MicroMandarin: Mobile Language Learning in Context

Darren Edge[1], Elly Searle[1,2], Kevin Chiu[1,3], Jing Zhao[1,4], James A. Landay[1,2]

[1]Microsoft Research, Beijing, People's Republic of China
[2]University of Washington, Seattle, WA, USA
[3]MIT Media Lab, Cambridge, MA, USA
[4]Peking University, Beijing, People's Republic of China

{darren.edge, elly.searle}@microsoft.com, kgc@media.mit.edu, jingzi@pku.edu.cn, landay@cs.washington.edu

ABSTRACT
Learning a new language is hard, but learning to use it confidently in conversations with native speakers is even harder. From our field research with language learners, with support from Cognitive Psychology and Second Language Acquisition, we argue for the value of *contextual microlearning* in the many breaks spread across different places and throughout the day. We present a mobile application that supports such microlearning by leveraging the location-based service Foursquare to automatically provide contextually relevant content in the world's major cities. In an evaluation of Mandarin Chinese learning, a four-week, 23-user study spanning Beijing and Shanghai compared this contextual system to a system based on word frequency. Study sessions with the contextual version lasted half as long but occurred in twice as many places as sessions with the frequency version, suggesting a complementary relationship between the two approaches.

Author Keywords
Context-Aware, Mobile, Microlearning, Language Learning

ACM Classification Keywords
H.5 Information Interfaces & Presentation: User Interfaces

General Terms
Design, Experimentation, Human Factors

INTRODUCTION
Learning a second language is a significant lifetime goal for many people all over the world, but it is rarely easy to achieve. Even for skilled linguists, it takes around 2200 hours (88 weeks) of dedicated classroom instruction to reach general proficiency in languages dissimilar to their native language (e.g., Chinese and Japanese for English native speakers) [21]. Living in countries where these languages are spoken is often necessary to reach fluency.

Aside from people who need to use a second language in their chosen professional careers, many language learners simply do not have the time or resources to dedicate two years of their lives to intensive or immersive instruction. This is true both for people seeking to cross the "language divide" by gaining a world language and improving their socioeconomic mobility [19], and for second language learners with non-instrumental sources of motivation (e.g., those driven by the intellectual-challenge or desire to integrate with members of the target culture [29]). For all such learners, the challenge is to maintain the *executive motivation* [7] to spend time learning in the face of competing demands from family, work, and social life.

One solution is to use the mobile phone as a platform for "microlearning" [11], helping language learners to identify and exploit the many moments during the day where other distractions are temporarily halted and attention can be diverted to language learning. In contrast with PCs, the ultra-portability of mobile devices, combined with the low costs of retrieval, use, and stowage, makes them ideally suited to exploit opportunities for microlearning. Moreover, since such opportunities naturally occur in different places, context-aware mobile systems that provide contextually-relevant content (e.g., [1, 2, 6, 15, 26, 27, 28]) could provide immediately valuable learning material (Figure 1).

Figure 1. Contextual Microlearning

In this paper, we present a mobile application for the microlearning of language in its context of use, in a way that encourages use of that language in meaningful interactions with native speakers. Our model of context is based on location, leveraging the location-based social networking service Foursquare to find the venue closest to the learner's mobile GPS coordinates. We use the type of this venue (e.g., "Café") to automatically suggest the study of language (e.g., "Cappuccino") tagged with that context.

Contextual microlearning is thus the frequent but brief study of material whose meaning is relevant in the current context. Our approach is based on flashcards, with the user prompted to recall the translation of each contextual card.

Our initial rationale for such contextual microlearning was based on three findings from Cognitive Psychology:

1. *Encoding specificity* [37] means that recall is best when the contexts of learning and retrieval share perceptual cues. If we learn language in its context of use (e.g., the word "menu" in a restaurant), it should help us to recall that language when and where we need to use it.
2. *Spaced repetition* [12] means that learning is best when content is presented over time rather than in quick succession. If we learn language in short bursts in real-world locations, we will naturally space repetitions over both time and the places where we need it most.
3. *Situated cognition* [5] supports the notion that knowledge cannot be fully abstracted away from the activities, contexts, and cultures in which it is developed. Language is therefore learned implicitly through use in context as well as through explicit study.

We expand on this initial rationale in the next section of the paper, through analysis of prior systems, our own user research, and theories from Second Language Acquisition (SLA) research. We then draw on this new understanding to articulate the key obstacle to language learners making progress: shortage of spoken interactions with native speakers in the pursuit of real-world goals. This was most evident in our user research of English-speaking learners of Mandarin Chinese in China, and prompted the development of a mobile application for the contextual microlearning of Chinese. Finally, we describe our implementation of this novel system for city-wide contextual learning using Foursquare, and its evaluation over a 4 week, 23-user study spanning the Chinese cities of Beijing and Shanghai.

UNDERSTANDING LANGUAGE LEARNING ACTIVITY

The activity of second language learning has many different forms, varying with the ages, motivations, aptitudes, personalities, learning styles and strategies, locations, and native and target languages of learners [10]. The technologies used to support language learning are equally diverse, covering all language modalities (speaking, listening, reading, and writing) and a range of media types.

In this section, we first present established and emerging technologies for language learning, before going on to describe how we came to develop our own understanding of what it is like to learn the two biggest world languages—English and Chinese. We then discuss our findings in light of research on Second Language Acquisition (SLA).

Established Technologies for Language Learning

Two of the oldest methods for language learning are the Pimsleur system (1967) and Rosetta Stone (1992). The Pimsleur approach is audio-based, with the learner listening to new language items before being asked to recall them in phrases of increasing complexity and at increasing intervals. In contrast, the "dynamic immersion" approach of Rosetta Stone is audiovisual, with the learner inferring mappings between constructions in the target language and photographic representations of the concepts they stand for (without any translation). Together with podcasts of situational dialogues (e.g., ChinesePod), these approaches represent the lesson-based method of delivering content in 20-30 minute chunks of concentrated study.

Unlike lesson-oriented learning, *flashcards* operate at the granularity of facts (e.g., word translations) and measure the learner's ability to provide the correct response (e.g., a Chinese word) in the presence of a stimulus (e.g., an English word). Recent research into the *testing effect* exploited by flashcards—that learning is enhanced when the learner is required to recall information rather than restudying it—has shown that even covert retrieval practice, with no observable user response or immediate system feedback, suffices to increase the degree of learning and reduce the rate of forgetting [30]. Other work has attempted to model human memory to predict future performance based on a history of learning events [31]. However, most algorithms for automated flashcard review are derived from the Leitner system [12], developed in the 1970s for managing piles of physical flashcards. In this system, successive piles represent increasing degrees of knowledge, with a correct response to a flashcard promoting it to the next pile, and an incorrect response relegating it to the first pile. Review is based on the principle of *spaced repetition*—given that humans exhibit a negatively exponential forgetting curve [8], repetitions of things-to-be-remembered should occur at increasingly spaced intervals, just before they are likely to be forgotten.

A typical Leitner schedule is to review each flashcard pile at increasing intervals (e.g., every 1, 2, 4, 8, 16 days). Extending this algorithm, SuperMemo (1987) and Mnemosyne (2003) add graded recollection on a scale of 0–5, Smart.fm (2008) requires selection of an answer from a growing number of alternatives, Anki (2008) allows many cards to be created from multi-field facts, and Lingt (2009) adds levels and achievements to create a social game.

Another example of this shift towards social language learning is the community-based website LiveMocha (2010), which helps complementary native speakers and language learners find and help one another around lessons matched to both learners' abilities. The Hanjamaru system [13] for Koreans learning Chinese, on the other hand, is an example of a 2D platform game built around core learning mechanics of hearing the sounds of Chinese characters each time enemies labeled with those characters are attacked, and building new powers from 'conquered characters'. The Tactical Language and Culture Training System [18] supports role play in a 3D game world, using AI for speech processing, behavior interpretation, and content generation.

Emerging Technologies for Language Learning

Whereas most established systems for language learning have clustered around the methods of content delivery—dialog-based lessons and vocabulary-based flashcards—research has focused on how learning can be made more mobile, more multimodal, more ubiquitous, and more fun.

Much research has targeted learners of English as a Second Language (ESL), including systems for mobile audioblogging of language use experiences in response to oral assignments [16], AI-based chatterbots that engage the learner in simulated human-human conversation [34], and second language interaction within the virtual worlds of computer games (e.g., the SIMs [23]). Other ESL systems have attempted to leverage the context of learning, such as TenseITS [6]—an intelligent tutoring system (ITS) for Chinese learners of English tenses that adapts to context through the learner's *explicit input* of their location, concentration level, interruption frequency, and available time. More automatic location-based learning using *indoor positioning* of PDAs over Wifi has also been prototyped, but not deployed or evaluated as a system [15].

Aside from ESL systems, many research projects have tackled the particular challenges of learning Chinese. For Chinese children learning the stroke order of the 3500 or so characters required for literacy, "Let's play Chinese Characters" [36] gives examples of mobile games that aim to be social and fun compared to rote learning. For foreign students learning Chinese in China, the idea of a Context-Aware Mobile Chinese Language Learning system that would provide language based on the learner's level and location has been proposed but not implemented [1].

Two systems that use location to support teacher-directed learning are the LOCH system for tasks that require interaction in particular places with native speakers [28], and the CLUE system for the association of sentences and language questions with the GPS coordinates of *specific locations* [26]. Two further systems that use RFID for context detection are the JAPELAS system [27] for learning how to talk in tagged rooms and the TANGO system [27] for learning the how to talk about tagged objects.

A similar approach has also been used to support learning *in the home* [2]. A bracelet RFID reader and various sensors are used to detect learners' interactions with objects and appliances throughout an instrumented home, responding with English to Spanish noun translations.

Although we see a number of interesting context-aware systems for language learning, none offers automatic, contextual language at locations across entire cities that would fit into the everyday activities of language learners.

Study of Chinese & English learners in Beijing & Seattle

To gain a deep understanding of specific language learning practices and experiences, we decided to focus our initial research on the world languages of English and Chinese.

We conducted a web survey of English and Chinese native speakers, learning Chinese and English respectively, in both the U.S. (Seattle) and China (Beijing). The survey was circulated on both mailing lists and community forums. In Beijing we had 103 respondents learning English and 64 learning Chinese, whereas in Seattle we had 77 and 59 respectively. All were educated at college level or higher.

Overall, our learners of English had more self-reported experience and ability (more than 7 years, over 4000 words) than our learners of Chinese (less than 3 years, around 1000 words). All groups stated they most wanted to improve their speaking and listening. Drawing from this initial pool, we conducted 23 open interviews in preferred study locations, covering 7 learners of English and 4 learners of Chinese in Seattle, and 6 of each in Beijing. Questions addressed the study, practice, and goal-directed use of the second language as well as existing and potential tool support.

Our main finding was that satisfaction was *not* based on the language or location of learning. Rather, salient attributes were if the learner needed to use the language to achieve other goals (e.g., school or work) and if they had a supportive relationship with one or more native speakers.

The learners with the lowest satisfaction levels were those who studied language but were not able to practice it with supportive native speakers. They were primarily expatriates in Beijing studying Chinese because they lived in China, rather than for work or formal education. This group reported intense levels of frustration, embarrassment, and panic when trying to use Chinese in Beijing (e.g., "If people came with floating subtitles, I probably wouldn't panic as much" and "Their sentence structure is so fucked up. It's infuriating."). They did not use dictionaries or other media to mediate communication with native speakers, but would look up vocabulary before or afterwards. They reported becoming more confident after a study session and then being more likely to take the emotional risk of speaking.

In comparison, the learners with the highest satisfaction levels were those who were happy to learn the language by using it, despite the mistakes and misunderstandings that would sometimes arise. These were predominantly people learning a language spoken natively by friends, partners, or family members who were willing to implicitly recast or give explicit feedback on the learner's incorrect utterances.

Although this latter group of *use-directed* learners was in general more advanced than the former group of *study-directed* learners, language study does not naturally translate into use over time. We learned that our study-directed learners would often avoid using language that they knew unless it was absolutely necessary, and refrain from speaking at all unless they knew in advance what to say and how to say it. This was exemplified by one learner who was reasonably fluent when speaking with the language partner he trusted, but unwilling to negotiate basic exchanges in shops if he didn't know the names for things.

Drawing on theories of Second Language Acquisition

The *usage-based model* of SLA is based on the premise that "language is learned from participatory experience of processing input and producing language during *interaction* in social contexts where individually-desired non-linguistic outcomes are goals to be achieved" [32]. The slow development of our target learners can be explained by their habitual avoidance of such participatory social interactions.

We analyze this behavior by drawing on the accepted five-stage model of SLA [10], which moves from *apperceived input* (noticed) and *comprehended input* (understood), via *intake* (hypothesis formation) and *integration* (grammar formation), through to *output* (speech). The two input stages are strongly affected by the learner's *executive motivation* [7] to place themselves in situations where they can learn the language by using it. It is through second language interactions with native speakers whereby intake is integrated into the learner's unique *interlanguage* [32], which shares elements of both their native and target languages. In our study, embarrassment over the breadth, depth, correctness, comprehensibility, and automaticity of interlanguage were all widely reported, leading to learners raising their *affective filter* [32]. This filter determines the extent to which learners actively seek input, as well as how, when, and where they will choose to speak based on their emotional experiences of prior communication attempts.

Implications for design

In the introduction, we outlined a vision for contextual microlearning based on three principles from Cognitive Psychology—encoding specificity, spaced repetition, and situated cognition. In this section, we have reviewed prior systems for language learning, our own preliminary user research, and theories from Second Language Acquisition (SLA). Given the evidence from these sources, we decided to tackle the challenge of getting our study-directed learners to adopt a use-directed attitude towards learning.

To promote such behavior change, our key design goal became to support *contextual microlearning* that helps study-directed learners to study in context, with the effect of exposing them to more opportunities for speaking.

Contextual language

Learning is best when performed in the context of future use (*encoding specificity*), and this principle is implicit in the many systems for context-aware language learning (e.g., [1, 2, 6, 15, 26, 27, 28]). By providing learners with language that they might need to use in any given location, we can potentially develop their *interlanguage* in a manner more appropriate for the achievement of real-world goals. The implication for design is therefore to build a system that contains language tagged with its context of use and which presents language according to the learner's context.

Contextual study

Learning is best when performed at expanding intervals over time (*spaced repetition*), and this principle is implicit in the many systems for flashcard-based learning. By allowing learners to naturally space their visitation of language content across moments of free time throughout the day, we can potentially create microlearning opportunities that help with learners' *executive motivation* to spend time on language learning. The implication for design is therefore to extend the conventional Leitner flashcard system to space repetitions not only in time, but also across the places that the learners visits during the day.

Contextual use

Learning is best when performed through participatory experience (*situated cognition*), and this principle is implicit in the many systems for conversation-oriented systems (e.g., [6, 16, 34]). By encouraging learners to use the language they are studying, we can potentially help lower learners' *affective filters* in a way that helps them both notice and take advantage of contextual opportunities for speaking with native speakers. The implication for design is therefore to incentivize what we might call the "micro-use" of language at the level of individual flashcards, which we can track and give feedback on along with conventional flashcard study statistics.

DESIGN

Of our three implications for design, *contextual study* and *contextual use* describe scenarios predicated on the realization of *contextual language* tagged with its context of use and triggered when learners enter that context.

Implementation of contextual language

The context-aware systems for language learning reviewed in this paper, as well as many other context aware systems for activities such as understanding virus outbreaks [24], wildlife behavior [3], and climate change [24], are predominantly based on teachers or researchers manually associating learning content with the GPS coordinates of real-world locations from within tightly constrained geographic regions. For a scalable approach to the automatic presentation of language based on a learner's context, we decided to leverage the web API of Foursquare to retrieve nearby "venues" and their types.

Foursquare venue types as contextual language tags

There are many advantages to using an established location-based service for context-aware systems research: the ontology of venue types is pre-constructed, the location-to-venue mapping is pre-populated, and the location data in the new system will automatically improve over time. No existing system for language learning has exploited these advantages for the automatic presentation of contextual language at a city-wide scale across the major world cities.

For our system, we decided to use a model of context based on *venue type*, obtained from the category attributes of the Foursquare venue closest to the learner's GPS location. Foursquare has 8 top level categories (e.g., Food, Shopping, Arts & Entertainment) and 235 subcategories (e.g., Chinese Food, Bookstore, Stadium), some of which branch into a third level of distinction (e.g., types of Stadium: Baseball, Basketball, Football, Hockey, Track, etc.).

Human computation for contextual language tagging

Inspired by human computation for language translation [17, 35], we created a database of English-Chinese translations associated with their context of use. Our native English speaking team members first entered at least five English language words for each venue subcategory of Foursquare and then translated that into Mandarin using Bing Dictionary, Google Translate, and Nciku. The Chinese-English bilinguals on our team then verified the resulting flashcards for accuracy and relevance. In addition, we added context tags to existing flashcards that we imported by spreadsheet from the Mnemosyne project, resulting in around 3000 contextual flashcards in total.

Design of the MicroMandarin mobile application

Our MicroMandarin application supports 4 key functions: studying language based on where you are, using language you have learned based on where you are, browsing all language you have seen through the application, and tracking progress by referring to statistics of flashcards seen, correct, incorrect, learned, and used. These are implemented as four application screens (see Figure 2).

Study screen

The study screen (Figure 2a-b) is responsible for showing flashcards according to the current context. Our extension of the Leitner algorithm to multiple contexts is as follows:

1. For each flashcard, maintain a *correct count* that represents the number of consecutive correct responses to that flashcard. Increment this count when the learner knows the answer, reset to zero when not. The mapping of Chinese and English to the front and rear of flashcards alternates with increasing correct count, and a flashcard is "learned" when this count reaches four[1].
2. For each context, maintain a set of 10 related but unlearned flashcards and a *Leitner count* that determines which flashcards are shown in the current *Leitner session* for that context. Flashcards with a correct count of N are shown to the learner every N^{th} Leitner session such that better known cards are spaced further apart. After every 10 flashcards, present quantitative feedback about progress made in that *microlearning session*.
3. Whenever one of the 10 current flashcards is learned, replace it with the next available flashcard for the current context. For any given context, flashcards are presented according to attributes of that context in the order [*subcategory*, *category*, *general*], where *subcategory* is all flashcards tagged with that subcategory, *category* is all flashcards tagged with the "Other" subcategory of the parent category and *general* is the non-contextual language in the database (~5%).

As in the Leitner system, learners self-report whether after seeing the word on the flashcard front they correctly recalled the translation on the rear.

Use screen

The use screen (Figure 2c) presents flashcards for the current context that have already been learned. Below each flashcard are two buttons, *skip* and *used today!* Skipping a card moves onto the next—the idea is to provide users with a means of browsing the language they have learned in a way that encourages them to use it. If they have used it earlier in the day, or in response to seeing the card, they can reward themselves by hitting the *used today!* button. This results in a dialog box congratulating the learner with the total number of flashcard-based language items they have now used in second language exchanges. When all speaking challenges are exhausted, the learner is invited to study more. This *used today!* button also appears on the study screen once a card has been flipped (Figure 2b), aiming to maximize thinking about language use.

Seen screen

The seen screen (Figure 2d-e) presents a scrollable list-based history of all flashcards seen using the application, summarized by their English translation. Selecting a list-item expands it into a full card, below which are buttons for returning to the list view and for "used today!".

Stats screen

The stats screen (Figure 2f) shows overall statistics of the learner's interactions: how many flashcards have been seen, recalled correctly, recalled incorrectly, learned, and used. The goal of this screen is to encourage thinking about studying and use as related components of success.

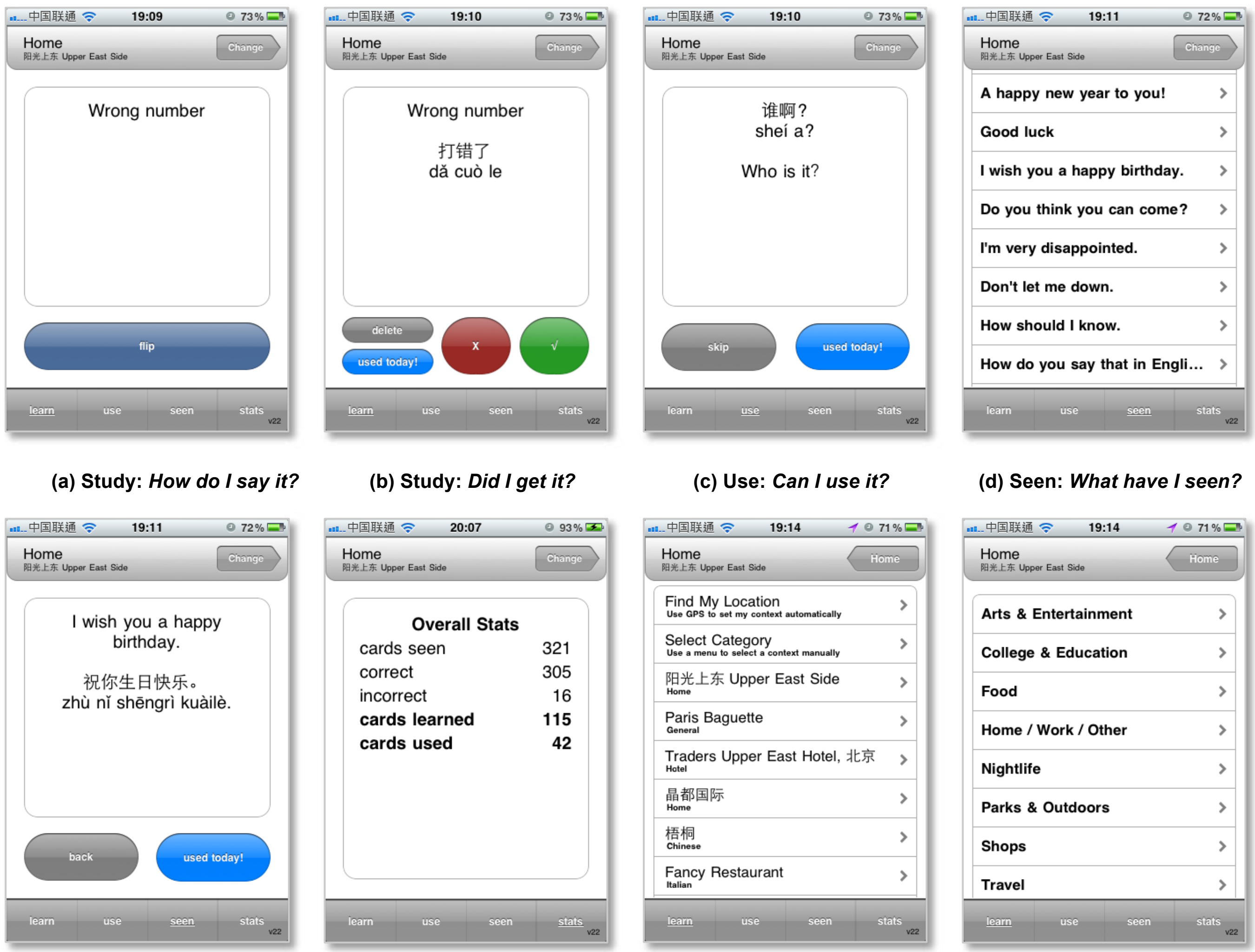
Context bar

At the top of the user interface is a persistent context bar that indicates the name and type of the venue closest to the learner's current location. For example the context bar in Figure 2 lists the venue "Upper East Side" of subcategory "Home", resulting in home-related language being shown[2].

Hitting the *Change* button results in a scrollable list of the three ways in which the context can be changed (Figure 2g). These are *Find My Location*, using GPS to update the context automatically; *Select Category*, using a hierarchy of scrollable lists to set the category and subcategory of the context manually (Figure 2h); or selecting one of 10 nearby locations as returned by the API call to Foursquare. The top such result, representing the closest venue, is automatically selected as the learner's context when the application loads. Although the language content is associated only with the subcategory of the closest venue, not the learner's specific GPS coordinates, we present the venue name to communicate the reasoning behind our selection.

[1] 5-16 exposures are typically needed to learn a word [25].

[2] Upper East Side is an apartment complex in Beijing.



(a) Study: *How do I say it?* (b) Study: *Did I get it?* (c) Use: *Can I use it?* (d) Seen: *What have I seen?*

(e) Seen: *What does it mean?* (f) Stats: *What have I done?* (g) Context: *What is nearby?* (h) Context: *What do I want?*

Figure 2. MicroMandarin User Interface Screens

Technical Implementation

We implemented the user interface as an HTML5, CSS3, and JavaScript mobile client that uses JSONP to communicate in a RESTful way with our Ruby on Rails Web service. We exploit geolocation, offline caching, and animation in Webkit-based mobile browsers, and Rails at the back end to connect with our language database, Foursquare, and mobile carriers. Our client is instantly deployable and updatable through the Web browser, with an additional administrative interface for live interaction with language content, notifications, and user logs.

EVALUATION

To evaluate the effectiveness and potential of contextual microlearning for second language acquisition, we conducted a four-week field trial in which participants were exposed to two versions of our mobile application: the *contextual version* described above, and a *frequency-based version* in which flashcards from the same database were presented in order of most to least common by frequency[3]. The two versions thus shared the same contextually-tagged flashcards but differed in their selection of cards to present to the learner. Matching the difficulty of flashcards with learners' ability levels was beyond the scope of this work.

The *natural order* [25] in which the vocabulary of a language is learned is highly correlated with word frequency, and so being exposed to language in order of frequency is likely to be optimal in terms of pure learning. However, as presented in this paper, many similar such systems exist and form the mainstream approach to language learning. The purpose of this comparison is *not* to

[3] Word frequencies were derived from a Web corpus [33]. Phrases were assigned the frequency of their rarest word.

demonstrate the superiority of learning highly contextual language over language of high frequency. Rather, it is to understand how these approaches complement one another and how contextual microlearning creates experiences that transcend conventional study tools and habits.

Participants

We recruited participants for our study using a Web form that we circulated via direct email to previous survey respondents, postings to social network groups, messages to mailing lists, and Twitter. We received 115 responses in total, which we filtered down to 28 based on the constraints of our study: that participants were fluent English speakers actively learning Mandarin Chinese, were based in Beijing or Shanghai (the only two Foursquare cities in mainland China), used an iPhone as their primary mobile device (the most popular phone in our target user group and which runs an HTML5 mobile browser), and were not travelling for more than 6 days of the four week study period. Ages of the 23 participants who completed the study[4] ranged from 23–42, with 14 males and 9 females. Self-reported language levels were distributed novice low=2, novice mid=6, novice high=9, intermediate low=1, intermediate mid=4, and advanced=1. Internet connection was by Wifi only for 5 participants and by Wifi or data connection for the other 18.

Materials

The frequency-based version of our application was the same as the contextual version except that the *Change* button was removed from the context bar; the context label was filled in automatically based on the subcategory of the flashcard being presented; and the *Use* screen, *used today!* buttons, and *cards used* statistic were removed (since the language was not presented in its context of use).

We augmented both versions with initial dialog boxes that would help us to verify the accuracy of our Foursquare-based location lookup. When the application first loads, we take the top result from the venues nearby the learner and ask whether this is their current location. In the contextual version, we also ask the same question whenever the user manually changes their context (as in Figure 2g). This provided data on the extent to which our participants studied and used language based on their context.

Method

To track real-world speaking as well as application use, we adapted the Experience Sampling Method (ESM) [14] to capture the frequency, quality, content, and purpose of participants' interactions in Chinese with native speakers[5]. SMS links to an online Web survey were automatically sent to participants' phones 4 times a day throughout the study. Participants completed these based on Chinese interactions since the previous survey. Completion of 75% or more of these surveys was a condition of entry into a prize draw for 3 copies of Rosetta Stone Mandarin Levels 1–3. All participants received a gift worth $100 for responding to ESM surveys, trying both versions of the application, and completing a post-study survey. No conditions were placed on application use, other than to use it if they wanted to.

We adopted a within-subjects study design comparing the two application versions, dividing participants into two groups and counterbalancing the condition order (i.e., contextual then frequency for group one and frequency then contextual for group two). Groups were balanced first for self-reported Chinese ability, second for internet connection, third for city, and fourth for gender. The structure of our four-week study was based on an initial week of ESM for everyone, followed by 10 days using one application version, then another 10 days using the other application version. On the remotely-triggered switch between versions, all of the learner's active flashcards were archived such that each card was only ever seen in a single condition. After the study, participants contributed feedback through a Web survey (22/23 completion rate).

Hypotheses

We organize our hypotheses and subsequent presentation of study results around the key components of contextual microlearning that we outlined as implications for design following our initial user research: the need to suggest *contextual language* for *contextual study and use.*

Our hypotheses target six possible differences between the contextual and frequency versions, based on the various ways in which user experiences and usage patterns could vary: language novelty[6] (*H1*), session frequency (*H2*), session length (*H3*), session locations (*H4*), flashcards seen (*H5*), and flashcards learned (*H6*). We used 2-tailed t-tests at P=0.05 with N=23 for all comparisons. Since we only ran tests against this small number of a priori hypotheses, it was not necessary to apply a correction (e.g., Bonferroni) for controlling the family-wise error rate.

RESULTS AND DISCUSSION

We begin by revisiting our preliminary findings on the frustrations of speaking a second language with additional evidence from our post-study survey and over 1700 ESM responses from participants who spanned all ability levels.

Evidence of speaking difficulties

Our ESM data indicated that our participants had around 3 interactions per day with native Chinese speakers (M=3.3, SD=2.5), talking almost exclusively with one person at a

[4] Five participants dropped out due to technical problems or unexpected travel, resulting in 12 of the finishing participants starting with the contextual version and 11 with frequency version. There was no substantial loss of balance.

[5] Guidelines from the American Council on the Teaching of Foreign Languages (ACTFL) [4] were used to craft questions and responses tied to speaking proficiency levels.

[6] The proportion of flashcards seen but not already known (i.e. they were answered incorrectly at least once).

time through a few sentences (45%) or isolated words and memorized phrases (36%), with much fewer extended exchanges (17%) and instances of substantial discourse (2%). Immediate needs were by far the most common purpose (52%), following by social small talk (28%), with work/school and professional being rare (10% and 3%, with 7% other). Speaking mostly felt OK (50%), followed by good (33%), bad (14%), awful (3%) and great (1%).

In our post-study survey, 18% of participant reported almost never becoming frustrated in their interactions with native speakers, 32% about a quarter of the time, 18% half of the time, 9% about three quarters of the time, and 23% almost always. Reasons for frustration include lack of effort from the listener and lack of appropriate words to say:

> *I find talking to new people is the worst case. Because i am a foreigner they already assume they won't understand what I say and they stop trying to figure it out.*

> *Sometimes I think I'm doing well, but then I realize later I made lots of 'rookie' mistakes, relying on set phrases when the situation calls for more specific vocabulary.*

Overall, our study participants experienced many of the same conversational difficulties as our earlier interviewees.

Evidence of context detection accuracy

The greatest number of participants thought that the system was "ok" in its ability to detect their exact location (9% very bad, 23% bad, 50% ok, 4% good, 14% great) and generally better at finding it within list of nearby venues (4% very bad, 18% bad, 36% ok, 27% good, 14% great). This was supported by our logs of context switching: of 435 automatic requests for nearby Foursquare venues, the closest venue was the correct one 208 times (48%). When participants manually switched away from the automatically selected context, they more often selected a context that represented where they were (94 times, 57%) rather than an unrelated context (72 times, 43%). The value of such automatic location detection now and in the future was also supported by participant quotes, such as:

> *I could see using it in nearly every situation - further at work, when shopping, at restaurants, etc., particularly if [the location detection] were highly accurate.*

Although not everybody thought the location detection was good enough—*"the context wasn't accurate enough, however it was useful when it got the location right"*—we found Foursquare to provide an immediate and reasonably accurate solution for our location-based research needs.

Evidence of contextual language appropriateness

We found that 75% of language presented in the contextual version was novel, compared with 55% for the frequency version (P=0.015, *H1* significant: contextual M=0.75, SD=0.25 vs. frequency M=0.55, SD=0.30). This disparity could cause frustration with either version: the frequency version wasting time with language already known; the contextual version wasting time showing cards too difficult for the learner. Learner modeling may resolve this difficulty, but the following quotes and ratings indicate that our system can be considered a successful proof-of-concept at providing contextually-appropriate content in the wild:

> *I really loved the specific vocabulary provided by the Contextual version. For instance, I once logged into the application from a coffee shop ... although I already knew the basics, MM also offered words for items on the menu ("latte," "hazelnut," etc.).*

> *I really like the contextual version because I felt like I was studying language points that were relevant but that I might not have chosen to study on my own.*

In the post-study survey, the greatest number of participants reported that our system was "good" in its quality of language content (4% very bad, 5% bad, 18% ok, 64% good, 9% very good) and in the variety of that content (0% very bad, 9% bad, 18% ok, 64% good, 9% great). However, for the purposes of the study our language set was limited, closed, and impersonal. To function at the level required by the following participant, we would need to open up our means of flashcard creation to the broader user community and investigate the detection of richer contextual features:

> *If it could not only detect location (and from that type of conversation) but also be more specific as to what I want to say. Being at the hospital and getting phrases for "he has a cold" etc. was not useful when we were there to talk about our baby being born prematurely.*

Evidence of contextual study and use

Analysis of our log data showed that the two application versions did not significantly differ in the number of sessions they were used for (P=0.77, *H2* not significant: contextual M=9.7, SD=6.7 vs. frequency M=9.2, SD=11.1). However, usage patterns did differ: the contextual version was used for half as many flashcards per session (P=0.04, *H3* significant: contextual M=10.7, SD=6.3 vs. frequency M=23.6, SD=26.9), but these sessions occurred in twice as many places (P=0.01, *H4* significant: contextual M=9.7, SD=5.6 vs. frequency M=5.1, SD=5.6). In other words, the contextual version resulted in behavior that closely resembled episodes of microlearning constrained by time demands and motivated by concrete needs:

> *Studying the words while in the location (and then, using them immediately) has helped me retain and remember them. I had a similar experience after logging into the app from home to review the names of household objects.*

> *On several occasions I learned office related terms that I then immediately used in my workplace.*

> *I would use the Contextual version to prepare myself for a specific situation (shopping, in a restaurant) or to simply review or learn relevant words on the go.*

> *As a unique learning tool, the contextual version is far more effective and provokes someone to study relative to*

what they're doing. Students might find the frequency version more useful, as they set aside specific time to study, but for a busy businessperson ... having the ability to learn terms that are contextual to your location or activities is fantastic and a differentiator for this tool.

However, in terms of the number of flashcards presented to the user and "learned" according to our implementation of the Leitner algorithm, the significantly longer frequency-based sessions resulted in more cards being both seen (P=0.02, *H5* significant: contextual M=119.0, SD=138.1 vs. frequency M=176.0, SD=181.7) and learned (P=0.01, *H6* significant: contextual M=19.1, SD=30.2 vs. frequency M=41.3, SD=63.2) with the frequency version. Based on other qualitative feedback and quantitative results, we interpret these negative results as follows: (1) studying was substantial in both versions; (2) our implementation of the Leitner algorithm made it possible to "learn" a flashcard in a single sitting, even though such "cramming" is rarely beneficial for real retention [30]. A further longitudinal study is therefore needed to fully understand the implications of contextual microlearning on language competence, performance, motivation, and affect.

Limitations and future work

Our *used today!* button was not used at all by 8 of our participants, and was only applied to 84 of the 2736 cards seen in the contextual version. The main feedback was that while for some people "*it seemed to be effective ... I still remember the words!"*, for others, "*it didn't really work that way, I got to use many of the words, but not always within the same day*". More research is needed on how to explicitly motivate people to use the language they are studying, and we anticipate that more persuasive, game-based, social systems will help provided such motivation.

When asked to submit a preference for either application version, the split was 12:10 only slightly in favor of the contextual version. For beginners, the frequency version simply gave them all they needed:

I would use the frequency version to occupy time when waiting at home or out and do not have specific language needs, but just want to improve my language skills.

Given my level (which is pretty basic), I found [the frequency version] the best because it was a continual reminder to do more and sneak some in. The repetitiveness of it helped me learn/remember.

For learners beyond the mid-novice stage, the frequency-only version was less well-regarded:

I didn't seem to benefit very much from the Frequency version ... the cards were not challenging enough.

Contextual appears to be more useful at my level.

I found the contextual version much more effective. The frequency version was still a very useful learning aid.

I got much more out of the Contextual version because it was able to immediately apply what I was learning and it fulfilled a gap in my language-learning routine. The Contextual version was like having a smart phrasebook.

In a future system, we imagine combining frequency and contextual approaches to language presentation, as well as conventional features like a dictionary, to create a more balanced learning tool providing personalized learning content across a wide range of levels. As suggested:

I'd like the option of both, to be honest! I think they both have their advantages and can't really be compared - they are different products for different needs. An app that can switch between the two would be best.

Given that the greatest number of participants agreed that both contextual and frequency versions encouraged them to study more often than usual compared to their previous habits (contextual: 5% strongly disagree, 31% disagree, 14% neutral, 36% agree, 14% strongly agree; frequency: 9% strongly disagree, 23% disagree, 32% neutral, 36% agree, 0% strongly agree), we hope that such a combination will help promote a "micro" approach to language learning.

CONCLUSION

In this paper we have presented MicroMandarin—a system based on the principle of *microlearning* in which opportunities for language study and use are created through the automatic suggestion of contextual language.

We were interested in how such contextual learning can complement existing language learning tools based on the principle of *spaced repetition*, by leveraging the phenomena of *encoding specificity* and *situated cognition*. To understand the potential trade-offs and synergies between contextual and frequency-based learning, we conducted a 4-week study of expatriate language learners in China.

Through our prior user research, we had already identified expatriates in China as needing more *executive motivation* to spend time learning, since they typically lead lives in which language learning is not a strongly instrumental motivation for career success. These primarily *study-directed* learners had problems interacting in real social interactions with native speakers—their frustrations with the limitations of the *interlanguage* raising their *affective filters* over time, leading them to pass over or even avoid opportunities to practice the language they had learned.

Through its exploitation of the kinds of location-based services that are fast becoming mainstream, our MicroMandarin application goes some way towards demonstrating the value of *contextual microlearning*. Although our current work targets learners who are immersed in Chinese in the developed cities of China, in contrast with prior work on supporting foreign language learning in underdeveloped regions (as in [19]), we hope that future work will be able to bridge these two extremes and improve language learning for all.

REFERENCES

1. Al-Mekhlafi, K, Hu, X. & Zheng, Z. (2009). An Approach to Context-Aware Mobile Chinese Language Learning for Foreign Students. Mobile Business, 340–346
2. Beaudin, J., Intille, S., Tapia, E., Rockinson, R. & Morris, M.E. (2007). Context-Sensitive Microlearning of Foreign Language Vocabulary on a Mobile Device. Ambient Intelligence, 55–72
3. Benford, S., Rowland, D., Flintham, M., Hull, R., Reid, J., Morrison, J., Facer, K. & Clayton, B. (2004). Designing a Location-Based Game Simulating Lion Behavior. ACE'04
4. Breiner-Sanders, K.E., Lowe, P., Mile, J. & Swender, E. (2000). ACTFL Proficiency Guidelines—Speaking. Foreign Language Annals, 33(1)
5. Brown, J.S., Collins, A. & Duguid, P. (1989). Situated Cognition and the Culture of Learning. Educational Researcher, 18(1), 32–42
6. Cui, Y & Bull, S. (2005). Context and learner modeling for the mobile foreign language learner. System, 33(2), 353
7. Dörnyei, Z. & Ottó, I. (1998). Motivation in Action: A Process Model of L2 Motivation. Working Papers in Applied Linguistics 4, 43–69
8. Ebbinghaus, H. (1885). Memory: A Contribution to Experimental Psychology. Translated by Henry A. Ruger & Clara E. Bussenius (1913), accessed from http://psychclassics.yorku.ca/Ebbinghaus/index.htm
9. Ellis, N. (1995). The Psychology of Foreign Language Vocabulary Acquisition: Implications for CALL. CALL, 8(2), 103–128
10. Gass, S.M. & Selinker, L. (2009). Second Language Acquisition: An Introductory Course (3rd ed.). Routledge
11. Gassler, G., Hug, T. & Glahn, C. (2004). Integrated Micro Learning – An outline of the basic method and first results. Interactive Computer Aided Learning '04
12. Godwin-Jones, R. (2010). Emerging technologies from memory palaces to spacing algorithms: approaches to second-language vocabulary learning. Language, Learning & Technology, 14(2)
13. Hanjamaru. http://www.hanjamaru.com/. See also http://www.virtualworldsnews.com/2009/03/hanjamaru-combines-language-education-with-mmorpg.html
14. Hektner, J.M. & Csikzentmihalyi, M. (2002). The Experience Sampling Method: Measuring the context and content of lives. In R.B Bechtel & A. Churchman (eds) Handbook of Environmental Psyhcology. Wiley
15. Hsieh, H., Chen, C. & Hong, C. (2007). Context-Aware Ubiquitous English Learning in a Campus Environment. ICALT'07, 351–353
16. Hsu, Wang, S. & Comac, L. (2008). Using audioblogs to assist English-language learning: an investigation into student perception. CALL, 21(2), 181–198
17. Hu, C., Bederson, B.B. & Resnik, P. (2010). Translation by iterative collaboration between monolingual users. SIGKDD Workshop on Human Computation 54–55
18. Johnson, W.L. (2007). Serious Use of a Serious Game for Language Learning. AI in Education, 67–74
19. Kam, M., Ramachandran, D., Devanathan, V., Tewari, A. & Canny, J. (2007). Localized iterative design for language learning in underdeveloped regions: the PACE framework. CHI'07, 1097–1106
20. Kukulska-Hulme, A. (2009). Will mobile learning change language learning? ReCALL, 21(2)
21. Language Learning Difficulty for English Speakers. http://web.archive.org/web/20071014005901/http://www.nvtc.gov/lotw/months/november/learningExpectations.html
22. Ma, Q. & Kelly, P. (2006). Computer assisted vocabulary learning: Design and evaluation. CALL, 19(1), 15–45
23. Miller, M. & Hegelheimer, V. (2006). The SIMs Meet ESL: Incorporating authentic computer simulation games into the language classroom. Interactive Technology and Smart Education, 3(4)
24. MIT Scheller Teacher Education Program (2008) Augmented Reality Games. http://education. mit.edu/drupal/ar/projects
25. Nation, P. (1990). Teaching and learning vocabulary. Heinle & Heinle
26. Ogata, H. & Yano, Y. (2003). How Ubiquitous Computing can Support Language Learning. KEST'03
27. Ogata, H. & Yano, Y. (2004). Context-Aware Support for Computer-Supported Ubiquitous Learning. WMTE'04
28. Ogata, H., Paredes, J, Saito, N., Yano, Y., Oishi, Y. & Ueda, T. (2006). Supporting Mobile Language Learning Outside Classrooms. ICALT'06
29. Oxford, R. & Shearin, J. (1994). Language Learning Motivation: Expanding the Theoretical Framework. Modern Language Journal, 78(1)
30. Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. Psychonomic Bulletin and Review, 14, 187–193
31. Pavlik Jr., P.I., Bolster, T., Wu, S., Koedinger, K.R., & MacWhinney, B. (2008). Using optimally selected drill practice to train basic facts. Intelligent Tutoring Systems
32. Robinson, P. & Ellis, N.C. (2008). Handbook of Cognitive Linguistics and Second Language Acquisition. Routledge
33. Selection of Chinese Corpora. http://corpus.leeds.ac.uk/query-zh.html
34. Sha, G. (2009). AI-based chatterbots and spoken English teaching: a critical analysis. CALL, 22(3), 269–281
35. Shahaf, D. & Horvitz, E. (2010). Generalized task markets for human and machine computation. AAAI'10
36. Tian, F., Lv, F., Wang, J., Wang, H., Luo, W., Kam, M., Setlur, V., Dai, G. & Canny, J. (2010). Let's play chinese characters: mobile learning approaches via culturally inspired group games. CHI'10, 1603–1612
37. Tulving, E. & Thomson, D. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80(5), 352–373