

Data-driven learning in informal contexts?

International Perspectives on Corpus Technology for
Language Learning Seminar Series

Pascual Pérez-Paredes





Informal contexts in DDL?

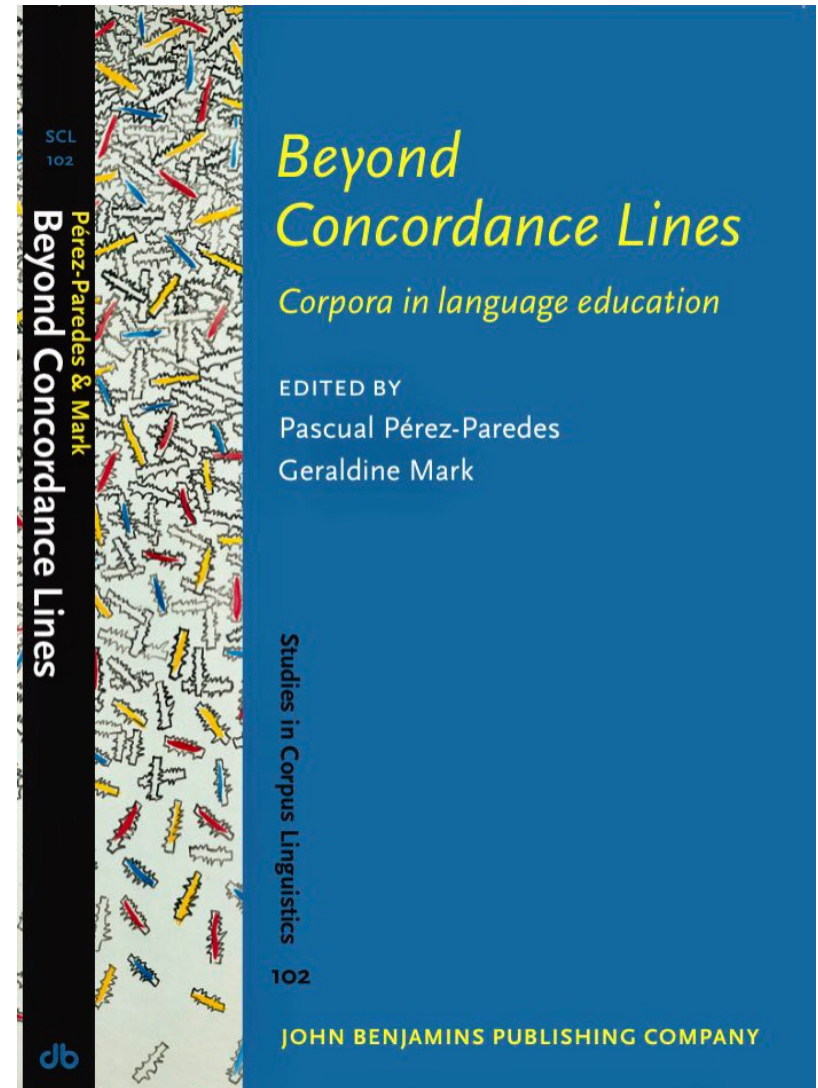
What can they reveal about data-driven learning?

CL, DDL & language learning

- Pérez-Paredes, P. (2010). Corpus Linguistics and Language Education in Perspective: Appropriation and the Possibilities Scenario. In T. Harris & M. Moreno Jaén (Eds.), *Corpus Linguistics in Language Teaching* (pp. 53-73). Peter Lang.
- Pérez-Paredes, P. & Mark, G. (2022). What can corpora tell us about language learning? In O'Keeffe, A. & McCarthy, M. (eds.) *The Routledge Handbook of Corpus linguistics, 2nd Edition*. Routledge.

In this talk

- Some thoughts on DDL in 2022
- Two scenarios for DDL in informal contexts: statistical learning & use of apps
- Concluding remarks





The scope of DDL in 21st century

Some ideas to begin with

Boulton (2021)

- Room for greater originality rather than repeating existing work with minor variations, partly due to ignorance of large quantities of work leading to reinventing the wheel (Gillespie, 2020, p. 128).
- Needs: free resources (as opposed to producing and testing one's own corpora or software), including even the web-as corpus; DDL for mobile tools; 'regular' teachers who are not the researchers/authors for greater integration into existing courses; and outside university classrooms, whether during a course or at some later time.

Boulton & Vyatkina (2021)

- Long-term, higher-level, non-language skills such as critical thinking, independent learning, and learner autonomy feature as alleged benefits from DDL BUT no direct exploration of these concepts as research objectives.
- The main procedures—examining concordances and frequency information—have remained largely unchanged since the early research days.
- Research has overwhelmingly been conducted with university students in language-for-general-purposes classes.

O'Keefe (2021)

- Teacher mediated learning << >> student mediated learning
- Criticisms of the free-range approach include its unanchored nature relative to language syllabi (where does it fit with syllabi?), the lack of control over the learning outcome and the teacher's loss of control (how can the teacher control what the students are using the corpus for?) (see Bernardini, 2004; Mukherjee, 2006; Kennedy & Miceli, 2017).

O'Keefe (2021)

- The UB model:
 - 1. frequency (the amount of times a construction is experienced and used)
 - 2. recency (the more recently we experience a construction, the stronger our memory of it)
 - 3. context (a given context triggers an association and mental categorisation of a frequently experienced construction)
- Central to DDL has been the notion of grappling with raw data (Cobb, 2005). As Cobb notes, the DDL paradigm offers a methodology to help perform this grappling through adaptive tools and methods.

DDL (Chambers, 2022)

- A DDL approach implies a level of active participation in the learning process, which corpus-based dictionaries and grammars cannot provide in the same way as learner (and teacher) interaction with the corpus itself.
- DDL involves either indirect access, where they learn about language use by studying concordances prepared in advance by the teacher, or direct access, where they use concordancing software to carry out their own searches of corpora.
- Thus DDL, while encouraging learners to take an active part in their learning, is also contributing to a revival of interest in frequency, not the mindless repetition of behaviourism, but rather what Ellis (2002: 177) calls 'mindful repetition in an engaging communicative context by motivated learners'.



Is DDL in line with how learners access information?



- Today's learners exist within a digital world, or are 'digital natives' (Prensky, 2009)
- While DDL was intended to "cut out the middleman" (i.e. the teacher) between the learner and the data (Johns, 1990, p. 18), in many classroom contexts this has already happened
- DDL tools need to evolve to accommodate the ways learners are used to sourcing information, rather than DDL practitioners trying to force learners to adopt KWIC concordancing
- DDL approaches where concordances are the start and finish of DDL pedagogy should be supplemented, augmented – or even replaced entirely

Source: <https://www.youtube.com/watch?v=SRIaoJ8faWU>

Conclusion



- DDL is certainly not 'dead': increasing numbers of research publications over time and positive results of various syntheses and meta-analyses in recent years.
- Nonetheless, we could be doing more to bring it into the mainstream – where that is warranted - but this depends on local conditions and immediate and long-term goals for teaching and individuals
- Nobody should expect DDL to be a panacea across the board.
- DDL research may be treading water, but that there is plenty of scope for innovation if we only take a step back and look at new possibilities without the blinkers of dogmatic reflexes such as “that isn’t DDL!”

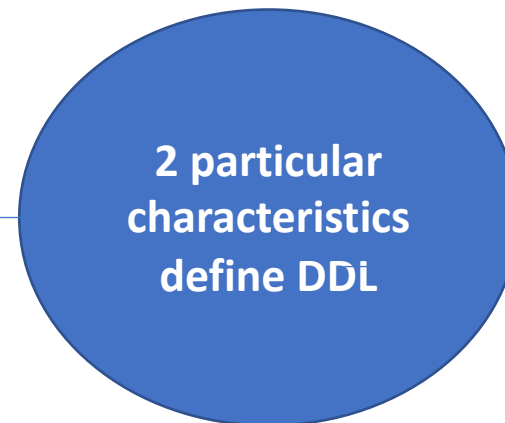


A new equation?

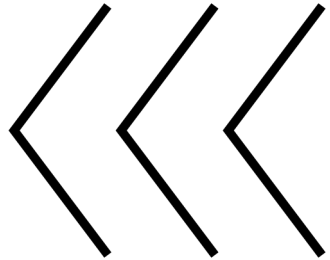
Computers, corpora, mobile devices &
language learning in the wild

Smart (2014)

- 1) real language data are used as sources of language learning materials or reference resources;
- 2) learning activities are student-centered and focus on language discovery.



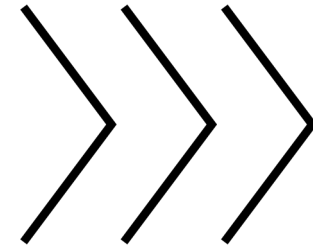
Highly controlled
activities in
instructed contexts



learning

Experiential learning

Learning in the wild



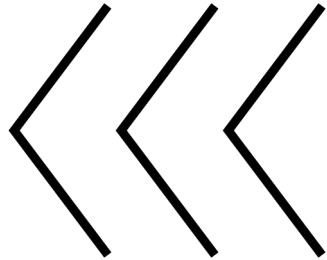
Statistical learning
(SLA)

Highly controlled
activities in
instructed contexts

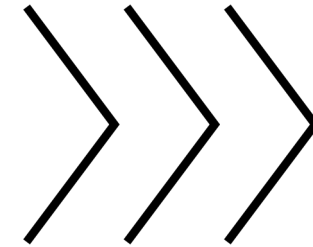
MALL
(CALL)

Experiential learning

Learning in the wild



learning



A woman with long dark hair is sitting on a brown couch in a living room. She is holding a smartphone in her hands and looking at it with a smile. A laptop is open on her lap. The background shows a cozy living room with a fireplace, a framed picture on the wall, and a shelf with various items. The text "Informal language learning" is overlaid in white on the image.

Informal language learning

Introductory remarks

Informal learning (Godwin-Jones, 2020)

- Informal language learning underrepresented in L2 research
- Informal language learning “in the wild” is difficult to analyze because it’s difficult to document.
- Input in informal environments is not structured.
- Chik (2020). 5 main dimensions: location, formality, pedagogy, locus of control, and trajectory.

Kukulska-Hulme & Lee (2020: 170)

Locations outside the classroom represent social spaces or “settings” that offer a variety of affordances for language learning (Benson and Reinders 2011), but these opportunities will not be fully realized unless we make efforts to propose and try out new designs for learning in these settings.

Researching statistical learning

Scenario 1

Detecting regular patterns

Boulton (2010, p. 535): constructivism leading to learners using adaptive behaviour in detecting regular patterns in the data that are meaningful to them, rather than attempting to learn and apply rules they are given, a more 'artificial' intellectual activity.

Abstracting categories from ambient input

...the process of language learning means first abstracting constructions from the meaningful input and then understanding the relationships between constructions. This is largely determined by frequency – the more often constructions are experienced and understood together, the more entrenched they become. It is predicted that learners subconsciously acquire first the constructions that they come across most frequently in the input that they receive.

Pérez-Paredes, O’Keeffe & Mark (2020)

Cognitive advantages in adults & young adults

- Statical learning has been described as automatic, incidental, and spontaneous.”
- Domain-general probabilistic mechanisms
- Abilities to extract statistical regularities from the ambient language input and to apply them to L2 grammars (Hudson., Kam & Newport, 2005; Scovel, 2000).
- Metalinguistic awareness favours L2 learning (learners’ variation)
- Metalinguistics awareness helps SL categorize L2 input.
- Metalinguistically aware learners are more likely to id underlying rule-governerd systems.
- Multilinguals are particularly well prepared for this.

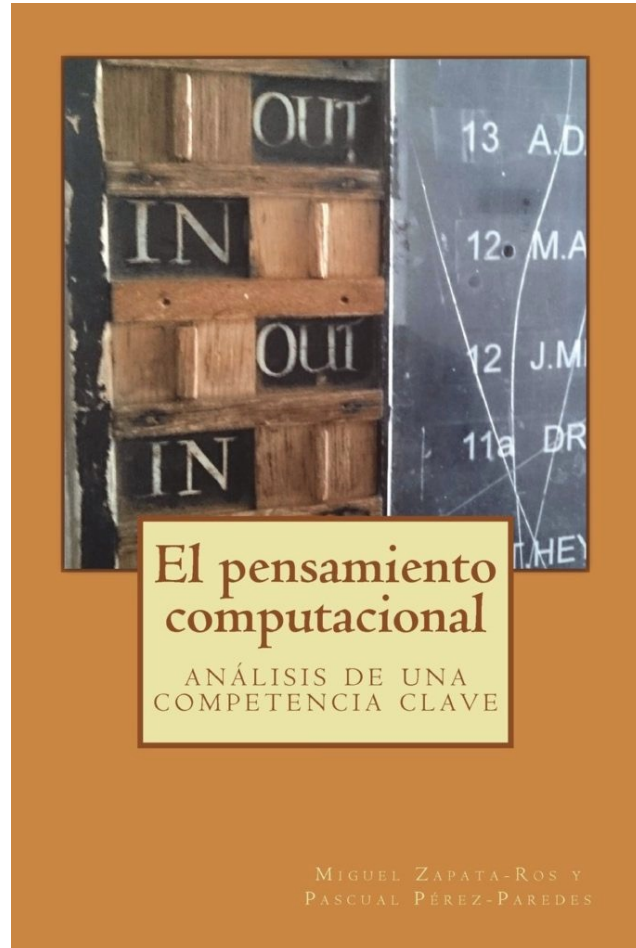
Statistical Learning

- Research has shown that both explicitly reinforced learning and implicit learning both facilitate learning (Shin and Christianson, 2012)
- Ellis (2005) both explicit and implicit learning and memory interface and suggest different level of linguistic representation.
- Adults can handle larger input data sets in language learning (Lev-Ari, 2018)
- Adults are better (than children) at integrating multiple sources of info in L2 learning
- Virtual learning environments:
- Virtual informal environments pose substantial opportunity not only for language learners, but also for researchers” (Christianson & Deshaies, 2020: p.34).

Noticing is poorly treated in DDL research (Pérez-Paredes, 2020 TaLC plenary)

- Generalization (domain general)
- Search for regularities and structures
- 2 abilities: perceptual ability and their symbolic inferential ability.
- Initial representations of the pattern (e.g., pictorial, verbal and symbolic) can influence students' performance.
- The development of specific language that assists students to describe/represent patterns
- Three types of generalization that emerged from the exploration of figural pattern tasks: factual; contextual; and symbolic (Radford, 2006).

Computational thinking and DDL



SL and language learning: new (roles for) corpora?

- Digital spaces: Fan fiction reading and writing (Sauro and Zourou, 2019; Sauro, 2020)
- Collaboration spaces outside the classroom and MALL (Kukulska-Hulme & Lee (2020) >> Díez-Bedmar & Pérez-Paredes (2012)
- Video and streaming (Vanderplank, 2020)

Yus (2021)

- A positive development of new digital narratives is the fact that they offer a “potential of online storytelling for enhanced interactivity and for multiple modes of user involvement that can decisively shape a story” Georgakopoulou 2013 : 700).
- Only advanced users who are capable of obtaining all the reward from these advances will be pleased to see evolutions in these interfaces and happy to be granted a more active role in these new digital narratives.

A hand holding a black smartphone with a white home button. The screen shows a grid of colorful app icons. The background is a dark, semi-transparent overlay on a light-colored surface with architectural blueprints and a pen.

Researching mobile (apps) use

Scenario 2

Using mobile devices for language learning

- The promise of technology: there is an app for everything
- Self-directed uses : (1) personalization and autonomous, customized learning; (2) mobile devices used to engage in authentic learning experiences; and (3) connectivity, where learners use mobile devices to connect with speakers of the L2. (Lai and Zheng, 2018)

Godwin-Jones (2020)

- Mobile apps development (Refugees, migrants, displaced populations)
- Mobile devices provide real-world contexts for language and culture learning.
- Superdiversity (Blommaert, 2021) >>> Multilingual realities where the L1/L2 binary opposition has lost its validity.

DDL & (mobile/online) language learning

- Quan, Z. (2016). Introducing “mobile DDL (data-driven learning)” for vocabulary learning: an experiment for academic English. *Journal of Computers in Education*, 3(3), 273-287.
- Pérez-Paredes, P., Guillamón, C. O., Van de Vyver, J., Meurice, A., Jiménez, P. A., Conole, G., & Hernández, P. S. (2019). Mobile data-driven language learning: Affordances and learners' perception. *System*, 84, 145-159.
- Crosthwaite, P. (2020). Taking DDL online: Designing, implementing and evaluating a SPOC on data-driven learning for tertiary L2 writing. *Australian Review of Applied Linguistics*, 43(2), 169-195.

Boulton (2021)

- One problem here is that many test instruments are highly constrained and artificial; more work on user-generated queries is crucial.
- Longer, experimental studies are present, but the methodological design does not sufficiently cater for this: longitudinal and delayed post-tests are essential if we are to test the claim that DDL leads to long-term learning, along with other alleged benefits such as increased autonomy, noticing and language awareness, and that the skills acquired can be transferred to new, self-initiated areas.

User generated DDL

- Web services
- AI & machine learning: exponential GROWTH in services and affordances
- Instant feedback
- Learners' interests
- Learners' needs

Reimagining “the learner” in DDL

Socially contextualized MALL places the learner’s lives and their digital and material experiences as the central point for their language learning (Leung, 2005; Wagner, 2019) and favours a usage-driven and user-centered L2 pedagogy (Eskildsen & Theodórsdóttir, 2017).

In the context of language learning and language education, governments and educational institutions retain almost absolute power when deciding what counts as competence, and institutional ideologies about language learning (Farr & Song, 2011) determine how learners frame their experiences about instructed language learning and how artefacts can possibly mediate the use of technology for language learning. Pérez-Paredes & Zhang (forthcoming)



ELEXIS EU @elexis_eu · 21s

The [#ELEXIStool](#) Lexonomy is a dictionary writing & publishing system enabling users to easily

- ◆ create a new dictionary - push/pull lexicographic data from [@SketchEngine](#)
 - ◆ clone/edit entries
 - ◆ customize your configuration
 - ◆ publish your [#dictionary](#)
- 👉 lexonomy.elex.is



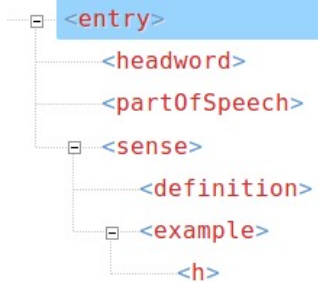
1



Měchura, M. B. (2017) '[Introducing Lexonomy: an open-source dictionary writing and publishing system](#)' in *Electronic Lexicography in the 21st Century: Lexicography from Scratch*. *Proceedings of the eLex 2017 conference, 19-21 September 2017, Leiden, The Netherlands*.

Save

Cancel



Element

entry

Attributes

Add...

Content

☒ Child elements ☐ Text ☐ Text with markup ☐ Value from list ☐ Empty

Child elements

<headword>	min	<input type="text" value="1"/>	max	<input type="text" value="1"/>				
<partOfSpeech>	min	<input type="text"/>	max	<input type="text" value="1"/>				
<sense>	min	<input type="text" value="1"/>	max	<input type="text"/>				

Add...

How it works

Automatically extract data from website with 1 click



1

Enter URL

Manual input or automatic generation,
unlimited URLs and pages



2

Click Start

ScrapeStorm will automatically detect
content and next page button



3

Download

You can choose to download results
automatically or manually

Pushing the boundaries of DDL

- Perhaps this is the main challenge ahead for learner corpora: to imagine a method (a discipline perhaps?) that builds on Granger's (2015) reconceptualisation of CIA, but at the same time can successfully relate to current language learning characterised by the overlapping of material and digital social contexts in a multilingual world (Douglas Fir Group 2016; Ortega 2013) across language learning theories. (Pérez-Paredes & Mark, 2022)
- O'Keeffe (2020, 2021): We will see more in the process of adopting a more pedagogically-aware and SLA-focused approach to researching DDL and by doing so we can open up opportunities for SLA to use DDL within its methodological repertoire.
- Opportunities for the reconceptualization and expansion the concept of:
 - Corpora
 - Learners' engagement with (linguistic) data
 - Widening the scope of research methods and research designs

Some references

- Aslin, R.N. & Newport, E.L. (2014), Distributional Language Learning: Mechanisms and Models of Category Formation. *Language Learning*, 64: 86-105.
- Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology*, 25(3), 66-89.
- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language learning*, 60(3), 534-572.
- Boulton, A. (2021). Research in data-driven learning. *Beyond Concordance Lines: Corpora in language education*, 9.
- O’Keeffe, A. (2021). Data-driven learning, theories of learning and second language acquisition. *Beyond Concordance Lines: Corpora in language education*, 35.
- Pérez-Paredes, P. & Zhang, D. (forthcoming). Mobile assisted language learning: scope, praxis and theory. *Porta Linguarum*.



Thanks!!!

@perezparedes

Data-driven learning in informal contexts?

