EUROCALL

UNIVERSITY OF COPENHAGEN

Development of an online tense and aspect identifier for English

John Blake[1]

Abstract. This article describes the development of a tense and aspect identifier, an online tool designed to help learners of English by harnessing a natural language processing pipeline to automatically classify verb groups into one of 12 grammatical tenses. Currently, there is no website or application that can automatically identify tense in context, and the tense and aspect identifier fills that niche. Learners can use the tool to see how grammatical tenses are used in context. Finite verb groups are automatically identified, and the relevant words in the verb group are highlighted and colorized according to the tense identified. The latest deployed system can identify tenses in simple, compound, and complex sentences. False positive results occur when there is ellipsis of auxiliary verbs or when the tagger assigns the incorrect part-of-speech tag. The user interface of the tense identifier is a web app created using the Flask framework and deployed from the Heroku platform. The tool can be used for inductive and deductive teaching approaches, or even to check for tense consistency in a thesis.

Keywords: tense, aspect, iCALL, visualization.

1. Introduction

Basic internet searches identify websites that explain the form and function of tenses, and numerous examples of each tense can be found. However, when teachers or learners want to identify the tense and aspect in a particular sentence, no online tool was discovered that could automatically identify the grammatical tense of sentences submitted by users.

1. University of Aizu, Aizuwakamatsu, Japan; jblake@u-aizu.ac.jp; https://orcid.org/0000-0002-3150-4995

How to cite: Blake, J. (2020). Development of an online tense and aspect identifier for English. In K.-M. Frederiksen, S. Larsen, L. Bradley & S. Thouësny (Eds), *CALL for widening participation: short papers from EUROCALL 2020* (pp. 36-41). Research-publishing.net. https://doi.org/10.14705/rpnet.2020.48.1161

As pointed out by Blake (2019, p. 76), the tense and aspect identifier is designed to fill that niche by harnessing a natural language processing pipeline to automatically classify verb groups by grammatical tense. A tool that could identify tense would help learners recognize tenses in context, raising awareness of their usage. As Schmidt (1990) asserts in his noticing hypothesis, learners need to pay attention to particular language features before they can be learned.

The following section introduces the relevant grammatical concepts that underpin this research. The development and evaluation methods are then briefly described. The main deliverable, the tense and aspect identifier, is presented in the results and discussion section. The conclusion describes the usage of the tool, its limitations and future work.

2. Background

The majority of published coursebooks designed for English language learning are structured so that learners are exposed to grammatical tenses in a systematic manner, frequently starting with present tenses. The choice of tense and verb impacts meaning, and so learners need to understand these grammatical tenses. There are two tenses in English: present and past; and three grammatical aspects, namely future, perfect, and progressive. These combine to create 12 verb forms as shown in Table 1. Eight of these forms are tensed (past or present forms) while four are modalized (future forms).

Table 1. Twelve verb forms (grammatical tenses)

	Simple	**Progressive**	**Perfect simple**	**Perfect progressive**
Past	Past simple	Past progressive	Past perfect simple	Past perfect progressive
Present	Present simple	Present progressive	Present perfect simple	Present perfect progressive
Future	Future simple	Future progressive	Future perfect simple	Future perfect progressive

Textbooks designed to help non-native speakers of English make extensive use of these 12 verb forms or grammatical tenses (Yule, 1998). According to Biber et al. (1999), 85% of all verb forms in spoken English are tensed. The frequency distribution of tenses is skewed to favor simple forms that account for 90% of all cases. The four verb forms that are classed as perfect progressive make up less

than 0.5% of all verb forms. Despite their relative rarity, they regularly feature in language tests that focus on grammar, such as Japanese university entrance examinations. Tense usage varies greatly between languages. Table 2 shows three sentences using present progressive tense in Japanese, which when translated into English result in three different tenses. This lack of a one-to-one mapping causes significant challenges for learners.

Table 2. One-to-many mapping between verbs in present progressive in Japanese and English

Japanese	Transliteration	English translation	Tense for translation
本を読んでいる。	hon o yonde iru	She is reading a book.	Present continuous
愛している。	ai shite iru	I love you.	Present simple
日本に十年滞在している。	Nihon ni jyuunen saizai shite iru	He has stayed in Japan for ten years.	Present perfect simple

3. Method

Automatic identification of grammatical tense is a non-trivial task. Six different systems were created using different approaches. The approaches included a rule-based approach, a hybrid machine learning approach, a deep learning model, and different design structures using the Natural Language ToolKit (NLTK) (Bird, Loper, & Klein, 2009). The NLTK is a popular collection of libraries and programs that can be used by software or scripts written in Python to process English texts.

A mobile-first approach was adopted for the design of the graphical user interface in all the systems. Scalability was also prioritized to cope with large numbers of simultaneous users. The systems were compared on usability, accuracy, and processing speed. A series of problem-discovery usability tests were conducted with 30 Japanese computer science majors undertaking a course in computational linguistics and educational technology. Each test involved students attempting tasks with the aim of uncovering problems with the interface and system.

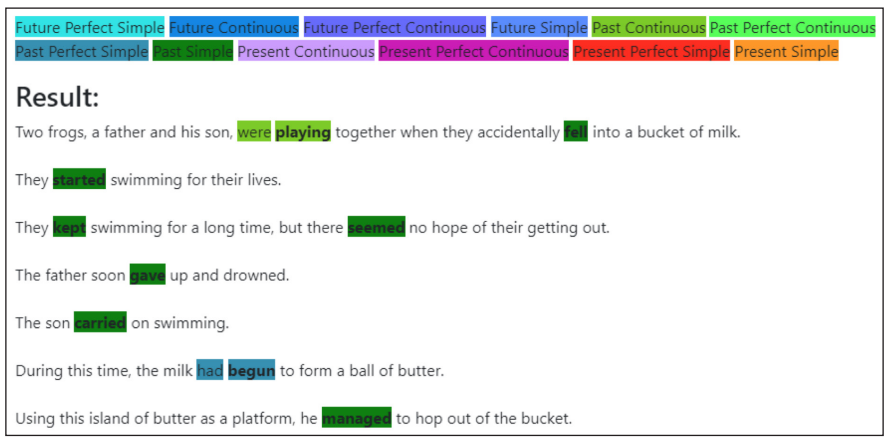
Accuracy tests were also carried out by the undergraduates using two tailormade datasets: one dataset balanced by tense and voice, and the other balanced by tense and function. Processing speed was measured coarsely as *fast* when results were displayed immediately or *slow* when users had to wait for results to load.

4. Results and discussion

The system with the highest scores for usability and accuracy was achieved via a pipeline created using NLTK and a parse tree stored server-side in a Python dictionary. This system is the focus of this article. The systems that processed only single sentences, however, were significantly faster with no noticeable wait time before the results were displayed. The user interface of the NLTK tense and aspect identifier is a web app created using the Flask framework and deployed from the Heroku platform.

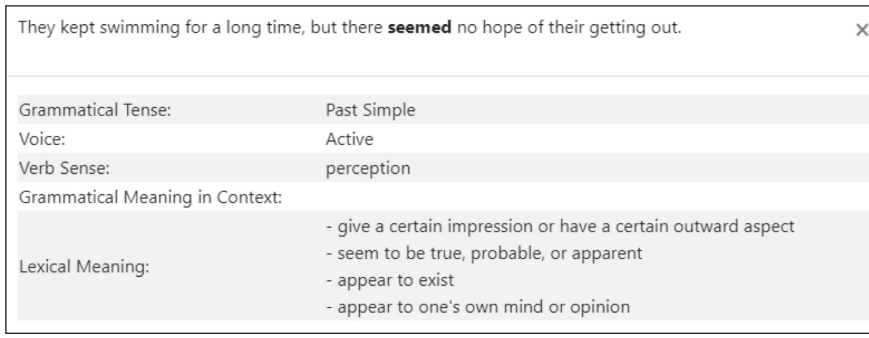
This tool automatically identifies finite verb groups, highlights the relevant words, and colorizes the background of the verb groups according to their grammatical tense. Teachers and learners of English can use this tool to identify the 12 grammatical tenses in any sentence, simply by typing or copy-and-pasting sentences into the input box. The result is displayed once the text is submitted. Figure 1 shows the output of a simple short story.

Figure 1. Screenshot of result showing grammatical tenses detected in a simple short story



Additional functionalities are being added into this interface. When clicking on a colorized verb group, further details such as voice and lexical meaning are displayed in a pop-up box, as shown in Figure 2. Multimedia explanations in English and Japanese have also been prepared but are not yet deployed. There is also space for grammatical meaning in context, but that functionality is not yet implemented.

Figure 2. Screenshot of additional information available in pop-up box



Feedback received during usability testing on the currently deployed version with both teachers and learners was positive. One usability study participant actually checked his own graduation thesis with this tool and found a number of errant uses of present tense in place of past tense in the method section.

5. Conclusion

The tense and aspect identifier can find finite verb groups, highlight relevant words, classify tense into one of 12 categories, and colorize the verb groups according to grammatical tense. The latest deployed system (https://www.jb11.org/tense-identifier.html) can identify tenses in simple, compound, and complex sentences.

Students and teachers of English may find this tool useful to identify not only the tense of the sentence, but also to understand the complex interaction between tense, aspect, and modality. The tense identifier, as mentioned elsewhere (Blake, 2019, p. 76), can be used to learn the English tense system either inductively by inputting numerous sentences and working out the rules, or deductively to confirm whether sentences input conform to the rule. The tool can be used to raise learners' awareness of how grammatical tenses are used in context. A discovery learning approach, in which learners work out the tenses in a text and then compare those with the colorized output, can also be used.

In the next release, the parse tree will be refined to take account of elided auxiliary verbs for transitive verbs as the current system can only deal with elisions for intransitive verbs or transitive verbs that share the same direct object. The accuracy of alternative part of speech taggers will be compared and the tagger with the highest

accuracy will be incorporated. Additional information will also be displayed in the pop-up box, including grammatical meaning in context, and video explanations of the tense.

6. Acknowledgments

This research was supported by a JSPS *Kakenhi* Grant-in-aid for Scientific Research (C) entitled *Feature visualizer and detector for scientific texts*, Grant Number 19K00850. Additional funding was also received from the University of Aizu Competitive Research Fund. I would like to thank Mr Tran Vu Duc and Mr Simon Pavlik for their invaluable contributions to the codebase.

References

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* Longman.

Bird, S., Loper, E., & Klein, E. (2009). *Natural language processing with python.* O'Reilly Media Inc.

Blake, J. (2019). *Tense and aspect identifier* [book of abstracts]. JALTCALL. https://conference2019.jaltcall.org/files/2018/09/JALTCALL2019_Handbook.pdf

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11,* 129-158.

Yule, G. (1998). *Explaining English grammar: a guide to explaining grammar for teachers of English as a second or foreign language*. Oxford University Press.

Published by Research-publishing.net, a not-for-profit association
Contact: info@research-publishing.net

CALL for widening participation: short papers from EUROCALL 2020
Edited by Karen-Margrete Frederiksen, Sanne Larsen, Linda Bradley, and Sylvie Thouësny

Publication date: 2020/12/14

Typeset by Research-publishing.net
Cover theme by © 2020 Marie Flensborg (frw831@hum.ku.dk), based on illustration from freepik.com
Cover layout by © 2020 Raphaël Savina (raphael@savina.net)

ISBN13: 978-2-490057-81-8 (Ebook, PDF, colour)

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: décembre 2020.