

東京外国語大学アジア・アフリカ言語文化研究所主催フィールド言語学ワークショップ

### 第11回文法研究ワークショップ

Lecture and Practical session on GRAID:

Annotating corpora for research on referentiality across typologically diverse languages

講義と実習：コーパスに基づく対象指示方略の類型～アノテーションシステムGRAIDを用いて～

#### 開催のお知らせ・募集要項

東京外国語大学アジア・アフリカ言語文化研究所では、以下の要領でワークショップを開催します。皆様のご参加をお待ちしております。

#### 記

1. 開催日時：2017年3月21日（火）・22日（水）10:30-16:30（両日とも）
2. 開催場所：東京外国語大学アジア・アフリカ言語文化研究所（AA研）  
マルチメディア会議室（3F・304）
3. 講師：Stefan Schnell 博士（メルボルン大学ポスドク研究員）
4. スケジュール：

	3月21日	3月22日
10:30-12:00	講義： 指示方略の選択と情報構造・イベント構造との相互作用に関する諸問題の紹介	講義： 一致システムの類型論・一致システムの歴史的発展
13:00-15:00	実習： アノテーションシステムGRAIDの紹介	実習： GRAIDを用いて一致システムの歴史的発展を調べる
15:15-16:30	実習： 参加者自身の調査データにGRAIDアノテーションを付与	

・ 内容の詳細は募集要項3ページ目の説明をご参照ください。

- ・ 午前中の講義のみの出席も可能です。
- ・ 実習に参加される方はパソコンをお持ちください。

5. 参加申込方法：次の情報を下記「7. 申込先」宛に電子メールでお寄せ下さい。

(1) 氏名（ローマ字） (2) 所属 (3) 連絡先メールアドレス

※件名は必ず、文法研究ワークショップ「コーパスに基づく対象指示方略の類型」受講希望として下さい。

6. 申込締切：2017年3月16日（木）午前11時59分

7. 申込先：

「多言語・多文化共生に向けた循環型の言語研究体制の構築（LingDy3）」事務局  
lingdy-apply[at]aacore.net（[at]を@に変えて送信ください）

8. その他：

- ・ 講義は英語で行います。通訳はありません。
- ・ 参加は無料です。

※ご不明な点がございましたら、上記「7. 申込先」までご連絡ください。

※過去の文法研究ワークショップにつきましては、

<http://www.aa.tufs.ac.jp/ja/training/fieldling-ws/grammar-wr-ws> をご覧ください。

主催：頭脳循環を加速する戦略的国際研究ネットワーク推進プログラム「危機言語・少数言語を中心とする循環型調査研究のための機動的国際ネットワーク構築」

後援：東京外国語大学アジア・アフリカ言語文化研究所基幹研究「多言語・多文化共生に向けた循環型の言語研究体制の構築 (LingDy3)」

以上

## Workshop Description

This two-day workshop deals with the empirical treatment of research questions in the area of referential choice and event representation, drawing on corpora from typologically diverse lesser-described languages. On the first day, we will first discuss the basic set of problems and questions concerning the expression of events and participants therein in connected discourse. Two crucial insights are connected to this line of research: firstly, that patterns of referential choice for participant roles in different types of events are accountable for in terms of general principles of information packaging; and secondly, that patterns of referential choice feedback into processes of grammaticalisation and language evolution. This forms a usage-based emergentist approach to language evolution and diversification, where grammatical systems essentially derive from principles of discourse organisation.

We then turn to the practicalities of addressing these research questions in a corpus-based cross-linguistic approach. For these purposes, we introduce the annotation conventions "GRAID" (Grammatical Relations and Animacy In Discourse) that have been developed for the specific purpose of research in the Chafe'an tradition. We will also introduce our multilingual corpus project Multi-CAST ([lac.multi-cast.de](http://lac.multi-cast.de)) that consists of GRAID-annotated narrative texts from currently 7 languages. After introducing the basic principles of GRAID annotations, participants will have the opportunity to trial these annotations on their own discourse data. For this practical part of the workshop, participants are kindly asked to bring their own recorded discourse data, see specifications and some requirements below.

On the second day, we will first discuss one of the most prominent recurring topics within the Chafe-Givónian research agenda, namely the development of agreement systems. We will first outline the three major accounts in terms of 1. overuse of topicalisation, 2. marking of high accessibility of speaker and hearer, and 3. mere frequency, before turning briefly to the typology of agreement systems and remaining challenges to account for attested cross-linguistic patterns in a usage-based approach. We also briefly take up questions about alignment in agreement patterns.

We then turn again to the practical treatment of these issues in a corpus-based approach. We first outline the treatment of different more or less canonical patterns of agreement in GRAID, focussing in particular on the issue of fully obligatory versus variable agreement. We then take a look at first results from participants' annotation work in comparison to existing datasets from the Multi-CAST collection.

### Requirements for text data:

- Should comprise monologous narratives rather than conversational data.
- Should be in ELAN, Toolbox or FLEx.
- Should contain at least transcription and free translation into English.
- Ideally is also morpheme-by-morpheme glossed.