

Identifying negative language transfer in writing to increase English as a Second Language learners' metalinguistic awareness

Leticia Farias Wanderley

EdTeKLA Research Group, Department of Computing Science, University of Alberta
fariaswa@ualberta.ca

Carrie Demmans Epp

EdTeKLA Research Group, Department of Computing Science, University of Alberta
demmanse@ualberta.ca

ABSTRACT: This workshop paper describes the design and experimental application of an automated error detection tool that informs English as a Second Language (ESL) learners about possible negative language transfer effects in their writing. As these effects are known to hinder learners' proficiency in a second language, we hope to increase learners' metalinguistic awareness by explaining the latent causes of their misconceptions. The study utilizes an error annotated dataset of ESL learners' essays to validate common negative language transfer cases in different native languages and inform the error detection tool's feedback.

Keywords: writing analytics, language transfer, second language acquisition

1 INTRODUCTION AND BACKGROUND

English has become the *lingua franca* that is used in interactions between native speakers of distinct languages (Cenoz & Jessner, 2000). Most online content, business transactions, and international communications employ English. Given its spread and applicability, it is no wonder that people are interested in learning how to communicate in English. However, English learners struggle to acquire the language, especially when it differs from their mother tongues. In this work, we seek to provide feedback on learners' writing, elucidating errors that could be caused by discrepancies between English and their native languages. By doing so, we hope to raise learner awareness of the languages' grammatical distinctions and improve their writing skills.

When learning a novel language, learners often rely on their native languages' grammatical structures to form utterances in that new language. This phenomenon is known as language transfer, and it can be observed in all linguistic levels (Selinker, 1969). In the process of second language acquisition, language transfer is one of the strategies used by learners, intentionally and unintentionally, to communicate in that second language. This phenomenon usually occurs when the learners are unsure about the correct way to express themselves and can lead to them making grammatical errors due to a mismatch between rules in the two languages. The type and frequency of language transfer observed in learners' utterances vary according to their first language (L1) and proficiency in the second language (L2). The more proficient the learners, the more aware of L2 rules and their application they are. A less

proficient learner will rely on transfer more often. The nature of transfer observed, negative or positive, will depend on the amount of overlap between L1 and L2 rules. When grammatical rules between the L1 and L2 differ, the transferred language structure may result in an invalid utterance, according to the L2's rules. This negative language transfer effect can be defined as a lack of metalinguistic awareness in the L2 that forces learners to fall back on their L1s.

According to Gombert (1992), metalinguistic awareness is the ability to analyse language as an object, reflecting on its form and rules. Learners' metalinguistic awareness increases as they employ the language and receive feedback about their utterances. In this work, we set out to explore how error feedback, informed by the learners' L1s, can improve their L2 writing. We hope to increase the learners' L2 metalinguistic awareness by explaining the potential negative language transfer effects that lead to incorrect utterances.

2 LANGUAGE TRANSFER: AUTOMATING ITS DETECTION

The initial phase of this work consists of using automated methods to find negative language transfer evidence in English as a Second Language (ESL) learner writing. This step is essential to identify which aspects of each native language are transferred to English. Once negative language transfer evidence is established, we need to determine how to include it in the learner feedback and then measure whether it affects the learners' metalinguistic awareness and writing skills.

To automate negative language transfer detection, the writing assistant tool will leverage existing error correction systems, such as the LanguageTool¹ API, to identify and suggest corrections for errors in the learners' utterances. Our tool will then analyse the incorrect and corrected utterances' structures by identifying the grammatical categories, or parts-of-speech, of the words that form the utterances, in a process called part-of-speech tagging. Once the utterances' grammatical structures are found they can be compared against structures commonly used in the learner's L1.

The same part-of-speech tagging process will be applied to written corpora in the learner's native language to model a distribution of structures in that language. By comparing the frequency of the incorrect and corrected part-of-speech structures to part-of-speech sequences of the same length in the learner's native language, it is possible to determine whether the learners' incorrect utterances employ a commonly used sequence of parts-of-speech from their native languages and whether the error correction system's suggestion is valid in the learner's L1. If the learner's utterance structure, which is incorrect in English, is found to be common and valid in the learner's native language or the system corrected structure is not valid in the learner's L1, the system will flag a negative transfer error and provide feedback that references the relevant distinction between English and the L1. If there is not a clear correlation between the incorrect and corrected part-of-speech sequences in English and the distribution of sequences in the L1 part-of-speech tagged corpora (e.g., the incorrect utterance's structure used is not common in the learner's native language) the system will still provide error

¹ <https://languagetool.org/>

feedback. However, this feedback will not reference the learner's L1 nor will it reference language transfer effects.

To validate and test this methodology we will use a learner essay dataset from Cambridge's First Certificate in English exam (FCE). This error annotated dataset contains essays from 1,244 learners who, in total, have 16 different L1s (Yannakoudakis et al., 2011). The learners' errors are annotated and corrected following the error-tagging system described by Nicholls (2003). Each essay data point is not only annotated with writing errors and their respective correction but also complemented by metadata about the native language of its author and the score assigned by the annotator. The FCE dataset will be employed in the identification of common syntactic negative language transfer effects, as it contains the error annotated and corrected essays, and the learners' L1. This data will confirm the existence of language transfer and the tool's capacity to detect it. One of the advantages of using this dataset is the fact that it was manually tagged and corrected, hence there is no need to employ any extra error detection. It also contains essays that were designed to assess learners' writing skills in English at an intermediate level (Yannakoudakis et al., 2011), which enables the analysis of more complex syntactic structures. Another critical application of this dataset is in the language transfer detection model tuning, as the annotated errors can be applied to determine the distribution thresholds for what it means for a structure to be common in different languages.

Our negative transfer detection tool is under development. The current status of the development is that the FCE dataset has been processed and part-of-speech tagged. The errors have been grouped into syntactic and non-syntactic clusters. The next step of the implementation is to process corpora in different languages, initially Spanish and Chinese, to obtain the distribution of part-of-speech sequences necessary to identify language transfer effects.

3 PROPOSED STUDY: UNDERSTANDING THE SUPPORT TOOL'S EFFECT

To understand whether the negative language transfer feedback supports language learning and improves the learners' writing skills, we will recruit ESL learners into two groups. Both groups will use an online editor to write short essays in English. One group will receive negative language transfer informed feedback, while the other will receive generic feedback about the grammatical errors. Before and after the essay writing sessions, we will assess learner awareness of English syntax through cloze tests, in which they select the best option among negative language transfer and correct alternatives; error correction tasks, in which they correct ungrammatical sentences that contain negative transfer; and selection tasks, in which learners choose the grammatical version of a sentence from a set of options (Chireac et al., 2019).

In the proposed study the negative transfer feedback will be delivered by a web browser extension that is under implementation by a team of software developers. This extension will use the error correction API to detect errors in the learners' utterances as they write on an online editor. It will call attention to writing errors by highlighting the incorrect utterances in the users' texts. If the users wish to understand the nature of the error, they can select the highlighted text to make the extension display a pop-up box

containing an explanation about why the utterance is incorrect. For errors caused by negative language transfer, the feedback explanation will contrast specific rules in the user's L1 and English that are associated with the identified error. The extension users will be able to select their native languages to enable language transfer feedback.

As negative language transfer is a typical phenomenon in second language acquisition, we propose an automated error detection tool that will give learners more insight about writing errors that potentially arise from differences between their native languages and English. With that, we hope to support ESL learners' language development by giving them more opportunities to understand the causes of their errors. We expect that the addition of feedback based on language transfer effects will enhance learners' English metalinguistic awareness and, hence, their writing skills.

REFERENCES

- Cenoz, J., & Jessner, U. (Eds.). (2000). *English in Europe: The acquisition of a third language* (Vol. 19). Multilingual Matters.
- Chireac, S. M., Francis, N., & McClure, J. (2019). Awareness of Form and Pattern in Literacy Assessment: Classroom Applications for First and Second Language. *The Reading Matrix: An International Online Journal*, 19(1).
- Gombert, J. E., & Gombert, J. É. (1992). *Metalinguistic development*. University of Chicago Press.
- Nicholls, D. (2003, March). The Cambridge Learner Corpus: Error coding and analysis for lexicography and ELT. In *Proceedings of the Corpus Linguistics 2003 conference* (Vol. 16, pp. 572-581).
- Selinker, L. (1969). Language transfer. *General linguistics*, 9(2), 67.
- Yannakoudakis, H., Briscoe, T., & Medlock, B. (2011, June). A new dataset and method for automatically grading ESOL texts. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* (pp. 180-189). Association for Computational Linguistics. <https://dl.acm.org/doi/10.5555/2002472.2002496>