

City University of New York (CUNY)

CUNY Academic Works

Dissertations, Theses, and Capstone Projects

Graduate Center

2-2020

Analysis of PRO-drop Errors in L2 English by L1 Spanish Speakers

Marcos R. Ynoa

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

 Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#), [Computational Engineering Commons](#), and the [Educational Technology Commons](#)

This Thesis is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.

ANALYSIS OF PRO-DROP ERRORS IN L2 ENGLISH BY L1 SPANISH SPEAKERS

by

MARCOS YNOA

A master's thesis submitted to the Graduate Faculty in Linguistics in partial fulfillment of
the requirements for the degree of Master of Arts, The City University of New York

2020

© 2020

MARCOS YNOA

All Rights Reserved

Analysis of PRO-drop errors in L2 English by L1 Spanish Speakers

by

Marcos Ynoa

This manuscript has been read and accepted for the Graduate Faculty in Linguistics in
satisfaction of the thesis requirement for the degree of Master of Arts.

Date

Martin Chodorow

Thesis Advisor

Date

Juliette Blevins

Acting Executive Officer

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

ANALYSIS OF PRO-DROP ERRORS IN L2 ENGLISH BY L1 SPANISH SPEAKERS

by

MARCOS YNOA

Adviser: Martin Chodorow

Although educators and linguists have described and studied L2 learner errors, often from the perspective of positive and negative transfer (Krashen, 1981), there is relatively little empirical data on the frequencies of L2 error types. Learner error rankings do exist (Donahue, 2001) (Cambridge learner corpus; (Nicholls, 1999)), but these rankings are often too general and may overlook specifics when it comes to particular language learner groups. The goal of our work is to develop a tool that can be used to explore error patterns, address educator needs, and help answer research questions in L2 language learning. The tool we have developed combines an online corpus of corrected learner sentences (Lang-8; Mizumoto, T. et. al. 2013), an Error Annotation Toolkit (ERRANT; (Bryant, C. et. al. 2017)), and a dependency parser (SpaCy; (Honnibal, M. et. al. 2015)) to provide a more informative view of learner errors.

To test the usefulness of the tool, we asked educators of Spanish-speaking English learners from an international high school to provide a list of error types that they commonly encounter in the classroom. One of the most frequently mentioned error types was Pro-drop of the 3rd person neuter singular (It), which is the focus of this thesis. Using our tool, we investigated Pro-drop instances within learner sentences relative to the syntactic role of the pronoun and its corresponding correction. Results showed a significantly higher error rate of Pro-drop with the 3rd person neuter singular particularly in nominative position in

relation to other pronoun types. Investigating errors, such as Pro-drop produced by specific L1 groups, has implications for discourses formed around such topics and may prove to be beneficial to educators, researchers, and developers of educational technology.

ACKNOWLEDGMENTS

First, I would like to express my gratitude to my advisor Professor Martin Chodorow. The foundation of this thesis largely came to fruition due to his insight and guidance. Not only was his advice invaluable from a theoretical standpoint, but also when it came to computational and statistical methodologies. It was a great pleasure to work with him and I am grateful for the growth I have experienced under his advisement. I would also like to thank Hussein Ghaly and Kyle Gorman for helping with initial error analysis approaches that would eventually find their way into the thesis.

I would also like to extend my gratitude to the PiLaCS team (Participating in Literacies across Computer Science): Kate Menken, Chris Hoadley, Laura Ascenzi-Moreno, Sara Vogel, Jasmine Ma, Sarah Radke, Jeanette Marte, and the rest of the educators, researchers, and administrators that I have had the pleasure of working with. Being part of the research team has expanded my understanding of Education in respect to my background, the population of Bilinguals/Language Learners and Educators that I have had the pleasure of working with. Being part of the PiLaCS has taught me what it means to approach student agency as a conduit to help make Computer Science education accessible to all.

I would also like to thank the amazing group of teachers from Pan American International High School, who were generous enough to provide their insights for this analysis. Their input was crucial for identifying what were the most important needs of the student population they (and I) have been serving as educators. It is my hope that the results emerging from this thesis may be useful to at least some degree.

I would like to thank my friends, colleagues, and family for their support throughout the duration of my master's program. The multitude of support made a considerable difference, academically and emotionally. *Te quiero, mamá.*

Contents

| | |
|--|------------|
| Contents | vii |
| List of Tables | ix |
| List of Figures | x |
| 1 Introduction | 1 |
| 2 Methodology | 8 |
| 2.1 Teacher Observations of Error Types | 8 |
| 2.2 Error Analysis Toolkit | 9 |
| 2.3 The Lang-8 Corpus | 10 |
| 2.4 ERRANT(ERRor ANnotation Toolkit) | 14 |
| 2.5 spaCy | 17 |
| 2.6 Realignment of Edits | 17 |
| 2.7 Toolkit Output | 18 |
| 3 Results | 20 |
| 3.1 Learner Sentence Breakdown | 20 |
| 3.2 Results: Pro-drop by person, gender and number | 21 |
| 3.3 Results: Pro-drop by case | 23 |
| 4 Discussion | 24 |
| 4.1 Pro-drop in L2 English writing by a Spanish L1 | 24 |
| 4.2 Limitations | 25 |
| 4.2.1 Lang-8 Learner Profiles | 25 |
| 4.2.2 Quality of Corrections | 26 |

| | | |
|----------|---|-----------|
| 4.2.3 | Unannotated sentences | 27 |
| 4.2.4 | Part-of-speech (POS) and Parsing issues | 27 |
| 4.3 | Questions for future investigation | 28 |
| 4.4 | Future Work | 31 |
| 5 | Conclusion | 33 |
| | References | 35 |

List of Tables

| | | |
|---|---|----|
| 1 | Table adapted from Errors Made by U.S College Students and by ESL College Students (Donahue, 2001) | 5 |
| 2 | Table adapted from Proportion of errors in the <i>Cambridge Learner's Corpus</i> (Nicholls, 2003) | 6 |
| 3 | Teacher observations of error types, ranked by frequency | 9 |
| 4 | A partial list of the error types ERRANT can identify (Bryant et al., 2017) . | 15 |
| 5 | Frequencies and error rates by pronoun person, gender, and number | 21 |
| 6 | Frequencies and error rates by pronoun person, gender, number, and case . . | 22 |

List of Figures

| | | |
|----|---|----|
| 1 | The Iceberg Representation of Language Proficiency | 2 |
| 2 | Representations of the " <i>Dual-iceberg</i> " and <i>CUP</i> (<i>Common Underlying Proficiency</i>) | 3 |
| 3 | The data organization process | 10 |
| 4 | source: https://lang-8.com/ | 11 |
| 5 | source: https://sites.google.com/site/naistlang8corpora/ | 12 |
| 6 | Data format of entry | 13 |
| 7 | Flowchart detailing the data curating process | 14 |
| 8 | An example of an ERRANT output | 16 |
| 9 | An example of dependency parsing | 17 |
| 10 | Final output of the toolkit after realignment | 18 |
| 11 | Distribution of unique learner sentences, with and without annotations | 20 |
| 12 | An example of a non-referential 'it' insertion by the corrector | 25 |
| 13 | An example of a correction that could lead to an ungrammatical sentence | 27 |
| 14 | An example of a correction with inadequate information for the learner | 27 |
| 15 | Examples of ERRANT problems with multi-token edits | 28 |
| 16 | An example of an unnecessary or extraneous pronoun (U:PRON) | 29 |
| 17 | Another example of an unnecessary or extraneous pronoun (U:PRON) | 29 |
| 18 | An example of an unnecessary or extraneous pronoun (U:PRON) possibly resulting from negative transfer from Spanish | 29 |
| 19 | An example of a pronoun replacement (R:PRON), 'i' for 'It' | 30 |
| 20 | Another example of a pronoun replacement (R:PRON), 'it' for 'them' | 31 |

1 Introduction

This thesis aims to explore the error type of Pro-drop produced in learner English, specifically in relation to transference issues stemming from the native language (in this case, Spanish). Within the field of Second Language Acquisition (SLA), specifically Teaching English as a Second Language, a substantial amount of research has taken place to analyze error production in the language actively being learned (the second language or L2) with respect to the learner's native language (L1). It has been posited that English Language Learners (ELLs) will often use whatever resources they have available, which can also include their knowledge of the L1, in order to reinforce routines to facilitate their understanding of the L2. This is commonly known as *translanguaging* (García and Wei, 2013). The *translanguaging* theory suggests that ELLs draw upon rich linguistic and semiotic repertoires in order to solidify their understanding. An ELL's L1 is one of these resources that learners utilize according to the theory. For example, a Spanish-speaking ELL may draw upon their use of cognates (words that share common origins) to understand phonetically and/or orthographically similar words in English. In his 1981 study, Krashen categorized L1 transfers into two categories, *Positive* and *Negative* transfers, which could pertain to lexical items and/or grammatical structures transferred from one language to another. In the context of a Spanish-speaking ELL, an example of a positive transfer could be the acquisition of the English determiner *the*, since the equivalent can be found in Spanish. An example of a negative transfer would be the learner having difficulty discerning that *the* has only one form and is not gendered as *el* and *la* are for masculine and feminine, respectively, in Spanish. Ultimately, Krashen argues that there is a double-edged nature to L1 transference to the L2. On one hand, a student's native language literacy may help them transfer concepts from the native language and apply them to the language being acquired. On the other hand, Krashen also recognizes that not every student's L1 shares the same structure or patterns

with their L2, so one must anticipate a variety of transfer errors that may occur.

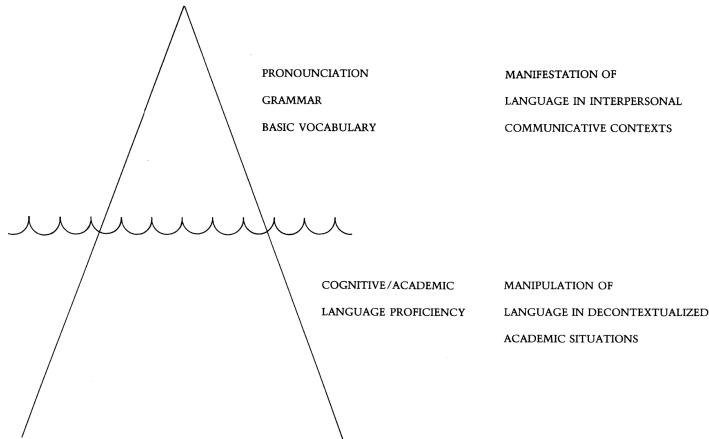


Figure 1: The Iceberg Representation of Language Proficiency
by Cummins (1981)

Jim Cummins, who has been influential in the development of SLA and TESOL (Teaching English as a Second or Other Language) methodologies, has also contributed to the topic of language transfer. Cummins is best known for introducing the distinction between Basic Interpersonal Communication Skills (BICS) and Cognitive Academic Language Proficiency (CALP) (Cummins, 1981). BICS is defined as a speaker's proficiency in an everyday setting, that is, cognitively undemanding language use within the context of purposeful interpersonal interactions. CALP, on the other hand, is defined as more of a manipulation of linguistic resources. It is cognitively demanding by nature, removed from meaningful interpersonal interaction and connected to literacy skills which predominantly concern, but potentially are not limited to, academic settings. Essentially, Cummins argues that everyone is capable of obtaining BICS regardless of a person's intellectual or academic prowess, although when it comes to CALP, there may be a variety of factors that affect to what degree the proficiency develops. Cummins expresses his argument via the Iceberg Theory. The iceberg itself is a representation of a speaker's language proficiency. According to the theory, on the surface of the iceberg rests the speaker's BICS, which is apparent in all speakers of a given language.

However, a speaker's CALP, the language required outside of the interpersonal context, is hidden below the surface of the iceberg. (Cummins, 1981).

Cummins introduced the concept of *Linguistic Interdependence* between the L1 and L2 with a Dual-iceberg representation. The idea behind *Linguistic Interdependence* is that proficiency in a learner's L1 promotes various opportunities for development in the learner's L2. Cumming states:

[...] instruction by means of a minority language in the early grades is not just promoting proficiency in the surface manifestations of that language; it is also promoting deeper cognitive and academic skills that underlie development of literacy in *both* the bilingual's languages. (Cummins, 1981)

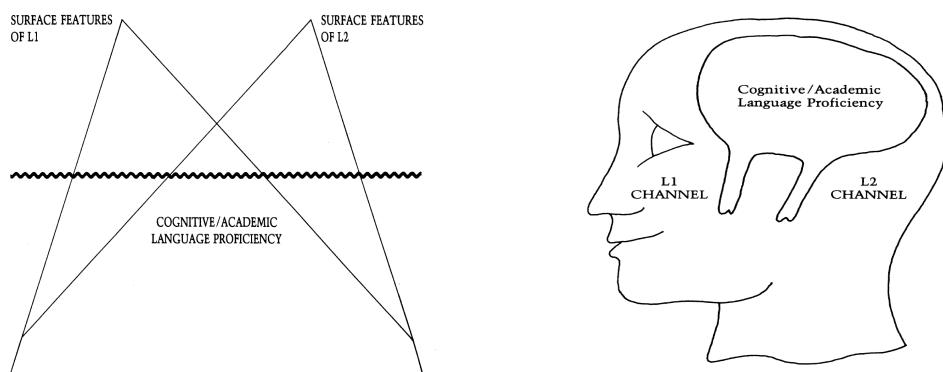


Figure 2: Representations of the "Dual-iceberg" and CUP (*Common Underlying Proficiency*) by Cummins (1981)

The Common Underlying Proficiency (CUP) theory was introduced to further elaborate on the idea that a learner's experience with any language can promote the CALP of both languages within the right context where proper exposure is given to the languages in question (Cummins, 1981). The implications of these theories are that there is a two-sided nature of transference from a learner's L1 to their L2. Krashen's analysis highlights the possibility that both positive and negative transfer can surface, which in turn could impact the student's

learning of the L2. Cummins' theory, however, views this as opportunistic, as a learner's L2 CALP has the opportunity to expand (and vice versa), provided that the proper exposure is granted for the languages to develop.

In 2011, Lee carried out a study that analyzed the feedback practices of 26 English teachers in Hong Kong based on the feedback they provided on student essays and based on interviews with a few of the teachers. The goal was to analyze underlying problems that can affect teachers' feedback practices. This connects to Lee's main argument and motivation for the study, which is that in order to improve teachers' feedback strategies, it is essential to tackle common assumptions and confront possible shortcomings with current practices. Lee was quick to identify a generalized attitude towards correcting student writing, which tends to require a copious amount of time, often beyond normal teaching hours. Results of the analysis of approximately 5,353 feedback instances from the study showed a preference for teachers correcting errors (94.1%) in contrast to feedback on other areas such as organization (0.4%) or stylistic choices (3.8%). Lee argues that such a concentration on error correction feedback can have detrimental effects on teachers in general, leading to stress, change in attitude towards the profession and obligations, and consideration of career changes.

Although the number of teachers involved in Lee's study is not representative of educators as a whole (in both location and in number), based on our experience, the teachers I have worked with and I agree with Lee's arguments. Educators, particularly within the public school system, are faced with an increasing workload, both in and out of the classroom and not solely just providing students with meaningful feedback. A challenge arises when trying to address the various types of learner errors; if feedback focuses on predominantly one dimension of writing that leaves little to no time for other crucial areas in which students would need equal, if not more, attention from teachers. Approaches that help educators with error correction could allow them to know what the most prominent needs of their students are and could free up time for other tasks, such as providing feedback in other areas of

student writing or being able to manage other responsibilities throughout any given day.

A possible solution to the problem of limited time and resources is to focus effort on the most frequent errors, based on ranked lists of types of errors produced by language learners. Such lists do exist as the result of previous efforts to categorize common error types. Two examples are shown in Table 1, an analysis of errors made by U.S. college students in general compared to ESL college students (Donahue, 2001), and Table 2, an analysis of learner errors from the *Cambridge Learner's Corpus* (Nicholls, 2003).

| Error Type | Rank for U.S. College Students | Rank for ESL College Students |
|-------------------------------------|-----------------------------------|----------------------------------|
| No Comma after introductory element | 1 | negligible |
| Vague pronoun reference | 2 | negligible |
| ... | ... | ... |
| Wrong word | 4 | 2 |
| ... | ... | ... |
| Missing words | negligible | 3 |
| ... | ... | ... |
| Wrong pronoun | negligible | 16 |

Table 1: Table adapted from Errors Made by U.S College Students and by ESL College Students (Donahue, 2001)

Although these lists provide some insight as to the types of errors ELLs are likely to produce in the classroom setting, their categorizations of error types are vague and do not provide information about errors of various L1 subgroups. To the best of our knowledge, there are no lists detailing the relative frequencies of error types by L1. For example, in the list of errors from Donahue's study, we can see categories such as "Wrong Word", "Wrong Pronoun", or "Missing Words". The vagueness of the categories makes it difficult to precisely identify what could be giving language learners difficulty with the items in question. Furthermore, we are not informed of the difference in size between adjacent ranks. Essen-

| Rank | Error Type | Prop | Example Sentence |
|------|---------------------------|-------|--|
| 1 | Content word choice error | 0.199 | We need to deliver the merchandise on a daily *base / basis. |
| 2 | Preposition Error | 0.134 | Our society is developing *in/at a high speed. |
| ... | ... | ... | ... |
| 9 | Agreement Error | 0.040 | I *were / was in my house. |
| ... | ... | ... | ... |
| 18 | Quantifier Error | 0.007 | It doesn't give them too *much/many problems. |

Table 2: Table adapted from Proportion of errors in the *Cambridge Learner's Corpus* (Nicholls, 2003)

tially ranks do not provide information about relative frequencies, information that could be extremely useful in an educator’s hands and could help inform best practices based on an error proportion analysis, and also possible negative transfer that stems from the L1. This would make it possible to allocate time more optimally towards classroom instruction.

A ranked list of error types that include relative frequencies based on the learner’s L1 could have a multitude of benefits. I briefly mentioned how this type of information could be useful for educators not only to inform best practices moving forward, but also to potentially help educators strategically commit to an otherwise time consuming process. Additionally, such an analysis could be of benefit to the research community in terms of answering questions related to SLA, linguistics, and the educational sciences. Finally, for software developers, it could help shape approaches to the development of educational applications geared towards students and/or teachers in order to help create both robust and inclusive platforms within educational technology. A toolkit for exploring error patterns and reporting frequencies of error types for different language learner populations could be useful in these cases, which is what initiated this study.

When it comes to English and Spanish, there are noticeable differences in regard to pronouns in subject and object positions. With this in mind, instances of pronoun dropping are prohibited to a greater degree in English (e.g., the English sentence *Washed the dishes* constitutes a syntactic violation whereas the equivalent in Spanish is acceptable). Taking these differences between English and Spanish into account, we can imagine a Spanish-speaking learner of English being prone to dropping English pronouns – a case of negative transfer. We can also imagine this providing an example of how the toolkit and corresponding data can be used to investigate transfer patterns in learner English. For this study, the primary focus of the toolkit will be on Pro-drop patterns in writing by ELLs with a Spanish L1.

We focus on the following research questions:

1. Do gender and number of the pronoun have an effect on drop rate?
2. Do the rates of Pro-drop depend on pronoun case? Specifically, are Subject (nsubj) pronouns more likely to be dropped than Direct Object (dobj) pronouns?

2 Methodology

2.1 Teacher Observations of Error Types

Before testing the usefulness of a toolkit for exploring error patterns, we asked eight educators of ELLs to provide up to ten error types that their students have produced in their classrooms. The eight teachers had an average of 7.6 years of teaching experience, and each had taught for at least one year. At the time of the survey, they taught at a public high school that is part of the International Network of schools. These schools are designed to assist students who have recently arrived (within the last four years) in the United States. The students receive additional language support, such as standalone ESL classes, while at the same time satisfying core high school courses which are designed with strategies to assist language learners studying specialized content areas. Although most high schools within the International Network have a diverse population of language learners, the teachers' school serves an almost entirely Latinx population; this is a reflection of the neighborhood in which the school is situated. Geographically, the school is located between two neighborhoods that are home to diverse populations. One of the neighborhoods in particular is home to a large Latinx community which is reflected in the student demographics. The teachers' observations were collected via an anonymous online questionnaire through Qualtrics. Once all the observations were collected, the responses were merged and ranked by the number of teachers that identified a particular error type (see Table 3).

Though we have a large number of singletons among the error types, we can see that some errors were more commonly listed by the teachers. Run-on sentences ranked highest, followed by subject-verb agreement, Pro-drop (the 3rd person singular neuter pronoun *it* being specifically identified twice), question formation, pronoun case, omission or misuse of *do*, and conjugation errors. Although a thorough analysis of all error types present within this list is beyond the scope of this project, we chose to investigate Pro-drop with our toolkit

| Nature of problem | Count | Category |
|--|----------|--------------------------|
| Run-on sentences | 4 | Grammar/ Punctuation |
| Subject-verb agreement | 3 | Grammatical |
| Pro-drop (it-2) | 3 | Grammatical |
| Question formation | 2 | Grammatical |
| Omission of "do"/ Misuse in variation | 2 | Grammatical |
| Conjugation | 2 | Grammatical |
| Lexical Errors (e.g: story/history, make/do) | 1 | Grammatical/ Phonetic |
| Word choice (direct translation) | 1 | Grammatical |
| ... | ... | ... |

Table 3: Teacher observations of error types, ranked by frequency

as a test of its usefulness.

2.2 Error Analysis Toolkit

The toolkit that we assembled to help us explore the relative frequencies of error types and answer our research questions is a product of a combination of resources. The first is a learner corpus from Lang-8, a language exchange website, extracted and put together by Mizumoto et al. (2011). Next is the ERRor ANnotation Toolkit (ERRANT; (Felice et al., 2016)); (Bryant et al., 2017)). Finally, we made use of the dependency parser and Part-of-Speech (POS) tagger in spaCy (Honnibal and Johnson, 2015) to provide linguistic features that are relevant to the Pro-drop phenomenon.

Each of these resources is explained in more detail in the following sections, along with the methods in place to curate the data that was essential for our analysis.

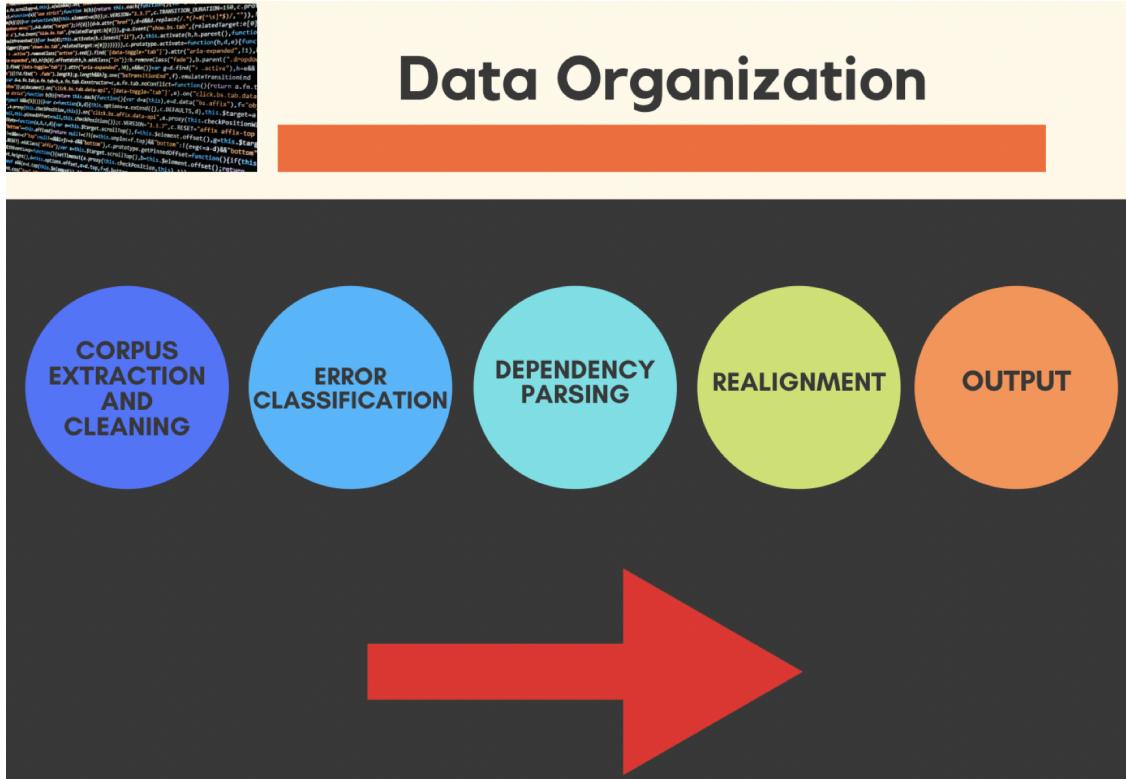


Figure 3: The data organization process

2.3 The Lang-8 Corpus

Lang-8 is a free-to-use language-exchange website. In any given target language, native language speakers can volunteer to help check and provide corrections of a learner’s written journal entries. Conversely, if a learner is a native speaker of a target language that another user is trying to learn, they can also volunteer to check that user’s journal entries and offer feedback/corrections. This is illustrated in Figure 4.

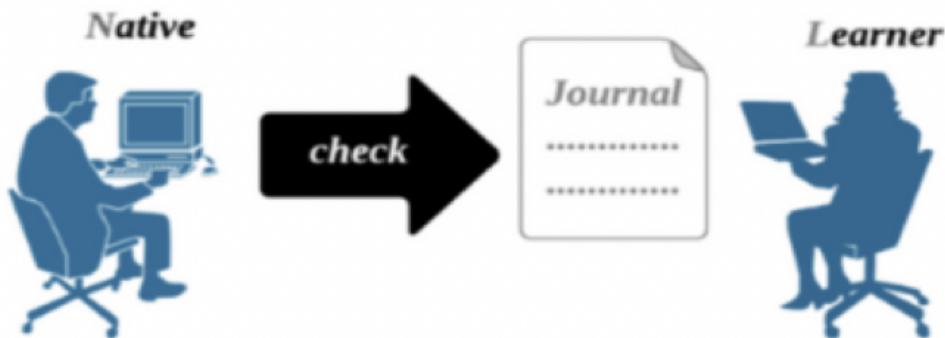
Mizumoto et al. (2011) scraped the website for an earlier study and made the Lang-8 learner corpus publicly available. The corpus contains all of the languages supported by Lang-8. (See Figure 5 for the top 20 learner languages by number of entries in the corpus.)

The entries are stored in JSON format (see Figure 6) and hold informative elements such as: the learning language, the native language, and learner sentences with their corresponding

What is Lang-8 ?

Lang-8 is a free language-exchange social network.

Just write a journal in any language you're learning, and it will appear to native speakers of that language. They'll correct it and leave useful comments and feedback.



In return, you check other users' journals written in your native language.



Figure 4: source: <https://lang-8.com/>

corrections, which we are most interested in. If no correction/feedback was supplied for a learner sentence, then its correction/feedback is represented as an empty element. From this information, we were able to narrow down and extract journal entries of users who were learning English and had Spanish as their native language. In total, there are 2,354 such journal entries containing a total of 29,648 learner-correction sentence pairs (where some of the corrections could be empty).

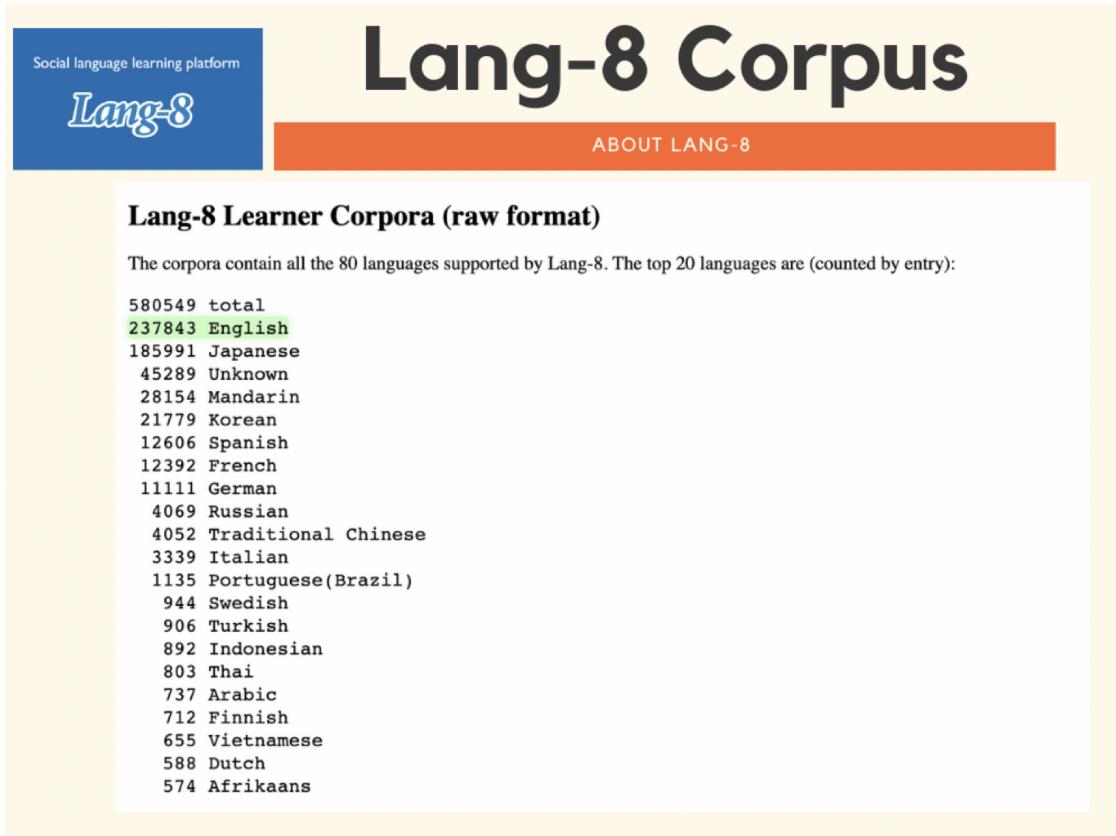


Figure 5: source: <https://sites.google.com/site/naistlang8corpora/>

Although the Lang-8 data overall are ideal for our analysis, the meta tags that were scraped with the corrections proved to be an initial challenge. These tags most likely were used as tools by the users on the Lang-8 website in order to color-code, bold, or strike out certain items for emphasis. Originally we cleaned the learner-corrector sentences using a variety of regular expressions. Unfortunately, many of the meta-tags scraped were incomplete. Furthermore, it seems that the meta-tags that accompanied the corrections were used arbitrarily by users, which made it difficult to map meta-tags to specific corrections, which in turn made it difficult to handle with regex expressions. Basically, the meta-tags are not as useful in this analysis and contribute to a fair share of noise. Fortunately in 2017, Tomoya Mizumoto, one of the authors responsible for originally extracting data from Lang-8, released an extractor that returns clean learner-corrector sentence pairs (<https://github.com/tomo>

The screenshot shows the Lang-8 Corpus extraction and cleaning interface. At the top left is the Lang-8 logo with the text "Social language learning platform". To the right is the title "Lang-8 Corpus" and below it, a red bar with the text "EXTRACTION + CLEANING".

DATA FORMAT
=====

The data is in json format. The structure is

```
[{"journal_id": "772869", "sentence_id": "227504", "learning_language": "English", "native_language": "Spanish", "learner_sentence1": "My prefer color", "learner_sentence2": "Hello people.", "correction1_to_sentence1": "Today I didn't know to tell us.", "correction2_to_sentence1": "My prefer color is red.", "correction1_to_sentence2": "Because is funny and different.", "correction2_to_sentence2": "The red can to pretend dangeruis and it's sexy.", "correction1_to_sentence1": "The color red is chosen for people with self-confidence."}, {"learner_sentence1": "Today I didn't know how to say it this:", "learner_sentence2": "My favourite color is red.", "correction1_to_sentence1": "Because it is funny and different.", "correction2_to_sentence1": "Red can pretend to be dangerous and it's sexy.", "correction1_to_sentence2": "The color red is chosen by people with self confidence."}]
```

Example:

```
[{"journal_id": "772869", "sentence_id": "227504", "learning_language": "English", "native_language": "Spanish", "learner_sentence1": "My prefer color", "learner_sentence2": "Hello people.", "correction1_to_sentence1": "Today I didn't know to tell us.", "correction2_to_sentence1": "My prefer color is red.", "correction1_to_sentence2": "Because is funny and different.", "correction2_to_sentence2": "The red can to pretend dangeruis and it's sexy.", "correction1_to_sentence1": "The color red is chosen for people with self-confidence."}, {"learner_sentence1": "Today I didn't know how to say it this:", "learner_sentence2": "My favourite color is red.", "correction1_to_sentence1": "Because it is funny and different.", "correction2_to_sentence1": "Red can pretend to be dangerous and it's sexy.", "correction1_to_sentence2": "The color red is chosen by people with self confidence."}]
```

Please note that corrections may contain tags like [f-red]...[/f-red], [f-blue]...[/f-blue], and [sline]...[/sline] (meaning deletion).

Statistics:

- "EN", "SP"
- 2354 elements,
- 29648 Sentence pairs

Figure 6: Data format of entry

wb/Lang8-NAIST-extractor). If the learner sentence had no corresponding correction(s), then the learner sentence is duplicated in the correction field (which would then be marked for no changes via the ERRANT toolkit, described below). We chose to use Mizumoto’s extractor for the purposes of this analysis.

In order to produce proper edits, ERRANT requires a 1-to-1 mapping of each learner sentence to its corrections. There was a considerable number of sentence pairs in which two or more alternative corrections were provided for a given error. Additionally, there was a lot of meta-linguistic conversation between the corrector and the learner that appeared outside of the original correction, which was usually noted by parenthesis, dashes, or some other special symbol; these discussions are both rich and interesting, but for the purposes

of this analysis, we had to segment any multi-corrections to match the single learner sentence. To do this, we created a separate file which contains the original learner sentence, the original correction(s) and/or discussions attached to the correction, and manually chose the corrector's first alternative correction or the edit that best matched the intentions of the learner.

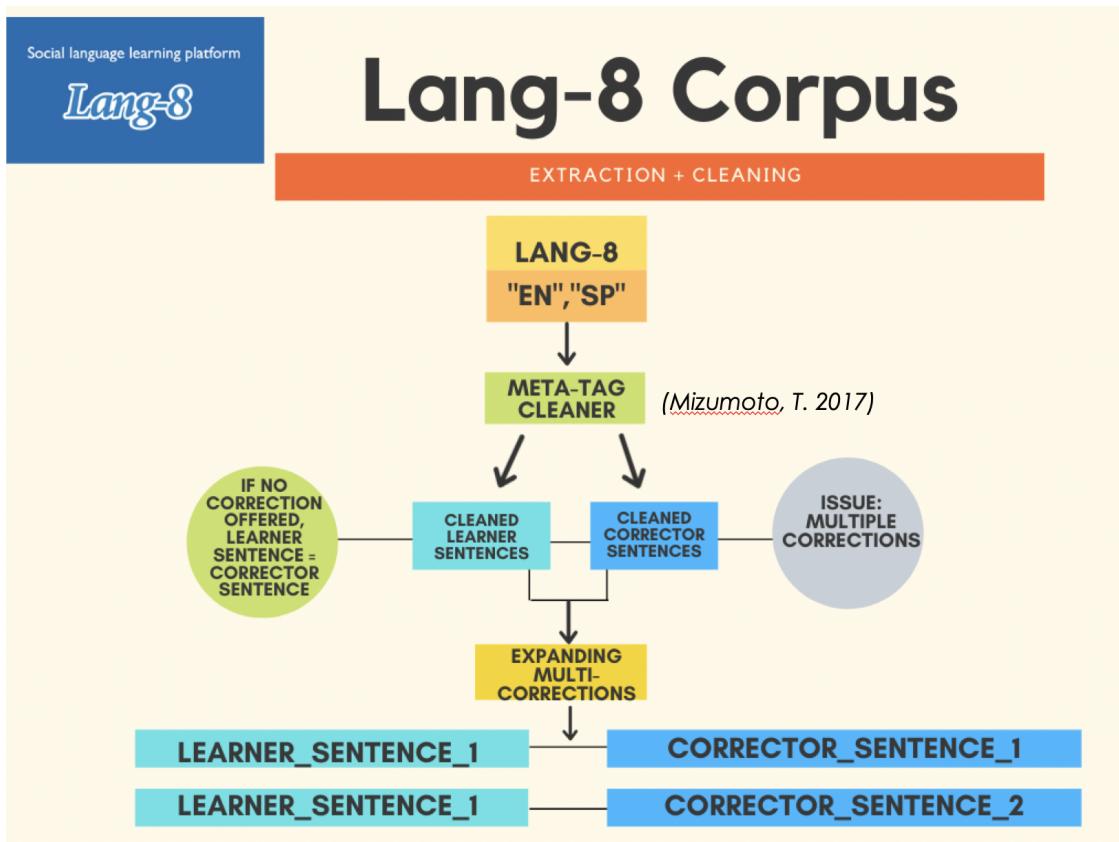


Figure 7: Flowchart detailing the data curating process

2.4 ERRANT(ERRob ANnotation Toolkit)

ERRANT is a toolkit developed by Bryant et al. (2017). It is able to annotate parallel English sentences (original and corrected sentence pairs) with error type information. ERRANT uses an algorithm based on Damerau-Levenshtein distance, a variant of the original Levenshtein

| Code | Meaning | Example |
|-------------|----------------|-------------------------------------|
| ... | ... | ... |
| DET | Determiner | the ->a |
| NOUN | Noun | person ->people |
| PREP | Preposition | of ->at |
| PRON | Pronoun | ours ->ourselves |
| VERB | Verb | ambulate ->walk |
| VERB:FORM | Verb Form | to eat ->eating dancing ->danced |
| ... | ... | ... |

Table 4: A partial list of the error types ERRANT can identify (Bryant et al., 2017)

edit distance formula. The formula was created by Vladimir Levenshtein in 1967 to calculate the least costly of edits necessary to match one string to another.

According to Bryant et al. (2017), a significant shortcoming of the original Levenshtein distance is that the derived algorithm accounts for token transpositions (edits involving token order) as the sum of an insertion and deletion, which are more costly. The Damerau-Levenshtein distance was introduced in order to be able to factor in token transpositions at a reduced cost. (Felice et al., 2016).

ERRANT uses Part-of-Speech (POS) tags to classify edits that can fall under any of the following three categories: *Missing* (M, which implies an insertion took place), *Replacement* (R, which implies a substitution took place), and *Unnecessary* (U, implying the deletion of an item).

ERRANT uses a rule-based merging framework in order to process its annotations effectively. The rules are applied in a top-down manner recursively until either all token sequences are exhausted or no more rules can be applied. It is important to note that although the merging rules are applied in order, some rules can be overwritten by rules which have a higher priority on the list. For a detailed explanation on ERRANT, complete with a list of rules and an example of the framework being applied, please refer to Bryant et al.’s paper.

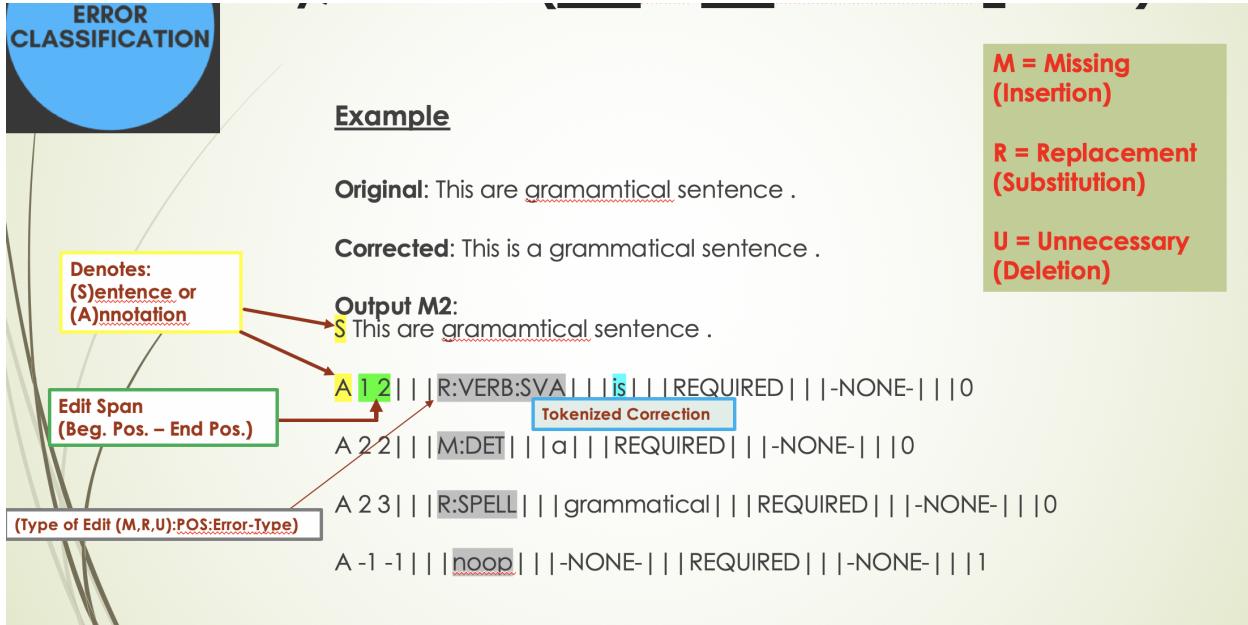


Figure 8: An example of an ERRANT output

As shown in Figure 8, the output gives us various error type information. For each learner-corrector error annotation, the user is presented with the original uncorrected sentence, the corrected sentence, and an output segment, which shows each edit undertaken in order to match the learner’s string to the corrector’s based on the original sentence. S denotes the original sentence (the correction isn’t part of the output, only the edits), and A denotes subsequent annotations. The number pair immediately to the right of the A is the edit span, or rather, the range within the sequence of tokens in which the edit has occurred. Following that, we see a cluster of information starting with the type of edit (M,R, or U) and the error code of the original token undergoing the edit (e.g: R:VERB:SVA means a replacement edit has occurred to a verb, and that the error falls under the category of SVA (subject-verb agreement)). A partial list of other error types is shown in Table 4. The next field shows the token that is affected by the edit. If M or R is the type of edit, then the token displayed is what is being added. In the case of U, the field will be empty since nothing is added.

2.5 spaCy

spaCy is a free, open-source library for Python that can be used for various Natural Language Processing (NLP) related tasks (Honnibal and Johnson, 2015). For the purposes of this analysis in particular, we are interesting in spaCy’s dependency parsing capabilities as it can give us additional linguistic information about the type of pronouns that are being dropped and their syntactic roles. Dependency parsing is the process of defining a given sentence with a syntactic structure based on head-dependent relations, as shown in Figure 9.

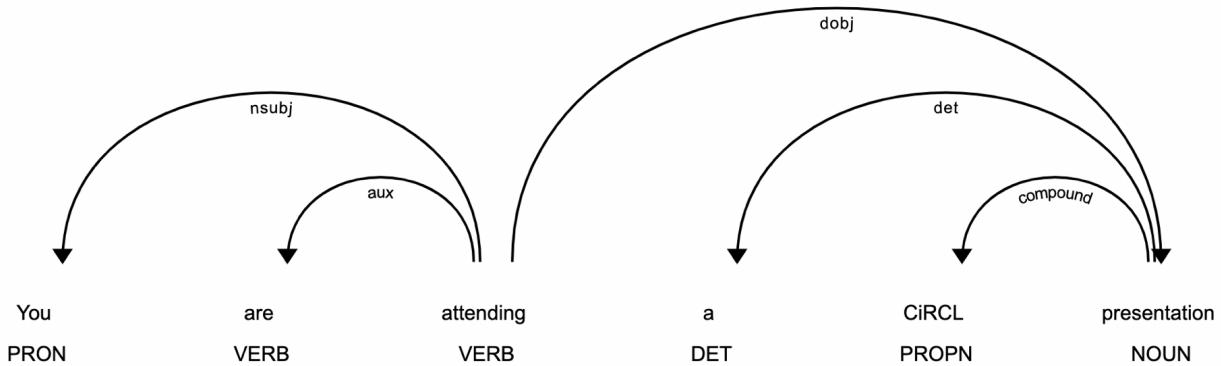


Figure 9: An example of dependency parsing

2.6 Realignment of Edits

Although the error type information produced by ERRANT is useful, the position of the edit spans denoted by ERRANT are relative to the original learner sentence and not the corrected sentence. That means that if we want to locate the position of a token in the corrected sentence (such as a pronoun that was inserted), the position of that token could vary depending on other changes made in the corrected sentence. Considerable efforts were taken to realign edit spans from ERRANT to their corresponding positions in the corrected sentence, a process which is crucial for analyzing cases of Pro-drop as well as for other kinds

of errors.

2.7 Toolkit Output

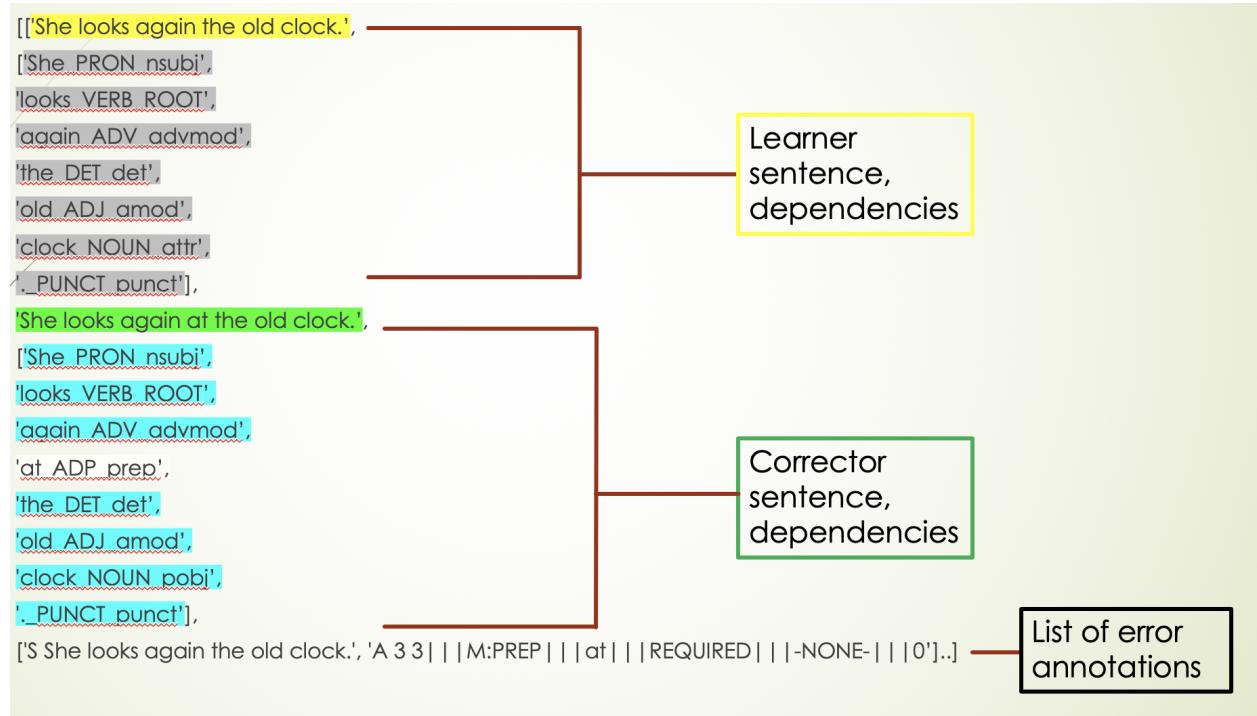


Figure 10: Final output of the toolkit after realignment

Using Figure 10 as a reference, the toolkit we have assembled curates the data in JSON format¹, with a collective list holding each learner-corrector pair as one unique item. Within any given list item we have the following: the learner sentence along with its POS and dependency information, and the corrector sentence with its POS and dependency information. Finally, we have an embedded list of all the error annotations produced by ERRANT for the given learner-corrector sentence pair. The rationale behind this is that we can look into Pro-drop cases by focusing on the error type of M:PRON, which denotes a missing pronoun.

¹The data object is publicly available and can be accessed here: https://github.com/menoah/lang8_sp_eng_data_object. For more information as to the error annotation types supplied by ERRANT please visit: <https://github.com/chrisjbryant/errant>.

If this is found in the list of error annotations, it means the corrector inserted a pronoun into his or her correction, which implies the pronoun was missing in the learner sentence. We can use the dependency information supplied by spaCy to investigate the missing pronoun's syntactic role within the corrected sentence, which gives us information that is relevant to our research questions about case and/or person and number affecting the relative frequencies of Pro-drop.

3 Results

3.1 Learner Sentence Breakdown

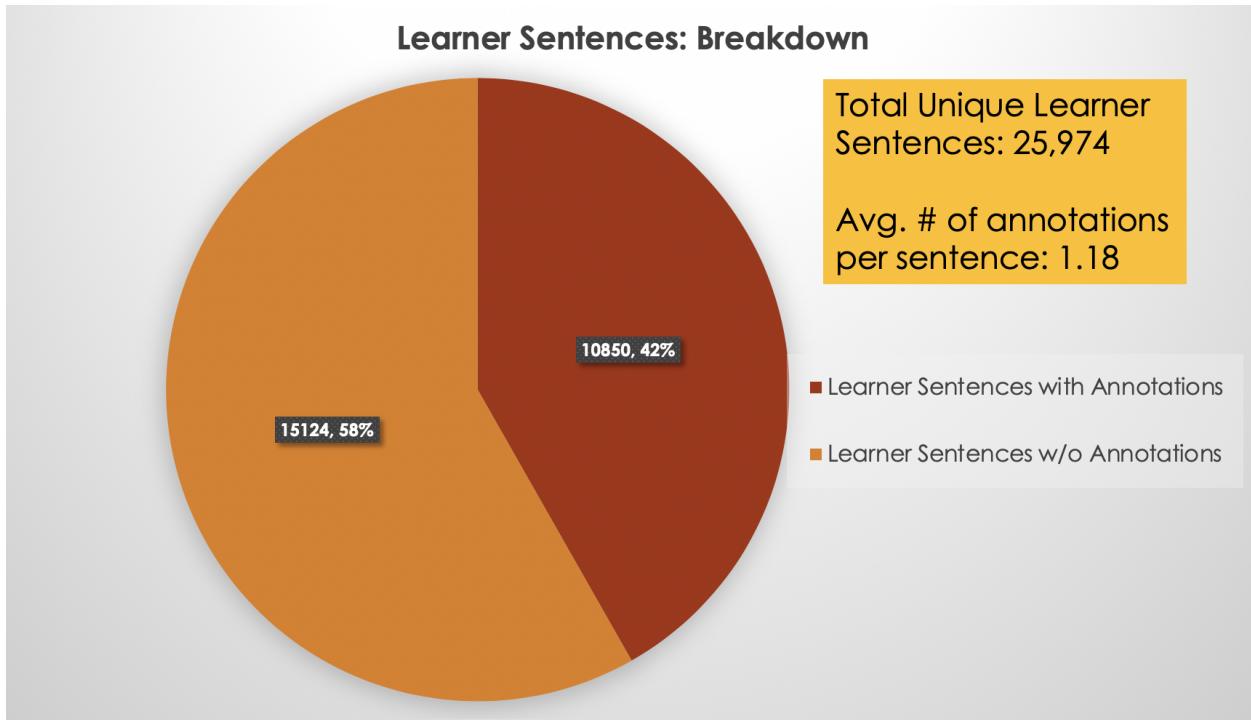


Figure 11: Distribution of unique learner sentences, with and without annotations

Given the format of the output, we could now get a general breakdown of the learner sentences within the Lang-8 corpus. In Figure 11 we see a total of 25,974 unique sentences after removing duplicates. Some sentences appeared multiple times in the corpus, once for each corrector who annotated the sentence. This resulted in a total count of 29,648 sentence instances. When there were annotations by multiple correctors, only the first corrector’s annotation was retained. A little less than half of the learner sentences have at least one annotation with an average of just 1.18 annotations per learner entry. The proportion of learner sentences without any sort of annotation is surprising and could be for a number of reasons, which will be addressed in the discussion.

| Frequencies and Error rates by Pronoun person, gender, and number | | | | | | |
|---|----------------|--------------------------------------|------------------------------------|-------------------------------------|----------------|--|
| Pronoun person number and gender | # of Sentences | # of Sentences in Unannotated subset | # of Sentences in Annotated Subset | # of sentences w/ M:PRON Annotation | Error Rate (%) | Chi Square test of 3rd NS vs. Other pronouns (p-value) |
| 1st Singular (I, me) | 21984 | 8196 | 13788 | 90 | 0.65 | |
| 1st Plural (We, us) | 2545 | 973 | 1572 | 14 | 0.89 | |
| 3rd MS (he, him) | 1746 | 712 | 1034 | 15 | 1.45 | |
| 3rd FS (she, her) | 1432 | 589 | 843 | 15 | 1.78 | |
| 3rd PL (they, them) | 2252 | 915 | 1337 | 43 | 3.22 | |
| 3rd NS (It) | 4463 | 1973 | 2490 | 255 | 10.24 | <0.0001 *** |

Table 5: Frequencies and error rates by pronoun person, gender, and number

3.2 Results: Pro-drop by person, gender and number

First we wished to determine whether person, gender and number affect the error rate of Pro-drop, specifically when it came to the 3rd person neuter singular. We used the chi square test (with p-value of <0.05 showing statistical significance). Based on the results in Table 5, the error rate of the 3rd person neuter singular pronoun *it* is significantly greater than the error rates of the other pronouns. We also see that *it* ranks second in overall frequency (with 1st person singular *I, me* being most frequent). When it comes to the number of corrections and percent corrected though, 3rd NS is more frequent, with 255 instances of Pro-drop corrections for an error rate of 10.24% .

| Frequencies and Error rates by Pronoun number, gender, and case | | | | | | |
|---|----------------|--------------------------------------|------------------------------------|-------------------------------------|----------------|--|
| Pronoun person number and gender | # of Sentences | # of Sentences in Unannotated subset | # of Sentences in Annotated Subset | # of sentences w/ M:PRON Annotation | Error Rate (%) | Chi Square (c) or Fisher exact (f) test Nominative vs Accusative (p-value) |
| I | 18230 | 7166 | 11064 | 73 | 0.66 | 0.84 (c) |
| me | 3754 | 1030 | 2724 | 17 | 0.62 | |
| we | 2200 | 828 | 1372 | 11 | 0.80 | 0.41 (f) |
| us | 345 | 145 | 200 | 3 | 1.50 | |
| he | 1371 | 540 | 831 | 13 | 1.56 | 0.75 (f) |
| him | 375 | 172 | 203 | 2 | 0.99 | |
| she | 1077 | 418 | 659 | 15 | 2.28 | 0.052 (f) |
| her | 355 | 171 | 184 | 0 | 0.00 | |
| they | 1485 | 611 | 874 | 26 | 2.97 | 0.49 (c) |
| them | 767 | 304 | 463 | 17 | 3.67 | |
| it (nom) | 3225 | 1451 | 1774 | 196 | 11.05 | 0.03* (c) |
| it (acc) | 1238 | 522 | 716 | 59 | 8.24 | |

Table 6: Frequencies and error rates by pronoun person, gender, number, and case

3.3 Results: Pro-drop by case

Next we wanted to investigate if our results could answer our second question of whether case affects the rate of Pro-drop. We used the Chi square or Fisher exact test (depending on the sparsity of our data) to see if case had any effect on the rate of Pro-drop (a p-value of $< .05$ again showing statistical significance). Table 6 shows no significant differences in error rates between nominative and accusative case for first person singular, first person plural, third person singular masculine, third person singular feminine, and third person plural pronouns. However, the difference between nominative and accusative case for third person singular neuter *it* is significant ($p < .03$), with a higher Pro-drop rate for nominative (nsubj) than accusative (dobj) *it*.

4 Discussion

4.1 Pro-drop in L2 English writing by a Spanish L1

As previously discussed, there is a difference between English and Spanish when it comes to allowing for Pro-drop. In regard to Spanish, Pro-dropping is generally allowed to a much greater extent both in nominative and accusative positions. A posited reason for this is that qualities usually reserved for pronouns are conjugated onto the verb. For example, a sentence such as *lavé los platos* ((I) washed the dishes) is not only a common example of a Pro-drop in nominative position but we can also see that the verb *lavé* is conjugated to agree with the corresponding pronoun, although it is not explicitly mentioned. In addition, Spanish pronouns can take the form of clitics, morphemes which are not independent and thus must attach to a “host” verb (Zagona, 2002). Clitics can also influence Pro-drop in nominative or accusative positions. If we look at the following sentence: *él se lo comió* (he ate (it)) we can observe the object clitic *lo* relying on the verb *comió* to imply the presence of an object. There are even sentences in Spanish that allow for Pro-drop in both positions such as *se lo bebió* ((he/she/it/etc...) drank (it)). With this in mind, we can see how the flexibility of Pro-drop manifesting in a learner’s L1 Spanish can become a negative transfer in the learner’s L2 English.

Referring to the results of Table 6, although we can see that overall Pro-dropping seems to be occurring more often in the nominative position, we do see some cases of Pro-drop occurring in the accusative position. What is even more interesting is that in addition to seeing higher error rates regarding Pro-dropping with the 3rd NS (*it*), we also see a higher rate of Pro-dropping in the accusative position for the 3rd NS. We ask ourselves why would it be the case that we find significantly higher rates of Pro-dropping with the 3rd NS as opposed to the other pronouns. One explanation for the higher rates of Pro-dropping could be the use of non-referential (or existential) *it*. The closest similarity to Spanish are the

```

S I have read and heard that would be raining all week starting today =/
A 6 6||| M:PRON|||it|||REQUIRED|||-NONE-|||0
A 12 12|||M:PREP|||from|||REQUIRED|||-NONE-|||0
A 12 13|||R:NOUN|||today.|||REQUIRED|||-NONE-|||0

```

Figure 12: An example of a non-referential ‘it’ insertion by the corrector

verbs *es* and *estar*, which are verbs indicating *to be* although they are used a bit differently. For example, a sentence such as *It is raining* demonstrates the use of the existential *it* in English and is equivalent to the Spanish sentence *Esta lloviendo*, which literally translates to *Is raining*. The existential ‘it’ is required in order to satisfy a syntactic role in English; for Spanish, it can potentially shift over to the verbs *es* and *estar*. Figure 12 is one such case of a potential negative transfer involving the existential *it*, resulting in a Pro-drop in the learner’s L2 production.

Although the results we have gathered here are not representative of all Spanish speaking L2 learners as a whole, this data certainly gives enough insight to demonstrate Pro-drop error rates for potential learners. Furthermore, the prevalence in error rate for the 3rd person neuter singular correlates with its relatively high ranking from the teacher observations. Findings such as the effect of case on the 3rd person neuter singular could be useful information for educators. This shows promise in terms of the toolkit’s usefulness, which in theory could help aggregate results in a relatively shorter amount of time.

4.2 Limitations

4.2.1 Lang-8 Learner Profiles

Unfortunately there is no data in the corpus that gives any insight into the learner’s level of proficiency in the target language. This could have proved useful for longitudinal studies to see how users’ practices improve over time. There are instances in which users self-identify with their level of proficiency via meta-linguistic discussion, but of course, this too can result

in further limitations.

Another factor that is not accounted for is the variation among different varieties of Spanish. Spanish is one of the most popular/widely used languages in the world. It is spoken as the primary language by about 360 million people globally, in more than 18 countries. Various dialectal differences exist depending on geographic region between countries and even within the same country (Posner and Sala, 2019). Keeping this in mind, conventions of language use can differ from region to region and some errors may be more prominent depending on the dialect when it comes to learner production. Furthermore, many of the correctors seem to have some degree of background knowledge of Spanish, which could also be influenced by where they acquired Spanish. Unfortunately, there is no information provided in the corpus that can inform us as to the corrector's dialect or their proficiency for that matter.

4.2.2 Quality of Corrections

The quality of corrections have also served as limitations to some degree. Although most of the corrections seem to point the learner in the right direction, there are some instances in which corrector's choice of corrections impacted the quality of the correction. Figure 13 and Figure 14 are two examples of such instances. For the first example, we can see that the learner's sentence was actually correct, with the corrector offering an alternative way to construct the sentence (within parentheses). However, the resulting edits would make the sentence ungrammatical, creating a revision that would produce "...can also helps...". In the second example, we see suggested corrections offered for a correct learner sentence. In this case, there appears to have been an incomplete edit with missing letters, which could be confusing for the learner. In addition, the word "blushed" is offered as an alternative for the learner to use, but it is not clear which word(s) "blushed" should replace.

```

Learner: Learning about suffixes and prefixes helps children to uncover word meanings.
Corrector: Learning about suffixes and prefixes (can also) helps children to uncover word meanings.

```

Figure 13: An example of a correction that could lead to an ungrammatical sentence

```

Learner: Johnson became frozen and at the same time turned red.
Corrector: Johnson became frozen and e time began to turn red / blushed.

```

Figure 14: An example of a correction with inadequate information for the learner

4.2.3 Unannotated sentences

Though we were able to successfully extract about twenty-five thousand unique learner sentences, close to half of them were without annotations and hence, without corrections. We are not certain as to what percentage of the unannotated sentences are actually error free. Many of these sentences simply may have not been seen by any correctors. For this reason, the statistical tests reported earlier were based solely on pronoun frequencies in the subset of sentences that were annotated.

4.2.4 Part-of-speech (POS) and Parsing issues

Although ERRANT proved effective in being able to provide error-type information, there are limitations worth noting involving its part-of-speech tagging and parsing.

Figure 15 shows a few instances in which ERRANT had difficulty handling multi-token edits. For example the token "It doesn't" gets labeled as an M:PRON (which is partially correct since the pronoun was missing from the original sentence). Unfortunately, "doesn't" is also part of the missing pronoun insertion. This seems to be due to the fact that the first word of the sentence (*Doesn't*) required a change in the case of its initial letter when it became the second word of the sentence – a change that is unrelated to the missing pronoun error type. In the examples of "where English" and "and I want", we can see that the error is classified as a Replacement (R) but the part-of-speech tag is simply "Other",

Examples of problems with multi-token edits

Labels might not necessarily correlate exclusively with type of edit:

S Doesn't matter now, may be nevertheless that his dead being a legacy of life for my.

A 0 1 | | M:PRON | | It doesn't | | REQUIRED | | -NONE- | | 0

Missing and uninformative labels

S I want to go to a country which its official language is English to learn and improve my English.

A 5 6 | | R:DET | | the | | REQUIRED | | -NONE- | | 0

A 7 8 | | R:OTHER | | where English | | REQUIRED | | -NONE- | | 0

A 8 9 | | R:OTHER | | is | | REQUIRED | | -NONE- | | 0

A 11 13 | | R:OTHER | | and I want | | REQUIRED | | -NONE- | | 0

Figure 15: Examples of ERRANT problems with multi-token edits

which is uninformative. Unfortunately these annotation issues may result in mislabeling or overlooking errors that would otherwise be useful for analysis.

4.3 Questions for future investigation

Given the scope and time frame of this study, other types of analysis could not be explored. Even so, there are future questions that we (or others following this work) would hope to answer.

This study has only dealt with instances of Pro-drop which ERRANT detects as insertions on the corrector's part, and which it diagnoses as missing pronouns (M:PRON). In the future, we would also like to explore instances of pronouns being deleted by the corrector, which ERRANT diagnoses as unnecessary or extraneous (U:PRON), and instances of pronouns being replaced (R:PRON). These annotations may provide valuable evidence of other types of negative transfer effects from the learner's L1.

In Figures 16 and 17 we can see over-generation of pronouns, which could possibly be due

```

S Aside we beat eggs with a little of milk and we add salt to him to pleasure.
A 0 2 |||R:OTHER|||Meanwhile,|||REQUIRED|||-NONE-|||0
A 3 3 |||M:DET|||the|||REQUIRED|||-NONE-|||0
A 7 8 |||U:PREP|||||REQUIRED|||-NONE-|||0
A 10 11 |||U:PRON||| REQUIRED|||-NONE-|||0
A 13 15 |||U:OTHER||| REQUIRED|||-NONE-|||0
A 16 17 |||R:OTHER|||taste.|||REQUIRED|||-NONE-|||0

```

Figure 16: An example of an unnecessary or extraneous pronoun (U:PRON)

```

S Hi, i am an Spanish guy that i came to Dublin on last saturday night.
A 3 4 |||R:DET|||a|||REQUIRED|||-NONE-|||0
A 7 8 |||U:PRON||| REQUIRED|||-NONE-|||0
A 11 12 |||U:PREP|||||REQUIRED|||-NONE-|||0
A 13 14 |||R:ORTH|||Saturday|||REQUIRED|||-NONE-|||0

```

Figure 17: Another example of an unnecessary or extraneous pronoun (U:PRON)

```

S For the next class I have to listen a Korean song and try to write the lyrics, I want to
choose "I can't let you go, even if I die" of 2AM, can you recommended me another song?
A 8 8 |||M:PREP|||to|||REQUIRED|||-NONE-|||0
A 9 10 |||R:ORTH|||Korean|||REQUIRED|||-NONE-|||0
A 15 15 |||M:PART|||down|||REQUIRED|||-NONE-|||0
A 30 31 |||R:PREP|||by|||REQUIRED|||-NONE-|||0
A 34 35 |||R:SPELL|||recommend|||REQUIRED|||-NONE-|||0
A 35 36 |||U:PRON|||||REQUIRED|||-NONE-|||0
A 37 38 |||R:NOUN|||song|||REQUIRED|||-NONE-|||0
A 38 38 |||M:OTHER|||to me?|||REQUIRED|||-NONE-|||0

```

Figure 18: An example of an unnecessary or extraneous pronoun (U:PRON) possibly resulting from negative transfer from Spanish

```

S i was great to hear from you!
A 0 : |||R:PRON|||t|||REQUIRED|||-NONE-|||0

```

Figure 19: An example of a pronoun replacement (R:PRON), 'i' for 'It'

to the learner's attempt to hyper-correct him/herself. Figure 18 captures a possible instance of negative transfer from the learner's L1. In English, when using the verb *recommend*, the indirect object, if present, requires the preposition (*to*), or if absent, is understood as *me*. In Spanish, one of the possible constructions does involve the expression of an overt indirect object (for example, '*me*') without a preposition, which could explain the learner's error production in English. Keeping extraneous pronouns in mind, the investigation would involve looking at the original learner sentences to determine which extraneous pronouns are being used, and at what rate. It would be interesting to see if certain types of pronouns have higher rates of extraneous use than others.

When it comes to pronoun replacements, Figures 19 and 20 illustrate some examples of errors in pronoun choice. In the first example, one assumption is that in Spanish, such a construction (*it*) does not seem possible (as to my knowledge). In place of the '*it*', Spanish operates with a reflexive pronoun, which presents itself as a possible negative transfer from the learner's L1. The second example shows a replacement from '*it*' to '*them*'. Although a concrete explanation cannot be formulated based on the learner's choice of pronoun for the second example, it would be interesting to see which types of pronouns seem to be used in error, as well as which types of corrections are favored in terms of the corrector's suggested pronoun. Since the edits here are replacements, it would be beneficial to look at both the learner and corrector sentences in order to get the full scope of the pronouns being replaced and their substitutes.

There are also other phenomena in Spanish worth exploring which could give insight into word order errors. For example, it is common knowledge that in Spanish the adjective

```
S We stood watching it and we noticed there weren't two, but three, four, five, almost six
surfers!!!!
A 3 4 |||R:PRON|||them|||REQUIRED|||-NONE-|||0
```

Figure 20: Another example of a pronoun replacement (R:PRON), 'it' for 'them'

follows the head that it modifies (e.g: "The car green" vs. "The green car"). It would be interesting to explore the prominence of any word order (WO) errors and the token types that are corrected; doing so could give us a better idea of what types of syntactic rules may be transferring over from the learner's L1.

In the teacher judgments, "do omission/misuse of a variation", and "conjugation" rank relatively high. This may be due mainly to the differences in conjugation systems between English and Spanish. It would be worthwhile to investigate the contexts in which SVA errors occur, since any resulting data could be informative for learners, educators, and researchers. Interestingly enough, two teachers ranked "Question Formation" as a common error type. English and Spanish exhibit a few differences in terms of question formation, one example being the inversion of the verb relative to the subject (Zagona, 2002). Consider the English sentence (*What did John need?*). In Spanish, the equivalent would be (*¿Qué necesitaba John?*). We wonder if phenomena such as this inversion may be a potential negative transfer that could affect the learner's L2 question formation in English. Although data on this issue were not examined in the current study, it would be interesting to see what dynamics may transfer over and affect question formation patterns in English.

4.4 Future Work

Keeping the previous questions in mind, a future goal is to attach rankings and relative frequencies to a variety of learner errors. Such efforts may not only inform teaching strategies but could also serve as a basis to facilitate meta-linguistic discussion among teachers and students. Basturkmen et al. (2002) has observed that meta-linguistic discourse can be an

important means for students to explore language forms in the classroom. In fact, research has shown that the more students engage in meta-linguistic conversations, the more likely they are to incorporate their knowledge into their production. And so it is our hope that the use of this toolkit in tabulating the relative frequencies of various error types for a variety of L1s will aid in promoting meta-linguistic discussion and thus increase the likelihood of student achievement and the effectiveness of pedagogical practices.

5 Conclusion

This thesis mainly discussed the following questions in regard to learner English with a Spanish L1:

1. Do gender and number of the pronoun have an effect on drop rate?
2. Do the rates of Pro-drop depend on pronoun case? Specifically, are Subject (nsubj) pronouns more likely to be dropped than Direct Object (dobj) pronouns?

We discussed earlier the need for more specific relative frequency information about error types produced by English Language Learners, especially considering that the learner's L1 may influence the types of errors that are produced via negative transfer (Cummins, 1981). This provided the rationale for our study.

We were able to collect observations of error types from teachers that are a part of an international public high school serving a predominantly Latinx population in order to test the usefulness of an error analysis toolkit. Based on the frequencies of the teachers' observations, Pro-drop was one of the highest ranked error types, which is why we chose to focus on it for the purpose of this study.

We used data consisting of learner/corrector sentence pairs from Lang-8 (Mizumoto et al., 2011), specifically a subsection of English L2 speakers with a Spanish L1, in order to analyze error types related to Pro-drop. We then made use of ERRANT (Bryant et al., 2017) to extract error type information which we added to the dependency parse and part-of-speech information provided by spaCy (Honnibal and Johnson, 2015). Analyzing instances of Pro-dropping based on pronoun type and position in the corrected sentences required a realignment of edits since the position of the pronoun insertion could be affected by prior corrections of the sentence. Essentially, we created a complex data object that held part-of-speech, and dependency information for the learner and corrector sentence pairs, alongside

corresponding error type information.

When looking to see if results showed that Pro-drop error rates are affected by gender and number, we found that the 3rd person singular neuter pronoun *it* had a significantly higher error rate than other pronouns. The answer to the first question above is: Yes, *it* is dropped more often than other pronouns.

When investigating whether or not pronoun case affects Pro-drop rates, we found that Pro-drop occurred significantly more often in the nominative than in the accusative position and more often with *it*. Pro-drop of the 3rd person singular neuter pronoun in subject position is typically an allowed construction in Spanish. Our assumption is that negative transfer is responsible for the error rates in the L2 production. The answer to the second question then is: Yes, *it* in subject position is more likely to be dropped than in object position.

Though there is future work anticipated, we hope findings from this study may prove beneficial to the educational, NLP, and research communities. We feel that educators specifically can take advantage of using current and future analyses of error types to complement their pedagogical practices in order to ensure achievement for language learners as a whole.

References

- Basturkmen, H., Loewen, S., and Ellis, R. (2002). Metalanguage in focus on form in the communicative classroom. *Language awareness*, 11(1):1–13.
- Bryant, C., Felice, M., and Briscoe, E. J. (2017). Automatic annotation and evaluation of error types for grammatical error correction. Association for Computational Linguistics.
- Cummins, J. (1981). Empirical and theoretical underpinnings of bilingual education. *Journal of education*, 163(1):16–29.
- Donahue, S. (2001). Formal errors: Mainstream and esl students. In *Conference of the Two-Year College Association (TYCA)*.
- Felice, M., Bryant, C., and Briscoe, T. (2016). Automatic extraction of learner errors in esl sentences using linguistically enhanced alignments. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 825–835.
- García, O. and Wei, L. (2013). *Translanguaging: Language, bilingualism and education*. Springer.
- Honnibal, M. and Johnson, M. (2015). An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1373–1378, Lisbon, Portugal. Association for Computational Linguistics.
- Lee, I. (2011). Working smarter, not working harder: Revisiting teacher feedback in the l2 writing classroom. *Canadian modern language review*, 67(3):377–399.
- Mizumoto, T., Komachi, M., Nagata, M., and Matsumoto, Y. (2011). Mining revision log of language learning sns for automated japanese error correction of second language learners. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 147–155.
- Nicholls, D. (2003). The cambridge learner corpus: Error coding and analysis for lexicography and elt. In *Proceedings of the Corpus Linguistics 2003 conference*, volume 16, pages 572–581.
- Posner, R. and Sala, M. (2019). Spanish language. <https://www.britannica.com/topic/Spanish-language#ref331588>. Accessed: 2020-01-20.
- Zagona, K. (2002). *The syntax of Spanish*. Cambridge University Press.