

日本語文末表現による推論について

田辺 利文, 首藤 公昭 福岡大学工学部電子情報工学科
{ tanabe , shudo }@tl.fukuoka-u.ac.jp http://www.fukuoka-u.ac.jp



研究の背景

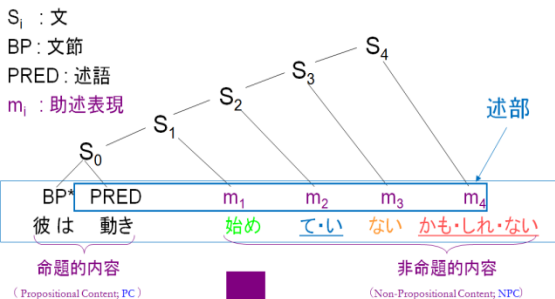
「Multiword Expressions: A pain in the Neck for NLP」(Sag et al., 2002)の発表 → MWE(複単語表現)の重要性が世界的に認知
以後ACLやCOLINGなどでほぼ毎年workshopが開催されるなどMWEの研究が盛んに
自立語性MWE:「油・を・売る」「水・に・流す」「責任・を・取る」など 付属語性MWE:「かも・しれ・ない」「なけれ・ば・なら・ない」など (“”は短単位の単語境界)

目的 日本語文末表現だけで推論してみる

文末表現辞書の概要

収録対象のMWEがもつ性質 f1: 非構成(イディオム)性(構成性原理の成り立ちにくさ) f2: 要素語間の強い共起性 の1つ以上をもつ
文末表現辞書は 約1,450個の付属語性MWEと約 50個の助動詞、終助詞を収録
それぞれの文末表現には対応する意味タグ列を定義(現段階で133種) 例)「なけれ・ば・なら・ない」→〈必要性₁〉、「まい」→〈否定・蓋然性〉
(意味タグ列中の“-”は意味タグの接続。また本研究では、これ以降、文末表現を「助述表現」とよぶ)

非命題的意味構造(NPS: Non-Propositional semantic Structure)と助述表現の統計データ



助述表現列 n	述部の生起	単一の単語 助述表現の生起 (A)	複単語 助述表現の生起 (B)	A+B	B/(A+B)
0	4,899	-	-	-	-
1	3,131	1,852	1,279	3,131	0.41
2	966	1,128	804	1,932	0.42
3	178	276	258	534	0.48
4	34	63	73	136	0.54
5	2	7	3	10	0.30
合計	9,210	3,326	2,417	5,743	0.42

EDRコーパス中の任意の述部(9,210個)を調査

述部に少なくとも1個の助述表現が含まれる割合: 47% (= 9,210 - 4,899 / 9,210)

助述表現が複単語表現である割合: 42% (= 2,417 / 5,743)

NPS 推量[否定₁完了[起動[彼は動く]]]

NPSの作成: 入力文に対し長単位の形態素解析を行った後、述部内に存在する助述表現列を意味タグ列に変換し、命題的内容部を引数として逆向きに意味タグ列を配置
NPSの特徴: 入れ子構造をしており、言語依存性がない → 機械翻訳や言い換えなどに応用可能 (Shudo et al., 2004; Tanabe et al., 2001)

推論ルール(の一部)

推論ルール: 推論(含意)によって、NPSを別のNPSに書き換えるルール

推論ルールは現段階で約180種

("n"は長単位の単語境界)

例	書き換え後	書き換え前NPS	書き換え後NPS	適用条件	ルールの特徴
彼/が/誘う/のだもの	彼/に/誘わ/れ/たく/なかつ/た	理由付け[x]	過去[否定[願望[受動態1[x]]]]	xが他動詞	ヴォイスが変わる
ご飯/を/食べる/とも	ご飯/を/食べる/つもりだ	同意[x]	方針・予定[x]	xが他動詞	自動詞か他動詞で書き換える推論ルールが変わる
成績/は/上がる/とも	成績/は/上がる/はずだ		推量[x]	xが自動詞	
食べ/にくい	食べ/たく/ない	困難性[x]	否定[願望[x]]	xが他動詞	「てくれる」がなくなる
上がり/にくい	上げ/たい		願望[x]	xが自動詞	
食べ/てくれる/な	食べる/な	禁止[受動態2[x]]	禁止[x]	xが他動詞	モダリティとアスペクトが絡む
うどん/を/食べ/なげればならない	うどん/を/食べ/て/いい/ない	必要性1[x]	願望[否定[x]]	xが他動詞	
ナマコ/を/食べ/なげればならなかった	ナマコ/を/食べ/たく/なかつ/た	過去[必要性1[x]]	過去[否定[願望[[x]]]]	xが他動詞	モダリティと主観表現が絡む
飲み/すぎる	飲む/な	過剰性[x]	傾向[x]	xが他動詞、主体が2人称	基本的には、主体の人称によって適用できる推論ルールが異なる。また、上司と部下の関係などによっても、適用に制約を付加させることができる
	飲み/けいこうがある		禁止[過剰性[x]]	xが他動詞、主体が2人称	
飲み/すぎ/た	飲み/すぎ/ない/ほうがいい	過去[過剰性[x]]	必要性4[否定[過剰性[x]]]	xが他動詞、主体が2, 3人称	〈不足性〉と〈過剰性〉は関連した意味タグであり、主体の人称によって適用できる推論ルールが異なる点は同じだが、書き換え後のNPSの体系は、〈過剰性〉には禁止があるなど、異なった体系になるようである
飲み/たりない	飲み/ない	不足性[x]	否定[願望[x]]	主体が1人称	
	飲み/なさい		願望[x]	xが他動詞、主体が1人称	命令[x]
	飲み/ほうがいい		必要性4[x]	主体が2, 3人称	
	飲む/だ		過去[x]	xが他動詞	使役態[x]
	飲ま/せる		使役態[x]	xが他動詞	
飲み/たりない/です	飲み/ましよう	丁寧[不足性[x]]	勧誘[x]	xが他動詞	

本研究によってもたらされるメリット

- 高精度な主観感情情報の処理の実現 詳細な意味分類、網羅性のある推論ルールにより、より精密な主観感情情報の処理が可能に
- 円滑な対話システムの実現 相手のいいたいことをより早く認識
- 通信トラフィックの削減 円滑な対話により、無駄な対話部分の省略が可能になることから通信のコストダウンが可能に
- 新感覚的教育ツールの構築 語用論的な側面をとらえた、新たな観点での日本語ネイティブネスの判定も
- 高度な文章校正 文脈に応じて、適切ないい回しに言い換えたり、提案するサジェスト機能の組み込み
- 高い実現可能性 辞書と推論ルールが共にコンパクトなため、既存システムへの組み込みも比較的容易

これからの課題

推論ルールのカバレッジの向上、推論ルールの人称および動詞のカテゴリによる適用条件の検証、推論ルールのレベル定義、性能評価

参考文献

- Sag, I. A., Baldwin, T., Bond, F., Copestake, A. and Flickinger, D. (2002). Multiword Expressions: A Pain in the Neck for NLP. Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics, CICLING2002, pp.1-15.
- Kosho Shudo, Toshifumi Tanabe, Masahito Takahashi, and Kenji Yoshimura. (2004). MWE as non-propositional content indicators. Proceedings of second Association for Computational Linguistics (ACL) Workshops on Multiword Expressions: Integrating Processing, pp.32-39
- Toshifumi Tanabe, Kenji Yoshimura, and Kosho Shudo. (2001). Modality expressions in Japanese and their automatic paraphrasing. Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS), pp.507-512.