論文内容の要旨

博士論文題目　Construction and Analysis of Multiword Expression-Aware Dependency Corpus
(複単語表現を考慮した依存構造コーパスの構築と解析)

氏名　加藤明彦

(論文内容の要旨)

自然言語表現における複単語表現（Multiword Expression, MWE）とは，統語的または意味的にまとまった言語表現であり，それを構成する個々の単語の品詞や意味から，その全体的な文法役割や意味が予測できないという非構成性を有する複数の単語からなるまとまりである．統語的な依存構造の情報を利用し，かつ意味理解が必要なタスクでは，単語をベースとした依存構造よりも，MWE を考慮した依存構造，すなわち MWE を統語的な単位とする依存構造の方が好ましい．英語の依存構造コーパスは句構造コーパスからの自動変換によって構築されることが多いが，ほとんどの句構造コーパスでは，MWE が句構造木内の独立な部分木になっていることは保証されていないため，MWE を考慮した依存構造に容易に変換することはできない．

そこで本研究では，MWE が句構造木の部分木になるように木を修正する手続きを定式化し，複合機能語，形容詞 MWE，および，固有表現を考慮した依存構造コーパスを Ontonotes 上に構築した．また意味理解が必要なタスクでは，複合機能語や形容詞 MWE のように，固定された単語列という形で使用される MWE（連続 MWE）だけでなく，句動詞のように他の単語や句を内部に含むような非連続な出現を持ちうる MWE（動詞 MWE）の認識も重要であるため，クラウドソーシングを用いて Ontonotes コーパスに対して動詞 MWE のアノテーションを行なった．最後に，上記コーパスを利用し，連続 MWE を考慮した依存構造と動詞 MWE の双方を予測する問題に取り組んだ．これは，依存構造の情報は動詞 MWE 認識で有効な特徴量として働くという直感に基づいている．評価実験の結果，連続 MWE 認識，連続 MWE の範囲を依存関係ラベルとして符号化した依存構造解析，動詞 MWE 認識の階層的マルチタスク学習に基づくモデルの有効性が確認された．この結果は，連続 MWE を考慮した依存構造解析器が捉えた特徴量が，動詞 MWE 認識で有効であることを示唆している．

氏　名	加藤明彦

（論文審査結果の要旨）

令和元年10月21日に開催した公聴会の結果を参考に令和元年11月20日に本博士論文の審査を行った．以下のとおり，本博士論文は，提案者が独立した研究者として，研究活動を続けていくための十分な素養を備えていることを示すものと認める．

加藤明彦は，本博士論文において，英語の複単語表現に関する研究に取り組み，標準的なタグ付きコーパスとして利用されているOntonotesコーパスの句構造アノテーションを利用して，複合機能表現，形容詞MWE，固有表現，句動詞などのMWEを依存構造木として半自動的にアノテーションする方法を提案した．さらに，構築したコーパスを用いて，MWE認識を伴う依存構造解析法を提案し，その評価実験を行った．

本論文の貢献は以下のようにまとめることができる．

1. 複単語表現のうち，複合機能表現，形容詞MWE，固有表現などを，Ontonotesコーパスの句構造木アノテーションを利用して，依存構造木コーパスとして半自動的にアノテーションコーパスを構築する方法を提案した．

2. 構築した依存構造木コーパスをLDC(Linguistic Data Consorsium)を通じて公開した．

3. 句動詞の一部には，目的語となる名詞句を表現内に取り込むため，構成語が必ずしも連続して出現しないものが存在する．句動詞には，字義通りの意味で用いられるものと複単語表現として用いられるものが存在するため，その区別を行う必要がある．句動詞の各利用事例を正しく分類するため，クラウドソーシングを利用して，アノテーションを行い，その結果を公開した．

4. 複単語表現の認識が依存構造解析にどのように影響を与えるかを調査するため，複単語表現の認識と依存構造解析を同時に行う手法をいくつか提案し，その有効性について実験によって検証を行った．

複単語表現の情報をコーパス中にアノテーションする方法を提案し，かつ，広く利用可能なデータとして公開したこと，また，複単語表現の認識を行うことにより，より整合性が高く，解析精度の高い依存構造解析法を提案した本研究は，独創性が高く，しかも実用的であり，自然言語処理の分野において高い貢献があると評価する．

よって，本論文は，博士（工学）の学位論文として価値あるものと認める．