

# Sentence Recall Game: A Novel Tool for Collecting Data to Discover Language Usage Patterns

Jun Wang  
School of Information Studies  
Syracuse University  
Syracuse, New York 13244  
junwang4@gmail.com

Bei Yu  
School of Information Studies  
Syracuse University  
Syracuse, New York 13244  
byu@syr.edu

## ABSTRACT

Recently we ran a simple memory test experiment, called *sentence recall*, in which participants were asked to recall sentences that they had just seen on the screen. Many participants, especially non-native English speakers, made various deviations in their recalled sentences. Some deviations represent alternative ways to express the same meaning, but others suggest that there are missing pieces in the participants' language knowledge. The deviation data, on the one hand, can provide individual users valuable feedback on their language usage patterns that they may never notice, on the other hand, can be used as training data for automatically discovering language usage patterns in a subpopulation of language learners. This paper presents our attempts to create an enjoyable sentence recall game for collecting a large amount of deviation data. Our results show that the game is fun to play and the collected deviation data can reveal common language usage patterns among non-native speakers.

## Categories and Subject Descriptors

H.5.m [Information Systems]: Information Interfaces and Presentation (I.7)—*Miscellaneous*

## General Terms

Design, Experimentation, Human Factors

## Keywords

Human computation, Games with a purpose, Computer-assisted language learning, Language and memory, Educational gaming

## 1. INTRODUCTION

Common language usage patterns in a subpopulation of language learners have proven to be a very valuable language learning resource. For example, the misuse of English articles like "a" and "the" is very common among speakers who do not speak western European language (e.g. Russian or

Chinese speakers)[6]. These common error patterns span all aspects of language usages, from syntactic level article and preposition usage to deep semantic level collocations and idiomatic expressions. Capturing these common error patterns of particular language learner groups can help language teachers prepare more targeted teaching materials as well as increase language learners' awareness of their own habitual usage patterns.

A traditional way to identify these patterns is to build large-scale annotated corpora of non-native speakers. Because of the high cost to build such corpora, only a few of them have been made available[4], including the proprietary ones like the Cambridge Learner Corpus. The high cost and low accessibility prevents regular language researchers, educators, and learners from using them as research or learning materials.

Motivated by the ESP game and generally the idea of games with a purpose[7, 8], we introduce a game, called *sentence recall game*, to help collect large-scale data for discovering language usage patterns. In a sentence recall task, players are asked to recall sentences that they have just seen on the screen. During the recall, a player may generate a different sentence from the original one. The deviations between the recalled and the original sentences, such as word omission, insertion, or substitution, could mean random noises. However, if the similar deviations occur again and again among different players, we have reason to believe that these deviations can reveal common language usage patterns among them.

We have developed a game prototype and ran a pilot data collection experiment using the Amazon Mechanic Turk. Our results demonstrate that the game is fun to play and the collected data can reveal common language usage patterns among non-native speakers.

## 2. GAME DESIGN

A game consists of  $N$  rounds, and each round involves one or more sentence recall trials. In each trial, a sentence is briefly presented to the player, and then the player is asked to type it by recalling the sentence from his memory. If the player makes a successful recall, he will enter into the next round. A recall is successful if its normalized string edit distance from the original sentence is less than a threshold. If a player fails to recall a sentence, it could mean he does not pay enough attention or the sentence is difficult for him

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

KDD-HCOMP'10, July 25, 2010, Washington, DC, USA.  
Copyright 2010 ACM 978-1-4503-0222-7 ...\$10.00

to grasp the meaning or the wording. In this case, he will be given another recall trial, on a different sentence. The game is over when the player successfully completes the last round or fails  $K$  times in one round.

To increase the enjoyability of the game and to elicit as much as possible useful recall deviations from players, we have employed the following techniques.

### Animating sentences.

A sentence is displayed to a player with the trace of its words gradually fading with time (see Fig. 1). This way of dynamically displaying a sentence is used to mimic the sound effect of speaking a sentence, and in particular, to make it challenging for the player to memorize the specific wording of the whole sentence. As a side effect of the challenge, we could collect more deviation data because a player may remember the meaning of the sentence but forget the exact linguistic form, and thus he has to reconstruct the sentence using the language patterns stored in his long-term language knowledge[5].



Figure 1: An illustration of sentence animation.

### Varying sentence display speed.

Controlling the display or animation speed of a sentence is an important design factor. If the speed is too slow, players may find the game boring and can easily make a perfect recall; if it is too fast, players may not understand the sentence at all and thus generate very poor recalls, and as a result, they might get frustrated. Also, from the perspective of the goal of our game design, we are neither interested in collecting those perfect recalls that are identical to the originals, nor those poor recalls like random noises.

In our design, in the first round of a game, a sentence is displayed at a speed of 400 words per minute (wpm)<sup>1</sup>. When a player enters into the next round, the sentence display speed will increase by 20%. If he fails, the speed will decrease by 10%. We have noted that when a player switches back and forth between successful and failed recalls, such abrupt changes in speed could make one really confused. To avoid this problem, we will consider in the future the use of a sigmoid function to adjust the speed[2].

### Immediate deviation feedback.

Immediately after a recall attempt, a player will be shown both the original and the recalled sentence with deviations

<sup>1</sup>Words per minute is a common measure of reading or writing speed. For the purpose of measurement, a word is standardized to be of the size of five characters, including the space key. The reading speed of an average American adult is around 300 wpm, and with the use of RSVP (rapid serial visual presentation), the speed can easily exceed 400 wpm. Considering that the way of animating sentences in our game bears a resemblance to the RSVP technique, we set 400 wpm as the starting speed.

highlighted. For example, Fig. 2 illustrates that when playing the sentence “*He seems to have washed his hands of the job*”, a player recalled the preposition “*of*” as “*off*”. Such feedback could help players, especially for non-native speakers, to discover their habitual language usage patterns and increase their awareness of the need to improve their language knowledge.

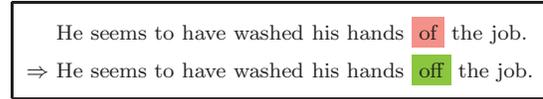


Figure 2: Immediate deviation feedback.

### Scoring.

A player will get points for each successful recall. The amount of points that a player can get depends on the string edit distance of his recall from the original sentence.<sup>2</sup> It also depends on the sentence display speed: the faster the speed, the more the points. The scoring mechanism is designed to encourage players to try their best to make good recalls in the setting of fast sentence display speed—a perfect recall in a low speed setting is less worthy than a near perfect recall in a high speed setting.

## 3. EVALUATION

We seek answers to the following three questions: (1) will people enjoy playing such a simple sentence recall game? (2) will people like to use the game for language learning purposes? and (3) can the game play data reveal the common language usage patterns among the players especially those non-native speakers? In this section we focus on the first two questions, and leave the third one to the next section.

Our evaluation is based on the game play data and the survey data, collected from 33 native English speakers and 57 non-native speakers (including 46 Indian players<sup>3</sup>), who were mainly recruited through Amazon Mechanical Turk. We recruited twice. In the first time we put no restriction on the eligibility of the players, and in the second time we required that the players be from India since we wanted to collect more data from non-native English players—according to a recent report by Ipeirotis (2010), one third of turkers are from India[3].

About 300 sentences are used in the game. They were culled from the Google English Dictionary, with each sentence containing a number of between 8 and 12 words.

### Enjoyability.

A simple analysis of the game log data shows that each player played for an average of 8.9 minutes. Over 1 out of 4 non-native players played the game at least twice—note that a turker only needed to play once to get paid, and about 1 out of 7 native players played twice. Specifically, 3 Indian players played for more than 40 minutes, including one who

<sup>2</sup>Considering that some word pairs, such as ‘can’ and ‘could’, are exchangeable, the distance between them will be 0.

<sup>3</sup>According to their self-reports, our Indian players speak the following 6 (native) languages: Gujarati, Hindi, Kannada, Malayalam, Tamil, and Telugu.

was referred to the game by his (her) turker friend.<sup>4</sup> We believe these numbers show the game was fun to play.

In addition, many turkers also left comments saying that the game was fun to play. Here are a few examples:

*“The words moved across the screen faster as the proficiency rose, and I was able to fill in some words that I did not catch clearly due to the connotation of the sentence. It’s a very addictive game.”*

*“Great simple game. Glad there was a ‘top 5’ at the end to gauge my score. Liked the increasing challenge.”*

*“This was actually one of the more entertaining psych-survey-type games that I’ve encountered. I don’t think I’ve ever seen a memory-style game use that particular method of obscuring the information to be remembered; it made it much less samey than other similar games.”*

*“It is fantastic playing the game. It is a game to increase your memory as well. Very interesting to do this even without earning money.”*

### Designing for learning purposes.

An important feature of our game is that it is also designed as a language learning tool. We asked turkers whether the game would be helpful for improving their language skills, and the responses are summarized in Table 1: 80% of non-native English speakers responded positively, while only 50% native speakers thought so. After receiving comments from an anonymous reviewer, we realized that a better survey question might be: *would you like to use the game for language learning purposes?* Regarding the evaluation of whether the game is effective for improving players’ language skills, we need to run a controlled experiment, but this remains to be future work.

**Table 1: Responses about whether the game could be used for improving language skills.**

	yes	maybe	no
native English speakers (NS)	81%	6%	13%
non-native English speakers (NNS)	49%	27%	24%

We also want to point out that quite a few non-native speakers specifically mentioned that the game is good *because* it can help improve their English. This is an encouraging sign since the game is mainly designed for discovering habitual language patterns among non-native speakers. We believe that, in the long run, an enjoyable game with educational purposes would be more likely to be adopted in a sustainable way.

## 4. LANGUAGE USAGE PATTERNS

Now we move on to the third question, raised in the last section, of whether the collected data can reveal common language usage patterns, especially common usage errors among non-native speakers. We have collected some 2,000

<sup>4</sup>We know this because the turker left a comment: *“I have bookmarked it and will certainly refer my friends to play it.”* And then, 20 minutes later, a player from India, unidentified, played for about an hour, with two short breaks each of a few minutes.

recalled sentences, about 40% of which are perfect recalls. Given that we have 300 source sentences in the pool, so each sentence, on average, only introduces a few recalls. Luckily, thanks to the use of the strategy of random-sampling sentences during the play, we have some 20 sentences being played 20 times or more. By generally analyzing the recalls on all the sentences and specifically examining the recalls on those frequently played sentences, we will demonstrate that some common language patterns can be revealed.

**First**, the collected data shows that non-native players tend to make various grammatical mistakes, which native players rarely did. For example, Table 2 demonstrates that non-native players have a tendency to neglect articles—a result consistent with what happened in the Japanese Learners of English corpus[4]. This shows that the omission of articles poses a serious problem for those non-native English speakers who do not speak western European languages[6].

**Table 2: Non-native speakers tend to omit articles. NS: native speakers; NNS: non-native speakers.**

omission	insertion		substitution					
	NS	NNS	NS	NNS				
the	0	10	the	4	11	a → the	4	7
a	1	8	a	2	1	the → a	2	2

The 9 sentences that were involved in the omission of the article *“the”* by non-native speakers:

He believes food prices will go up in ~~the~~ future.  
 Next door to ~~the~~ garage is a small orchard area.  
 We shall take ~~the~~ treasure away to a safe place.  
 The bullet had entered ~~the~~ top part of the brain.  
 She went into her bedroom and lay down on ~~the~~ bed.  
 He seemed to be just ~~the~~ right sort for the job.  
 He dreams of changing the world for ~~the~~ better.  
 Amy stood quietly in ~~the~~ doorway watching him.  
 One of the office girls was down with ~~the~~ flu. (2 NNS)

**Second**, the game play data shows that non-native players are subject to making mistakes on idiomatic expressions. For example, for this sentence:

*“To all intents and purposes the case was closed”,*

all native speakers recalled the phrase *“to all intents and purposes”* perfectly, and all non-native speakers got it wrong, mutating it to: *“to all the intents and purpose”, “to all intent and purpose”, “to all intents and purposes”,* and *“to all intentions and purposes”*.

Another striking example is the common mistake made by non-native speakers when recalling the sentence

*“He seems to have washed his hands of the job.”*

It turns out that 7 among 11 of them recalled *‘of’* as *‘off’* while no English speaker made this mistake.

These examples reflect exactly how idiomatic expressions are processed and produced in human memory: they are holistically stored and accessible from memory as units, but for non-native speakers, the expressions are stored in pieces and therefore subject to mutation when reassembled[9].

**Third**, the data also shows that the players often provide alternative ways to express the same meaning or paraphrases (see Table 3 for a few examples). If a sentence was recalled by a large number of players, we would be able to show the distribution of alternative expressions and see which ones are more popular than others. Even though this game does not tell apart the right and wrong expressions, demonstrating various expressions and their occurrences would induce the players to compare them and think over which are more appropriate.

**Table 3: Examples of paraphrase.**

---

The boss covers all bases when he sets up a job.
⇒ covers all the bases (1 NS, 5 NNS)
⇒ a new job (3 NNS)
The man sat at his regular table near the window.
⇒ by the window (3 NS, 1 NNS)
⇒ at the window (2 NS, 1 NNS)
Sandman is a cult figure on the local music scene.
⇒ in the local music scene (2 NS, 2 NNS)
I had a pretty good idea what she was going to do.
⇒ a pretty good idea of (2 NS, 3 NNS)
It looked as though it might rain at any minute.
⇒ it might rain any minute (3 NNS)
So this could be the perfect time to buy a home.
⇒ to buy a house (2 NNS)
Within a matter of days she was back at work.
⇒ back to work (1 NS, 4 NNS)
I used to write during my free periods at school.
⇒ in school (3 NNS)

---

Currently, we have not employed any text mining algorithms to discover the language patterns in the collected data—only some simple statistics techniques were used. In the future, when a large amount of data is available, we would need to develop text mining algorithms to automatically summarize the common patterns.

## 5. DISCUSSIONS AND FUTURE WORK

The present work is closely related to the field of *computer-assisted language learning*. A problem for second language learners, especially for those intermediate and advanced learners who use the second language on a daily basis, is that many occurrences of their habitual mistaken language usages never get corrected. Even worse, these habitual usages are hard to detect in that the speakers or writers are not fully aware of them. Our game can be used as a tool to help learners improve their language skills. We believe recall deviation patterns, such as the recall of “wash one’s hands of the job”, as “wash one’s hands off the job”, would assist language teachers/learners to be better aware of the trends and nuances in language learning.

In addition, the game can also be used as a language assessment tool in that the recall deviations can be used to gauge the language proficiency of the players or language learners[10]. It is also worth to note that the emerged common (error) patterns or paraphrases can be used by language

educators to create fill-in-the-blank (cloze) tests.

Our work also contributes to the growing field of automatic error detection in the writings of non-native speakers[1]. Our game, if adopted by a large number of language learners, can generate large-scale training data that can be used in the development of writing error detection systems.

The sentences used in our current experiment were culled from the Google English Dictionary. An interesting extension of the game is to build a game platform that allows users to easily create, customize, and share their collections of sentences for use in different language use scenarios, such as academic writing and everyday conversation. For example, a teacher can load a customized game with sentences tailored to her teaching purposes, and then uses the game play data to find the common patterns of her students.

A limitation of the current form of the game is that it is only applicable to short sentences. One solution could be that a player only needs to recall part of a sentence, like the fill-in-the-blank tests. Good candidates for the blanks could be those formulaic phrases that are used regularly by native speakers[9]. The new approach also has the potential to speed up the pace of game play and greatly reduce players’ burden of typing a whole sentence.

## 6. ACKNOWLEDGEMENTS

We thank Veronica Maidel for helpful discussions and help setting up the Amazon Mechanical Turk task. We also thank the anonymous HCOMP 2010 reviewers for their insightful comments and suggestions.

## 7. REFERENCES

- [1] M. Gamon. Using mostly native data to correct errors in learners’ writing: A meta-classifier approach. In *HLT/NAACL*, pages 163–171, 2010.
- [2] S. Hacker and L. von Ahn. Matchin: eliciting user preferences with an online game. In *CHI*, pages 1207–1216, 2009.
- [3] P. G. Ipeirotis. Demographics of mechanical turk. <http://hdl.handle.net/2451/29585>, March 2010.
- [4] J. Lee and S. Seneff. An analysis of grammatical errors in non-native speech in english. In *IEEE Workshop on Spoken Language Technology*, 2008.
- [5] M. C. Potter and L. Lombardi. Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, 29:633–654, 1990.
- [6] M. Swan. *Practical English Usage*. Oxford University Press, 2005.
- [7] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *CHI*, pages 319–326, 2004.
- [8] L. von Ahn and L. Dabbish. Designing games with a purpose. *CACM*, 51(8):58–67, 2008.
- [9] A. Wray. *Formulaic language and the lexicon*. Cambridge Univ Press, 2002.
- [10] A. Wray and T. Fitzpatrick. Why can’t you just leave it alone? Deviations from memorized language as a gauge of nativelike competence. In *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins, 2008.