

## まとめと将来の展望

プラシャント・パルデシ

国立国語研究所

prashant@ninjal.ac.jp

### 1 まとめ

このワークショップでは、日本語初の統語・意味解析情報付きコーパスのアノテーション法や意義について議論を行った。本コーパスプロジェクトの最大の課題は、これまでに蓄積されてきた文法研究の成果をどのようにしてコーパスとして実現可能な形で盛り込むかということである。日本語研究者の諸氏には、それぞれの観点からのアドバイスを是非いただきたい。

本コーパスは日本語文法研究者や言語処理研究開発に取り組む人々の他に、日本語教育に携わる人々によっても使われることが想定される。これら広範囲の人々には本コーパスを使った上で、私たちコーパス開発者へのフィードバックを求めたい。

### 2 将来の展望

本プロジェクトではこれまでに、主として単語分割、形態素情報および統語解析情報のアノテーションを行ってきた。コーパスが完成すれば、それだけでも日本語研究に大きく貢献することは確実である。しかし、この狭義のコーパスから直接得られる情報に満足しては、それが持っているはるかに大きな可能性の扉を閉ざしてしまうことになる。

「アノテーション方式とコーパスの特徴」において、従来の解析木のカッコ表示に代わる XML によるエンコーディングについて解説した。これによってアノテーションをより豊かにし、コーパスを拡張していくことが可能になる。将来において計画しているのは、以下の諸点に関わるコーパスの拡張である。

- UniDic (Den et al. 2008) にもとづくレンマの付加。これにより、活用形や表記の変異に関わらず同じ語を検索することが可能になる。
- ローマ字表記を付加して、外国人研究者の便宜に供する。
- 英訳の付加。これは外国人研究者の他に、日本語教育での利用も意図している。
- PropBank (Bonial et al. 2010) 方式の格フレーム情報を付加する。これにより語彙意味論的リソースへのリンクを付ける。例えば、機械翻訳や意味抽出における単語の意味の曖昧性の解消に役立つ。
- アーギュメントや副詞にインデックス情報を付加する。複雑な構文の研究、例えば南 (1974) の文階層の検証に貢献する。
- 否定、モーダル要素および量化表現におけるスコープ情報を付加する。

## 参考文献

- Bonial, Claire, Olga Babko-Malaya, Jinho D. Choi, Jena Hwang and Martha Palmer. PropBank Annotation Guidelines, edition 3.0. Center for Computational Language and Education Research, Institute of Cognitive Science, University of Colorado at Boulder.
- Den, Yasuharu, Junpei Nakamura, Toshinobu Ogiso and Hideki Ogura. (2008) A proper approach to Japanese morphological analysis: Dictionary, model, and evaluation. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, pages 1019–1024. Marrakech, Morocco: European Language Resources Association (ELRA).
- 南不二男 (1974) 『現代日本語の構造』 大修館書店。