2T-05

英文意味解析システム ESage

石田港　　谷津元樹　　原田実

青山学院大学理工学部情報テクノロジー学科

1. はじめに

一般に、人工知能技術の中でも，深層学習は，英語日本語翻訳など、英文に示された内容を日本語に変換することには大きな成果を上げているが，質問応答や要約は，１対１の単語翻訳よりはるかに複雑な問題であることから，有効な解決方法が見つかっていない．英文構造解析において，最先端と言われている Stanford Parser [2]の単語レベルの依存関係では文の意味を効率的に比較するのが難しい．一方，日本語には文節という文の構成単位があり，文節に主辞という主とした意味を持つ形態素を配置することで，文全体の意味をとらえやすくなっている．

2. ESage の概要

ESage は Stanford Parser による係り受け解析の結果と EDR 電子化辞書 [3]の情報を基に意味解析を行っている．Stanford Parser からは，係り受け情報以外に単語の品詞，原型，固有名詞かどうかの情報も利用している．ESage における処理の概要を図 1 に示す．

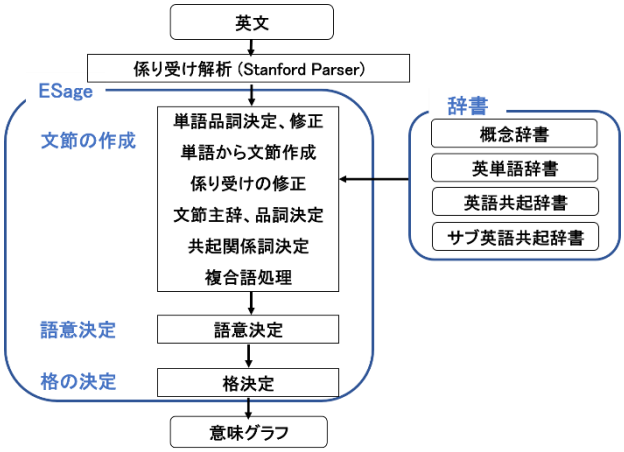

図 1. ESage のシステム構成

3. ESage における意味解析

入力文の係り受け解析結果から，意味グラフとして意味解析結果を出力するまでの流れを以下に説明する．

English semantic analysis system *ESage*

Ishida Minato, Yatsu Motoki and Harada Minoru
Department of Integrated Information Technology, College of Science and Engineering, Aoyama Gakuin University.

3.1 英単語の品詞の決定と修正

単語の品詞は Stanford Parser の品詞と EDR 電子化辞書の詳細品詞の対応付けを行い決定した．対応付けできないものはルールを作成し単語間の依存関係から品詞の決定を行った．中でも，Stanford Parser の品詞解析結果が予期しているものと異なった場合は，依存関係におけるルールからその解析結果を無視して品詞を正しいものに修正する．

3.2 文節の作成

Stanford Parser の解析結果は単語レベルの依存関係となっている．係り受け情報および品詞情報を用いて単語を日本語と同様に文節としてまとめた．文 “The train is due to arrive in Tokyo at ten.” の係り受け解析結果と文節作成後の係り受け構造の例を図 2 に示す．

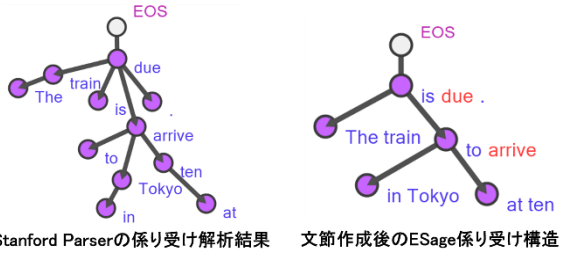

図 2. 文節作成の前と後における係り受け構造

3.3 係り受け・複合語の修正

EDR 詳細品詞を用いたルールから，係り受けの修正を行った．この修正により入力文によっては語順が変更される場合がある．また，”is able to”のように複数の語から 1 つの意味を持つ複合語になるものは結合し、逆にハイフンなどにより 1 語になっているもので、語意を得られない複合語は複数の単語に分割する修正を行った．

3.4 文節主辞・文節品詞・共起関係子の決定

文節の主辞・品詞・共起関係子の決定について説明する．共起関係子とは，語間の役割関係を示す前置詞，接続詞もしくは記号である．主辞と共起関係詞は 2 文節間の深層格を決定する際に辞書引きで用いる．主辞は主に EDR 品詞に優先順位を決め文節ごとに決定した．共起関係子はルールを決め，文節品詞・係り受け情報・語順を基に決定した．

Copyright ©2020 Information Processing Society of Japan.
All Rights Reserved.

3.5 語意と格の決定

語意と格の決定について説明する．語意は概念辞書の約 40 万種類から，格は共起辞書に出現する 37 種類から選択する．共起辞書から下記の 5 種類の方法で語意と格の候補を検索し出現確率に基づく評価を行い決定した．また冠詞の格など一意に決定できるものはルールにより決定した．

3.6 意味グラフの出力例

文“What time was Kenji Taira found in his car after he had attacked his sister-in-law?” の出力結果を図 3 に示す．”was found”や”after had attacked”を 1 つの文節としてまとめている．また，各文節の下にはその主辞の EDR 語意が，[EDR 語意 ID：日本語語意説明：英語語意説明] として示されている．各文節の中でも特に主要な意味を持つ主辞は赤く表示されている．また，各文節間の役割的関係は受け側文節から係り側文節に引かれた矢印線の上に赤字で表示されている．例えば，"he"は"after had attacked"の意志のある主語なので、'agent'が割り当てられている．一方，"Kenji Taira"は"was found"の目的語として'object'が割り当てられている．さらに，文の重要な述語を表す文節は，文末を表す EOS 文節にかかっている．例文では、"was found"がこの主述語にあたる．さらに、例文は質問文なので、この主述語にモダリティ"question"が与えられている．また，"was found"は受動態でもあるため、この文節にはモダリティ"passive"も与えられている．

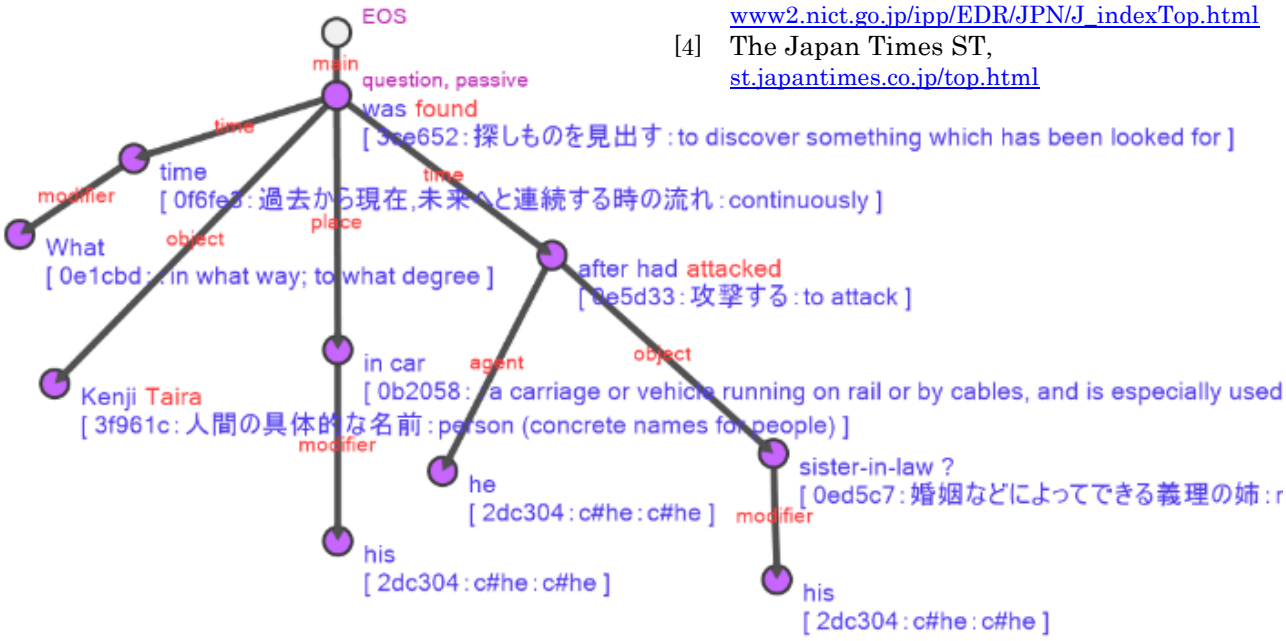

図 3. 意味グラフの出力例

4. 評価実験

The Japan Times ST [4]から 112 文を ESage で解析し,全形態素と全係り受けの深層格における正解率(=正解数/全出力数)を人手で評価した．

結果を表 1 に示す．主な品詞のうち、語意の精度が最も高いのは形容詞で，そのつぎに固有名詞，そして最も低いのは動詞であった．これは，形容詞や固有名詞などに比べ，動詞は語意の候補数が多いためである．

表 1. 評価実験結果

	語意の精度	格の精度
全出力数	2510	1185
正解率	91.2%	81.0%

5. まとめ

本研究によって 40 万を超える語意粒度において 90%以上の精度をもつ英文に関する意味解析システムを世界で初めて実現した．

今後の課題として，語意の精度に比べた深層格の精度の低さが挙げられる．深層格は共起辞書や決定ルール，語意は辞書の充実により精度上昇が期待できる．

参考文献

[1] 原田実,水野高宏: “EDR を用いた日本語意味解析システム SAGE", 人工知能学会論文誌,Vol.16, No.1, pp.85-93 (2001.1).
[2] Stanford Parser, nlp.stanford.edu/software/lex-parser.shtml
[3] EDR 電子化辞書, NICT. www2.nict.go.jp/ipp/EDR/JPN/J_indexTop.html
[4] The Japan Times ST, st.japantimes.co.jp/top.html

Copyright ©2020 Information Processing Society of Japan.
All Rights Reserved.