

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 13 日現在

機関番号：23901

研究種目：若手研究(B)

研究期間：2013～2015

課題番号：25770207

研究課題名(和文) Function word choice in Spanish learner language: error annotation, automatic detection and correction

研究課題名(英文) Function word choice in Spanish learner language: error annotation, automatic detection and correction

研究代表者

バルベルデ ピラル (Valverde, Pilar)

愛知県立大学・外国語学部・准教授

研究者番号：10588205

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：本研究の目的は、スペイン語を学ぶ日本語母語話者を対象に、機能範疇の言葉(冠詞、前置詞、接続詞等)の誤り(以後、エラー)の自動検出システムを開発することであり、以下の点について考察した。1) 外国語教師のエラーに対する許容度とはどのようなものか、またそれをどのように改善できるか、2) スペイン語母語話者の文と比較し、日本語母語話者はどの単語を多用するのか、又は活用できていないか、3) 日本語母語話者によって書かれたスペイン語文をどのように収集するか、4) エラーを自動検出するためにConstraint Grammar方式をどのように活用するか。

研究成果の概要(英文)：The goal of the research is to develop a system for automatic error detection in grammatical words (articles, prepositions, conjunctions, etc.), for learner of Spanish whose first language is Japanese. To do that, we have studied, for these type of learners, 1) the agreement level of language teachers when correcting errors in grammatical words, on how agreement can be improved, 2) which words are overused and underused in learner texts with respect to native texts, 3) how we can collect texts to study their interlanguage and 4) how we can use the Constraint Grammar formalism to automatically detect errors in grammatical words.

研究分野：コーパス言語学

キーワード：スペイン語 言語習得 コーパス 誤り検出 制約文法

1. 研究開始当初の背景

本研究は外国語教育・学習、コーパス言語学、自然言語処理という3つの分野に基づいている。

近年、外国語学習者の誤り自動検出システムへの関心が高まっているが、英語以外の言語ではほとんど存在していない。そしてこれらのシステムは、頻出単語であると同時に学習者に多い冠詞や前置詞といったエラーに特に焦点が置かれている。また学習者による文の量が非常に多い場合のみ使用可能な設定であるという点から、本研究ではスペイン語学習者のデータを収集する必要があった。

2. 研究の目的

本研究では、日本語母語話者がスペイン語を学ぶ際によく見られる冠詞、前置詞、接続詞等のエラー自動検出システムを設計することを目的としており、主に以下の点について考察していく。

- a) 外国語教師のエラーに対する許容度とはどのようなものであるか、またどのようにそれを改善できるか。
- b) エラーの他に、スペイン語母語話者の文と比べて、日本語母語話者の文中にはどの単語が多用されているか、又は活用されていないか。
- c) 日本語母語話者のスペイン語を分析するためのデータをどこから収集すべきか。
- d) エラー自動検出のために、Constraint Grammar 方式をどう応用することができるか。

3. 研究の方法

- a) スペイン語学習者のエラーに対するスペイン語母語話者の許容度について：
本研究で用いる定冠詞・不定冠詞エラーの登録方式を改善させるための実験を行った。日本語母語話者によって書かれた300の文を4名のスペイン語母語話者に提示し、その中に見られる冠詞の使い方について尋ねた。各文には少なくとも定冠詞、不定冠詞、無冠詞のいずれかの名詞句があり、これらのデータより4名の間に見られる許容度の違いについて分析した。

- b) 機能範疇の言葉の使用頻度について：
日本語母語話者が書くスペイン語学術文と、スペイン語母語話者の学術文に見られる機能範疇の言葉の使用頻度について比較した。具体的には、愛知県立大学のスペイン語を専攻する日本人学生によって書かれた29の卒業論文で構成されているスペイン語学習者のコーパス(およそ6万4千語)を、以下の3つの学術コーパスと比較した。スペイン人用の大学入試で構成されているPAAUコーパス(およそ35万語)、メキシコ国立自治大学(UNAM)の学生による学術

文で構成されているCLAEコーパス(およそ23万語)、そしてスペイン国立遠隔教育大学(UNED)の図書館にある15の博士論文で構成されているコーパス(およそ250万語)である。

- c) 日本語母語話者のスペイン語コーパス分析に必要なデータ収集について：

これまで、愛知県立大学の学生の許可を得て非公式に集められた文と、CORANEコーパスの文を用いた。しかしながらこの2つでは詳細な研究をするには不十分であったため、他のデータを探すことにした。このコーパス不足を早急に解決するため、ウェブ上で公開されているブログから使用可能な文を集めた。またここからブログ執筆の動機や、ブログによく見られる主題、また彼らの言語的特徴についても考察することが可能となった。

- d) Constraint Grammar 方式を、エラー自動検出のためにどのように応用できるかについて：

これまでの経験の中で、特に一致エラー(動詞の活用一致、形容詞変化等)を検出するための式を作成した経験を活用しながら、現在は冠詞エラーを自動検出する式の作成に取り組んでいる。この式により、どのような文脈で誤りが起こるのか、その誤りはどのような特徴があるか、またその誤りが見つかった場合、どのようなタグをその文に付けるべきかを知ることができる。例えば以下の式は、N-PROP-D というリスト内にある固有名詞に必要な限定詞が欠けている場合、(限定詞を付与するという意味の)“%0->el”というタグを付けることを意味する。ここには、La Habana のように常に限定詞が必要な固有名詞が含まれている。

```
ADD (%0->el) TARGET N-PROP-D (-I NOT ART);
```

4. 研究成果

- a) エラーに対するスペイン語母語話者の許容度について：

今回の冠詞に関する実験から、スペイン語母語話者間に見られるエラーに対する許容度はいくつかの要因に左右されるということが明らかになった。第一に、スペイン語母語話者の経験である。外国語教師であるスペイン語母語話者(k=0.79)は、一般のスペイン語母語話者(k=0.76)に比べてわずかに許容度が高く、また一般のスペイン語母語話者は外国語教師に比べてより多くの文を誤りであるとみなす傾向がある。第二に、冠詞の用法が語彙的・意味論的な要因に左右される場合、外国語教師の許容度は上がり(k=0.84)、語用論的な要因に左右される場合は低くなる

(k=0.51)。またスペイン語母語話者間の許容度の差は判断の登録方式にも因ると言える。その理由は、本調査で2人以上の母語話者が冠詞の用法として「どちらとも言えない」を選択したケースが1度もなかったため、「どちらとも言えない」場合ほどその許容度は下がるからである。最後に、次の a) と b) のように明白な文脈がなく、解釈が2通り可能な名詞句がある場合、全員が許容できないと判断した。

- i) 名詞句が「総称としての人物または物事」(冠詞)とも、「不特定の量の何か」(無冠詞)とも解釈できる場合。
 - ii) 名詞句が「不特定の量の何か」(無冠詞)とも、「特定の種であるが不特定のもの」(不定冠詞)とも解釈できる場合。
- b) 機能範疇の言葉の使用頻度について：
以下のような学習者の傾向が見られた。

- より使用頻度の高い語彙を多く用いる傾向にある。例えば、量を表す限定詞の mucho や、pues の代わりに原因を表す接続詞の porque、de acuerdo con の代わりに情報源を指す接続詞の según、tras の代わりに順序を表す成句の después de 等である。
- 関係詞の使用は予想より低く、したがって形容詞節はほとんど用いられない。
- 使用頻度が異なるだけでなく、異なった用法で用いられている機能範疇の言葉もある。スペイン語母語話者は、主に前述のものに言及する場合に、ese や aquel 等の限定詞もしくは指示代名詞、また ello や ellos 等の人称代名詞を用いるのに対し、日本語母語話者はこれらの語をほとんど用いない、もしくは別の用法で多用している。例えば成句である por eso の中で eso を用いたり、主語として人称代名詞 ellos を用いている。
- スペイン語母語話者は学術文において1人称複数形(nuestro, nosotros, nos等)をよく用いるが、日本語母語話者は用いない。

c) 日本語母語話者のスペイン語コーパス分析に必要なデータ収集について：

本研究では2つの方法でデータ収集を行った。1つ目の方法は、41名の学習者による46のブログから短時間でおよそ60万語のコーパスを集める方法である。そこから、これらのブログは相反する性質を持つことから、外国語教師にとって興味深い研究テーマであることがわかった。例えば、ブログは私的であると同時に公的であり、また発話で

あり文章でもあるという独特な性質を併せ持っており、さらに独り言とも会話とも捉えることができるという点である。またブログ執筆者のほとんどが女性であり、日本もしくはスペイン語圏に住んでいることがわかった。主にブログを執筆する動機は、個人的な日記のように日常生活を記すことと、外国人の読み手に日本文化を伝えることの2点にあると読み取れる。こういった文の言語学的特徴については、一定の文法分野(動詞、副詞、人称代名詞)、感嘆詞、俗語、未完了文を多用する等の口語表現の特徴を有していることがわかる。さらに、日本文化について言及する、何かを引用する、強調、コードスイッチングといった様々な用途で、スペイン語文に日本語の単語を使用する点や、絵文字を使用する点も興味深い。

2つ目のデータ収集方法は、第8回国際コーパス言語学会で研究発表をしたのをきっかけに、CEDEL2コーパス(英語母語話者のスペイン語コーパス)を作成したグラナダ大学の研究者たちと共同で手掛けることとなったプロジェクトである。このプロジェクトは日本語母語話者のスペイン語コーパスを構築するものであり、現在その作成に取り掛かっている。

d) 現在では、ブログをコーパスデータとしながら、冠詞エラーを自動検出するシステムの開発に取り組んでいる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計4件)

Valverde, M.P. (to appear), Japanese L1 speakers blogging in Spanish: motivation, topics and linguistic properties. In 8th International Conference on Corpus Linguistics, Easy Chair's EPIc Series in Language and Linguistics, 査読有, 2016.

Valverde, M.P. Frecuencia de uso de palabras gramaticales en textos académicos: Comparación de un corpus aprendices de ELE con tres corpus de referencia. In *Hispanica*, 査読有, 59. 2015. 127-154, 日本イスタニヤ学会.

Valverde, M.P. and A.Otani. Annotating article errors in Spanish learner texts: design and evaluation of an annotation scheme. In Aroonmanakun, W., Supnithi, T. & Boonkwan, P. (Eds.), Proceedings of the 28th Pacific Asia Conference on Language, Information, and Computation (PACLIC 28), Chulalongkorn University. 査読有.

2014. 234-243.
Valverde, M.P. Challenges in the annotation of article errors in Spanish learner texts. In Liu, C.-C. et al. (Eds.) Workshop Proceedings of the 22nd International Conference on Computers in Education. 1st Workshop on Natural Language Processing Techniques for Educational Applications (NLP-TEA-1), Asia-Pacific Society for Computers in Education. 査読有 . 2014. 24-34.

[学会発表](計7件)

Valverde, M.P. A blog corpus of learner Spanish with automatic article error annotation. 8th International Conference on Corpus Linguistics. March 2- 4, 2016. University of Málaga, Málaga, Spain.

Valverde, M.P. Herramientas de la lingüística de corpus para profesores de ELE. Workshop "El léxico en la enseñanza y el aprendizaje de ELE: el uso de diccionarios electrónicos y digitales y algunas herramientas de la lingüística de corpus". CANELA 日本・スペイン・ラテンアメリカ学会, June 21, 2015. Aichi Prefectural University, Nagoya, Japan.

Valverde, M.P. & Otani, A. Annotating article errors in Spanish learner texts: design and evaluation of an annotation scheme. 28th Pacific Asia Conference on Language, Information, and Computation (PACLIC 28), December 12-14, 2014. Chulalongkorn University, Phuket, Thailand.

Valverde, M.P. Challenges in the annotation of article errors in Spanish learner texts. 22nd International Conference on Computers in Education, 1st Workshop on Natural Language Processing Techniques for Educational Applications (NLP-TEA-1), Asia-Pacific Society for Computers in Education, November 30-December 4, 2014. Nara Prefectural New Public Hall, Nara, Japan.

Valverde, M.P. Frecuencia de uso de palabras gramaticales en textos académicos escritos por aprendices japoneses de español: un estudio basado en corpus. 日本イスタニヤ学会 第 60 回大会, October 11-12, 2014. Osaka University, Osaka, Japan.

Valverde, M.P. Overuse and underuse of Spanish function words in

college-level writing by NNS writers of Japanese L1. American Association of Applied Linguistics (AAAL) Conference, March 22-25, 2014. Portland Marriott Downtown Waterfront, Portland, Oregon.

Valverde, M.P. Detección automática de errores en textos escritos por aprendices de español. XXV Congreso de la Confederación Académica Nipona, Española y Latinoamericana (CANELA) 日本・スペイン・ラテンアメリカ学会 25 回大会, May 28-29, 2013. International Christian University (ICU), Tokyo, Japan.

6 . 研究組織

(1) 研究代表者
バルベルデ ピラール (Valverde Pilar)
愛知県立大学・外国語学部・准教授
研究者番号 : 10588205

(2) 研究分担者

(3) 連携研究者