

## 二国間交流事業 共同研究報告書

平成 23年 4月 1日

独立行政法人日本学術振興会理事長 殿

共同研究代表者所属・部局 国立情報学研究所・コンテンツ科学研究系

職・氏名 (ふりがな) 准教授・宮尾祐介 みやお ゆうすけ

1. 事業名 相手国(ドイツ)との共同研究 振興会対応機関 ( DFG )
2. 研究課題名 多言語における頑健な文法の開発および意味記述の深化

3. 全採用期間

平成 21 年 4 月 1 日 ~ 平成 23 年 3 月 31 日 ( 2 年 0 ヶ月 )

4. 研究経費総額

(1) 本事業により交付された研究経費総額 5,000 千円

初年度経費 2,500 千円、 2年度経費 2,500 千円、 3年度経費            千円

(2) 本事業による経費以外の国内研究経費総額 0 千円

## 5. 研究組織

### (1) 日本側参加者

氏名 <small>(ふりがな)</small>	所属・職名	研究協力テーマ
みやおゆうすけ 宮尾祐介	国立情報学研究所・准教授	文法開発手法の研究
まつぎたくや 松崎拓也	東京大学・助教	文法開発手法，構文解析技術の研究
わんしゃんりー 王向莉	東京大学・特任研究員	中国語文法の設計・開発
はなおかひろき 花岡洋輝	東京大学・大学院生	日本語文法の設計・開発
ごらん とびち Goran Topic	東京大学・大学院生	クロアチア文法の設計・開発
いわさしゆんや 岩澤俊弥	東京大学・大学院生	中国語文法の設計，ツリーバンク開発手法の研究

### (2) 相手国側研究代表者

所属・職名・氏名 Saarland University・Senior Lecturer・Valia Kordoni

### (3) 相手国参加者（代表者の氏名の前に○印を付すこと）

氏名	所属・職名（国名）	研究協力テーマ
○Valia Kordoni	Saarland University・Senior Lecturer（ドイツ）	文法開発手法，意味表現の研究
Yi Zhang	Saarland University・Research Post-Doc（ドイツ）	文法開発手法の研究，中国語の文法開発
Feiyu Xu	DFKI・Senior Researcher（ドイツ）	情報抽出，機械翻訳の研究
Hans Uszkoreit	Saarland University・Professor（ドイツ）	情報抽出，機械翻訳の研究

## 6. 研究概要（研究の目的・内容・成果等の概要を簡潔に記載してください。）

本交流事業は、複数言語において実用的な構文解析器を構築することを目指し、ドイツ・ザールランド大学およびDFKIとの共同研究により、頑健な文法の開発および意味記述の設計について研究・情報交換を行うことを目的とする。具体的には、日本語、中国語、ドイツ語などの言語を対象とし、それぞれの言語において頑健な（実世界の文を高精度・高被覆で解析できる）文法を開発するため、文法開発手法や文法理論について情報交換・技術移転を行う。また、機械翻訳や情報抽出などのアプリケーションを想定し、これらの言語における意味記述について検討を行う。これにより、頑健かつ高精度な構文解析器を複数言語において実現することを目指す。

本研究における文法開発手法は、研究代表者がこれまでに英語文法の開発で実践してきたコーパス指向の文法開発手法をベースとする。コーパス指向の文法開発とは、文法を直接人手で開発するのではなく、既存の構文木付きコーパス（ツリーバンク）を利用し、そこから目的とする文法を獲得する。一般に、既存のツリーバンクは表層的な構文構造しか与えないため、そのままでは意味構造を計算できる文法は獲得できない。しかし、そこを出発点としてより深い構造付きのコーパスを作成することができる。具体的には、表層的構造を手掛かりとして、文法理論に基づく構造（深層構文構造や意味構造など）、および文法理論から要請される文法的制約を追加することで、文法理論に基づく深い構造が付与されたツリーバンクを得ることができる。深い構造付きのツリーバンクが得られたら、そこから文法（文法規則と辞書）を自動獲得することができる。本手法では既存の大規模ツリーバンクを再利用することができるため、意味構造を計算できかつ頑健な構文解析器を低コストで構築することができる。

ドイツ語、日本語、中国語などでは既にいくつかの大規模ツリーバンクが存在する（TIGER Treebank, 京都テキストコーパス, Penn Chinese Treebank など）。これらのツリーバンクはすべて表層的構文構造のコーパスであるが、英語文法の開発で用いた文法開発手法をこれらの言語に応用することで、頑健かつ実用的な構文解析器を低コストで実現できると期待される。ただし、本研究グループは英語以外の文法理論・文法開発について知見が少ない。したがって、ドイツ語、中国語、日本語、ギリシャ語の文法の開発を経験しているザールランド大学とDFKIとの共同研究によりこの点を克服する。

また、ザールランド大学およびDFKIで開発されている文法は、Minimal Recursion Semantics という意味記述の枠組みを利用している。この枠組みは言語非依存な記述形式（言語独立な抽象的意味表現という意味ではなく、意味を記述するためのデータ構造が言語非依存であるということ）であり、機械翻訳や複数言語における情報抽出などへの可搬性が高い。この枠組みについて情報交換を行うことにより、本研究グループで開発している日本語や中国語などの文法の意味記述の設計を行う。

本事業により、2009年10月および2010年8月～9月に本研究グループがドイツ・ザールランド大学・DFKIを訪問し、また相手研究グループが2010年3月および2011年2月に東京大学を訪問し、情報交換や技術移転作業を行った。これらの研究交流により、中国語と日本語を中心として文法理論の基本設計を確立し、現在文法開発を進めている。中国語については高被覆な文法の開発が完了し、既に構文解析器が動作している段階である。また、中国語文法は相手研究グループでも開発を行っているため、文法リソースや例文データなどを共有し、比較検討を行うことができるように整備している。日本語についてはここで検討した理論に基づき現在文法開発作業を進めている。また、構文解析技術についても情報交換を行い、構文解析の高速化・高精度化技術である supertagging について、他研究グループが研究に参画できるよう共同コンペティションを企画することとなった。他の文法枠組み（CCG, LFG など）の研究グループも含めて議論を進めており、近いうちに共同コンペティションを提案する予定である。