Modeling Language Learners' Knowledge State: What Are Language Students' Free Written Productions Telling Us?

Sylvie Thouësny and Françoise Blin

School of Applied Language and Intercultural Studies
Dublin City University, Dublin, Ireland
sylvie.thouesny@mail.dcu.ie, francoise.blin@dcu.ie

Language learner models usually provide intelligent tutoring systems with information about the learner's knowledge state, i.e. the individual's weaknesses and strengths in the target language. The information is normally collected via answers to pre-defined written production types. However, storing information from essay-type questions is more challenging, as this requires the use of instruments that can differentiate errors from mistakes. This paper investigates whether observing incorrect, as well as correct forms in the learners' input provide us with a better insight into the learners' competence. It explains how an analysis of an error-tagged and part-of-speech encoded learner's input help compute the ratio of incorrect to correct forms. The results of a preliminary analysis focusing on morpho-syntactic errors show that the exclusive use of this ratio may, in some cases, be inadequate to discriminate errors from mistakes and, therefore, to represent the learner's knowledge in terms of competence.

1. Introduction

Providing diagnostic feedback to language learners is a controversial, yet enticing and challenging topic for intelligent computer assisted language learning systems, not to mention language teachers. Computers as well as humans have to determine somehow which learners' incorrect forms are worth spending time on. Indeed, a deviance from the norm due to tiredness or emotional states will be regarded as a slip of the pen. From a sociocultural perspective, incorrect forms may be considered as genuine slips of the pen when learners rely on themselves to correct their own written texts without help or minimum help (Aljaafreha & Lantolf, 1994).

Mistakes, or slips of the pen, are considered as occasional lapses in a learner's performance as opposed to errors that represent gaps in a learner's knowledge or competence (Ellis, 1997). Therefore, differentiating errors from mistakes in students' free written productions will help model the learners' competence, i.e. the underlying knowledge that will provide information on learners' strengths and weaknesses.

While learner modeling enables intelligent tutoring systems to observe, record, analyse and even infer reasons of an ill-formed word (Heift & Schulze, 2007), it has been noted "that student performance cannot be directly mapped to knowledge", this being due to variables such as slips of the pen that affect the knowledge representation in terms of competence (Beck & Chang, 2007, p. 138).

For example, Michaud & McCoy (2003) aim to capture the learner's performance in terms of grammar proficiency by comparing user-written essays to stereotype expectations. Even if the learner's performance is (a) compared to an expert's knowledge and (b) considered as a subset of this expert's knowledge, slips of the pen are not taken into consideration.

In this paper, learners' natural language input is, with regard to form-related features, analysed to shape the extendibility to which correct as well as incorrect forms provide information about the learner's knowledge. More specifically, we are interested in investigating the following question:

How can we interpret the scores, i.e. the ratio of incorrect to correct forms in language learner's input?

Measuring the distance between scores and amount of assistance learners require when self-editing their own incorrect forms will assess the validity of this ratio. In the following, we present the participants and their self-editing tasks. The paper then analyses the data encoding processes and, how a computer assisted error encoding program, Markin [1], can be used in conjunction with a part-of-speech tagging tool, TreeTagger[2] to tag a corpus of free texts produced by language learners. Finally, results and future work are discussed.

1 www.cict.co.uk/software/markin/index.htm

2 www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger

2. Participants and their Tasks

During spring 2008, a study was conducted at a university in a French language course that included 76 learners of French in their first year, in which 13 students consented to participate in this study. All participants had English as their first language.

Participants were asked to write a 200 word film review. To do so, students were provided with four short films that were relatively unknown in order to maintain learners' originality and authenticity in terms of written productions. Afterwards, participants were asked to self-edit their own work twice during a laboratory session. They were provided with two electronic copies of their own text. For the first self-editing exercise, all errors were highlighted. For the second self-editing exercise, all errors were highlighted and information about the error types were included. This information was visible with a mouse roll-over action on the incorrect forms.

Participants were asked to fill in the blanks for each highlighted incorrect form. Figure 1 shows how error types were provided to students during the second self-editing exercise.

Fig. 1. Second self-editing exercise

3. Procedure

The learner's input is initially error-encoded then part-of-speech-encoded. Both encoding processes combined together provide a representation of the learner's performance, which is then compared to the self-editing data to shape the learner's competence. A diagram of data encoding processes is provided in Fig. 2.

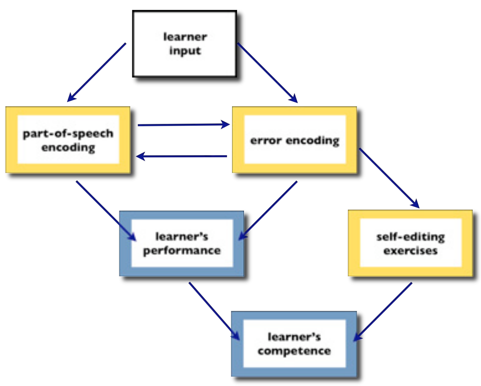

Fig. 2. Data encoding process

3.1 *Error encoding*

The error classification includes (a) selection errors such as the use of an incorrect lexeme or gender, (b) syntactic errors that relate to the syntax of the sentence, including omission and addition, (c) morpho-syntactic errors that are characterised by an incorrect, missing or misplaced morpheme in a semantically correct word and finally (d) spelling errors that refer to semantically correct word, however incorrectly spelled. For the purpose of this study, the analysis primarily focuses on morpho-syntactic errors, listed in Table 1.

Table 1. Error types considered

category	*subcategory*	*error type*
morpho-syntactic	agreement	determinant/noun noun/adjective pronoun/antecedent past participle subject/verb
	formation	plural conjugation partitive determinant

Only one correction per word, or group of words, was carried out. Therefore, a level of precedence over the different error categories was defined; from high to low: selection, syntactic, morpho-syntactic and spelling.

Error tags were inserted using an application whose utilisation is targeted to help teachers correct texts submitted electronically, namely Markin, see Fig. 3 for a screenshot.

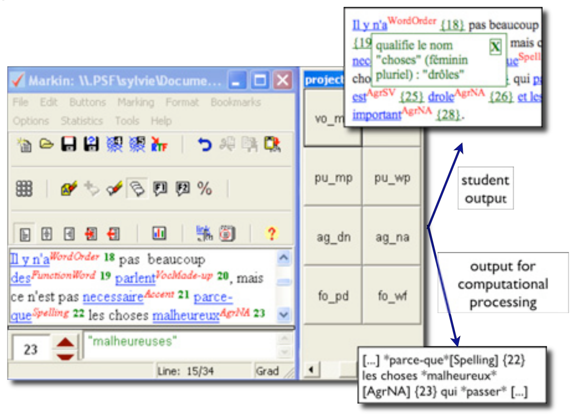

Fig. 3. Error tag insertion and output

After tagging all students' texts with the appropriate error tags, the corrected versions were exported as (a) HTML pages, as a mean to provide participants with immediate feedback, and (b) as unformatted texts in order to computationally process the information on error types.

3.2 *Part-of-speech encoding*

TreeTagger was used as a tool to assign the most probable part-of-speech tag to each token in the input text.

Although TreeTagger achieved a tagging accuracy of 96.34% (Schmidt, 1994), a major problem with pos-tagging is that it depends on the input text correctness (Dickinson, 2006). For this reason, TreeTagger's accuracy was evaluated when processing language learners' ill-formed written productions. To do so, the system output and the hand-annotated output of the same corpus were compared by means of the Kappa's coefficient (Cohen, 1960). This measure is accepted as being a more reliable measure than simple percentage as it takes into account the percentage of agreement that could have occurred by chance (Jurafski & Martin, 2000).

Kappa's coefficient was increased with (a) an extended set of commonsense rules based on recurrent tagging errors, and (b) a cross-reference between the pos-tagged and error-tagged data, see Fig. 4.

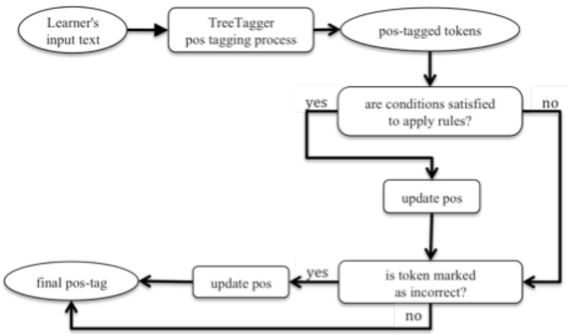

Fig. 4. Process to improve TreeTagger's accuracy

The proportion of agreement between human and machine rose from 91.5% to 96.6% after improvements, i.e. after applying the set of rules and cross-referencing the pos-tagged and error-tagged corpus.

3.3 *Self-editing encoding*

All alternatives proposed by the students during first and second self-editing exercises were manually reviewed, marked either as acceptable or not, and stored in a database.

4. Preliminary Results

To represent the learner's performance, the ratio of incorrect to probable correct observations in the learners' essay-type question is computed by counting error types and part-of-speech tags.

For example, one student named "Jane" wrote 5 instances of noun adjective agreement incorrectly, whereas 11 other adjectives, not marked as incorrect, occur in the same text. This student had, therefore, the possibility of writing 11 other occurrences of noun adjective agreement correctly. Consequently, the ratio of incorrect to correct forms is equivalent to 5:11, which means that the percentage of success in writing noun adjective agreements equals 68.75%. Jane's score seems to indicate that she performed rather well in terms of noun adjective agreements. A representation of Jane's performance is displayed in Fig. 5.

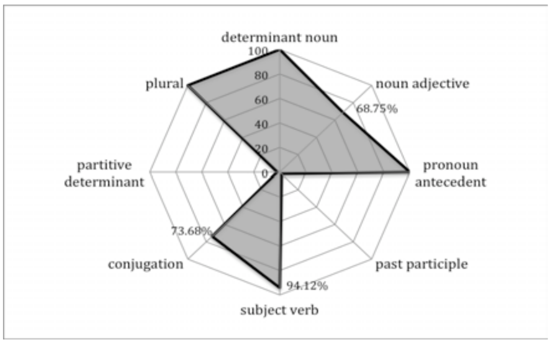

Fig. 5. Learner's performance

To determine whether the incorrect forms are to be deemed as errors or mistakes, one may interpret these results by investigating the amount of assistance students require to correct their ill-formed words.

For example, Jane was unable, neither during the first nor the second self-editing exercises, to correct the noun adjective agreement errors she wrote except one. Her self-editing corrections are listed in Fig. 6.

A cross beside the word indicates that the replacement proposed by the student was an unfruitful attempt, and the unique tick in this figure represents one correct alternative to the only agreement error she was able to notice and correct.

	noun adjective agreements	first self-editing		second self-editing	
1	l'eau *chaud	*blank*	✗	chaude	✓
2	les choses *malheureux	*malheureaux	✗	malheur	✗
3	les choses *drole	drôle	✗	drôle	✗
4	les choses *important	important	✗	important	✗
5	le film n'est pas trop *longue	*blank*	✗	*blank*	✗

Fig. 6. First and second self-editing exercises

As a result, Jane's self corrections reveal that she probably requires more specific feedback from the teacher to be able to notice the error type and correct herself. Consequently, the incorrect forms she wrote are more likely to be considered as gaps in her knowledge rather than occasional lapses or slips of the pen.

Therefore, two students, "Louis" and "Marie", with almost identical performance representations may be interpreted differently in terms of competence, since incorrect forms for one student may be considered as errors, whereas for the other as mistakes.

Both students Louis and Marie achieved a relatively good score of over 90% success when conjugating verbs. However, both students were unable to correct themselves without assistance, which suggests a gap in their competence. Louis was able to correct his incorrect forms with assistance, Marie, on the other hand, was unable to correct herself even with assistance. Even with an honorable score of more than 90% success in one specific form-focused feature, the 10% of incorrect forms cannot be interpreted as mistakes.

Marie definitely requires overt feedback from the teacher to be able to correct her incorrect forms, which are more likely to be interpreted as a more serious gap in her knowledge than Louis', since he was capable of correcting himself with little assistance.

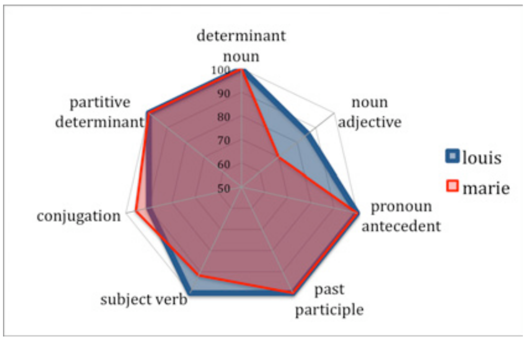

Fig. 7. Knowledge performance of two students

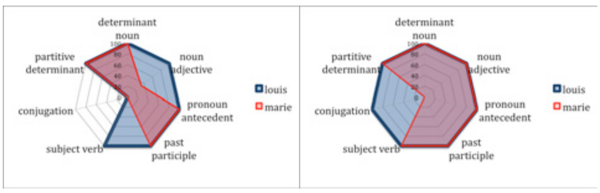

Fig. 8. First and second self-editing exercises

5. Conclusion and Further Research

The ratio of incorrect to correct forms was compared to the amount of assistance learners required when self-editing themselves. This helped determine whether this ratio was a valid indicator in differentiating error from mistake. We have demonstrated that errors from mistakes could not be discriminated with neither excellent nor poor scores. A low score does not necessarily mean that the incorrect forms are to be deemed as errors as opposed to mistakes when the score is high. As stated by Chapelle & Douglas (2006) when critically reviewing the use of computer-assisted language testing, a low score does not demonstrate "with all certainty that the examinee's […] level is low" (p. 97).

However, as an attempt to model the learner's knowledge in terms of competence, that is differentiating errors from mistakes, the amount of assistance learners require when correcting themselves is valuable information. If learners still rely on teachers to correct themselves, then the incorrect form is probably an error.

One limitation of this analysis is that it has been evaluated with only two different types of assistance. Specific focus in further investigations will be placed on level and types of feedback and how a learner model will integrate students' level of development.

References

Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal*, 78(4), 465-483.

Beck, J. E., & Chang, K. (2007). Identifiability: a fundamental problem of student modeling. In C. Conati, K. McCoy, & G. Paliouras (Eds.), *User Modeling 2007: 11th international conference, UM 2007, Corfu, Greece, Proceedings* (pp. 137-146). Berlin: Springer Verlag.

Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge: Cambridge University Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.

Dickinson, M. (2006). *An investigation into improving part-of-speech tagging.* In Proceedings of the Third Midwest Computational Linguistics Colloquium (MCLC-06). Urbana-Champaign, IL.

Ellis, R. (1997). *Second language acquisition*. Oxford: Oxford University Press.

Heift, T., & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning; parsers and pedagogues.* New York: Routledge.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: an introduction to natural language processing, computational linguistics and speech recognition*. New Jersey: Prentice Hall.

Michaud, L. N., & McCoy, K. F. (2003). Evaluating a model to disambiguate natural language parses on the basis of user language proficiency. In P. Brusilovsky, A. Corbett, & De Rosis, F. (Eds.), *User modeling 2003: 9th international conference, UM 2003 Johnstown, PA, USA, June 2003, Proceedings* (pp. 96-105). Berlin: Springer-Verlag.

Schmidt, H. (1994). *Probabilistic part-of-speech tagging using decision trees*. Paper presented at the International Conference on New Methods in Language Processing, Manchester, UK.