

構文解析システムを利用したスペイン語品詞体系の設定

栗林ゆき絵¹

フリー

1999/03/16

1 研究の目的

従来、辞書や文法書においては、伝統的な品詞のリストを提示し、それを各語に付与する形で、品詞の記述とすることが一般的であった。限られた数の既に定められた品詞を個別の語に与えるため、複数の品詞の特徴を併せ持つ場合は、「品詞Aの品詞B的用法」、のように記述されたり、辞書では一つの見出しの下に複数の品詞の下位見出しを設けて記述するなどの方法がとられている。しかし、品詞とは何かを改めて考えると、それを統語的性質を同じくするカテゴリとする立場をとると、上記のようにある1つの品詞が付与された1つの語に別の品詞の特徴を認めるのは不自然であり、言語の学習上も効率的とは言えない。また、表記・音型の同じ1つの語を複数の品詞の多品詞語とみなすのが必ずしもいつも適切であるとは限らない。

本研究では、品詞を純粋に統語的なカテゴリとし、構文解析システムを利用して、その文法規則から品詞を定義することで、統語的性質のラベルとしての品詞体系を提示することを試みたい。

2 方法

品詞体系の設定は以下の手順で行った。

1. コーパスからの例文の収集
2. 例文の解析実験による品詞の追加、修正、およびルールのチューニング
3. 品詞体系とルールリストの確定

例文の収集は偏りのないよう、短編小説の冒頭から100文を抜き出し、最初のチューニングに利用した。初期に設定した品詞を例文の各語に付与し、それに対するルールを記述して例文の解析実験を始めた。解析実験を行いながらルールのチューニングを行っていく際に、新しいカテゴリすなわち新しい品詞を設けたほうが効率的な場合は、品詞を新たに用意してそれに対するルールを記述するという方法で、ルールの作成と品詞体系の作成を同時に並行して行っていた。最初の100文の解析実験が終了した時点で、解析能力の不十分だった限定詞および数量詞まわりを重点的にチューニングするために、その後、限定詞および数量詞を含む文約100文を小説、新聞等から追加して収集し、それらの解析実験を行いながら、限定詞、数量詞まわりの品詞体系とルールの作成を行った。収集した例文のうち、対象とした構文(詳細は文献[5]参照)のものがひととおり解析できるようになった時点でチューニングを終了し、品詞体系とルールのリストを確定した。

ルールの作成、品詞の設定にあたっては、「できるだけ少ないルール、できるだけ少ない品詞で、できるだけ多くの文の解析が可能となるようにする」という基本的な方針のもとで行う。したがって、解析精度を上げるために、品詞を細分化したり、品詞の数をむやみに増やすことは避ける。また、同じ見出しで複数の品詞を持つ多品詞語も原則として認めない立場をとる。

本構文解析システムは、品詞とルールの体系を考察することを目的とするため、辞書は作成せず形態素解析のフェーズは持たない。見出し語とあらかじめ付与された品詞の記述された入力ファイルにより構文解析を行う。

¹e-mail:kuriy@bk.iij4u.or.jp

3 実際の品詞体系設定作業の例: 限定詞・数量詞まわりの品詞体系の設定

本研究では特に、冠詞、不定語などの限定詞と、数量詞の周辺の品詞体系を重点的に整理した。

分類にあたっては、今回は対象とする語を以下のものに絞って検討した。(性数変化するものは変化形も扱う。)

mucho, poco, todo, demasiado, bastante, algo, nada, más, menos, uno(unも含む), alguno(algúnも含む), ninguno(ningúnも含む), otro
dos(tres, cuatro,...)[数詞,numeral],²
el(la, los, las)[定冠詞,artículo definido],
este(esta, estos, estas), ese(esa, esos, esas), aquel(aquella, aquellos, aquellas) [指示語, demostorativo],³
mi,tu, su[所有詞,posesivo],⁴

3.1 関連する文法規則

上記の語の解析には以下のルールが使用された。(Xは任意の品詞、SXは任意の品詞Xを主要部とする句を表す。)

3.1.1 既存の文法規則(あらかじめ記述してあったもの)

1. $X+SN \rightarrow SX/SN$ (SN:名詞句, /:orを表す。SXになるかSNになるかはXの品詞によって決まる。)
2. $SV+SX \rightarrow SV$ (SV:動詞句)
3. $X+sap \rightarrow SAP$ (sap:位置関係語, SAP:sap句)
4. $SX+SP_{sn} \rightarrow SX$ (SP_{sn}:名詞修飾前置詞句)
5. $pr+SX \rightarrow SP$ (pr:前置詞, SP:前置詞句)

3.1.2 新規の文法規則(限定詞・数量詞まわりの対応に今回新規に追加)

1. 規則 a: $todo+SX \rightarrow STODO$ (todo:品詞 todo, STODO:todo句)
規則 b: $todo \rightarrow STODO$
2. $X+SP_{sn} \rightarrow SX+SP_{sn}$
3. 規則 a: $X+[NOT(SN,SD)] \rightarrow SX+[NOT(SN,SD)]$ (SD:限定詞句, [NOT(A)]:A以外を表す)
規則 b: $X+SD \rightarrow SD$
4. $SV+SX+SY$ のとき、 $SX+SY$ の修飾関係(修飾可能性)を先にチェック
5. $X+mas \rightarrow SMAS$ (mas:品詞 mas, SMAS:mas句)
6. $SD+X \rightarrow SD$

²unoはいわゆる不定冠詞の用法と合わせて扱うため、数詞には含めない。

³esto, eso, aquelloは代名詞とし、ここでは扱わない(p.ej. *Eso debe de ser marciano.* (Corpus[1]:25))。アクセントなしのeste, ese, aquelが代名詞機能を持つ場合は対象とする。

⁴名詞に後置されるmío, tuyo, suyo, nuestro, vuestro、およびアクセントつきのmíはここでは扱わない。

3.2 限定詞・数量詞まわりの品詞体系

前述のルールと併せて下記の品詞を設定した。(品詞、(実際の語)、適用規則、の順に示す。)

1. 定冠詞: dtad(*el, la, los, las*):適用規則=既存1, 4⁵, 新規1a,2,3b
2. 指示語: dtdm(*este, ese, aquel, etc.*):適用規則=既存1, 4,5⁶, 新規1a,2, 3a, 3b, 6
3. 所有詞: dtps(*mi, ti, su*):適用規則=既存1, 新規1a,3b
4. 不定冠詞: dtai(*un, uno, una, unos, unas*):適用規則=既存1, 4,5, 新規1a,2,3a,3b
5. 不定語: dtid(*alguno, ninguno, otro*):適用規則=既存1, 4,5, 新規2,3a,3b
6. 数詞: dtnu(*dos, tres, cuatro..., etc.*):適用規則=既存1, 4,5, 新規2,3a
7. 数量語: sac(*mucho, poco, un poco, demasiado, bastante*):適用規則=既存1,2, 4,5, 新規4,5
8. 品詞 todo(*todo*):適用規則=既存1,2, 4,5, 新規1b
9. 品詞 algo(*algo, nada*):適用規則=既存2, 4,5, 新規4,5
10. 品詞 mas(*más, menos*):適用規則=既存1,2,3, 新規4

3.3 現時点でのスペイン語の品詞体系

品詞体系は現時点では、全体で次のようになっている(計30品詞)。今後さらにルールを改良し、解析できる構文を増やしていく予定であるので、その過程で品詞体系もさらに修正されることになる。

s(名詞), sap(位置関係語), sat(時間語), sac(数量語), v(動詞), inf(不定詞), pp(過去分詞), gr(現在分詞), cp(コピュラ), aj(形容詞), dtad(定冠詞), dtdm(指示語), dtps(所有詞), dtid(不定語), dtai(不定冠詞), dtnu(数詞), mas(品詞 mas:比較語), mejor(品詞 mejor:比較語), todo(品詞 todo), algo(品詞 algo), av(副詞), cl(接辞), prnp(名詞修飾前置詞), prvp(動詞修飾前置詞), praj(形容詞修飾前置詞), que(補文標識), rel(関係詞), cj(接続詞), pa(並列詞), no(品詞 no:否定辞)

4 作業過程で得られた品詞の特性等

品詞体系をルール作成の過程で確定していくなかで、動詞の機能語性や、形容詞と数詞の連続性など、品詞に関する興味深い事象がいくつか見られた。これらは、自然言語処理の分野にとどまらず、言語学の研究としてこれから注目していく価値のある点と思われる。ここではそういった注目すべき事象の1例として、形容詞と数詞の連続性について触れておく。

4.1 形容詞と数詞の連続性

形容詞と数詞は一見まったく関係のない別の品詞のように思われるが、ルールを実際に記述してみると、この2つの品詞は共通のルールをかなり持っていた。(現時点で形容詞のルールが15、数詞のルールが15あり、そのうち9つは共通のルールである。) このことは形容詞と数詞という2つの品詞の連続性を表すものであり、今後の研究で、これらの類似点および相違点について調査を進め、この2つの品詞の位置づけを明らかにしていくことが、文法記述における1つの重要な課題であることが示された。

⁵既存4は左辺のSXがこの品詞の語単独で形成された句である場合を想定している。

⁶既存5:注5と同様。

5 まとめ

本研究では、スペイン語の品詞体系の設定を、構文解析システムのルールを利用するという方法で行った。品詞は、それに適用されるルールの集合によって定義されるものとし、統語的ふるまいを同じくする語は同じ品詞に属するような品詞の設定を行うことができた。この方法を用いると、複数の品詞の性質をもつ語に対して、「品詞Aの品詞B的用法」、のような例外の記述をする必要がなくなり、品詞がその語の統語的性質を過不足なく表すことになる。品詞体系だけでなく文法規則全般の記述にあたっても、構文解析システムを利用することにより、形式的でかつ例外のない記述が可能となり、同時にシステムによる検証が可能であるので、今後、スペイン語文法研究に継続して利用していきたい。

Bibliography

- [1] Bosque, Ignacio. 1997. "Sobre los complementos de medida". *Homenaje a Josse de Kock*. Universidad de Lovaina.
- [2] 出口厚実. 1995. 「不定語と否定語」. [11] 所収.
- [3] 影山太郎. 1993. 『文法と語形成』. ひつじ書房
- [4] 栗林ゆき絵. 1998. 「自然言語処理からみたスペイン語における機能語」. 第18回スペイン語学研究会夏期セミナーハンドアウト
- [5] 栗林ゆき絵. 1998. 「数量を表す語の品詞分類について(自然言語処理を利用した品詞分類)」. 第44回日本イスパニア学会ハンドアウト
- [6] Leonetti Jungl, Manuel. 1990. *El artículo y la referencia*. Taurus. Madrid
- [7] 中岡省治. 1995. 「数量詞」. [11] 所収.
- [8] Takagaki, Toshihiro. 1997. "El participio adjetivo en español". *Lingüística Hispánica*. vol.20. pp.143-164.
- [9] 寺崎英樹. 1998. 『スペイン語文法の構造』. 大学書林
- [10] 瓜谷望. 1995. 「スペイン語限定形容詞の範囲」. 『語学研究第80号』. 拓殖大学語学研究所
- [11] 山田善郎他. 1995. 『中級スペイン文法』. 白水社.
- [12] 桑名一博他編. 1990. 『西和中辞典』. 小学館.

Corpus

- [1] Mateos, Pilar, 1983. *Lucas y Lucas*. Ediciones SM. Madrid
- [2] Martín Gaité, Carmen, 1953. "Un día de Libertad". en *Cuentos completos*. 1978. Alianza Editorial. Madrid
- [3] Martín Gaité, Carmen, 1958. "Lo que queda enterrado". en *Cuentos completos*. 1978. Alianza Editorial. Madrid
- [4] El país digital (<http://www.elpais.es>)