

## **JFLCorp (Japanese as a Foreign Language Corpus): Building a new Spanish L1 – Japanese L2 digital learner corpus following CEDEL2 standards**

Nobuo Ignacio López-Sako and Cristóbal Lozano

University of Granada

The development of second/foreign language (L2/FL) learner corpora has gained momentum in the last two decades. These corpora are nowadays essential in second language acquisition research (Granger et al., 2015; Tracy-Ventura & Paquot, 2020) and the development of teaching/learning resources (Granger, 2017; Hawkins & Filipovic, 2012). However, most learner corpora have focused on L2 English, and East-Asian languages as L2 are still under-represented. Some important exceptions are two subcorpora of L2 Japanese (C-JAS and I-JAS) included in the NINJAL platform (<https://www.ninjal.ac.jp/english/database/subject/jsl/>), the Jinan Chinese Learner Corpus (Huang, Malmasi & Huang, 2015) or the Korean Learner Corpus (Park & Lee, 2016), but they mainly focus on L1 English, the L1 Spanish - L2 Japanese combination being almost inexistent (see the Université Catholique du Louvain [Learner Corpora Around the World](#) database for a full list).

Aiming at filling the above gap, a new project has been launched ([ANACOR R&D project](#)) to build an L1 Spanish - L2 Japanese learner corpus called *Japanese as a Foreign Language Corpus* (JFLCorp) as part of an on-going project featuring a wide variety of L1s (English, Spanish, Japanese, Arabic, Chinese, Dutch, French, German, Greek Italian, Portuguese and Russian) and L2s (English, Spanish, Japanese). The L2 Spanish corpus (Corpus Escrito del Español L2, CEDEL2) (Lozano, 2009; Lozano, under review; Lozano & Mendikoetxea, 2013) is already available at <http://cedel2.learnercorpora.com/>, and the L2 English version (Corpus of English as a Foreign Language, COREFL) (Lozano, Díaz-Negrillo & Callies, 2020) is about to be launched (<http://corefl.learnercorpora.com>).

In our presentation, we will introduce the criteria that have been strictly applied for JFLCorp following the 10 corpus-design principles stated by Sinclair (2005) and adapted to L2 corpora (Lozano & Medikoetxea, 2013). JFLCorp follows the same design principles as the other corpora within the Project (CEDEL2 and COREFL), which allows for multi-layered inter- and cross-linguistic comparisons amongst (sub)corpora. We will illustrate the online data collection method and present the different sections of the digital data-gathering tool, including a test of Japanese grammar to establish the level of competence of the participants. The transliteration criteria (Minami, 1998a, 1998b) will also be discussed. Finally, some preliminary corpus data will be reported and a demonstration will be done of the CEDEL2 search engine to illustrate how the future web-based JFLCorp interface will operate.

### **References**

- ANACOR R&D Project: <https://www.researchgate.net/project/ANACOR-A-corpus-based-approach-to-anaphora-resolution-in-second-language-acquisition-beyond-the-interfaces>
- Centre for English Corpus Linguistics (27-11-2019). *Learner corpora around the world*. Université Catholique de Louvain. <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>
- Granger, S. (2017). Learner corpora in foreign language education. In S.L. Thorne & S. May (Eds.), *Language, Education and Technology, Encyclopedia of Language and Education, 3<sup>a</sup> ed.* (pp. 427-440). Springer. doi 10.1007/978-3-319-02237-6\_33
- Granger, S., Gilquin, G. & Meunier, F. (Eds.). (2015). *The Cambridge Handbook of Learner Corpus Research*. Cambridge University Press. doi.org/10.1017/CBO9781139649414
- Hawkins, J. A. & Filipovic, L. (2012). Criterial features in L2 English: Specifying the reference levels of the Common European Framework. *English profile studies, Vol. 1*. UCLES/Cambridge University Press.
- Lozano, C. (2009). CEDEL2: Corpus Escrito del Español como L2. In C. M. Bretones *et al.* (Eds.), *Applied Linguistics now: Understanding language and mind* (pp. 197–212). Universidad de Almería. Version 1.0: <http://cedel2.learnercorpora.com>
- Lozano, C. (submitted 2020, under review 2021). CEDEL2: Design, compilation and web interface of an online corpus for L2 Spanish acquisition research.
- Lozano, C., Díaz-Negrillo, A. & Callies, M. (2020). Designing and compiling a learner corpus of written and spoken narratives: COREFL. In C. Bongartz & J. Torregrossa (Eds.), *What's in a narrative. Variation in story-telling at the interface between language and literacy*. Peter Lang.
- Lozano, C. & Mendikoetxea, A. (2013). Learner corpora and second language acquisition: the design and collection of CEDEL2. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data*. (pp. 65–100). John Benjamins. Version 1.0: <http://cedel2.learnercorpora.com>
- Minami, M. (1998a). ローマ字表記：ヘボン式と訓令 [Romanization of Japanese: Hepburn y Kunrei]. In Y. Oshima-Takane, B. MacWhinney, H. Sirai, S. Miyata & N. Naka (Eds.), *CHILDES for Japanese. 2nd. Ed.* (pp. 21-32). The JCHAT Project Nagoya: Chukyo Univerisity.
- Minami, M. (1998b). 分かち書き [Wakachigaki]. In Y. Oshima-Takane, B. MacWhinney, H. Sirai, S. Miyata & N. Naka (Eds.), *CHILDES for Japanese. 2nd. Ed.* (pp. 39-43). The JCHAT Project Nagoya: Chukyo Univerisity.
- National Institute for Japanese Language and Linguistics (NINJAL) (27-11-2019). Corpora. National Institute for Japanese Language and Linguistics [Kokuritsu Kokujo Kenkyūjo]. <https://www.ninjal.ac.jp/english/database/type/corpora/>

- Park, J. & Lee, J.H. (2016). A Korean learner corpus and its features. *EONEOHAG: Journal of the Linguistic Society of Korea*, 75, 69-85.
- Sinclair, J. (2005). How to build a corpus. In M. Wynne, (Ed.). *Developing linguistic corpora: A guide to good practice* (pp. 79-83). Oxbow books.
- Tracy-Ventura, N. & Paquot, M. (Eds.) (2020). *The Routledge Handbook of Second Language Acquisition and Corpora*. Routledge. doi.org/10.4324/9781351137904
- Wang, M., Malmasi, S. & Huang, M. (2015). The Jinan Chinese Learner Corpus. *Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications*, 118-123. Denver, CO, USA.