

# CALLIG: Computer Assisted Language Learning using Improvisation Games

Luis Morgado da Costa , Joanna Ut-Seong Sio 

♣ Nanyang Technological University, Singapore

♡ Palacký University, the Czech Republic

lmorgado.dacosta@gmail.com

joannautseong.sio@upol.cz

## Abstract

In this paper, we present the ongoing development of CALLIG – a web system that uses improvisation games in Computer Assisted Language Learning (CALL). Improvisation games are structured activities with built-in constraints where improvisers are asked to generate a lot of different ideas and weave a diverse range of elements into a sensible narrative spontaneously. This paper discusses how computer-based language games can be created combining improvisation elements and language technology. In contrast with traditional language exercises, improvisational language games are open and unpredictable. CALLIG encourages spontaneity and witty language use. It also provides opportunities for collecting useful data for many NLP applications.

**Keywords:** improvisation; computer-assisted language learning; language games; creativity; divergent thinking; remote association

## 1. Introduction

The system introduced in this paper is part of a larger project entitled iTELL – a suite of applications looking into applying deep computational parsers to intelligent Technology Enhanced Language Learning environments. iTELL includes several applications, for both English and Mandarin Chinese, exploring how to leverage the broad linguistic knowledge available to deep computational parsers and apply it to pedagogical settings. In particular, this paper focuses on CALLIG (Computer Assisted Language Learning using Improvisation Games). CALLIG comprises a series of fun language games, integrating the principles of improvisation comedy with grammatical error detection and other language technologies in order to create a fun language learning environment.

The main motivation for this project was to create a platform where we could explore improvisation principles as a dimension to gamify certain aspects of second language learning for advanced learners of English. In addition, we were also enticed by the ability to collect new kinds of data that are extremely rare, which can facilitate research in certain niche fields of linguistics and psychology, such as humor and creativity.

The remainder of this paper is structured as follows: Section 2 provides a brief introduction to improvisation, as well as some anchors between improvisation and the rest of the paper; Section 3 discusses some aspects of gamification of learning, followed by Section 4, which discusses the current state of Computer Assisted Language Learning; Section 5 discusses Grammatical Error Detection (GED) in CALLIG; Section 6 provides an overview of the games currently available in CALLIG, followed by a more descriptive description of how we simulate certain aspects of these games in Section 7; Section 8 discusses our current plans for future work; Section 9 discusses some applied usages of the data collected by our system; Section 10 provides release notes; Section 11 provides some concluding remarks and Section 12 contains the acknowledgements.

## 2. Improvisation

Improvisation is a type of performance where performers create the content of the performance as it is performed. There is no predetermined content. Everything is made up on the spot. Such performances can be of music, theater or dance, to name a few possibilities.

Improvisational comedy is a branch of improvisational theatre. There are two main types of improvisational comedy: long form and short form. Long form improvisational comedy consists of a sequence of improvised scenes. A few suggestions would be elicited from the audience for inspiration, which act as the launching pad for the show. These scenes are often related. The thread that links them is discovered and developed as the performance progresses. Short-form improvisational comedy consists of games (generally a few minutes in length). Each game has its own built-in constraints. For example, in the game “Numbers”, players can only speak in sentences with a given number of words. Every game requires inputs from the audience, e.g., an occupation, a location, an emotion, a number, etc.. These suggestions would be used in the scene. We have been using the term “improvisational comedy”, but in fact one of the rules in improvisation is that improvisers do not try to be funny in a performance, contrary to what one would expect. The comic effect produced is a side-effect. In improvisational comedy, the suggestions and the constraints in the games are often incongruous and the comedy often comes from the unexpected connections that improvisers make to link seemingly unrelated ideas together. This, we believe, is one of the sources of humour in improvisational comedy. The popular American TV show “Whose line is it anyway?” is a well-known performance of short-form improvisational comedy. The show consists of a panel of four performers who engage in a number of games where they create characters, scenes, and songs on the spot.

Improvisation promotes, among other things, collaboration, spontaneity, risk-taking and creative language use. The ap-

plied value of such an art form hasn't gone unnoticed. The techniques, the principles, tools, practices, skills and mindsets developed in improvisation have been used for non-performance purposes, such as language learning and corporate training. Many of the major players in tertiary education have improvisation programs for business schools or for communication training (e.g., UCLA <sup>1</sup>, Stony Brook University <sup>2</sup>, and MIT <sup>3</sup>).

## 2.1. Improvisation elements in CALLIG

Improvisation games are regularly performed as theatre performances, involving not just witty language use, but also physicality and most often than not, collaboration with multiple players. For CALLIG, only verbal improvisation is relevant. At this stage, we are only building single-user games, though collaboration is an implementation on our agenda. There are a lot of online resources for improvisation games, though such games might not be directly usable and need to be adapted or designed anew due to the aforementioned reasons.<sup>4</sup>

Excluding physicality and collaboration (for the time-being), both existing and future games of CALLIG (will) contain the following improvisation elements: (i) spontaneity; (ii) random suggestions; (iii) creativity. We discuss each item in turn below.

Improvisation performances are spontaneous. In a performance, improvisers have to react and respond on the spot. Any delay in response due to over-thinking is considered bad improvising. In CALLIG, spontaneity is attained by having a time limit within which the user must finish the task. The time limit differs in different games depending on the difficulty level. We tested multiple time limits with multiple users to decide on a length that is long enough to create tension but not too short to finish the task at hand.

In an improvisation performance, suggestions are elicited from the audience and are incorporated into the performance to highlight both the unscripted nature of the performance and the skills of the performers. In CALLIG, each game begins with a randomly generated prompt to guide the user's input. The prompts could be random words, phrases, numbers, etc. In an improvisation performance, the performers can ask for many suggestions and select among them. In CALLIG, users can also refresh and get a new prompt if they don't like the one they are given.

Improvisation activities are celebrated for their creativity. Creativity contains many aspects. For our purposes, we focus on two cognitive processes, which exist in a lot of improvisation games: remote association and divergent thinking. Creative thinking is the process of putting associative elements into new combinations which either meet specific requirements or are in some way useful (Mednick, 1962). The more mutually remote the elements of the new combination, the more creative the process or solution. Divergent thinking is the process of generating multiple related

ideas for a given topic or solutions to a problem. (Guilford, 1967). Divergent thinking occurs in a spontaneous, free-flowing, "non-linear" manner. In improvisation training, improvisers are told to stop filtering themselves. This inhibition of self-judgment enhances the ability to generate a large number of ideas. All our current and future games (will) require remote association and divergent thinking. Users have to connect words/phrases in an unusual way, forcing them to generate uncommon ideas.

## 2.2. Improvisation in language learning

The most effective learning occurs when the learners are free to explore and discover with the support of scaffolds (the learning paradox) (Sawyer, 2011a). Similarly, in teaching, teachers must allow themselves the freedom to explore within plans, routines and structures (the teacher paradox) (Sawyer, 2011b). This makes improvisation an excellent tool in teaching and learning. Improvisation contrasts with the traditional way of teaching as transmission of knowledge and skills. Instead of a prescribed curriculum and a fixed execution plan, improvisation celebrates openness and unpredictability (Kurtz, 2011). On the other hand, improvisation is never completely free, it occurs within a network of structures, rules and frameworks (Sawyer, 2011b). Each short-form improvisational comedy game comes with its own set of rules and restrictions, these constraints provide a nice platform to anchor and scaffold teaching and learning. Furthermore, improvisational comedy games are highly malleable. The constraints can be customized for various training programs, especially those pertaining to language. In addition to providing contexts for witty language use, improvisation games also provide possibilities of testing particular language skills, for instance, they can be adapted for the teaching of linguistics, covering areas in phonetics, syntax, semantics and pragmatics (Sio and Wee, 2012). Improvisation activities provide varied contexts of language use that do not appear in traditional language classrooms. The entertaining nature of such games makes language learning less repetitive and more enjoyable. CALLIG can thus function as a useful complement to regular classroom teaching and learning.

## 3. Gamification of Learning

Despite being a relatively young topic, gamification of learning has become a trending topic in recent years. As the number of papers published on gamification of learning is fastly growing (Hamari et al., 2014), so is general public awareness and peer scrutiny of its effectiveness.

Gamification is broadly understood as the *use of game design elements in non-game contexts* (Deterding et al., 2011). These can include game mechanics, game dynamics, and frameworks, such as badge or point reward systems, time constraints, limited resources, turn taking, interaction, competition, roleplaying, etc. – integrated in a way that encourages users to achieve some desired learning goals (Tu et al., 2015; Deterding et al., 2011).

An extensive literature review presented by Hamari et al., aiming to answer the question *Does gamification work?* (Hamari et al., 2014), suggests that gamification works,

<sup>1</sup><http://www.npr.org/2012/12/05/166484466/it-s-improv-night-at-business-school>

<sup>2</sup><http://www.centerforcommunicatingscience.org/improvisation-for-scientists/>

<sup>3</sup><http://tll.mit.edu/design/improv-workshops>

<sup>4</sup><http://improvencyclopedia.org/>

despite also suggesting that more rigorous methodologies ought to be used to further research on gamification. Moreover, gamification can be used for multiple domains of learning, including declarative knowledge, conceptual knowledge, rule-based knowledge, and procedural knowledge (Kapp, 2012).

The inherent benefit of gamification is often deemed to come from a positive, intrinsically motivating, “playful” experience – an experience that relate well with improvisation games.

#### 4. Computer Assisted Language Learning

The field of Computer Assisted Language Learning (CALL) had its birth around the 1980s and has been gaining momentum ever since. Throughout the last four decades, Artificial Intelligence’s contributions to CALL applications have been mainly focused on problems like error classification and correction, user modeling, expert systems, and Intelligent Tutoring Systems (Schulze, 2008; Gamper and Knapp, 2002). Individual systems differ immensely. Some focus on one basic language skill (e.g., reading, writing, listening, or speaking), while others look for broader coverage. Some systems have a larger focus on grammar, others on vocabulary, dialogue interaction, pronunciation, etc.

Within the written dimension, the tasks of automated Grammar Error Detection (GED) and Correction (GEC) have attracted much attention from the field in recent years. This is especially true for English, where a myriad of shared-tasks periodically compare and attest the impact of the latest available technology (Dale and Kilgarriff, 2011; Dale et al., 2012; Ng et al., 2013; Ng et al., 2014; Daudaravicius et al., 2016; Bryant et al., 2019).

Gamification in CALL, even though not entirely new, is still widely unexplored. Nevertheless, a few CALL platforms must be acknowledged due to their popularity. Duolingo<sup>5</sup> is one of such applications. Duolingo is a free mobile and web-based platform, where users can learn dozens of different languages through vocabulary and translation-based exercises (Garcia, 2013). It presents gamification elements such as badges, point systems, leaderboards, a skill tree for users to progress through, to name a few. Two other systems, very similar in nature, are Memrise<sup>6</sup> and Quizlet<sup>7</sup>. These two are free mobile and web-based platforms focusing on learning through digital flashcards. Learning through flashcards is widespread in language learning, though in and on its own, it is not specific to language learning. This kind of learning method has been acknowledged concerning the benefits to vocabulary retention (Kornell, 2009), which has undoubtedly contributed to their popularity. Both platforms also include gamification elements such as point systems, leaderboards, time constraints, along with a few different games to explore and learn the content of the flashcards. In addition to these applications, a great number of other similar applications could also be made reference to, with some minor differences. Most language learning platforms available today share, in great part, a lot of the mechanics and goals of the applications mentioned above.

<sup>5</sup>www.duolingo.com/

<sup>6</sup>www.memrise.com/

<sup>7</sup>www.quizlet.com/

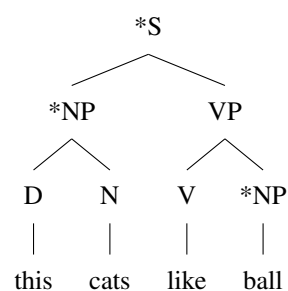
Platforms like Duolingo, Memrise or Quizlet focus on language learning at earlier stages of second language acquisition, in particular vocabulary and simple sentence structures. Our system has different goals. We aim at the training of language skills in the domain of semantics and pragmatics, with unrestricted language, which are more suitable for advanced second language learners. These skills include but are not limited to the understanding of lexical semantics, semantic association, conceptual retrieval, different registers of language use and witty language use. Our system is built based on improvisational principles so it also enhances spontaneity, flexibility and potentially creativity. Improvisational games are also engaging and fun to play because of the accidental generation of humour. All these provide users strong intrinsic motivation to use CALLIG for language learning.

#### 5. Grammatical Error Detection (GED) in CALLIG

CALLIG uses symbolic parsers, such as computational grammars, to perform GED. Symbolic parsers take a long time to develop before being able to compete against statistical parsers on coverage aspects. When coverage is acceptable, however, symbolic parsers generally provide much higher quality and richer analyses of the language. Our system takes advantage of this rich semantic and syntactic information to perform error detection and select feedback based on a concept known as *mal-rules*.

*Mal-rules*, as first proposed by (Schneider and McCoy, 1998), extend descriptive grammars in order to allow specific ungrammatical phenomena, while reconstructing structures that were violated. Although the design of *mal-rules* is time consuming, they can enable fine-tuned error distinctions that statistical parsers would have a hard time dealing with. Consider example (1), below:

- (1) \* *This cats like ball.*



Diagnosing (1) as ungrammatical is just the first step. The subsequent decision of how to correct this sentence is a much harder task. Without context, at least four corrections (2 to 5) should be considered (but more exist).

- (2) *These cats like the ball.*  
 (3) *These cats like balls.*  
 (4) *This cat likes the ball.*  
 (5) *This cat likes balls.*

From a pedagogical point of view, each of these corrections should elicit different kinds of corrective feedback. While

dealing with this ambiguity might seem daunting for some statistical systems, a few *mal-rules* would allow this sentence to be parsed while reconstructing all of the meanings shown above. Describing the inner workings of *mal-rules* is outside the scope of this paper – a fuller account of how *mal-rules* and semantic reconstruction can be used in Computer Assisted Language Learning can be found in Morgado da Costa et al. (2016).

However, the important aspects of using *mal-rules* is that they can apply to syntactic and semantic structures, as well as individual lexical items. This enables *mal-rules* to be used for both error detection and error correction.

From a pedagogical perspective, however, providing a corrected form for an ungrammatical sentence is not enough to engage students in active learning. Because of this, within CALLIG we perform only grammatical error detection. Detected error can then be used to provide feedback messages and guide students towards a successful correction of a problematic sentence.

CALLIG's error detection technology is mostly inspired by and builds on previous work by Suppes et al. (2014) and Flickinger and Yu (2013), who have showed that the use of computational parsers, such as the English Resource Grammar (Copestake and Flickinger, 2000; Flickinger, 2000), to evaluate the grammatical correctness of the written work of individual students yields significant positive results. The use of this kind of technology can be used to reduce the workload of teachers in their attempt to evaluate written sentences produced by each of their students in a timely manner.

## 6. Improvisation Games in CALLIG

CALLIG's ultimate goal is to build a collection of fun improvisation games, explore language learning contexts and provide opportunities for spontaneous and witty language use. We currently have four implemented games. We will give an account of each of these games below.

### 6.1. Sex with Me

**Sex with Me** is a one-liner game. The player will be given a prompt with the form: "*Sex with me is like a/an [object]!*" The object is randomly generated by the system. For words that we consider not common, definitions would be provided. The player can read the definition by hovering the cursor over the word.

The goal of the game is to justify why *sex with me* is like the randomly generated object. The player has to come up with a justification and type it in the answer box within 40 seconds.

Some examples are given below:

Prompt: *Sex with me is like legos...*

Answer: *You need to be imaginative to make it fun.*

Prompt: *Sex with me is like depression...*

Answer: *It makes you want to sleep.*

This game, though a bit risqué, is fun and challenging. It requires the player to quickly find features shared by both sex and the object. The output is often humorous due to the unlikely combination.

There are many similar one-liner games which will likely be added to CALLIG in the future – **Famous Last Words** and **Pick up Lines** are some examples. In **Famous Last Words**, the prompt would be the name of a famous figure (dead or alive; real or fictional), and the input would be the last line that the figure utters before dying, making use of common knowledge of such figures. In **Pick up Lines**, the prompt would be an occupation, and the input would be a pick-up line uttered by someone with that occupation, playing with stereotypes of different occupations..

### 6.2. Haiku on Demand

Haiku is a short form of Japanese poetry, containing 3 lines and comprising 17 syllables: 5 (1st line), 7 (2nd line) and 5 (3rd line). The 3rd line often contains an observation about a fleeting moment in nature. It is simple, direct and intense. It focuses on the juxtaposition of images and a sudden revelation at the end with a sense of enlightenment.

In this game, a random poem title is generated by the system. The generation of title follows one of multiple predefined patterns using a mix of parts-of-speech and frequency information. For example, one of such patterns is the combination of a determiner, an adjective and a noun into a noun phrase (e.g., "*my oversized urinal*", "*the hysterical assumption*"). Another of such patterns is a modified verb phrase, comprised of a uninflected verb and an adverb (e.g., "*conjugate cold-bloodily*", "*internalize pungently*").

After the random title is generated, the user is then prompted to input the three lines of the haiku. A custom-made syllable-checker is ran after the Haiku is completed to confirm that the input has the desired number of syllables. The user has to come up with a haiku of the given title within 90 seconds. Here is an example:

Prompt: *The rude bug*

Answer: *Small and poisonous*

*It lies on the floor, panting*

*And the light turns green*

In the future we would like to explore some variations to this simple setting. These variations include, for example, in addition to a randomly-generated title, there would be a randomly generated word that needs to be placed at the last line of the haiku to force an unexpected ending. Yet another variation could be to reduce the allocated time with every completed Haiku – making it increasingly harder to complete the game.

### 6.3. Wicked Proverbs

A proverb is a well-known piece of wisdom that advises you on how to live properly, for example "*The squeaky wheel gets the grease.*" (intended meaning: those who complain will get attention). Proverbs exist in all languages, but are often language/culture specific (e.g., similar messages are often expressed using different concepts).

The goal of this game is to invite the user to create a proverb-style piece of wisdom using randomly generated *must-use* words and provide an explanation. Some examples include:

Prompt: *Must use: "chocolate" and "chopsticks"*

Proverb: *Sex is like eating chocolate with chopsticks*

Explanation: *if it gets too hot you have a mess*

Proverb: *A good marriage is like chocolate chopsticks.*

Explanation: *always a sweet pair*

The game is timed at 90 seconds per proverb plus explanation, which is deemed to be enough for the user to come up with an idea without allowing the users to think too much about it (which would be undesirable).

Interesting variations of this game could include a randomized required length for the Wicked Proverb (e.g., a number from 5 to 20), which would also influence the number of *must-use* words generated. There could also be a repetition mode, where users are prompted to provide three Wicked Proverbs using the same constraints.

## 6.4. Forced Links

**Forced Links** is an association game. The user is given two unrelated words as a prompt (e.g., *tea* and *koala*), and is asked to come up with a chain of words that would connect the two given words within 20 seconds (e.g., *tea, England, Australia, koala*). It is essentially a game based on semantic association. It reveals patterns of relatedness among different words among different speakers. The two related words given as the prompt can be nouns or adjectives. There is no restriction on the part of speech of the linking words nor on the number of linking words.

## 7. The System

CALLIG is still under development, and it is mainly developed on top of existing open-source platforms. At its core, it is a modular web system developed using Python, Flask<sup>8</sup> and Bootstrap.<sup>9</sup> The system is fully open-source, and easy to expand in scope. The use of flexible web technologies such as Bootstrap also ensures that it can easily be played on mobile devices.

Each game in CALLIG has an introduction page with instructions on how to play the game, as well as a randomized sample of responses by previous players on the top of the page. These responses include information about the author (username) and the time it took them to complete that particular game, which can be used as a competitive measure among players (i.e., being able to come up with a funny response under time pressure can be seen as an achievement). Figure 1 shows the introduction page for the game **Wicked Proverbs**.

The game page differs for each game, but generally include a prompt (e.g., the title, in the case of **Haiku on Demand**; or words that must be used, in the case of **Wicked Proverbs**), some input boxes for the answers and a timer. The duration of this timer varies from game to game, and the user will lose the ability to submit an answer once the timer runs out. Examples of how the user plays these games can be seen in Figures 2 and 3.

### 7.1. Simulation of Audience Suggestions

In an improvisation performance, suggestions are elicited from the audience. In CALLIG, suggestions are randomly

<sup>8</sup><http://flask.pocoo.org/>

<sup>9</sup><https://getbootstrap.com/>

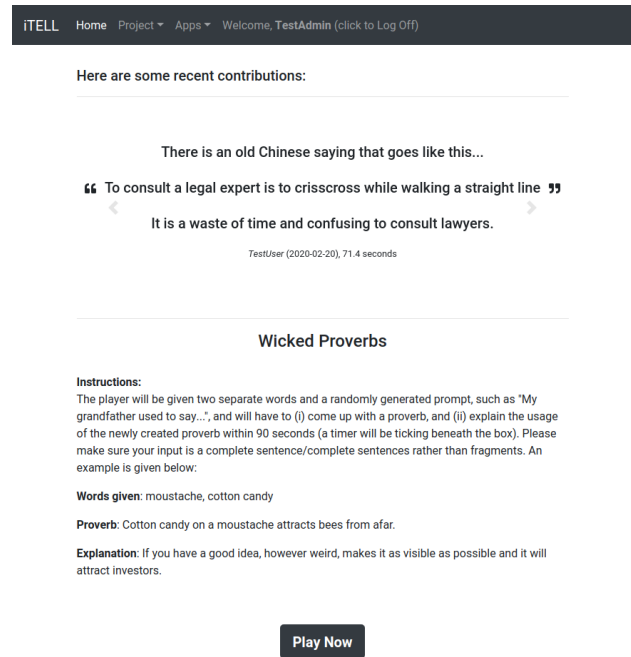


Figure 1: Introduction page for **Wicked Proverbs** game

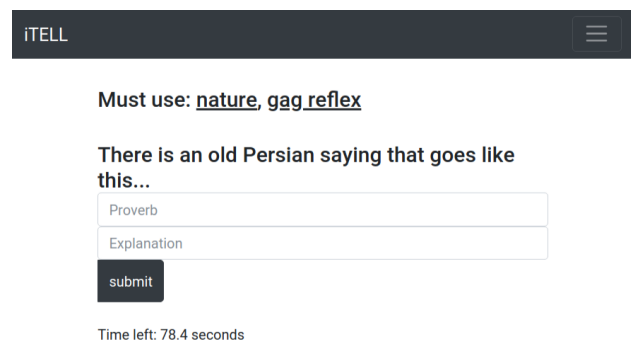


Figure 2: Example of **Wicked Proverbs** being played

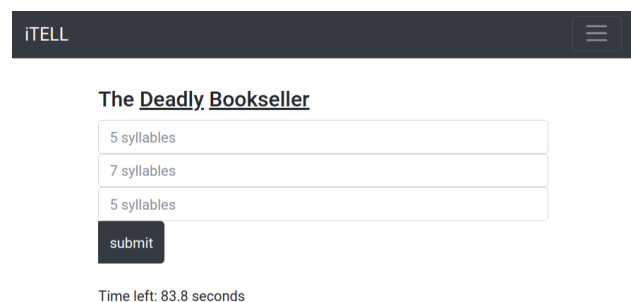


Figure 3: Example of **Haiku on Demand** being played

generated by the system. These two types of suggestions are not identical. Suggestions elicited from an audience are almost always interesting (and potentially amusing) since audience members suggest ideas they want to see developed on stage. Furthermore, the host of the game has the option of choosing a suggestion among the many given. This also gives the option of getting rid of undesirable suggestions. Within CALLIG, audience suggestions (e.g., for the title of the Haiku, or for *must use words* in other games) are gener-

ated with the help of the Princeton English Wordnet (Fellbaum, 1998), which is accessed using the API provided by the Natural Language Toolkit (Bird, 2006).

Wordnets are often large lexical databases, where open class words (i.e., nouns, verbs, adjectives and adverbs) are grouped by sets of synonyms into semantic concepts. These concepts are linked to each other by semantic relations, such as hyponymy (i.e., *a type of*) and meronymy (i.e., *part-whole*). This rich semantic graph also allows the encoding of some measure of semantic distance, which is useful for certain games (i.e., **Forced Links**).

The Princeton English Wordnet is used in tandem with curated wordlists designed specifically for each game. While the wordnet is able to provide a level of true randomness, curated lists of words and expressions are used to maintain a level of familiarity and humor that would be expected from a real-life audience. It should be noted that the system can function perfectly without such curated lists. True randomness sometimes generates concepts that are infrequent, and possibly unknown to the user. Concepts like this come with a definition, provided by the wordnet, that is accessible to users by hovering the mouse on top of the suggested words. This can also be used as a way to introduce new vocabulary to second language learners. The mixture of randomized items from Wordnet and curated wordlists ensure that users won't be given too many unfamiliar words consecutively, which might lead to frustration.

Despite our attempts to control these simulated suggestions the best we can, there is no guarantee that all suggestions are meaningful or sensible. For examples, in **Haiku on Demand**, the system has generated titles like “the weak pisha paysha” and “the handsewn welterweight”. The generation of this kind of nonsensical titles often has to do with semantic mismatch that is too far apart for the user's interpretative accommodation. The current way to address this is to allow users to refresh the game and get a new prompt if they don't like the one they are given. These infelicitous suggestions are kept by the system, and can be used to prevent similar suggestions in the future.

## 7.2. Linguistic Adequacy and Feedback

Whenever appropriate, CALLIG tries to enforce certain degrees of linguistic adequacy. This is the pedagogical dimension of the system. It tries to use each game to enable “learnable moments” throughout the user experience. The system tries to be as precise as possible, ignoring problems when it isn't prepared to provide useful feedback.

This linguistic adequacy takes different forms in different games. In the **Haiku on Demand** game, for example, only answers that respect the syllable count for each line are accepted as a valid answer. If the user fails to follow the 5-7-5 syllable constraint, then they will be notified and prompted to try again. Our hope is that this will raise the user's awareness of how to count syllables, a skill that can help with pronunciation and fluency in a foreign language.

Given **Haiku on Demand**'s poetic nature, there would not be much sense to perform strict grammatical checks in this game. For other games, however, such as **Sex with Me** and **Wicked Proverbs**, grammatical checks are appropriate.

Following the discussion presented in Section 5, CALLIG

is able to identify around 50 different classes of grammatical errors using a special version of the English Resource Grammar (Copestake and Flickinger, 2000; Flickinger, 2000) expanded with *mal-rules*. The selection of these error classes was done using corpora that identified common grammatical errors among undergraduate student population, such as the NTU (Nanyang Technological University) Corpus of Learner English (Winder et al., 2017) and the NUS (National University of Singapore) Corpus of Learner English (Dahlmeier et al., 2013). Our system is currently able to detect a wide variety of common errors. These error classes include: problems with subject-verb agreement; the omission of articles for singular count nouns; the use of indefinite articles with mass nouns; and the use of the wrong form of the indefinite article “a/an”; “their/there” confusion; “its/it's” confusion; irregular forms of past tense, etc.. More than one error can exist in each sentence. And for each error identified in a sentence, the system will generate a constructive feedback message that aims to explain the error and help the user to avoid it in the future. When the system is unsure what is wrong with a sentence, then the error is completely ignored. This is done with the user's experience in mind, as flagging too many ungrammatical sentences might be demotivating for the user. The available error checks and constructive feedback messages present in CALLIG were adapted from the work presented in Morgado da Costa et al. (2020).

Similar to what happens with **Haiku on Demand**, after submitting and answer to **Sex with Me** or **Wicked Proverbs**, the user's answer is checked for grammaticality. Figure 4 shows an example of an answer that was deemed ungrammatical by the system. In this case, the system is able to correctly identify the lack of a determiner before the noun “jungle”.

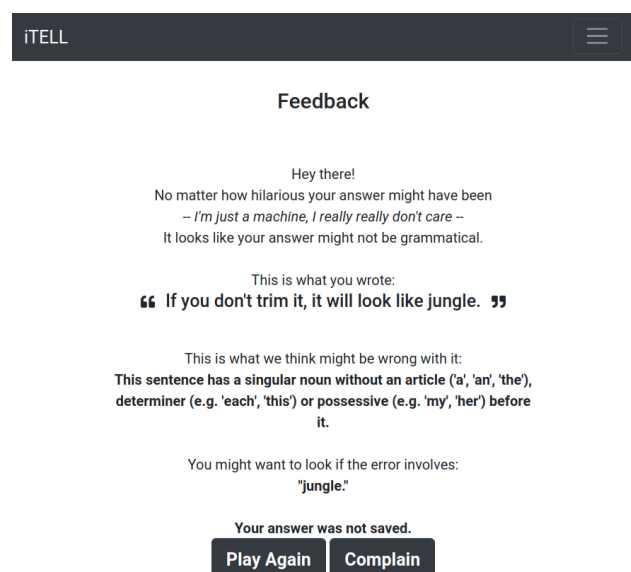


Figure 4: Example of constructive feedback provided for an ungrammatical answer in **Sex with Me**



## 8. Future Extensions

As mentioned before, CALLIG is still very much under development. As such, in this section we will outline the main dimensions we would like to expand our system to cover in the near future.

### 8.1. Future Games

As the central aspect of CALLIG, we are constantly researching possible games to adapt and make available through our system. In principle, games must satisfy at least one of two requirements: a) they must be able to focus on some aspect of language learning that CALLIG is able to control and diagnose; or b) they must produce relevant linguistic data related to creativity or humour that can be used in further research on creativity, humour or language learning. In addition to variations of already existing games, we are currently considering the implementation of the following games:

#### Give Me Ten

This is a game where the user is asked to produce a list of 10 items. This game takes a few forms, for example: a) “*10 ways to describe a/an entity*”; b) “*10 things you cannot do in a structure or geographical location*”; or c) “*10 things you cannot say to a/an person*”

Each pattern contains a semantic category, namely *entity*, *structure* or *geographical location* and *person*. To generate prompts, the system needs only to randomly extract a hyponym of the semantic category in the given pattern, for example: “*10 things to describe a banana*”, “*10 things you cannot do in a church*”, “*10 things you cannot do in China*”, “*10 things you cannot say to a priest*” (etc.).<sup>10</sup>

A potential variation for this game, likely increasing its difficulty, would be to have a pattern containing two slots for randomly drawn items belonging to the same category, for example, “*10 things you can do in France but not in China*”; or “*10 things you can say to Jesus not to but Buddha*”.

#### Reverse Trivial Pursuit

In this game, the system prompt should be interpreted as an answer. The goal of the game is for the user to provide as many possible questions as possible. For example, if the answer generated is “*my intellect*”, some potential questions would be “*What is your most valuable possession?*”, “*What is the sharpest thing in the world?*”, “*What is the thing that makes you unattractive?*”, etc. The system can generate single words or phrases as prompts. As usual, a timer would be used to create pressure and the user would try to input as many questions as possible.

This game would be especially interesting to test and help with the formation of questions, as all the user answers should be in the form of a question.

#### Famous Conversations that Never Existed

<sup>10</sup>This is similar to the challenge “Scenes we’d like to see” in the British panel show “Mock the Week” where comic contestants take turn coming up with witty one-liners on different given scenarios, except that in our game, 10 answers are needed consecutively.

Dialogues introduce different language registers that contrasts greatly with poetic or narrative styles introduced by other games. Dialogues are often casual, and elicit the use of linguistic constructions that are more frequent in this style of writing (e.g., questions, orders, interjections, etc.)

The goal of this game is for the user to create a dialogue between two characters meeting in a specific location. Both characters and locations should be randomly generated. These characters can be fictional characters or famous people (e.g., Batman, Shrek, Jesus, Gandhi) or regular occupations (e.g., plumber, doctor). The location can be geographical locations (e.g., China) or structures (e.g., a submarine, a closet, etc.).

A random number of lines for each character would be generated, and the user would then be prompted to fill the empty lines with a coherent dialogue between the two characters based on possible relations between the characters and their current location. This game tests the user’s ability to relate randomly selected items (characters and location) and weave all the elements into a sensible whole (the dialogue). Ideally, this game would be played line-by-line, and once submitted a line would be irreversible. As usual users would have a timer to complete each line (e.g., 15 seconds). This game allows a great number of variations, which would resort to restricting the users’ input in some way: e.g., disallowing all questions, allowing only wh-questions, providing sentence length restrictions (e.g., minimum, maximum or exact number of words), and provision of *must-use* words for a given line.

### 8.2. Advanced Linguistic Constraints

With the use of the English Resource Grammar (ERG), CALLIG is also able to impose and check for certain classes of advanced linguistic constraints. For example, similar to what happens with **Haiku on Demand**, where the system is checking the number of syllables per line of input, using the ERG would allow our system to check if specific syntactic phenomena had been used. For example, the system could request and check if a passive construction or a definite noun phrase was used in a specific input. These specific linguistic requirements can also be incorporated in game design, e.g., an improvisational language game that focuses on question formation (see **Reverse Trivial Pursuit**). In other words, we would like to further explore the full range syntactic and semantic information provided by the ERG to improve our game design and to more tightly relate our games with certain aspects of language structure and fluency.

### 8.3. Social and Collaborative Gaming

Despite being hosted online, where users can see other people’s answers, the current implementation of our games focuses on a single player environment.

Currently, the system takes the role of the host and the audience in an improvisation game, providing instructions as well as suggestions for the games. Nevertheless, adding social features to it would enable us to dwell deeper into performance style improvisation games, as well as allow meaningful interaction between users. We would like to extend the social and collaborative setting of CALLIG in

two ways:

1. We would like to build a social platform that will enable users to save, post and share the results of their interaction with the system. This platform would allow users to publish the writings they feel the most proud of, and share them on social media. The published works will be accessible to all, and registered users will be able to upvote or downvote other users' writings (possibly even on a scale).

2. Improvisation, as a performance, is generally collaborative in nature, resorting to the use of a "group mind" to create something unpredictable. In the future, we would like to have spaces to explore this "group mind" by introducing collaborative gaming (i.e., chat-room style gaming). Different users can be playing the same game where multiple users are required to work together to create a coherent whole, and each active user takes turns in guiding the development of the narrative. Having this feature would allow us to add games that have been left out because they only make sense in a collaborative setting, as well as different variations of already implemented games. An example of this would be a variation of the game **Famous Conversations that Never Existed**, where instead of having a single user writing the dialog between two characters, we can have two users taking the role of each character and build a dialogue together. This brings in the element of unpredictability and requires flexibility on the user to adjust depending on what the other user has contributed.

#### 8.4. Multilingual Support

Given the mostly language agnostic design of our games, we believe that most of the games currently implemented in CALLIG could easily be supported in other languages. With some exceptions (e.g., the syllable counter for *Haiku on Demand*), most of the language technology we use revolves around the semantic hierarchy provided by a WordNet. Fortunately, resources such as the Open Multilingual Wordnet (Bond and Foster, 2013) include parallel semantic data for hundreds of languages, facilitating this process.

The first language we would like to experiment with is Mandarin Chinese. Mandarin Chinese is a fairly well-resourced language including, for example, the Open Chinese Wordnet (Wang and Bond, 2013) – also integrated in the OMW. Also, despite being a completely optional resource within CALLIG, Mandarin Chinese also has computational grammar – ZHONG (Fan et al., 2015) – which is also being used to build an error CALL (Morgado da Costa et al., 2016). The existence of both a wordnet and a computational parser enhanced with *mal-rules* make Mandarin Chinese an ideal candidate to test CALLIG's ability to support other languages.

### 9. Applied Usages

The creative outcomes (different formats of spontaneous writings) produced by users' interaction with CALLIG will generate a lot of spontaneously written data (e.g., semantic association, humour ranking of different forms of creative writings, etc.). This in turn can serve as a rich resource for both creativity studies and linguistic studies. For instance, games that require complete sentence input (when Grammatical Error Detection is performed) can generate data on

grammatical errors, the game **Forced Links** provides association data between words, **Give Me Ten** can be used to derive commonsense knowledge, or to enrich semantic hierarchies, such as wordnets, by generating definitions and semantic associations for different entities. Games like **Give Me Ten** can potentially also provide data relevant for studies on phenomena such as the Serial Order Effect (Beaty and Silvia, 2012), by confirming whether remote associations are, as expected, reached later than obvious associations.

Improvisation often generates humour. However, improvisation performances are not generally transcribed, humour studies based on improvisation data are rare, if they exist at all. In CALLIG, we hope to have, in the near future, improvisation data with funniness ranking readily available for humour research. The data can then be subject to different kinds of text analysis, or repackaged for experimental use.

### 10. Release Notes

This application is released as part of a larger project entitled iTELL. All its components, including CALLIG, are released under an MIT License. The project is available on Github at <https://github.com/lmorgadodacosta/iTELL>.

### 11. Conclusions

In this paper, we introduced the on-going development of CALLIG (Computer-Assisted Language Learning using Improvisational Games), a system that uses language technology to create online games with elements of improvisation. We have presented four available games, and discussed how we are integrating CALL technology to perform grammatical error detection and are able to provide timely feedback to advanced learners of English.

Improvisation provides opportunities to exercise the muscles of creativity, especially in the area of divergent thinking and remote association. Improvisation games have been used in second language teaching and learning due to its openness and flexibility. Improvisation allows the co-existence of *structure* and *freedom to explore* for both teachers and students, and is deemed an excellent tool for language training. Improvisation exercises in language classrooms, as of now, require the physical presence of a group of people. And despite its possible benefits, it is fair to state that there are people who do not feel comfortable physically performing these games (in public or in a classroom). CALLIG provides a platform for playing language games in a more private and less labour-intensive setting. It can be useful to build confidence before leaping to physical performances, or as training ground for important skills such as spontaneity, collaboration, and risk-taking.

Despite being in its early stages of development, CALLIG is now fully functional, and it has started to generate data in a closed beta environment. We hope this data will soon be useful to multiple lines of research – including but not limited to research on second language learning, lexical semantics, common sense reasoning, humor and creativity.



## 12. Acknowledgements

Even though CALLIG is not yet a public platform at this moment, we have opened it for a selected group of people for testing purposes. This includes, but not limited to, students and staff of the improvisation team from Palacký University, a few linguistic colleagues around the world as well as a few improvisation and stand-up comedians from Singapore. Even though it won't be possible to thank them all individually here due to space, we would like to thank them as a group and we are very grateful for their help and input.

An earlier version of this system was discussed at the 8th International Biennial Conference of the French Association for Cognitive Linguistics (AFLiCo) - "Language, Cognition, and Creativity" in Mulhouse, France (5-7 Jun 2019). We would like to thank the Department of English and American Studies at Palacký University and the Interdisciplinary Graduate School at Nanyang Technological University for their financial support.

## 13. Bibliographical References

- Beatty, R. E. and Silvia, P. J. (2012). Why do ideas get more creative across time? an executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts*, 6(4):309.
- Bird, S. (2006). NLTK: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 69–72. Association for Computational Linguistics.
- Bond, F. and Foster, R. (2013). Linking and extending an Open Multilingual Wordnet. In *51st Annual Meeting of the Association for Computational Linguistics: ACL-2013, Sofia*, pages 1352–1362.
- Bryant, C., Felice, M., Andersen, Ø. E., and Briscoe, T. (2019). The bea-2019 shared task on grammatical error correction. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 52–75.
- Copestake, A. and Flickinger, D. (2000). An open source grammar development environment and broad-coverage English grammar using HPSG. In *In proceedings of LREC 2000*, pages 591–600.
- Dahlmeier, D., Ng, H. T., and Wu, S. M. (2013). Building a large annotated corpus of learner English: The NUS corpus of learner English. In *BEA@ NAACL-HLT*, pages 22–31.
- Dale, R. and Kilgarriff, A. (2011). Helping our own: The HOO 2011 pilot shared task. In *Proceedings of the 13th European Workshop on Natural Language Generation*, pages 242–249. Association for Computational Linguistics.
- Dale, R., Anisimoff, I., and Narroway, G. (2012). HOO 2012: A report on the preposition and determiner error correction shared task. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 54–62. Association for Computational Linguistics.
- Daudaravicius, V., Banchs, R. E., Volodina, E., and Napoles, C. (2016). A report on the automatic evaluation of scientific writing shared task. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 53–62.
- Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15. ACM.
- Fan, Z., Song, S., and Bond, F. (2015). An hpsg-based shared-grammar for the chinese languages: Zhong [—]. In *Proceedings of the Grammar Engineering Across Frameworks (GEAF) Workshop*, pages 17–24.
- Fellbaum, C. (1998). *Wordnet: An electronic lexical database*. MIT Press Cambridge.
- Flickinger, D. and Yu, J. (2013). Toward more precision in correction of grammatical errors. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task, CoNLL 2013, Sofia, Bulgaria, August 8-9, 2013*, pages 68–73.
- Flickinger, D. (2000). On building a more efficient grammar by exploiting types. *Natural Language Engineering*, 6(01):15–28.
- Gamper, J. and Knapp, J. (2002). A review of intelligent CALL systems. *Computer Assisted Language Learning*, 15(4):329–342.
- Garcia, I. (2013). Learning a language for free while translating the web. does duolingo work? *International Journal of English Linguistics*, 3(1):19.
- Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill.
- Hamari, J., Koivisto, J., and Sarsa, H. (2014). Does gamification work?—a literature review of empirical studies on gamification. In *2014 47th Hawaii International Conference on System Sciences*, pages 3025–3034. IEEE.
- Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons.
- Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, 23(9):1297–1317.
- Kurtz, J. (2011). Breaking through the communicative cocoon: Improvisation in secondary school foreign language classrooms. *Structure and improvisation in creative teaching*, pages 133–161.
- Mednick, S. (1962). The associative basis of the creative process. *Psychological review*, 69(3):220.
- Morgado da Costa, L., Bond, F., and Xiaoling, H. (2016). Syntactic well-formedness diagnosis and error-based coaching in computer assisted language learning using machine translation. In *Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA2016)*, pages 107–116, Osaka, Japan. The COLING 2016 Organizing Committee.
- Morgado da Costa, L., Winder, R. V. P., Li, S. Y., Liang, B. C. L. T., Mackinnon, J., and Bond, F. (2020). Automated writing support using deep linguistic parsers (in print). In *Proceedings of the 12th Conference on Language Resources and Evaluation*, Marseille, France, may. European Language Resources Association (ELRA).
- Ng, H. T., Wu, S. M., Wu, Y., Hadiwinoto, C., and

- Tetreault, J. (2013). The CoNLL-2013 shared task on grammatical error correction. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–12, Sofia, Bulgaria, August. Association for Computational Linguistics.
- Ng, H. T., Wu, S. M., Briscoe, T., Hadiwinoto, C., Susanto, R. H., and Bryant, C. (2014). The CoNLL-2014 shared task on grammatical error correction. In *CoNLL Shared Task*, pages 1–14.
- Sawyer, R. K. (2011a). *Structure and improvisation in creative teaching*. Cambridge University Press.
- Sawyer, R. K. (2011b). What makes good teachers great? the artful balance of structure and improvisation. *Structure and improvisation in creative teaching*, pages 1–24.
- Schneider, D. and McCoy, K. F. (1998). Recognizing syntactic errors in the writing of second language learners. In *Proceedings of the 17th International Conference on Computational Linguistics - Volume 2, COLING '98*, pages 1198–1204, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Schulze, M. (2008). AI in CALL – artificially inflated or almost imminent? *Calico Journal*, 25(3):510–527.
- Sio, U. S. J. and Wee, L. H. (2012). Teaching linguistics using improvised comedy. In *Language Arts in Asia: Literature and Drama in English, Putonghua and Cantonese*, pages 283–302. Cambridge Scholars Publishing.
- Suppes, P., Liang, T., Macken, E. E., and Flickinger, D. P. (2014). Positive technological and negative pre-test-score effects in a four-year assessment of low socioeconomic status k-8 student learning in computer-based math and language arts courses. *Computers & Education*, 71:23–32.
- Tu, C.-H., Sujo-Montes, L. E., and Yen, C.-J. (2015). Gamification for learning. In *Media Rich Instruction*, pages 203–217. Springer.
- Wang, S. and Bond, F. (2013). Building the Chinese Open Wordnet (COW): Starting from core synsets. In *Proceedings of the 11th Workshop on Asian Language Resources, a Workshop at IJCNLP-2013*, pages 10–18, Nagoya.
- Winder, R. V. P., MacKinnon, J., Li, S. Y., Lin, B., Heah, C., Morgado da Costa, L., Kuribayashi, T., and Bond, F. (2017). NTUCLE: Developing a corpus of learner English to provide writing support for engineering students. In *Proceedings of the 4th Workshop on NLP Techniques for Educational Applications (NLPTEA 2017)*, Taipei, Taiwan. (IJCNLP 2017 Workshop).