく必要がある．また，言語モデルの訓練は，それらのセグメント可能性をすべて評価するように工夫する必要がある．

3-3．言語モデルのパープレキシティ

このようにして自動生成された単語（４４規則のみ適用）を用いた場合のテストセットパープレキシティを新聞約１週間分(約４万文)に対して求めた．文字および形態素と比較した結果を表３に示す．なお，いずれの場合も訓練テキストデータは新聞記事３ヶ月分で，語彙数は相対累積度数(カバレージ)９７%に設定してある．また，言語モデルはトライグラムモデルであり，テストデータとは異なる１週間文の新聞記事でheld-out補間処理[4]を行っている．

単語単位の場合のパープレキシティは１４０程度であり，特に，離散発声を想定する場合には，音響モデルに対する識別要求性能としては実用上問題のないレベルと言える．一方，語彙数は形態素の半分近くまで削減されており，言語モデル，音響モデル共に，必要とする資源は，形態素に比べかなり少なくて済む．

なお，各単位の長さが異なるため，単純にパープレキシティの比較は出来ないが，一文あたりのエントロピーがある程度目安になる．これを見る限りでは，形態素単位が最も優れているが，これは，形態素単位が多くの複合語を語彙として持つため，複合語に関する統計データの不足をカバー出来ているためではないかと考えている．

表3 各単位の統計量の比較

	文字	形態素	単語
のべ出現数	19,719,375	12,111,558	11,772,585
異なり表記数	1,028	40,136	21,626
一文あたりの平均単位数	44.6	27	26.2
Test-set Perplexity	41.9	108	143.2
一文あたりのEntropy	240.3	182.3	187.5

3-4．音響モデルに対する要求性能

日本語では多数の同表記異音語，あるいは同音異義語が存在するため，表記の言語モデルで推定されたパープレキシティがそのまま音響モデルの対象語彙数に相当するわけではない．表記と読みの間の関係をどのようなモデルで表現するかによるが，音響モデルの観点からは，読みの異なり語の数が認識の難しさの一つの目安になる．そこで，先の新聞記事３ヶ月分の訓練テキストデータ(単語単位)について，異なり読みの個数を調査した．また，出現頻度の影響を見るため，モノグラムのパープレキシティを表記のモデルと比較した．結果を表記との対比の形で表４に示す．同表記異音語に比べ，同音異義語の出現数が多く，結局読みの種類は表記よりも１割程度少ないことが分かる．このため，音響モデルに要求される性能は，表記のパープレキシティから推定されるものより，さらに低くてよいことになるが，一方で同音異義語の処理に関しては課題が残ると予想される．

表4 単語の表記数と読みの数の比較

	異なり数	Perplexity(1-Gram)
単語の表記	21,626	902.8
単語の読み	19,001	768.7

4．終わりに

日本人が潜在意識的に持っている単語単位に近いものを，形態素解析結果から自動推定出来る可能性を示し，その単語単位が，言語モデルの観点からも有効な単位であることを示した．また，この単位は，離散，連続いずれの認識システムにおいても基本認識単位となりうる．ただ，人間が考える単語単位をすべて模倣するにはまだ多くの処理が必要であり，今後も改良を続けたい．また，読みを含めた言語モデルの可能性についても検討を行う予定である．

謝辞

データ使用を許可していただいた日本経済新聞社に感謝する．また，形態素解析プログラムを提供していただいた，自然言語処理グループの方々，ならびに，セグメンテーション実験に関して有意義な助言をいただいたT.J.Watson Reseach CenterのS.Roukos氏に感謝する．

参考文献

[1]S.Furui,M.M.Sondhi編，「Advances in Speech Signal Processing」，第21章,Marcel Dekker,Inc(1992)

[2]時枝誠記，「日本文法 口語編」，岩波全書(1950)

[3]丸山宏，荻野紫穂，「正規文法に基づく日本語形態素解析」，情報処理学会論文誌，Vol35，No. 7，pp1293-1299(1995)

[4]中川聖一，「確率モデルによる音声認識」，電子情報通信学会(1988)