

「21 世紀コーパスの使い方」

第 167 回関西スペイン語教授法ワークショップ (TADESKA) 例会

日時：2023 年 12 月 10 日 (日) 10:30 - 12:20

場所：関西学院大学梅田キャンパス(ハブスクエア) 1401 教室

担当：長由佳

¿Cómo manejar el corpus del siglo XXI?

CLIVII Reunión del Taller de Didáctica de Español de Kansai (TADESKA)

Fecha y hora: Domingo, 10 de diciembre de 2023, de 10:30 a 11:30

Lugar: Universidad Kwansei Gakuin, Campus de Umeda, Aula 1401

Ponente: Yuka OSA KONDO

\*\*\*\*\*

本発表では 21 世紀コーパスの基本的な使い方を、参加者とともに実践的に確認した。21 世紀コーパスとはスペイン王立アカデミーが Web ページ (<https://www.rae.es/corpes/>) に公開しているスペイン語圏における書き言葉と話し言葉のテキストをデータベース化して検索できるようにした **Corpus del Español del Siglo XXI (CORPES)** のことである。また様々なジャンルである地域、年代、男女、本、ウェブ、再生できる音声データなどに **Filtro** (絞り込み) 検索をすることも可能であり、スペイン語のコーパスとしては最大級のものであると考えられる。

21 世紀コーパスの教育的活用の利点としては、今のリアルなスペイン語をできるだけ伝えられることが考えられる。例えば、”tengo”に後続する名詞を探して使用頻度の高い例文を授業で扱うことができる。また、学生からの質問に対しても、教科書にはないが、実際には使用されていると答えることが可能である。そして、メキシコの若者ではよく使われているといったスペイン語の使用について伝えられる、などである。本発表では、21 世紀コーパスを授業で活用する目的で、参加者たちの質問を受けながら、実際に検索を行って丁寧に進めていった。

まず、はじめの項目 **Búsqueda** (検索) には二種類の下位分類があり、**Palabras gramaticales** (文法) と **Palabras ortográficas** (綴り字) のうち、文字や文を入力してそのまま探す **Palabras ortográficas** (綴り字) 検索から扱った。まず探したい”¿Qué tal?”をそのまま入力して検索するとエラーが出ることを確認し、その理由として正規表現としての?が「一字以上」を意味し、”¿Qué tal?”の?は正規表現として解釈されてしまったことを述べた。そして「正規表現ではなく文字として解釈せよ」という意味の/を?の前に加えて”¿Qué tal/?”と検索したところ”¿Qué tal?”が見つかった。このとき追加条件として **Sensibilidad** の **Acentos** と **Mayúsculas** (アクセントと大文字にこだわる) のどちらもチェックを入れない

と、”¿Que tal?”や”¿qué tal?”も検索してしまうので注意が必要である。

さらに他の正規表現についても触れ、\*は「文字 0 字以上」を表し”\*mente”は”mente”や”alegremente”を、“sub\*”は”subjativos”や”subscripción”を検索する。? は「文字を一字以上」で”?specto”は”aspecto”を”??specto”は”respect”を探す。|は「どちらか」を表し”manzana|pera”は「manzana あるいは pera」を、!は「ではない」を意味して”!manzana”は「manzana を探さない」という指示になる。なお正規表現の|（どちらか）と！（ではない）は後でも扱った。

つづいて **Palabras gramaticales**（文法検索）には三種類の下位分類、**Lema**（見出し）と **Forma**（形式）と **Categoría gramatical**（文法範疇）がある。**Lema**（見出し）検索で”comer”を入力すると”comía”や”comiendo”などの全ての活用が、**Forma**（形式）検索では”comer”のみが見つかる。つづいて **Forma**（形式）で”como”を探すと、”como”の動詞と副詞がどちらも検索されるが、その場合 **Categoría gramatical**（文法範疇）で **verbo**（動詞）に絞ることができる。

なお **Categoría gramatical**（文法範疇）については、大きく二十種類（**adjetivo** 形容詞、**adverbio** 副詞、**afijo** 接辞、**artículo** 冠詞、**conjunción** 接続詞、**contracción** 縮約、**cuantificador** 量化詞、**demostrativo** 指示詞、**desconocido** 不明、**extranjerismo** 外来語、**interjección** 間投詞、**interrogativo** 疑問詞、**numeral** 数詞、**posesivo** 所有詞、**preposición** 前置詞、**pronombre personal** 人称代名詞、**puntuación** 句読点、**relativo** 関係詞、**sustantivo** 名詞、**verbo** 動詞）に分類されている。この中の **demostrativo**（代名詞）を取り出し、さらにより細かく **demostrativo determinante** を検索すると”este libro”の”este（この）”など、“**demostrativo núcleo**”を探すと、”este（これ）”などが、さらに詳しく **demostrativo neutro núcleo** と探すと”esto”が検出される。このような詳細な文法範疇についてリストを配って説明を行った。

さらに **Categoría gramatical**（文法範疇）検索で必要な **Operador**（オペレーター）の **no** と **o** についても言及した。まず **Operador**（オペレーター）の **no**（正規表現”!”「ではない」にあたる）を選択すると、「検索条件は何も指示されずに全ての例が検索」され、**Operador** の **no** と **verbo**（動詞）を選ぶと「動詞を除いて検索」される。また **Operador**（オペレーター）の **o**（正規表現”|”「あるいは」にあたる）そして **verbo**（動詞）と **adverbio**（副詞）を選択すると、「動詞か副詞のどちらかを含む」全ての例が表示される。

さらに文法範疇の高度な検索方法として **Añadir categoría gramatical**（文法範疇を加える）について扱った。これまで **Operador**（オペレーター）の **o** で動詞と副詞を選択すると「動詞か副詞」のどちらかが含まれる例が見つかったが、**Añadir categoría gramatical**（文法範疇を加える）で動詞と副詞を選択すると、動詞と副詞のどちらも含む例が見つかる。

この **Añadir categoría gramatical**（文法範疇を加える）を使って、最後に二語以上の高度な検索を行うが、その前に **Distancia**（距離）と **Intervalo**（区間）を理解しておく必要がある。まずは **Distancia**（距離）について、例えば「A から右 2 語目に B」を検索したいとき、B の位置を A からみて <Distancia（距離）1 Derecha（右）> と指示する。この場合

A からみて右 2 語目の B のみを探し、右 1 語目あるいは右 3 語目以降の B や左の B は省かれる。つづいて Intervalo (区間) について、例えば「A から左 3 番目までの B」を見つきたいときに B の位置を A からみて <Intervalo (区間) 3 Izquierda (左)> と検索する。このとき A からみて左の 1 番目 2 番目 3 番目全ての B を探して、左の 4 番目以降の B と右の B を省く。

最後に二つ以上の高度な検索として、まず「ir の活用+a+不定詞」を扱った。まず一語目”ir”は Lema (見出し語) 検索で”voy”や”fueran”などの活用を全て探す、一語目の”ir”からみて右 1 語目 <Distancia (距離) 1 Derecha (右)> に”a”をそのまま探す Forma (形式) 検索をし、一語目”ir”からみて右 2 語目 <Distancia (距離) 2 Derecha (右)> に Categoría gramatical (文法範疇) 検索で infinitivo (不定詞) を全て見つける。

つづいて二つ以上の高度な検索として、「quizá と接続法」を扱った。まず一語目の”quizá”は”quizás”などの異形態があることから、念のため Sensibilidad の Acentos と Mayúsculas (アクセントと大文字にこだわる) を選択せず、Lema (見出し語) 検索をする。つづいて Categoría gramatical (文法範疇) で動詞の Modo (叙法) の subjuntivo (接続法) を選択し、quizá からみて左右 5 語以内に接続法が現れることを <Intervalo (距離) 5 Izquierda (左) Derecha (右)> と指示した。

検索結果の表示方法としては、Estadísticas (統計) Concordancias (コンコーダンス) Documentos (ドキュメント) Coapariciones (共起) Inventarios (一覧) の 5 種類がある。Estadísticas (統計) は使われている地域や国、年代による変遷などがグラフとともに、Concordancias (コンコーダンス) は検索ワードを含む前後の文、Documentos (ドキュメント) は国、地域、ジャンル、タイトル、Coapariciones (共起) は検索ワードと一緒にあらわれる語彙を使用頻度順に、Inventarios (一覧) は検索ワードを使用頻度順に、表されている。Inventarios (一覧) での検索ワードの頻度順については、Categoría gramatical (文法範疇) Forma (形式) Lema (見出し語) でソートすることができ、検索ワード”comer”を、Lema でソートすると、”comer, come, comiendo...”というように使用頻度順に並べられ、Concordancias (コンコーダンス) でも同様にソートが可能である。そして数量的結果として、Número de resultados (実例数) Documentos (文書数) Frecuencia absoluta (絶対頻度) Frecuencia normalizada (相対頻度) が表示される。ここでの相対頻度は、検索語の実例数をコーパスの総語数で割り 100 万をかけて算出する 100 万語あたりの語数のことである。

本発表では 21 世紀コーパスの使い方について、参加者からの質問に答えながら、これまでもよりも詳細な説明を行い、様々な検索機能を使った高度な検索までを扱うことができた。