Developing the Linguistic Features of Academic Writing in Spanish

Part 1: Background and Research Questions

One of the goals of collegiate foreign language (FL) studies programs is for students to develop advanced language capacities, defined as the “ability to read, write, listen, reflect, and communicate intelligently about a culture’s multiple facets” (Swaffar, 2006, p. 248). As the *sine qua non* of evidence of learning in FL studies programs, writing is central to both the development and assessment of advanced language capacities. And while numerous reports detail the shortcomings of university FL majors’ oral proficiency (Carroll, 1967; Magnan, 1986; Swender, 2003), fewer details are available regarding writing outcomes and development. The current project seeks to explore how the writing of Spanish majors develops in terms of complexity, both lexical and syntactic, in a traditional Spanish studies major.

While some propose that academic writing must be detached and objective (e.g., Schleppegrell, 2004; Zwiers, 2008), recent scholarship has foregrounded the persuasive nature of academic writing, highlighting that language is used to “acknowledge, construct, and negotiate social relations” (Hyland, 2005, p. 173). Building convincing arguments and constructing a credible stance requires linguistic sophistication and nuance, which are realized through a variety of linguistic devices. The syntactic and lexical features of academic discourses differ from those of everyday discourses (e.g., Gee, 2002; Schleppegrell, 2004). Among the features that differ are grammatical intricacy and lexical density. Written language has low grammatical intricacy but high lexical density; this happens through the use of technical, discipline-specific terminology and grammatical metaphor (clauses in which syntactic and semantic content are incongruent). The proportion of content words to non-content words, i.e. grammatical words, is greater in written text than spoken text. In an opposite trend, the grammatical intricacy is less in written text, with fewer clauses per syntactic unit.

University foreign language studies students transition from everyday, informal, personal discourses in introductory and intermediate language classes to public, formal, discipline-specific discourses in major courses. While numerous studies detail linguistic accuracy with respect to specific grammatical structures, with some considering development via cross-sectional samples (e.g., Blackwell & Quesada, 2012; Malovrh & Lee, 2010), no studies consider longitudinal development of syntactic lexical complexity over the course of a Spanish studies major.

Recent work coming out of Georgetown University’s German Department has documented development in the syntactic features of academic register during the course of an academic major in German (cf. Byrnes, 2009; Byrnes, Maxim & Norris, 2010; Ryshina-Pankova, 2010, 2011), yet the generalizability of the findings is questionable given the program’s unique curricular model. Whether similar development occurs in more traditional curricular models has yet to be determined. The research coming out of the Georgetown University German Department has spurred dialogue on FL pedagogy and literacy through its focus on functional linguistic features, not just linguistic accuracy, when examining FL writing. The purpose of this study is to apply these methods of analysis to a more traditional context that better represents the realities of most university FL studies departments in the United States. This project seeks to (1) expand the research base on foreign language academic writing to include Spanish and (2) explore how features of academic writing develop over the course of a traditional foreign language studies major. The specific research questions that guide the study are:

- Does the lexical complexity of written texts produced by Spanish studies students change over the course of the major? If so, how?
- Does the syntactic complexity of written texts produced by Spanish studies students change over the course of the major? If so, how?

Identifying the features of written academic discourse students acquire with little to no explicit instruction can then inform pedagogical decisions at both the course and program level.
Part 2: Project Description
The goals of this study are to apply recent advances in analytical tools to the writing of university Spanish studies students, focusing on syntactic and lexical complexity through longitudinal analysis of a variety of measures. Three measures will be used to explore each construct: amount of subordination, types of clause embedding, and clause-combining strategies for syntactic complexity and lexical density, lexical diversity, and lexical sophistication for lexical complexity. Each measure is described below together with its relationship to written academic discourse.

• Amount of subordination. Written, academic discourse has higher levels of clause embedding or subordination than oral language, which is comprised primarily of main clauses joined together through linking. Quantitative analysis of syntactic complexity rely on the t-unit, a unit of measurement which consists of a main clause and all subordinate and non-clausal units attached to it, from this the number of clauses per t-unit can be calculated and writing can be classified as more (more clauses per t-unit) or less (fewer clauses per t-unit) syntactically complex.

• Type of clausal embedding. The type of clauses that are subordinated is also indicative of higher levels of complexity in language production as some types of clauses, such as object of comparison (i.e., He is the man whom no one is smarter than.), are less frequently embedded. The Noun Phrase Accessibility Hierarchy (Keenan & Comrie, 1977) will be used to guide this analysis, following the assumption that clauses higher in the hierarchy are acquired later (cf. Lantolf, 1998; Byrnes & Sinicrope, 2008). This analysis complements the amount of subordination as it addresses the quality or type of subordination, not just the quantity.

• Clause-combining strategies. Clausal units can be linked to the main clause in a variety of ways; paratactic and hypotactic clausal combination strategies, which build on and add to the main clause, are more indicative of oral, informal registers. Written, academic discourse thus contains less parataxis (e.g., Veni, vidi, vici.) and hypotaxis (e.g., I wrote a play, which was directed by a young schoolteacher, who then taught me what “real” plays are.) and more embedding. The frequency of each type of clause - main, paratactic, hypotactic, and embedded - will be calculated to determine the strategies collegiate writers use.

• Lexical density. Academic language becomes more complex by becoming more lexically dense (Halliday, 1994); in other words more lexical items are packed into each clause. In order to calculate lexical density, the total number of content words will be divided by the total number of words produced in a text.

• Lexical diversity. This measure examines whether a variety of words are used or whether words are repeated. More academic texts tend to have higher rates of lexical diversity. The Guiraud Index will be used to calculate this number: square root of number of distinct content words / total number of content words.

• Lexical sophistication. Academic discourse requires precision and technical vocabulary; many of these words are restricted to specific contexts of use and are consequently less frequent in corpora. Words will be classified according to three bands: 1-2000 (most frequent), 2000-3500 (mid-frequency), and 3500-5000 in line with previous studies (Bardel, Gudmundson & Lindqvist, 2012).

Each of the measures incorporated into the current analysis has been applied to writing of other groups of learners, most notably Spanish heritage learners (Achugar & Colombi, 2008; Colombi, 2000, 2002), university German foreign language students (e.g., Byrnes, 2009), and English as a second/foreign language speakers (e.g., Drury, 1991; Martin, 1996). With the exceptions of Byrnes (2009) and Ryshina-Pankova (2011), most studies of foreign language writing consider performance at a given point in time or development over an academic semester; they do not consider how syntactic and lexical complexity change over longer periods of time. Exploration of longitudinal development as well as the focus on students in a traditional foreign language studies major are the significant contributions to be made by the present project.
A corpus of 72 texts, from 8 different students, comprise the data set for the current project. The corpus represents writing from each course in the Spanish studies major, from the first course in the major to the senior capstone project. Understanding which features develop without explicit pedagogical interventions can inform both theory in second language acquisition as well as decisions regarding instructional practices. Because the program at the University of Minnesota adheres to a typical university foreign language studies model, findings will be representative of other undergraduate foreign language studies majors and implications will be applicable to instruction in other languages and at other institutions.

Funding for a graduate research assistant is requested through this grant. The research assistant will be responsible for coding and calculating each of the six features that comprise this analysis. The proposed research assistant, Tripp Strawbridge, is a graduate student in Hispanic Linguistics with a specialty in Spanish second language acquisition. Although he has conducted pilot studies on foreign language writing previously, four of the analytical measures will be new to him. In addition, this will be his first opportunity to consider development within individual learners. By working under my guidance on this project, he will not only gain familiarity with new methods of analysis but will develop a deeper understanding of development and how it can be documented via linguistic analyses. This collaborative project will benefit his professional preparation in the form of conference presentations and publications as well as informing his future scholarship on Spanish as a foreign language.

Grant funds will exclusively used for summer RA funding. My personal research funds will be used to purchase the software programs necessary for analysis as well as the statistical consulting as part of the analysis.

This project has been approved by the University’s Institutional Review Board. Its study number is 1403E49010, and it fell within the exempt category of review.

The estimated timetable for the project is outlined in the table below.

<table>
<thead>
<tr>
<th>Dates / Hours</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Already complete</td>
<td>Data collection</td>
</tr>
<tr>
<td>June 5-14 30 hours</td>
<td>Mentoring RA on analytical measures</td>
</tr>
<tr>
<td>June 14-Aug 11 170 hours</td>
<td>Developing coding scheme</td>
</tr>
<tr>
<td></td>
<td>Learning software programs</td>
</tr>
<tr>
<td></td>
<td>Data coding (~2.5 hours/text)</td>
</tr>
<tr>
<td></td>
<td>Statistical analysis &amp; interpretation</td>
</tr>
<tr>
<td>Aug 14-18 20 hours</td>
<td>Continued statistical analysis &amp; interpretation</td>
</tr>
<tr>
<td></td>
<td>Preparation of presentations / manuscripts</td>
</tr>
</tbody>
</table>

**Part 3: Relevance and Plans for Dissemination**

Given the growing focus on advanced language capacities and academic literacy development in the fields of second language acquisition, educational linguistics, and foreign language pedagogy, the results will be of interest to a wide audience. The results of the proposed project will provide valuable insights for researchers as well as university foreign language studies departments. In addition, the findings will inform curricular decisions in the Department of Spanish and Portuguese Studies.

Results will first be presented via conference presentations at the Second Language Research Forum, American Association of Applied Linguistics Annual Conference, and Symposium of Second Language Writing. Based on feedback from presentations, publications will be developed during Spring and
Summer 2018. It is anticipated that findings will be submitted to The Modern Language Journal, Journal of Second Language Writing, and Foreign Language Annals for publication.

References


