

「統語・意味解析コーパスの開発と言語研究」研究発表会

プロジェクト名・リーダー名

統語・意味解析コーパスの開発と言語研究

プラシヤント・パルデシ (国立国語研究所 理論・対照研究領域 教授)

開催期日

平成29年3月4日 (土) 10:00～18:00

開催場所

東北大学 川内北キャンパス「川内合同研究棟」5階、531-532号室

キャンバスマップ (A07の建物)

↑リンク <http://www.tohoku.ac.jp/japanese/profile/campus/01/kawauchi/areaa.html>

※どなたでも参加可能ですが、参加人数を確認するため、事前に [kasai.yoko\[at\]ninjal.ac.jp](mailto:kasai.yoko[at]ninjal.ac.jp) 宛にお申し込みください。

●「統語・意味解析コーパスの開発と言語研究」I

発表者:プラシヤント・パルデシ(国立国語研究所)

題目:「統語・意味解析コーパスの開発と言語研究」

要旨:「統語・意味解析コーパスの開発と言語研究」プロジェクトでは言語研究・言語処理に資する統語・意味解析コーパスを開発・公開し、それに基づく言語研究を行うことを目標としている。本発表ではプロジェクトの今までの活動を振り返り、見えてきた課題や将来の展望について報告を行う。

発表者:岸本秀樹(神戸大学)

題目:「教材開発と構造体コーパス:現状と課題」

要旨:本発表では、統語・意味解析コーパスプロジェクトに関する神戸大学ユニットでの作業の目標について概観した上で、日本語統語論のテキストの教材(書籍出版)および構造体コーパスのウェブページで公開予定の補助教材(練習問題等)の開発状況をレポートする。統語論の教材の学習から構造体コーパスの利用に向かうために生じる具体的な課題についても検討する。

発表者:Yoshimoto, Kei (東北大学)

題目:“How to annotate what”

要旨:Now that the foundations of NPCMJ have been laid, we confront many difficulties in

improving our corpus by subdividing tags. The finer the semantic distinctions that we make, the more ambiguity and vagueness that the annotators must face. In this talk, I propose a way to tell what we should annotate from what we shouldn't. Furthermore, I attempt to offer a set of tests that help the annotators choose one tag rather than another.

発表者: Butler, Alastair (国立国語研究所)

題目: “The Keyaki Treebank and the NPCMJ: Bridging a growing divide”

要旨: The Keyaki Treebank has a simple format for encoding structural and semantic information with bracketed trees. The NINJAL Parsed Corpus of Modern Japanese (NPCMJ) has started as a (reversible) repackaging of the Keyaki Treebank content into XML markup. As the NPCMJ grows in the richness of the information offered, the Keyaki Treebank itself can continue to be enlarged and refined in its native bracketed format and yet preserve its role as the pivotal component for a widening NPCMJ enterprise.

発表者: Horn, Stephen Wright

題目: “Grammatical principles for annotation and query”

要旨: This presentation explores how a recognition of basic grammatical principles and processes can be sufficient for assigning structures to linguistic expressions in the Keyaki Treebank format. Starting from the notion of the projection of phrases, this talk connects intuitive notions like headedness, predication, selection, modification, co-reference, focus, etc., with concrete structural relations exemplified in the Treebank. It is hoped this approach will be useful for both building and querying the Treebank.

発表者: 堀田智子 (東北大学)

題目: 「日本語学習者の話し言葉データに見られる中間言語の諸相」

概要: 本発表では、まず NINJAL に提供した話し言葉データの概要について報告する。特に、調査目的、調査方法、発話内容、話者情報(日本語母語話者と日本語学習者)などについて説明する。また本データを基に発表者が、第二言語習得、語用論、認知言語学のアプローチから行ってきたこれまでの研究成果について報告する。さらに、近日公開予定の「日本語学習者対照コーパス」の概要を紹介する。

発表者: 檜山祥太(東北大学) 佐藤亮輔(東北大学) 周 振(東北大学)

題目: 「アノテーション研究 (東北チーム)」

発表者: 金城由美子 (NINJAL)

題目: 「言語資料としての国会会議録」

発表者:岸山健(東京大学)、折笠誠(上智大学)、金城由美子(NINJAL)

題目:「アノテーション研究(NINJAL チーム)」

発表者:Butler, Alastair (国立国語研究所)

題目:“Introducing Techniques for better Tregex searching”