（様式６号）「課程博士用」

学 位 論 文 の 要 旨

専　攻　名	システム工学　　専　攻	氏　名（ふりがな）	松下　真也（まつした　しんや）　㊞

学位論文題目

少数言語における保存活動の支援に関する研究

（Study of support to preservation activities for minority languages）

世界には約 7,000 言語が現存し，いくつかは消滅の危機に瀕している．特に，そのような言語は少数言語と呼ばれ，言語学者らは保護活動に努めている．多くの保護活動は手作業により行われており，その支援が必要とされている．本研究では，このような保護活動のなかで，録音した音声を収集しその意味を取ったうえで保存する保存活動に着目し，その支援をした．その中でも「音声の書き起こし・書き起こしテキストの解析」という作業工程に着目した．いずれの作業も，精度を高めるためには，対象言語固有の事前知識や，機械学習をするための多量のデータが必要である．しかし，保存活動ではいずれも乏しい状況であり，既存の手法の適用は困難である．そのことをふまえつつ，これらの作業の一部を自動化することで，作業の支援を試みた．

「音声の書き起こし」では，音声データを，その発音を表すテキストに変換する．これを実現する技術として，音声認識や音声ラベリングが挙げられる．しかし，これらの技術は，保存活動で必要とされている以上の解析をする．本研究では，事前知識・データ量が不十分である状況をふまえて，「音声の書き起こし」を，特定の音声を連続音声から探し出すようなタスク(音声マッチング)と単純化してとらえ，この状況下での支援方法を検討した．

「書き起こしテキストの解析」では，与えられたテキストから最終的にはその意味をとらえる．一般的にこの解析は，分かち書きや形態素解析，構文解析，意味解析などを組み合わせて行う．本研究では，これらの解析のうち，最も基本的な分かち書きに焦点を当て支援する．特に，事前知識・データ量が不十分である状況をふまえて，教師なし分かち書き手法について検討した．

以上の結果として，今回対象とした作業の効率化・高精度化が実現した．これは，残された作業はあるものの，少数言語の保存活動の一助となるだろう．

本論文は以下 6 章で構成される．

1 章では，本研究の背景について述べ，音声処理およびテキスト処理に関する用語を確認するとともに，保護活動の意義・活動内容について説明し，支援対象の活動内容・支援内容を示す．そして，支援内容に関連する従来研究を示しながら，本研究の目的を述べる．

2 章では，少数言語音声の書き起こしを支援するため，画像処理手法を利用した音声マッチングの自動化手法を提案した．一般的な音声マッチングでは，周波数を特徴量に変換して用いる．この変換においては，フーリエ変換を基にした手法が主流であり，変換対象に対する時間的制約がかかる．特に，この制約を許容するためには多量のデータが必要となるため，一般的な変換方法では少数言語データに適さない．また，保存活動においては音素が完全に特定できていないため，任意の特徴量と一致または類似という緩やかなマッチングが必要である．提案法では，音声マッチングで用いる周波数の特徴を画像として表現し，画像処理におけるマッチング手法を適用することで，保存活

動のための緩やかなマッチングを検討した．その結果，画像処理手法を利用することで，緩やかなマッチング処理が可能となった．

3 章では，少数言語のテキストの分かち書き手法を検討した．具体的には，事前知識なしでさまざまな言語の分かち書きを行う手法の一つである Nested Pitman-Yor Language Model (NPYLM)に着目し，不十分なデータ量での挙動を分析し，その結果に基づいて，少数言語の単語分割における NPYLM の適用上の問題点を明らかにした．具体的には，NPYLM は統計的手法であるため，学習テキストが不足する場合は分割精度が低くなる傾向にあることを示した．特に，分割精度の低下は，過剰な分割を主に引き起こすことを示した．

4 章では，3 章で述べた過剰分割の改善を目的として，「NPYLM の 2 段階適用」を提案した．分割の必要がない文字列を一文字に置き換えることで，データ量が不十分なときの NPYLM における挙動(過剰分割)を抑制する．また，2 回 NPYLM を適用することで，その 1 回目で置き換える文字列の検出をし，2 回目で分かち書きを行う手法を提案した．これにより，言語に依らず，テキストの不足に起因する過剰分割を改善した．

5 章では，4 章で提案した手法を改善した．4 章の手法は，過剰分割を改善したが，極端な分割不足を引き起こすことがあったので，これを改善する．そこで，4 章の手法における置き換えにおいて，選択的な置き換えを提案した．これは，置き換えにおいて，置き換えるべき文字列と，置き換えるべきでない文字列を分別することで，分割不足を引き起こす置き換えを行わないようにするものである．分析の結果，置き換えるべき・できなでないは，置き換え候補の文字列の長さにより判断することができることを明らかにした．所定の長さ以下の文字列のみを置き換えることで，過剰分割の改善と分割不足の抑制を両立することができた．

6 章では全体のまとめを述べ，今後について示す．

続紙　有□　無□

（様式６号－続紙）「課程博士用」

ふりがな 氏　名	まつした　しんや 松下　真也　　㊞