システムディスクリプション

|  |
| --- |
| 記入日：2012年7月31日  対象トラック：　一致  システム名：　規則に基づくagreement error detectionシステム  チーム名：　NAIST  性能：　ドライラン　 R=0.723 P=0.556 F=0.628  　　　　フォーマルラン　R=0.913 P=0.553 F=0.689 |

1. 概要

本システムは依存構造解析の結果をもとに、主語と動詞の一致に関する誤りを規則に基づいて（ルールベースで）検出するシステムである。

Stanford Parserによる依存構造を解析結果および作成した規則をもとに、スクリプトによる自動タグ付けを行った。

1. 実行方法

実行環境・使用言語：Ubuntu 10.04.3 LTS

実行方法：

0. CorrChaをインストールする。

1. CorrChaディレクトリの直下にedcw\_detect\_agree.pyを配置する。

2. 検出対象となる\*.edcファイルのみを含むディレクトリ(e.g. all\_edc)を作成する。

3. 検出結果ファイルを保管するディレクトリ(e.g. all\_sys)を作成する。

4. edcw\_detect\_agree.pyを実行する。末尾のスラッシュに注意。

例）python edcw\_detect\_agree.py -i all\_edc/ -o all\_sys/

5. 3で作成したディレクトリに検出結果が保管されていることを確認する。

6. zipコマンドで出力結果をzipファイルに圧縮する。

例）zip test.zip all\_sys/\*.sys

1. 使用データ，ツール，辞書

ツール名：

Stanford Parser 2.0.2

CorrCha (NAISTの林部さんが作成したもの。)

Python 2.6.5 + python-argparse (Python 2.7 以降には標準添付)

EDCW(KJ-コーパス)フォーマットのデータ

1. システムの詳細

処理の流れは次の通りである

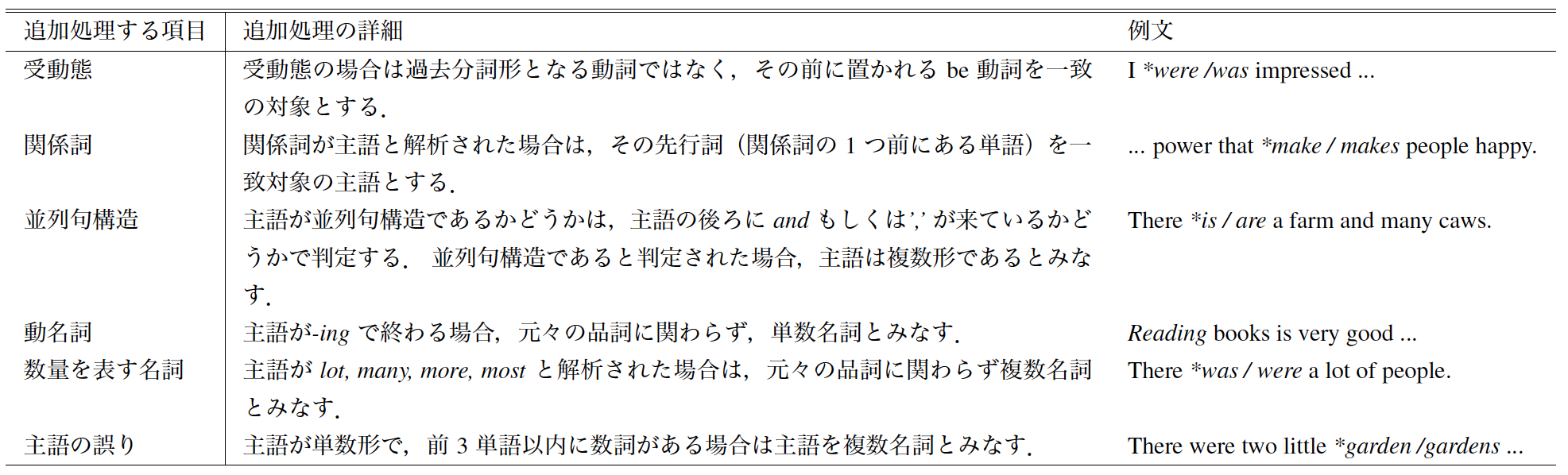
（1）Stanford Parserによる依存構造解析を行う。

（2）上記解析結果をもとに一致させるべき主語と動詞を同定する。ただしこのとき、Stanford Parserで主語と動詞を単純に取得できない場合については表１の追加処理に従って、主語と動詞のペアを取得する。

（3）（2）で取得した主語と動詞のペアが文法的に一致しているかを判定する。

（4）（3）の判定結果をもとに自動でタグを付与する。

表１　追加処理一覧



1. セールスポイント

Stanford Dependencyでは単純に主語と動詞の一致を見ることができないような文（受動態、完了形、関係詞節）に対して、一致させるべき動詞を取得している点。

そして主語が並列句構造については、表層的な特徴（主語の後ろにandか“,”が来ている場合は主語が並列句構造をなしている場合が多い）を用いて実現している。いずれの方法もナイーブでありながら、誤り検出に対して頑健に働くことが示されている。（追加処理を行わなかった場合、ドライランの性能は、R=0.530, P=0.355, F=0.425であった。）