# W-1



# クラウドソーシングを用いた言語分析
# Large-scale Crowdsourced Analyses for Linguistic Researches

企画・司会：国立国語研究所　浅原　正幸
National Intitute for Japanese Language and Linguistics, Japan, Masayuki ASAHARA

## ワークショップの趣旨

クラウドソーシングは，不特定多数の作業者（ワーカー）への簡単な仕事の外注（アウトソーシング）である。Yahoo! クラウドソーシングやランサーズなどのクラウドソーシングサイトでは，非常に小さい単位でのタスク（マイクロタスク）を数千人規模で依頼するサービスを提供しており、大量の語彙・例文に対する大人数のアンケート調査を安価に実施することができる。事例として，反対語 7,658 対に対する 1,597 人規模 (1 対あたり 20 人) の反対語らしさ・置き換え可能性調査 (荻原他, 2019) や，分類語彙表の見出し語 100,830 語に対する 3,392 人規模 (1 語あたり最低 16 人) の単語親密度調査 (浅原, 2019) などがあげられる。

本ワークショップでは，現代日本語の新しい調査方法として，クラウドソーシングを用いた大規模語彙調査事例について紹介する。1 つ目は形態素解析などで利用される形態論情報付与付き辞書整備を行った事例である。2 つ目は形容詞の叙述用法と修飾用法のちがいの認識について評価実験として検証した事例である。3 つ目は多義語の語義のちがいについて評価実験として検証した事例である。いずれの調査も各事例について数十人規模の評定値を得る数千人規模の調査である。

しかしながら，クラウドソーシングでは機械処理によりチェック設問（きちんと調査内容を理解しているかを判定する設問）に答えるものもある。こういったものをどのように排除するかについて紹介する。さらに分析時に一般化線形混合モデル（ベイジアン線形混合モデルも含む）により実験協力者間の揺れをモデル化する方法について紹介する。

会場では調査結果を可視化したものを席で見られるようにする。最後にコメントの時間をとるとともにフロアも含めた全体質疑を行う。

## ワークショップの構成

- 10:00-10:10 趣旨説明
  国立国語研究所　浅原正幸
- 10:10-10:30 クラウドソーシングによる形態論情報付与付き辞書整備
  国立国語研究所　岡照晃
- 10:30-10:50 クラウドソーシングによる述定・装定の用法分析
  筑波大学・国立国語研究所　西内沙恵
- 10:50-11:10 クラウドソーシングによる語義調査
  国立国語研究所　加藤祥
- 11:10-11:30 クラウドソーシング結果の可視化手法と統計処理
  国立国語研究所　浅原正幸
- 11:30-11:45 コメント
  国立国語研究所　山崎誠

- 11:45-12:00 全体質疑
  質問紙を配布して Q/A

### 各発表の概要

「クラウドソーシングによる形態論情報付与付き辞書整備」　国立国語研究所　岡照晃

本発表では，計算機データベース上の形態論情報付き辞書へ，クラウドソーシングを使い，新情報を大規模に追加する試みを紹介する。当該の辞書は 20 万語規模もの見出し語を有しているが（活用展開や異語形まで含めると 70 万語規模），ここへの新情報追加さえ，例えば見出し語間の関係として「複合語となっている見出し語へ，それを構成している別の見出し語たちはどれか？」といった情報の追加さえ，クラウドソーシングはわずか数日で可能にする。反面，ウェブ上の作業者の大多数は辞書整備に携わったことがなく，中には作業意図を無視する悪質なユーザも存在する。ここでは，こうした作業者たちをどのようにコントロールし辞書整備を行なっているか，具体例をあげて解説する。

「クラウドソーシングによる述定・装定の用法分析」　筑波大学・国立国語研究所　西内沙恵

本発表では，統語情報が意味の解釈にどのように影響しているのか，という問題の解明に向けて，クラウドソーシングによる実験調査からアプローチする。文法と意味の関係を探るために，クラウドソーシングを通した大規模被験者実験がどのように役立てられるかを，形容詞用法の事例から検討する。最も基本的な意味クラスに属する多義的形容詞の用法・意味別の実例について類似度を評定してもらう調査を実施し，用法の意味解釈への影響を分析した。述定・装定には，結合する名詞を説明・限定するという文法機能だけでなく，形容詞の多義解釈にもかかわる意味機能が備わっていることを調査結果から論じる。

「クラウドソーシングによる語義調査」　国立国語研究所　加藤祥

本発表では，多義語の意味調査例を紹介し，意味の調査における一般的な解釈としてのクラウドソーシング実験の活用可能性を考えたい。文脈における多義語の意味判定は，読み手によって，あるいは一人の読み手であっても揺れが生じるものである。そのため，用例を用いて多義語の意味判定のグラデーションを調査し，多義のネットワークと派生関係の解明を試みている。被験者実験では，調査対象とする多義語について，実験協力者に指標文と判定文を提示し，指標文と対照して用例における語の意味の類似度を判定してもらう。この結果，語義の関係がどのように用例に現れているのか整理することが可能となる。これらの調査手法と「一般的な」意味解釈の調査事例を示す。

「クラウドソーシング結果の可視化手法と統計処理」 国立国語研究所　浅原正幸

本発表では，クラウドソーシング結果のグラフによる可視化手法について紹介する。また，クラウドソーシングにおける実験協力者のバイアスを低減させるための統計処理について紹介する。ベイジアン線形混合モデルのランダム効果による実験協力者のゆれの処理について議論する。

**Acknowledgement**

本研究は国立国語研究所コーパス開発センター共同研究プロジェクト・所長裁量経費プロジェクト 2018 および科研費 17H00917, 18H05521, 18K18519, 19K13173, 19K00591, 19K00655 によるものです。

## References

荻原亜彩美・森山奈々美・浅原正幸・加藤祥・山崎誠 (2019) 「『分類語彙表』に対する反対語情報付与」,『言語処理学会第 25 回年次大会発表論文集』, 1061–1064 頁.

浅原正幸 (2019) 「クラウドソーシングによる単語親密度推定」,『言語処理学会第 25 回年次大会発表論文集』, 45–48 頁.