

Syracuse University

SURFACE at Syracuse University

International Programs

International Programs

8-27-2024

Computational Approaches to Bilingualism in Spanish Education: a Database for Natural Science Students

Federico Ortega Riba

Follow this and additional works at: <https://surface.syr.edu/eli>



Part of the [Education Commons](#)

The views expressed in these works are entirely those of their authors and do not represent the views of the Fulbright Program, the U.S. Department of State, or any of its partner organizations.

Recommended Citation

Ortega Riba, Federico, "Computational Approaches to Bilingualism in Spanish Education: a Database for Natural Science Students" (2024). *International Programs*. 275.

<https://surface.syr.edu/eli/275>

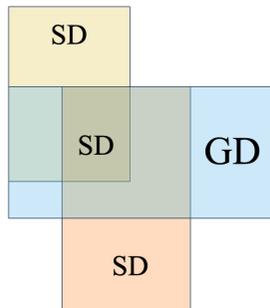
This Poster is brought to you for free and open access by the International Programs at SURFACE at Syracuse University. It has been accepted for inclusion in International Programs by an authorized administrator of SURFACE at Syracuse University. For more information, please contact surface@syr.edu.

COMPUTATIONAL APPROACHES TO BILINGUALISM IN SPANISH EDUCATION: A DATABASE FOR NATURAL SCIENCE STUDENTS

1. INTRODUCTION

One major issue in creating dictionaries, especially for young students is **balancing general and specialized language**. Computational tools can provide the most frequent single and multi-word terms in a specialty field to help students learn. However, in primary education, the most frequent words might not appear specialized at first glance.

Within **specialized discourse**, sub-discourses can develop based on the knowledge of the student. For example, the language level between two ten-year-old students differs from that between a ten-year-old and a seven-year-old. Cabré (1993) argues there is an **intersection between general and specialized languages** due to these pragmatic dependencies.



Specialized discourse (SD) and General discourse (GD). Adapted from Gómez (2005, 46).

2. METHODS

SINGLE-WORDS ✓

MULTI-WORD TERMS ✓



Keyword function in Sketch Engine & logo

- A comparable corpus has been compiled for each grade level of the Spanish Anaya (2019) edition for the subject of **Natural Sciences** and its English version.
- Additionally, each compiled academic unit follows this labeling format: *topic_publisher_language_grade_subdomain_extension*.
- After extracting the units in .txt format for each grade, they were uploaded to the **Sketch Engine (SK)** tool. Each corpus follows this labeling format: *publisher_grade_language*, and both single and multi-word terms in Spanish and English have been extracted.

Lemma	Lemma	Lemma	Lemma	Lemma
1 inerte	11 oralmente	21 ooooh	31 sentidos	41 sastrería
2 oviparo	12 aspereza	22 tallo	32 cascanueces	42 escama
3 vertebrado	13 carnívoro	23 viviparo	33 lupa	43 grapar
4 colorear	14 mural-silueta	24 comunicarse	34 farola	44 inventor
5 coleccionar	15 herbívoro	25 foca	35 hala	45 tic
6 esqueleto	16 ducharse	26 ordenadores	36 lince	46 recuadro
7 dibujar	17 exprésate	27 desenroscar	37 saltamontes	47 reutilizar
8 asear	18 extremidades	28 fijate	38 invento	48 adivinanza
9 olfato	19 pulmones	29 apalabrar	39 omnívoro	49 cepillar
10 invertibrado	20 jardinera	30 rosál	40 balancín	50 chopo

Single words	Frequency
viviparous	6
oviparous	6
omnivore	6
vertebrate	6
invertebrate	6
amphibian	6
reptile	6
herbivore	5
carnivore	5
pollute	5
intestine	5

- The number of keywords has been limited to **120** for first grade and **200** for the other levels, in both the English and Spanish corpora.
- The resulting symmetrical terms in both collections have been compared in Excel to extract those with a **frequency range from 3 to 6** in all corpora.

Above: Keywords function in SK
Left: Results of keywords using the COUNTIF function in Excel

There is **more than a 50% correlation** between the frequency of terms in the Spanish corpus and their equivalents in the English corpus: specifically, **60.31%** for single terms and **69.33%** for multi-word terms.

3. RESULTS

A public domain on the **WordPress.com** platform was used to create the database. The design of the entries is organized as follows:

- Definition and example of use
- Equivalent
- Collocations and example of use
- Image
- References

RegEx

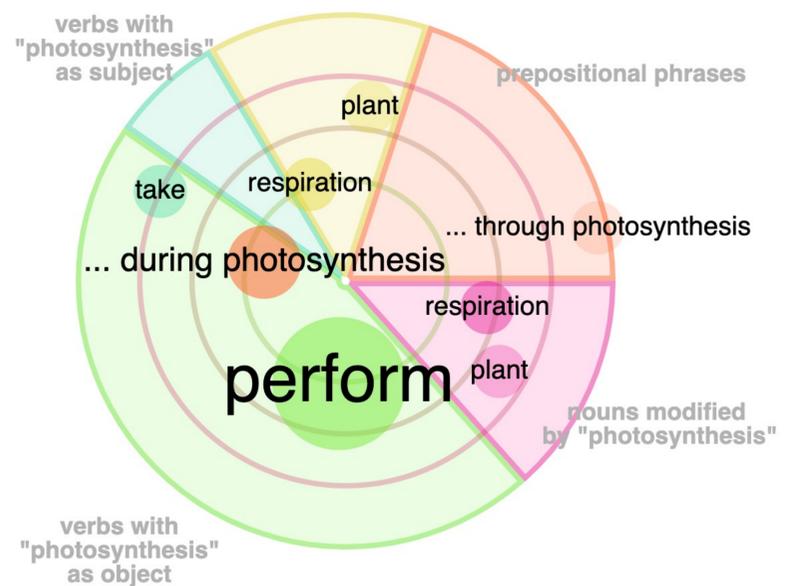
This illustration was made by M0tty

Let us study how an example of entry for the term **fotosíntesis** and its English equivalent is outlined.

1. A search is conducted in **CQL** using the following RegEx: `[lemma="fotosíntesis"]][{0,5}[lemma="ser|generar|transformar|obtener|utilizar|mantener"]]` within `<s/>`.

- Collocations** are extracted using the Word Sketch function.
- Definitions** are extracted from the CQL search.
- The **image** has a Creative Commons license and it is extracted from the internet.

photosynthesis



visualization by SKETCH ENGINE

Above: Concordance function in SK
Bottom: Word Sketch function in SK

4. CONCLUSION

- There is currently a need for the creation of **open educational resources**.
- The database is proposed as a reference work that will address the **limitations of bilingual education** in Spanish Primary Education.
- This research is seen as a **precursor** to a Master's dissertation or even a doctoral dissertation.

5. REFERENCES

