

Soundscapes in English and Spanish: a corpus investigation of verb constructions

SND-ID: snd1159-1. **Version:** 1.0. **DOI:** <https://doi.org/10.5878/xr43-3155>

Download data

SND1159.1.0.zip (896.16 KB)

Citation

Paradis, C., & Caballero, R. (2020) Soundscapes in English and Spanish: a corpus investigation of verb constructions (Version 1.0) [Data set]. Lund University. Available at: <https://doi.org/10.5878/xr43-3155>

Creator/Principal investigator(s)

[Carita Paradis](#) - Lund University, Center for Language and Literature

[Rosario Caballero](#) - Universidad de Castilla-La Mancha, Centre for Languages and Literature

Research principal

[Lund University](#) - Center for Language and Literature

Description

This corpus study explores how SOUND events are communicated in English and Spanish. The aims are to (i) contribute production data for a better understanding of the couplings of meanings and their realizations in the realm of SOUND, (ii) account for typological differences between the two languages, and (iii) further the theoretical discussion of how SOUND is conceptualized through the window of language. We found that while there are significant differences between the languages with respect to how SOUND events are communicated, they are similar with respect to what domains the sound descriptions are instantiated in, namely PERCEPTION (including the more specific domain of sound), MOTION, MANIPULATION, EMOTION-REACTION, CONSUMPTION and COGNITION. One striking difference has to do with the conflation of SOUND FOR ACTION, e.g., creak, squeak, and SOUND FOR MOTION, e.g., slam, crash. This finding supports the received view of English as a language that may lexicalize MANNER in those kinds of verbs, while Spanish expresses MANNER through qualifiers outside the verb. Moreover, both languages employ three different perspectives on the soundscapes: Producer-, Experiencer- and Phenomenon-based. While English favours the Producer perspective, Spanish features an even distribution between Producer and Experiencer. Phenomenon-based descriptions are relatively few in both languages.

A corpus of 951,903 words (415,594 in English and 536,309 in Spanish) with narratives from three different popular genres in English and Spanish, namely fantasy (Throne of glass by Sarah Maas and El último Catón by MThrone of glass by Sarah Maas and El último Catón by Matilde Asensi), romance (Beyond sunrise by Candice Proctor and El tiempo entre costuras by María Dueñas) and thriller (The silkworm by Robert Galbraith and El verano de los juguetes rotos by Toni Hill).

Data format / data structure

[Text](#)

Resource type

Corpus

Foreseen use

Human use

Text corpus

- Linguality
 - Bilingual
- Language
 - English (eng)
Items: 951,903
 - Spanish (spa)
[More..](#)
- Modality
 - Written Language
- Size
 - Items: two pdf documents and two txt files
- Annotation
 - Semantic Annotation
 - Manual annotation

Responsible department/unit

Center for Language and Literature

Contributor(s)

Universidad de Castilla-La Mancha, Centre for Languages and Literature

Commissioning organisation

El Ministerio de Economía Industria and Competitividad de España MINECO - FFI2017-86359-P

Research area

[Humanities and the arts](#) (Standard för svensk indelning av forskningsämnen 2011)

[Languages and literature](#) (Standard för svensk indelning av forskningsämnen 2011)

Keywords

[Cognitive semantics](#), [Sensory perceptions](#), [Hearing](#), [Concept driven](#), [Lexicalization](#), [Motion](#), [Manipulation](#), [Synaesthesia](#), [Embodiment](#)

Publications

Caballero, R. & Paradis, C. (forthcoming) Language and Cognition

Caballero, R. & Paradis, C. (forthcoming) Language and Cognition.

If you have published anything based on these data, [please notify us](#) with a reference to your publication(s). If you are responsible for the catalogue entry, you can update the metadata/data description in DORIS.

Accessibility level

Access to data through SND
Data are freely accessible

Use of data

[Things to consider when using data shared through SND](#)

License

[CC BY-NC-ND 4.0](#)

Versions

Version 1.0. 2020-04-07

CLARIN Virtual Collection Registry

[Add to collection](#)

A virtual collection is connected to a specific research purpose and contains links to data resources from various digital archives. It is easy to create, access, and cite the collection.

Read more about virtual collections on the [CLARIN website](#).

Download metadata

[DataCite](#)

[MetaShare](#)

[MetaShare-CMDI](#)

[DDI 2.5](#)

[DDI 3.3](#)

[DCAT-AP-SE 2.0](#)

[JSON-LD](#)

[PDF](#)

[Citation \(CLS\)](#)

[File overview \(CSV\)](#)

Published: 2020-04-07

Last updated: 2020-04-14