

研究種目：基盤研究（C）

研究期間：2006～2008

課題番号：18500102

研究課題名（和文） 動的シソーラスを用いた特許文の解析システム

研究課題名（英文） Patent Analysis System using Dynamic Thesaurus

研究代表者

横山 晶一（YOKOYAMA SHOICHI）

山形大学・大学院理工学研究科・教授

研究者番号：20250946

研究成果の概要：長くて複雑な構造を持つ特許文に対して、その並列性や連体修飾の係り受けに注目し、ある程度の構造を解明することができた。その際に、動的シソーラスに当たる格フレームを用いて、その性質が解明された。また、英語の特許文と対照することで、訳し分けの問題にも寄与した。さらに、構文解析器（パーサ）を自動適用することで、特許文に対応したパーサの作成にもある程度成功した。

交付額

（金額単位：円）

	直接経費	間接経費	合計
2006年度	1,400,000	0	1,400,000
2007年度	900,000	270,000	1,170,000
2008年度	800,000	240,000	1,040,000
年度			
年度			
総計	3,100,000	510,000	3,610,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：①自然言語処理 ②シソーラス ③特許文 ④人工知能 ⑤計算機システム

1. 研究開始当初の背景

(1) 特許文の要約中の「課題」、「解決手段」、「請求範囲」といった部分は、長い一文で書かれることが多く、機械翻訳や要約、キーワード抽出などの自然言語処理における構文、意味解析の支障となることが指摘されてきた。長い特許文の並列構造や係り受け構造をどのように解析するかは、開始当初は全く不明の状態であった。

(2) 意味解析には、シソーラスなどの階層辞書が有用であることはすでに分かっていたが、そのようなシソーラスが特許文に対してはどのように有効に働くのかは分かっていた。そこで、シソーラスを作成して、

その有効性を確かめる必要性があった。

2. 研究の目的

(1) 複雑な特許文を解析するために、並列構造がどのようになっているかを解明する。また、連体修飾節や、係り受けの構造を解明し、それらを扱うシステムを構築する。

(2) 意味解析のためのシソーラスを作成して、それを特許文に適用し、上記の構造の解明に寄与するかどうかを確かめた上で、それをシステムに組み込んで有効性を検証する。

3. 研究の方法

(1) 特許文の並列構造の解明のために、まず

既存の形態素解析システムである茶釜の出力を手作業で分類した。それに基づいて、誤りを自動修正するシステムを作成し、その過程で、並列構造を解明する手がかりを得る。

また、法律文の並列構造と特許文の並列構造を比較して、その違いを解明する。

さらに、連体修飾節の範囲を自動的に定められるかどうかを検討し、その結果をシステムに組み込む。

(2) 既存のシソーラスを用いてどの程度並列構造や係り受け構造が解析できるかを確かめる。その後、格フレームといった、動的シソーラスに類似する構造を導入して、英訳結果と比較し、どの程度格フレームが英訳の訳し分けに役立っているかを解明する。

4. 研究成果

(1) ①特許文特有のパターンである「本発明は～である」という定型文と、「A と B との C が」という形で表される並列構造文を見つけて、係り受けの誤りを自動修正するシステムを構築し、この種の誤りが自動修正できることを確認した。

②特許文を自動的に分割するために、連体修飾節の範囲を定めることを自動的に行った。シソーラスを用いることによってある程度の成果が得られた。

③法令文における並列接続詞の上下関係、すなわち、「または」と「もしくは」では、「または」の方が優先するという関係が特許文にも成り立つかどうかを調査し、特許文ではほとんどの場合に同等の並列接続詞として扱われていることを確かめた上で、それに基づいて、別の形態素解析システムである Juman に基づいた誤り修正システムを構築した。その結果、約 56%について修正することができた。

④構文解析器（パーサ）の中に、学習機能を持ったものがある。これを用いて、特許文を学習させ、特許文に適合したパーサを構築することができるかどうかを実験した。その結果、ある程度のパーサ適応を実現できることが解明された。

(2) ①シソーラスに類似した辞書を用いる結合価（用言がどのように他の要素と結合するかを解明した辞書）を用いて、特許文の連体修飾構造を分析した。その結果、結合価を使用すると、連体修飾節の範囲をある程度定めることができることが分かった。

②動的シソーラスの一種である格フレーム辞書を用いて、特許文の英訳の際に、格フレ

ームの区別が訳し分けに反映されているかどうかを調査した。その結果、サ変動詞の一部に、格フレームの区別が英訳に反映されているものがあることが判明した。これについては、なお、細かな分析が必要である。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 6 件）

①横山晶一：特許文への構文解析適応、平成 20 年度 AAMT/Japio 特許翻訳研究会報告書、pp.65-70 (2009) 無

②横山晶一：特許文における接続詞と係り受けの構造、Japio 2008 Year Book, pp.68-73 (2008) 無

③横山晶一、橋本力、小野裕太、戸坂圭介：並列接続詞を含む特許文の係り受け修正システムと、特許文の連体修飾節の構造解析、平成 19 年度 AAMT/Japio 特許翻訳研究会報告書、pp.68-76 (2008) 無

④横山晶一：特許文解析誤り自動修正システムと、正確な翻訳のための特許文の分割、Japio 2007 Year Book, pp.228-233 (2007) 無

⑤横山晶一、見年代茂大、吉田節行：特許文解析誤り自動修正と、正確な翻訳のための特許文の分割、平成 18 年度 AAMT/Japio 特許翻訳研究会報告書、pp.73-83 (2007) 無

⑥横山晶一：特許文解析誤りの分類と自動修正の可能性、Japio 2006 Year Book, pp.188-191 (2006) 無

〔学会発表〕（計 10 件）

①吉田節行、横山晶一：パーサ適応と格フレームによる特許文解析精度の向上、情報処理学会第 71 回全国大会、2009.3.10、立命館大学びわこくさつキャンパス

②奥山真澄、横山晶一：格フレームを用いた特許文の訳し分け、情報処理学会東北支部平成 20 年度研究会、2009.3.9、山形大学工学部

③ C.Hashimoto, D.Kawahara, T.Yoshida, H.Goto, S.Yokoyama: Parser Adaptation to Patent Domain Exploiting Case Frames, ISUC2008, 2008.12.16, 大阪国際会議場

④橋本力、河原大輔、吉田節行、後藤広樹、横山晶一：特許文書の構文解析、第 51 回自動制御連合講演会、2008.11.22、山形大学工

学部

⑤橋本力、河原大輔、吉田節行、後藤広樹、横山晶一：日本語特許文書へのパーサー適応、NLP若手の会第3回シンポジウム、2008.9.23, 熱海金城館

⑥横山晶一、小野裕太、橋本力：並列接続詞を含む特許文の係り受け修正システム、言語処理学会第14回年次大会、2008.3.19, 東京大学駒場キャンパス

⑦小野裕太、横山晶一、橋本力：特許文の接続詞係り受け修正システム、情報処理学会東北支部平成19年度研究会、2008.3.11, 山形大学工学部

⑧ S.Yokoyama, S.Kennendai: Error Correcting System for Analysis of Japanese Patent Sentences, Machine Translation Summit XI, Workshop on Patent Translation, 2007.9.11, Copenhagen Business School

⑨吉田節行、横山晶一：特許文の機械翻訳における正しい係り受け判定のための文章分類、情報処理学会東北支部平成18年度研究会、2007.3.9, 山形大学工学部

⑩見年代茂大、横山晶一：特許文解析誤りの修正システム、情報処理学会第69回全国大会、2007.3.8, 早稲田大学大久保キャンパス

〔図書〕(計 0件)

〔産業財産権〕

○出願状況(計 0件)

○取得状況(計0件)

〔その他〕

6. 研究組織

(1) 研究代表者

横山 晶一 (YOKOYAMA SHOICHI)
山形大学・大学院理工学研究科・教授
研究者番号：20250946

(2) 研究分担者

なし

(3) 連携研究者

なし