（ 続紙 1 ）

京都大学	博士（情報学）	氏名	Mijiti Abulimiti
論文題目	Automatic Speech Recognition of Agglutinative Language based on Lexicon Optimization （単語辞書の最適化に基づく膠着言語の自動音声認識）		

（論文内容の要旨）

In agglutinative languages, selection of lexical unit is not obvious and it is one of the important issues in designing a language model for automatic speech recognition (ASR). Choice of lexicon unit affects unit length, frequency and also ASR performance. Words in agglutinative languages have a variety of derivatives, and increase the vocabulary size explosively, causing OOV (out-of-vocabulary) and data sparseness problems. Therefore, the morpheme unit is conventionally adopted in many inflectional languages, such as Japanese, Korean, Turkish, Finnish, German and Arabic. However, morphemes are short, often consisting of one or two phonemes, thus they are more likely to be confused in ASR than the word unit. The goal of this study is to design an optimal lexicon by balancing the characteristics of the word unit and the morpheme unit.

As opposed to the previous work, we investigate approaches that are directly related to ASR performance or WER (word error rate), considering phonetic confusability and unit length. A novel discriminative approach is proposed to select word or sub-word entries which are likely to reduce the WER. We compare segmentation and concatenation approaches for lexicon optimization in our Uyghur large-vocabulary continuous speech recognition (LVCSR) system. The proposed concatenation method based on the proposed discriminative approach significantly reduces WER and the lexicon size compared to the word-based model. Furthermore, the method preserves linguistic information including both word and morpheme boundaries, which is useful for many natural language applications such as machine translation and information retrieval.

Chapter 1 provides an overview of the problems in lexicon design in agglutinative languages and approaches investigated in the thesis.

Chapter 2 briefly reviews previous studies on lexicon optimization. They are based on statistical measures, such as co-occurrence frequency, mutual information and language model likelihood. Unsupervised morpheme extraction methods are also reviewed.

Chapter 3 presents our morphological analyzer based on rules and statistical models. The morpheme segmentation accuracy of 97.6% is achieved and a baseline ASR system is built based on the morpheme and word units. Statistical characteristics and baseline ASR performance of various unit sets are compared.

Based on the baseline ASR results, Chapter 4 presents an approach which extracts problematic morpheme sequences. It is realized by aligning and comparing the ASR results by the morpheme-based model with those by the word-based model. The manually selected features reduce both WER and the lexicon size while maintaining the high coverage of the morpheme unit. The combined effect of these features is also confirmed.

In Chapter 5, we present a discriminative approach to automate the feature extraction from the two layers of ASR results. We describe each word by a set of features and define an evaluation function. Then, the weights of the features are learned by the difference of the two ASR results to select "critical" word entries which generate different (probably correct) hypotheses from the morpheme-based unit. This learning mechanism is applicable to any unseen words or sub-words. Specifically, we apply the Support Vector Machines (SVM) and Logistic Regression (LR) model as well as simple perceptron. The SVM and LR are more robustly trained and SVM results in the best performance, significantly reducing both WER and the lexicon size. The proposed learning scheme is realized in an unsupervised manner in that it does not need correct transcription for training data.

Chapter 6 compares various sub-word optimization approaches in the Uyghur LVCSR system. Both segmentation and concatenation methods with manual and automatic approaches are compared. Although the best performances of both methods are comparable, the proposed morpheme concatenation approach is advantageous in that it preserves linguistic information.

Chapter 7 concludes the thesis with discussions of the generality of the proposed approach in other agglutinative languages.

（続紙 ２ ）

（論文審査の結果の要旨）

本論文は、多様な接頭辞・接尾辞によって単語が構成される膠着言語、特にウイグル語の自動音声認識を指向して、機械学習に基づいて最適な単語辞書を構成するための研究をまとめたものであり、得られた主な成果は次の通りである。

1. 統計的モデルに基づいてウイグル語の形態素解析システムを構成し、形態素カバー率及び形態素区分化精度ともに約98%を実現した。これに基づいて、ウイグル語の大語彙連続音声認識システムを初めて構築し、形態素単位と単語単位の各ベースラインモデルについて、カバレージと認識精度に関する利点・欠点を定量的に示した。

2. 高いカバレージを持つ形態素の単位をベースとしながら、音声認識誤りの頻度が高い系列については単語の単位を取り入れる枠組みを系統的に行うために、単語単位と形態素単位の音声認識結果の対応付け（アライメント）に基づいて、特徴抽出及び単語エントリの追加を行う方法を提案した。これによりベースラインモデルから、辞書サイズを大きく増やすことなく、認識精度を改善できることを示した。

3. 上記の過程を機械学習で行う方法を定式化した。音声認識誤りを基準として、各特徴の重みをパーセプトロンやSVMなどで識別学習することにより、最適化が行われ、かつ部分単語系列への汎化適用も可能となった。この手法により、認識精度の有意な改善を実現した。また、他の統計的な手法とも比較して、認識精度及び辞書サイズの両面で提案手法の有効性を確認した。

以上のように本論文は、単語辞書の構成が自明でない膠着言語に対して、音声認識の観点から最適化を行う方法を提示するもので、学術上・実用上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成 25 年 1 月 23 日に論文とそれに関連した内容に関する口頭試問を行った結果、合格と認めた。