INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Using Alexa for Flashcard-based Learning

Lucy Skidmore, Roger K. Moore

Speech and Hearing Research Group, Dept. Computer Science, University of Sheffield, UK
{lskidmore1,r.k.moore}@sheffield.ac.uk

Abstract

Despite increasing awareness of Alexa's potential as an educational tool, there remains a limited scope for Alexa skills to accommodate the features required for effective language learning. This paper describes an investigation into implementing 'spaced-repetition', a non-trivial feature of flashcard-based learning, through the development of an Alexa skill called 'Japanese Flashcards'. Here we show that existing Alexa development features such as skill persistence allow for the effective implementation of spaced-repetition and suggest a heuristic adaptation of the spaced-repetition model that is appropriate for use with voice assistants (VAs). We also highlight areas of the Alexa development process that limit the facilitation of language learning, namely the lack of multilingual speech recognition, and offer solutions to these current limitations. Overall, the investigation shows that Alexa can successfully facilitate simple L2-L1 flashcard-based language learning and highlights the potential for Alexa to be used as a sophisticated and effective language learning tool.

Index Terms: Alexa, voice assistants, computer-assisted language learning, spaced-repetition, flashcards

1. Introduction

The variety of uses for VAs such as Amazon Alexa continue to grow rapidly along with the number of people adopting this technology [1]. As the field continues to expand, one area of interest that has recently emerged is the use of VAs for educational purposes. [2, 3].

One educational application of Alexa that is proving to be popular is flashcard-based learning: the conversational framework of Alexa is well-suited to the short exchanges required for flashcard-based learning, and it is already possible for developers to create flashcard skills using Alexa Skill Blueprints.[1]

Flashcard-based learning is a well-studied and popular method for L2 vocabulary memorisation [4]. Using Alexa for language learning in this way could offer learners endless opportunities for practice, enabling them to develop listening and speaking skills that are otherwise supplementary in other modes of flashcard-based learning but nonetheless deemed important by learners [5]. However, as it stands, it is currently not possible to integrate multilingual functionality into Alexa's Skill Blueprint for flashcards and so the problem of creating a flashcard skill for language learning remains non-trivial.

With this in mind, this paper describes the development process for creating a flashcard-based language learning skill. It answers the three following research questions:

1. *What features of Alexa can be used to facilitate a flashcard-based language learning skill?*
2. *What adaptations to traditional flashcard-based models are required for VAs?*
3. *What additional features could be introduced to Alexa to better facilitate flashcard-based language learning?*

This paper first introduces flashcard-based learning and how it is used in language learning applications. This is followed by an overview of the design and implementation processes for creating a language learning flashcard skill for Alexa. Next, a discussion of the developmental advantages and disadvantages of Alexa provides answers to the research questions outlined above. Based on the answers to these questions, limitations of the study are noted alongside suggestions for future research. To conclude, this investigation is considered in the wider context of computer-assisted language learning (CALL).

2. Flashcard-based Learning

2.1. What is flashcard-based learning?

In a language-learning context, flashcards are commonly used to study L2 vocabulary in what is known as a 'paired-associate' format [6]. One common example of a learning pair would be an L2 word and its associated meaning in the learner's native language (L1), e.g. the Japanese word *ichi* and its English meaning *one*. Traditionally, these pairs are written on either side of a paper flashcard. A learner is presented with one side of the flashcard and tries to recall what is on the other side – this is known as 'cued-recall learning' [7].

More recently, digitised flashcards have become a popular mode of learning and can be found in well-known language learning applications such as Duolingo [8]. Studies have shown that digitised flashcards outperform traditional paper-based flashcards in some cases thanks to computer-assisted enhancements such as immediate feedback and personalisation [9, 10]. Other research has shown that learners often prefer digitised flashcards due to their portability and accessibility [11, 12]. Thanks to the integration of Alexa into devices such as mobile phones and car dashboards [1], it is logical to assume that using flashcard-based Alexa skills would have similar positive effects on learners.

2.2. Leitner System for flashcard-based learning

The Leitner System [13] is a common approach to implementing flashcard-based learning in applications [7, 8]. Flashcards are introduced using 'spaced repetition', a learning method that came from Ebbinghaus' research findings that reviewing material at exponentially spaced intervals aids memorisation [14].

The Leitner System is represented as N piles of flashcards. Words enter into the system at deck 1. During a flashcard session, learners start with the flashcard at the top of the current deck i that they are studying. As Figure 1 depicts, if the learner gets the answer on the back of the flashcard correct, that card is promoted to the bottom of the subsequent deck (deck $i+1$), or removed from the system altogether if deck i = deck N. If the learner is incorrect, the flashcard is demoted to the bottom

[1] https://blueprints.amazon.co.uk

http://dx.doi.org/10.21437/Interspeech.2019-2893

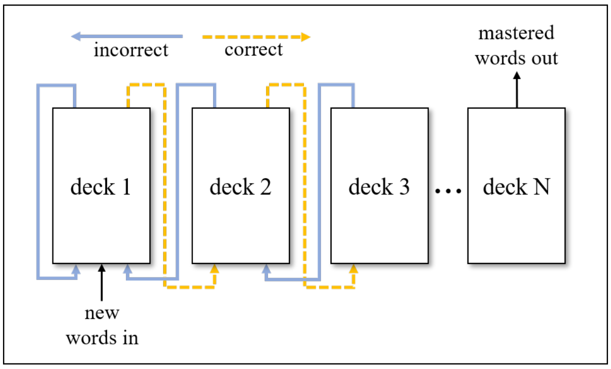

Figure 1: *Diagram of the Leitner System*

of the previous deck (deck $i-1$), or to the bottom of the same deck if deck i = deck 1. Words that have been removed from the system are deemed to be 'mastered'.

The frequency of practice for each deck is set at pre-determined, exponentially spaced intervals. For example, the learner may look at flashcards in deck one every day, flashcards in deck two every other day, flashcards in deck three every four days, and so on. This allows words that the learner finds easy to be seen less often as they progress further up the decks and difficult words to be repeated more frequently as they remain in the lower decks. This dynamic adaption of flashcard practice order depends on learner performance, which means no two learners' Letiner decks will be the same. This creates a personalised learning experience.

3. Design

The skill designed for this investigation is called 'Japanese Flashcards' and is available to use on all Alexa-enabled devices.

3.1. Japanese Vocabulary

Japanese Flashcards has 100 available Japanese words for the learner. This vocabulary was taken from a freely available resource provided by Tofugu[2], a Japanese educational resource platform, which outlines 100 of the most useful phrases for a beginner learner of Japanese. There are ten topics covered including greetings, basic verbs and numbers.

3.2. Adapting the Leitner System for Alexa

As outlined in Section 2.2, the Leitner System facilitates long-term retention of vocabulary through repeated testing of words over time. This mode of learning is essential for memorisation, but it is also in contradiction to the way that consumers tend to use VAs: short actions completed within seconds such as asking a question or checking the weather. Flashcard-based learning has also been shown to be an effective but 'boring' task for students [15]. One way to mitigate these problems is to design 'microlearning' activities which reflect the way that users adopt these kinds of technologies [7]. With this in mind, it is important to ensure that voice applications designed for language learning can facilitate both long-term retention of vocabulary as well as a satisfying short-term experience for the learner. Therefore, based on the presumption that users will likely interact with Japanese Flashcards sporadically for short periods of time, a heuristic approach to the traditional Leitner System model was applied in order to facilitate both short-term 'microlearning' and long-term memorisation. This approach provides a baseline for implementing spaced repetition in an Alexa skill and further testing on how learners use the skill over time will be necessary for optimisation of the system.

In order to satisfy the requirement for short-term experience, the Leitner System model was adapted to be used in one microlearning session rather than over a number of days. In order to achieve this, the Leitner System model for Japanese Flashcards contains four decks and each deck is seen at exponentially decreasing intervals per session. In one session, flashcards in deck 1 are tested eight times, flashcards in deck 2 are tested four times, flashcards in deck 3, twice, and flashcards in deck 4, once. Each session introduces five new Japanese words to the learner. This allows for the users to be introduced to the vocabulary at a steady but satisfying pace.

In order to facilitate long-term retention, the Leitner System model is also applied *across* sessions as well as *within* sessions. At the end of each session, any flashcards that have completed the Leitner decks (i.e. deemed as mastered) are recorded in the user's database and re-introduced in subsequent sessions at exponential intervals. For example, words that were deemed as mastered in a user's session one are reintroduced in session two, session four, session eight and so on.

3.3. Skill architecture

A skill's functionality is separated into two parts: 'intents', which represent user utterances and 'intent handlers', which contain the rules for how the skill will respond to these utterances. Intents are defined in what is known as an interaction model: "*the voice interface through which users interact with the skill*."[3] Intent handlers are written by the developer as functions in code.

3.3.1. Interaction model

Table 1 summarises the interaction model design for Japanese Flashcards. The 'Intent' column represents the name of the intent that the user's utterance is mapped to, and 'sample utterances' are a set of phrases that the user might use to invoke the intent. The variety of sample utterances coded as **answerIntents** is a good example of the flexibility of production allowed in Alexa skills; developers are able to account for 'ums and ahs' that are likely to occur as a learner tries to recall the meaning of a word, avoiding the potential frustration of a user giving a correct answer but having the VA interpret it as incorrect. The term **[answerSlot]** acts as a placeholder for all possible answers in the vocabulary list.

Table 1: *Interaction model for Japanese Flashcards*

Intent	Sample Utterances
launchIntent	'Alexa, open Japanese Flashcards'
	'Alexa, launch Japanese Flashcards'
practiceIntent	'Let's go'
answerIntent	[answerSlot]
	erm [answerSlot]
	is it [answerSlot]

[2]https://www.tofugu.com

[3]https://developer.amazon.com/docs/custom-skills/create-the-interaction-model-for-your-skill.html

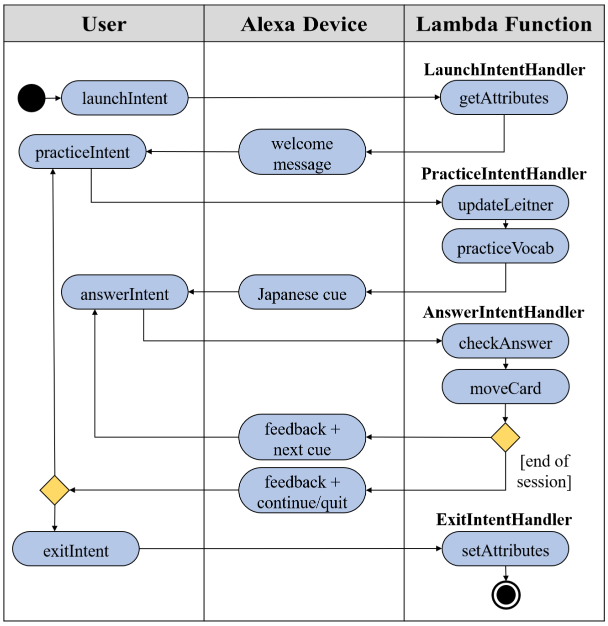

Figure 2: *UML activity diagram of Japanese Flashcards*

In addition to the skill specific intents and utterances, the interaction model also includes pre-built intents: AMAZON.CancelIntent, AMAZON.StopIntent and AMAZON.HelpIntent. These account for general functionality required in all Alexa skills.

3.3.2. UML activity diagram for Japanese Flashcards

Figure 2 outlines the functional interaction between the user, the Alexa-enabled device and the skill's code when engaging in a Japanese Flashcards session. When the user says "*Alexa, open Japanese Vocabulary*", a **launchIntent** is sent to the Lambda function and Alexa responds with a welcome message, dependent on the user's saved attributes from previous sessions. Alexa then invites the user to start their flashcard practice by prompting the utterance "*Let's go*". This **practiceIntent** invokes the **PracticeIntentHandler**, which calls the **createLeitner** function, generating the Leitner decks for the session based on the user's session history. Next, the **practiceVocab** function is called. The function iterates through the flashcards in the user's Leitner decks. Alexa provides the appropriate word in Japanese, to which the user replies in English. Upon hearing the user's response, received as an **answerIntent**, the user's utterance is compared to the correct answer, stored in a dictionary. If the user's answer matches the correct answer, a correct chime is played and the flashcard is moved up the user's Leitner deck. If the user's answer is incorrect, an incorrect chime is played and the card is moved down the Leitner decks, as described in Section 3.2. This process continues until the user reaches the end of the allotted session practice, or says "*Exit*" - whichever comes first. When the **exitIntent** is called, the user's data is saved as persistent attributes for subsequent sessions.

4. Implementation

4.1. Development environment

Japanese Flashcards was developed as an 'Alexa-hosted skill', using the Alexa Developer Console (ADC). Launched in January 2019, Alexa-hosted skills (Beta) have automatic access to an AWS Lambda endpoint, Amazon S3 bucket and an Amazon S3-backed key-value table. The automatic management of these AWS cloud services allows developers to publish skills efficiently without leaving the ADC. For Japanese Flashcards, the Amazon S3-backed key-value table was used to implement persistence, allowing data from users' sessions to be stored and accessed in subsequent sessions. This is of particular interest for L2 learning applications as it creates the opportunity for personalisation, a feature which has been shown to be desired by language learners [16]. At the end of each session, the user's personal Leitner deck, along with the words deemed as 'mastered' during the session are saved as 'attributes' in an Amazon S3-backed key-value table. The next time that the user launches Japanese Flashcards, these attributes are retrieved and used to create the Leitner deck for the session.

The ADC allows for the development of both the interaction model and intent handlers required for a skill. Firstly, the interaction model was generated using the ADC's 'Skill Builder', which enables developers to build an interaction model using an intuitive GUI, and subsequently generates a JSON representation to be deployed along with the skill's intent handlers. The code for the intent handlers was written using the Alexa Skills Kit (ASK) Software Development Kit (SDK) for Node.js[4] and can be found on GitHub.[5]

```
var switchLang = function(text) {
  if (text){
    return "<prosody volume ='x-loud'>" +
    "<voice name='Mizuki'><lang xml:lang='ja-JP'>" +
    text +
    "</lang></voice></prosody>"
  }
}
```

Figure 3: ***switchLang*** *function*

4.2. Alexa Skills Kit Sound Library

Sound effects were used from the ASK Sound Library[6], a set of freely available sounds created by Amazon for developers to use in skills. Correct and incorrect chimes were taken from the Game Show Sounds within the Sound Library to indicate whether or not the user responded correctly.

4.3. Japanese TTS with Amazon Polly

The Japanese vocabulary used in the skill was generated using AWS Amazon Polly, Amazon's text to speech (TTS) service. In order to switch from Alexa to Amazon Polly's Japanese voice 'Mizuki', a **switchLang** function was written. As shown in Figure 3, this function takes the speech text as an argument and returns the text wrapped in speech synthesis markup language (SSML) tags. The tags required to generate Japanese speech are 'voice name' (Mizuki) and 'language' (Japanese). In addition, as volumes across Amazon Polly languages are not normalised, the 'prosody' (x-loud) tag was used to boost the volume of the Japanese speech. Further information on SSML tags can be found in the Amazon Polly Developer Guide[7].

[4]https://github.com/alexa/alexa-skills-kit-sdk-for-nodejs
[5]https://github.com/lucyskidmore/JapaneseFlashcards
[6]https://developer.amazon.com/docs/custom-skills/ask-soundlibrary.html
[7]https://aws.amazon.com/polly

5. Discussion

5.1. Research question one

The areas where development of Japanese Flashcards was successful answer this investigation's first research question: *What features of Alexa can be used to facilitate a flashcard-based language learning skill?*

Namely, skill persistence proved to be an invaluable feature for implementing flashcard-based learning with Alexa, allowing for enhancements such as personalisation, progress tracking and user-adaption [17]. There are other ways not explored in this investigation for persistence to be effectively used in language learning skills. For example, users' persistence data could be analysed to highlight words that they are frequently getting wrong or could extend to more sophisticated analysis such as modelling their memory and subsequently predicting likely forgotten words [8], or optimising the order of word practice [18].

Integration of Amazon Polly voices with Alexa was essential for the successful development of Japanese Flashcards. Using TTS processed directly in the cloud in this way instead of using pre-recorded Japanese utterances saves both time and space, both of which are essential if this model were to be applied at scale. However, there is still some room for improvement. We suggest that built in functions for switching languages are created in the ASK SDK allowing for automatic normalisation of volume and ease of implementation. It is also worth noting that it is not clear whether using TTS technology in a language learning context is effective, and this warrants further investigation.

Finally, the flexibility of responses allowed from users thanks to input variation in the interaction model is a powerful feature of Alexa that could be used beyond flashcard-based learning. In the case of Japanese Flashcards, it allows users to respond to the skill in a natural way (e.g. including words like 'erm', and other filler words), both enhancing the user experience and mitigating user frustration. This could be extrapolated to flashcards that require longer answers, for example words and their definitions.

5.2. Research question two

Implementing the adapted Leitner System that facilitates short-term experience and long-term learning outlined in Section 3.2 answers this investigation's second research question: *What adaptations to traditional flashcard-based models are required for VAs?*

This heuristic approach to the Leitner System provides a starting point for research in the area of using flashcard-based, spaced repetition language learning with VAs. To further improve and validate this approach, it would be beneficial to carry out a longitudinal study following learners using Japanese Flashcards, including linguistic testing to quantitatively assess learners' language improvement after using the skill for a period of time. Various pre-set spaced-repetition configurations could be tested, as well as configurations explicitly set by learners.

5.3. Research question three

Findings from this investigation also illuminated areas of Alexa that either currently do not exist or could be improved for a better learning experience, offering an answer to research question three: *What additional features could be introduced to Alexa to better facilitate flashcard-based language learning?*

As it stands, there is currently no built-in functionality with Alexa for multilingual speech recognition. The implication for Japanese Flashcards is that learners are only able to listen to Japanese, with flashcards presented in L2–L1 format. This allows no opportunity for Japanese production, which is an essential part of language learning [19]. This issue could be solved in two ways: either making raw speech data available for developers to use external voice recognition tools such as Google's Cloud Speech-to-Text (SST) API[8] (which would also enable additional features such as using confidence scores for pronunciation training), or by integrating multilingual ASR within the Alexa development environment. Including multilingual functionality in this way would also allow users to create their own flashcards to learn and practice, which has been shown to be a positive feature of flashcard software that increases student performance [6, 20].

6. Future work

This investigation is limited to evaluating the development process of a flashcard-based language learning skills, and further empirical testing is necessary to establish its validity as a successful tool for language learners. One example of a validity test could be to compare the efficacy of voice-based flashcard learning with that of visual only or audio-visual flashcard learning. Another evaluation method could be to gather data on learners' attitudes towards voice-based systems. In addition, a comparative investigation of other voice assistants such as Google Assistant and Siri would be valuable. This study also gives way to broader interesting research questions, including investigating how learners interact with and adapt to systems that use ASR as well as testing the impact of using TTS voices in language learning applications.

7. Conclusion

This investigation sought to understand how to develop a non-trivial flashcard-based Alexa skill for language learning and answered three research questions. Firstly, existing Alexa development features that facilitate flashcard-based language learning were established: skill persistence for logging users' learning data, TTS with Amazon Polly for generating L2 words and a flexible interaction model to account for variation in users' answers. Secondly, a heuristic adaptation to the traditional Leitner System used in flashcard software was introduced in order to create opportunities for 'microlearning', accounting for the way that VAs are commonly used. Finally, limitations of Alexa were identified, namely the lack of multilingual speech recognition which remains a barrier for effective L2 speaking practice with Alexa.

With better integration of multilingual functionality as well as other features highlighted in this paper, Alexa could go far beyond facilitating simple exercises such as flashcard-based learning, and move towards other valuable learning paradigms such as conversational role play and pronunciation training, providing an invaluable resource to language learners in the future.

In summary, this paper acts as a starting point for investigations into developing Alexa skills for educational purposes and contributes to the emerging field of 'voice-assisted language learning' [21, 22].

[8] https://cloud.google.com/speech-to-text/docs/

8. References

[1] B. Kinsella and A. Mutchler, "Voicebot.ai Voice Assistant Consumer Adoption Report," Voicebot.ai, US, Sep. 2018. [Online]. URL: https://voicebot.ai/voice-assistant-consumer-adoption-report-2018

[2] C. Delgado-Kloos, C. Catalan, P. Merino and C. Alario-Hoyos, "Design of a Conversational Agent as an Educational Tool," in *Learning with MOOCS, September 26-28, Madrid, Spain, Proceedings*, 2018, pp.27–30.

[3] S. Babic, T. Orehovacki and D. Etinger, "Perceived user experience and performance of intelligent personal assistants employed in higher education settings," in *41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), May 21-25, Opatija, Croatia, Proceedings* 2018, pp. 0830–0834.

[4] R. Godwin-Jones, "Emerging Technologies: From Memory Palaces to Spacing Algorithms: Approaches to Second-Language Vocabulary Learning," *Language Learning & Technology*, vol. 14, no. 2, pp. 4-11, 2010.

[5] F. Rosell-Aguillar, "Autonomous language learning through a mobile application: a user evaluation of the busuu app," *Computer Assisted Language Learning*, vol. 31, no. 8, pp. 854-881, 2018.

[6] T. Nakata, "Computer-assisted second language vocabulary learning in a paired-associate paradigm: a critical investigation of flashcard software," *Computer Assisted Language Learning*, vol. 24, no. 1, pp. 17-38, 2011.

[7] D. Edge, S. Fitchett, M. Whitney and J. Landay, "MemReflex: Adaptive Flashcards for Mobile Microlearning," in *14th International conference on Human computer interaction with mobile devices and services (MobileHCI'12), September 21-24, San Francisco, CA, USA, Proceedings* 2012, pp. 431–440.

[8] B. Settles and B. Meeder, "A Trainable Spaced Repetition Model for Language Learning," in *54th Annual Meeting of the Association for Computational Linguistics, August 7-12, Berlin, Germany, Proceedings* 2016, pp. 1848–1858.

[9] R. Hirschel and E. Fritz, "Learning vocabulary: CALL program versus vocabulary notebook," *System*, vol. 41, pp. 639-653, 2013.

[10] M. Sccott and K. Chung-JA, "Benefits of Using CALL Vocabulary Programs to Provide Systematic Word Recycling," *English Teaching*, vol. 63, no. 1, pp. 1-13, 2008.

[11] J. Nikoopoura and A. Kazemib, "Vocabulary Learning through Digitized & Non-digitized Flashcards Delivery," *Procedia – Social and Behavioral Sciences*, vol. 98, pp. 1366-1373, 2014.

[12] G. Dizon and D. Tang, "Comparing the efficacy of digital flashcards versus paper flashcards to improve receptive and productive L2 vocabulary," *The EuroCALL Review*, vol. 25, no, 1, pp. 3-15, 2017.

[13] S. Leitner, "*So lernt man lernen. Angewandte Lernpsychologie ein Weg zum Erfolg*," Freiburg im Breisgau, Germany: Verlag Herder, 1972.

[14] H. Ebbinghaus, "*ber das Gedchtnis: Untersuchungen zur experimentellen Psychologie*," Darmstadt, Germany: Wissenschaftliche Buchgesellschaft, 1885.

[15] A. Sebert Hanson and C. Brown, "Enhancing L2 learning through a mobile assisted spaced-repetition tool: an effective but bitter pill?," *Computer Assisted Language Learning*, pp. 1-23, 2019. [Online]. doi: 10.1080/09588221.2018.1552975

[16] K. Sage, J. Rausch, A. Quirk and L. Halladay, "Pacing, Pixels, and Paper: Flexibility in Learning Words from Flashcards," *Journal of Information Technology Education: Research*, vol. 15, pp. 431-456, 2016.

[17] J. Markowitz, "*Robots that talk and listen - technology and social impact*," Chicago: De Grutyer, 2014.

[18] S. Reddy, I. Labutov and S. Banerjee, "A Queueing Network Model for Spaced Repetition," in *Third (2016) ACM Conference on Learning @ Scale, April 25-26, Edinburgh, Scotland, UK, Proceedings* 2016, pp. 289–292.

[19] M. Eskenazi, "Using Automatic Speech Processing for Foreign Language Pronunciation Tutoring: Some Issues and a Prototype," *Language Learning & Technology*, vol. 2, no, 2, pp. 62-76, 1999.

[20] A. Luczak, "Using Memrise in Legal English Teaching," *Studies in Logic, Grammar and Rhetoric*, vol. 49, no, 62, pp. 141-152, 2017.

[21] G. Dizon, "Using Intelligent Personal Assistants for Second Language Learning: A Case Study of Alexa," *TESOL Journal*, vol. 8, no, 4, pp. 811-830, 2017.

[22] S. Moussalli and W. Cardoso, "Are commercial "personal robots" ready for language learning? Focus on second language speech," in *CALL communities and cultureShort papers from EUROCALL 2016, April 24-27, Limassol, Cyprus, Proceedings* 2016, pp. 325–329.