e-Dubba: a Learning Platform for Sumerian Exposing Cuneiform Signs as Text

Louis LECAILLIEZ

University of Strasbourg, France
louis.lecailliez@outlook.fr

Abstract: *A wide variety of computer-assisted language learning systems have been developed for English and other living languages such as French, Chinese or Japanese. Comparatively, few systems, if any, have been developed for extinct languages, in particular for Sumerian. This language has the added difficulty for students to be written in cuneiform, a script in which a sign can have multiple readings and meanings and for which computer support is low. In this paper, we present e-Dubba, the prototype of a web-based learning platform that offers Sumerian courses in French. It is the first system of its kind to offer copy-paste-able cuneiform signs, which allows the reuse of them conveniently in other software applications or for communication. In addition, it integrates a dictionary that is capable of giving meaning and translation of words in the context in which they appear. Finally, the platform features learning analytics by logging user's page consultations and dictionary lookups. Teacher is provided with a page of its own to monitor progress.*

Keywords: Sumerian, CALL, smart dictionary, learning analytics.

1. Introduction

Numerous computer-assisted language learning (CALL) systems have been built for living languages (Gamper & Knapp, 2002). This paper lists software systems for English, Japanese, French, German, Spanish, Italian, Russian, Greek, Chinese, Arabic, Hebrew, Thai and Malay. There is however an imbalance in the representation of the languages, as most systems were developed to support the learning of English (14 platforms), Japanese (5) or French (4) or a combination (5) including one of those languages. That lack of language support diversity is hardly surprising given that most CALL publications emanate from the USA (30%), Taiwan (11%), UK (8%), and in a smaller percentage Canada, Australia and France (Gillespie, 2020).

Most world's languages, including the extinct ones, are thus lacking representation in the CALL landscape. One extinct language is the exception, classical Chinese, thanks to being studied as part of the school curriculum in China, Taiwan and Japan (Iino, 2002) and having a large population of speakers of a

language (Standard Chinese) with which it shares its writing system. Consequently, there exist some digital learning systems for classical Chinese like the one presented by Wu & Chen (2018). On the other hand, languages written using the cuneiform script, that is Sumerian, Akkadian, Elamite, Hurrian, Urartian and Hittite (Daniels & Bright, 1996) has no educational electronic system developed for that we are aware of.

Sumerian, which is offered in dedicated classes or as part of a history curriculum in higher institutions around the world, sometimes remotely, have the added difficulty of being written with a logographic script called cuneiform (Cooper, 1996). As the offering is still quite limited, online software solutions have the potential to reach a larger public that wouldn't be able to learn the language presently. The cuneiform script has been integrated into the Unicode standard since version 5.0 (Andries, 2008) but the general electronic support (input methods, fonts, software support) is lacking. Moreover, learning materials even in digital form relies almost exclusively on transliterations or images of the characters instead of using text. This poses an obvious issue for learners, as no text-based educational application like a flashcard software can be used for reviewing the course content. Communication with teachers or other students is also impaired by the need to use an alphabetic transcription instead of the script itself.

In this paper we present e-Dubba (*Edubba* means «house from which the tablets are coming out» in Sumerian), a web-based platform for learning Sumerian containing learning material in French and embedding a dictionary. In addition to hosting course content and tracking learner progress as any other learning analytics (Clow, 2013) solution would do, it includes the following innovations: Sumerian characters as text and a dictionary that gives the exact translation of a word in the context in which it appears.

2. Pedagogical Problematic

Cuneiform, the script created and used by Sumerian, is a big hurdle for learners because of its logographic nature. A cuneiform sign can be used as a logogram (it then designates a *signified* and *signifier* pair as a whole), a phonogram that is used only for its phonetic value or as a determinative (Jagersma, 2010), that is a character that helps determining the semantic field of a word by prefixing or suffixing it. Like Chinese, words are not separated by visible boundaries. Those two properties of the writing system pose an evident pedagogical issue where the burden of guessing the value of a sign, as well as text segmentation, is left to the student.

The burdensome impact of the script on one's studies is multiple, for instance: more learning items to remember (both sign value and their meaning in addition to words of the language), difficulty to read a text without a dictionary, trouble or impossibility to input text on a computer and more a convoluted dictionary lookup process. Even the conventional transcriptions, which use letters

with diacritics like ⟨ĝ⟩ and ⟨š⟩ are uneasy to input on a computer. Some of those aspects are shared with learners of Chinese and Japanese.

While there are digital dictionaries for both of these East-Asian languages which can be used using the native scripts and often with a romanization scheme (like hanyu pinyin for Chinese, or Hepburn romanization for Japanese), in Sumerian the reference dictionary is a paper dictionary, the *Manuel d'épigraphie Akkadienne* (Labat & Malbran-Labat, 1976). It contains cuneiform signs and words, their readings in Akkadian and Sumerian as well as their translations in French.

On the digital side, the ePSD (ePSD, 2022) dictionary in which an entry is presented in Figure 1 provides English translations but displays cuneiform signs as images instead of text. While this choice is understandable from an accessibility point of view, an option to display them as text is not present, despite the fact that the images of signs were clearly generated with the font *CuneiformComposite* used in the Unicode standard and by e-Dubba. In addition, it doesn't support lookup from Unicode characters.
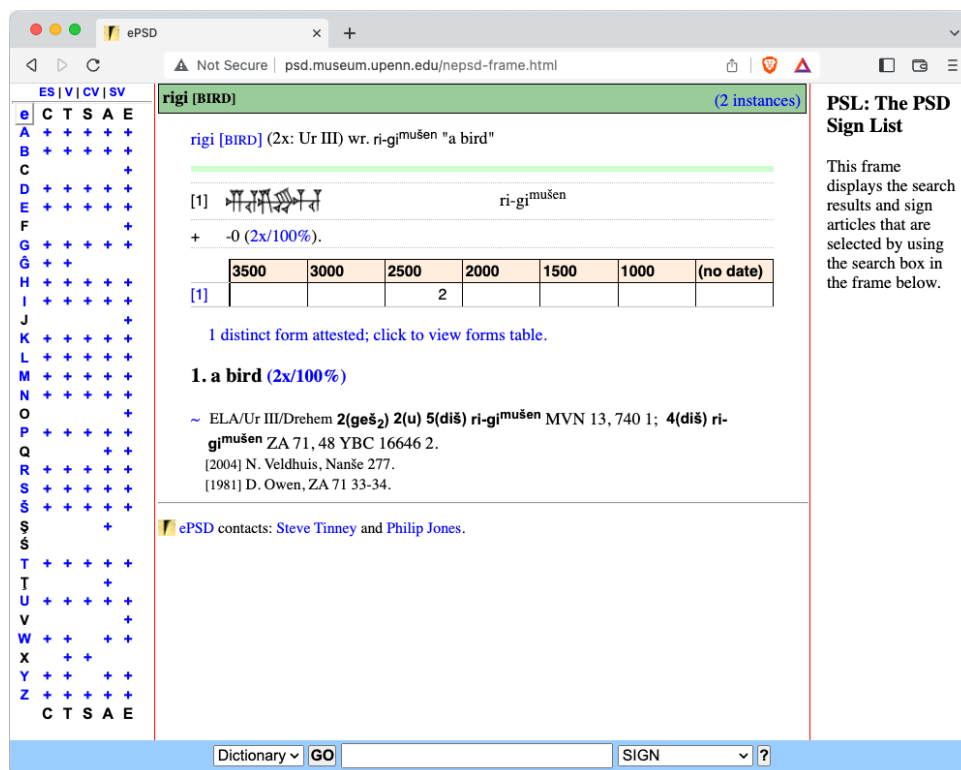

Figure 1. An entry in the ePSD dictionary

This situation motivated us to create a platform where: (1) cuneiform text is selectable so students can handle them as any other text on a computer and (2) create a Sumerian-French lexicon which is accessible directly within the pages of

the course. Finally, logging capabilities (3) allow us to perform learning analytics. This benefits the teacher responsible for the class by monitoring the progress of the cohort of students. Dictionaries lookups are logged as well and students can track their individual progress.

3. E-Dubba Platform

3.1. Course Content

The content hosted by e-Dubba is based on the learning material of the first year of a university diploma offered by University of Strasbourg (France). This two-year diploma can be either taken on-site or remotely. Each semester is divided into 12 weeks, for which material is handed to students as PDF documents, in the case of remote students. The platform was developed with the remote students as primary users, and on-site students as secondary users. The learning material includes for each lesson a main lesson file, and depending on the week various exercises, an epigraphic document to translate and answer files to exercises of the previous week.
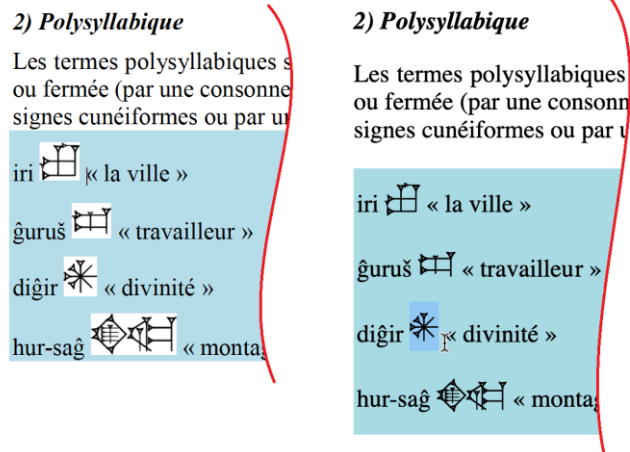

Figure 2. Original Learning Material (left), HTML Version (right)

The prototype of e-Dubba is built to host content of the main lessons. Five lessons have already been converted to HTML. A big difficulty of the task lies in the fact that in the original material cuneiform characters are images acquired from ePSD. Given the scarcity of resources containing cuneiform signs encoded with Unicode, some of those characters had to be manually searched in the Unicode standard table of characters (block U+12000). Figure 2 shows an extract of the original material (PDF) on the left with its HTML e-Dubba counterpart on the right.

It is visible on Figure 2 that the third Sumerian character (*diĝir*) is selected in the HTML version. This capability is an innovation in the landscape of Sumerian where even recent research works (Chiarcos et al., 2018; Punia & Schenk, 2020)

seem to ignore the cuneiform characters themselves and relies on transcriptions exclusively. For pedagogical purpose, the issue caused by the lack of signs' glyph variation support in Unicode which is problematic for researchers is greatly counterbalanced by the educational affordances offered by those characters being accessible in text form. That way, they can inputted easily into flashcard software and can be used into written communication, for example when corresponding with the teacher responsible of the course.

3.2. Integrated Dictionary

All lessons on e-Dubba feature a dictionary panel on the right of the course content, formatted in HTML as described in Section 3.1. When a Sumerian word from the lesson is clicked, the data displayed in the dictionary panel is updated. Figure 3 is a screen capture of a lesson page. The yellow part of the dictionary contains information of the word or sign in the context in which it appears. Meaning that can appears in other contexts are listed in the red part.

Figure 3. Lesson and Dictionary Panel

To provide the dictionary feature, we created a small Sumerian-French lexicon containing the signs, words and their values as used through the course formatted as a lexical network, using the methodology presented by (Lecailliez & Mangeot, 2018). As the script is logographic, we used special hyper-edges (see Figure 4, top) to encode information in a way that distinct meanings can be addressed by a unique id. Usage of hyper-edges allows to avoid the pitfalls of

automated logographic character processing in linked data as explained in (Lecailliez, 2017). Figure 4 (top) shows a cuneiform sign (in black) with two meanings (shown in blue) with their respective transcriptions (red). Each of the hyper-edge links a character, its meaning and its transcription is associated with a unique id, shown in purple as a digit in the illustration. In the real dictionary GUID are used for the identity of edges and nodes of the lexical network.
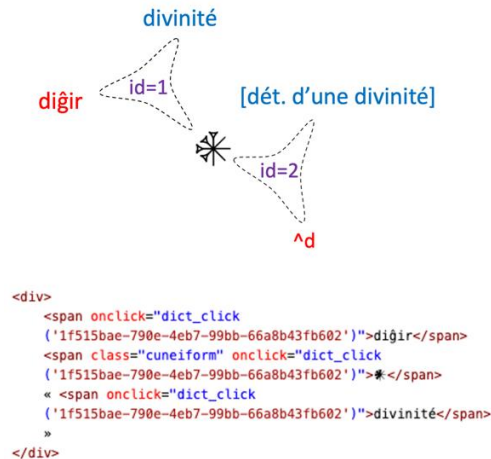


```
<div>
    <span onclick="dict_click
    ('1f515bae-790e-4eb7-99bb-66a8b43fb602')">diĝir</span>
    <span class="cuneiform" onclick="dict_click
    ('1f515bae-790e-4eb7-99bb-66a8b43fb602')">𒀭</span>
    « <span onclick="dict_click
    ('1f515bae-790e-4eb7-99bb-66a8b43fb602')">divinité</span>
    »
</div>
```

Figure 4. Hypergraph Modeling (top), HTML Encoding (bottom)

The bottom of the Figure 4 shows how the three parts (sign, meaning, transcription) are associated to each other in the HTML file of the lesson by linking the unique id (here *1f515bae-790e-4eb7-99bb-66a8b43fb602*) instead of the raw data itself. This method guarantees that when a click is performed on a word to query the dictionary, the data returned (here those associated with the meaning divinity) are the one expected in that context. Other meanings are displayed separately in a dedicated box of the dictionary.

3.3. Learning Analytics

A pedagogical problem common to courses taught online is that without analytics, the teaching staff doesn't know about the progression of students. The student itself might also be confused in where he stands in its learning.

For this reason, we integrated learning analytics into e-Dubba. After the user creates an account, he can log into the platform. Two user actions are automatically recorded by the system: pages that are browsed and dictionary queries performed by clicking on a Sumerian word in the text of a lesson. In comparison to the PDF version of the learning material, content is divided into more pages. This fine-grained division allows for more precise tracking of the student reading progression of the learning material. This could be reused for additional system features such as an integrated vocabulary review system, because repetition is one of the most important factors of vocabulary learning (Nation, 2017).

At this stage, the software components for dictionary lookup monitoring are implemented but visualization of progression as a student is not. So is the teacher page that should gather and present the reading progression of the student cohort, a simpler but similar in spirit of the dashboard presented in (Majumdar et al., 2019).

3.4. Possible Developments

The system currently focuses on the lesson material which is mostly in French with some Sumerian examples. Not included in the platform are exercises, most of which are translation homework's from a given Sumerian tablet. A possible and planned extension of the platform will be pages displaying the content of tablets in text form in addition to their photography. This would allow for additional features similar to those present in commercial online graded readers for Chinese that help readers to understand the text.

In particular, three types of information about a sign would be immediately relevant to the user: its pronunciation, translation and category (logogram, phonogram or determiner). The first two types of information are already present in the dictionary and the technique presented in Section 3.2 can be reused to provide contextual information of the words. The third information could be provided using a coloring scheme. The main difficulty lies in creating the text, which can be addressed using a web IME (input method). In addition, the integrated dictionary can be used to provide a provisional segmentation of a text thus removing the burden from the editor to manually segment the text and add most of the information pertaining to signs.

Finally, the platform can be adapted for other languages written using cuneiform script, such as Akkadian which is also offered remotely at the University of Strasbourg.

4. Conclusion

In this paper we presented e-Dubba, a web-based learning platform for French-speaking learners of Sumerian. The platform provides three main features which are innovative in the context of learning a language written using the cuneiform script. First, cuneiform signs are encoded using the Unicode standard making them easily paste-able into other applications. Secondly, we added a Sumerian-French dictionary to the system. By using it, learners don't have to rely on external systems for most of their vocabulary needs. In addition, it alleviates the complexities of the cuneiform script as used by Sumerian by providing the exact definitions of words and signs in the context in which they appear. Finally, a student can track its progression within the course thanks to the logging capabilities of the system. The teacher responsible for the course can monitor progression of the whole cohort of students.

The system is currently in development, with some of its features missing. When completed, it will be proposed to the students taking an online degree of Sumerian at University of Strasbourg for evaluation.

References

Andries, P. (2008). *Unicode 5.0 en pratique: codage des caractères et internationalisation des logiciels et des documents*. Dunod.

Chiarcos, C., Pagé-Perron, É., Khait, I., Schenk, N. & Reck- ling, L. (2018). Towards a linked open data edition of Sumerian corpora. In *Proceedings of the 11th International Conference on Language Resources and Evaluation*.

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, 18(6), 683-695.

Cooper, J. S. (1996). *Sumerian and Akkadian*. In P. T. Daniels & W. Bright (Eds.), The World's Writing Systems (pp. 37-57). Oxford University Press.

Daniels, P. T. & Bright, W. (1996). *The World's Writing Systems*. Oxford University Press.

ePSD (2022). http://psd.museum.upenn.edu/nepsd-frame.html.

Gamper, J. & Knapp, J. (2002). A review of intelligent call systems. *Computer Assisted Language Learning*, 15(4), 329-342.

Gillespie, J. (2020). Call research: Where are we now? *ReCALL,* 32(2), 127-144.

Iino, M. (2002). Language and English education in Japan. *English language education in China, Japan, and Singapore*, 91-93.

Jagersma, B. (2010). *A descriptive grammar of Sumerian*. Doctoral thesis: Leiden University.

Labat, R. & Malbran-Labat, F. (1976). Manuel d'épigraphie akkadienne: (signes, syllabaire, idéogrammes). *Librairie Orientaliste Paul Geuthner*.

Lecailliez, L. (2017). Preliminary thoughts on issues of modeling Japanese dictionaries using the ontolex model. In A. Horák, P. Rychlý, & A. Rambousek (Eds.). *Proceedings of the Eleventh Workshop on Recent Advances in Slavonic Natural Languages Processing (RASLAN)* (pp. 11- 19). Tribun EU.

Lecailliez, L. & Mangeot, M. (2018). A hypergraph data model for building multilingual dictionary applications. In *Proceedings of ASIALEX* (p. 138-146).

Majumdar, R., Akçapınar, A., Akçapınar, G., Ogata, H. & Flanagan, B. (2019). Laview: Learning analytics dash- board towards evidence-based education. In *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge*.

Nation, P. (2017). How vocabulary is learned. *Indonesian Journal of English Language Teaching*, 12(1), 1-14.

Punia, R., Schenk, N., Chiarcos, C. & Pagé-Perron, É. (2020). Towards the first machine translation system for Sumerian transliterations. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 3454-3460).

Wu, T.-T. & Chen, A.-C. (2018). Combining e-books with mind mapping in a reciprocal teaching strategy for a classical Chinese course. *Computers & Education*, 116, 64-80.