Vol.2016-NL-226 No.14
Vol.2016-SLP-111 No.14
2016/5/17

NLTK のタグ付けに対する修正及び文法情報の付加

山岡幸高[†1]

機械翻訳，特に英日翻訳がうまくいっていないと言われて久しい．タグ付けの自動化がうまくいっていないことが主な原因である．また，高い精度で文法情報を付加するシステムもない．
そこで，まず NLTK でタグ付けをし，その誤りを自作の構文解析機で修正し，文法情報を付加することを考えた．その結果，タグ付けの制度に著しい向上が見られた．また学校文法による解説の付加と文型表示にも成功した．

1. はじめに

機械翻訳の精度向上等のため，自動でタグ付けをするシステムの需要は多い. 以前私は, 単語の品詞などを表示し，学校文法における解説を付加するシステムを発表した[1]. 精度は高かったものの，一般の文章に適用するためには，専用の辞書を準備しなければならないため，膨大な時間がかかる．

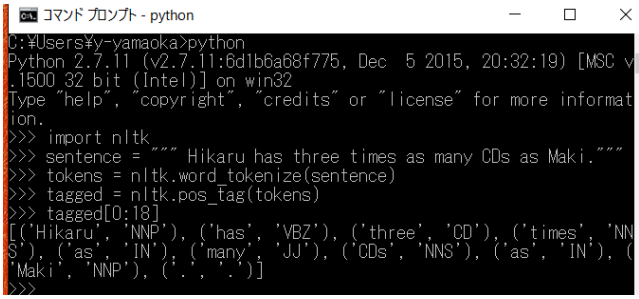
そこで今回，ユーザーにまず NLTK [2] でタグ付けをしてもらい，その結果を私の構文解析機にかけてもらうことにより，タグ付けの誤り修正し，情報を付加することを考えた．

2. 方法

まず

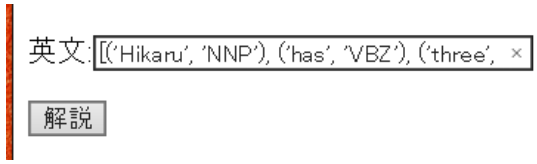
Python 2.7
NLTK
numpy-1.10.2-win32-superpack-python2.7.exe [3]

をダウンロード，インストールし，タグ付けしてもらう．
そうすると例えば



となる．（
[('Hikaru', 'NNP'), ('has', 'VBZ'), ('three', 'CD'), ('times', 'NNS'), ('as', 'IN'), ('many', 'JJ'), ('CDs', 'NNS'), ('as', 'IN'), ('Maki', 'NNP'), ('.', '.')]
であるが，この例では RB(副詞)である最初の as が IN と表示されている）

そして
http://yamaoka.rosx.net/yamaoka-nltk.html
に上の結果をコピー・アンド・ペーストし



解説ボタンを押すと

品詞
hikaru(N) have(V) three times(副) as(副) many(形) CDs(N) as(接) maki(N) .
文型 S V O {as N .
文法 A … X times as 原級 as B「の X 倍」
副詞的目的格

となる（最初の as は副(RB)と修正されている）.

使い方の詳細は Web サイト [4] にあるので見てほしい.

文法項目

解説で表示させる文法項目を以下のように設定した．略号は表 1 にまとめた.

1 文型
There is[are] 名詞
be 名詞[形容詞] {V C} keep 形容詞 {V C}
become 名詞 {V C} get 形容詞 {V C}
look 形容詞 {V C} seem 形容詞 {V C}
feel 形容詞 {V C}
smell 形容詞 {V C} taste 形容詞 {V C}
give 名詞 名詞 {V O O} lend 名詞 名詞 {V O O}
buy 名詞 名詞 {V O O} make 名詞 名詞 {V O O}
cook 名詞 名詞 {V O O}
make 名詞 形容詞 {V O C} keep 名詞 形容詞 {V O C}
paint 名詞 形容詞 {V O C}
call 名詞 名詞 {V O C} name 名詞 名詞 {V O C}
find 名詞 形容詞 {V O C}

†1 九州大学
Kyushu University

2　時制
am/is/are V-g {現在進行形}
was/were V-g {過去進行形}
be going to V
will be V-g {未来進行形}
have[has] V-n {現在完了形}
have[has] been V-g {現在完了進行形}
had V-n {過去完了形}
had been V-g {過去完了進行形}
will have V-n {未来完了形}

3　助動詞
may have V-n「したかもしれない」
must have V-n「したにちがいない」
cannot[can't] have V-n「したはずがない」
should[ought to] have V-n
cannot[can’t] help V-g

4　受動態
be being V-n {進行形の受動態}
have[has] been V-n {完了形の受動態}

5　不定詞
It is 形容詞[名詞] to V {不定詞の名詞用法}
S V it 形容詞[名詞] to V {不定詞の名詞用法}
名詞 to V {不定詞の形容詞用法}
不定詞の副詞用法
It is 形容詞[名詞] for A to V
It is 形容詞 of A to V
allow 名詞 to V
make 名詞 V　let 名詞 V　have 名詞 V　see 名詞 V
to have V-n {完了形の不定詞}
to be V-n {受動態の不定詞}
what to V {疑問詞＋to V}
too 形容詞/副詞 (for A) to V
形容詞/副詞 enough (for A) to V
be to V

6　動名詞
動名詞
having V-n {完了形の動名詞}
being V-n {受動態の動名詞}

7　分詞
V-g 名詞 {分詞の形容詞用法}
V-n 名詞 {分詞の形容詞用法}
名詞 V-g … {分詞の形容詞用法}
名詞 V-n … {分詞の形容詞用法}
remain V-n　come V-g
keep 名詞 V-g　see 名詞 V-g　have 名詞 V-n
分詞構文
V-n {受動態の分詞構文}
having V-n {完了形の分詞構文}
名詞 V-g {独立分詞構文}
with 名詞 V-g {付帯状況}

8　接続詞
so 形容詞 that S V

9　関係詞
名詞 who V {主格の関係代名詞}
名詞 which V {主格の関係代名詞}
名詞 whom/who S V {目的格の関係代名詞}
名詞 which S V {目的格の関係代名詞}
名詞 S V {目的格の関係代名詞の省略}
名詞 whose 名詞 … {所有格の関係代名詞}
前置詞 which {前置詞＋関係代名詞}
名詞 where S V {関係副詞}　名詞 when S V {関係副詞}
the reason why S V {関係副詞}
what
whoever V {名詞節を導く複合関係代名詞}
whoever S V {名詞節を導く複合関係代名詞}
whatever S V {名詞節を導く複合関係代名詞}
whatever V {名詞節を導く複合関係代名詞}
whoever V {副詞節を導く複合関係代名詞}
whoever S V {副詞節を導く複合関係代名詞}
whatever S V {副詞節を導く複合関係代名詞}
whatever V {副詞節を導く複合関係代名詞}
whenever S V {副詞節を導く複合関係副詞}
wherever S V {副詞節を導く複合関係副詞}
however 形容詞[副詞] S V {副詞節を導く複合関係副詞}

10　仮定法
If S V-d …, S would[could] V ….「もし S が…するなら、S は…する[できる]だろうに」
If S had V-n …, S would[could] have V-n ….「もし S が…したなら、S は…した[できた]だろうに」
If it were not for …, S would[could] V …
If it had not been for …, S would[could] have V-n …
S V … as if S V-d ….
S V … as if S had V-n ….
S wish S V-d ….「S は S が…すればよいのにと思う」
S wish S had V-n ….「S は S が…すればよかったのにと思う」
It’s (about) time S V-d ….

Vol.2016-NL-226 No.14
Vol.2016-SLP-111 No.14
2016/5/17

11　比較
A … as 原級 as B「A は B と同じくらい」
A … as many 名詞 as B　A … as much 名詞 as B
A … not as[so] 原級 as B
A … twice as 原級 as B「A は B の 2 倍」
A … X times as 原級 as B「A は B の X 倍」
A … 比較級 than B「A は B より」
A … the 最上級 (名詞) (in/of B)
No (other) 名詞 … as[so] 原級 as A「A ほど な 名詞 はない」
Nothing … as[so] 原級 as A
No (other) 名詞 … 比較級 than A「A より な 名詞 はない」
Nothing … 比較級 than A
A … 比較級 than any other 名詞「A はほかのどの 名詞 よりも」
the 比較級 of the two (名詞)
比較級 and 比較級「ますます…」
The 比較級 S V …, the 比較級 S V …「すればするほど、ますます」
(all) the 比較級 for 名詞　(all) the 比較級 because S V …
not so much A as B「A というよりはむしろ B」
no more than 数詞
A is no more B than C is D「A が B でないのは C が D でないのと同様である」

12　否定
no 名詞「1 つも…ない」
be no (形容詞) 名詞「決して…でない」
not all {部分否定}「すべてが…というわけではない」
not always {部分否定}
never V … without V-g … {二重否定}「…すれば必ず…する」
anything but …「決して…でない」

V : 動詞,　V-d : 動詞の過去形,
V-g : 現在分詞, 動名詞,　V-n : 過去分詞
v : 助動詞,　v-d : 助動詞の過去形
N : 名詞,　所 : 代名詞の所有格,　冠 : 冠詞
形 : 形容詞,　副 : 副詞
比 : 比較級,
形比 : 形容詞の比較級,　副比 : 副詞の比較級
疑 : 疑問詞,　前 : 前置詞,　接 : 従属接続詞
関代 : 関係代名詞,　関副 : 関係副詞
　文法解説での略号
(　) : 省略可能　[　] : 入れ替え可能　/ : 入れ替え可能
{　} : 追加説明
　文型表示での略号
S : 主語　C : 補語　O : 目的語
[　] : 名詞句, 名詞節　(　) : 形容詞句, 形容詞節　{　} : 副詞句, 副詞節

表 1　品詞の略号

3. 結果

総合英語参考書で重複の多かった例文 [5] に対し，解説（文法項目）の表示が正しくできたかどうかを示す.

文型
13 文中 12 文○，1 文×
時制
20 文中 20 文○
助動詞
9 文中 9 文○
受動態
11 文中 11 文○
不定詞
25 文中 21 文○，3 文△，1 文×
分詞
20 文中 17 文○，1 文△，2 文×
動名詞
7 文中 6 文○，1 文×
関係詞
18 文中 18 文○
仮定法
14 文中 14 文○
比較
24 文中 24 文○
否定
6 文中 6 文○

○ :　正解
△ :　A or B と 2 つ表示して片方が正解
× :　不正解

4. おわりに

動詞が文頭にある場合のタグ付けに問題があった．また長く複雑な文に対しては，句や接の品詞の識別が必要なもの（準動詞や複合関係代名詞など）に関する精度が下がった．しかし総合英語参考書レベルの文に対しては，上のように全般的に満足のいく結果が得られた.
cgi をアップロードしているので，ぜひ試してほしい.
http://yamaoka.rosx.net/yamaoka-nltk.html

謝辞 ご協力頂いた皆様に，謹んで感謝の意を表する．

参考文献

1 山岡幸高: 構文解析機による英文法解説，研究報告自然言語処理 (NL) , 2015-NL-223(9),1-4 (2015-09-20) , 2188-8779
2 Natural Language Toolkit
http://www.nltk.org/
3 Numerical Python
https://sourceforge.net/projects/numpy/files/NumPy/
4 http://yamaoka.rosx.net/yamaoka-nltk.html
5 http://yamaoka.rosx.net/yamaoka.html
6 田中省作，小林雄一郎，徳見道夫，後藤一章，冨浦洋一，柴田雅博: 学校英文法の学参例文データベースとその応用，情報処理学会研究報告，第 2012-CH-93 巻第 5 号:1-8 (2012).
7 重藤優太郎，東藍，近藤修平，北裏龍太，坂口慶祐，光瀬智哉，久本空海，吉本暁文, Frances Yung，松本裕治: 英語の複単語表現辞書の構築と品詞タグ付けへの応用，情報処理学会研究報告，Vol.2012-NL-209 No.7 (2012).
8 丹生伊佐夫, Graham Neubig，小林和也, Sakriani Sakti，戸田智基，中村哲: 構文情報が機械翻訳に及ぼす影響の分析，情報処理学会研究報告, Vol.2013-NL-212 No.8 (2013).
9 南條浩輝，吉見殻彦，岡田真也: 機械翻訳のための統計的手法に基づく前編集，情報処理学会研究報告, Vol.2009-NL-291 No.1 (2009).
10 大野一樹，波多野賢治: 係り受け関係の階層化とその共起に基づいた構文木モデルを利用した構文解析手法の提案，情報処理学会研究報告, Vol.2013-NL-214 No.6 (2013).
11 野村恵造: Vision Quest 総合英語，新興出版社啓林館 (2013)
12 石黒昭博: 総合英語 Forest 7th Edition，桐原書店 (2013)
13 小寺茂明: デュアルスコープ総合英語 四訂版，数研出版 (2011)
14 吉波和彦，北村博一: ブレイクスルー総合英語 改訂二版，美誠社 (2011)
15 瓜生豊，篠田重晃: Next Stage 英文法・語法問題，桐原書店 (2011)
16 Studyplus
http://studyplus.jp/home
17 東京都教科書委員会
http://www.kyoiku.metro.tokyo.jp/press/2015/pr150827a.html
http://www.metro.tokyo.jp/INET/OSHIRASE/2014/08/20o8s500.htm