Journal of Technology and Chinese Language Teaching Volume 9 Number 1, June 2018
http://www.tclt.us/journal/2018v9n1/colpaert.pdf pp. 1-14

Exploration of Affordances of Open Data for Language Learning and Teaching

Jozef Colpaert
Universiteit Antwerpen
(安特卫普大学)
jozef.colpaert@uantwerpen.be

Abstract: The new wave of huge data sources, also referred to as big data, entails new forms of information and content for Computer Assisted Language Learning (CALL). Open data is the latest appeal in the open data philosophy and trends. In this article, we explore the affordances of (linked) open data for education in general and for CALL in particular. Learning analytics, adaptive learning environments, and open educational resources are the obvious beneficiaries, but contextualization of the learning process appears to be a relatively unexplored affordance. And this while available literature already touts its merits.

Keywords: Open data, contextualization, computer assisted language learning (CALL)

1. Introduction

The most vulnerable aspect of Computer Assisted Language Learning (CALL) as a discipline is its multifaceted nature (Levy, 1997), which I have coined the "pluriness" of CALL (Colpaert, 2018). CALL is dealing with a multitude of different disciplines, actors, contexts, theories, cultures, languages, skills, levels, subjects, technologies, learning goals, content types, and standards.

The underlying foundations—technology and pedagogy—are subject to rapid changes. The technological evolution is obvious: from mainframe, personal computer, multimedia CD-ROM, networking, Internet, mobile devices to current smart devices. Recent pedagogical paradigms are mushrooming, rooted in theories and approaches such as Constructivism (sociocultural theory and activity theory, among others), Project-Based Learning, Connectivism, Problem-based Learning, Connectionism, Task-Based Language Teaching (TBLT), Deep Learning, Complex Dynamic Systems, Self-Determination Theory, Universal Design for Learning, Dörnyei's L2 SELF model, 4CD/ID (van Merriënboer & Kirschner, 2007) Complex Dynamic Systems (Ellis & Larsen-Freeman, 2009), skills (Bloom digital taxonomy, 21st century skills, and higher order thinking skills), or technology (TAM, TPACK, and SAMR).

A third stream, data, strongly linked to available technology, is currently changing the CALL landscape imperceptibly. Such phenomena include big data, smart cities, the Internet of Things and augmented reality. In this context, a recent movement has emerged which has not been linked frequently yet with language learning and teaching: open data. In this article, we explore the affordances of open data for CALL.

As this article is a continuation of an earlier article in this journal (Colpaert, 2016), and a small step in our long-term research project, we apologize for the inevitable extensive self-referencing.

2. CALL and Data

We will first provide an overview of the various instantiations of data in the world of CALL, and the problems associated with some of them.

Traditional textbooks by publishers (Decoo, 2010), or self-authored materials produced by teachers for use with their own students only, have been the predominant language-learning content in the 20th century.

Since the advent of affordable digital technology in the 1980s, and of the Internet around the turn of the century, more artefacts—developed by design for education or not—have emerged. Such artefacts include: authentic documents found on the Web, adapted (easy) readers, corpora, interactive (tutorial) courseware, Massive Open Online Courses (MOOCs), Open Educational Resources (OERs), learning analytics, virtual worlds (such as Second Life and OpenSim), serious games, augmented reality (the view of a real-world environment whose elements are augmented by sensory input such as sound, video, graphics, or GPS data), and the Internet of Things (real-world objects and artefacts which carry readable data that can be used as content in tasks).

This wide array of educational artefacts opens up many possibilities for language teachers worldwide. But at the same time, we already know that affordances also entail limitations and challenges, certainly for interactive courseware, MOOCs, OERs and learning analytics. Interactive language courseware, also called tutorial CALL, has become an endangered species due to the labor-intensiveness of content and software development, the complexity of the required linguistic-didactic functionality, and the lack of generic, reusable, and exchangeable content (Colpaert & Decoo, 1999; Colpaert, 2004; Colpaert, 2013). MOOC platforms, such as Moodle and OpenLearning, allow teachers to create and use language courses that are accessible worldwide. But to what extent are these materials massive, open, online, and interactive enough to be called a genuine course (Colpaert, 2014)? OERs are materials that are shared, reused, improved and shared again. They are supposed to reduce workload for teachers and to increase learning efficacy considerably. But their success seems to be hampered (Colpaert, 2012) by various challenges, namely psychological ("what will others say about my content?"), technological ("what should I use to share my content?"), epistemological ("what does 'open' mean exactly?") and juridical ("is Creative Commons enough to protect my content?'"). Finally, learning analytics: in the early years of CALL, referred to as "Tracking and Logging" and later known as (e-)portfolios (Little, 2002), the latest trend is in gathering and collecting data about the learner. But how should we analyze and use all these data? Is it just about testing and certification? Or should we not focus more on how learners build knowledge together in their cultural and social settings (Ferguson & Buckingham Shum, 2012)?

3. Data, Information, and Content

This brings us to the definition of data. As already explained in Colpaert (2016), data are series of raw tokens which can turn into information when they reduce uncertainty (Shannon & Weaver, 1949), or into content when they contribute to learning.

We distinguish four types of information:

- Information about the learner (learn**er** analytics): information provided by the learner, the teacher, the school, parents, e-portfolios, social networks (e.g. through data scraping) and smart phones (e.g. by providing the learner's geotemporal location). This information can be useful and even necessary for the adaptation and personalization of the learning process.
- Information about the learning process (learn**ing** analytics): data gathered by a system or electronic learning environment with a view to: a) assessing and supporting the learner; b) analyzing the learning process; c) improving the learning environment, and; d) predicting learner behavior (in the case of big data or educational data-mining on a large scale).
- Pedagogical metadata: sets of data that may facilitate the reusability and discoverability of digital learning resources, with a view to supporting the learner (hyperlink glosses, captions, just-in-time information, and procedural information) or the teacher (tags, readability indexes, CEFR levels, and pedagogical instructions). These metadata are mostly integrated in content for learning or teaching.
- Research data: with recent open access policies also comes the tendency to facilitate access to the datasets with any publication.

We distinguish four types of content:

- Published materials: textbooks, courseware, and MOOCs.
- Self-created (or co-authored with colleagues or students) materials: OERS, LMS-embedded exercises, sound files, subtitles, captions, corpora, knowledge clips, fan fiction, and textual or audiovisual content produced in online communities of practice.
- Authentic documents found on the Web, especially the semantic web, or level-adapted materials (e.g., easy readers).

- Content found in virtual worlds, serious gaming, ambient intelligence, augmented reality and the Internet of Things.

This availability of huge amounts of data for language learning and teaching under the form of information or content is often called big data. But how discoverable, accessible, usable, useful, interoperable, reusable, and sustainable are these data? A recent phenomenon sheds new light on this question: open data.

4. The Open Philosophy

Before exploring open data in more detail, it is interesting to mention other frequent collocations with "open."

The term "open" probably appeared for the first time in a discussion about open versus closed courseware in the 1980s (Van Elsen et al., 1991). In closed courseware, learners and teachers were able to make sophisticated selections, but they were not allowed to make any changes to the content itself. With open courseware (such as Adam and Eve and Hot Potatoes later), teachers could enter their own content.

The term "open source" originated in the context of software development to designate a specific approach to creating and distributing source code for computer programs. It is known as the Open Source Initiative[1]. It mainly applied to licensing and redistribution aspects. Currently, it also refers to any artefact users can modify and share because its design is publicly accessible. It "designates a broader set of values—what we call 'the open source way.' Open source projects, products, or initiatives embrace and celebrate principles of open exchange, collaborative participation, rapid prototyping, transparency, meritocracy, and community-oriented development" (c.f., Open Source Way[2]).

In this respect, the "open" attitude was applied very quickly to content development in general and in educational content development more specifically. Open educational resources "are any type of educational materials that are in the public domain or introduced with an open license. The nature of these open materials means that anyone can legally and freely copy, use, adapt and re-share them. OERs range from textbooks to curricula, syllabi, lecture notes, assignments, tests, projects, audio, video and animation" (UNESCO, 2018).

Open education is a philosophy and a movement about the way people should produce, share, and build on knowledge. Everyone in the world should have access to high-quality educational experiences and resources, and proponents work to eliminate barriers

[1] c.f. https://www.opensource.org
[2] c.f. https://www.opensource.com

to this goal. Open universities, open courses, such as OpenLearn, and many publicly accessible MOOCs are concrete instantiations of this movement.

Open access (OA) refers to online research outputs that are free of restrictions for access and free of restrictions for use. Open access can be applied to all forms of published research output, including articles, papers, theses, book chapters, and monographs. The FAIR principle summarizes the criteria: research findings should be Findable, Accessible, Interoperable and Reusable.

Many other collocations arise in this respect, such as open standards, open government, open science, open gaming, open hardware, and open knowledge.

5. Open Data Initiatives

The Open Knowledge Initiative[3] is a global non-profit organization focused on realizing the value of open data to society by helping civil society groups access and use data to take action on social problems. Open Knowledge International does this by showing the value of open data, by providing organizations with the tools and skills to effectively use open data, and by making government information systems responsive to civil society. “Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)” (Open Definition, 2018).

The Global Open Data Index[4] is an annual effort to measure the state of open government data around the world. The crowdsourced survey is designed to assess the openness of specific government datasets according to the open definition.

The Belgian Open Data Initiative[5] offers more than 7,000 datasets, including environment, public sector, science, technology, economy, finance, population, transport, regional matters, culture, sports, energy, health, education, international matters, agriculture, fisheries, and justice.

China is opening up as a potential huge data repository for information and content. “With the release of the Big Data Development Action Plan by the State Council in September 2015, open data was officially recognized and listed as one of ten key national projects in the plan. The Plan not only gave a concrete timeline for opening government data, but also explained the motivation behind the national initiative: it is expected that big data and open data may drive economic transition, improve China’s competitiveness, and create a modern governance model” (Ma, 2017).

Open Data China[6] is an advocate for open data by providing information on open data initiatives, supporting training and dissemination activities, and by stimulating

[3] c.f. https://okfn.org/about
[4] c.f. https://index.okfn.org
[5] c.f. https://data.gov.be/en
[6] c.f. http://opendatachina.com/en/

innovation for citizens. Many challenges, however, lie ahead, such as the fact that open data initiatives are limited to major cities (Liu et al., 2015), and “the absence of clear guidance and strong enforcement, the lack of transparency culture in the government, weak data literacy among both officials and the public, and limited tracking of data usage and impact” (Ma, 2017).

Opening up existing data sets is just the first step and does not automatically lead to a democratic government: “Publishing open data is of course not sufficient for open governments or open societies. It is just one ingredient in the mix, and no replacement for other vital elements of democratic societies, like robust access to information laws, whistleblower protections and rules to protect freedom of expression, freedom of the press and freedom of assembly” (Jonathan Gray, as cited in Ross, 2015).

6. (Linked) Open Data

Before looking into the affordances of open data for education in general and for Computer Assisted Language Learning in particular, let us first have a closer look at the more technical side of open data, without losing ourselves in irrelevant details.

Strangely enough, open data are not open to the extent that we can access a data source and change its content. Open data is based on the principle that a system or app contacts a data source using its URI (Uniform Resource Identifier) and RDF (Resource Description Framework).

The client system requests a dataset from the data server by calling an API (similar to web services) or by “scraping” the data source. The discussion between advocates of both approaches is still quite lively. Data sources can be in various formats, such as XLS, JSON, CSV, XLSX, HTML, PDF, XML, ODS, or TXT. The client system can now use these data and link them to other data sources. These linked open data can become complex and powerful sources of information for new services. The Linked Open Data Cloud[7] shows a diagram of all linked open data, also per subcloud such as geography, government, life sciences, linguistics, media, publications, social networking, or a user-generated subcloud. A good example is DBpedia[8], a public data infrastructure for a large, multilingual, semantic knowledge graph.

Besides the mere legal requirement of the open definition mentioned above, the goal of open data has also been described as maximizing the reuse of data. The more machines can access public datasets, the more value it creates for the public.

When creating applications or answering queries with public datasets today, a developer will make a pragmatic decision on reusing a limited set of datasets. Each dataset

[7] c.f. https://lod-cloud.net
[8] c.f. http://wiki.dbpedia.org

comes with a certain cost to adopt, even when publicly and freely available, as the developer needs to invest time to understanding how to reuse the data. The field of open data research has been characterized by looking for how technical principles can foster data adoption, cost-efficiently scaling up the average number of adopted datasets per application. This research happens on the four data interoperability levels: legal, technical, syntactical, and semantic. Principles such as the FAIR principles, the data on the Web best practices, or the linked data principles, advocate raising interoperability when publishing data on the Web (Colpaert, 2017).

The idea of open data, now well adopted by governments globally as evidenced by the many open data portals that can be found worldwide, ultimately puts forward a vision where the data published on the Web by many organizations can create power for the many, not the few.

7. Affordances of (Linked) Open Data for Education

The (linked) open data phenomenon entails several affordances for education in general. Large learning management systems, interactive courseware applications, and mobile apps will benefit from open data as new functionalities will be added and new systems will be conceptualized.

On the level of data as information we can distinguish:

- *Personalization* of the learning process: adaptation of parameters such as level of difficulty, activity type, speed, topic, feedback, and help functionalities on the basis of open learner data (Walkington, 2013; Brusilovsky & Millán, 2007). This adaptation is formulated in the form of explicit rules.
- *Intelligence* of the learning process: traditional neural networks become far better performing through the availability of huge data sources, which give a boost to Artificial Intelligence (Luckin et al., 2016). The intelligence of a system in terms of complex tutoring depends on the data of many learners. Rules remain implicit.
- *System upgrades* on the basis of extensive data mining about learner behavior and performance (Koedinger et al., 2008).
- Better *integration of research findings* in education and in the development of educational systems.
- Evidence-based *educational policy*: all data sources combined (learner, school, and community) should allow policy makers to draw up a more justifiable education policy plan.

On the level of content we can distinguish:

- By linking existing content to other data sets, learning content can be *enriched.* A typical example would be to link sentences of existing exercises to grammar topics in another data set.
- Existing content can also be *reused* in order to generate new systems, services, and products. Just as textbook content has been transformed into CD-ROM databases for interactive courseware and later into SQL repositories for online distributed systems, this content can now also be regenerated into other products as open data. The generation of mobile (and even personalized) exercises on the fly based on old textbook content is technologically speaking quite straightforward.
- Open data may rejuvenate the old concept of *open educational resources*. We mentioned epistemological, psychological, technological, and juridical barriers for OERs to break through. New systems based on an open data approach may offer a safer, comprehensible, and trusted environment for finding, accessing, using, editing, and sharing learning content.
- *Contextualization* of the learning process. Depending on the geotemporal location of the learner, a system can look up content, interactive or not, from open data sources, which is relevant for learners regarding their learning goals, interests, and preferences. This approach fits perfectly within current phenomena such as Augmented Reality, Smart Cities, and the Internet of Things.

In recent literature on the topic, the main focus is on two of these functionalities: learning analytics and open educational resources (Mouromtsev & d'Aquin, 2016). Let us now have a closer look at affordances for CALL.

8. Open Data for CALL: Focus on Contextualization

The considerations on affordances for education seem to be valid for CALL as well. Literature on the topic also mainly focuses on learning analytics and OERs (e.g. De Meester et al., 2018). As far as Chinese is concerned, two examples are worth mentioning. Zhishi.me is an effort to build Chinese linking open data, covering the three largest Chinese encyclopedias: Baidu Baike, Hudong Baike and Chinese Wikipedia (Niu et al,. 2011), and an online database for Chinese verb compounds (Zhan et al., 2015).

There is, however, one function that further deserves our special attention: the contextualization of the learning process.

Let us first look at an example of a new, imaginary app for Contextualized Language Learning:

I am in the neighborhood of a fisherman's village on the coast. It is about noon. The app gives me information on the type of fish sold here. It

> *explains that I first have to buy the fish and then have to take it to one of the local restaurants across the street. It provides reviews of the restaurants so I can choose. It explains that the waiter, who does not speak English, will ask me if I want knife and fork or chopsticks, and if I want my fish steamed, grilled, or cooked. The waiter will also ask if the fish should be cut in small pieces or not, and served with cheese or not. It will explain how to do the dishes with tea. Finally, it will explain how to ask for the bill and thank the staff.*

This app will not only turn my meal into the best possible lunch, but also into an experience with maximum exposure and learning effect.

The context of the learner can be defined as the collection of data available for, about, and at a specific location at a given moment in time. These data can become information or even content. Information enables the learner to perform a task and to focus on task execution (e.g., when the learner knows the departure time, bus number, and platform number, she or he will be able to take that bus), while content refers to data the learner has to process in some way.

Contextualization refers to the adaptation of the learning process to the specific context of the learner. Mobile devices such as smart phones and tablets can establish our location, date, and time exactly, while related services can inform us about weather and traffic conditions, cultural events, where to eat, and what to visit at that specific location (de Jong, Specht, & Koper, 2008). When we approach a railway station, these devices can automatically list arriving and departing trains, depending on our plans. When we are in the neighborhood of an Italian restaurant around noon, they can show the lunch menu, even when the learner does not ask for it. Given the potential of current paradigms such as Web 3.0 (semantization and enrichment), open knowledge and open data, Augmented Reality, Ambient Intelligence, and Sustainability of Learning Content, we need to investigate under which conditions online data can become useful as information or content.

The notion of context, as described above, becomes particularly important in light of the current pedagogical evolutions. Constructivist and task-based approaches strongly focus on learner autonomy with some degree of teacher and system support. The Situated Learning approach (Lave & Wenger, 1991) states that learning should be situated in a specific context and embedded within a particular social and physical environment. Complexity Theory and the Dynamic Systems Approach (Larsen-Freeman, 2002) see learning as a non-linear process, where rich input is considered crucial for learning to occur. Based on these arguments, we need to investigate to what extent the context of the learner, as defined above, can be implemented in a meaningful, useful, and enjoyable way within a state-of-the-art pedagogical approach in a mobile language learning environment.

The main objective of CALL research in this respect should be to create new opportunities for language learners to practice their language skills, wherever and whenever they are, “temporarily halted… they can divert their attention to language learning” (Darren, Searle, Chiu, Zhao, & Landay, 2011). Literature seems to suggest (there is no substantiated evidence in terms of generalizable findings yet) that learners are tempted to learn and practice more frequently when the learning content is adapted to their immediate surroundings.

Contextual language learning via mobile devices is a relatively new research topic. The didactic origins for contextual learning can be retraced to Situated Learning and Communities of Practice (Lave & Wenger, 1991). Contextual learning is also based on findings from cognitive psychology: Tulving and Thompson refer to contextualization as “encoding specificity” (Tulving & Thompson, 1973), which means that we link a word to the location where we learned it, thus making it easier to recall when being asked for it at the same or in a similar location. Godwin-Jones (2010) also advocates for learning language in real-world locations, as the learner will naturally space repetitions over both time and the places where we need it most (Darren et al., 2011). Godwin-Jones refers to the algorithms for review as specified in the Leitner system, developed in the 1970s.

Contextualization is also found in the concept of situated cognition (Brown, Collins, & Duguid,1989), which argues that knowledge is situated in activities, contexts, and culture in which it is developed and used. However, Larsen-Freeman (2012) warns that the dynamism as expressed in her Complexity Theory and the Dynamic Systems Approach is not a natural characteristic of many online language learning courses. In a similar vein, de Jong, Specht, and Koper (2008) point out the shortcomings of current learning content. They refer to Ogata and Yano, who identified the characteristics of contextualized learning that make it suitable for learning: permanency of the recorded learning processes, immediate access and accessible content, interactivity, and the situating of instructional activities (Ogata & Yano, 2004). Next, they translate these characteristics into a technical framework, which specifies context as information about objects in the real world (including the learner), time, location, activity (goals and tasks), and relations.

Much like exploratory projects, such as MicroMandarin (Darren et al., 2011) and LOCH (Ogata, Hui, & Yin, 2008), we should define context in terms of time and location. But we should also distinguish ourselves from previous attempts to contextualize content, as it also will take into account problems encountered when trying to use various data sources, such as authentic documents, enriched data, open data, Open Educational Resources, and existing interactive materials, as well as psychological issues, such as self-efficacy, locus of control, resistance, and motivation. Motivation will be a key topic, confronting Self-Determination Theory (Deci & Ryan, 2000; Deci & Vansteenkiste, 2004) with the L2 SELF model (Dörnyei & Ushioda, 2009) and our Personal Goal Theory (Colpaert, 2010).

Information-based tasks, such as “This museum is now open. Go inside and ask for a museum hand-out” or “The next bus will arrive in three minutes. Ask the driver when the next bus will arrive, ” are exciting to conceptualize, but they are difficult to evaluate and often entail the involvement of native speakers in a real-world setting. They cannot be defined in advance and should be generated "on-the-fly" based on an analysis of the context. Content-based tasks, on the contrary, involve some action to be undertaken by the learner based on the content itself. The advantage of content-based tasks is that they can be prepared in advance, and that any learner action can be monitored and logged in detail.

9. Consequences and Caveats

The open data phenomenon affords several unique opportunities for education worldwide. It will allow teachers and publishers to find, retrieve, edit, and share content worldwide without having to change the structure of the content.

However, based on our experience with previous technological "revolutions," we need to point out the following caveats. Teachers will still need to fundamentally rethink their attitude towards learning content and their innate notion of intellectual property. Researchers might need to focus more on designing interfaces for exchanging data in the most effective way. Policy makers may need to stop providing for repositories of learning content, and instead promote the emergence of start-ups around the development of learning apps. Publishers should finally understand that they have to fundamentally rethink their obsolete production chain, adopt a new business-to-business model, and radically select for sustainable learning content. This choice will not only entail consequences for authoring content, but it will also open up a new product range based on the same data sets: traditional textbooks, tailor-made ("on-the-fly") textbooks, learning content on demand, learning content for specific purposes, Open Educational Resources, interactive app(lication)s, mobile intelligent apps (with personalization and contextualization of the learning process), learning support, business-to-business products, and research data.

In this respect, there is a need for more transdisciplinary research focusing on the ontological specification of common concepts for guiding the design of interfaces for accessing learning content.

Only a transdisciplinary approach can succeed. If we want to learn from the past, then we should know that interdisciplinarity is not the solution to solve the "pluriness" of CALL. Open data is a focal point where it all comes together: technology, pedagogy, policy, business, linguistics, and psychology. Transdisciplinarity defined by Colpaert (2018) implies the construction of mental and physical artefacts on a higher boundary-transcendent level of abstraction. We can start with conceptualizing the learner, the teacher, and the learning environment in a new way.

References

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Brusilovsky, P., & Millán, E. (2007). User Models for Adaptive Hypermedia and Adaptive Educational Systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The adaptive web* (pp. 3-53). Berlin: Springer-Verlag.

Colpaert, J. (2004). *Design of online interactive language courseware: Conceptualization, specification and prototyping. Research into the impact of linguistic-didactic functionality on software architecture* (Unpublished doctoral dissertation). University of Antwerp, Antwerpen, Belgium.

Colpaert, J. (2010). Elicitation of language learners' personal goals as design concepts. *Innovation in Language Learning and Teaching*, *4*(3), 259-274.

Colpaert, J. (2012, March 29-30). Open Educational Resources for language teachers: A goal-oriented approach. Paper presented at the Joint CMC and Teacher Education EuroCall SIGS Workshop Bologna. Retrieved from http://eurocallsigsbologna.weebly.com/papers.html

Colpaert, J. (2013). Ontological specification of an authoring interface for creating sustainable language learning content. In L. H. Wong, C. C. Liu, T. Hirashima, P. Sumedi & M. Lukman (Eds.), *Proceedings of the 21st International Conference on Computers in Education* (pp. 644-653). Indonesia: Asia-Pacific Society for Computers in Education.

Colpaert, J. (2014). Reflections on present and future: towards an ontological approach to LMOOCs. In E. Martín-Monje & E. Bárcena (Eds.), *Language MOOCs: Providing learning, transcending boundaries* (pp. 162-173). Berlin: Mouton de Gruyter.

Colpaert, J. (2016). Big content in an educational engineering approach. *Journal of Technology and Chinese Language Teaching, 7*(1*)*, 1-14.

Colpaert, P. (2017). *Publishing transport data for maximum reuse* (Unpublished doctoral dissertation). Ghent University, Flanders, Belgium.

Colpaert, J. (2018). Transdisciplinarity revisited. *Computer Assisted Language Learning 31*(1-2). doi:10.1080/09588221.2018.1437111

Colpaert, J., & Decoo, W. (1999). The role of didactic functions in CALL design. In K. Cameron (Ed.), *CALL & the learning community* (pp. 64-74). Exeter: Elm Bank Publications.

De Jong, T., Specht, M., & Koper, R. (2008). Contextualized media for learning. *Educational Technology & Society*, *11*(2), 41-53.

De Meester, B., Lieber, S., Dimou, A., & Verborgh R. (2018). Interoperable user tracking logs using linked data for improved learning analytics. In *CALL your Data, XIXth International CALL Research Conference-Proceedings*. Antwerp: Universiteit Antwerpen.

Deci, E. L., & Vansteenkiste, M. (2004). Self-determination theory and basic need satisfaction: Understanding human development in positive psychology. *Ricerche di Psichologia*, *27*, 17-34.

Deci, E. L., & Ryan, R.M. (2000). The "what" and "why" of goal pursuits: Human needs and self-determination of behavior. *Psychological Inquiry*, *11*, 227–268.

Decoo, W. (2010). *Systemization in foreign language teaching: Monitoring content progression*. London: Routledge Research.

Dörnyei, Z., & Ushioda, E. (2009). *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters.

Edge, D., Searle, E., Chiu, K., Zhao, J., & Landay, J. A. (2011). MicroMandarin: Mobile language learning in context. *Conference on Human Factors in Computing Systems – Proceedings* (pp. 3169-3178).

Ellis, N.C., & Larsen-Freeman, D. (2009). *Language as a complex adaptive system*. Chichester; West Sussex, U.K: Wiley-Blackwell.

Ferguson, R., & Buckingham Shum, S. (2012). Social learning analytics: Five approaches. In *2nd International Conference on Learning Analytics & Knowledge-Proceedings*, (29 Apr-2 May, Vancouver, Canada). ACM Press: New York.

Godwin-Jones, R. (2010). Emerging technologies from memory palaces to spacing algorithms: approaches to second-language vocabulary learning. *Language Learning & Technology*, *14*(2), 4-11.

Koedinger, K., Cunningham, K., Skogsholm, A., & Leber, B. (2008). An open repository and analysis tools for fine-grained, longitudinal learner data. In *1st International Conference on Educational Data Mining — Proceedings* (pp. 157-166).

Larsen-Freeman, D. (2002). Language acquisition and language use from a chaos/complexity theory perspective. In C. Kramsch (Ed.), *Language acquisition and language socialization: Ecological perspectives* (pp. 33-46). London: Continuum.

Larsen-Freeman, D. (2012). Complex, dynamic systems: A new transdisciplinary theme for applied linguistics? *Language Teaching*, *45*(2), 202-214.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: University of Cambridge Press.

Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford: Oxford University Press.

Little, D. (2002). The European language portfolio: structure, origins, implementation and challenges. *Language Teaching*, *35*(3), 182-189.

Liu, X., Song, Y., Wu, K., Wang, J., Li, D., & Long, Y. (2015). Understanding urban China with open data. *Cities*, *47*, 53–61.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. London: Pearson.

Ma, Y. J. (2017). Why Open Data is good for China. *Global Investigative Journalism Network*. Retrieved from https://gijn.org/2017/03/28/why-open-data-is-good-for-china/.

Mouromtsev, D., & d'Aquin, M. (Eds.) (2016). *Open Data for education. Linked, shared, and reusable data for teaching and learning*. Springer, Switzerland. Retrieved from https://link.springer.com/book/10.1007%2F978-3-319-30493-9.

Niu, X., Sun, X., Wang, H., Rong, S., Qi, G., & Yu, Y. (2011) Zhishi.me–Weaving Chinese linking Open Data. In L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal,...E. Blomqvist (Eds.), *The Semantic Web–ISWC 2011. Lecture Notes in Computer Science*, *7032*. Berlin; Heidelberg: Springer.

Ogata, H., & Yano, Y. (2004). Context-aware support for computer-supported ubiquitous learning. In *Proceedings of the 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education* (pp. 27–34). IEEE Computer Society, Washington, DC.

Ogata, H., Hui, G. L., & Yin, C. (2008). LOCH: supporting mobile language learning outside classrooms. *International Journal of Mobile Learning and Organisation, 2*(3), 271-282.

Open Definition (2018). The Open Definition. Retrieved from https://opendefinition.org/.

Ross, E. (2015, December 2). Why open data doesn't mean open government. *The Guardian*. Retrieved from https://www.theguardian.com/media-network/2015/dec/02/china-russia-open-data-open-government.

Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68-78

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: The University of Illinois Press.

Tulving, E., & Thompson, D. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352-373.

UNESCO. (2018). What are Open Educational Resources (OERs)? Retrieved from http://www.unesco.org/new/en/communication-and-information/access-to-knowledge/open-educational-resources/what-are-open-educational-resources-oers/.

Van Elsen, E., Van Deun, K., & Decoo, W. (1991). Wordchip: The application of external versatility to an English lexical CALL program. *System, 19*(4), 401-417.

Van Merriënboer, J. G., & Kirschner, P. (2007). *Ten steps to complex learning*. New York: Taylor & Francis.

Walkington, C. A. (2013). Using adaptive learning technologies to personalize instruction to student interests: The impact of relevant contexts on performance and learning outcomes. *Journal of Educational Psychology*, *105*(4), 932-945.

Zhan, W., Ma, T., Tian, J., & Sunaoka, K. (2015). Development of an online database for Chinese resultative verb compounds and its application based on data visualization technology. *Journal of Technology and Chinese Language Teaching*, *6*(1), 1-15.

