

多義語用例の類似度による語義の分類 — 「遠い」と「近い」を例に—

加藤祥 (国立国語研究所)・西内沙恵 (筑波大学 [院] / 国立国語研究所)・浅原正幸 (国立国語研究所)

『現代日本語書き言葉均衡コーパス』(以降 BCCWJ) に『分類語彙表』番号の付与が行われた(以降 BCCWJ-WLSP, 加藤ら 2019a)。意味情報が付与されたコーパスにより、多義語を語義別に頻度調査することが可能となった(加藤ら, 2019b)。しかし、読み手によって、あるいは一人の読み手の中でも場合によって、文脈による意味の判定には揺れが考えられ、実際、BCCWJ-WLSP の作業でも(1)のような例では作業者が意味判定に迷っている。

(1) 両手を胸の前でぎゅっと握りしめて、加護は遠い目をした。(サンプル ID : PB39_00372)

そこで、読み手の意味の判断に揺れや迷いの生じる用例と語義的に類似度の高い用例を取得し、多義語の語義と用例を整理することにした。具体的には、『分類語彙表』で複数の番号(語義)を有する語(多義語)について、各用例間の語義的な類似度を調査する被験者実験を行う。まず、BCCWJ-WLSP を用い、各語義において典型的な用例と、同コーンケーションや類似文脈ながら付与済みの語義に判定揺れが見られる用例群を収集した。このほか、BCCWJ から先行研究などに語義の記述があるものの BCCWJ-WLSP から取得できなかった種類の用例や、語義の付与判定が困難と考えられる用例を加え、各用例間の意味的な類似度(0~5 の 6 段階)を調査した(図 1)。実験協力者は用例 1(指標文) × 1(判定文)の各組合せにつき 20 名以上(日本語 ID を有する 20 歳以上の男女)をクラウドソーシングで募集した。

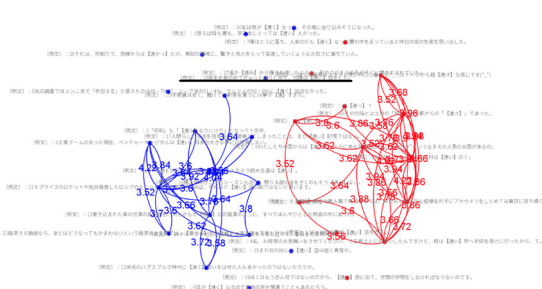
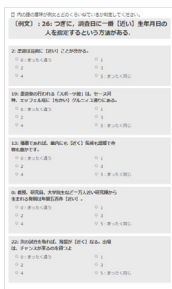


図 1 (左) 実験画面例, 図 2 (中央) 「遠い」の語義の高類似度用例群(類似度 3.5 以上),

図 3 (右) 「近い」の語義グラデーション(ここでは表の一部を图示した)

調査の結果、語義的に高い類似度と判定された用例群を可視化できる(例: 図 2 中下線が用例(1))。また、用例と語義との対応を確認し、『分類語彙表』や先行研究の記述を検証可能である。語義ごとの意味的なグラデーションを取得し(図 3)、語義判定の尺度を分析するとともに、多義語の各語義の派生関係を明らかにする。

本発表では、位置を示す「遠い」(33 用例, 56,100 組, 実験協力者 560 名(異なり))と「近い」(42 用例, 90,300 組, 実験協力者 1,453 名(異なり))の調査結果を例として報告する。「遠い」と「近い」は、意味分類や多義の派生において同様の傾向が見られるのではないが、意味的な重複の可能性が確認される。語義的類似度による用例群の分類例(図 2)のほか、用例を語義的な類似度によって語義グラデーションとして整理した表(図 3)を示す。

今後、この他の語の調査結果ともあわせ、多義語の意味分類と派生関係の整理を目指している。

謝辞

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、JSPS 科研費 19K00591, 19K00655, 国立国語研究所所長裁量経費 2018 によるものです。

参考文献

加藤祥, 浅原正幸, 山崎誠(2019a). 分類語彙表番号を付与した『現代日本語書き言葉均衡コーパス』の書籍・新聞・雑誌データ. 日本語の研究, 15 (2), p. To Appear./加藤祥, 田邊絢, 浅原正幸, 古宮嘉那子, 新納浩幸(2019b). 多義語の語義分布と語義間の派生関係調査の試み—一相の類を中心に—. 言語処理学会第 25 回年次大会./国立国語研究所(2004). 分類語彙表. 増補改訂版, 大日本図書.