

Luncheon Linguistics, 24 April, 2019

2019 (平成 31) 年 4 月 24 日

「マレー語・インドネシア語コンコーダンサーMALINDO Conc の開発について」

発表者：野元裕樹 (東京外国語大学大学院総合国際学研究院准教授)

MALINDO Conc は、マレー語・インドネシア語のコーパス検索のために構築されたオンライン・コンコーダンサーである。MALINDO Conc では、現在、以下の 3 種類のコーパスが検索可能である。

1. ライプツィヒコーパスコレクション (Goldhahn et al. 2012) のマレー語・インドネシア語部分の言語判定をやり直したの再編版 (Nomoto et al. 2018a)
2. インドネシア語カエルの話 (Moeljadi 2014)
3. マレー語変種コーパス (標準口語、サバ方言) (Nomoto 2018; Kartini & Nomoto 2018)

これらのコーパスにはすべて語の形態情報 (接頭辞、接尾辞、周接辞、重複の種類) のアノテーションが付与されている。アノテーションを可能にするために、事前にマレー語・インドネシア語の形態情報辞書 MALINDO Morph (Nomoto et al. 2018b) を構築した。MALINDO Conc では、コーパスに付与したアノテーションを利用し、「ingin (欲する) + 受動標識 di-を含む語」のように、形態情報を使った検索語句の指定ができる。これにより、コーパスを使った本格的な統語論の研究が可能になった。

【言語資源 URL】

<https://malindo.aa-ken.jp>

マレー語・インドネシア語コンコーダンサーMALINDO Conc、マレー語・インドネシア語形態情報辞書 MALINDO Morph の説明と言語資源へのリンクがある。MALINDO Conc については、日本語のユーザーガイドも用意されている。