So, H. J. et al. (Eds.) (2020). Proceedings of the 28th International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Automatic identification of tense and grammatical meaning in context

John BLAKE
Center for Language Research, University of Aizu, Japan
jblake@u-aizu.ac.jp

Abstract: This paper describes the development and evaluation of the Tense and Meaning Identifier (version 2.0). This prototype pedagogic tool categorizes finite verb groups in simple sentences into one of twelve grammatical tenses, and categorizes the verb into classes, such as stative and dynamic, and where applicable subclasses, such as punctual or durative. Using the results of the tense and verb class categorization, the grammatical meaning in context is predicted. Drawing on the Natural Language Toolkit, a program was written to classify finite verbs by their tense and aspect. A tailor-made list of verbs and their associated verb classes and subclasses was created by crawling the web and extracting lists from grammar books. The tailor-made lists are stored in a Python dictionary. A tense-class Python dictionary was also created to look up the corresponding meaning in context. A web app with a submission form was created to enable online submissions and show the tense, verb class and meaning in context for simple declarative sentences. The tense identifier is able to relatively accurately (69% to 100%) identify tenses, but further development is necessary to reduce false positive results. The limitations of this prototype are detailed and suggestions for further work provided.

Keywords: grammatical meaning, tense identification, aspect identification, NLTK

1. Introduction

Learners of English language dedicate many hours to understanding the complex tense and aspect system. Grammatical aspect is inextricably intertwined with lexical aspect. One realization of this association is that stative verbs do not normally take progressive forms. The combination of tense and verb class can be used to predict the meaning of the grammatical tense in context. For example, the most common grammatical meaning of past progressive used with durative verbs (e.g. *walking*) is to describe an action in progress at a point or period in the past as illustrated in the example: *I was walking when I tripped.*

Should language learners wish to know the lexical meaning of a verb, they can look it up in an online dictionary. However, should they wish to know the grammatical meaning of a verb in context, there is currently no way to look it up. In fact, there is no online tool that can display the tense and aspect, or even verb class on demand. The Tense and Meaning Identifier (version 2.0) aims to fill this niche by identifying and displaying the tense, verb class and grammatical meaning in context for finite verb groups in simple declarative statements.

2. Background

Tense, strictly speaking, refers to either the past or the present. However, textbooks also include future, which to grammarians is usually considered as an aspect, due to its inherent subjectivity. There are multiple ways to categorize verb forms, but the twelve-tense system is one of the most popular. It should be noted that the name of the tense refers to the type of words in the verb form and not necessarily the time zone that the tense is used to refer to. For example, present simple tense can be used to refer to the past or future as shown in the following examples:

- A man **walks** into a bar and **says** “ouch”. [referring to past]
- I **can do** it tomorrow. [referring to future]

Textbooks designed to help non-native speakers of English make extensive use of the twelve verb forms or grammatical tenses (Yule, 1998, p.54). The twelve grammatical tenses, exemplified in Table 1, combine tense, aspect and mood to convey 26 common grammatical meanings in context. These 26 categories can be more finely subdivided into 45 categories (Quirk and Greenbaum, 1993).

Table 1. *Table showing the dynamic verb swim in twelve grammatical tenses*

	Past	**Present**	**Future**
Simple	swam	swim	will swim
Progressive	w(as|ere) swimming	(am|is|are) swimming	will be swimming
Perfect simple	had swum	ha(s|ve) swum	will have swum
Perfect progressive	had been swimming	ha(s|ve) been swimming	will have been swimming

Figure 1 exemplifies the process to determine tense and meaning in context. Users input a sentence. Using the part-of-speech (POS) tags, the verb group is identified. The verb tense and the verb class are matched. The grammatical (and not the lexical) meaning is then determined.

Sentence:	I **was driving** my car then.
Expected POS tags:	VBD + VBG
Verb group:	was driving
Verb tense:	past progressive
Verb class	dynamic, durative
Grammatical meaning:	an action in progress at an earlier time

Figure 1. Key concepts required to identify grammatical meaning in context

3. Development

The server-side of the system architecture uses scripts created in Python to identify tense, determine verb class, and predict the grammatical meaning in context while the client-side is a web app with a simple submission form.

A program in Python that utilizes tokenization and part-of-speech (POS) tagging function of the Natural Language Toolkit (NLTK) (Loper and Bird, 2002, Bird, Loper, and Klein, 2009) was created. The Penn Treebank tag set (Marcus, Santorini, and Marcinkiewicz,1993) was used. The verb group can be identified by the POS tags, and regular expressions can then used to match particular permutations of parts of speech (POS) and word forms to identify the grammatical tense of the verb group.

A Python dictionary of keys (verbs) and values (classes) was created by collating lists of verbs collected and harvested from various language learning resources. As the largest class of verbs is the durative subclass of dynamic verbs, this is assigned as default. The dictionary items then override this assignment. Polysemy causes issues, since some verbs are stative with one lexical meaning but dynamic with another meaning.

A matrix of the twelve grammatical tenses and two main classes (stative and dynamic) was created. This resulted in 24 cells; however, stative verbs were not used in four cells. Following this step, the possibility of different meanings for the five subclasses of verbs was considered for each of the cells. When a change of meaning was discovered while testing prototypical verbs in each of these classes, these subclasses were appended. The final matrix served as the basis of the design of the layout of a Python dictionary that stores the tense-class and meaning in context.

The web app is hosted on Heroku and deployed directly from Github using the Flask framework harnessing WTforms. Figure 2 shows a screenshot of the output of a submission.

4. Evaluation

The accuracy of the tense identification was evaluated using a balanced dataset of 218 simple declarative sentences. The results for the less common tenses, e.g. future perfect progressive were excellent, but the results for the most common tenses ranged from around 70% to 90%.

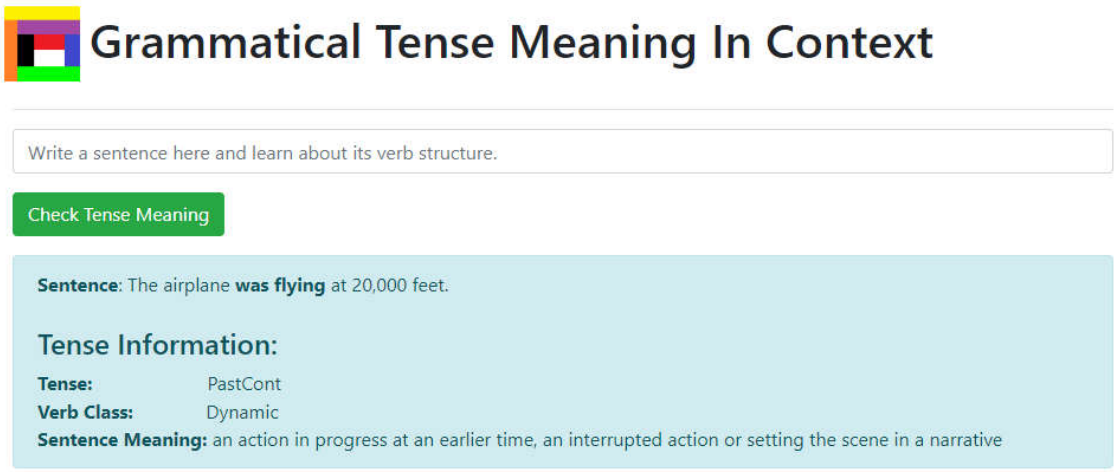

Figure 2. Output for simple sentence showing tense, class and meaning

Table 2. *Table showing percentage of tenses identified accurately for each of the twelve forms*

	Past	**Present**	**Future**
Simple	88.89%	88.00%	72.72%
Progressive	100.00%	100.00%	69.23%
Perfect simple	100.00%	100.00%	100.00%
Perfect progressive	100.00%	100.00%	100.00%

The current caveats with this system include (1) part-of-speech tagging errors resulting in misidentification of the finite verb, (2) the lack of a comprehensive dataset for verb classes, (3) the inability to disambiguate between the lexical meanings of polysemic verbs, which in turn affects the ability of the system to determine the grammatical meaning in context. Additional limitations include that (4) only simple declarative sentences are addressed, and (5) just one verb group is processed.

5. Conclusion

To address the limitations of the system the next prototype will use a parse tree that matches the POS tags rather than a series of conditional statements. The verb class dictionary will be replaced by drawing on WordNet to identify supersenses for verbs, one of which is stative. A more finely grained matrix of grammatical meanings in context will be constructed using the 180 permutations of fifteen supersenses and twelve tenses. In addition, the system will be extended to deal with multiple verb groups. The Tense and Meaning Identifier (version 2.0) is available at https://tense-identifier.herokuapp.com/ .

Acknowledgements

I would like to thank Mr. Simon Pavlik for his invaluable contributions to the codebase. 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. Addition funding was also received from the University of Aizu Competitive Research Fund.

References

Bird, S, Loper, E. & Klein, E. (2009). Natural Language Processing with Python. Sebasopol, CA: O'Reilly Media.
Loper, E. & Bird, S. (2002). NLTK: The Natural Language Toolkit. arXiv preprint cs/0205028.
Marcus, M.P., Santorini, B. & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. Computational Linguistics, 19 (2), 313—330.
Quirk, R. & Greenbaum, S. (1993). A University Grammar of English. Hong Kong: Longman.
Yule, G. (1998). Explaining English Grammar: A Guide to Explaining Grammar for Teachers of English as a Second or Foreign Language. Oxford: Oxford University Press.