

Strategies and
Analyses of Language
and Communication
in Multilingual and
International Contexts

Strategies and Analyses of Language and Communication in Multilingual and International Contexts

Edited by

David Levey

Cambridge
Scholars
Publishing



Strategies and Analyses of Language and Communication in Multilingual
and International Contexts

Edited by David Levey

This book first published 2020

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2020 by David Levey and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-4324-2

ISBN (13): 978-1-5275-4324-9

TABLE OF CONTENTS

Introduction	viii
David Levey	

Part 1 Language Acquisition

Chapter One.....	3
Empowering Advanced Vocabulary Learners of Spanish: An Application of the SCAP Tool Patrick Goethals	
Chapter Two	13
The Lexico-syntactic Continuum in the Academic Writing of Students of Spanish: A Comparative Study in Bilingual Programmes Magdalena Jiménez and Esperanza Falcón	
Chapter Three	25
The Acquisition of Ser and Estar in 2L1 English-Spanish Data Radoslava Stankova Laykova and Raquel Fernández Fuertes	
Chapter Four.....	39
A Comparative Study of the Acquisition of the Spanish Past Tense Aspect by Mandarin Chinese Learners using Guided and Semi-Guided Written Productions Yuliang Sun, Lourdes Díaz and Mariona Taulé	
Chapter Five	49
Spontaneous Translanguaging in the Basque Autonomous Community Amaia Fernandez, Pilar Sagasta and Nagore Ipiña	
Chapter Six.....	61
Pauses and Mental Processes during a Transaction Task: The Case of English (L1/L2), Spanish (L1/L2) and Heritage Spanish Núria Enríquez, Lourdes Díaz and Yuliang Sun	

Part 2 Language Teaching and Assessment

Chapter Seven.....	75
Creating Linguistic Diversity in the Classroom: The Beliefs of English Teaching Assistants Angela George	
Chapter Eight.....	83
From Teaching <i>English for International Business</i> to Teaching <i>International Business in English</i> : Considerations for an ESP Teacher to Start Providing CLIL Teacher Training Candela Contero Urgal	
Chapter Nine.....	91
CEFR-related Proficiency Tests of Productive Skills (Oral and Written): Towards Reliable Ratings Caroline Shackleton and Nathan Turner	
Chapter Ten	101
Not up to Standard! Creating a Global Assessment Framework and Universal Proficiency Test for Maritime English Alison Noble, Ludwina Van Son and Peter John	

Part 3 Psycholinguistics, Pragmatics and Discourse Analysis

Chapter Eleven	111
Rapid Naming, Reading, Comprehension and Syntactic Production by Italian Dyslexic Children With and Without Specific Language Impairment Emanuele Casani	
Chapter Twelve	123
Strategic Communication in English-Medium Lectures at the University of Zaragoza: Pre-Emptying/Remedying Breakdowns in Communication and Establishing Solidarity among ELF Speakers María Ángeles Velilla Sánchez	

Chapter Thirteen 131
The Use of Standardised Maritime English: A Study of a Small Corpus
of VHF Communications in Southern Spain
María Dolores Perea Barberá and Daniel Parada Galindo

Chapter Fourteen 141
Trump vs. Clinton: An Analysis of the US Presidential Debates of 2016
Carla Fernández Melendres and Antonio Moreno-Ortiz

Part 4 Corpus Linguistics and Translation Studies

Chapter Fifteen 153
Exploiting Multiparallel Corpora as a Measure for Semantic Relatedness
to Support Language Learners
Johannes Graën and Gerold Schneider

Chapter Sixteen 169
Researching the Phraseological Units of the Academic Genre through
a Domain-Specific Corpus of Engineering, Maritime Navigation
and Applied Sciences Research Articles in English
María-Araceli Losey-León

Chapter Seventeen 179
From Translation Professionals to Translation-Related Professionals or
“Non Translators”
Marián Morón Martín

Chapter Eighteen 187
The Identification of Potentially Ambiguous Noun Phrases in Scientific
English: A Crucial Aspect of Translator and Post-Editor Training
François Maniez

Chapter Nineteen 197
Creating Materials for Autonomous Practice and Reflection in Business
Interpreting
Francisco J. Vigier-Moreno

Chapter Twenty 205
Draw Me a Tale: A Translational Analysis of the Illustrations
in *James and the Giant Peach*
Bruno Echauri Galván

INTRODUCTION

DAVID LEVEY

UNIVERSITY OF CADIZ (SPAIN)

This collection offers an insight into some of the new tendencies and current research in Applied Linguistics which will be of interest to scholars as well as language practitioners and professionals. Covering a variety of aspects related to language and communication in international situations and plurilingual contexts, this volume presents empirical findings by researchers from ten countries working in 20 different universities and centres of excellence.

Some of the chapters have been penned by one sole author, while others are the result of inter-university and international collaborative projects involving two or more experts. The focus is broad, with samples and reports taken from various languages, contexts and communities (English, Spanish, Basque, Catalan, Mandarin Chinese and Italian). The contributions touch on a wide range of stimulating topics with far-reaching implications. These include learner empowerment, intercultural learning, translanguaging, linguistic diversity, sociolinguistic competence, learning/teaching for special purposes, career and professional contexts and outlets, the need for reliable evaluation, the Common European Framework of Reference and developments in machine translation, to name just a few.

The 20 papers have been ordered under four broad headings, although inevitably a certain overlap exists and some papers straddle more than one linguistic field. A brief introductory summary of each of the chapters is provided below.

Part 1: Language Acquisition

The first contribution, by Patrick Goethals, presents a vocabulary learning application which forms part of a wider project being developed at Ghent University (Belgium). Using state-of-the-art methodology based on Natural Language Processing, Corpus Analysis and Computer-Assisted Language Learning, the aim is to produce an effective customised tool for teachers and advanced learners of Spanish.

The second chapter provides a qualitative and quantitative analysis of academic writing. In their study, Magdalena Jiménez (Università Roma Tre) and Esperanza Falcón (Universidad Pablo de Olavide) compare the syntactic and lexical complexity of the writings of Spanish L1 and Spanish L2 students studying in bilingual secondary schools in Spain and Italy.

In the following chapter, Radoslava Stankova Laykova and Raquel Fernández Fuertes from Valladolid University look at the ways in which English-Spanish bilinguals acquire the Spanish copula verbs *ser* and *estar*. In particular, the authors explore whether they are acquired simultaneously or consequentially and gauge the extent to which level and linguistic context play a role in the acquisition process.

Using informants studying in Nanjung (China) and Barcelona and Granada (Spain), Yuliang Sun, Lourdes Díaz and Mariona Taule analyse how aspectual past tenses in L2/L3 Spanish are acquired by Mandarin Chinese students. Based on written data, the authors aim to identify the use and acquisition patterns of the Spanish *pretérito indefinido* and *pretérito imperfecto* under different pragmatic-discursive constraints.

Amaia Fernández, Pilar Sagasta and Nagore Ipiña from Mondrogón University analyse the linguistic repertoire and discourse practices of 153 emergent trilingual children (Basque, Spanish and English) in primary schools in the Basque Autonomous Community, focussing, in particular, on their use of translanguaging.

In the final chapter of this section, Núria Enriquez (University of Houston), Lourdes Díaz (Pompeu Fabra University) and Yuliang Sun (University of Barcelona) analyse the role of pauses as processing indicators in the oral production of L1/L2 Spanish and English native speakers as well as speakers of Spanish as a heritage language.

Part 2: Language Teaching and Assessment

In the first of the four chapters dedicated to teaching and assessment, Angela George (University of Calgary) explores the importance of linguistic diversity and dialectal awareness in the classroom based on an analysis of the views and attitudes towards teaching different geographical varieties of English from a sample of thirty-three English language and culture teaching assistants working in Spain.

In the next chapter, Candela Contero Urgal (University of Cádiz) considers the difficulties of implementing Content and Language Integrated Learning (CLIL) in Andalusian universities. She argues that clearer guidelines are needed to help English for Specific Purposes (ESP) teachers adapt to the new CLIL environment. Using survey data provided

by English for International Business (EIB) students, the author aims to raise awareness of those teaching techniques which can be applied to both content subjects and language courses.

The following two chapters deal with language testing. In the first, Caroline Shackleton and Nathan Turner from the Centro de Lenguas Modernas (University of Granada) argue that, not only must language test development constructs accurately represent target language use, but that reliable test scores must also be provided. The authors exemplify how Rasch statistical analysis is being used to ensure rating reliability for productive skills in the University of Granada's bi-level CertAcles exam.

Given the internationalisation of shipping and the importance of English language competence and accuracy in the maritime industry, Alison Noble and Ludwina Van Son from the Antwerp Maritime Academy (Belgium) in collaboration with Peter John from Jade University of Applied Sciences (Germany) discuss realistic options for creating an assessment framework for Maritime English, using authentic items to test communicative competence.

Part 3: Pragmatics and Discourse Analysis

The first contribution of Part 3 analyses the reading, comprehension and production skills of dyslexic children. Emanuele Casani (Università Ca' Foscari Venezia) argues that slowness in rapid naming of visual stimuli is a typical problem for children with developmental dyslexia (DD) and specific language impairment (SLI) and is longitudinally related to written decoding, rather than to written comprehension. Comprehension and production of syntactically complex structures also prove problematic. The study explores these skills and their relationship to reading in Italian primary schoolchildren.

English as a Lingua Franca (ELF) has become a major and expanding field of research within Applied Linguistics. The next contribution by María Ángeles Velilla Sánchez (University of Zaragoza) examines the dynamics of ELF in academic interactions and the approaches and strategies involved. Based on the analysis of 12 English-medium lectures recorded in two different degrees at the University of Zaragoza, the author analyses communicativeness at pragmatic and discourse analysis levels.

In Chapter 13, M^a Dolores Perea Barberá and Daniel Parada Galindo (University of Cádiz) examine a corpus of oral VHF radio communication messages to gauge the extent to which the language employed in these exchanges follows the International Maritime Organization (IMO) recommendations, and whether the "Standard Marine Communication

Phrases”, as established by the IMO, are actually used by shipboard crews navigating in the Bay of Algeciras (Spain) which is one of the busiest areas in Europe for maritime traffic.

The final chapter in this section deals with political rhetoric and discourse analysis. Carla Fernández Melendres and Antonio Moreno-Ortiz from the University of Málaga analyse the 2016 US presidential debates using various methodologies (corpus-based analysis, discourse analysis and sentiment analysis) with the aim of determining to what extent the language choices of Donald Trump and Hillary Clinton influenced voters and the subsequent election results.

Part 4: Corpus Linguistics and Translation Studies

Incorporating recent developments in computational linguistics, Johannes Graën and Gerold Schneider from the University of Zurich demonstrate how a large multilingual resource, combined with statistical methods, can be used to directly generate lists of items which are useful for language learners, translators and linguists. To demonstrate their effectiveness and reliability, potentially easily confused non-compositional German particle verbs, as well as potential false friends, are tested.

In the second contribution in this section, María-Araceli Losey-León (University of Cádiz) presents the preliminary findings of an ongoing corpus-based investigation project of the prototypical Academic English phraseology found in research articles in the fields of Science and Technology. Here, the author discusses the corpus design, criteria and compilation method before going on to illustrate the extraction techniques.

In the light of reforms in the education system, Marián Morón Martín (Universidad Pablo de Olavide) considers the different routes into non-specific translation employment available to Translation graduates in a globalised and boundary-free market. The author reflects on non-translation as a transversal manifestation of the translation profession today, and presents a summary of key research findings on graduates working in non-translation posts.

Chapter 18 explores machine translation (MT) processing of English ambiguous structures in the translation of scientific texts into French and Spanish, and more specifically <Adj_N_and_N_of_N> structures. François Maniez (Université Lyon 2) argues that while MT is successful in cases where noun phrases with coordinated nouns have been stored in the translation memory, other cases require caution on the part of post-editors. The author argues that in order to improve future translators’ performance and employability, it is crucial to teach them to recognise potentially

ambiguous noun phrases.

In the next contribution, Francisco J. Vigier-Moreno (Universidad Pablo de Olavide) argues that, while Business Interpreting offers an attractive career opportunity for graduates in Translation and Interpreting, training resources are difficult to come by. The author presents a recent project which aims to meet the increasing demand by creating training materials which provide trainees with autonomous and reflective practice within Business Interpreting settings, using the most prominent features of Business Interpreting and real-life cases.

In the final chapter, Bruno Echaui Galván (University of Alcalá) focuses on transmutation processes from text into illustration, building on the premise that translating and illustrating a book share common mechanisms. For this end, the author analyses illustrations from Roald Dahl's *James and the Giant Peach* using translation procedures such as literal translation, explicitation, omission, and transcreation in order to determine the type of dynamics established between the text and its illustration, and the implications such interactions may have on the reader.

PART 1

LANGUAGE ACQUISITION

CHAPTER ONE

EMPOWERING ADVANCED VOCABULARY LEARNERS OF SPANISH: AN APPLICATION OF THE SCAP TOOL

PATRICK GOETHALS
GHENT UNIVERSITY (BELGIUM)

1. Introduction

This paper will present a vocabulary learning application that was created within the broader Research & Development SCAP project (Spanish Corpus Annotation Project, www.scap.ugent.be). The overarching aims of this project are (a) to strengthen the links between Natural Language Processing (NLP), Corpus Analysis (CA) and Computer-Assisted Language Learning (CALL), (b) to promote the use of tagged and lemmatised corpora among learners and teachers of Spanish, and (c) to promote the autonomy of advanced learners (Little 2007) by exploring computer-driven possibilities to customise learning materials.

The vocabulary learning application was designed to fulfil the pedagogical needs of advanced learners of Spanish (B2+) who wish to develop their vocabulary knowledge in a specific domain (e.g. business, health services, science, etc.). In a classroom setting, it is very difficult to satisfy these individual needs, as the interests of advanced learners may be very diverse, and the pedagogical answer should be given “just-in-time”, especially when the learners are (semi-)professionals. Therefore, the application is designed to generate customised learning materials, adapted to the advanced proficiency level and the field of interest of the learner. The learner can work with the application without the intervention of a teacher, and, although the application integrates state-of-the-art technology, no advanced technological or computer skills are required.

2. Vocabulary learning

In this section, I will discuss two questions that are crucial in vocabulary curriculum design, namely the selection of the target vocabulary items, and the distinction between informal and formal learning activities. I will focus especially on the way in which the SCAP application helps to bring into practice the principles that have been extensively discussed in academic bibliography.

2.1. Vocabulary selection

Since vocabulary learning is a time-consuming task, one of the main challenges consists of selecting the appropriate vocabulary items for successive learning activities (Nation 2016). There are at least four principles that should play a role in specialised vocabulary selection for advanced learners: (a) items that are supposedly already known should be excluded from explicit learning activities, (b) when an advanced learner chooses to work on a specialised semantic domain, the chosen items should be highly specific to this domain, (c) items that have cognates in other languages known by the learner can be excluded from comprehension tasks, and (d) infrequent specialised items can be excluded from initial production tasks, or at least should be first introduced in a comprehension task. I will show how the SCAP tool integrates these four principles.

The starting point is a corpus that is part-of-speech-tagged and lemmatised. SCAP includes several specialised corpora, but the user will also be able to input a new corpus. The tool will automatically tag and lemmatise the corpus. Especially for a morphologically rich language such as Spanish, it is crucial to lemmatise the data. As a first output, the user can call up an Excel file including several types of metadata for all lemmas of the main part-of-speech categories (nouns, verbs, adjectives and adverbs). Table 1-1 shows a small selection of results for the noun category of a 230K-word corpus of speeches at shareholder meetings (the pedagogical context could be an interpreter who is preparing him/herself for an interpreting task). The nouns appear in decreasing order of frequency, but, importantly, this is not the only and not even the main criterion that will guide the item selection.

First, the column “Level-TWS” defines whether the lexical item appears in a popular thematic word list among Dutch- and German-speaking learners of Spanish, namely the *Thematischer Grund- und Aufbauwortschatz Spanisch* (TWS), originally published by Ernst Klett Verlag for German, and adapted for the Dutch market by Intertaal. This

publication contains around 5000 vocabulary items which are subdivided into elementary and intermediate levels. SCAP assigns to the (lemmatised) items from the target corpus a label “elementary” or “intermediate” if the lemma is included in TWS, or “new” if it does not appear in the list. Activities designed for learners at B1-B2 level could still include the intermediate items in production tasks but exclude them from comprehension tasks. Activities for students at a B2+ level should clearly focus on the “new” category. Obviously, this coding is especially relevant for those students who first worked with this vocabulary learning method, but SCAP can be combined with other initial learning methods, provided that an index list of the vocabulary items is created.

Secondly, the tool defines the semantic specificity of the items by combining two measures. The first measure is called “%Diff” (Gabrielatos and Marchi 2011), and compares the (normalised) frequency of the lemma in the target corpus with its (normalised) frequency in a reference corpus. The higher the score, the higher the specificity (or “keyness”) of the lexical item. The design allows the user to choose between different reference corpora, which, in this case, is a 7M-word corpus of youth literature: given the very different nature of the corpora, it can be expected that %Diff will have high values. (The score “+100.000” means that there were 0 cases in the reference corpus, which is an ad hoc solution for the problem that it is not possible to divide by zero.) %Diff complements the frequency data, since it allows the user to identify those words that are not only frequent in the target corpus, but are also more frequent than in a reference corpus. In this sense, *sostenibilidad* or *endeudamiento* are more “exclusive” than *caso* or *año*, although the latter are more frequent.

However, as Gries (2008) has emphasised, overall corpus frequency data should be complemented by data on frequency in the different parts of the corpus (called “diffusion”), since it may be the case that a high overall frequency is due to a special use in one subpart. The diffusion score varies between 0 (equal distribution) and 1 (unequal). In practice, the threshold to distinguish between both categories seems to be situated around a score of 0.8. From a pedagogical point of view, it seems reasonable to start with more equally distributed words, and then proceed to unequally distributed words—as such, we would start with words which can be expected to appear in many examples of this text type, and later on discuss words which appear only in specific instances. Applying this to the nouns in Table 1-1 with an overall frequency lower than 100 and a %Diff score higher than 15000 means that we would start with items such as *sostenibilidad*, *diversificación*, *integración*, *gestor*, *endeudamiento*, *liquidez*, or *sinergia*, and not with items such as *adjudicación*, *desinversión* or

aerolínea. This seems especially relevant in the case of the last word, whose high frequency is due to a specific subsample of the corpus (namely a shareholder meeting of an airline company).

Finally, we find a “SequenceMatcher” score, which measures the similarity between the Spanish word and its possible equivalents in Dutch. This score ranges between 0 (extremely non-similar) and 1 (exact similarity). This helps to exclude cognates from a comprehension task, since words such as *dividendo*, *integración*, *diversificación* or *sinergia* seem to be easily understandable for Dutch-speaking learners who know the words *dividend*, *integratie*, *diversificatie* or *synergie*. Yet, it can be relevant to include these words in a production task.

LEMMA	Level-TWS	FREQ	%Diff_Ref	Diffusion	Match (Dutch)
año	elementary	1643	711	0,12	
crecimiento	elementary	539	92453	0,31	
desarrollo	intermediate	318	17451	0,36	
accionista	elementary	256	52650	0,27	
dividendo	new	251	+100.000	0,35	0,94 (dividend)
caso	intermediate	100	4	0,47	
integración	elementary	81	83352	0,62	0,76 (integratie)
sostenibilidad	new	78	+100.000	0,52	
aerolínea	new	72	222441	0,94	
diversificación	new	43	+100.000	0,61	0,83 (diversificatie)
endeudamiento	new	37	+100.000	0,67	
liquidez	new	32	+100.000	0,74	0,74 (liquiditeit)
gestor	new	29	+100.000	0,66	
desinversión	new	28	+100.000	0,85	
sinergia	new	27	+100.000	0,74	0,75 (synergie)
adjudicación	new	26	+100.000	0,82	
desempleo	intermediate	9	13808	0,83	
rapidez	new	9	-21	0,89	

Table 1-1: Selection of lemmatised nouns from a corpus of speeches at shareholder meetings

The different metadata help to make a meaningful vocabulary selection that is well adapted to the proficiency level, the intended semantic domain, the comprehension or production task, or the L1-L2 interferences. The vocabulary selection can be done manually, by manipulating the Excel filters, or by making use of predefined filters combining several parameters (e.g. a filter that defines the “200 most interesting words for a first production task” would focus on the 200 most frequent items excluding items with a relatively low %Diff or a relatively high Diffusion score). This selection will be the input for the learning activities described below.

2.2. Learning activities

One of the main academic debates about vocabulary learning concerns the pedagogical advantages of explicit versus implicit learning activities, for example, cloze exercises, translation exercises, etc. versus extensive reading (e.g. Laufer and Rozovski-Roitblat 2011, Nation 2018). The latter has the advantage of being more authentic and more stimulating for learners eager to use language for content-related purposes, but the former is more time efficient (Groot 2000, Nation 2018). SCAP facilitates both implicit and explicit learning activities by generating different kinds of didactic output. The examples below are ordered from more reading-oriented towards more instruction-oriented. A very short list of target items is used for reasons of space, but this can be extended to various hundreds of target items.

a. text reading order suggestion: the tool identifies the corpus texts that contain the highest proportions of the target words (also related to their length). Thus, the learner who wishes to read a full-length text can choose the most efficient input. For example, in case we are interested in all the nouns included in Table 1-1, the tool suggests a shareholder speech (of +/- 5000 words) that contains 72% of the target items.

b. reading contexts: selecting this option, the user will be able to download a WORD-file with a list of sentences that are retrieved from the selected corpus and illustrate the target items. Sentences appear in the same order as in the original texts and target items are highlighted. Obviously, this considerably reduces the amount of text that a learner has to read to find all the target words in context. A key is added with translation suggestions which illustrate the meaning of the word (see next application).

Léxico en contexto

1. • Contamos con unas holgadas ratios de liquidez que exceden ampliamente los niveles requeridos .
2. - Familias con mayor capacidad de consumo e inversión , por la mejora en el empleo y el menor endeudamiento .
3. Tenemos la suerte de contar con un Consejo comprometido , que vela por el futuro y sostenibilidad de la Compañía y que trabaja para asegurar que Telefónica continuará ganando relevancia con un modelo de negocio responsable .

Glosario

1 liquidez : liquiditeit. 2 endeudamiento : schuldenlast. 3 sostenibilidad : duurzaamheid.

Fig. 1-1: Automatically generated sentence-based reading activity

c. glossary: the tool scrapes various open access dictionaries and uses the API of machine translation tools (DeepL, Google, Windows). By comparing the translation suggestions, it proposes the most plausible candidates to be included in a glossary. If there are no translations proposed by more than one tool, a “?” symbol is added. The output can be read in WORD (see below) or in Excel, which in turn allows the user to import the data into flashcard vocabulary training programs. For the moment, the application is designed for Dutch-speaking users, but this will be extended to English and French.

Glosario

adjudicación (sust): [?] *onderscheiding, toekenning*

endeudamiento (sust): *schuldenlast*

gestor (sust): *manager*

liquidez (sust): *liquiditeit*

sostenibilidad (sust): *duurzaamheid*

Fig. 1-2: Automatically generated Translation Glossary (Spanish-Dutch)

d. cloze sentences: the tool creates a list of cloze sentences for the target items. In the example, the first letter and a translation suggestion are given. At the end of the document, a key is included.

Rellenar huecos

1. No obstante , los más de 500 millones de caja , netos de deudas , nos permite mantener un balance saneado y una política de *d.* estable como anticipamos hace ya 10 años cuando nos presentamos a los mercados por primera vez . (*deelta, dividend, winstaandeel*)

2. El resultado operativo antes de amortizaciones también crece , un 4,7 % en términos orgánicos , gracias a la contención de los gastos , las *s.* en Brasil y Alemania , la vuelta al crecimiento en España y la aceleración en Reino Unido . (*synergie*)

3. Por ello , Mediaset España mantiene año a año su compromiso de hacer accesible su programación a las personas con discapacidad visual o auditiva , como instrumento de *i.* social y cultural de estos colectivos . (*integratie*)

Clave

1 dividendo 2 sinergia (sinergias) 3 integración

Fig. 1-3: Automatically generated cloze exercise.

e. Corpus query interface: finally, through a web interface, the user has access to the tagged and lemmatised corpus where several advanced query possibilities are available in order to deepen the knowledge of the lexical targets. It goes beyond the scope of this paper to illustrate the full possibilities, but the query below nicely shows the added value of working with tagged and lemmatised data. Concretely, the query takes the noun *crecimiento* as its input, and looks for all adjectives preceding or following the noun. The result of the query is an Excel file where the adjectives appear in descending order of total frequency (the ordering can also be by position). In a second sheet, the user can find all the examples. Similar queries can be realised in order to find frequent n-grams, combinations of nouns and verbs, verbs and prepositions, etc.

	A	B	C	D	E
1	Lem_1 ▾	Lem_2 ▾	Ante ▾	Post ▾	Total ▾
2	crecimiento	rentable	0	31	31
3	crecimiento	económico	0	19	19
4	crecimiento	sostenido	0	12	12
5	crecimiento	importante	8	0	8
6	crecimiento	sostenible	0	7	7
7	crecimiento	anual	0	6	6
8	crecimiento	futuro	0	6	6
9	crecimiento	fuerte	5	0	5
10	crecimiento	mayor	4	0	4
11	crecimiento	sólido	3	1	4
12	crecimiento	significativo	0	4	4
13	crecimiento	alto	3	0	3
14	crecimiento	moderado	0	3	3

Fig. 1-4: Sample query result.

3. Conclusion

By applying state-of-the-art NLP and CA methodologies, SCAP allows the user to guide vocabulary selection in a scientifically meaningful and systematic way. Specially designed to support advanced learners (and their teachers) to broaden their vocabulary knowledge of specific semantic fields, it integrates measures that help to exclude words that are possibly already known by advanced learners, or words that are not specific for the semantic domain. Moreover, by automatising the generation of a set of didactic activities, it allows the user to customise the learning materials, which makes it possible for learners to choose their field of interest, and work autonomously combining both extensive reading tasks, a translation glossary and explicit instruction tasks.

References

- Gabrielatos, C., Marchi, A. 2011. Keyness. Matching metrics to definitions. *Corpus Linguistics in the South*, University of Portsmouth, 5 November 2011.
- Goethals, P., Lefever, E., Macken, L. 2017. SCAP-TT: Tagging and lemmatising Spanish tourism discourse, and beyond. *Ibérica* 33: 279-288.
- Gries, S. T. 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13: 403-437.
- Groot, P. 2000. Computer assisted second language vocabulary acquisition. *Language Learning & Technology* 4: 56-76.
- Laufer, B., Rozovski-Roitblat, B. 2011. Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research* 15: 391-411.
- Little, D. 2007. Language learner autonomy: Some fundamental considerations revisited. *International Journal of Innovation in Language Learning and Teaching* 1: 14-29.
- Nation, P. 2016. *Making and Using Word Lists for Language Learning and Testing*. John Benjamins.
- Nation, P. 2018. Reading a whole book to learn vocabulary. *ITL-International Journal of Applied Linguistics* 169: 30-43.

CHAPTER TWO

THE LEXICO-SYNTACTIC CONTINUUM IN THE ACADEMIC WRITING OF STUDENTS OF SPANISH: A COMPARATIVE STUDY IN BILINGUAL PROGRAMMES

MAGDALENA JIMÉNEZ
ROMA TRE UNIVERSITY (ITALY)
AND ESPERANZA FALCÓN
UNIVERSITY PABLO DE OLAVIDE (SPAIN)

1. Introduction

Writing is already difficult in L1, let alone in L2. This research aims to analyse academic writing in the context of History classes. In particular, we have studied the interplay between syntactic and lexical complexity in texts written by Spanish L2 students in Italy and Spanish mother-tongue students in Spain. To this end, we analysed the narratives produced by these students within the framework of the BIMAP, a state-funded research project (Proyectos I+D de Excelencia), which aims to describe CALP (Cognitive Academic Language Proficiency), in order to obtain a better understanding of advanced language structures in bilingual contexts.

2. Rationale and theoretical framework

Considering the general linguistic difficulties involved in writing, our initial purpose was to compare the performance of L1 and L2 students in Spanish to shed light on the process of writing by studying its lexico-syntactic interface. In this respect, cognitive linguists consider that syntax,

semantics and pragmatics do not constitute modules or levels with independent analysis, nor are they connected through interfaces, but they are on a continuum and can be analysed simultaneously (Maldonado 2012, 214).

Complexity can be of several kinds: syntactic, lexical, phraseological, morphological, and semantic. In this study, we were particularly interested in syntactic complexity, a construct first researched in the sixties when Hunt began to study syntactic maturity in English L1 (Hunt 1965) in the context of generative grammar. The seventies saw the beginning of research in English L2 and this continued until the end of the nineties. Some authors started to refer to syntactic complexity in the field of English as L2. In particular, Skehan (1996) began to relate complexity to accuracy and fluency creating the acronym CAF, which is a construct that comprises these three sub-constructs which maintain a dynamic relationship over time in the development of the learner's interlanguage.

It is also important to distinguish between syntactic maturity and syntactic complexity because language acquired later is not necessarily always structurally more complex (Pallotti 2009, 593). Moreover, it is important to be aware that syntactic complexity has different levels or dimensions. We, therefore, need to distinguish complexity at different levels: multiclausal, interclausal and intraclausal. According to the type of complexity, different indices or measures should be used. For instance, multiclausal complexity can be assessed by calculating sentence length; interclausal complexity can be appreciated by analysing subordination density; and intraclausal complexity can be measured using finite clause length, that is to say, the ratio between the total number of words and finite verb forms. This last index is important because, in the development of academic writing, research has shown the increase in intraclausal complexity which depends to a large extent on the development of nominalisation. In fact, as Hunt (1965) already indicated, writers move from main clauses to reduced phrases (see Wolfe-Quintero et al. 1998, 72).

In order to better understand the lexico-syntactic interface in academic writing, the contribution of systemic-functional linguistics (SFL) is essential. In fact, such an interface relates ultimately to the well-known distinction between dynamic and synoptic styles first coined by the founding father of the SFL school, Halliday (1985).

Dynamic styles meet oral needs in low levels of formality and show an increase of subordination. At the other end of the spectrum, synoptic styles can be considered as an evolution in written language, related to high levels of formality in academic communication and they stand out due to

their conceptual density, the development of nominalisation, greater cohesion and the lengthening of clauses and phrases.

As for lexical complexity, we decided to measure the number of phrasemes, also known as phraseological units. From a psycholinguistic point of view, they are “chunks” or “bundles” internalised by speakers in an automatic way. In brief, they are a non-conscious repertoire resulting from the high frequency of co-occurrence of several elements (at least two) and they are fixed through the different types of memory that are involved in the learning process. Their acquisition facilitates the codification and decodification of language information and is related to the fluency and authenticity of speakers and learners (Lee 2007).

In this regard, Gries’ (2008) definition, which emphasises co-occurrence and one semantic unit, is very useful. For this reason, the concept of the phraseme is considered here in a broad sense, and two types can be distinguished: first, a compositionality type such as collocations, and second, a non-compositionality type which is more idiomatic or metaphoric such as proverbs.

In this study, we have followed the phraseme classification of Granger and Paquot (2008) who distinguish three main types of phrasemes: i) referential phrasemes which are used to convey a content message. They refer to objects, phenomena or real-life facts, e.g. *sembrar plantas* (Eng.: to sow crops); ii) textual phrasemes which are used to structure and organise content, e.g. *en cuanto a* (Eng.: as for); and iii) communicative phrasemes which express feelings or beliefs towards a propositional content. They are used to explicitly address interlocutors, either to focus their attention, include them as discourse participants or influence them. This type of phraseme can be a whole sentence as in the case of proverbs or slogans, e.g. *En la Guerra Civil no hubo ni buenos ni malos* (Eng.: In the Spanish Civil War there were neither “good guys” nor “bad guys”). The reader can clearly see that the distinction resembles the SFL metafunctions: ideational, textual and interpersonal.

Drawing on Granger and Paquot (2008), this study considers three major types of phrasemes: i) referential phrasemes: lexical collocations, grammatical collocations, idioms, irreversible bi- and trinomials, similes, compounds and verbal periphrases; ii) textual phrasemes: complex prepositions, complex conjunctions, linking adverbials, textual sentence stems; and iii) communicative phrasemes: speech act formulae, attitudinal formulae, concluding attitudinal formulae, proverbs, common places, slogans and idiomatic sentences.

The results section will show the prevalence of referential and textual phrasemes over those of a communicative nature. In general, communicative

phrasemes were the least used. The reason appears to be related to the nature of the task.

3. Methodology

In this research, we analysed a sample of 20 texts in Spanish L1 and L2 from two larger studies that had been carried out previously. It is important to note that Spanish L2 had been placed in an immersion¹ setting akin to the Content and Language Integrated Learning model² (CLIL) so widespread in Europe now. More precisely, the students took Literature and History courses in Spanish L2 at Liceo Virgilio (Rome), and students' narratives were collected by Jiménez in 2015. As for the Spanish L1 students³, a single data collection was carried out in class by Falcón in a secondary school in southern Spain in 2016.

Spanish L1 texts were written in digital format whereas Spanish L2 were transcribed into digital format and analysed by hand, considering two indices: i) the finite clause length: i.e. the ratio between words and finite verb forms; and ii) the density of phrasemes: i.e. the ratio between phrasal units (PHU) and finite clauses.

As far as the genre type was concerned, both studies intended to describe the evolution of narratives. More precisely, the description focused on the evolution from narratives as anecdotal, naïve renditions of the past, as for example in fairy tales, to more mature accounts of past events which were more personal and elaborated.

4. School context

The students had the following features in common: the bilingual context of the schools (Spanish L2/Italian L1 in Italy and Spanish L1/English L2 in Spain), the academic writing in the context of History classes, their age (18 years old), the time allocated (50 minutes), similar instructions, and a topic from the syllabus: the Industrial Revolution (in Spanish L2) and the Spanish Civil War (in Spanish L1).

As for the differences, we need to emphasise that, in terms of access to that particular school, the Italian students had different selection criteria

¹ See Jiménez (2017) for a complete analysis of CALP in Spanish as a foreign language as opposed to Spanish as an immersion language.

² See Dalton-Puffer et al. (2014) for a characterization of CLIL.

³ See Falcón (2017) to delve into the evolution of CALP over the years in mid and late adolescence L1 Spanish students.

and on average presented a higher-level academic profile in contrast to the Spanish students. Other features included years of schooling (penultimate in L2 / last year in L1); and linguistic level (B2-C1 in L2 / native speakers in L1).

5. Results

The results indicate significant differences in all the measures used, starting with text length. In fact, Spanish L2 participants on average wrote longer texts by about 50 words.

As for the indices considered, there is variability and some correlation between the indices in both conditions. In Figures 2-1 and 2-2, the performance of Spanish L1 students can be observed.

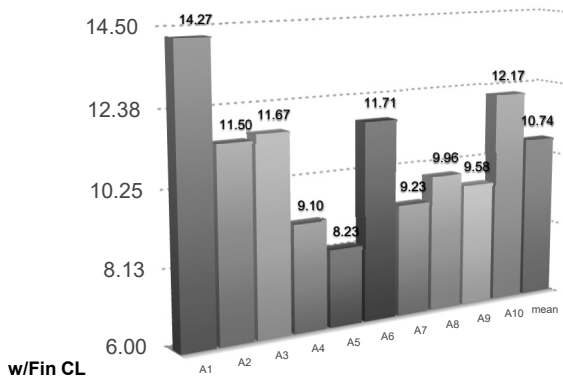


Fig. 2-1. Finite clause length in Spanish L1 students
See centrefold for this image in colour

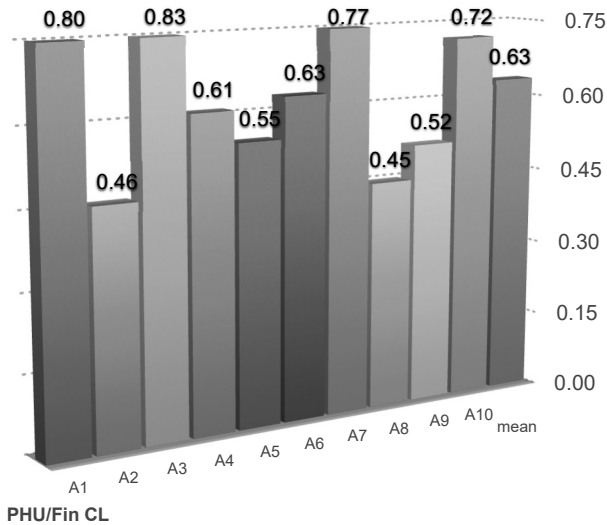


Fig. 2-2. Phraseological unit density in Spanish L1 students
See centrefold for this image in colour

By the same token, the performance of Spanish L2 students was also characterised by variability with some correlation between both indices in the whole group of students, as can be seen in Figures 2-3 and 2-4.

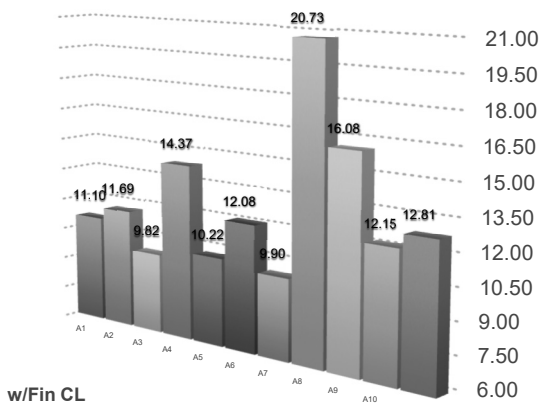


Fig. 2-3. Finite clause length in Spanish L2 students
See centrefold for this image in colour