VIAL

Vigo International Journal
of Applied Linguistics

Universidade Vigo

VIAL. Vigo International Journal of Applied Linguistics.

Editorial Advisory Board
Michelle Aldridge (Cardiff University)
Allison Beeby (Universitat Autònoma de Barcelona)
Martin Bygate (Lancaster University)
Heidi Byrnes (Georgetown University)
Teresa Cadierno (University of Southern Denmark)
Jasone Cenoz (Universidad del Pais Vasco)
Guy Cook (King's College London)
Michael Daller (University of the West of England, Bristol)
Jean-Marc Dewaele (Birkbeck College, University of London)
Ana Fernández Dobao (University of Washington)
Mª del Pilar García Mayo (Universidad del País Vasco)
Peter Garrett (Cardiff University)
Kimberly Geeslin (Indiana University)
Joan Kelly Hall (Penn State University)
Zhaohong Han (Columbia University)
Linda Harklau (University of Georgia)
Alberto Hijazo-Gascón (University of East Anglia)
Alex Housen (Vrije Universiteit Brussel)
Scott Jarvis (Ohio University)
Iraide Ibarretxe-Antuñano (Universidad de Zaragoza)
Nan Jian (University of Maryland)
Alan Juffs (Pittsburgh University)
Jim Lantolf (Pennsylvania State University)
Carmen Muñoz Lahoz (Universitat de Barcelona)
Terence Odlin (Ohio State University)
Ignacio Palacios (Universidade de Santiago de Compostela)
Amos Paran (University of London)
Sagrario Salaberri (Universidad de Almería)
Luc van Doorslaer (Lessius University College, Belgium)
Roberto A. Valdeón (Universidad de Oviedo)
Fernando Vilariño Freire (Universitat Autònoma de Barcelona)
Joanna Weatherby (Universidad de Salamanca)
Scientific Advisory Board
Stuart Campbell (University of Western Sydney)
Michael Hoey (University of Liverpool)
Enric Llurda (Universitat de Lleida)
Rosa Mª Manchón (Universidad de Murcia)
Rafael Monroy (Universidad de Murcia)
Aneta Pavlenko (Temple University)
Martha Pennington (University of Durham)
Carmen Pérez Vidal (Universitat Pompeu Fabra)
Félix Rodríguez (Universitat d´Alacant)
Barbara Seidlhofer (Universität Wien)
Michael Sharwood-Smith (University of Edinburgh)
John Swales (University of Michigan)
Elaine Tarone (University of Minnesota)
Krista Varantola (Tampere University, Finland)

Editor-in-chief
Mª Rosa Alonso Alonso (Universidade de Vigo)

Managing editor
Marta Dahlgren (Universidade de Vigo)

VIAL is an international and peer-reviewed journal which is indexed and abstracted in CINDOC (Centro de Información y Documentación Científica), EBSCO, ERIH PLUS, JCR, Web of Knowledge (Arts and Humanities Citation Index, Social Sciences Citation Index), Linguistics Abstracts, Linguistics and Language Behavior Abstracts, Latindex, MLA Directory of Periodicals, Scimago Journal Rank Scopus, Ulrich's Database

Webpage: www.webs.uvigo.es/vialjournal/
© Servizo de Publicacións da Universidade de Vigo, 2019
Printed in Spain - Impreso en España
I.S.S.N.: 1697-0381 - Depósito Legal: VG-935-2003
Imprime e maqueta: Tórculo Comunicación Gráfica, S.A. Deseño da portada: Elena Gómez.

Reservados tódolos dereitos. Nin a totalidade nin parte deste libro pode reproducirse ou transmitirse por ningún procedemento electrónico ou mecánico, incluíndo fotocopia, gravación magnética ou calquera almacenamento de información e sistema de recuperación, sen o permiso escrito do Servicio de Publicacións da Universidade de Vigo.

VIAL

Vigo International Journal of Applied Linguistics

Number 16 - 2019

Editor-in-chief
Rosa Alonso

Managing Editor
Marta Dahlgren

Contents:

Do prospective primary school teachers suffer from Foreign Language Anxiety (FLA) in Spain?

Marian Amengual-Pizarro
Universitat de les Illes Balears (UIB), Spain
marian.amengual@uib.es

Abstract

The main aim of this study is to investigate Foreign Language Anxiety (FLA) in relation to the teaching and learning of English as a foreign language. A total of 75 prospective primary school teachers at the University of the Balearic Islands (UIB) took part in this study. A small questionnaire that included the Foreign Language Classroom Anxiety Scale (FLCAS) (Horwitz et al., 1986) was used to collect data. The results of this study show that most participants experience average and high anxiety levels in the language classroom. Communication apprehension was reported to be the main source of FLA, followed by fear of negative evaluation and test anxiety. The findings also revealed the strong association between FLA, motivation, language proficiency and degree of self-confidence. Furthermore, the data indicate that the primary source of speaking anxiety is related to participants' lack of English proficiency. This may have potential adverse effects on the confidence levels of L2 teachers, their target language use, and their instructional competence (Horwitz, 1996).

Keywords: Foreign Language Anxiety (FLA), prospective primary school teachers, foreign/second language learning, English teaching, Spanish EFL students.

Resumen

El principal objetivo de este trabajo es investigar la ansiedad lingüística vinculada a la enseñanza y aprendizaje de la lengua inglesa como lengua extranjera. Un total de 75 futuros profesores de primaria de la Universitat de les Illes Balears (UIB) participaron en este estudio. Se diseñó un breve cuestionario que incluía la *Foreign Language Classroom Anxiety Scale* (FLCAS) (Horwitz et al., 1986) para recoger los datos. Los resultados

de este estudio demuestran que la mayoría de participantes experimentan niveles de ansiedad medios y altos en la clase de lengua extranjera. La aprensión comunicativa es la fuente principal de dicha ansiedad seguida por el miedo a la evaluación negativa y la ansiedad ante los exámenes. Los resultados también demuestran la fuerte asociación entre la ansiedad lingüística, la motivación, el dominio de la lengua y el nivel de auto-confianza. Además, los datos indican que la principal fuente de ansiedad está relacionada con el escaso dominio de la lengua inglesa. Esto puede tener efectos potencialmente negativos en los niveles de confianza de los profesores de segundas lenguas, su uso de la lengua meta en clase y su competencia pedagógica (Horwitz, 1996).

Palabras clave: Ansiedad lingüística, futuros profesores de primaria, aprendizaje de segundas lenguas, enseñanza del inglés, aprendices españoles de inglés como lengua extranjera.

1. Introduction

In the field of language education, the relevance of affective variables has long been recognised (Phillips, 1992; Dewaele, 2005; Pavlenko, 2011). Among them, Foreign Language Anxiety (FLA) has been the focus of much research attention since this affective variable is believed to exert negative influence on foreign or second language (L2) learning by interfering with the successful development of the target language (Krashen, 1985; MacIntyre & Gardner, 1991; Onwuegbuzie et al. 2000; Scovel, 2001; Horwitz, 2001; Williams & Andrade, 2008; Liu, 2013).

FLA is described as a complex psychological construct linked to a specific type of context or situation (Horwitz et al., 1986; MacIntyre & Gardner, 1989). Horwitz et al. (1986: 128) defined FLA as "a distinct complex construct of self-perceptions, beliefs, feelings, and behaviours related to classroom language learning arising from the uniqueness of the language learning process". FLA is then characterised as a situation-specific type of anxiety (independent of any other type of anxiety) prompted by L2 learning contexts. Although various studies (Spielmann & Radnofsky, 2001; Dörnyei, 2005) reveal that some degree of FLA may have a positive or beneficial influence on language learning (i.e. *facilitating* anxiety), most research findings point to the negative effects of language anxiety (i.e. *debilitating anxiety)* on L2 effective development (Cheng et al., 1999; Horwitz, 2000, 2001; Onwuegbuzie et al., 2000; Scovel, 2001; Gregersen 2003; Elkhafaifi, 2005). These negative feelings may be manifest in physical symptoms such as sweating, headaches, heart palpitations or other emotional reactions such as frustration, lack of concentration, excessive worry, forgetfulness, absenteeism, etc., which may have an adverse effect on the learning of the L2 regardless of language

teaching methodology (Horwitz & Young, 1991; Aida, 1994; MacIntyre 1999; Casado & Dereshiwsky, 2001; Gregersen, 2007; Williams & Andrade, 2008; Liu, 2013).

FLA has been associated with specific language skills such as speaking, listening, writing or reading (Sellers, 2000; Cheng, 2002; Elkhafaifi, 2005; Woodrow, 2006). However, speaking is usually regarded as the greatest source of FLA in the classroom since L2 students must try to express and communicate themselves in a language they do not fully master (Horwitz et al., 1986; MacIntyre & Gardner, 1991; Phillips, 1992; Young, 1992; Aida, 1994; Matsuda & Gobel, 2004; Woodrow, 2006; Criado & Mengual, 2017). This may challenge the image students have of themselves as competent speakers, causing tension, worry and frustration (Ortega-Cebreros, 2003; Arnaiz-Castro & Guillén, 2013; Al-Saraj, 2014). As a result, anxious students are more likely to avoid participating and engaging in activities that may be perceived as more complex or risky when using the L2 (MacIntyre & Gardner, 1991, 1994; Aida, 1994; Sheen, 2008), thus missing crucial opportunities to improve their oral communication skills in the target language (Gregersen, 2007; Kim, 2009; Arnaiz & Guillén, 2012). Due to their especially communicative-oriented approach and interactive nature, foreign language classroom situations are therefore regarded as particularly anxiety-inducing contexts (Yang, 2012).

Horwitz et al. (1986) designed a self-report instrument, the Foreign Language Classroom Anxiety Scale (FLCAS), to measure the level of anxiety experienced by learners in L2 classroom settings. Although the FLCAS deals with general foreign language anxiety, particular emphasis is placed on speaking and listening, since these skills have been found to be the most substantially affected by FLA. Horwitz et al. (1986) associate FLA with three main interrelated anxiety constructs: communication apprehension, test anxiety, and fear of negative evaluation. Communication apprehension refers to the type of anxiety experienced when communicating with people in the L2, which may result in frustration and self-consciousness in some students. Test anxiety is related to fear of failing in tests. Test-anxious students are likely to be afraid of making mistakes since they associate the L2 classroom to testing situations rather than to communicative learning environments (Gregersen & Horwitz, 2002). Finally, fear of negative evaluation arises from an excessive worry about being evaluated negatively by others. Students who suffer from this latter type of performance anxiety tend to avoid using the L2 to prevent negative personal and academic judgements from other people. According to Horwitz et al. (1986), these three main interrelated factors are shown to influence the effectiveness of language learning.

Language anxiety also seems to be interrelated with other affective variables such as motivation. In fact, numerous studies have found significant negative correlations

between FLA and motivation (Clément et al., 1994; Rodriguez & Abreu, 2003). Research suggests that motivated students tend to be less anxious when learning the foreign language, and are also willing to put more effort into learning tasks. In other words, motivation appears to promote the learning of the target language whereas FLA is likely to hinder language development and achievement. Other variables such as gender (Aida, 1994; Bekleyen, 2009; Park & French, 2013; Öztürk, 2016), course grades (Arnaiz-Castro & Guillén, 2012, 2013), prior experience, and perceived self-proficiency have also been shown to have an influence on FLCA (Horwitz et al., 1986; Matsuda & Gobel, 2003; Phillips, 2003; Elkhafaifi, 2005; Kongchan & Wareesiri, 2008).

Most of the research studies on FLA conducted so far have focused on language anxiety experienced by L2 learners rather than by L2 teachers. However, Horwitz (1996) placed special emphasis on the harmful effects of FLA on non-native language teachers, who are also identified as foreign language learners regardless of their advanced level of English proficiency. Horwitz explains that the process of learning an L2 is never finished, and many non-native teachers of the language suffer from anxiety when using the foreign language. Teachers with high levels of anxiety are likely to make little use of the target language in class, which may affect the quality of classroom instruction, and contribute to reduce teachers' self-confidence. According to Horwitz (1996: 366): "Foreign language anxiety can inhibit a teacher's ability to effectively present the target language, interact with students, and serve as a positive role model as a language learner". In the same vein, Williams (1991) concluded that anxious teachers are shown to be less effective in their instructional practices. Thus, FLA seems to be a potentially negative factor influencing the quality of language instruction.

Today's emphasis on communicative-oriented approaches has placed increasing pressure on L2 teachers who need to demonstrate a high level of proficiency in the target language so as to be able to respond effectively to the linguistic needs of L2 students (Zheng, 2008; Amengual-Pizarro, 2013). In fact, research has related FLA experienced by L2 teachers to general language proficiency and oral communication skills, which have been found to be especially anxiety-provoking factors (Horwitz et al., 1986, Horwitz, 1996). Thus, Kim & Kim (2004) concluded that, among other variables, the limited linguistic proficiency of L2 student teachers as well as their lack of self-confidence while being compared to native teachers were identified as some of the main sources of teaching anxiety. İpeck (2007) also related FLA experienced by L2 teachers to making mistakes while using the target language as well as to being compared with native speakers of the language. Among the different language skills, speaking and listening seem to be the main causes for concern to student teachers while teaching a foreign language. Other stress-related factors associated to the level of

anxiety of L2 teachers are also motivation and teaching experience (Ortega-Cebrero, 2003; Rodríguez & Abreu, 2003; Gagliardi & Maley, 2010; Arnáiz-Castro & Guillén, 2013).

In light of the importance of FLCA on L2 development and achievement, there is a need to conduct further research on foreign language teaching anxiety. To date, investigations into FLA experienced by prospective primary school teachers are clearly scarce. However, understanding and identifying the causes of this affective variable is of paramount importance in order to enhance L2 teaching and learning (Williams, 1991; Horwitz, 1996). Furthermore, within the trend towards the internationalization of Spanish universities, knowledge of the English language has become a compulsory degree requirement for all student teachers, regardless of their chosen specialised subjects (i.e. Tutorial Action, Language and Audition, Physical Education, Arts and Musical Education, etc.). The generalist orientation of teacher education programmes in Spain also allows primary school teachers holding a B1 or B2 certificate of English proficiency (Common European Framework of Reference for Languages, CEFR, 2001) to teach English at primary school level, irrespective of the specialist training they may have completed. On this basis, and given the key role the learning of English plays in today's world, it is essential to explore the FLA construct from a L2 student teachers' perspective in order to promote the English proficiency of primary school teachers, and encourage the development of anxiety-reducing strategies that may lead to the creation of more supportive and effective L2 learning environments.

2. Research questions

In spite of the relevant role FLA plays in L2 teaching and learning, there has been a shortage of research that addresses FLA in Spain. Most of these studies have been conducted in secondary schools (Ortega-Cebreros, 2003; Martínez Agudo, 2013; Criado & Mengual, 2017) and Official Language Schools (Pérez-Paredes & Martínez-Sánchez, 2000-2001). Only a few studies have explored FLA towards English as a foreign language in a university context (Arnaiz-Castro & Guillén, 2012, 2013). In fact, to the best of the author's knowledge, studies which specifically investigate foreign language teaching anxiety in Spanish universities are practically absent (see Arnaiz-Castro & Guillén, 2013). Therefore, this study aims to address this research gap by examining the level of FLCA experienced by Spanish pre-service teachers while learning English as a foreign language and the main variables associated with their language anxiety levels. Specifically, the following research questions were posed:

1) What level of FLCA do prospective primary school teachers have towards the learning of English?

2) What is the main source of FLCA among future primary school teachers?

3) Are there any differences in the degree of FLCA that participants experience across different formative itineraries or specialised education programmes?

4) Is FLCA affected by participants' gender, and level of English language proficiency?

5) What are the main sources of FLCA reported by student teachers when using the English language?

3. Method

3.1. *Participants*

The participants in this study were all second-year prospective teachers studying on a four-year university degree in Primary School Education at the University of the Balearic Islands (UIB), Spain. A total of 75 students enrolled in a compulsory English degree course, 'English language and teaching II', awarded 6 ECTS (European Credit Transfer System) credits, took part in this study. This English language course aims to help student teachers to satisfy the mandatory requirement for all undergraduates to attain a B2 level of English proficiency (CEFR) before graduation, regardless of their chosen specialised subjects. Previously, participants had all taken a compulsory English course during their first year studies, 'English language and teaching I' (of 6 ECTS credits) at B1 level of English proficiency (CEFR), as part of their bachelor's degree requirements. None of the respondents in this study were native speakers of English.

The sample consisted of 63 (84%) females and 12 (16%) males. With regard to age, the majority of respondents (84.9%) were between 19 and 25 years of age, 9.6 % of the respondents were between 26 and 35 years, 4.1 % were between 36 and 45 years old and 1.4% were older than 45 years.

3.2. *Instrument and data collection*

The data for this study were collected by means of a questionnaire which was administered to all participants in mid-February 2018. The questionnaire included three different sections. The first section consisted of background information regarding participants' gender, age, and their university entrance examination marks

in the English Test (ET). The Spanish translated version of the Foreign Language Classroom Anxiety Scale (FLCAS) developed by Horwitz et al. (1986) was included in the second section of the questionnaire (see Pérez-Paredes & Martínez Sánchez, 2000-2001). The FLCAS is the most widely accepted self-report measure of general foreign language anxiety. The Scale consists of 33 items which are rated on a 5-point Likert scale ranging from 1 ('strongly disagree') to 5 ('strongly agree'). The FLCAS is intended to assess three types of performance anxieties related to foreign language anxiety, namely: Communication apprehension (items 1, 4, 9, 14, 15, 18, 24, 27, 29, 30 and 32), test anxiety (items 3, 5, 6, 8, 10, 11, 12, 16, 17, 20, 21, 22, 25, 26 and 28), and fear of negative evaluation (items 2, 7, 13, 19, 23, 31 and 33). An anxiety score is calculated for each respondent by adding up their scores on the 33 items. Total scores on the FLCAS range from 33 (low level of anxiety) to 165 (high level of anxiety). Reverse coding is carried out for negatively worded items (T= 9) so, in all instances, a high score indicates a high degree of anxiety.

The third section of the questionnaire asked participants to state the formative itinerary or specialised education programme (Educational Support, Language and Audition, Foreign Language (English), etc.) they intended to follow. Participants were also asked to report their feelings towards the English language course (whether they liked it or not), as well as towards the use of the target language in class (whether they felt either confident or, on the contrary, anxious or nervous while using the L2).

Respondents took approximately 25 minutes to complete the questionnaire during a regular class period time. The quantitative data collected were analysed with the Statistical Package for the Social Sciences (SPSS) 22.0. The internal consistency measure of the FLCAS in this study was .92, which is very similar to the Cronbach's alpha coefficient of the Scale obtained by Horwitz et al. (1986) in their study (= .93). The results, therefore, indicate a high level of internal consistency for the present sample (75 respondents).

4. Results and discussion

4.1. *What level of FLCA do prospective primary school teachers have towards the learning of English?*

Total scores were calculated for each respondent in order to assess their level of FLA towards the learning of English. Anxiety scores in this study ranged from 54 to 141 points. Participants were then classified into one of the three following groups according to their level of language anxiety: high anxiety, average or moderate

anxiety, and low anxiety. The mean language anxiety score (M) of the Scale for the 75 participants was 105.06, and the standard deviation (SD) was 21.87. Following other studies (Tum, 2012; Öztürk, 2016), respondents who obtained scores of more than one standard deviation (SD = 21.87) above the general mean language anxiety score (M = 105.06) were classified as students with high anxiety (i.e. scores between 127 and 141). Students with low anxiety were those whose scores were more than one SD below the M (i.e. scores between 54 and 83). Finally, students with average or moderate anxiety were those who obtained scores of one SD above and one SD below the M (i.e. scores between 84 and 126). Analysis of the descriptive data showed that the majority of pre-service teachers (63.2%) appear to be moderately anxious in the language classroom. However, it is worth noting that 22.1% of the student teachers reported experiencing high levels of language anxiety as opposed to 14.7% of the respondents who admitted feeling low levels of anxiety. Therefore, overall, results indicate that 71.9% of the participants seemed to experience average and high levels of anxiety in the L2 classroom (Liu, 2006; Arnáiz & Guillén, 2012; Tum, 2012).

4.2. *What is the main source of FLCA among future primary school teachers?*

In order to explore the main source of FLCA among participants, the mean scores and standard deviations were computed for each type of performance anxiety associated with FLCA, namely, communicative apprehension, test anxiety and fear of negative evaluation. The descriptive results reveal that communicative apprehension is awarded the highest mean score (M = 3.39, SD = .756), followed next by fear of negative evaluation (M = 3.25, SD = .874), and, finally, test anxiety (M = 3.04, SD = .600).

Table 1 shows the mean and standard deviation for all the items on the general FLCA Scale (33 items). The findings have been arranged in descending order of importance (after responses to negative items were reverse-coded) so as to facilitate their interpretation. Therefore, in all cases, the highest scores represent a high level of FLCA.

Table 1. Participants' level of FLCA

Items: FLCA	**N**	**Mean**	**SD**
10. I worry about the consequences of failing my foreign language class.	75	4.33	1.004
9. I start to panic when I have to speak without preparation in language class.	75	4.03	1.090
1. I never feel quite sure of myself when I am speaking in my foreign language class.	75	3.96	1.058
13. It embarrasses me to volunteer answers in my language class.	75	3.69	1.185
15. I get upset when I don't understand what the teacher is correcting.	75	3.61	1.150
33. I get nervous when the language teacher asks questions which I haven't prepared in advance.	75	3.60	1.151
12. In language class, I can get so nervous I forget things I know.	75	3.59	1.104
7. I keep thinking that the other students are better at languages than I am.	75	3.53	1.212
23. I always feel that the other students speak the foreign language better than I do.	75	3.52	1.143
24. I feel very self-conscious about speaking the foreign language in front of other students.	75	3.49	1.267
27. I get nervous and confused when I am speaking in my language class.	75	3.47	1. 119
18*. I feel confident when I speak in foreign language class.	75	3.45	1.119
11*. I don't understand why some people get so upset over foreign language classes.	75	3.39	1.312
26. I feel more tense and nervous in my language class than in my other classes.	75	3.36	1.382
30. I feel overwhelmed by the number of rules you have to learn to speak a foreign language.	75	3.35	1.180

22*. I don't feel pressure to prepare very well for language class.	75	3.33	1.057
6. During language class, I find myself thinking about things that have nothing to do with the course.	75	3.31	1.134
29. I get nervous when I don't understand every word the language teacher says.	75	3.27	1.308
8*. I am usually at ease during tests in my language class.	75	3.20	1.193
20. I can feel my heart pounding when I'm going to be called on in language class.	75	3.19	1.270
14*. I would not be nervous speaking the foreign language with native speakers.	75	3.16	1.272
2* I don't worry about making mistakes in language class.	75	3.09	1.377
25. Language class moves so quickly I worry about getting left behind.	75	3.05	1.262
3. I tremble when I know that I'm going to be called on in language class.	75	2.99	1.307
31. I am afraid that the other students will laugh at me when I speak the foreign language.	75	2.92	1.402
32*. I would probably feel comfortable around native speakers of the foreign language.	75	2.91	1.093
17. I often feel like not going to my language class.	75	2.89	1.203
16. Even if I am well prepared for language class, I feel anxious about it.	75	2.85	1.249
4. It frightens me when I don't understand what the teacher is saying in the foreign language.	75	2.76	1.403
28*When I am on my way to language class, I feel very sure and relaxed.	75	2.54	1.196
19. I am afraid that my language teacher is ready to correct every mistake I make.	75	2.45	1.349
21. The more I study for a language test, the more confused I get.	75	2.27	1.155

5*It wouldn't bother me at all to take more foreign language classes.	75	1.77	.909
Overall mean = 105.06			

*Reversed values: The highest scores represent at all times a high anxiety level.

As can be seen, three of the 7 highest-scoring items within the general Scale were related to communicative apprehension (items 9, 1, and 15). The other 4 items were associated with fear of negative evaluation (items 13, 33), and test anxiety (item 10 and 12). However, it is noteworthy that the top-ranking scoring item across the three types of anxiety scales (item 10: 'I worry about the consequences of failing my foreign language class', $\bar{x} = 4.33$) is associated with test-anxiety. Therefore, failing the L2 course is considered to be the most anxiety-inducing factor related to FLCA. This is clearly attributable to the pressure most student teachers feel to pass the English language course in order to satisfy the English language requirement (B2 level, CEFR) before graduation. In addition to this, the findings reveal that the main anxiety-provoking factors are associated with communicative apprehension and speaking anxiety: 'I start to panic when I have to speak without preparation in language class' (item 9, $\bar{x} = 4.03$) and 'I never feel quite sure of myself when I am speaking in my foreign language class' (item 1, $\bar{x} = 3.96$). Oral communication skills are also linked to fear of negative evaluation since using the L2 was found to cause embarrassment ('It embarrasses me to volunteer answers in my language class', item 13, $\bar{x} = 3.69$) and affect the self-image of many respondents ('I get nervous when the language teacher asks questions which I have not prepared in advance', item 33, $\bar{x} = 3.60$). These results concur with those of MacIntyre & Gardner (1991) who state that communicative apprehension and fear of negative evaluation are two closely related types of anxiety constructs. In addition to item 10 related to test anxiety, many other respondents also regarded the practice and use of the L2 in class as a permanent test situation (see Gregersen & Horwitz, 2002): 'In language class, I can get so nervous I forget things I know' (item 12, $\bar{x} = 3.59$) and, consequently, ranked this aspect among the six top scoring items associated to FLCA. As can be observed, all the items of the Scale except for three (items 19, 21, and 5) were rated above 2.5 points on a 5-point scale, which indicates that respondents experienced a considerable degree of FLCA.

Overall, the data show that, in line with previous research findings (Al-Saraj, 2014; Aida, 1994; MacIntyre & Gardner, 1991, 1994; Yang, 2012), most student teachers seem to feel self-conscious and experience negative feelings while communicating and using the L2 in class, which may result in embarrassment and apprehension for most of them. Furthermore, results indicate that participants are really concerned about not being able to fulfil the compulsory English language requirement before

graduation, which may lead them to regard the English language classes as especially anxiety-inducing contexts.

4.3. Are there any differences in the degree of FLCA that prospective primary school teachers experience across different formative itineraries or specialised education programmes?

With regard to the main specialised subjects intended to be chosen by participants (third section of the questionnaire), results indicate that the most popular formative itineraries were Physical Education (27.5%) and Language and Audition (24.6%), followed by Educational Support (20.3%), Foreign Language (English) (14.5%), and, finally, Arts and Music (13.0%).

An independent-sample *t*-test was used to determine whether there were significant differences in FLCA levels between respondents who intended to become future Primary school English teachers and those who were planning to follow other formative itineraries (Physical Education teachers, Language and Audition teachers, Educational Support teachers, and Arts and Music Education teachers). The data indicate there were significant differences ($t = -2.312$, $p = .024 < .05$) between both group of students. Prospective primary school English teachers ($M = 90.44$, $SD = 19.61$) experienced significantly lower levels of FLCA than future teachers specialising in other subjects ($M = 108.04$, $SD = 21.42$). These results came as no surprise since pre-service primary school English teachers are expected to show a higher degree of motivation towards the learning of English than their fellow counterparts.

Interestingly, the independent *t*-test analysis conducted between participants who reported liking English and those who admitted not liking the course subject, that is, those by whom English was considered just a compulsory degree requirement, revealed statistically significant differences ($t = -3.019$, $p = .004 < .05$). Thus, participants who liked English ($M = 100.52$, $SD = 20.51$) were found to be less anxious than students who did not like the subject ($M = 117.67$, $SD = 21.08$). Furthermore, these differences were not only significant for the general anxiety Scale but also for the three different performance anxiety scales: communicative apprehension ($t = -3.191$, $p = .002 < .05$), fear of negative evaluation ($t = -2.204$, $p = .031 < .05$), and test anxiety ($t = -2.715$, $p = .008 < .05$). Therefore, the results of this study confirm that students who are unmotivated and who do not like the English language tend to feel more uncomfortable when using the L2, are likely to be more concerned about negative evaluations of their linguistic performance by others, and appear to be excessively worried about failing in test situations. These findings are consistent with that of other studies which advocate that motivated students experience lower degrees of FLCA than students who lack

motivation (Clément et al., 1994; Liu & Huang, 2011), highlighting the close links between FLCA and motivation to learn the target language.

4.4. *Is FLCA affected by participants' gender, and level of English language proficiency?*

An independent sample *t*-test was run to examine potential differences in terms of FLCA levels as a function of gender. In line with previous research findings (Aida, 1994; Bekleyen, 2009; Park & French, 2013; Öztürk, 2016), the data showed that gender does not have a statistically significant effect on the level of FLCA, although female participants (M = 105.53, SD = 21.87) reported a slightly higher mean level of FLCA than their male counterparts (M = 102.64, SD = 22.80). Therefore, the null hypothesis that there is no difference in the level of FLCA across gender cannot be rejected.

On the contrary, participants' university entrance examination mark in the English Test (ET) was found to be a determining factor affecting FLCA. The one-way analysis of variance (ANOVA) (see Table 2 and 3) revealed that the results of the university entrance ET had a statistically significant effect on the level of FLCA reported by student teachers (F (2, 58) = 9.052, p <.000, p2= .23). Although the sample effect size is relatively small (.23), the data indicated that the observed power, that is, the likelihood that the test will produce a significant result, was very strong (.96). Bonferroni *post hoc* tests (Table 4) showed that there was a significant difference in anxiety levels between respondents who obtained scores between 7 and 8.9 points on a 10-point scale in the university entrance ET (T = 18.03%), and both those who achieved less than 5 points (T = 34.42%, *p* = .001) and those who obtained between 5 and 6.9 points (T = 47.55%, *p* = .001) in this examination. In other words, the higher the mark in the ET, the lower the level of FLA experienced by participants. None of the respondents were awarded more than 8.9 points in the ET. In fact, it is a cause of great concern that a considerable number of student teachers (T = 34.42%) did not pass the ET at A2-B1 level of English proficiency (CEFR) to enter the university. The lack of proficiency in English may lead prospective teachers to experience high levels of FLCA since these students are aware of the fact that they need to demonstrate a B2 level of English proficiency in the four traditional language skills (reading, writing, speaking and listening) at the end of their second year of study in the Teacher Training Faculty (UIB).

Table 2. Test of homogeneity of Variances

Levene Statistic	df1	df2	Sig
2.251	2	58	.114

Table 3. ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6791.890	2	3395.945	9.052	.000
Within Groups	21759.094	58	375.157		
Total	28550.984	60			

Table 4. *Post hoc* test. Bonferroni

Dependent variable: TOTALFLCA

Bonferroni

(I) mark in the SUEE**	(J) mark in the SUEE	Mean Difference (I-J)	Std Error	Sig.	95% Confidence Inetrval	
					Lower Bound	Upper Bound
less than 5	between 5 and 6.9	2.718	5.550	1.000	-10.97	16.40
	between 7 and 8.9	28.840*	7.209	.001	11.07	46.61
between 5 and 6	less than 5	-2.718	5.550	1.000	-16.40	10.97
	between 7 and 8.9	26.122*	6.859	.001	9.21	43.03
between 7 and 8	less than 5	-28.840*	7.209	.001	-46.61	-11.07
	between 5 and 6.9	-26.122*	6.859	.001	-43.03	-9.21

* The mean difference is significant at the 0.05.

**SUEE: Spanish University Entrance Examination

In any event, these results seem to confirm the strong link between proficiency and anxiety, and are congruent with other research findings that suggest that language proficiency in the L2 is a strong predictor of language teaching anxiety (Horwitz et al., 1986; Matsuda & Gobel, 2003; Phillips, 2003; Elkhafaifi, 2005; Kongchan & Wareesiri, 2008; Gagliardi & Maley, 2010).

4. 5. *What are the main sources of FLCA reported by student teachers when using the English language?*

Finally, participants were asked to specify their degree of confidence while using the target language in class. The data reveal that the majority of student teachers (T = 55, 73.3%) admitted feeling nervous or anxious when doing oral communication tasks in the context of the L2 classroom, as opposed to a small minority of respondents (T= 19, 25.3%) who reported feeling confident. Only 1 participant did not answer this question (1.3%). Furthermore, a *t*-test analysis was carried out to determine the relationship between degree of confidence and FLCA. The findings indicate that there were statistically significant differences between both group of participants ($t = -4.840$, $p = .000 < .05$). Thus, confident students ($M = 86.12$, $SD = 16.03$) were found to suffer from lower levels of FLCA than nervous or anxious students ($M = 111.86$, $SD = 19.80$).

In order to obtain a deeper understanding of the data, respondents were also asked to provide the main causes of their anxiety during oral communication tasks (Table 5). The main categories identified (see Nunan, 1992) were the following:

Table 5. Sources of speaking anxiety

Responses	Frequency	Percentage %
- Little practice in speaking tasks.	3	5.27
- Lack of vocabulary and grammar knowledge.	5	8.77
- Lack of English proficiency.	25	43.85
- Fear of committing mistakes.	7	12.28
- Poor pronunciation.	4	7.02
- Inadequacy of previous education.	1	1.75
- Self-consciousness, fear of being laughed at.	3	5.27
- Test anxiety.	2	3.51
- Blank-mind anxiety when being forced to speak.	5	8.77
- Costly registration fees.	2	3.51
Total	**57**	**100**

As can be seen, the majority of respondents who admitted feeling anxious or nervous while using the L2 in class (T= 55, 73.3%) thought their level of anxiety stemmed mainly from their overall lack of proficiency in the target language (see also Matsuda & Gobel, 2004; Tanaka & Ellis, 2003; Cheng et al., 1999; Clément et al., 1994). More specifically, pre-service teachers emphasised their limited knowledge of grammar, vocabulary and pronunciation, which led them to feel too nervous and self-conscious when having to express themselves in the L2. In fact, some authors (Horwitz, 1996; Tum, 2010, 2012) point out that a limited and insufficient command of the English language is likely to cause feelings of linguistic apprehension and insecurity, which may lead L2 teachers to develop chronic feelings of anxiety. Therefore, the findings of this study concur with that of other researchers who highlight the close relationship between language teaching anxiety, self-confidence, and L2 proficiency (Yashima, 2002; Matsuda & Gobel, 2004; Kongchan & Wareesiri, 2008; Koul et al., 2009; Gagliardi & Maley, 2010; Mak, 2011; Aydin, 2016).

Other causes of anxiety mentioned by participants were: fear of committing mistakes, lack of language skills to convey their intended messages, and fear of being an object of laughter or being ridiculed by their fellow mates, which is clearly associated with fear of negative evaluation as an anxiety construct. Further sources of anxiety highlighted by respondents were related to instructional methodologies, such as lack of sufficient practice in oral tasks or inadequacy of previous education. Participants' personal characteristics, such as fear of going blank when being forced to speak were also reported, suggesting that speaking in public and being put on the spot are considered major anxiety-inducing factors for L2 speakers (Young, 1990; Ewald, 2007). Finally, language tests, as well as the costly registration fees involved when failing the language course were also found to contribute to a lesser extent to potential causes of language anxiety.

5. Conclusion

The main purpose of this study was to explore FLCA experienced by prospective primary school teachers in relation to the teaching and learning of English as a foreign language. In accordance with previous research findings, this study provides evidence that the majority of prospective teachers (T = 63.2%) experience moderate levels of FLCA (Liu, 2006; Arnáiz & Guillén, 2012; Tum, 2012; Elaldi, 2016; Öztürk, 2016). However, a considerable number of student teachers (22.1%) also admit feeling highly anxious in the English classroom. The data revealed that the main source of FLCA is associated with communicative apprehension, followed next by fear of negative evaluation and, finally, test anxiety. Thus, student teachers report feeling self-conscious and uneasy when using the L2 in class. This seems to be closely related

to the excessive concern participants feel about being evaluated negatively by others, which has detrimental effects on the practice and development of their communicative skills. These findings are consistent with those of other studies (MacIntyre & Gardner, 1991) which point to the strong relationship between these two anxiety constructs: communicative apprehension and fear of negative evaluation. However, it is worth noting that the data show that failing the English language course is the most anxiety-inducing factor across the three main types of anxiety scales. This is clearly attributable to the pressure prospective teachers feel to pass the language course in order to fulfill the compulsory graduation requirement in English at B2 level (CEFR) for all undergraduates, regardless of their chosen specialized subjects (Physical Education, Tutorial Action, English Language teaching, etc.).

The data from this study also confirmed other research findings that suggest close links between FLCA and motivation (Liu & Huang, 2011; Clément et al. 1994). Thus, participants who chose to specialize in English, and who liked the English language course, reported experiencing statistically significant lower levels of FLCA across the three performance scales: communicative apprehension, fear of negative evaluation, and test anxiety. On the contrary, no statistically significant differences were found between FLCA and gender (see also Aida, 1994; Bekleyen, 2009; Park &French, 2013; Öztürk, 2016), although female participants exhibited a slightly higher level of FLCA than their male counterparts. The findings also revealed that the university entrance examination mark in the English Test (ET) was found to be a determining factor affecting FLCA. ANOVA results showed statistically significant differences in anxiety levels between both those respondents who obtained scores between 7 and 8.9 points on a 10-point scale, and those who were awarded less than 5 points and between 5 and 6.9 points in this latter examination. This indicates that FLA tends to increase as L2 proficiency decreases. The fact that a considerable number of student teachers (34.42%) did not pass the ET at A1-B1 English level (CEFR) in the university entrance examination should also be a cause of great concern, since these students will have to demonstrate a B2 level of English proficiency in the four traditional language skills (reading, writing, speaking and listening) at the end of their second year of study in the Teacher Training Faculty (UIB). This may constitute a major cause of FLCA since the results of this study show that language proficiency in the target language is a significant predictor of prospective teachers' anxiety (see also Horwitz et al., 1986; Matsuda & Gobel, 2003; Phillips, 2003; Elkhafaifi, 2005; Kongchan & Wareesiri, 2008; Koul et al., 2009; Gagliardi & Maley, 2010; Mak, 2011). Likewise, the findings suggest that confidence is a major variable affecting FLCA. Thus, confident students experienced statistically lower levels of FLCA than nervous or anxious students. The data also indicated that the primary source of speaking anxiety was mainly related to participants' limited command of the English language, with special emphasis on

particular areas such as grammar, vocabulary and pronunciation. According to some authors (Horwitz, 1996; Kim & Kim, 2004), feelings of linguistic insecurity in the L2 and lack of confidence may negatively affect L2 teachers' ability to interact in the target language. These results are also consistent with that of other researchers who highlight close links between language teaching anxiety, self-confidence, and L2 proficiency (Clément et al., 1994; Cheng et al., 1999; Tanaka & Ellis, 2003; Kim & Kim, 2004; Matsuda & Gobel, 2004). Therefore, as Horwitz (1996: 370) advocates: "we must be supportive of prospective foreign language teachers as they prepare to enter the profession; we do not want to perpetuate feelings of anxiety in future generations of language learners and teachers".

Overall, the findings of this study highlight the importance of FLCA in relation to L2 teaching and learning. Since FLCA has clearly shown to have adverse effects on L2 achievement as well as on L2 teachers' instructional practices and personal well-being (Horwitz, 1996; Kim & Kim, 2004; Bekleyen, 2009; Tum, 2012), all necessary steps should be taken to ensure the promotion of more supportive and successful learning environments that may help student teachers to increase their motivation, improve their L2 proficiency and enhance their self-confidence in the use and development of the L2 in order to relieve the debilitating effects of FLA.

References

Aida, Y. (1994) Examination of Horwitz, Horwitz & Cope's construct of Foreign Language Anxiety: The case of students of Japanese. *The Modern Language Journal* 78(2): 155– 168.

Al-Saraj, T. M. (2014) Revisiting the foreign language classroom anxiety scale (FLCAS): The anxiety of female English language learners in Saudi Arabia. *L2 Journal* 6: 50– 76.

Aydin, S. (2016) A qualitative research on foreign language teaching anxiety. *The Qualitative Report* 21(4): 629 – 642.

Amengual-Pizarro, M. (2013) Primary education degrees in Spain: Do they fulfill the linguistic and pedagogic needs of future teachers? *Vigo International Journal of Applied Linguistics* (VIAL) 10: 9– 27.

Arnaiz, P. and Guillén, F. (2012) Foreign language anxiety in a Spanish university setting: Interpersonal differences. *Revista de Psicodidáctica* 17(1): 5– 26.

_____ . (2013) Anxiety in Spanish EFL students in different university degree programmes. *Anales de Psicología* 29(2): 335–344.

Bekleyen, N. (2009) Helping teachers become better English students: Causes, effects, and coping strategies for foreign language listening anxiety. *System* 37: 664–675.

Casado, M. A. & Dereshiwsky, M. I. (2001) Foreign language anxiety of university students. *College Student Journal*, 35(4): 539-551.

Clément, R., Dörnyei, Z. & Noels, K. (1994) Motivation, self-confidence and group cohesion in the foreign language classroom. *Language Learning* 44: 417-448.

Cheng, Y, S., Horwitz, E. K. & Shallert, D, L. (1999) Language anxiety: Differentiating writing and speaking components. *Language Learning* 49: 417–446.

Council of Europe. (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Criado, R. & Mengual, Y. (2017) Anxiety and EFL speaking in Spanish compulsory and non-compulsory secondary education: A mixed method study. *Miscelánea: A Journal of English and American studies 55*: 13-35.

Dewaele, J-M. (2005). Investigating the psychological and emotional dimensions in instructed language learning: obstacles and possibilities. *The Modern Language Journal 89(3)*: 367– 380.

Dörnyei, Z. (2005) *The psychology of the language learner: Individual differences in second language acquisition*, Mahwah, NJ: Lawrence Erlbaum Associates.

Elaldi, S. (2016) Foreign language anxiety of students studying English language and literature: A sample from Turkey, *Educational Research and Reviews 11(2)*: 219– 228.

Elkhafaifi, H. (2005) Listening comprehension and anxiety in the Arabic language classroom. *Modern Language Journal* 89(2): 206 – 220.

Ewald, J. D. (2007) Foreign language learning anxiety in upper-level classes: Involving students as researchers. *Foreign Language Annals* 40(1): 122–142.

Gagliardi, C. & Maley, A. (2010) *EIL, EFL, Global English: Teaching and learning issues. Linguistic insights*. Switzerland: Bern.

Gregersen, T. S & Horwitz, E.K. (2002) Language learning and perfectionism: Anxious and non-anxious language learners' reactions to their own oral performance. *The Modern Language Journal 86(4)*: 562 – 570.

Gregersen, T. S. (2003) To err is human: A reminder to teachers of language-anxious students. *Foreign Language Annals 36(1)*: 25 – 32.

_____ . (2007) Breaking the code of silence: A study of teachers' nonverbal decoding accuracy of foreign language anxiety. *Language Teaching Research 1(2)*: 209–221.

Horwitz, E. K. (1996) Even teachers get the blues: Recognizing and alleviating language teachers' feeling of foreign language anxiety. *Foreign Language Annals* 29(3): 365-372.

_____ . (2000) It ain't over 'till it's over: On foreign language anxiety, first language deficits, and the confounding of variables. *The Modern Language Journal 84*: 256 – 259.

_____ . (2001) Language anxiety and achievement. *Annual Review of Applied Linguistics 21*: 112– 126.

Horwitz, E. K. Horwitz, M, B. & Cope, J. (1986) Foreign language classroom anxiety. *The Modern Language Journal 70*: 125– 132.

Horwitz, E, K. & Young, D.J. (eds). (1991) *Language Anxiety: From theory and practice to classroom implications*. Englewood Cliffs, NJ: Prentice-Hall,

İpek, J. H. 2007. *Foreign language teaching anxiety*. Eski□ehir: T.C. Anadolu Üniversitesi Yayınları, No: 1730.

Kim, S. Y. (2009) Questioning the stability of foreign language classroom anxiety and motivation across different classroom contexts. *Foreign Language Annals*, 42(1): 138– 157.

Kim, S.Y. & Kim, J.H. (2004). When the learner becomes a teacher: Foreign language teaching anxiety as an occupational hazar. *English Teaching 59(1):* 165 – 186.

Kongchan, C. &. Wareesiri, W. (2008) Teacher's anxiety using L2 in EFL classrooms. *Proceedings of the 16th Annual Korea TESOL International Conference: Responding to a changing* world. 2008. Seoul, Korea. 145 –149.

Koul, R., Roy, L., Kaewkuekool, S. & Ploisawaschai, S. (2009) Multiple goal orientations and foreign language anxiety. *System 37*: 676– 688.

Krashen, S. D. (1985) *The input hypothesis: Issues and implications.* New York: Longman.

Liu, H. J. (2013) Effects of Foreign Language Anxiety and Perceived Competence on Learning Strategy Use. *International Journal of English Linguistics 3(3):* 76– 87.

Liu, M. (2006) *Anxiety in Chinese EFL students at different proficiency levels. System 34(3):* 301– 316.

Liu, M. & Huang, W.. (2011) An Exploration of Foreign Language Anxiety and English Learning Motivation. *Education Research International*, vol. 2011, Article ID 493167, 8 pages.

MacIntyre, P. D. (1999) Language anxiety: A review of the research for language teachers. In D. J. Young (ed.) *Affect in foreign language and second language learning: A*

practical guide to creating a low-anxiety classroom atmosphere. Boston: McGraw-Hill. 24– 45.

MacIntyre, P. D. & Gardner, R. (1991) Methods and results in the study of anxiety and language learning: A review of literature. *Language Learning* 41: 85– 117.

_____ . (1994) The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning 44*: 283– 305.

Mak, B. (2011) An exploration of speaking-in-class anxiety with Chinese ESL learners".*System*, *39(2)*: 202– 214.

Martínez Agudo, J. (2013) An investigation into Spanish EFL learner's anxiety. *RBLA, Belo Horizonte*, 13(3): 829-851.

Matsuda, S. & Gobel, P. (2004) Anxiety and predictors of performance in the foreign language classroom. *System 32 (1)*: 21– 36.

Nunan, D. (1992) *Research methods in language learning*, Cambridge: Cambridge University Press.

Onwuegbuzie, A. J., Bailey, P. &. Daley,C.E. (2000) Cognitive, affective, personality, and demographic predictors of foreign-language achievement. *The Journal of Educational Research 94*: 3– 15.

Ortega-Cebreros, A. (2003). Measuring language anxiety perceived by Spanish university students of English. Bells: Barcelona English language and literature studies, 12, Retrieved from http://www,raco,cat/index,php/bells/article/viewArticle/82928/0/

Öztürk, G. (2016) Foreign language teaching anxiety among non-native teachers of English: A sample from Turkey. *Sakarya University Journal of Education 6 (3)*: 54 – 70.

Park, G-P. & French, B. F. (2013) Gender differences in the Foreign Language Classroom Anxiety Scale. *System*, 41(2): 462-471.

Pavlenko, A. (ed.).(2011) *Thinking and speaking in two languages*. Clevedon, UK: Multilingual Matters.

Pérez-Paredes, P. & Martínez-Sánchez, F. (2000-2001) A Spanish version of the foreign language classroom anxiety scale: Revisiting Aida's factor analysis. *Revista Española de Lingüística Aplicada RESLA 14*: 337– 352.

Phillips, E. M. (1992) The effects of language anxiety on students' oral test performance and attitudes. *The Modern Language Journal*, 76: 14– 26.

Phillips, J. (2003) Foreign languages, literatures, cultures: Moving toward unity. *ADFL Bulletin 35 (1)*: 15 –19.

Rodríguez, M. & Abreu, O. (2003) The stability of general foreign language classroom anxiety across English and French. *The Modern Language Journal*, *87 (3)*: 365 – 374.

Sheen, Y. (2008) Recasts, language anxiety, modified output, and L2 learning. *Language Learning 58 (4)*: 835–874.

Scovel, T. (2001) *Learning new languages: A guide to second language acquisition*. Boston: Heinle & Heinle.

Spielman, G. & Radnofsky, M. (2001) Learning language under tension: New directions from a qualitative study. *The Modern Language Journal*, 85: 259– 278.

Tanaka, K. & Ellis, R. (2003) Study abroad, language proficiency, and learner beliefs about language learning *JALT Journal 25*: 63–85.

Tum, D. O. (2010) *A study of non-native teachers' and student teachers' feelings of foreign language teaching anxiety* (Unpublished Master's Thesis); The University of Texas, Austin.

_____ . (2012) Feelings of language anxiety amongst non-native student teachers. *Procedia Social and Behavioural Sciences 47*: 2055 – 2059,

Yang, H.C. (2012) Language Anxiety: From the classroom to the community. *English Teaching and Learning 36 (4)*: 1– 28.

Yashima, T. (2002) Willingness to communicate in a second language: The Japanese EFL context. *The Modern Language Journal* 86 (1), 54– 66.

Young, D. J. (1990) An investigation of students' perspectives on anxiety and speaking *Foreign Language Annals 23*: 539–553.

_____ . (1992) Language anxiety from the foreign language specialist's perspective: Interviews with Krashen, Omaggio Hadley, Terrell, and Rardin. *Foreign Language Annals 19*: 439 – 445.

Williams, L. S. (1991) The effects of a comprehensive teaching assistant training program on teaching anxiety and effectiveness. *Research in Higher Education* 32 (5): 585 – 598.

Williams, K. E. & Andrade, M.R. (2008) Foreign language learning anxiety in Japanese EFL university classes: Causes, coping, and locus of control. *Foreign Language Teaching: 5(2)*: 181– 191.

Woodrow, L. (2006) "Anxiety and speaking English as a second language". *RELC Journal 37(3),* 308– 328.

Zheng, Y. (2008). Anxiety and Second/Foreign language learning revisited. *Canadian Journal for New Scholars in Education 1(1)*: 1 – 12.

Readability indices for the assessment of textbooks: a feasibility study in the context of EFL

Pascual Cantos Gómez
Universidad de Murcia, Spain
pcantos@um.es

Ángela Almela Sánchez-Lafuente
Universidad de Murcia, Spain
angelalm@um.es

Abstract

Readability indices have been widely used in order to measure textual difficulty. They can be useful for the automatic classification of texts, especially in language teaching. Among other applications, they allow for the previous determination of the difficulty level of texts without the need of reading them through. The aim of this research is twofold: first, to examine the degree of accuracy of the six most commonly used readability indices, and second, to present a new optimized measure. The main problem is that these readability indices may offer disparity, and this is precisely what has motivated our attempt to unite their potential. A discriminant analysis of all the variables under examination has enabled the creation of a much more precise model, improving the previous best results by 15%. Furthermore, errors and disparities in the difficulty level of the analyzed texts have been detected.

Keywords: Readability indices, text difficulty, EFL, EFL textbook, automatic classification of texts.

Resumen

Los índices de legibilidad se han utilizado de forma extensiva para determinar la dificultad textual, y pueden resultar muy útiles para la clasificación automática de textos, en especial en el ámbito de la enseñanza de lenguas. Entre otras de sus aplicaciones, está la de poder determinar la dificultad de texto sin necesidad de leerlo previamente. El objetivo de estudio es doble: por un lado, analizar el grado de precisión de los seis índices de legibilidad más utilizados, y por otro lado, partiendo de estos datos, intentar diseñar una nueva medida de legibilidad optimizada. El principal

problema es que estos índices pueden ofrecer disparidad, y es precisamente lo que ha motivado nuestro intento de unificar su potencial. Un análisis discriminante de todas las variables examinadas ha permitido la creación de un modelo mucho más preciso, mejorando los resultados previos en un 15%. Además de ello, es importante destacar que se han detectado errores y disparidades en el nivel de dificultad de los textos analizados.

Palabras clave: índices de legibilidad, dificultad textual, inglés como lengua extranjera, libro de texto de inglés, clasificación textual automática.

1. Introduction: formalizing text difficulty by virtue of readablility indices

Readability indices allow measuring how difficult it is to read a text based on its properties, by using constructs known to reflect complexity, such as average sentence length and number of complex words (Fry, 1968; Ash & Edgell, 1975). In the 1950s, these readability indices became increasingly popular, and researchers in the field devoted great effort to devising a substantial number of new formulae, since they can be useful for the automatic classification of texts, especially within language teaching.

Among other applications, readability indices allow for the previous determination of the difficulty level of texts without the need of reading them through. This is precisely what distinguishes readability formulae from comprehensibility tests, such as cloze tests: the former are determined only by the text itself, offering a value which indicates the complexity of the text only on the basis of quantitative elements, while the latter, first described by Taylor (1953), measures the comprehensibility of a text, that is to say, how understandable a text is to an actual reader. In other words, cloze tests give an actual measure of comprehension while readability formulae make a prediction. Precisely for this reason, even though they have been important in traditional readability research and readability formulae have been based on their results, comprehensibility tests have not been used in the present study, mainly quantitative in nature.

As this study is not intended to provide an extensive review of all the readability formulae, only a brief overview and description of the most commonly used readability indices is offered below.

1.1. Traditional Approaches to Readability

First of all, the Flesch/Flesch–Kincaid readability tests include two indices: the *Flesch Reading Easiness Score* and the *Flesch–Kincaid Grade Level*. The first system was devised by Rudolf Flesch in 1948. After several attempts at simplification (Farr, Jenkins, & Paterson, 1951; Kincaid, Fishburne, Rogers, & Chissom, 1975), this is the resulting formula:

$$FRE = 206.835 - 1.015 * \left(\frac{total\_words}{total\_sentences} \right) - 84.6 * \left(\frac{total\_syllables}{total\_words} \right)$$

In 1976, a revision commissioned by the U.S. Navy resulted in a modification of this index to generate a grade-level score, enabling the translation from the 0–100 score to a U.S. grade level. Nowadays, the ensuing formula is known as the *Flesch–Kincaid Grade Level*:

$$F\_KGL = 0.39 * \left(\frac{total\_words}{total\_sentences} \right) + 11.8 * \left(\frac{total\_syllables}{total\_words} \right) - 15.59$$

As can be seen, it uses the same core measures as the Reading Easiness test, namely word length and sentence length. However, the weighting factors are different, and the results of the two tests hence correlate inversely. In this way, a text with a relatively high score on the former test normally achieves a lower score on the latter.

A further index whose score corresponds to U.S. grade level is the *Gunning Fog Index*, or simply *Fog Index*. It was developed by R. Gunning (1952), becoming particularly popular owing to its easy calculation without a calculator (DuBay, 2004). *GFI* gets its index from mean sentence length (in words) and average number of complex words (words with three and more syllables):

$$GFI = 0.4 * \left(\frac{words}{sentence} \right) + 100 * \left(\frac{complex\_words}{words} \right)$$

Subsequently, *Automated Readability Index (ARI)* was worked out by Smith and Senter (1967) for the U.S. Army, and its validity on technical materials was proved by Smith and Kincaid (1970). The formula uses mean word length (in characters) and mean sentence length (in words):

$$ARI = 4.71 * \left(\frac{characters}{word} \right) + 0.5 * \left(\frac{words}{sentences} \right) - 21.43$$

In 1969, G. H. McLaughlin published *SMOG (Simple Measure of Gobbledygook)* in an attempt to make *Gunning Fog Index* calculation even easier. Indeed, in his work the author describes it as "laughably simple" (McLaughlin, 1969, p. 639). It is based upon the conviction that word length and sentence length are to be multiplied rather than added. The formula used at present is the following one:

$$SMOG\_grade = 1.043\sqrt{30 \times \frac{number\_of\_polysyllables}{number\_of\_sentences}} + 3.1291$$

where polysyllable count refers to the number of words of more than two syllables. The resulting score corresponds to the years of education needed to thoroughly understand a given piece of writing.

Finally, the *Coleman-Liau Index* was devised by Coleman and Liau (1975). Like the *ARI*, this measure relies on characters instead of syllables per word, which, as commented on above, is not the trend in readability indices. A further point of similarity between the *ARI* and the *CLI* which is also shared by the Flesch–Kincaid readability tests and the *GFI* is that the ensuing score stands for U.S. grade level. The *CLI* is calculated with the following formula:

$$C\_LI = 5.89 * \left(\frac{characters}{words}\right) + 29.5 * \left(\frac{sentences}{words}\right) - 15.8$$

1.2. Current Research in Readability

One of the main criticisms of the features used for the calculation of traditional readability indices is that they are considered to be linguistically shallow. However, as DuBay (2004) puts it, they are surprisingly effective and widely used at the present moment. Some attempts to combine classical features with other linguistic components for the prediction of text complexity have been recently made. Such is the case of Vajjala and Meurers (2012; 2013; 2014a; 2014b), Crossley, Yang, and McNamara (2014), Flor and Beigman (2014), and Fitzgerald et al. (2015), among others, who take into account language-specific morphological features or the quantification of coherence and cohesion in a text.

Some researchers have tried to validate traditional readability indices for EFL use, like Brown (1998) and Greenfield (1999). The former examined their performance administering cloze tests to 2,300 Japanese learners of EFL and comparing the results with scores predicted by traditional readability indices. Greenfield measured the performance of 200 Japanese college students on cloze tests, this time on a selection of academic passages. Interestingly enough, these two studies yielded contradictory

results: while Greenfield (1999) found traditional formulae to be predictive of reading difficulty, this was not the case of Brown (1998). As Crossley, Allen, and McNamara (2011: 87) put it,

Greenfield (2004) argued that Brown's (1998) passage set was not sufficiently variable in difficulty and too difficult overall to provide a robust passage set for L2 learners. Overall, these studies offer some evidence that classic readability measures discriminate reading difficulty reasonably well for L2 students, but are limited to the appropriate academic texts for which they were designed and do not reach the level of accuracy achieved in L1 cross-validation studies (Greenfield, 1999).

Along these lines, Crossley et al. (2011) compared the classification potential of some of the traditional readability indices mentioned above to more modern readability formulae based on psycholinguistic and cognitive accounts of text processing in discriminating between levels of L2 reading texts, exploring which readability index best classifies text level. However, to our knowledge, no study has compared the performance of the whole set of traditional readability indices with the further purpose of optimizing the results.

No doubt, the level of usage of readability formulae in educational contexts has diminished hugely; yet they are still used heavily to judge the readability of medical patient education materials (e.g. Freda, 2005; Cronin, O'Hanlon, & O'Connor, 2011). However, the main critique of the use of these formulae is limited to the observation that there is no consensus as to which readability formula is best suited for assessing patient education materials. Guo, Zhang, and Zhai (2011) argued that it is preferable to use more than one readability method to improve the validity of the results. Thus, although they have their limitations, such as overemphasis on observable character/word counts, morphological, syllabic features, etc., they are becoming more popular than ever (see Guo, Zhang & Zhai, 2011: 103). It appears that, despite the critiques, readability formulae are still perceived to have a useful function in a number of fields. It was partly to re-examine this functionality that the present study was carried out.

The main trouble with using readability indices is their disparity, and this is precisely what has motivated this paper: attempting to unite their potential. It is certainly true that the limitations of these indices have provoked much discussion and debate, and that in the last decades of the 20th Century there was serious criticism on their extensive use in areas such as law, journalism or health care. Some representative instances of this scholarly controversy are Maxwell (1978) and Connaster (1999), who offered some reasonable alternatives to readability indices like usability testing. Nevertheless, as DuBay (2004:3) puts it, "although the alternatives are useful and even

necessary, they fail to do what the formulas do: provide an objective prediction of text difficulty".

2. Research goal

The aim of this investigation is twofold: first, to examine the accuracy of six of the most commonly used traditional readability indices: *Flesch Reading Ease*, *Flesch–Kincaid Grade Level*, *Gunning Fog, Automated Readability Index*, SMOG, and *Coleman-Liau*; and second, by means of the data obtained, to present a new optimized measure using *Discriminant Function Analysis*.

These six formulae have been chosen because they represent the traditional approach whose performance the present authors aimed to test on EFL materials. Readability indices like the Lorges and Dale-Chall formulae have been excluded from this study because they do not only use quantitative parameters such as average sentence length and number of different words, but also lists of the most common words in English, mainly subsets of the *Dale list of 3000* (Dale & Chall, 1948). Such parameters would entail an external element, and we were mainly interested in the combination of parameters which could be calculated from the text itself.

Although some comparative studies on readability indices (e.g. Crossley et al., 2011) suggest that the *Flesch–Kincaid Grade Level* index is a revision of the *Flesch Reading Ease* index in order to ease its interpretation, we have decided to keep both in our study as intermediate tests obtained spoke against this observation. Partial correlation depending on the linguistic level was only significant for B1 texts (0.99), but not for A2.1 (0.14), A2.2 (0.02) or B2 (-0.36); and this is also applicable to the overall correlation (0.14).

3. Methodology

3.1. Task and Procedures

In order to test the accuracy of the six readability indices mentioned above, the indices of 20 already graded texts have been calculated, five texts for each linguistic level from the coursebook series *Innovations* (Dellar & Walkley, 2005a; Dellar & Walkley, 2005b; Dellar, Walkley, & Hocking, 2004; Dellar, Hocking, & Walkley, 2004). *Innovations* is a five-level general English course for foreign students. For this research, we have only taken the first four books and randomly extracted five text samples for each linguistic level (A2.1, A2.2, B1 and B2).

The randomly chosen texts for each linguistic level have been arranged and analysed according to their chronological order in each of the textbooks they belong to. That is, the elementary/A2.1 text with OD-code=1 occurs in the textbook previous to the one with OD-code=2, etc. (see Table 1).

3.2. Data Analysis

Before calculating the six readability indices, we first obtained the essential quantitative counts for each text, necessary for the various readability indices calculations: number of characters, sentence count, number of complex words (word of more than two syllables), and syllable count. Except for the latter, all these parameters were calculated with *WordSmith Tools 6.0*. As for syllable count, a reliable piece of freeware has been used: *WordCalc*.

In addition, each text sample was typified with a reading-order difficulty code (*OD-Code*: 1 to 20), according to its occurrence sequence in the textbooks and its corresponding linguistic level code (LL-Code: 1 = elementary/A2.1, 2 = pre-intermediate/A2.2, 3 = intermediate/B1, and 4 = upper-intermediate/B2). The preliminary data of all 20 texts are given in Table 1 below. Intuitively, the order of the text samples in Table 1 corresponds to its sequence of appearance in the various textbooks. Therefore, we might assume that text with OD-code = 1 and LL-code = 1 is, in principle, easier to read than text with OD-code = 5 and LL-code = 1.

Table 1. Data summary

LL	OD-Code	LL-Code	Tokens	Characters	Sentences	Syllables	Complex words
A2.1	1	1	289	1161	37	312	5
A2.1	2	1	278	1112	25	330	7
A2.1	3	1	322	1426	29	399	12
A2.1	4	1	233	1014	41	270	4
A2.1	5	1	268	1104	21	320	3
A2.2	6	2	306	1174	24	347	12
A2.2	7	2	564	2089	43	608	6
A2.2	8	2	444	1772	27	482	8
A2.2	9	2	608	2453	44	676	15
A2.2	10	2	661	3062	44	854	32
B1	11	3	543	2249	39	631	16
B1	12	3	648	2771	41	773	16
B1	13	3	291	1196	19	347	5
B1	14	3	606	2548	29	755	28
B1	15	3	506	2267	28	653	17
B2	16	4	408	1930	29	561	25
B2	17	4	383	1774	30	508	22
B2	18	4	596	2661	32	746	27
B2	19	4	555	2665	26	744	34
B2	20	4	564	2695	23	782	28

Next, all readability indices for each text sample were calculated (Table 2); and finally all texts were ordered according to the respective readability indices (Table 3).

Table 2. Readability indices

LL	OD-Code	LL-Code	ARI	C-LI	FRE	F-KGL	GFI	SMOG
A2.1	1	1	1.3	11.6	107.5	0.1	4.8	5.2
A2.1	2	1	2.9	10.4	95.1	2.7	6.9	6.1
A2.1	3	1	4.9	12.9	90.7	3.3	8.1	6.8
A2.1	4	1	1.9	15.0	103.0	0.3	3.9	4.9
A2.1	5	1	4.3	10.7	92.8	3.4	6.2	5.2
A2.2	6	2	3.0	9.1	97.9	2.7	9.0	7.1
A2.2	7	2	2.5	8.2	102.3	2.2	6.3	5.2
A2.2	8	2	5.5	9.5	98.3	3.6	8.3	6.2
A2.2	9	2	4.4	10.0	98.7	2.9	7.9	6.4
A2.2	10	2	7.9	13.4	82.2	5.5	10.8	8.0
B1	11	3	5.0	10.7	94.3	3.5	8.5	6.7
B1	12	3	6.6	11.2	89.8	4.6	8.7	6.6
B1	13	3	5.5	10.3	90.4	4.4	7.8	6.0
B1	14	3	8.8	10.3	80.2	7.2	12.9	8.7
B1	15	3	8.7	12.2	79.3	6.6	10.5	7.5
B2	16	4	7.8	14.1	76.2	6.1	11.7	8.4
B2	17	4	6.7	13.7	81.6	5.0	10.8	8.0
B2	18	4	8.9	12.0	82.0	6.4	11.9	8.3
B2.	19	4	11.8	13.8	71.7	8.5	14.6	9.6
B2	20	4	13.3	13.5	64.6	10.3	14.7	9.4

Table 3. Texts ordered according to readability ease

LL	OD-Code	LL-Code	ARI	C-LI	FRE	F-KGL	GFI	SMOG
A2.1	1	1	1	11	1	1	2	2
A2.1	2	1	4	7	7	4	5	6
A2.1	3	1	8	14	10	7	8	11
A2.1	4	1	2	20	2	2	1	1
A2.1	5	1	6	9	9	8	3	4
A2.2	6	2	5	2	6	5	12	12
A2.2	7	2	3	1	3	3	4	3
A2.2	8	2	11	3	5	10	9	7
A2.2	9	2	7	4	4	6	7	8
A2.2	10	2	15	15	13	14	14	14
B1	11	3	9	8	8	9	10	10
B1	12	3	12	10	12	12	11	9
B1	13	3	10	5	11	11	6	5
B1	14	3	17	6	16	18	18	18
B1	15	3	16	13	17	17	13	13
B2	16	4	14	19	18	15	16	17
B2	17	4	13	17	15	13	15	15
B2	18	4	18	12	14	16	17	16
B2	19	4	19	18	19	19	19	20
B2	20	4	20	16	20	20	20	19

Furthermore, in order to find out significant discrepancies among the readability indices, they were normalized into *z*-scores (Figure 1 and Table 4). A brief examination of the data reveals that the *CL-I* index (*Coleman-Liau*) is the only readability index that exhibits significant deviations compared to the other five ones: in 5 out of 20 texts the deviation of this index exceeded in more than 2 standard deviation measures (texts 1, 4, 7, 14 and 20). Because of this divergence from the rest of the readability indices, we have decided to discard the *Coleman-Liau* readability index for this research.

Figure 1. Z-score normalization of text readability variables

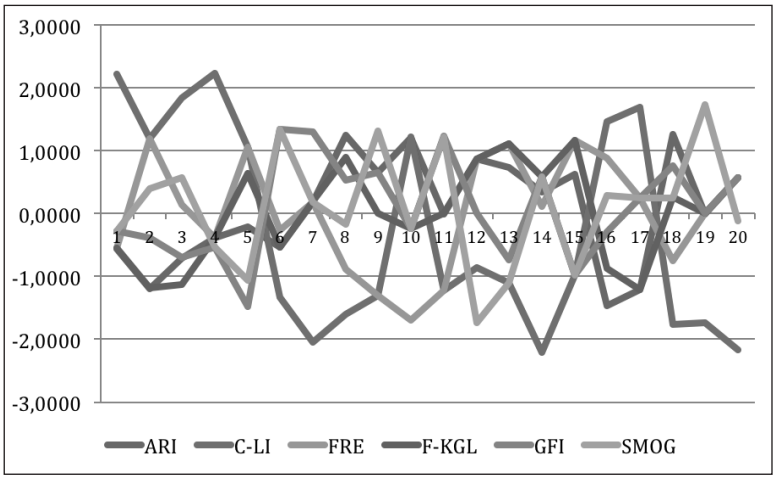


Table 4. Z-score normalization of the ordinal text readability variables

LL	OD-Code	LL-Code	ARI	C-LI	FRE	F-KGL	GFI	SMOG
A2.1	1	1	-0.55	2.21	-0.55	-0.55	-0.27	-0.27
A2.1	2	1	-1.19	1.19	1.19	-1.19	-0.39	0.39
A2.1	3	1	-0.70	1.83	0.14	-1.13	-0.70	0.56
A2.1	4	1	-0.38	2.23	-0.38	-0.38	-0.53	-0.53
A2.1	5	1	-0.21	1.05	1.05	0.63	-1.48	-1.05
A2.2	6	2	-0.53	-1.33	-0.26	-0.53	1.33	1.33
A2.2	7	2	0.18	-2.04	0.18	0.18	1.29	0.18
A2.2	8	2	1.24	-1.59	-0.88	0.88	0.53	-0.17
A2.2	9	2	0.65	-1.30	-1.30	0.00	0.65	1.30
A2.2	10	2	1.21	1.21	-1.69	-0.24	-0.24	-0.24
B1	11	3	0.00	-1.22	-1.22	0.00	1.22	1.22
B1	12	3	0.86	-0.86	0.86	0.86	0.00	-1.73
B1	13	3	0.73	-1.10	1.10	1.10	-0.73	-1.10
B1	14	3	0.34	-2.20	0.11	0.57	0.57	0.57
B1	15	3	0.62	-0.98	1.16	1.16	-0.98	-0.98
B2	16	4	-1.46	1.46	0.87	-0.87	-0.29	0.29
B2	17	4	-1.21	1.69	0.24	-1.21	0.24	0.24
B2	18	4	1.26	-1.76	-0.75	0.25	0.75	0.25
B2	19	4	0.00	-1.73	0.00	0.00	0.00	1.73
B2	20	4	0.56	-2.16	0.56	0.56	0.56	-0.11

Table 5 shows that textbook sample 1 (OD-code =1) is typified by the readability indices with the lowest score (*ARI, FRE* and *F-KGL*) or with the second lowest one (*GFI* and *SMOG*). In contrast, textbook sample 3 is, according to the readability indices, the 8th, 10th, 7th, 8th or 11th highest score. This is a striking case, as this text seems clearly misplaced, though it is placed at the beginning of the A2.1 EFL textbook. According to its RI, this text should not have been placed in the A2.1 book, but in a more advanced level, depending on the readability measures used: pre-intermediate (*ARI, FRE, FKGL* and *GFI*) or even intermediate one (*SMOG*).

Table 5. Text ordered according to readability ease

LL	OD-Code	LL-Code	ARI	FRE	F-KGL	GFI	SMOG	MD
A2.1	1	1	1	1	1	2	2	0.4
A2.1	2	1	4	7	4	5	6	3.2
A2.1	3	1	8	10	7	8	11	5.8
A2.1	4	1	2	2	2	1	1	-2.4
A2.1	5	1	6	9	8	3	4	1
A2.2	6	2	5	6	5	12	12	2
A2.2	7	2	3	3	3	4	3	-3.8
A2.2	8	2	11	5	10	9	7	0.4
A2.2	9	2	7	4	6	7	8	-2.6
A2.2	10	2	15	13	14	14	14	4
B1	11	3	9	8	9	10	10	-1.8
B1	12	3	12	12	12	11	9	-0.8
B1	13	3	10	11	11	6	5	-4.4
B1	14	3	17	16	18	18	18	3.4
B1	15	3	16	17	17	13	13	0.2
B2	16	4	14	18	15	16	17	0
B2	17	4	13	15	13	15	15	-2.8
B2	18	4	18	14	16	17	16	-1.8
B2	19	4	19	19	19	19	20	0.2
B2	20	4	20	20	20	20	19	-0.2

In order to determine the divergences between the textbook placing of the texts and the readability indices, we have calculated the mean divergences (*MD*) of all texts:

$$MD = \left(\frac{\sum RI}{\#RI} \right) - OD\_Code$$

where $\sum RI$ stands for sum of the various readability indices used, *#RI* for the number of readability indices applied and *OD-code* for the reading-order difficulty code within the textbook sequences.

According to the *MD*s, we find four misplaced texts, probably presented to students too early: Text 3: MD 5.8; Text 10: MD 4; Text 14: MD 3.4; and Text 2: MD 3.2.

Similarly, some apparently linguistically less demanding texts are also misplaced, appearing too late in the textbooks: Text 13: MD -4.4; Text 7: MD -3.8; Text 17: MD -2.8; Text 9: MD -2.6; Text 4: MD -2.4.

Data also reveal that some texts seem to have been improperly placed, as their indices are higher/lower for the textbook in which they appear:

- Text 3 – *A2.1*; should be *A2.2*
- Text 14 – *B1*; should be *B2*
- Text 15 – *B1*; should be *B2*
- Text 11 – *B1*; should be *A2.2*
- Text 17 – *B2*; should be *B1*
- Text 13 – *B1*; should be *A2.2*
- Text 7 – *A2.2*; should be *A2.1*

In order to determine the accuracy of the readability indices, we shall first order the texts according to their *Index Means* (*IM*s) and re-typify them as being elementary/A2.1 (*IM*≤5), pre-intermediate/A2.2 (*IM*≥ 5 and ≤10), intermediate/B1 (*IM*≥10 and ≤15) and upper-intermediate/B2 (*IM*≥15). The re-typification (*New LL-Code*) is given in Table 6. *SMOG* and *C-LI* are the least precise ones, although their correlation values are highly significant.

Table 6. Texts re-typified according to IMs

OD-Code	IM	LL	New LL-code
1	1.5	A2.1	1
4	2	A2.1	1
7	2.67	A2.1	1
2	5.33	A2.2	2
3	10	A2.2	2
5	6.5	A2.2	2
6	7.17	A2.2	2
8	7.83	A2.2	2
9	6.5	A2.2	2
11	9.17	A2.2	2
13	8.67	A2.2	2
10	14.5	B1	3
12	11	B1	3
17	14.33	B2	3
14	16.67	B2	4
15	15.17	B2	4
16	16.5	B2	4
18	15.5	B2	4
19	19.33	B2	4
20	19.67	B2	4

Regarding wrong linguistic level assignment, *ARI* and *F-KGL* accounted for five errors; *C-LI* for six errors, although text 13 was two-level wrongly assigned to *B2* instead of *A2.2* (see Table 7); *GFI* for seven errors; *SMOG* for ten errors (and a two-level wrong assignment); and *FRE* for eleven errors.

Table 7. LL-assignment errors

Textbook	New LL-code	ARI	C-LI	FRE	F-KGL	GFI	SMOG
1	1	Correct	Correct	Correct	Correct	Correct	Correct
2	1	Correct	Correct	Correct	Correct	Correct	Correct
3	2	Incor. (1)	Incor. (1)	Incor. (1)	Incor. (1)	Incor. (1)	Incor. (1)
4	1	Correct	Correct	Incor. (1)	Correct	Correct	Incor. (1)
5	2	Correct	Correct	Incor. (1)	Correct	Correct	Correct
6	2	Correct	Correct	Correct	Correct	Incor. (1)	Incor. (1)
7	1	Correct	Correct	Incor. (1)	Correct	Incor. (1)	Incor. (2)
8	2	Incor. (1)	Correct	Incor. (1)	Correct	Incor. (1)	Correct
9	2	Correct	Correct	Correct	Correct	Correct	Correct
10	3	Incor. (1)	Correct	Incor. (1)	Incor. (1)	Incor. (1)	Correct
11	2	Correct	Correct	Incor. (1)	Incor. (1)	Correct	Incor. (1)
12	3	Correct	Correct	Correct	Correct	Correct	Incor. (1)
13	2	Incor. (1)	Incor. (2)	Incor. (1)	Incor. (1)	Incor. (1)	Incor. (1)
14	4	Incor. (1)	Correct	Incor. (1)	Incor. (1)	Incor. (1)	Incor. (1)
15	4	Correct	Incor. (1)	Correct	Correct	Correct	Incor. (1)
16	4	Correct	Incor. (1)	Incor. (1)	Correct	Correct	Correct
17	3	Correct	Incor. (1)	Incor. (1)	Correct	Correct	Incor. (1)
18	4	Correct	Incor. (1)	Correct	Correct	Correct	Correct
19	4	Correct	Correct	Correct	Correct	Correct	Correct
20	4	Correct	Correct	Correct	Correct	Correct	Correct
Total errors		5 (5)	6 (7)	11 (11)	5 (5)	7 (7)	10 (11)

Surprisingly enough, the three readability indices that best adjust to the *New LL-Code* use different measures. As commented on above, *ARI* uses mean word length and mean sentence length, and to obtain the *F-KGL* index, we need mean sentence length and mean syllable per word. On the contrary, *GFI* gets its index from mean sentence length and average number of complex words. In this way, the calculation of the *ARI* and of *CLI* is straightforward; some easy text processing by means of any

standard concordance program will output the information required to calculate this index (e.g. *WordSmith Tools*). Nonetheless, *F-KGL* and *GFI* are more demanding, as we need reliable software syllable counting (i.e. *WordCalc* or *Syllable Counter*). These applications are less consistent and the resulting data might vary significantly.

Regarding complex word count (words with three and more syllables), we performed some preliminary experimenting and evidenced that 95% of all English words with eight or more characters do entail at least three syllables; this is the measure which has been used to calculate the *GFI* index.

4. Modeling a new index

To attempt the modeling of a new readability index able to classify text samples according to reading ease, we shall take:

- The data on the various texts analyzed (Table 1), entailing all the distinct measures required by the individual readability indices examined, and
- The *New LL-Code*, as this is a sort of average measure of all individual readability indices we have considered.

We shall try to model an index by means of *Discriminant Function Analysis* (*DFA*, hereafter). *DFA* is concerned with the problem of assigning individuals, for whom several variables have been measured, to certain groups that have already been identified in the sample. It is used to determine the variables that discriminate between two or more naturally occurring groups (Cantos, 2013). Thus, our aim is not just to measure and model reading ease, but also to look at the dataset that best describes it.

The *DFA*, using all variables (tokens, characters, sentences, syllables and complex words) outputs very promising results: only two errors (see Table 8). One A2.1 text has been assigned to A2.2 (text 2) and a B2 one has been classified as a B1 one (text 15). This gives an overall precision of 90% compared to the best precision scores of two readability indices above (*ARI* and *F-KGL*) of 75%. A further use of *DFA* is that, if it has turned out to be positive, it is possible to generate a predictive discriminant model to classify new cases.

Table 8. Preliminary *DFA*

		Predicted Group Membership				
New LL-code		**A2.1**	**A2.2**	**B1**	**B2**	**Total**
Count	A2.1	3	1	0	0	4
	A2.2	0	7	0	0	7
	B1	0	0	3	0	3
	B2	0	0	1	5	6
%	A2.1	75.0	25.0	0.0	0.0	100.0
	A2.2	0.0	100.0	0.0	0.0	100.0
	B1	0.0	0.0	100.0	0.0	100.0
	B2	0.0	0.0	16.7	83.3	100.0

By means of the *Fisher Coefficients*, we obtain a table (Table 9) with a constant value and a number of coefficients for each of the variables (tokens, characters, sentences, syllables and complex words) with reference to each readability-ease level.

Table 9. Fisher Coefficients

	Readability-ease level			
	A2.1	**A2.2**	**B1**	**B2**
Tokens	-0.14	-0.11	-0.26	-0.26
Characters	-0.06	-0.04	-0.05	-0.05
Sentences	1.36	0.79	1.01	0.63
Syllables	0.35	0.26	0.44	0.46
Complex words	-0.43	-0.20	-0.18	0.00
(Constant)	-21.86	-13.09	-31.81	-31.43

This yields four equations, one for each readability-ease level. To illustrate the potential applicability of the equations above, we can take, for example, a randomly chosen text with tokens = 300; characters = 1,200; sentences = 40; syllables = 400; and complex words = 10, which will be assigned to the readability-ease level with the largest resulting value according to the four functions above. Thus, maximizing the four coefficients we find that this text is most likely to be an A2.1 text, as *Elementary/*

A2.1 is the highest resulting coefficient (44.338); in second place, it would be classified under *Intermediate/B1* (34.239). The least likely group membership would be *Upper-intermediate/B2* (30.672), as the coefficient obtained in the corresponding equation is the lowest one.

5. Conclusions

Readability indices can be useful for the automatic classification of texts, especially within language teaching. Among other applications, they allow for the previous determination of the difficulty level of texts directly extracted from the Internet. The problem is that these readability indices may offer disparity, and this is precisely what has motivated our attempt to unite their potential, utilizing all the variables used by them. A discriminant analysis of all the variables under examination has enabled the creation of a much more precise model, improving the previous best results by 15%. It is also worth noting that errors or disparities in the difficulty level of the analyzed texts have been detected. Specifically, the *DFA* has helped us examine whether the linguistic features contained within the formula were significant predictors of level classification, and what is more, *DFA* has also optimized the predictors by means of re-weighting them (Fisher coefficients), resulting into four new readability indices, one for each LL, with not just new weighting but also a new "combination" of variables.

Our intention is to delve more deeply into the refinement and use of readability indices for tasks such as automatic classification of texts, especially within the area of language teaching, comparing different languages and confirming whether these readability indices offer a similar degree of precision or if they require any adjustment for its calculation as far as variables are concerned.

References

Ash, R. A., & Edgell, S. L. (1975). A note on the readability on the Position Analysis Questionnaire (PAQ). *Journal of Applied Psychology*, *60*(6), 765–766. http://dx.doi.org/10.1037/0021-9010.60.6.765

Brown, J. D. (1998). An EFL readability index. *Journal of the Japan Association for Language Teaching, 20*(2), 7–36.

Cantos, P. (2013). *Statistical Methods in Language and Linguistic Research*. Sheffield, UK: Equinox Publishing.

Coleman, M., & Liau, T. L. (1975). A Computer Readability Formula Designed for Machine Scoring. *Journal of Applied Psychology, 60*(2), 283–284. http://dx.doi.org/10.1037/h0076540

Connaster, B. F. (1999). Last rites for readability formulas in technical communication. *Journal of Technical Writing and Communication, 29*(3): 271–287.

Cronin, M., O'Hanlon, S., & O'Connor, M. (2011). Readability Level of Patient Information Leaflets for Older People. *Irish Journal of Medical Science, 180*(1), 139–142. https://doi.org/10.1007/s11845-010-0624-x

Crossley, S. A., Allen, D. B., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language, 23*(1), 84–101.

Crossley, S. A., Yang, H. S., & McNamara, D. S. (2014). What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language, 26*(1), 92–113.

Cunningham, J. W., & Mesmer, H. A. (2014). Quantitative Measurement of Text Difficulty: What's the Use? *The Elementary School Journal, 115*, 255–269. https://doi.org/10.1086/678292

Dale, E., & Chall, J. S. (1948). A Formula for Predicting Readability. *Educational Research Bulletin*, *27*, 37–53.

Dellar, H., & Walkley, A. (2005). *Innovations Elementary. Coursebook*. Boston, MA: Thomson ELT.

___. *Innovations Pre-Intermediate. Coursebook*. Boston, MA: Thomson ELT.

Dellar, H., Walkley, A., & Hocking, D. (2004). *Innovations Intermediate. Coursebook*. Boston, MA: Thomson ELT.

Dellar, H, Hocking, D. & Walkey, A. (2004). *Innovations Upper Intermediate. Coursebook*. (2nd ed.). Boston, MA: Thomson ELT.

DuBay, W. H. (2004). *The Principles of Readability.* Retrieved May 2, 2017, from http://www.impact-information.com/impactinfo

/readability02.pdf

Farr, J. N., Jenkins, J. J., & Paterson, D. G. (1951). Simplification of the Flesch Reading Ease Formula. *Journal of Applied Psychology, 35*(5), 333–357.

Fitzgerald, J., Elmore, J., Koons, H., Hiebert, E. H., Bowen, K., Sanford, E. E., & Stenner, A. J. (2015). Important text characteristics for early-grades text complexity. *Journal of Educational Psychology, 107*, 4–29. http://dx.doi.org/10.1037/a0037289

Flesh, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221–233.

Flor, M., & Klebanov, B. B. (2014). Associative Lexical Cohesion as a Factor in Text Complexity. *International Journal of Applied Linguistics: Special issue on Recent Advances in Automatic Readability Assessment and Text Simplification, 165*(2), 223–258. http://dx.doi.org/10.1075/itl.165.2.05flo

Freda, M. C. (2005). The Readability of American Academy of Pediatrics Patient Education Brochures. *Journal of Pediatric Health Care, 19*(3), 151–156. https://doi.org/10.1016/j.pedhc.2005.01.013

Fry, E. (1968). A Readability Formula that Saves Time. *Journal of Reading 11*(7), 513–516.

Greenfield, G. (1999). *Classic readability formulas in an EFL context: Are they valid for Japanese speakers?* Doctoral dissertation No. 99–38670. Philadelphia, PA: University Microfilms, Temple University.

Greenfield, J. (2004). Readability formulas for EFL. *Journal of the Japan Association for Language Teaching, 26*, 5–24.

Gunning, R. (1952). *The technique of clear writing*. New York, NY: McGraw-Hill.

Guo, S., Zhang, G., & Zhai, R. (2011). Integrating Readability Index into Twitter Search Engine. *British Journal of Educational Technology, 42*(5), 103–105.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (Automated Readability Index. Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel*. Memphis, TN: US Naval Air Station.

McLaughlin, G. H. (1969). SMOG grading - A new readability formula. *Journal of Reading, 22*, 639–646.

Maxwell, M. (1978). Readability: Have we gone too far? *Journal of Reading, 21*, 525–530.

Nelson, J, Perfetti, C., Liben, D., & Liben, M. (2012). *Measures of text difficulty: Testing their predictive value for grade levels and student performance*. Retrieved from the Council of Chief State School Officers website: https://achievethecore.org/content/upload/nelson_perfetti_liben_measures_of_text_difficulty_research_ela.pdf

Smith, E. A., & Senter, R. J. (1967). *Automated Readability Index*. Retrieved from the Defense Technical Information Center website: http://www.dtic.mil/dtic/tr/fulltext/u2/667273.pdf

Smith, E. A., & Kincaid, P. (1970). Derivation and validation of the Automated Readability Index for use with technical materials. *Human Factors, 12*(1), 457–464.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, *30*, 415–433.

Vajjala, S., & Meurers, D. (2012). On Improving the Accuracy of Readability Classification using Insights from Second Language Acquisition. In J. Tetreault (Ed.), *Proceedings of the 7th Workshop on the Innovative Use of NLP for Building Educational Applications* Quebec, Canada: Association for Computational Linguistics. 163-173.

__. (2013). On the Applicability of Readability Models to Web Texts. In S. Williams, A. Siddharthan, & A. Nenkova (Eds.), *Proceedings of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations*. Sofia, Bulgaria: Association for Computational Linguistics. 59-68.

__. (2014a). Assessing the relative reading level of sentence pairs for text simplification. In I. S. Wintner, S. Goldwater, & S. Riezler (Eds.), *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics* Gothenburg, Sweden: Association for Computational Linguistics. 288-297.

__. (2014b). Readability Assessment for Text Simplification: From Analyzing Documents to Identifying Sentential Simplifications. *International Journal of Applied Linguistics (Special Issue on Current Research in Readability and Text Simplification), 165*(2), 194–222. https://doi.org/10.1075/itl.165.2.04vaj

An n-gram based approach to the automatic classification of schoolchildren's writing

Jordi Cicres
Universitat de Girona, Spain
jordi.cicres@udg.edu

Sheila Queralt
Laboratorio SQ-Lingüistas Forenses, Barcelona, Spain
sheila.queralt@cllicenciats.cat

Abstract

This article focuses on the analysis of schoolchildren's writing (throughout the whole primary school period) using sets of morphological labels (n-grams). We analyzed the sets of bigrams and trigrams from a group of literary texts written by Catalan schoolchildren in order to identify which bigrams and trigrams can help discriminate between texts from the three cycles into which the Spanish primary education system is divided: lower cycle (6- and 7-year-olds), middle cycle (8- and 9-year-olds) and upper cycle (10- and 11-year-olds). The results obtained are close to 70% of correct classifications (77.5% bigrams and 68.6% trigrams), making this technique useful for automatic document classification by age.

Keywords: writing, n-grams, primary school, morphological categories, automatic classification

Resumen

Este artículo trata del análisis de la escritura de los escolares (a lo largo de la educación primaria) utilizando un conjunto de etiquetas morfológicas (n-gramas). Se han analizado un conjunto de bigramas y trigramas de un conjunto de textos literarios escritos por escolares catalanes con el objetivo de identificar qué bigramas y trigramas pueden discriminar los textos según los ciclos en los que se divide la educación primaria en España: el ciclo inicial (6 y 7 años), medio (8 y 9 años) y superior (10 y 11 años). Los resultados muestran cerca del 70% de clasificaciones correctas (el 77,5% en bigramas y el 68,6% en trigramas), lo que permite afirmar que la técnica es útil para la clasificación automática de los documentos según la edad.

Palabras clave: escritura, n-gramas, educación primaria, categorías morfológicas, clasificación automática

1. Introduction

The development of writing competence in the age corresponding to the primary school period (from 6 to 12 years old) is a key factor for both the expression of ideas (Graham, 2006) and cognitive development (Björk & Blomstrand, 2000). Writing is thus an essential tool for learning (Graham and Herbert, 2011). In view of this, the study of texts produced by schoolchildren is highly relevant and justified. Various approaches have been proposed for analyzing texts produced during the school period, including error analysis (Sofkova Hashemi, 2003); the analysis of the main textual properties —i.e. cohesion, coherence and adequacy (Sotomayor, Lucchini, Bedwell, Biedma, Hernández & Molina, 2013)—; the writing and revision processes (Flower & Hayes, 1981; Fitzgerald and Markham, 1987; Camps, 1990; Graham, 2006); and the analysis of the literary formal aspects of texts.

The present study focuses on the analysis of stylometric aspects. From this perspective, the center of attention of stylometry has concentrated, on the one hand, on analyzing the writing style of specific authors (ranging from the most famous controversies over the authorship of Shakespeare plays (Efron & Thisted, 1976; Lowe and Matthews, 1995; Merriam, 1996) to studies on the Federalist Papers (Mosteller & Wallace, 1964; Holmes & Forsyth, 1995; Tweedie, Singh & Holmes, 1996), or on constructing profiles that can help identify the author's gender, dialectal origin, educational level, etc., on the other.

For instance, in order to determine the author's gender in digital texts (specifically in tweets), it has been discovered that, in English, men use more determiners and prepositions, whereas women use more personal pronouns, auxiliary verbs and conjunctions. It has also been observed that women use more emoticons, ellipses (...), words with multiplied vowels (*nooo waaay*), repeated exclamation marks, combined punctuation marks (especially *?* and *!*), the *omg* abbreviation (from *Oh my God*) and onomatopoeic words (*ah, hmm, ugh, grr*), whereas the only common thing among male authors is the frequent use of *yeah* and *yea* (Bamman, Eisenstein & Schnoebelen, 2012).

As a result, several linguistic and computational approaches have been proposed, whose aim is to define a group of variables that can be used to discriminate the authors of texts according to sociolinguistic variables (gender, ethnicity, age, educational status...) or to identify the style of a specific author. For example, Cheng, Chandramouli

& Subbalakshmi (2011) used up to 545 parameters related to psycholinguistic and linguistic preferences according to gender, together with stylometric parameters, including character-based features (such as the ratio of letters, numbers, uppercase characters, spaces or the ratio of special characters in relation to the total number of characters), word-based features (including measures such as mean word-length, lexical richness, long- and short-word ratio, the ratio of hapax legomena and hapax dislegomena, etc.), syntactic features (with variables related to punctuation marks) and, lastly, structural features (with variables such as the number of sentences and paragraphs, mean number of sentences and of words per paragraph, the ratio of sentences starting with upper or lower case, etc.). This set of variables succeeded in correctly classifying texts according to the author's gender in 85.1% of the cases in extensive corpora (the Enron Corpus and the first volume of the Reuters Corpus).

Other techniques to describe the author's style consist in the analysis of n-grams, which are sets of *n* elements appearing together. In different areas of linguistics, in particular in studies on information theory and psycholinguistics (Jurafsky, 2003), learning theories (Anderson, 1982; Newell, 1990) and more recently computational linguistics, these categories have been employed not only for descriptive purposes, but also as classifiers (to classify genres or authors).

Although n-grams are usually based on lexical categories, in this article we concentrate on the use of sets of morphological labels, since several studies have shown their efficiency in describing the style of specific authors or literary genres.

Different studies have focused on extracting the syntactic information of texts. Baayen, van Halteren & Tweedie (1996) were the first to implement n-grams based on syntactic information on authorship attribution analysis using an annotated English corpus. Their study proved that authorship attribution analysis based on syntactic n-grams was more successful than the one based on lexical measures such as vocabulary richness. Stamatatos, Fakotakis & Kokkinakis (2000) later implemented sentence and chunk boundaries in order to discriminate between authors of Modern Greek texts. Their approach therefore used simpler information than the one by Baayen et al. (1996). Hirst and Feiguina (2007) used bigrams of syntactic labels and were able to obtain optimal results in authorship attribution with very short texts (about 200 words long). Other researchers, like Nazar & Sánchez Pol (2007), Spassova & Turell (2007) and Queralt & Turell (2013), obtained very successful results in authorship attribution through the application of a Part-of-Speech (POS) tagger to Spanish texts using bigrams and trigrams.

The use of n-grams has yielded very successful results in the determination of the authorship of written texts within the field of forensic linguistics, since this method

focuses on syntactic structure. Although syntactic variables are more complex and therefore present more obstacles for automatic analysis (compared to more superficial variables like sentence length or the use of punctuation marks, for instance), it is also more difficult for writers to modify them at their will. For this reason, they represent the concept of idiolectal style better than other variables (Turell, 2010; Queralt & Turell, 2013).

Nevertheless, these techniques have not yet been applied to the automatic analysis and classification of school texts. In this paper, the n-gram technique is used to analyse texts produced by primary school students (ranging between 7 and 12 years old) and to classify them according to their authors' ages. It is worth noting that children make hugely significant progress in their acquisition of reading and writing skills during the primary school years. In Spain, most schools initiate the teaching and learning of these skills when students are between 3 and 4 years old (Teberosky, 2001), so that most children have reached the alphabetic phase by the time primary education begins (at age 6). From then on, more intensive writing practices are introduced and the teaching of orthographic rules is initiated.

2. Objectives and hypotheses

The present article deals with the analysis of schoolchildren's writing (throughout the whole primary school period) using sets of morphological labels (n-grams). Our goal is to identify which bigrams and trigrams can help discriminate between texts written by children in each of the 3 cycles into which the Spanish primary education system is divided: lower cycle (6- and 7-year-olds), middle cycle (8- and 9-year-olds) and upper cycle (10- and 11-year-olds). An additional aim is to establish a means of automatically classifying new texts as belonging to one of the 3 cycles of primary school.

Therefore, the following hypotheses are considered:

1. It will be possible to find a combination of bi- and trigrams that characterize the writing style of each of the three cycles of primary education.

2. The combination of bi- and trigrams will allow us to correctly classify new texts, i.e., assign them to their corresponding age group.

3. Methodology

3.1. Corpus

The texts are written in Catalan by children attending school in the town of Balaguer (Catalonia). The native languages of all the participants are Catalan and Spanish.

The corpus used in this study comprises 169 fragments of literary texts in Catalan (a specific version of *Little Red Riding Hood)* written by 7- to 11-year-old children as an activity in their regular classrooms. We have not considered 6- and 7-year-old children because their command of the written language is still insufficient to write a long text, as required by the proposed exercise.

The children did not receive specific instructions on how to perform the writing task other than that they had to explain the *Little Red Riding Hood* story "in their own way". Table 1 shows the distribution of the corpus by the total number of samples in each class. Classes are grouped into cycles and the number of samples is divided by gender. Other measures shown are the mean number of words per text and the standard deviation.

Table 1. Dist,ribution of the corpus.

Cycle	Age	N	Boys	Girls	Average Length of Words	SD
Lower	7-8	42	24	18	157.90	60.779
Middle	8-9	39	12	27	202.69	74.836
	9-10	30	10	20	214.80	73.091
Upper	10-11	42	23	19	194.02	56.759
	11-12	16	7	9	289.50	124.125
Total		169	76	93	199.78	80.671

3.2. Analysis and labeling of the morphological categories

The text analysis process followed 4 steps. First, the texts were pre-processed. This step includes the digitalization of the texts (all the texts were originally written by hand). During the second step the researchers corrected spelling mistakes without

altering the syntactic structures in the texts (except for specific cases in which the syntactic mistakes would have prevented correct morphological analysis and labeling).

Next, the morphological labeling process was performed, which is initially an automatic process. In this case, however, we used HectorWWW (available on http://eines.iula.upf.edu/cgi-bin/hectorwww/hectormain.pl), a morphological analyzer and disambiguator developed by the University Institute of Applied Linguistics (IULA) at the Pompeu Fabra University in Barcelona. This tool works with three languages: Catalan, Spanish and English. The output provided is a text file with a list of all entries and their corresponding morphosyntactic label, with a standard format (Morel, Torner, Vivaldi, De Yzaguirre & Cabré, 1998). After automatic labeling, a manual revision was conducted in order to correct any errors in the labels. An extract of the linguistic characterization of the labels used by HectorWWW is shown in Table 2. An example of the output is illustrated in Figure 1 in which the Catalan sentence *La Caputxeta. Hi havia una vegada una nena que es deia Caputxeta /Once upon a time there was a girl called Little Red Riding Hood/* is tagged by HectorWWW.

Table 2. Linguistic characterization of HectorWWW's tag set.

Hectorwww's tag set	**Key**
AFP	Article-Feminine-Plural
EP12MS	Specifier-Possessive-1possessor-2nd Person-Masculine – Singular
JQ-FP	Adjective-Description- Feminine-Plural
N5-MS	Noun- Common- Masculine-Singular
REO-2MP	Pronoun-Personal-Strong-2nd Person-Masculine-Plural
VDP2S-	Verb Indicative Perfect 2nd Person-Singular

Figure 1. Output of the sentence *La Caputxeta. Hi havia una vegada una nena que es deia Caputxeta* tagged by HectorWWW.

```
<div1>
<head>
LA      AFS     el
CAPUTXETA       N4666   <unknown>
.       SENTF   .
</head>
<p>
<s>
Hi      REE8--- pr
havia   VDA3S-  heure:haver
una     E6--FS  un
vegada  N5-FS   vegada
una     E6--FS  un
nena    N5-FS   nen
que     RR---66 que
es      REEZZZS pr
deia    VDA3S-  dir
<name>
Caputxeta       N4666   <unknown>
</name>
```

Continuing with the third step, the labels were simplified following the model suggested by Bel, Queralt, Spassova and Turell (2012). In the case of conjugated verbs, only the number and the person were kept. For impersonal verb forms (V), only the type of form (infinitive or gerund) is kept, except for participles, where the number is also retained. As for nouns, they are classified into proper (N4) and common (N) nouns, with the latter also including information on their number (singular or plural). With respect to the other categories, such as articles (A), adjectives (J) and pronouns (R), only the numerical information is maintained. Categories which do not require any additional information are adverbs (D), conjunctions (C) and punctuation (DLD). Table 3 below shows the simplified tag set and its key meaning, while Figure 2 shows the previous example sentence with the new tag set.

Table 3. Linguistic characterization of tag set number 13. Source: Bel, N., S. Queralt, M. S. Spassova, and M. T. Turell. (2012: 196).

Hectorwww's tag set	**Tag set No. 13**	**Key**
AFP	AP →	Article-Plural
EP12MS	ES →	Specifier-Singular
JQ-FP	JP →	Adjective-Plural

N5-MS	NS →	Noun-Singular
REO-2MP	RP →	Pronoun-Plural
VDP2S-	V2S →	Verb 2nd Person-Singular

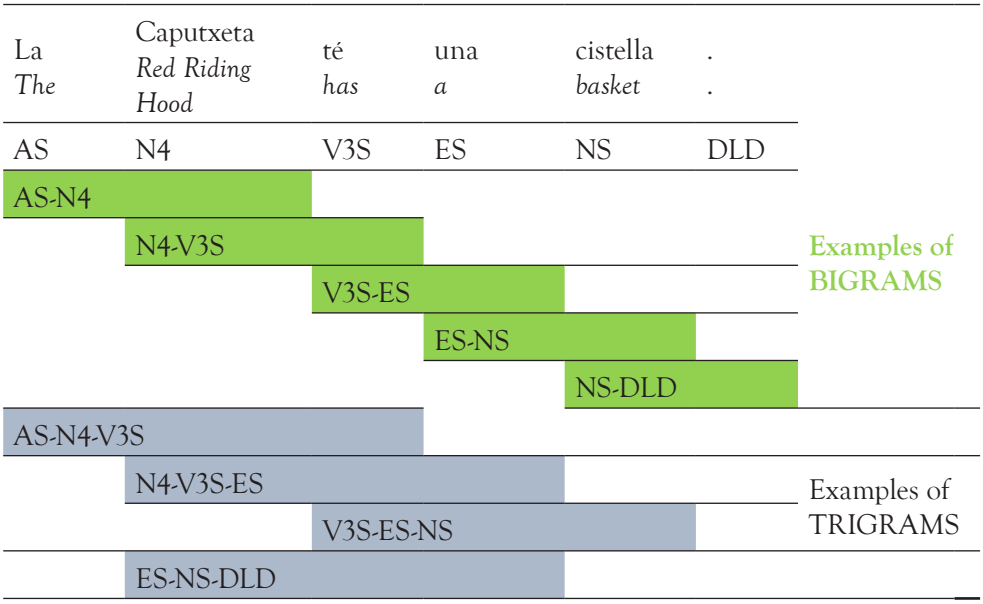
Figure 2. Output of the sentence *La Caputxeta. Hi havia una vegada una nena que es deia Caputxeta* with Tag set No. 13.

```
<div1>
<head>
LA      AS      el
CAPUTXETA       N4      <unknown>
.       SENTF   .
</head>
<p>
<s>
Hi      R-      pr
havia   V3S     heure:haver
una     ES      un
vegada  NS      vegada
una     ES      un
nena    NS      nen
que     R6      que
es      RS      pr
deia    V3S     dir
<name>
Caputxeta       N4      <unknown>
</name>
```

Finally, the last step consists in extracting n-grams (bigrams and trigrams) using the ForensicLab's private tool, known as Legolas software, specifically developed in the University Institute of Applied Linguistics (IULA) to work with output files produced by HectorWWW. At the end of this process, the n-gram results are obtained in a format which is adequate for their statistical treatment.

It must be stressed that only sets of two and three morphological categories (bi- and trigrams) have been examined since most researchers that have used n-grams have concluded that bigrams and trigrams are the combinations offering the highest performance (e.g. Baayen et al., 1996; Stamatatos, Fakotakis and Kokkinakis, 2000; Hirst and Feiguina, 2007; Nazar & Sánchez Pol, 2007; Spassova, 2007; Spassova & Turell, 2007; Grant, 2007; or Bel et al., 2012). Bel et al. (2012) also recommend the restriction of the number of bigrams and trigrams to the 40 most used (out of the hundreds possible) to facilitate analysis. An example of bigrams and trigrams is shown in Figure 3 below.

Figure 3. Examples of bigram and trigram.



3.3. Statistical analysis

Given the nature of the data and the goals pursued, two related statistical techniques were used. On the one hand, an ANOVA test was carried out to determine which variables showed significant-enough differences to classify the texts into the three age groups analysed. In addition, the post-hoc Dunnett's T3 test was applied to the data to reveal which groups differed from which. The Dunnett's T3 test was chosen due to the nature of the sample: unbalanced groups (containing a different number of individuals in each group) and unequal variances (the Levene test showed that the variances are not similar in all the groups).

On the other hand, a linear discriminant analysis (LDA) was carried out. This multivariable statistical technique has a twofold goal. Firstly, it identifies the features which can be used to differentiate two or more groups of cases and it constructs discriminant functions based on them. Secondly, it can classify new cases as belonging to one group or another (Pardo and Ruiz, 2002: 499).

The LDA technique consists in determining the characteristics that differentiate the distinct groups and, from those, finding the optimal plane where the projection of the observations best separates the groups. Subsequently, this optimal plane allows

us to classify new cases, i.e., to assign a new observation to one of the existing groups based on the values taken by their original variables.

In this study, conducted through SPSS software (19th version), we used the stepwise inclusion method and Lambda de Wilks, with the criteria of values being F: 3.84 as input and 2.71 as output (the standard values in SPSS). In order to check the discriminative power, we used the cross-validation method (leaving one out). With this method, one observation of the analysis is removed, and the discriminant functions are generated. After that, this observation is classified into one of the groups. Since the group to which the excluded observation of the analysis belongs is known beforehand, it is possible to check if the subsequent classification is correct. This process repeats itself for each observation.

4. Results

4.1. Bigrams

Table 4 below shows the combination of 2 grammatical categories which present significant differences between one or more than one group. In total, 25 bigrams were selected, 16 of which present differences between the lower and the middle cycles and 17 between the lower and upper cycles. 7 bigrams show differences between the middle and upper cycle. Only 2 of the variables (DLD-D and DLD_V3S) distinguish between the 3 cycles.

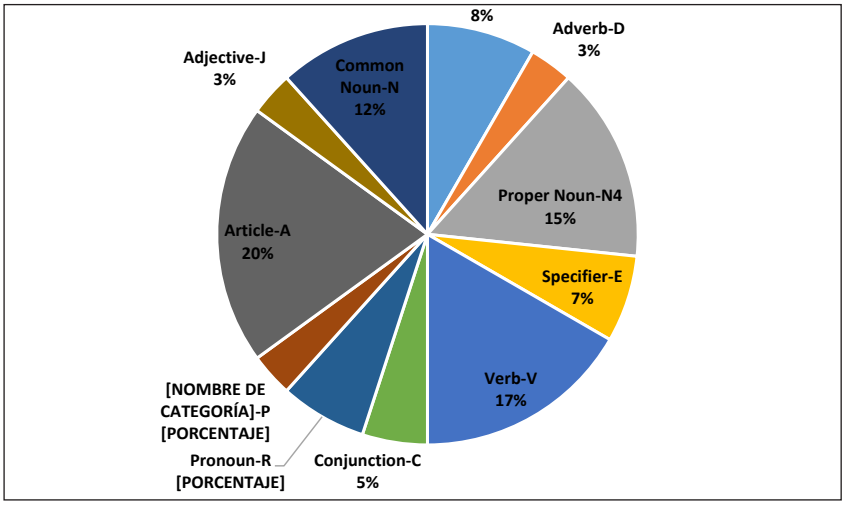
When observing the mean frequency of each bigram for each group, ascending and descending tendencies can be established, as well as cases in which the middle cycle presents small fluctuations. The bigrams which are most frequent in the lower cycle tend to be the most basic and common syntactic structures, while trigrams with ascending tendencies in their use are related to the use of punctuation (DLD) in 72% of the cases.

Table 4. Bigrams that present significant differences between groups.

Bigram	Differences lower-middle cycle	Differences lower-upper cycle	Differences middle-upper cycle
AS_N4	√	√	
P_AS	√		√
AS_NS			√
NS_P		√	
NS_C	√	√	
C_V3S	√		
NS_DLD	√	√	
C_AS		√	
P_VI			√
VI_AS	√	√	
N4_DLD	√	√	
C_RS		√	
DLD_AS	√	√	
DLD_D	√	√	√
N4_N4	√	√	
AS_ES	√	√	
DLD_C	√	√	
D_JS		√	
V3S_ES			√
VI_DLD	√		√
N4_RS	√		
NS_D		√	
D_DLD	√	√	
DLD_V3S	√	√	√
Total	**16**	**17**	**7**

From this table, we can also learn which categories are most significant when establishing differences between the cycles. As has already been explained, the most discriminant category is punctuation, followed by the use of articles, as well as proper nouns, verbs and conjunctions. The categories which show the smallest differences are adjectives, specifiers and pronouns. These results can be visualized in Figure 4:

Figure 4. Categories within the most significant bigrams for the detection of differences between cycles.



As concerns the results of the linear discriminant analysis, this was used to conduct a multivariate analysis of variance to test the hypothesis that the lower, middle and upper cycles would differ significantly on a linear combination of bigram variables. The overall Chi-square test turned out to be significant (Wilks = .290, Chi-square = 53.625, df = 8, Canonical correlation = .772, p <. 001). Regarding the variance explained in bigrams, 79% was explained by the first function and 21% by the second.

The stepwise discriminant analysis selected 9 of the 40 variables (Table 5). These 9 bigrams are variables which discriminate between the groups. As in the case of the bigrams, this satisfies the minimum sample size criterion (N=169) of having 10 cases per variable and the requirement that the number of cases in each group be equal to or exceed the number of variables recommended by Brown and Tinsley (1983), as well as Huberty's (1975) criterion to include at least 3 cases for every variable in each group. As shown, these results match the ones obtained for the ANOVA tests, since the bigrams are made up of the categories which presented a higher degree of significance in the ANOVA tests. The bigrams formed by punctuation marks (DLD), articles (A) and verbs (V) stand out.

Table 5. Variables included in the analysis.

P_AS	Preposition - Article Singular
NS_DLD	Noun Singular - Punctuation mark
P_VI	Preposition - Verb Infinitive
VI_AS	Verb Infinitive - Article Singular
N4_DLD	Proper Noun – Punctuation mark
DLD_D	Punctuation mark – Adverb
AS_ES	Article Singular – Specifier Singular
DLD_C	Punctuation mark – Conjunction
DLD_V3S	Punctuation mark – Verb Third Person Singular

Next, Table 6 presents the standardized discriminant function coefficients, and Table 7 shows the two functions at the group centroids.

Table 6. Standardized Canonical Discriminant Function Coefficient.

	Function	
	1	**2**
P_AS	-.230	.362
NS_DLD	.450	-.124
P_VI	.221	-.286
VI_AS	-.328	-.233
N4_DLD	.467	-.198
DLD_D	.094	.764
AS_ES	.546	.397
DLD_C	.322	-.047
DLD_V3S	.314	-.343

Table 7. Functions of Group Centroids.

Course	Function	
	1	2
lower cycle	-2.069	-.165
middle cycle	.879	-.594
upper cycle	.453	.826

Unstandardized canonical discriminant functions evaluated at group means

The classification results are shown in Table 8. Classification of cases based on the canonical variables was highly successful: 77.5% of the cases were correctly reclassified into their original categories. Cross-validation results were also successful, with 74% of the cases correctly classified. Based on the results, it can be observed that students in the lower and middle cycle are correctly classified in a very high percentage of cases (90.5% and 81.2%, respectively), and that those proving to be the most difficult to classify are students in the upper cycle, with a low 53.4% success rate in the classification. Therefore, the groups which are confused most frequently are the middle and upper cycles, while students in the lower cycle are clearly distinguished.

Table 8. Classification results.

		Course	Predicted Group Membership			Total
			lower cycle	middle cycle	upper cycle	
Original	Count	lower cycle	39	3	0	42
		middle cycle	1	58	10	69
		upper cycle	7	17	34	58
	%	lower cycle	92.9	7.1	.0	100.0
		middle cycle	1.4	84.1	14.5	100.0
		upper cycle	12.1	29.3	58.6	100.0
Cross-validated [b]	Count	lower cycle	38	3	1	42
		middle cycle	1	56	12	69
		upper cycle	8	19	31	58
	%	lower cycle	90.5	7.1	2.4	100.0
		middle cycle	1.4	81.2	17.4	100.0
		upper cycle	13.8	32.8	53.4	100.0

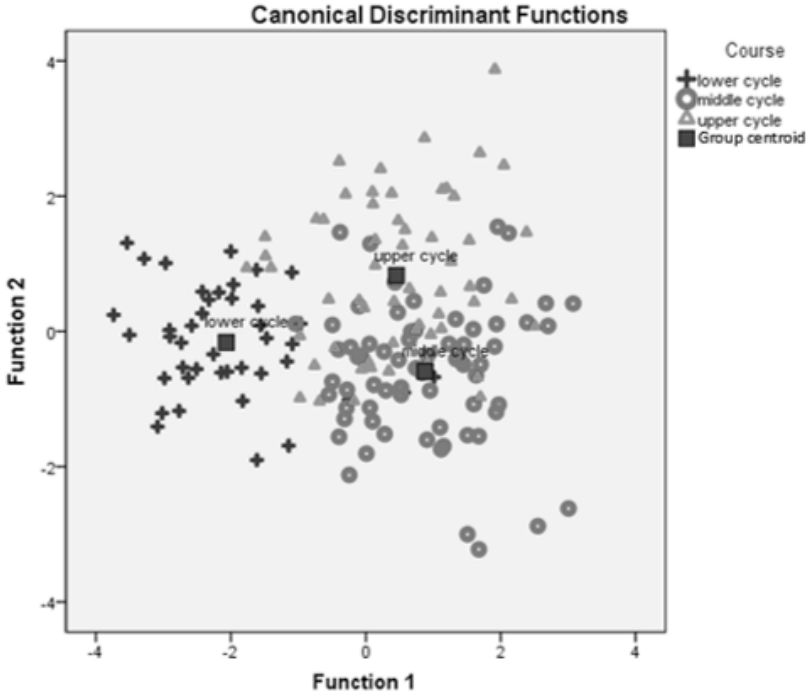
a. 77.5% of original grouped cases correctly classified.

b. Cross-validation is done only for those cases in the analysis. In cross-validation, each case is classified by the functions derived from all cases other than that case.

c. 74.0% of cross-validated grouped cases correctly classified.

Next, the results of the original classification are presented graphically. In Figure 5, each of the lower cycle samples are represented by a cross, the middle cycle samples are displayed as a circle, and the upper cycle ones as a triangle. It can be clearly observed that the centroid for the lower cycle is located far from the rest of the centroids and that there is only one case in which the sample finds itself nearer to another centroid. As regards the centroids for the upper and middle cycles, despite being closer to each other, both cycles can also be graphically distinguished, although there are more overlapping cases. Function 1 explained most of the differences (namely 79%), hence the importance of very dissimilar values in that function, as in the case of the lower cycle, which is located in negative values whereas the higher cycles show positive values which are close to each other. Nevertheless, Function 2 allows us to differentiate more clearly between the middle and the upper cycles, locating them in negative and positive values, respectively.

Figure 5. Canonical Discriminant Functions.



4.2. Trigrams

Table 9 contains the trigrams which present differences between the groups and shows between which groups the differences are found. There are a total of 21

sequences of 3 grammatical categories which show differences between the groups. Specifically, 13 of the variables present differences between the lower and the middle cycles, 14 variables show differences between the lower and upper cycles, and only 2 variables distinguish between the middle and upper cycle. Therefore, the most easily distinguished group is the lower cycle, both from the middle and upper cycles. The combination of middle and upper cycle presents very few differences. None of the variables show differences between the 3 cycles.

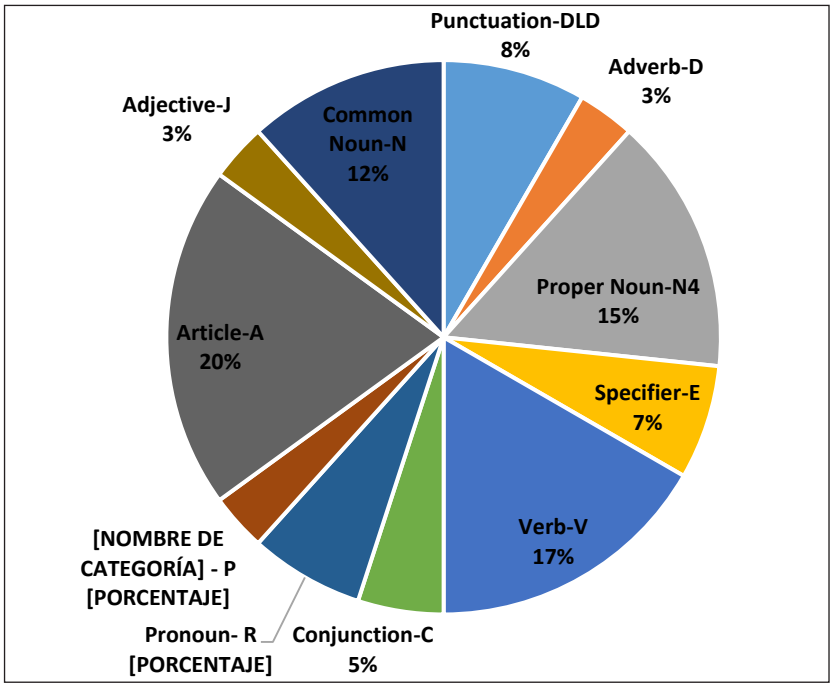
In the case of the mean frequency of each of the trigrams, in 62% of cases the students in the lower cycle repeat concrete trigrams which do not contain a punctuation label with a higher frequency and that their use diminishes as we ascend in the cycles: in other words, the higher the cycle, the higher the cycle, the less frequently are they used. This could be attributable to the fact that students in the lower cycle possess less syntactic richness and therefore tend to repeat a given structure more often. Another interesting fact is that in 50% of the trigrams whose frequency is inverse (that is, the higher the level, the higher the frequency of use), we notice that the trigrams are related to the punctuation label (DLD).

Table 9. Trigrams that present significant differences between groups.

Trigram	Differences lower-middle cycle	Differences lower-upper cycle	Differences middle-upper cycle
VI_P_AS	✓		
C_AS_N4		✓	
V3S_VI_AS		✓	
AS_ES_NS	✓	✓	
AS_N4_C	✓	✓	
VI_AS_N4	✓	✓	
DLD_AS_N4	✓	✓	
AS_N4_DLD	✓	✓	
V3S_ES_NS			✓
AS_NS_C	✓		
P_VI_RS			✓
AS_N4_RS	✓		
N4_RS_V3S	✓		
V3S_VI_DLD	✓	✓	
AS_N4_N4	✓	✓	
ES_NS_DLD	✓	✓	
AS_NS_DLD	✓	✓	
NS_D_JS		✓	
R_V3S_ES		✓	
P_AS_ES		✓	
NP_D_JP		✓	
Total	13	15	2

As regards the categories which form the trigrams with significant differences between the groups, it can be observed that they follow a pattern similar to that of bigrams, since the categories presenting more differences are verbs, articles and punctuation.

Figure 6. Categories within the most significant trigrams for the detection of differences between cycles.



Discriminant analysis was used to conduct a multivariate analysis of variance test of the hypothesis that the cycles would differ significantly on a linear combination of trigram variables. The overall Chi-square test proved to be significant (Wilks = .375, Chi-square = 29.354, df = 8, Canonical correlation = .742, p <. 001). Of the variance explained in trigrams, 86% was explained by the first function and 14% by the second.

The stepwise discriminant analysis discarded 9 of the 40 variables (Table 10). These 9 trigrams are discriminant variables between the groups. As in the case of bigrams, the results satisfy the minimum sample size criterion, the requirement that the number of cases in each group be equal to or exceed the number of variables, and also the criterion of at least 3 cases for every variable in each group. Again, the most frequent categories are verbs, punctuation and articles, although proper nouns and specifiers also prove to be interesting in the case of the most discriminant trigrams.

Table 10. Variables included in the analysis.

Trigrams	Key
ES_NS_DLD	Specifier Singular – Noun Singular – Punctuation mark
AS_NS_ DLD	Article Singular – Noun Singular – Punctuation mark
V3S_VI_ DLD	Verb Third Person Singular – Verb Infinitive – Punctuation mark
AS_N4_DLD	Article Singular – Proper Noun – Punctuation mark
AS_ES_NS	Article Singular – Specifier Singular – Noun Singular
VI_AS_N4	Verb Infinitive – Article Singular – Proper Noun
R_V3S_ES	Pronoun – Verb Third Person Singular – Specifier Singular
V3S_ES_NS	Verb Third Person Singular – Specifier Singular – Noun Singular
V3S_VI_C	Verb Third Person Singular - Verb Infinitive - Conjunction

Table 11 presents the standardized discriminant function coefficients. Table 12 shows the two functions at the group centroids.

Table 11. Standardized Canonical Discriminant Function Coefficient.

	Function	
	1	*2*
AS_ES_NS	*0.348*	*-0.428*
V3S_VI_C	*0.265*	*0.364*
VI_AS_N4	*-0.358*	*0.471*
AS_N4_DLD	*0.494*	*0.099*
V3S_ES_NS	*0.272*	*1.042*
V3S_VI_DLD	*0.466*	*0.271*
ES_NS_DLD	*0.349*	*-0.225*
AS_NS_DLD	*0.332*	*0.294*
R_V3S_ES	*-0.584*	*-0.667*

Table 12. Functions of Group Centroids.

Functions of Group Centroids		
Course	Function	
	1	2
lower cycle	-1.903	0.043
middle cycle	0.695	0.452
upper cycle	0.552	-0.569
Unstandardized canonical discriminant functions evaluated at group means		

The classification results are shown in Table 13. Classification of cases based on the canonical variables was successful in 68.6% of the cases, which were correctly reclassified into their original categories. Cross-validation results are also successful in 65.1% of the cases. From the results, it can be observed that the students in the lower cycle are correctly classified in a very high percentage of cases (88.1%). The classification presents issues with the middle and upper cycles, which are only correctly classified in 55.1% and 60.3% of the cases, respectively. Again, these two groups are confused with each other, while the samples by students in the lower cycle are clearly differentiated.

Table 13. Classification results.

		Course	Predicted Group Membership			Total
			lower cycle	middle cycle	upper cycle	
Original	Count	lower cycle	39	1	2	42
		middle cycle	6	39	24	69
		upper cycle	7	13	38	58
	%	lower cycle	92.9	2.4	4.8	100
		middle cycle	8.7	56.5	34.8	100
		upper cycle	12.1	22.4	65.5	100
Cross-validated [b]	Count	lower cycle	37	2	3	42
		middle cycle	6	38	25	69
		upper cycle	7	16	35	58
	%	lower cycle	88.1	4.8	7.1	100
		middle cycle	8.7	55.1	36.2	100
		upper cycle	12.1	27.6	60.3	100

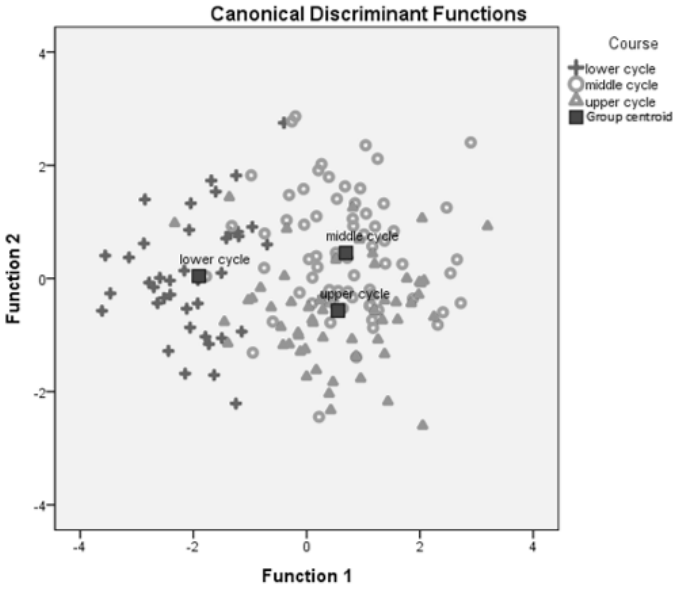
a. 68.6 % of original grouped cases correctly classified.

b. Cross-validation is done only for those cases in the analysis. In cross-validation, each case is classified by the functions derived from all cases other than that case.

c. 65.1% of cross-validated grouped cases correctly classified.

Next, the results of the original classification are displayed graphically. In Figure 7, as in the previous illustration, each of the samples of the lower cycle is shown as a cross, those of the middle cycle as a circle and the upper cycle ones as triangles. This graph shows that the centroid for the lower cycle is clearly distanced from the other cycles, while the middle and upper cycle centroids are located closer to one another. Therefore, the overlap between the samples of these two groups is also larger. It should be noted that Function 1 explained 86% of the variance, which means that the most notable differences must be found in the x-axis. Thus, we can see that the centroid for the lower cycle is located around the value -2, that is, very far from the other cycles. Function 2, with a low 14% of variance, presents weaker differences between the groups. However, it again allows us to distinguish between the higher cycles, since the middle cycle centroid is found in positive values whereas the upper cycle shows negative values.

Figure 7. Canonical Discriminant Functions.



5. Discussion

The analytical methods used in this study have shown once again that discriminant analysis is a very useful tool to analyse and describe the differences between groups of samples, as well as to classify new samples based on the differences and similarities presented by the use of their variables.

From a global perspective, it can be observed that bigrams which include a punctuation mark (DLD) present significant differences between the cycles on 18 occasions, followed by 12 bigrams which include a singular article (AS). The rest of the bigrams with other morphological categories present significant differences less frequently: ranging from a single instance of the singular adjective (JS) category, to seven instances of proper nouns (N4), adverbs (D) and conjunctions (C).

The discriminant analysis performed does not include all of the bigrams, but only those which can be used to better discriminate between the groups. The results of the discriminant analysis enable the classification of texts belonging to the three school cycles according to the frequency of bigram use. Thus, we can see that texts corresponding to the first cycle are correctly classified in 92.9% of instances (mainly due to the first discriminant function, as can be seen in Tables 6 and 7, since it clearly distinguishes this cycle from the middle and upper cycles). As to the frequency of bigrams, the main features include the low use of the combinations P_

AS (preposition-singular article), VI_AS (infinitive verb-singular article) and —with a lower frequency— DLD_D (punctuation mark-adverb). At the same time, the AS_ES (singular article- singular specifier), NS_DLD (singular noun-punctuation mark), N4_DLD (proper noun-punctuation mark), DLD_C (punctuation mark-conjunction) and DLD_V3S (punctuation mark-third person singular verb) bigrams are found with a highest ratio in the middle cycle (although they are also present in the upper cycle), but with little presence in the first cycle. Lastly, for the texts of middle and upper cycles, misclassifications represent around 30% of instances, where a text written by a student in the middle cycle was falsely attributed to the upper cycle. This process of misclassification occurred in the opposite direction in 14.5% of cases. The main variables which can be used to distinguish between the middle and upper cycles include DLD_D (punctuation mark-adverb), AS_ES (singular article-singular specifier) and P_AS (preposition-singular article), which characterize the texts in the upper cycle (and which are much less frequent in those by students of the middle cycle). These variables are related to a greater complexity in the texts. Thus, the students of the initial cycle tend to write shorter sentences, so that more bigrams appear in which a noun and a punctuation mark (mainly full stops) are combined, while the students of the middle and upper cycles make longer and more complex phrases. On the other hand, the bigrams that characterize the middle and upper cycles best are, on the one hand, the combination of a punctuation mark and an adverb, and on the other hand, a preposition and an article. It is shown, thus, that older children are more likely to use adjuncts (introduced by the adverb) at the beginning of the phrases, as well as introduce more complements of the name or adjuncts (introduced by the combination preposition and article).

These data suggest that students experience an important qualitative leap in their essay writing when they move from the first to the middle cycle. Furthermore, the change in the use of punctuation marks is significant (Hall, 1999; Sing & Hall, 2009).

As for trigrams, the morphological category found in those which present significant differences between the groups is the singular article (AS), with a total of 20 instances, followed by the proper noun (N4), with 13, and punctuation marks (DLD), with 10. Again, the first cycle is the easiest to distinguish (the discriminant analysis was successful in 92.9% of the cases). The texts in this group are characterized by the low use of the trigrams VI_AS_N4 and R_V3S_ES. Regarding the first of these two trigrams (VI_AS_N4), first-cycle students use it very repetitively (in most cases the AS_N4 is the object of the verb). Older children choose to introduce in their texts more clitic pronouns, so that the structures are more varied and richer. It is also interesting to emphasize the use of pronouns (R) in the second trigram: even though Catalan is a pro-drop language, younger children do not have enough variety

of linguistic resources to mark the subject, so they use significantly more personal pronouns with this syntactic function than the more competent writers.

The texts in the middle cycle are correctly classified in 56.5% of cases (most of the misclassifications occur with the upper cycle). The most frequent trigrams in this cycle are V3S_ES_NS, VI_AS_N4 and V3S_VI_C, as well as those containing punctuation marks. From this group of trigrams, the last one stands out because it includes a conjunction (C), which indicates that students at that age already use subordinate sentences frequently. This trigram is also frequent in the students of the upper cycle. Lastly, the texts by students in the upper cycle are correctly classified in 65.5% of cases and they are characterized by a lower frequency of the trigrams AS_ES_NS, R_V3S_ES and ES_NS_DLD. In this group, the low frequency of trigrams that include pronouns (R) stands out (which correspond mainly to personal pronouns with a subject function in the texts), since children in the upper cycle are already able to adopt other strategies to mark the subject or choose to omit it. In contrast, trigrams including punctuation are frequent, as they are increasingly more competent with the use of punctuation, especially commas. Once more, we can see that the trigram analysis reinforces the divide between the first and the other two cycles.

6. Conclusions

The results have shown that the analysis of bigrams and trigrams of morphological labels is useful for classifying texts according to the age of the children. The overall percentage of correct classifications is around 70% (77.5% in bigrams and 68.6% in trigrams). With regard to bigrams, it has been observed that those that include a punctuation mark are relevant for discriminating between groups. The differences between age groups with regard to some bigrams that include prepositions (and which usually include complements) and adverbs (which work as adjuncts) are also significant. Regarding the trigrams, once again those that include punctuation marks allow to discriminate between age groups. Also relevant are those that include conjunctions (which introduce mostly subordinate clauses) and personal pronouns (which mostly serve as the subject, and which the older children use in a greater proportion, since they do not have enough syntactic resources to avoid repetition of explicit subjects).

The majority of the bigrams and trigrams studied allow discrimination between the initial cycles (6-7 years), the middle cycles (8-9) and upper cycles (10-11). On the other hand, there are few differences between the middle and upper cycles. Thus, globally, the study can confirm the idea that when students turn 9 they experience a significant change in their writing competence, since they begin to use more complex

syntactic structures, they use the punctuation marks more efficiently and show more ability to avoid repetitions (mainly because they introduce the use of clitic pronouns).

References

Anderson, J. R. (1982) Acquisition of cognitive skill. *Psychological Review* 89: 369-406.

Baayen, R.H., van Halteren, H. & Tweedie, F.J. (1996). Outside the cave of shadows: Using syntactic annotation to enhance authorship attribution. *Literary and Linguistic Computing* 11(3): 121-131.

Bamman, D., Eisenstein, J. & Schnoebelen, T. (2012). *Gender in Twitter: Styles, Stances, and Social Networks*. Unpublished manuscript.

Bel, N., S. Queralt, Spassova, M.S. & Turell, M.T. (2012) The use of sequences of linguistic categories in forensic written text comparison revisited. In S. Tombilin, N. MacLeod, R. Sousa-Silva, & M. Coulthard (eds.) *Proceedings of the International Association of Forensic Linguists' Tenth Biennal Conference*. Birmingham: Center of forensic linguistics.

Björk, L. & Blomstrand, I. (2000) *La escritura en la enseñanza secundaria. Los procesos del pensar y del escribir*. Barcelona: Graó.

Brown, M. T. & Tinsley. H.E.A. (1983) Discriminant analysis. *Journal of Leisure Research* 15(4): 290-310.

Camps, A. (1990) Modelos del proceso de redacción: algunas implicaciones para la enseñanza. *Infancia y Aprendizaje* 49: 3-19.

Cheng, N., Chandramouli, R. & Subbalakshmi K.P. (2011) Author Gender Identification from Text. *Digital Investigation* 8: 78-88.

Efron, R. & Thisted, B. (1976). Estimating the number of unseen species: How many words did Shakespeare know?. *Biometrika* 63(3): 435-447.

Enron email dataset (2005, April). [Online]. Available: http://www-2.cs.cmu.edu/~enron/.

Fitzgerald, J. & Markham, L. (1987) Teaching children about revision in writing. *Cognition and Instruction* 4(1): 3-24.

Flower, L. & Hayes, J. (1981) Plans that guide the composing process. In C. Fredericksen and J. Dominic (eds.) *The nature, development and teaching of written communication, 2, Writing: process, development and communication*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Graham, S. (2006). Writing. In P. Alexander & P. Winne (eds.) *Handbook of educational psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Graham, S. & Hebert, M. (2011) Writing to Read: A Meta-Analysis of the Impact of Writing and Writing Instruction on Reading. *Harvard Educational Review* 81(4): 710-744.

Grant, T. (2007) Quantifying evidence in forensic authorship analysis. *The International Journal of Speech, Language and the Law* 14(1): 1-25.

Hall, N. (1999) Young Children's Use of Graphic Punctuation. *Language and Education* 13(3): 178-193.

Hirst, G. & Feiguina, O. (2007) Bigrams of syntactic labels for authorship discrimination of short texts. *Literary and Linguistic Computing* 22(4): 405-417.

Holmes, D. I. & Forsyth, R. (1995) The federalist revisited: New directions in authorship attribution. *Literary and Linguistic Computing* 10(2): 111-127.

Huberty, C. J. (1975) Discriminant analysis. *Review of Educational Research* 45: 543-598.

Jurafsky, D. (2003) Probabilistic Modeling in Psycholinguistics: Linguistic Comprehension and Production. In R. Bod, J. Hay, and S. Jannedy (eds) *Probabilistic Linguistics*. Cambridge, Massachusetts/London, England: The MIT Press.

Lowe, D. & Matthews, R. (1995) Shakespeare vs. Fletcher: A stylometric analysis by radial basis functions. *Computers and the Humanities* 29: 449-461.

Merriam, T. (1996) Marlowe's hand in Edward III revisited. *Literary and Linguistic Computing* 11(1): 19-22.

Morel, J.,Torner, S, Vivaldi, J. De Yzaguirre, L & Cabré, M.T. (1998). *El corpus de l'IULA: etiquetaris. Papers de l'IULA, sèrie informes, 18*. Barcelona: IULA, Universitat Pompeu Fabra.

Mosteller, F. & Wallace, D.L. (1964). *Inference and Disputed Authorship: The Federalist*. Reading, MA: Addison-Wesley Publishing Company, Inc.

Nazar, R. & Sánchez Pol, M. (2007) An extremely simple authorship attribution system. In M. T. Turell, M. S. Spassova, & J. Cicres (eds) *Proceedings of the Second European IAFL Conference on Forensic Linguistics/Language and the Law.* Barcelona: Publicacions de l'IULA. 197-203.

Newell, A. (1990) *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

OECD (2004) *Reviews of National Policies for Education. Denmark: Lessons from PISA 2000*. Paris, OECD.

Pardo, A. & Ruiz, M.A. (2002) *SPSS 11. Guía para el análisis de datos*. Madrid: McGraw–Hill.

Reuters corpora (2000). [Online]. Available: http://trec.nist.gov/data/reuters/reuters.html.

Queralt, S. & Turell, M.T. (2013). A semi-automatic authorship attribution technique applied to real forensic cases involving Judgments in Spanish. In R. Sousa-Silva, R. Faria, N. Gavaldà, & B. Maia. (eds.) *Bridging the Gap(s) between Language and the Law: Proceedings of the 3rd European Conference of the International Association of Forensic Linguists*. Porto: Faculdade de Letras da Universidade do Porto.

Shaoul, C. & Westbury, C. (2011). Sequences: Do they exist and do they matter?. *The mental lexicon* 6(1): 171-196.

Sing, S. & Hall, N. (2009). Listening to children think about punctuation. In A. Carter, T. Lillis, and S. Parkin (eds.) *Why Writing Matters. Issues of access and identity in writing research and pedagogy*. Studies in Written Language and Literacy, 12. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Sofkova Hashemi, S. (2003). *Automatic Detection of Grammar Errors in Primary School Children's Texts. A Finite State Approach.* PhD Thesis. Department of Linguistics. Göteborg University.

Sotomayor, C., G. Lucchini, P. Bedwell, M. Biedma, C. Hernández & D. Molina. (2013) Producción escrita en la Educación Básica: análisis de narraciones de alumnos de escuelas municipales de Chile. *ONOMÁZEIN* 27: 53-77.

Spassova, M. S. (2007) The relevance of inter and intra authorial variation in authorship attribution. Some findings on syntactic identification markers. *8th Biennial Conference on Forensic Linguistics/Language and the Law*. University of Washington, Seattle.

Spassova, M. S & Turell, M.T. (2007) The use of morpho-syntactically annotated tag sequences as forensic markers of authorship attribution. In M. T. Turell, M. S. Spassova, and J. Cicres (eds.) *Proceedings of the Second European IAFL Conference on Forensic Linguistics/Language and the Law*. Barcelona: Publicacions de l'IULA. 229-237.

Stamatatos, E., N. Fakotakis & G. Kokkinakis. (2000) Automatic text categorization in terms of genre and author. *Computational Linguistics* 26(4): 471-495.

Teberosky, A. (2001) *Proposta constructivista per aprendre a llegir i a escriure*. Barcelona: Editorial Vicens Vives.

Turell, M. T. 2010. The use of textual, grammatical and sociolinguistic evidence in forensic text comparison. *International Journal of Speech Language and the Law* 17(2): 211-250.

Tweedie, F. J., S. Singh & D. I. Holmes. (1996) Neural network applications in stylometry: The federalist papers. *Computers and the Humanities* 30(1): 1-10.

The Persistence of L1 Patterns in SLA: the Boundary Crossing Constraint and Incidental Learning

Amani Alghamdi
University of Swansea
amanialghamdi223@gmail.com

Michael Daller
Reading University
m.daller@reading.com

James Milton
Swansea University
j.l.milton@swansea.ac.uk

Abstract

Based on the theoretical framework of Talmy (1985 et passim) and Slobin (1987 et passim) the present study analyses the influence of L1 patterns on the description of motion events with boundary crossings. Arab speakers avoid the use of manner of motion verbs in the description of these events and use simple path verbs (e.g. enter, go etc.), whereas speakers of English mostly use manner verbs (run, crawl etc.). These deeply engrained differences between L1 and L2 are a learning challenge in SLA. We analyse the use of manner verbs by an intermediate and an advanced group of Arab EFL learners, who live in the UK. Most learners either avoid the description or use path verbs as in Arabic. As the learners do not produce ungrammatical sentences, they will not receive negative feedback (e.g. from a teacher) and rely entirely on incidental learning from the input. However, despite the high frequency of these manner verbs in the daily input of the learners, they do not acquire the patterns of the target language even at a high proficiency level. Implicit learning in this context is hardly possible and explicit teaching and learning is needed to overcome the influence of the first language.

Keywords: motion events, boundary-crossing, implicit statistical learning, linguistic typology, Arabic speakers of English

Resumen

Basado en el marco teórico de Talmy (1985 et passim) y Slobin (1987 et passim), el presente estudio analiza la influencia de los patrones de la L1 en las descripciones de eventos de movimiento con fronteras. Los hablantes de árabe evitan el uso de la forma de verbos de movimiento en la descripción de estos eventos y usan verbos más simples (por ejemplo, entrar, ir, etc.), mientras que los anglófonos usan principalmente verbos de movimiento (ejecutar, arrastrarse, etc.). Estas diferencias profundamente arraigadas entre la L1 y la L2 son un desafío de aprendizaje en la ASL. Analizamos el uso de verbos modales entre un grupo de árabe ILE (Inglesa como Lengua Extranjera) intermedio y avanzado, que viven en el Reino Unido. La mayoría de los estudiantes evitan la descripción o usan verbos de camino como en árabe. Como los estudiantes no producen oraciones agramaticales, no recibirán comentarios negativos (por ejemplo, de un profesor) y dependerán totalmente del aprendizaje incidental de entradas cotidianas. Sin embargo, a pesar de la alta frecuencia de estos verbos de movimiento en el aporte diario de los alumnos, no adquieren los patrones del idioma de destino, incluso en un nivel alto de competencia. El aprendizaje implícito en este contexto es casi imposible y se necesita enseñanza y aprendizaje explícitos para superar la influencia del primer idioma.

Palabras clave: eventos de movimiento, cruce de fronteras, aprendizaje estadístico implícito, tipología lingüística, hablantes árabes de inglés

1. Introduction

Speakers of different languages vary in their lexicalization of motion events. The study of these language-specific preferences led Talmy (1985, 1991, 2000a and 2000b) to propose his typology which depends on the ways the semantic components of motion are verbalized across the world's languages. In Talmy's original framework, languages are classified into two groups: Satellite-framed languages (S-languages) and Verb-framed languages (V-languages). According to Slobin (2004), S-languages can easily encode the use of *Manner* with motion verbs at boundary crossings, scenes where a figure crosses a spatial boundary, (he ran into the room) since they can encode the direction of the movement in the satellite (into), whereas this use is not licenced in V-languages (the boundary crossing constraint), such as French, Spanish and many others, where constructions such as "he entered the room running" or simply "he entered the room" are used. As a consequence, speakers of V-languages avoid the use of manner of motion verbs with boundary crossings in English because it is in conflict with their L1 patterns. Although Slobin's classification is useful to describe and predict these aspects of second language acquisition, it has been shown that the classification

is not a clear-cut dichotomy and that languages only have a general tendency for one category but also show structures that are more in line with the other category (Beavers, Levin, & Tham 2010, Slobin 2004). Arabic has typically been classified into a V-language (Saidi 2008), but how Arab learners describe *Manner* at boundary crossing motion events is under-researched. To this end, the present study attempts to compare the use of manner of motion verbs in boundary crossings of 64 Arab EFL learners with two proficiency levels (intermediate and advanced) who live in the UK with two monolingual groups: Arabic and English. First, we need to establish that English and Arabic natives follow different patterns. Then, we investigate the patterns produced by the learner groups. We assume that native groups will show different typological patterns and that both learner groups will face difficulty when using the manner verb in boundary crossing due to the effect of the cognitive constraint from L1.

As the learners do not produce ungrammatical structures with a path rather than a manner verb as the main verb (e.g. he entered the house), it is expected that they will not receive negative feedback. The teaching of motion event patterns has received little attention (exceptions are Attwood & Treffers-Daller (under review), and Cadierno & Robinson 2009). With the lack of negative feedback and deliberate teaching, the only possible way available for the learners is through incidental learning from the frequency in the input. Frequency of input has long been established in the literature as a factor which often leads to the acquisition of second language (L2) features. For example, lexical items which are frequently used tend to be learned relatively easier than the less frequent ones (Milton 2009). However, there are areas of Second Language Acquisition (SLA) that seem to be resistant to frequency effects and incidental learning (for a discussion see Ellis 2002, Gass and Mackey 2002). The boundary-crossing motion event domain could be one of these. To explore whether input frequency can play a role in learning manner verbs in boundary crossings, the frequency with which the learner groups use the target structures were compared with the frequency data of the same structures from the British National Corpus (BNC) and from the monolingual groups. We assume that, in our case, there are limitations for learning from frequency in the input and that the deeply ingrained cognitive constraint of L1 can only be overcome through explicit instruction. This paper is structured as follows: first we briefly discuss Talmy's typology and Slobin's thinking-for-speaking hypothesis with reference to the typological difference between Arabic and English. Second, the boundary crossing constraint is presented followed by a summary on the role of input frequency on incidental statistical learning. After that, the research questions and hypotheses are formulated, followed by a description of the study methods. Then, we present the results. Finally, the article is concluded by a discussion of the results.

2. Literature Review

2.1. Language Typology and Differences between L1 and L2

The typological frameworks of the present study are based on Talmy (1985, 1991, 2000a, 2000b) and Slobin's "thinking-for-speaking" (1987 et passim). Talmy's typology is based on different ways the semantic components of motion are used to describe motion events across the worlds' language. Generally, a motion event consists of a *Figure*, a *Path*, a *Ground or Landmark* and the *Motion* itself, e.g. the man (*Figure*) went (*Motion*) into (*Path*) the house (*Ground*). In addition, a co-event, such as *Manner of Motion* and *Cause of Motion*, can be expressed, e.g. the man ran (*Manner*) into the bank, the man threw (*Cause*) the ball. Talmy's typology makes a fundamental distinction between S-languages and V-languages. In S-languages such as English motion events can be described in a main verb and a satellite that indicates the *Path* (e.g. "into"). It is easy in these languages to express the *Manner* in the main verb (e.g. "ran into"). Most European languages, apart from the Romance languages belong to this type. Romance languages are V-languages where *Path* is typically expressed in the main verb and therefore *Manner* needs to be expressed in a different way, i.e. in a subordinate construction, for example, in Spanish "entrar corriendo a/en" (enters running) (Larrañaga, Treffers-Daller, Tidball and Ortega, 2012: 124). Slobin (1996) uses this typology to provide evidence for his thinking-for-speaking hypothesis in which he claims that people direct their attention while speaking to the components of motion events that are codable in their language, that is, the speakers' perspectives on motion events are defined by the options available in their language. This is also evident from studies on translations between typological different languages (Alonso 2018). The fact that V-language speakers express *Manner* outside the main verb when describing boundary crossings requires a heavier syntactic construction which could explain why speakers of V-languages tend to express *Manner* less frequently in these contexts (Slobin 2004). Özçalı□kan and Slobin (2003: 259) point out that there are no absolute rules but that it is about how "habitually" speakers of different languages describe motion events. In a similar vein, Slobin argues that languages can be ranked on a "cline of manner salience" and that "a number of factors contribute to the degree of salience of manner in languages" (2004:2). The picture is even more complex as satellite-framed patterns can be found in languages that are characterized as verb-framed, e.g. "Pierre s'est enfui de l'école" (Pierre ran away/ escaped from school), a satellite-framed structures that can be found in French, a typical V-language (Kopecka 2006: 83). Some languages, such as Mandarin, do not seem to fit into Talmy's simple dichotomy and therefore a third category has been suggested, serial-verb languages where one verb expresses *Manner* and another verb *Path*, e.g. Mandarin: *feil chul* "fly exit" (Slobin 2004: 8). Many studies have English as one part of a language pair in their methodology and it

is generally accepted that English is seen as an S-language (Özçalışkan & Slobin 2003, Slobin 2004, Alonso 2011, Gennari, Sloman, Malt & Fitch 2002). In Arabic, there are fewer studies. According to Talmy (1985) and Slobin (2006: 62), Arabic is classified as a V-language. This also holds for Arabic varieties other than Modern Standard Arabic (MSA). However, there seems to be a more complex situation. "Indeed, although Tunisian Arabic appears, as expected, to be strongly verb-framed, there are many other strategies which were not taken into account by Talmy's typology" (Saidi 2008: 202). Detailed studies of these alternative strategies in MSA are not available. However, bearing the complexity of the typological distinctions in mind, we assume as working hypothesis in the present study that the two languages involved here, English and Arabic, can be classified as either satellite-framed or verb-framed.

2.2. The boundary crossing constraint

This notion goes back to Slobin and Hoiting (1994) but the concept was first mentioned under a different name by Aske (1989). Aske compares the use of manner of motion verbs in Spanish and English and comes to the conclusion that the distribution of *Path* phrases is different in English and Spanish and that some constructions that are typical for English are not allowed in Spanish, e.g. "*ran into the house*". He suggests that this is the case because these telic phrases predict an end-state and in this context no manner of motion verbs are allowed in Spanish, instead a construction such as "entered the house running" is possible. Slobin and Hoiting coined the term boundary crossing constraint for V-languages. A more general version of this constraint is the notion "change of state generally" (Slobin 1997: 441), where not only movements but general changes of state, i.e. "*he kicked the door shut*" are not licensed in V-languages. Instead, constructions, such as "*he shut the door by kicking*", are used (see also Talmy 1991). In the context of the present study, we use the term "boundary crossing constraint" because our data consist of movement descriptions only. To our knowledge, there is only one study that includes boundary crossing events with Arabic speakers (von Stutterheim, Bouhaouos and Carroll, 2017). They found that "manner verbs do not combine with forms expressing a boundary crossing" (2017: 245) for MSA and for Tunisian Arabic. The literature also shows that there seems to be exceptions to the boundary crossing constraint for a variety of languages. Slobin (1997: 456) observes that manner of motion verbs are allowed in some V-languages in certain contexts, e.g. "he jumped from the branches" is possible in Portuguese. Slobin (2004: 7) argues that "verbs that encode particular forces that are more like punctual acts than activities, such as equivalents of 'throw oneself' and 'plunge'" might be an exception from the general constraint. In a similar vein, Naigles, Eisenberg, Kako, Highter and McGraw (1998) found that speakers of Spanish can use "jump" or "slide" when a figure jumps or slides into a pool. They conclude that "perhaps a boundary

crossing that is only the by-product of one's exertion, and not the original goal, is not viewed as a true boundary crossing" (1998: 453). In other words, the actor initiated the movement but after that the boundary crossing as such was out of his/her control and merely as result of gravity. Another possible explanation is that these events consist of a horizontal rather than a vertical motion (Naigles et al. 1998). These are of course speculative explanations, but it shows that the boundary crossing constraint is a complex issue and not just a simple dichotomy. It is beyond the scope of the present study, however, to investigate the exceptional character of these specific boundary crossings in detail but we include a picture with a figure that plunges into a pool and a figure that tumbles into a net in our data collection (see Appendix).

2.3. Incidental learning and input frequency

The question whether "it is possible for adults to learn linguistic regularities implicitly through exposure" has been raised by Kachinske, Osthus, Solovyeva and Long (2015: 391). Incidental language learning refers to "the acquisition of a word or expression without the conscious intention to commit the element to memory" (Hulstijn 2012: 1). It has been discussed for about 30 years (see Rebuschat 2015), mainly in first language acquisition research but increasingly also in SLA. In recent years, also the term "implicit statistical learning" has been used (Walk & Conway 2015: 191) to refer to the learners' ability to induce statistical regularities of language from the input automatically, unintentionally and without conscious awareness. We use this term in the present study because it combines both the notion of incidental learning and input frequency. According to Ellis (2002), frequency effects can be found in all aspects of second language learning, e.g. in the acquisition of phonology, syntax and lexis. Ellis (2002: 144) points out that "'rules'... are structural regularities that emerge from learner's lifetime analysis of the distributional characteristics of language input". The main argument here is not that language-specific innate structures are necessary to acquire language but rather frequency distribution in the input can provide us with cues to learn language structures. It is important to note that frequency is not the only factor and "moderating effects" (Ellis 2002: 147) also play a role in language acquisition. The question is what these moderating effects are and, for example, what the role of language transfer from L1 with regard to learning from frequency in the input is. Gass and Mackey (2002) respond to Ellis (2002) and state that although frequency in the input certainly has an influence on language learning, there are other important factors, such as saliency, the perception of patterns (2002: 253) and transfer from L1 (2002: 256). Arabic, for example, does not have copula and as a consequence Arabic learners of English find it difficult to learn and use this form in a consistent way. Another well-known example is the difficulties encountered in acquiring indefinite and definite articles by L2 learners whose L1s

do not have them (e.g. Chinese). Even Chinese learners at a high proficiency level in English seem to struggle with the use of articles because they are lacking in Chinese and this cannot be overcome in many cases despite the high frequency of articles in the input (see for example Robertson 2000).

When producing atypical but otherwise grammatical structures, negative feedback is likely not to be given. Thus, the learner relies solely on incidental learning from input frequency. One can assume that the mere fact that the correct structures are frequent in the input does not generally lead to the unlearning of ungrammatical or atypical L2 structures (see Gass and Mackey 2002: 255). The lack of direct negative evidence (e.g. correction by teachers) plays a crucial part in the discussion on implicit statistical learning. Some researchers argue that, in the absence of explicit feedback, it is possible that frequency distributions can provide learners with indirect evidence that certain structures are ungrammatical or atypical. Stefanowitsch (2008: 513) argues that "negative evidence can be inferred from the positive evidence in the linguistic input". According to Stefanowitsch, learners compare the expected frequency of a verb with the actual frequency, and if, for example, this verb is never encountered in a transitive construction, the learner will assume that it is intransitive. Likewise, Boyd, Ackerman and Kutaz argue that "learners are able to infer constraints specifying how a word cannot be used by considering how it is used" (2012: 1). Again, input frequency plays a crucial role in this context. For example, the verbs *disappear* and *vanish* are both intransitive, but *disappear* is more frequent and therefore learners are much more sure that it cannot take an object. As a consequence, overgeneralisations in a learner's language where these intransitive verbs take an object are much more frequent for *vanish* than for *disappear* (Boyd and Goldberg, 2011: 56). Thus, the frequency of correct input has an influence on the learners' grammatical judgements without any negative evidence. The frequency of *disappear* as intransitive verb in the input "blocks" (Boyd and Goldberg, 2011: 61) its use as a transitive verb. This *blocking* is less effective with less frequent verbs, such as *vanish*. Boyd and Goldberg (2011) argue that statistical pre-emption explains this type of learning. When children hear new words (e.g. the non-existing verb *cham*) in certain constructions, e.g. "The cow is chaming (intransitive) and Ernie's making the cow cham (periphrastic causative)", they are unlikely to use this verb in a transitive construction (Boyd and Goldberg 2011: 60). Thus, certain positive input is computed in an unconscious way to replace negative input and pre-empt the use of ungrammatical structures. One has to bear in mind that these examples are from children's first language acquisition. However, there are also studies that confirm that pre-emption also plays a role in adults (Boyd et al. 2012). In the domain of motion events, it can be assumed that if the learners are sensitive to the frequency in the input, they are likely to produce the type of motion verb which is highly frequent in the input available to them, be it a manner verb in the case of S-languages or a path verb in

V-languages. However, the case is more complicated with boundary crossing motion events (see Treffers-Daller & Calude, 2015).

One study which includes the notion of statistical learning with regard to the boundary crossing constraint is Treffers-Daller and Calude (2015). They found that adult learners of French with English as L1 are sensitive towards the frequency of motion verbs in the input and that their use of target like structures increases with higher proficiency, but that the learners at all levels fail to acquire the boundary crossing constraint in French because of lack of negative evidence. The English sentence "John runs into the house" is not the equivalent of the French sentence "Jean court dans la maison". The latter sentence means that John/Jean runs around inside the house because French does not licence manner of motion verbs with boundary crossings (Treffers-Daller and Calude 2015: 607). The learners in Treffers-Daller and Calude's study fail to acquire the boundary crossing constraint in French even though it is frequent in the input. Another study with English learners of Spanish (Larrañaga et al, 2012) shows a similar picture where the learners do not acquire the boundary crossing constraint in Spanish even at a higher proficiency level. This is in contrast to the findings of Cadierno and Ruiz (2006) who found that L1 plays only a limited role in advanced second language acquisition. The studies mentioned above are quite the reverse situation to our study as these learners of French or Spanish fail to acquire a constraint of the target language because they do not have this constraint in their L1, whereas our learners need to unlearn a constraint from L1 if they use the target language. One can assume that this is an even more challenging learning task since the learners in the present study produce grammatically correct sentences in the target language and are therefore very unlikely to receive negative feedback, whereas feedback in the case of English learners of Spanish or French is more likely as they produce sentences that have a different meaning than intended. Arab learners of English do not only need to notice the frequency of the motion verbs but to retreat from overgeneralizations that are transferred from their L1. They should notice that patterns like "run into" when describing boundary crossing scenes are frequent, but L1 patterns "enter running" are not expected in English. Hence, the frequency of manner verb with boundary crossing should block overgeneralization form L1. The fact that these L1 patterns are deeply entrenched makes us assume that frequency in the input is insufficient to overcome this difficulty. The present study seeks to investigate whether the boundary crossing constraint transferred from L1 can be overcome through input frequency and whether implicit statistical learning takes place with increasing proficiency.

3. Research questions

The following research questions are based on the discussion in the literature about the differences between Arabic and English in the expression of *Manner* in boundary crossing events. Research questions 1 and 2 refer to the differences between L1 and L2 and need to be answered to identify the potential learning burden. Research questions 3 and 4 ask whether and how this learning challenge can be overcome.

3.1. Research questions

1. Are there significant differences in the use of manner verbs in the description of boundary crossings between native speakers of Arabic and English?

2. Are there exceptions from the boundary constraint in Arabic that are similar to that in other languages, e.g. uncontrolled movement that are seen as "punctual acts" such as jumping?

3. To what extent does L2 proficiency influence the learnability of manner verbs with boundary by Arabic-speaking learners of English?

4. What role does input frequency and "implicit statistical learning" play in the acquisition of manner verbs expressing a boundary crossing?

4. Methodology

4.1. Participants

The participants in the present study are two groups of Arabic-speaking EFL learners and two control groups of native speakers of English and Arabic. The mean age of the English native speakers (n = 20) is 19.5 (4 males, 16 females), and that of the Arabic native speakers (n = 20) is 31.6 (1 male, 19 females). The first group of EFL learners consists of 34 participants in a pre-sessional course at a British university (mean age = 28.38; 19 males, 15 females). They are on an intermediate to upper intermediate level with IELTS scores ranging from 4.5 to 6.0. The second EFL learner group are 30 postgraduates at an advanced level with IELTS scores ranging from 6.5 to 8.5. Their mean age is 31.6 (9 males, 21 females). The imbalance in the gender distribution is partly due to the fact that monolingual Arabic speakers with limited contact to English are mainly found in the female Arabic population in the UK. As for their regular use of English, the learner groups report using the language between

'sometimes' and 'very frequently'. It is expected that since both learner groups live in the UK, they will have encountered the use of manner verbs in boundary crossing in the input in their daily contact. However, this is not explicitly taught and a contrastive analysis between the two languages is not part of their curriculum. Teachers normally give no negative feedback when the learners use correct structures, e.g. "he went into the house", where native speakers would say "he ran into the house" when a running figure is shown on a picture. Any learning of these structures seems to have taken place implicitly.

4.2. Measures

The material used was a free description task where participants were supposed to describe 12 pictures with boundary-crossing events. The task was designed and used by Özçalışkan (2015) and used by Cadierno (2010). The boundary-crossing events were of three types: 4 pictures depict a movement INTO a bounded space (e.g. *running into the house*), 4 pictures show a movement OUT OF a bounded space (e.g. *flying out of the cylinder*), and in the last 4 pictures the displacement is OVER a line or plane (e.g. *crawling over a carpet*). This particular material has been used because both *Manner* and *Path* components are salient in the pictures. A list of the figures with the different motion events is given in Table 1 below (see also Appendix).

Table 1. List of boundary-crossing motion events

	Type of Motion Event	Event Description
1	INTO a bounded space	Run into a house
2	OUT OF a bounded space	Fly out of a cylinder
3	OVER a plane or line	Crawl over a carpet
4	INTO a bounded space	Dive into a pool
5	OUT OF a bounded space	Dash out of a house
6	OVER a plane or line	Flip over a beam
7	INTO a bounded space	Tumble into a net
8	OUT OF a bounded space	Creep out of a house
9	OVER a plane or line	Leap over a hurdle
10	INTO a bounded space	Crawl into a house
11	OUT OF a bounded space	Sneak out of a pot
12	OVER a plane or line	Jump over a cliff

4.3. Procedure

The pictures were presented to the participants in the order found in Table 1 above. The participants were first introduced to the cartoon character Adam and then were asked to write a few sentences to describe what Adam is doing. The words for the landmarks in the pictures such as *house*, *pot* and *carpet* were provided and the learners were advised to use these words. The advanced Arab EFL learners and Arabic native speakers were met individually. The intermediate Arab EFL learners and English native speakers completed the task in a class setting. It took the participants 10 to 20 minutes to finish the task. The participants also filled in a background questionnaire prior to the task to obtain background information about the participants such as their age, how much they use English in their daily life and how long they live in the UK.

5. Data Analysis

In a first step, we analysed the 12 motion event construals according to an adaptation of the classification of Cadierno (2010) and Özçalıkan (2015) and put them into 6 different categories. An overview is given in Table 2.

Table 2. Event construal patterns

Category	Example
Manner verb + satellite	Ran into
Neutral verb + satellite	Go out
Neutral verb + manner adjunct Manner verb + neutral verb	Go into the class running Run and go into
Path verbs	Enter/ exit
Path verbs + manner adjunct or manner verb + path verb	Enter xxx running Run and enter
No boundary crossing or implicit boundary crossing	Go away from the carpet

After coding the data according to the categories in Table 2, two types of analysis were conducted: one where the verb type for each boundary crossing event (12 events) was computed for the native speakers in order to examine the effect of the boundary crossing event type on the use of manner verbs. In the other type of analysis, the number of the boundary-crossing events that fell in the classification above was computed for the learners in order to investigate the preferred event construal pattern.

A third analysis was also performed in order to explore the effect of input frequency on the descriptions of the investigated boundary crossing scenes. The frequency of all the manner verbs with boundary crossing (Manner verb + satellite) by the learner groups was compared to the frequency of the same patterns by Arab and English native groups and the frequency of these patterns in the input available to the learners which was obtained from the British National Corpus.

6. Results

In presenting the results, first, how native speakers of Arabic and English encoded the boundary crossing scenes is explored. Next, how learners encode the boundary-crossing situations is examined in comparison with the native groups. A comparison between the learner groups is also included. Finally, the role of input frequency is investigated.

6.1 The lexicalization patterns by the native groups

In order to find out whether native speakers of Arabic and English have different preferences for the 12 events, a comparison between these two groups was made on the basis of the 6 categories listed in Table 1. Because some of the categories have low frequencies, an analysis with a Chi^2 test would not have been appropriate. We therefore opted for the Fisher exact test. The test ("exact command" in SPSS with Monte Carlo option) produces for tables larger than 2x2 p-values based on the tests statistics of the Fisher-Freeman-Halton test, which is an extension of the original Fisher-exact test for larger tables. The results are given in Table 3.

Table 3. A comparison between the event construal by Arabic (n = 20) and English (n = 20) native speakers

Event	Value test statistics	p-value	Effect size (Cramers's V)
Event 1	36.998	< .001	.915
Event 2	30.857	< .001	.860
Event 3	29.260	< .001	.810
Event 4	Ns		
Event 5	44.252	< .001	.975
Event 6	Ns		
Event 7	Ns		
Event 8	39.065	< .001	.916
Event 9	Ns		
Event 10	36.108	< .001	.903
Event 11	29.957	< .001	.853
Event 12	10.035	< .01	.503

For four events, there is no significant difference between the two native speaker groups. For event 4 and 9, both groups use mainly manner verbs (event 4: NE: 20; NA: 18 manner verbs; event 9: NE: 18, NA: 16 manner verbs), for event 6 both groups avoid the boundary crossing (NE: 13; NA: 10 instances of avoidance) and for event 7 both groups mainly use simple Path constructions (NE: 10; NA: 12 path verbs). Event 4 describes a figure that dives into a pool, and event 9 describes a figure that jumps over a hurdle, both of which will be discussed later. One event (number 6) was probably difficult to interpret. Some participants said the figure was dancing and they probably did not see a boundary crossing there. For event 7, both native groups preferred the path verb (fall) over the manner verb (tumble), probably because the *Path* component is more salient than the *Manner* component in this event. For all other events, the differences between the two native speaker groups are significant with a large effect size (Cramer's V > .5; see Cohen 1988). For further analysis, the events with non-significant differences are excluded because they either are not a clear prompt for a boundary crossing (e.g. event 6) or pose no learning burden for Arab EFL learners as there is no difference in the event construal between L1 and L2 (e.g. event 4 and 9) . For the other events, a detailed overview of the preferred event construal by the native speaker groups is given in Table 4.

Table 4. Preferred event construal by native speakers (number in bracket = mode based on the categorisation in Table 2)

Event	Preferred construal Native Speaker of English	Preferred construal Native Speaker of Arabic
1	Manner verb + satellite (18)	Avoidance or implicit boundary crossing (9)
2	Manner verb + satellite (20)	Path verbs + manner verb or manner adjunct (9)
3	Manner verb + satellite (14)	Avoidance or implicit boundary crossing (11)
5	Manner verb + satellite (19)	Path verbs + manner verb or manner adjunct (13)
8	Manner verb + satellite (18)	Path verbs + manner verb or manner adjunct (18)
10	Manner verb + satellite (19)	Path verbs + manner verb or manner adjunct (11)
11	Manner verb + satellite (17)	Simple path verb (13)
12	Manner verb + satellite (19)	Manner verbs (11)

This means that the native speakers of English consistently use manner verbs + satellite (e.g. run into), whereas the native speakers of Arabic either avoid the description of the boundary crossing or use path verbs with or without adjunct (run and enter, enter running), for the exception of event (12) which will be discussed later, e.g:

daXala Adam i:la l-bait raki:Dan

entered Adam to the house running

Adam entered the house running

6.2 Lexicalization patterns among the Arab EFL learner groups

The further analysis focuses on the Arab EFL learners and we conflate the 6 categories used so far into a simple dichotomy: manner verb with satellite versus other event construal. Table 5 gives an overview over the number of event contruals with a

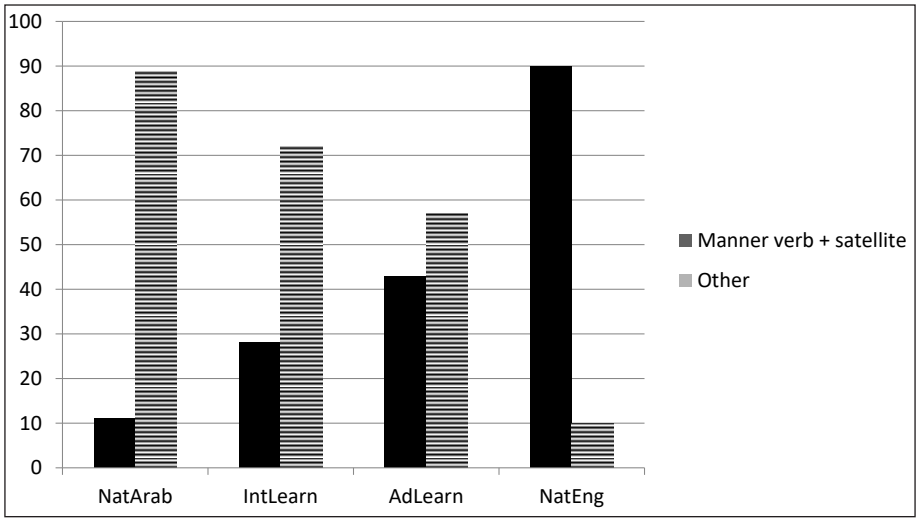
manner of motion verb plus a satellite or other alternative construals (which include categories 2 to 6 found in Table 2) for the four groups of participants. We excluded events from this computation where there was no significant difference between the event construals by native speakers, i.e. event 4, 6, 7, and 9.

Table 5. Event construal with manner of motion verbs + satellite and alternative construals

	Native speakers of English (n = 20)		Advanced Arab Learners (n = 30)		Intermediate Arab Learners (n = 34)		Native speakers of Arabic (n = 20)	
Event	Manner verb + satellite	other	Manner verb + satellite	Other	Manner verb + satellite	Other	Manner verb + satellite	Other
1	18	2	5	25	3	31	0	20
2	20	0	13	17	14	20	3	17
3	14	6	10	20	3	31	0	20
5	19	1	13	17	9	25	0	20
8	18	2	20	10	11	23	1	19
10	19	1	14	16	7	27	1	19
11	17	3	7	23	7	27	2	18
12	19	1	22	8	22	12	11	9

As mentioned before, there is a clear pattern where the use of manner verbs is preferred by native speakers of English and the use of other event construals is preferred by native speakers of Arabic. The patterns used by the learners lie between the preferred patterns of the native groups. Figure 1 visualises the preferred patterns of the four groups.

Figure 1. Preferred event construals by the four groups (percentages)



Overall, the differences between the four groups are significant (Fisher exact = 2156.976, $p < .001$, Cramer's V = 1.0). Again we used the "exact" command (Monte Carlo option) in SPSS, which produces the value for the Fisher-Freeman-Halton test for tables larger than 2x2. All differences between the individual groups are also highly significant at the .001 level (Native speakers of English and Advanced Learners: Fisher exact = 519.466; Advanced Learner and Intermediate Learners: Fisher exact = 686.824; Intermediate Leaners and Native speakers of Arabic: Fisher exact = 550.523). Figure 1 indicates a learning process, where learners at a lower level start with event construals that are closer to the Arabic native patterns, and where with increasing L2 proficiency more constructions are used that come closer to the English native-like pattern without reaching the native-like level totally.

6.3 The role of input frequency

As explicit teaching of these structures does not take place, only the learning difficulty of the patterns and the possible frequency of these structures in the input can explain the learning process or the lack of it. The differences between the two native speaker groups can explain in part the magnitude of the learning difficulty. This is an indication on how deeply entrenched these structures are in the L1. As an indication of the possible frequency of the target structures in the input (manner verb + satellite) such as 'run into', we use frequency data from the British National Corpus (BNC). We used three predictor variables: native Arabic use, native English use as found in the data, and the data from the BNC in two multiple regressions. The

dependent variable was the actual use by the intermediate and by the advanced learner groups. The results are given in Table 6.

Table 6. Predicting the use of manner verbs with boundary crossings (multiple regressions, method "enter")

Dependent variable Use of manner verbs with boundary crossing by	Anova	Explained Variance (R^2)	Predictor variables	Standardized Beta	Sig
Intermediate Learners	F (3, 16) = 72.571	.919	NatEng	.403	t = 5.312, p < .001
	p < .001		NatArab	.738	t = 10.072, p < .001
			BNC	-.063	t = -.880, p = .392
Advanced Learners	F (3, 16) = 37.831	.853	NatEng	.759	t = 7.442, p < .001
	p < .001		NatArab	.357	t = 3.625, p < .01
			BNC	-.152	t = -1.592, p = .131

Multicollinearity is not a problem as in both cases the value for Tolerance is greater than .2 and the Variance Inflation Factor is smaller than 10 (see Field, 2005: 175). Only the native Arabic and the native English patterns are significant in both models and the frequency based on the BNC does not predict the patterns used by the learners.

7. Discussion

Research question 1 about the differences between Arabic and English in the construal of motion events with boundary crossings can be answered on the basis of our findings. The results of the present study show a clear difference between native

speakers of English and Arabic. In line with expectations, Arabic native speakers use significantly fewer manner verbs with boundary crossing in their L1. Our findings lend support to Slobin's (2004) conclusion on thinking-for-speaking patterns found in depicting boundary-crossing motion events across S-languages and V-languages. Research question 2 which asks about exceptions from this categorical dichotomy can also be answered. In line with the literature, the avoidance of manner verbs by Arab native speakers does not hold for all events in the present study. This supports the revised typology of Beavers, Levin and Tham (2010), who show that in many languages both V-framed and S-framed patterns occur, but that languages differ in the extent to which these patterns can be found. In the present study, speakers of Arabic and English show clear tendencies but no exceptionless rules. A more fine-grained picture of these tendencies can be drawn through the analysis of event 4 which depicts a figure diving into a pool and event 9 which describes a figure jumping over a hurdle. In both cases, both native groups use mainly manner verbs. This is in line with similar event construals for Portuguese (Slobin 1997, 2004) and Spanish (Naigles et al. 1998). There seems to be a universal rule for speakers of different V-languages that manner verbs are licensed in these contexts. Naigles et al. (1998) argue that this is the case because the actor only initiates the act and that the actual boundary crossing is out of his/her control. However, in our case (event 4 and 9), the figure clearly plunges into the pool and leaps over the hurdle on purpose and the boundary crossing is clearly intended. We are therefore inclined to follow Slobin's (2004) explanation that a boundary crossing in a punctual act might be the reason for this exception. One might wonder why this is not applied to event 12, which depicts a punctual act (a figure jumps over a gap), where a significant difference between the native speakers of Arabic and English is found. A closer examination of the Arabic native speakers' description of this event, however, shows that the Arabic speakers produce more manner verbs than path verbs (11 manner verbs as opposed to 9 *Path* constructions), a result which shows that the use of manner verb is licensed in this type of boundary-crossing event. This is also in line with Slobin's (2004) interpretation. Further research is needed here with similar punctual acts in horizontal and vertical directions in order to get further insights into this pattern of exceptions across languages.

Research questions 3 and 4 ask about the factors that might influence the learnability of the English target structures. Two factors are investigated, the proficiency of the learners and the frequency of the relevant structures in the input. After excluding the events with punctual acts where the native groups show no significant difference, we are left with the events that seem to pose a learnability issue for the learners. For these events, our findings show that Arab EFL learners use significantly fewer manner verbs with boundary crossings than English native speakers. Instead, they tend to use simple path or neutral verbs or avoid the description of the boundary-crossing at all.

It seems that in the motion events where native speakers of English and Arabic show different tendencies, the production of manner verbs with boundary crossings by Arabic-speaking learners of English seems challenging. This can provide evidence for the influence of the learners' L1 in describing motion events with boundary crossings which goes in line with the results of Cadierno (2010), Alonso (2011), Larrañaga et al (2012), and Treffers-Daller and Tidball (2015). Moreover, this holds for different proficiency levels as both learner groups, the intermediate and the advanced, use fewer manner verbs in this context than English native speakers. There are, however, some indications that learning took place as the advanced learners use more manner verbs than the intermediate learners but both groups are significantly different from the native speakers. Therefore, it can be assumed that the difficulty in encoding manner verbs with boundary crossing motion events is not limited to the intermediate Arab learners but remains persistent with the advanced learners. This result is similar to the results of Larrañaga et al. (2012) in which they show that describing boundary crossing scenes is problematic for English learners of Spanish even in a later stage due to L1 influence. One has to bear in mind that in the current study the advanced learners are at a highly proficient level with IELTS scores above 6.5 and both learner groups live in the UK which increases the probability of exposure to the target structure in the input. This is an indication that the L1 patterns in the description of boundary crossings cannot be overcome simply by exposure to input frequency in L2. Even though the use of manner verb in depicting motion events with boundary crossing is assumed to be frequent in the L2 input, learners seem to fail to notice that patterns such as "run into" occur but not "enter running". Hence, input frequency could not help the learners to pre-empt L1 patterns and use manner verbs with boundary crossing in a native-like way. Through examining the role of frequency in the input, the two multiple regressions support this conclusion. The only significant variables that predict the use or the lack of use of manner verbs in boundary crossings are the preferences of the two languages involved. Frequency of the potential input as measured with data from the BNC is not significant in these regression models. The high explained variance (R^2) in the dependent variable solely on the basis of native Arabic and native English patterns shows how deeply ingrained these patterns are, which supports earlier findings for other language pairs (Treffers-Daller and Calude, 2015). Therefore, we can assume that implicit statistical learning based on frequencies is not possible for the patterns under discussion. Teacher feedback would be necessary to acquire these structures. Such negative feedback is, however, normally not given as the learners produce grammatically correct structures albeit different from the preferences of the target language. The findings of the present study are not only relevant from a linguistic viewpoint but also have clear pedagogical implications. Although foreign language teaching programmes cannot take every difference between L1 and L2 into account, it would be possible to include

the use of manner of motion verbs in boundary crossings in these programmes as this is a universal phenomenon and applies to many learners.

8. Conclusion

The analysis of the descriptions of motion events with boundary crossings reveals that English and Arabic native speakers behave differently where Arabic shows a preference to encode the traversal of boundary with path verbs normally without expressing *Manner* information and English systematically uses manner verbs with *Path* encoded in a satellite. Consequently, the boundary crossing constraint in L1 seems to affect the use of manner verbs by Arabic-speaking learners of English. The study shows that even Arabic-speaking learners with high proficiency in English tend to have difficulties in the use of manner verbs with boundary crossings which suggests that proficiency in general is not enough to overcome L1 preferences and to adopt target-like patterns. Through the analysis of the role of the input frequency as measured by the BNC, the study also shows that input frequency is not a significant factor in the learnability of manner verbs. Hence, implicit statistical learning from exposure to the input seems to be hardly possible. Future research is needed to investigate whether it is possible to overcome deeply ingrained cognitive constraints through explicit instruction in a classroom setting.

References

Alonso, R. A. (2011). The translation of motion events from Spanish into English: A cross-linguistic perspective. *Perspectives*, *19*(4), 353-366.

Alonso, R. A. (2018). Translating motion events into typologically distinct languages. *Perspectives*, *26*(3), 357-376.

Aske, J. (1989). Path predicates in English and Spanish: A closer look. *Proceedings of the Berkeley Linguistics Society, 15*, 1-14.

Attwood, T. & Treffers-Daller, J. (under review). Unlearning the boundary crossing-constraint.

Beavers, J., Levin, B., & Tham, S. W. (2010). The typology of motion expressions revisited. *Journal of linguistics*, *46*(02), 331-377.

Boyd, J. K., Ackerman, F., & Kutas, M. (2012). Adult learners use both entrenchment and preemption to infer grammatical constraints. In *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)* (pp. 1-2). IEEE.

Boyd, J.K. & Goldberg, A. E. (2011). Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language*, *87*(1), 55 – 83.

Cadierno, T. (2010). Motion events in L2 acquisition: The boundary-crossing constraint in English and Spanish. In Z. Han & T. Cadierno (eds.) *Linguistic relativity in SLA: Thinking for speaking* (pp. 1-33). Bristol: Multilingual Matters.

Cadierno, T., & Ruiz, L. (2006). Motion events in Spanish L2 acquisition. *Annual Review of Cognitive Linguistics*, *4*(1), 183-216.

Cadierno, T., & Robinson, P. (2009). Language typology, task complexity and the development of L2 lexicalization patterns for describing motion events. *Annual Review of Cognitive Linguistics*, *7*(1), 245-276.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hilsdale: NJ Erlbaum.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*(2), 143-188.

Field, A. (2005). *Discovering statistics with SPSS*. London: Sage.

Gass, S. & Mackey, A. (2002). Frequency effects and second language acquisition. *Studies in Second Language Acquisition*, *24*(2), 249-260.

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, *83*(1), 49-79.

Hulstijn, J. H. (2012). Incidental learning in second language acquisition. In C. A. Chapelle (ed.), *The Encyclopaedia of Applied Linguistics*. Blackwell Publishing Ltd.

Kachinske, I., Osthus, P., Solovyeva, K., & Long, M. (2015). Implicit learning of a L2 morphosyntactic rule, and its relevance for language teaching In P. Rebuschat (ed.), *Implicit and Explicit Learning of Languages*, (pp. 388 – 415). Amsterdam/ Philadelphia: John Benjamins.

Kopecka, A. (2006). Typological perspectives. In. M. Hickman & S. Robert (eds.), *Space in languages: Linguistic systems and cognitive categories* (pp. 83-102). Amsterdam/ Philadelphia: John Benjamins.

Larrañaga, P., Treffers-Daller, J., Tidball, F., & Ortega, M. C. G. (2012). L1 transfer in the acquisition of manner and path in Spanish by native speakers of English. *International Journal of Bilingualism*, *16*(1), 117-138.

Milton, J. (2009). Measuring Second Language Vocabulary Acquisition. Bristol: Multilingual Matters.

Naigles, L. R., Eisenberg, A. R., Kako, E. T., Highter, M., & McGraw, N. (1998). Speaking of motion: Verb use in English and Spanish. *Language and Cognitive Processes*, *13*(5), 521-549.

Özçalışkan, S. (2015). Ways of crossing a spatial boundary in typologically distinct languages. *Applied Psycholinguistics, 36*(2), 485-508.

Özçalışkan, Ş., & Slobin, D. I. (2003). Codability effects on the expression of manner of motion in Turkish and English. In A. S. Özsoy, D. Akar, E. Nakipoğku-Demiralp, E. Erguvanlı-Taylan & A. Aksu-Koç (eds.). *Studies in Turkish Linguistics*, (pp.259 – 270). Istanbul: Boğaziçi University.

Rebuschat, P. (2015) (ed). *Implicit and Explicit Learning of Languages*. Amsterdam/ Philadelphia: John Benjamins.

Robertson, D. (2000). Variability in the use of the English article system by Chinese learners of English. *Second Language Research*, *16*(2), 135-172.

Saidi, D. (2008). Typology of motion event in Tunisian Arabic. In M. Kokkonidis (ed.), *Proceedings of LingO 2007* (pp. 196-203).Oxford: Oxford University.

Slobin, D. I. (1987). Thinking for speaking. *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 435-444.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In J. J. Gumperz & S. C. Lenvinson (Eds.), *Rethinking Linguistic Relativity* (pp. 70-95). Cambridge: Cambridge University.

Slobin, D. I. (1997). Mind, code, and text. In J. Bybee, J. Haiman, & S. A. Thompson (eds.), *Essays on Language Function and Language Type: Dedicated to T. Givón* (pp. 437-467). Amsterdam/Philadelphia: John Benjamins.

Slobin, D. I. (2004). The many ways to search for a frog: linguistic typology and the expression of motion events In S. Strömqvist & L. Verhoeven (eds.), *Relating Events in Narrative: Typological and Contextual Perspectives* (Vol. 2, pp. 219-257). Mahwah, NJ: Lawrence Erlbaum Associates.

Slobin, D. I. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann & S. Robert (eds.), *Space in Languages: Linguistic Systems and Cognitive Categories* (pp. 59-81). Amsterdam/ Philadelphia: John Benjamins.

Slobin, D. I., & Hoiting, N. (1994). Reference to movement in spoken and signed languages: Typological considerations. *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society*, 487-505.

Stefanowitsch, A. (2008). Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics*, *19*(3), 513-531.

Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. In T. Shopen (ed.), *Language Typology and Syntactic Description* (Vol. 3, pp. 57-149). Cambridge: Cambridge University Press.

Talmy, L. (1991). Path realization: A typology if event conflation. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, 480-519.

Talmy, L. (2000a). *Toward a cognitive semantics*, Vol I, Cambridge, MA: MIT Press.

Talmy, L. (2000b). *Toward a cognitive semantics*, Vol II, Cambridge, MA: MIT Press.

Treffers-Daller, J., & Calude, A. (2015). The role of statistical learning in the acquisition of motion event construal in a second language. *International Journal of Bilingual Education and Bilingualism*, *18*(5), 602-623.

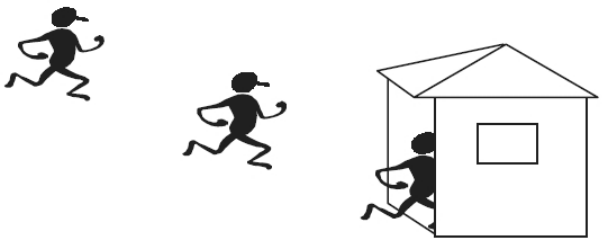
Treffers-Daller, J. & Tidball, F. (2015). Can L2 learners learn new ways to conceptualise events? Evidence from motion event construal among English-speaking learners of French. In P. Guijarro-Fuentes, K. Schmitz and N. Mullar (Eds.), *The acquisition of French in multi-lingual contexts* (pp. 145-184). Bristol: Multilingual Matters

von Stutterheim, C., Bouhaous, A., & Carroll, M. (2017). From time to space: The impact of aspectual categories on the construal of motion events: The case of Tunisian Arabic and Modern Standard Arabic. *Linguistics*, *55*(1), 207-249.

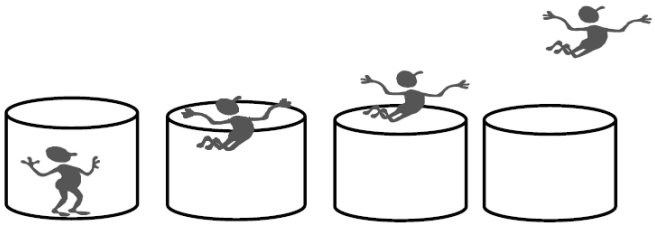
Walk, A. & Conway, Chr. (2015). Implicit statistical learning and language acquisition: Experience-dependent constraints on learning. In P. Rebuschat (ed.), *Implicit and Explicit Learning of Languages*, pp. 191 – 212. Amsterdam/ Philadelphia: John Benjamins.

Appendix

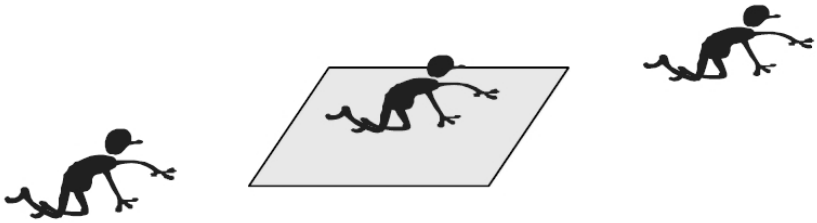
1. Run into a house



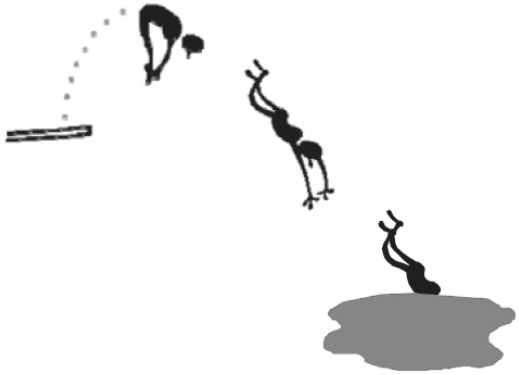
2. Fly out of a cylinder



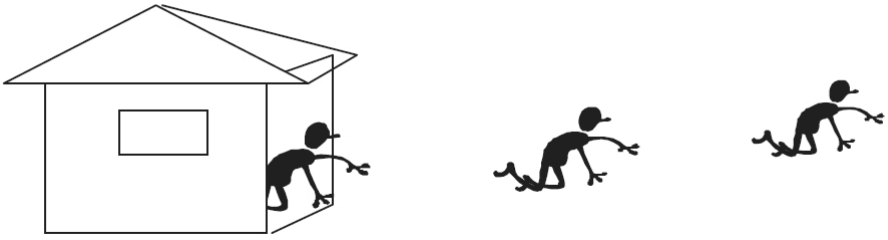
3. Crawl over a carpet



4. Dive into a pool

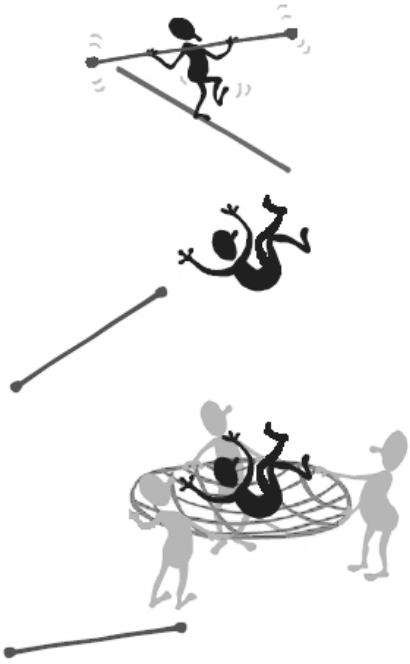


5. Dash out of a house

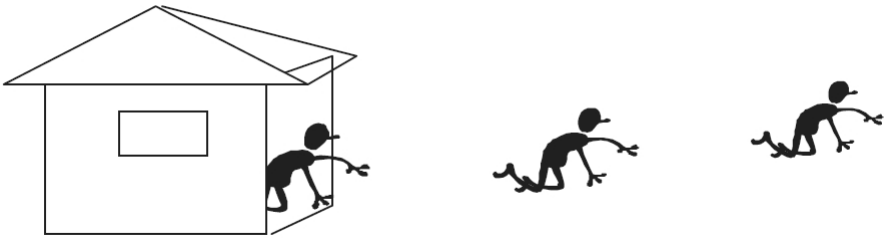


6. Flip over a beam

7. Tumble into a net

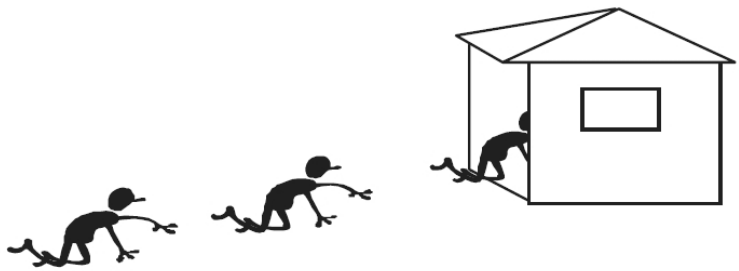


8. Creep out of a house

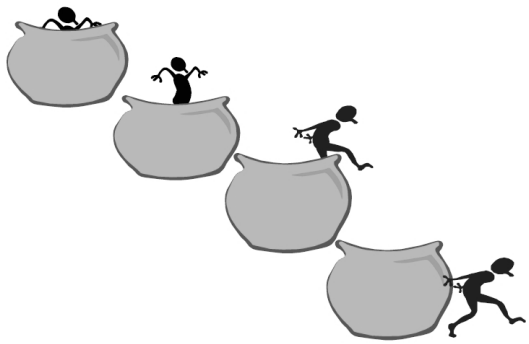


9. Leap over a hurdle

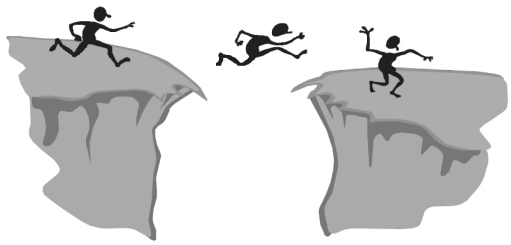
10. Crawl into a house



11. Sneak out of a pot



12. Jump over a cliff



Effects of L2 usage and L1 transfer on Turkish learners' production of English verb-argument constructions

Ute Römer
Georgia State University, USA
uroemer@gsu.edu

Selahattin Yilmaz
Georgia State University, USA
syilmaz2@gsu.edu

Abstract

Using data from the International Corpus of Learner English (ICLE) and the British National Corpus (BNC), this article examines what Turkish learners of English know about a set of frequent verb-argument constructions (VACs, such as 'V *with* n' as illustrated by '*I like to go with the flow*') and in what ways their VAC knowledge is influenced by native English usage and by transfer from their first language (L1), Turkish. An ICLE Turkish analysis gave us access to dominant verb-VAC associations in Turkish learners´ English, and provided insights into the productivity and predictability of selected constructions. Comparisons with the BNC and other ICLE subsets (ICLE German and ICLE Spanish) allowed us to determine how strong the usage effect is on Turkish learners' verb-VAC associations and whether Turkish learners differ in this respect from learners of other typologically different L1s. Potential effects of L1 transfer were explored with the help of a large reference corpus of Turkish, the Turkish National Corpus (TNC).

Keywords: Construction Grammar, verb-argument constructions, learner corpus; usage-based SLA, crosslinguistic transfer

Zusammenfassung

Dieser Aufsatz untersucht basierend auf Daten aus dem *International Corpus of Learner English* (ICLE) und dem *British National Corpus* (BNC), was türkische Lernerinnen und Lerner des Englischen über eine Auswahl von Verb-Argument Konstruktionen (VACs, z. B. 'V *with* n' illustriert durch '*I like to go with the flow*') wissen und welchen Einfluss englischer Sprachgebrauch und Erstsprache auf dieses Wissen haben. Eine Analyse von Daten aus dem türkischen Teil von ICLE (*ICLE Turkish*) ermöglichte

uns Zugang zu besonders dominanten Verb-VAC Assoziationen in türkischem LernerInnenenglisch sowie Einsicht in die Produktivität und Voraussagbarkeit verschiedener Konstruktionen. Durch Vergleiche mit dem BNC und anderen Teilen von ICLE (*ICLE German* und *ICLE Spanish*) konnten wir bestimmen, wie stark der Sprachgebrauchseffekt auf die Verb-VAC Assoziationen türkischer Lernerinnen und Lerner ist und ob es diesbezüglich Unterschiede zwischen türkischen Lernerinnen und Lernern und Lernerinnen und Lernern anderer typologisch unterschiedlicher Sprachhintergründe gibt. Mögliche interlinguale Interferenzeffekte wurden mit Hilfe eines umfangreichen türkischen Referenzkorpus, dem *Turkish National Corpus* (TNC), untersucht.

Schlüsselwörter: Konstruktionsgrammatik, Verb-Argument Konstruktionen, Lernerkorpus, gebrauchsbasierter Zweitspracherwerb, interlinguale Interferenz

1. Introduction

Research within the framework of Construction Grammar suggests that learning languages requires building a network of constructions, that is, conventionalized form-function associations, which can differ greatly from one language to another (Goldberg, 2006; Goldberg & Casenhiser, 2008; Hilpert, 2014). Specifically, verb-argument constructions (VACs) have received a great deal of attention as they are considered the "basic means of clausal expression in a language" (Goldberg, 1995: 3). For second language (L2) learners, typological differences between their first and second language may pose a challenge to their learning of constructions in the target language. Recent research on VACs in the language of advanced L2 learners of English of different first language (L1) backgrounds, for instance, shows that L1 Spanish learners produce VACs including 'verb plus preposition plus noun (phrase)' constructions (e.g. 'V *over* n' exemplified by '*she jumped over the fence*') with less target-like verbs than L1 German and L1 Czech learners at the same proficiency level (Ellis, O'Donnell & Römer, 2014; Römer, O'Donnell & Ellis, 2014). This especially applies to the production of directed motion constructions which are realized differently in verb-framed and satellite-framed languages (Cadierno, 2008, 2013; Slobin, 2004; Talmy, 2000). While a verb-framed language such as Spanish or Italian expresses the path of motion in the main verb (e.g. *saltar* in Spanish), a satellite-framed language such as English or German expresses the path of motion in a separate particle (the satellite) and manner in the main verb (e.g. *jump over* in English). To help learners of typologically different L1s in their acquisition of English, it is hence important to better understand these learners' knowledge of VACs and examine which verbs they most commonly associate with a particular construction.

Studies that have examined VACs in learner production data, both collected from learner corpora and in psycholinguistic experiments, have indicated that (i) L2 learners, especially at intermediate and advanced levels of proficiency, possess verb-constructional knowledge, that (ii) learners' VAC production is affected by verb frequencies in usage (the target language input they receive), that (iii) there is significant overlap between learners' and native speakers' verb-VAC associations, and that (iv) differences in VAC usage between L1 and L2 speakers can be explained on the basis of crosslinguistic transfer effects from the learners' first language (Ellis & Ferreira-Junior, 2009; Ellis, O'Donnell & Römer, 2014; Ellis, Römer & O'Donnell, 2016; Eskildsen & Cadierno, 2007; Römer & Garner, under review; Römer, O'Donnell & Ellis, 2014; Römer, Roberson, O'Donnell & Ellis, 2014; Römer, Skalicky & Ellis, 2018). These studies have focused on English language learners from a range of L1 backgrounds, including Czech, German, Italian, and Spanish.

However, to our knowledge, there are no related studies on L1 Turkish learners' productive knowledge of English VACs (with the recent exception of Babano□lu 2018), or studies that compare constructions in Turkish learners with those produced by learners from other L1 backgrounds. Given the typological properties of Turkish verb morphology (further described in Section 2), VACs that encode directed motion and include prepositions can be particularly difficult to acquire for Turkish learners. The goal of this article is therefore to use data from a corpus of L1 Turkish learner writing to gain a better understanding of what intermediate to advanced Turkish learners of English know about a subset of frequent VACs of the 'verb plus preposition plus noun (phrase)' type, and whether/how this knowledge is affected by L2 (target language) usage and by L1 transfer (Gass & Selinker, 1983; Jarvis, 2011, 2013; Jarvis & Pavlenko, 2008). The research questions we aim to address are:

RQ 1: How productive and how predictable are selected VACs in Turkish learner English compared to German and Spanish learner English?

RQ 2: In terms of dominant verb-VAC associations, do selected VACs in Turkish learner English differ from those in German and Spanish learner English? If so, in what ways?

RQ 3: Is the distribution of verbs in a set of high-frequency VACs in Turkish learner English influenced by English usage? Is this potential influence of usage stronger for Turkish than for German and Spanish learners?

RQ 4: Are there any noticeable effects of the first language on Turkish learners' use of English VACs?

Before we describe how these research questions are addressed methodologically, and what answers our analyses provide in response to them, we will provide some information on typological aspects of the Turkish language. This information offers the necessary background for our discussion of L1 Turkish learners' productive knowledge of English VACs and how it may be affected by crosslinguistic transfer.

2. Properties of verb constructions in Turkish

In Turkish, a verb-framed language, the path of motion is expressed within the verb, while manner is expressed separately and usually given less focus (Slobin, 2014). This typological difference leads to a less frequent expression of manner unless it is "the salient information in the discourse context" (Özyürek & Kita, 1999: 507). For example, the Turkish equivalents of *climb up* (*tırman-*) and *go down (in-)* focus on the path expressed in the main verb with manner being of secondary importance. However, Jessen (2014) notes that Turkish is different from most verb-framed languages in that path can also be expressed using other items, sometimes in combination with the main verb, such as nominals, adpositions (prepositions or postpositions), and case marking. Both Özyürek and Özçalışkan (2000) and Jensen (2004) also comment that L1 Turkish speakers may describe a motion event by using two verbal clauses, the first one being the subordinate clause that expresses manner, the second one being the matrix clause expressing path. For instance, in order to express *crawl up,* an L1 Turkish speaker is most likely to say *ascend by crawling (emekleyerek çık-).* Therefore, highlighting the different ways of expressing motion, Beavers, Levin, and Tham (2010) claim that it is hard to speak of a two-way typology with clear boundaries between the two.

Comparative research on the verbal expression of motion by L1 Turkish and English speakers has shown that these typological differences between the two languages lead to differences in the use and frequency of manner and path (Jessen, 2014, Özyürek & Kita, 1999; Özyürek & Özçalışkan, 2000; Özyürek, Kita, Allen, Furman & Brown, 2008; Slobin, 2004; Toplu, 2011). Nevertheless, much as these typological differences play a large role, there are also certain universals of expressing motion such as merging path and manner (Allen, Özyürek, Kita, Brown, Furman, Ishizuka, Fujii, 2007) and manner-dominant conceptualization of motion regardless of languages (Toplu, 2011).

Most of this comparative Turkish-English research made use of verbal production tasks that required the participants who were advanced learners to respond to audio-visual cues. In their bilingual production studies (L1 Turkish, L2 English), Demirtaş (2010) and D. Yilmaz (2018) showed that advanced Turkish learners were able to use manner verbs and path satellites to a great extent with only limited L1 influence such as using subordinating elements and avoiding the use of verbs that do not

express motion in Turkish. However, interestingly, the L1 Turkish production of the participants was found to include frequent use of manner verbs, which both authors attributed to the influence of the L2 (English) on the L1. Durun (2015) and Duruk (2016) also reported a similar finding on advanced Turkish learners, namely that they used path satellites very commonly like L2 English speakers with a limited amount of L1 influence affecting the use of path-dominant verbs (e.g. ENTER, CROSS). Looking at task and proficiency factors as well as L1 influence, İşler (2014), however, found that the participants' L1 led to limited use of path elements in general. While proficiency positively correlated with path expression in the results of the written task, gestures seemed to have played a much more important role in expressing path in spoken production.

In a study based on learner production data, Babanoğlu (2018) investigated the use of manner of motion verbs (e.g. WALK, RUN, FLY) and path verbs (e.g. ENTER, PASS, ARRIVE) across the L1 German and L1 Turkish subsets of ICLE. The study showed that L1 Turkish writers used both path and manner verbs significantly less than the German writers did. It also showed that the German writers made use of a remarkably higher use of satellites to express path especially with the verb RUN.

In addition to these typological aspects that affect the expression of motion, it is important to mention some general complexities of Turkish morphology. As documented by Durrant (2013), morphemic co-occurrences are much stronger in Turkish than lexical ones, which makes this a distinctive feature of constructional patterns in Turkish. In addition, English prepositions are known to pose difficulties for even advanced learners independent of the L1 background. English as a Lingua Franca (ELF) research has documented non-conventional uses of prepositions in contexts where English is used for oral communication by speakers of different L1s (Cogo & Dewey, 2010; Seidlhofer, 2004). Due to the combined influence of complex morphology in general (Durrant, 2013) and of expressing motion in particular (Jensen, 2014; Özyürek & Çalışkan, 2000), Turkish learners could have particularly marked challenges. For example, Çabuk (2009) and Özışık's (2014) analyses of the use of prepositions in Turkish learner English showed that Turkish learners tend to rely on L1 transfer, thus using prepositions that are similar to their L1, however mostly erroneously, such as *marry with someone* instead of *marry someone*. They also observe an overuse of the preposition *in* by Turkish learners. Likewise, differences between the two languages were found to pose challenges for learners of Turkish as a foreign language because of the multi-layered morphology (Özdemir, 2011). Thus, analyses of 'verb plus preposition plus noun (phrase)' constructions could shed light on several aspects of L1 influence on the L2 usage of Turkish learners.

3. Data and methods

To address our research questions, we collected VAC data from the Turkish subsection of the International Corpus of Learner English (ICLE Turkish; Granger et al., 2009). ICLE Turkish consists of argumentative essays written by intermediate to advanced level EFL learners who were undergraduate students majoring in English at three universities in southern Turkey. Overall ICLE Turkish contains 280 texts and 199,173 words. We also used data from previous analyses of VACs in L1 German and L1 Spanish learner writing and in native English usage. The learner writing for these analyses came from the German and Spanish subcomponents of ICLE (henceforth ICLE German and ICLE Spanish), whereas the 100-million-word British National Corpus (BNC; Burnard, 2007) served as a proxy for L1 English usage. In terms of text type and size, ICLE German (236,095 words) and ICLE Spanish (198,109 words) are comparable to ICLE Turkish. However, while the German learners who contributed to ICLE are mostly at advanced (C1) proficiency level, Turkish and Spanish texts in ICLE come from learners at intermediate and advanced levels (B1 to C1).

In addition to the L1 and L2 English language data, we also retrieved data from the Turkish National Corpus (TNC; Aksan & Aksan, 2009). The TNC is a 50 million-word general corpus of Turkish that includes mostly written data (98%) from a wide variety of genres ranging from scientific texts to fiction produced between 1990 and 2009. Modeled after the BNC and claiming a high degree of representativeness in terms of coverage, the corpus includes a proportional amount of texts from different domains and topics (Aksan, Aksan, Koltuksuz, Sezer, Mersinli, Demirhan, Yılmazer, Kurtoğlu, Atasoy, Öz, & Yıldız, 2012).

From ICLE Turkish, we exhaustively retrieved instances of the following 19 VACs, all covered in previous VAC analyses (Ellis, O'Donnell & Römer, 2014; Römer, Roberson, O'Donnell & Ellis, 2014) and originally selected from the *COBUILD Grammar Patterns* volume on verbs (Francis, Hunston & Manning, 1996): 'V *about* n', 'V *across* n', 'V *after* n', 'V *against* n', 'V *among* n', 'V *around* n', 'V *as* n', 'V *between* n', 'V *for* n', 'V *in* n', 'V *into* n', 'V *like* n', 'V *of* n', 'V *off* n', 'V *over* n', 'V *through* n', 'V *towards* n', 'V *under* n', and 'V *with* n'. We used the search interface provided on the ICLE corpus CD-ROM to extract the selected VACs from ICLE Turkish. Since the corpus is part-of-speech tagged, we were able to search for combinations of a verb (ICLE tags Vbe, Vdo, Vhave, Vlex, Vmod) directly followed by a preposition (*about*, *across*, etc.). We exported the resulting concordances to Excel for manual filtering for true hits of each VAC. In this filtering process we made sure that, in each concordance line, the word following the verb was used as a preposition and that the preposition was followed by a noun or noun phrase. For each VAC, we then created a lemmatized, frequency-sorted list of the verbs that occur in it.

Together with similar datasets based on ICLE German, ICLE Spanish, and the BNC that we generated in previous studies on the same VACs, the frequency-sorted ICLE Turkish verb lists served as the basis for type-token comparisons, and also as input for two types of quantitative analyses: normalized entropy analysis and correlation analysis. Normalized entropy (H_{norm}) is a measure of how uncertain a probability distribution is, in our case the distribution of verbs in a VAC (Kumar, Kumar & Kapur, 1986). H_{norm} values range from 0 to 1 with values closer to 1 indicating a more even distribution which makes it hard to predict what the verb in a new token of a VAC might be. On the other hand, H_{norm} values closer to 0 indicate an increasingly uneven and predictable distribution of items (here verbs), potentially with one or two types making up the lion's share of all tokens. Entropy has been shown to be less sensitive to Zipfian frequency patterns than type-token ratio (Eeg-Olofsson & Altenberg, 1994; Ellis & O'Donnell, 2014; Gries & Ellis, 2015). The correlation analysis allowed us to systematically compare for individual VACs how strongly its verb distributions overlap or differ among learner groups (e.g. ICLE Turkish vs. ICLE German), and between a learner group and L1 usage (e.g. ICLE Turkish vs. BNC), resulting in six types of comparisons (ICLE Turkish vs. ICLE German, ICLE Turkish vs. ICLE Spanish, ICLE German vs. ICLE Spanish, ICLE Turkish vs. BNC, ICLE German vs. BNC, ICLE Spanish vs. BNC). For each comparison, we calculated Pearson correlation coefficients (*r*) in R (R Development Core Team, 2017). We also used R to generate scatterplots that allowed us to visualize verb distributions and highlight which verbs contributed to a high or low correlation value. All calculations were based on the log10 transformations of the verb token frequencies. For selected verbs, we also carried out a more qualitative analysis of the top verb choices across datasets. Data retrieved from the TNC through concordance and collocation searches was used in tracing potential crosslinguistic transfer effects that may have had an impact on Turkish learners' verb usage in specific VACs.

4. Results and discussion

4.1 Productivity and predictability of VACs in Turkish learner English

We addressed RQ 1 by calculating type-token ratios and normalized entropy values for the selected VACs in ICLE Turkish and comparing the values with the same data retrieved for ICLE German and ICLE Spanish. Since entropy values vary across types of constructions (lower values do not always mean "better" or "more proficient"), we also calculated those for each target VAC in the BNC. The BNC values will serve as a reference point for comparison.

Table 1 gives an overview of the type numbers, token numbers, and type-token ratios for verbs in the 19 selected VACs across ICLE subsets. The type-token ratios indicate how productive each VAC is, with higher ratios indicating higher productivity. These values, however, of course have to be treated with caution when it comes to VACs with low token frequencies. The first thing we notice is that verb type and token numbers vary considerably across VACs. For ICLE Turkish, token frequencies range from 1 instance ('V *towards* n') to 679 instances ('V *in* n'). For the majority of VACs, token frequencies are fairly low and not robust enough for more detailed quantitative analyses. In terms of type-token ratios of the higher-frequency VACs, 'V *for* n' and 'V *with* n' are more productive than 'V *about* n' or 'V *in* n'. This means that L1 Turkish learners use the former two constructions with a comparatively wider range of verbs than the latter two, even though 'V *in* n' is the most frequent VAC in terms of tokens and more frequent in ICLE Turkish than in ICLE German and ICLE Spanish. Of these four VACs, 'V *about* n' is the most selective in its verb occupancy. We notice similar trends in the ICLE German and ICLE Spanish data. In both corpora, only four out of 19 VACs are used fairly frequently, and type-token ratios are higher for 'V *for* n' and 'V *with* n' than for V *about* n' or 'V *in* n'. VACs tend to be comparatively more productive in ICLE Spanish than in ICLE Turkish and ICLE German. We will see later how similar or different the actual verb profiles for these constructions are across the three L1 groups.

Table 1:.Verb type and token frequencies across VACs and ICLE subsets

	ICLE Turkish			ICLE German			ICLE Spanish		
VAC	Types	Tokens	TTR	Types	Tokens	TTR	Types	Tokens	TTR
V *about* n	31	209	14.8%	48	242	19.8%	42	178	23.6%
V *across* n	2	9	22.2%	6	9	66.7%	N/A	N/A	N/A
V *after* n	16	69	23.2%	4	15	26.7%	1	3	33.3%
V *against* n	7	17	41.2%	24	45	53.3%	13	61	21.3%
V *among* n	4	8	50.0%	4	5	80.0%	5	5	100.0%
V *around* n	9	12	75.0%	10	14	71.4%	10	15	66.7%
V *as* n	30	59	50.8%	30	56	53.6%	30	100	30.0%
V *between* n	13	15	86.7%	14	22	63.6%	11	19	57.9%
V *for* n	87	243	35.8%	91	338	26.9%	78	258	30.2%
V *in* n	152	679	22.4%	165	556	29.7%	163	647	25.2%
V *into* n	11	35	31.4%	62	175	35.4%	25	55	45.5%
V *like* n	25	87	28.7%	1	1	100.0%	2	2	100.0%
V *of* n	26	105	24.8%	35	149	23.5%	44	100	44.0%
V *off* n	2	2	100.0%	19	32	59.4%	4	4	100.0%
V *over* n	12	16	75.0%	29	45	64.4%	6	6	100.0%
V *through* n	10	13	76.9%	26	40	65.0%	15	20	75.0%
V *towards* n	1	1	100.0%	11	14	78.6%	3	3	100.0%
V *under* n	6	26	23.1%	11	14	78.6%	9	18	50.0%
V *with* n	98	311	31.5%	111	307	36.2%	97	269	36.1%

In addition to measuring the productivity of VACs in learner writing, we assessed how predictable each of them is. Normalized entropy scores for each selected VAC and (sub)corpus are provided in Table 2. The lower the entropy score, the more predictable the distribution of verbs in a VAC. For VACs with token frequencies of 10 or less, we did not calculate entropies (labeled "N/A" in Table 2). We see that in native English usage, normalized entropy values for the included VACs range from 0.51 to 0.80. Some VACs, such as 'V *of* n' or 'V *about* n', are much more predictable than others, for example 'V *around* n' or 'V *over* n'. Lower entropy scores tend to correspond with more Zipfian distributions in which the most frequent verb in a

construction takes up the largest share of the VAC tokens (e.g. TALK in 'V *about* n'). When compared to distributions in L1 usage, VACs in learner writing as captured in ICLE are overall much less predictable. For the majority of VACs, normalized entropy values are considerably higher in ICLE Turkish/German/Spanish than in the BNC. In ICLE Turkish, values are highest for 'V *around* n', 'V *as* n', 'V *between* n', 'V *into* n', and 'V *over* n' which means that these VACs are particularly unpredictable, and learners may not yet have a good sense of what the preferred verbs are in these constructions. Instead learners use these VACs with a variety of verbs while not favoring particular ones. With respect to these VACs, L1 Turkish learners are also further away from the usage norm than their German and Spanish peers.

Table 2. Normalized entropy values across VACs and corpora

VAC	BNC	ICLE Turkish	ICLE German	ICLE Spanish
V *about* n	0.55	0.78	0.75	0.74
V *across* n	0.78	N/A	N/A	N/A
V *after* n	0.72	0.46	0.62	N/A
V *against* n	0.77	0.76	0.92	0.68
V *among* n	0.56	N/A	N/A	N/A
V *around* n	0.80	0.97	0.90	0.94
V *as* n	0.72	0.92	0.88	0.77
V *between* n	0.73	0.98	0.91	0.86
V *for* n	0.63	0.88	0.87	0.83
V *in* n	0.67	0.74	0.79	0.79
V *into* n	0.74	0.93	0.88	0.88
V *like* n	0.57	0.79	N/A	N/A
V *of* n	0.51	0.61	0.71	0.86
V *off* n	0.73	N/A	0.95	N/A
V *over* n	0.78	0.96	0.87	N/A
V *through* n	0.74	0.93	0.95	0.96
V *towards* n	0.75	N/A	0.96	N/A
V *under* n	0.61	0.77	0.96	0.81
V *with* n	0.68	0.87	0.87	0.85

4.2 Dominant verb-VAC associations in Turkish learner English

To study dominant verb-VAC associations in Turkish learner language and to determine whether they differ from those in German and Spanish learners, we only focused on VACs which had token frequencies of at least 200 in ICLE Turkish. For VACs with (often considerably) smaller token numbers we found it difficult to determine clear association patterns. This left us with four VACs for a more detailed qualitative analysis: 'V *about* n', 'V *for* n', 'V *in* n', and 'V *with* n'. For these four VACs, Table 3 lists the ten most frequent verbs in L1 usage (BNC) and in the three ICLE datasets. Verbs that appear among the top-10 in usage as well as in a learner top-10 list are italicized. For 'V *about* n' we notice that the two verbs that appear most frequently in this construction, TALK and THINK, are shared across all four datasets. Of the other top-10 verbs in ICLE Turkish, only two (BE and KNOW) overlap with the BNC list. Additional verbs including CARE and COMPLAIN are shared with the ICLE German and ICLE Spanish lists. Two verbs that appear to be Turkish learner specific choices (at least as far as the top-10 verbs are concerned) are MENTION and DISCUSS. While both are verbs that are near synonyms of the strongly attracted verb TALK, their use leads to realizations of the construction ('*mention about* n', '*discuss about* n') that are not common in L1 usage and perhaps even considered unidiomatic. Overall, we see that Turkish learners most strongly associate verbs of communication and cognition with this VAC.

The most dominant verb associations of 'V *for* n' in ICLE Turkish are LOOK, WAIT, and BE, all three of which also appear in the top-10 lists for the other three corpora. Not shared with any of the other top-10 verb lists for this VAC are the verbs STUDY, STRUGGLE, and DO. Another item that appears among the top-10 verbs in this VAC in ICLE Turkish but is considerably more frequently used by German and Spanish learners is FIGHT. For the third focus VAC, 'V *in* n', the most strongly associated verbs in ICLE Turkish are BE and LIVE, both shared across all analyzed datasets. Apart from those two most frequent verbs, there is however little overlap between Turkish learner writing and L1 usage. Only WORK occurs in the ICLE Turkish top-10 and also in the three other lists. BELIEVE is shared across the three ICLE lists but does not appear among the top-10 verbs for this VAC in usage. Interestingly, the third most frequent verb in 'V *in* n' in ICLE Turkish is CHEAT – a verb that is not frequently used in this VAC in any of the other datasets. Another top-10 verb for this VAC in ICLE Turkish that does not occur among the most frequent verb choices in the other corpora is EXIST. We will examine potential reasons for the repeated use of these two verbs in Section 4.4.

For 'V *with* n', six of the top-10 most strongly associated verbs in ICLE Turkish are shared with the BNC list. They include DEAL, AGREE, LIVE, BE, COPE, and START.

Many of these are also frequent in ICLE German and ICLE Spanish. Verbs that are specific to ICLE Turkish and are not used frequently in this VAC by the German and Spanish learners include FACE, STRUGGLE, START, CHAT, and COMPETE. If we look at similarities and differences across all three L1 groups, we notice that, for this VAC, Turkish learners' verb preferences overlap more with those of German than with those of Spanish learners. This is also reflected in the higher correlation value for the ICLE Turkish-ICLE German comparison (r=0.65) provided in Table 4. For the same VAC, the correlation between ICLE Turkish and ICLE Spanish verbs is only 0.55.

The r-values in Table 4 also indicate that, similar to 'V *with* n', Turkish learner's verb associations for 'V *about* n' are slightly closer to German (r=0.79) than to Spanish learners (r=0.76). The high correlation values for 'V *about* n' suggest that all three learner groups have very similar verb associations for this VAC. For the other two high-frequency VACs in our sample, 'V *for* n' and 'V *in* n', correlations are stronger between ICLE Turkish and ICLE Spanish data (r=0.71 and r=0.62) than for ICLE Turkish and ICLE German (r=0.68 and r=0.59). Overall, we observe fairly high correlations across learner datasets for all four VACs which indicates an overall high degree of similarity in verb associations with some qualitative differences (as indicated in Table 3). Figure 1 serves to visualize this overlap in verb preferences between ICLE Turkish and ICLE German contributors for 'V *about* n'. The x-axis displays the logarithmic frequency of the verb type in 'V *about* n' in ICLE German; the y-axis shows the logarithmic frequency of the verb type in the same VAC in ICLE Turkish. If there were perfect overlap in verb choices between the two groups (both in terms of types and tokens), all verbs would be neatly placed along the diagonal through the middle of the plot. We see that for this comparison, this is not the case. However, most verbs appear fairly closely to the diagonal which indicates that both learner groups use them with similar token frequencies. This is also reflected in the high r-value of 0.79. Verbs that appear above (and further away from) the diagonal are markedly more frequent in ICLE Turkish than in ICLE German; verbs that are plotted below the diagonal are comparatively more frequent in ICLE German than ICLE Turkish. The plot confirms our earlier observations on verbs that are shared by both learner groups (e.g. THINK, TALK, CARE), and on the verb MENTION that is particularly frequent in this VAC in ICLE Turkish but less so in ICLE German.

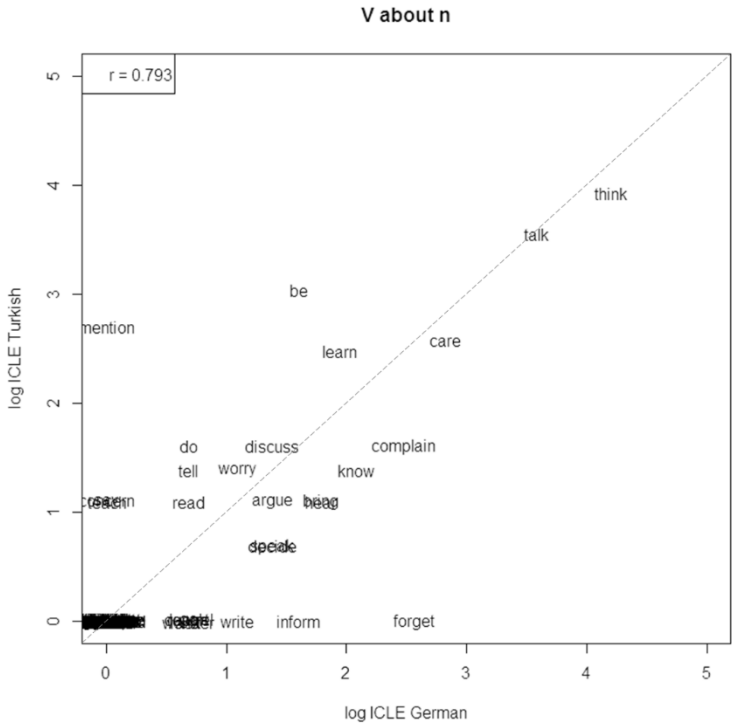
Table 3. Top ten verbs for selected VACs in L1 usage and across ICLE subsets

V about n								
Rank	**BNC**		**ICLE Turkish**		**ICLE German**		**ICLE Spanish**	
1	talk	3832	think	51	think	67	talk	49
2	think	3153	talk	35	talk	36	think	37
3	be	2827	be	21	care	17	CARE	8
4	know	1812	mention	15	forget	13	bring	7
5	worry	910	care	13	complain	12	SPEAK	7
6	say	721	learn	12	know	8	WORRY	7
7	bring	712	complain	5	learn	7	forget	6
8	hear	604	discuss	5	bring	6	know	6
9	forget	556	do	5	hear	6	hear	5
10	write	517	know	4	be	5	be	4
V for n								
Rank	**BNC**		**ICLE Turkish**		**ICLE German**		**ICLE Spanish**	
1	be	13039	look	21	wait	31	look	46
2	look	5242	wait	19	look	28	fight	24
3	wait	4247	be	16	fight	18	ask	15
4	go	3759	work	13	care	17	pay	14
5	ask	2659	study	10	pay	17	wait	9
6	pay	2541	pay	9	be	11	be	8
7	work	2398	struggle	9	ask	10	prepare	8
8	call	2329	do	8	search	10	work	7
9	come	1373	fight	8	blame	8	create	6
10	account	1369	search	8	decide	8	use	6
V in n								
Rank	**BNC**		**ICLE Turkish**		**ICLE German**		**ICLE Spanish**	
1	be	56466	be	158	be	87	be	107
2	live	4784	live	86	live	68	live	59
3	work	3305	cheat	44	sit	28	appear	33
4	go	2897	work	30	believe	25	believe	28
5	sit	2742	believe	15	work	24	happen	25
6	result	2707	happen	15	lie	22	work	17
7	come	2632	occur	14	stay	12	fall	16
8	appear	2175	do	12	happen	8	do	14
9	say	2007	learn	11	keep	8	think	14
10	stand	1951	exist	10	stand	8	show	12
V with n								
Rank	**BNC**		**ICLE Turkish**		**ICLE German**		**ICLE Spanish**	
1	be	8608	deal	32	deal	25	deal	31
2	deal	6407	agree	19	do	20	play	25
3	come	4091	live	18	agree	19	agree	22
4	go	3592	be	14	cope	19	finish	12
5	work	2854	face	13	play	13	happen	11
6	cope	2250	struggle	12	live	10	work	8
7	get	2114	cope	11	work	9	begin	6
8	agree	1844	start	11	associate	8	marry	6
9	start	1641	chat	10	communicate	8	do	6
10	live	1613	compete	7	go	8	be	6

Table 4. Correlations (*r*-values) for verb usage in selected VACs between learner groups (*p*-value for all correlations: < 0.01)

Comparison	V about n	V for n	V in n	V with n
ICLE Turkish vs. ICLE German	0.79	0.68	0.59	0.65
ICLE Turkish vs. ICLE Spanish	0.76	0.71	0.62	0.55
ICLE German vs. ICLE Spanish	0.83	0.75	0.62	0.60

Figure 1: Correlation of verbs in ICLE Turkish and ICLE German for 'V *about* n'



4.3 Examining the influence of native English usage on VACs in Turkish learner English

The correlation analysis of verb distributions across corpora also helped us to address RQ 3 and measure how strongly Turkish learners' VAC production is influenced

by L1 usage. Table 5 provides the *r*-values for comparisons of verb distributions in our four focus VACs between ICLE subsets and the BNC. Overall, *r*-values for these comparisons are much lower than for those discussed in the previous section. This is due to a large extent to the BNC-derived lists being much longer than those retrieved from the ICLE subsets. The result of this was that we included a large number of BNC verbs (data points) in our comparisons for which there were no occurrences in ICLE Turkish/German/Spanish.

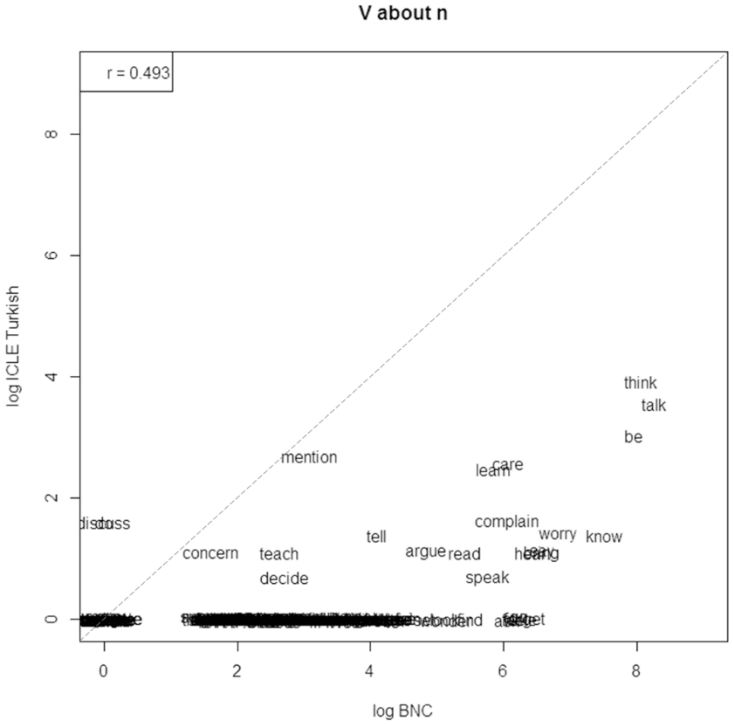
However, we still see that, especially for 'V *about* n' and 'V *for* n', correlations between learner verb-in-VAC production and verbs-in-VACs in usage are not trivial. Correlation values between 0.40 and 0.60 indicate that learners are indeed influenced by frequencies in usage and that their verb-VAC associations overlap with those of L1 speakers. The scatterplot for 'V *about* n' in Figure 2 illustrates that a large number of verbs that occur in this VAC in the BNC are also used by ICLE Turkish writers, even though not with the same relative frequencies. Like L1 speakers, Turkish learners frequently use verbs such as THINK, TALK, BE, CARE, and LEARN in this construction. The plot also confirms that one of the few verbs that occurs comparatively more often in this VAC in ICLE Turkish than in the BNC (plotted above the diagonal) is DISCUSS.

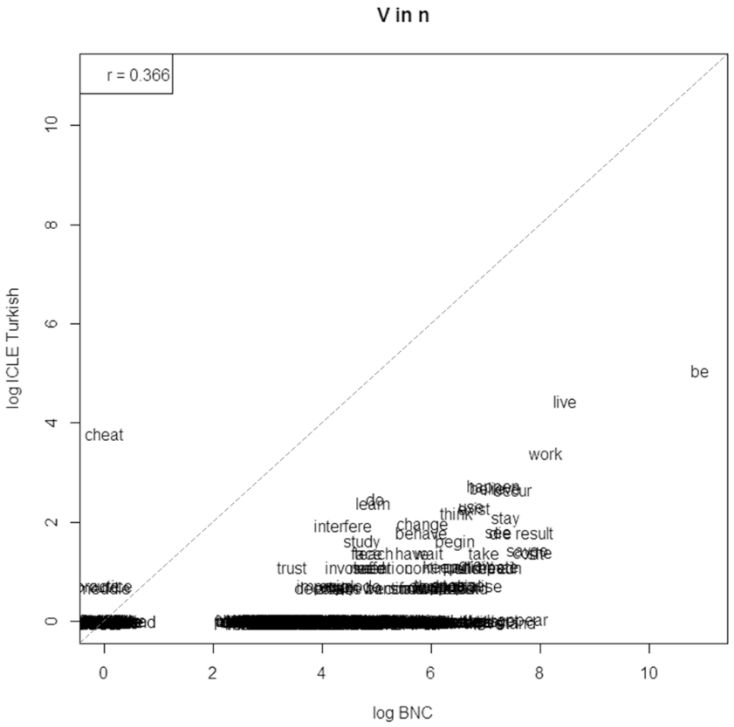
For 'V *in* n' and 'V *with* n', verb distributions in ICLE Turkish are less similar to those in the BNC, as reflected by the somewhat lower correlation values. This may point to difficulties Turkish learners experience with the appropriate use of the prepositions *in* and *with*. The bottom plot in Figure 2 visualizes the correlation between ICLE Turkish and BNC verbs used in 'V *in* n'. Compared to the plot for 'V *about* n', verbs here are plotted closer to the x-axis and further away from the diagonal indicating that Turkish learners' verb associations for this VAC are less in line with L1 usage than those for 'V *about* n'. The plot also highlights a verb we already commented on in the discussion of top-10 verbs for 'V *in* n' (Table 3), CHEAT, which is plotted above the diagonal because it is markedly more frequent in ICLE Turkish than in the BNC data. The influence of L1 usage on learner verb-VAC associations is fairly similar across learner groups for 'V *in* n' (see *r*-values in Table 5). For the three other focus VACs we observe slightly more variation with ICLE German having higher *r*-values than ICLE Turkish and ICLE Spanish for 'V *about* n' and 'V *for* n'. For 'V *with* n' the ICLE Turkish-BNC correlation is stronger than that between the other two ICLE sub-corpora and the BNC.

Table 5. Correlations (*r*-values) for verb usage in selected VACs between learner groups and L1 usage

Comparison	V about n	V for n	V in n	V with n
ICLE Turkish vs. BNC	0.49	0.40	0.37	0.36
ICLE German vs. BNC	0.60	0.49	0.39	0.33
ICLE Spanish vs. BNC	0.56	0.42	0.37	0.28

Figure 2. Correlations of verbs in ICLE Turkish and BNC for 'V *about* n' (top plot) and 'V *in* n' (bottom plot)





4.4 L1 Turkish transfer effects on Turkish learners' use of English VACs

In order to address our fourth and final research question, we examined whether certain verb-VAC association patterns that were found in Turkish learner writing could be the result of crosslinguistic transfer from Turkish, the learners' L1. We did this by extracting translation equivalents of selected verbs and prepositions from the Turkish National Corpus (TNC), and also by analyzing ICLE and BNC concordances of selected VACs in more detail.

Starting with 'V *about* n', we noted that MENTION was among the most strongly associated verbs in this VAC in ICLE Turkish. An ICLE concordance analysis shows that '*mention about*' is used altogether 15 times by 14 different student writers in the ICLE Turkish subcomponent. It also shows that of all ICLE writers only two others (one in ICLE Chinese and one in ICLE Polish) use this combination. The 100-million word BNC contains 22 instances of this verb-VAC combination, 14 of which occur in the spoken and the remaining eight in (non-academic) written English. The high frequency of '*mention about*' in ICLE Turkish has three possible explanations. First, given that '*mention about*' is more common in speech than in writing, the finding may point to a limited register awareness of the L1Turkish writers. Another explanation

would be a tendency of the learners to overgeneralize which verbs are appropriate choices for this VAC and select a verb (MENTION) that is semantically similar to verbs that fit the construction well (including TALK, THINK, WRITE, and ARGUE). Lastly, L1 influence could be another factor. In Turkish, a postposition *–den* (*from*) is attached to a noun (the argument) that precedes the verb MENTION. In the TNC example 'üç evreden bahset-' (*three phases+from mention*), the writer expresses the meaning 'mention three phases'. We counted 838 instances of this '*argument+postposition verb*' construction only in the academic subset of the TNC. The frequency of this construction and the general presence of complex verb-adposition patterns in Turkish usage (see also Jensen, 2014; Çabuk, 2009; Özışık, 2014) may lead to Turkish learners adding in prepositions in English where they are not required or even lead to unidiomatic expressions. The last two of these reasons could also explain the repeated use of '*discuss about*' by Turkish learners (five instances in four different student essays). Like '*mention about*' it can be a semantic extension of core exemplars of the VAC, as well as the result of preposition overuse. There are two equivalent constructions in Turkish: '*noun+-la/le ilgili tartış-*' (*noun+with about discuss*) and '*noun hakkında tartış-*' (*noun about discuss*). Even though we were only able to find a handful of examples of each in the TNC, their existence in their L1 may have affected our Turkish learner writers.

Another interesting observation we made on VACs in ICLE Turkish was the comparatively high frequency of 'V *in* n', the most common of the 19 analyzed VACs in terms of overall token number. This could be in part the result of learners overusing *in* (also reported by Çabuk 2009) because *in* serves as a translation equivalent of several lexicogrammatical items (case markers, adpositions) in Turkish due to its morphological richness being an agglutinative language (Durrant, 2013). Our analysis of the ICLE Turkish 'V *in* n' concordance brought to light several unidiomatic or erroneous uses of this construction, a sample of which we included in Table 6. We can argue that these examples are the result of crosslinguistic transfer from Turkish where translation equivalents of *in* often combine with nouns. For example, Turkish 'şüphe içinde' ('*hesitation in*(locative) *inside of*') might explain the learner's production of '*be in hesitation*' in English. The absence of a verb marker in Turkish may also explain the frequent use of forms of the semantically bleached verb BE in the examples in Table 6.

Table 6. Unidiomatic uses of 'V *in* n' in ICLE Turkish

Left context	Node (V *in*)	Right context (n)
or a place where the answer	is in	the teacher's authority
parents	are in	lack of money
she	may be in	hesitation to go out
they	were in	desire for big amounts of money
if anybody	becomes in	a difficult position
people	come in	a point that
they	survive in	Life
money	will talk in	Everywhere

One of the verbs frequently used in 'V *in* n' in ICLE Turkish but less so in ICLE German, ICLE Spanish, and the BNC was EXIST. A possible reason for this is again crosslinguistic transfer from Turkish where the verb has several approximate translations with a wide range of usages. The most common translation attested in the TNC is '*var ol-*' which literally means 'existent become'. In addition to expressing '*exist*', '*var ol-*' is also used to express the meaning '*there* BE' (existential use; Göksel & Kerslake, 2005, p. 111) with '*ol-*' serving as a support verb or light verb which does not carry much meaning itself (S□odowicz, 2007; Uçar, 2010). This means that an L1 Turkish learner of English may use forms of EXIST instead of using the more common structure '*there* BE'. '*Var ol-*' along with other similar forms and their inflections (*var*, *vardır*, *varsa*, *mevcut*, *mevcuttur*) is highly frequent in the TNC (2,456 instances per million words). In our discussion of the verb lists in Table 3, we also highlighted the verb CHEAT as occurring unusually frequently in 'V *in* n' in ICLE Turkish. A concordance analysis shows that the 44 instances of '*cheat in*' appear in 24 different learner argumentative essays on the value of university degrees. In these essays learners talk about students cheating in tests or exams. The frequent use of '*cheat in*' hence appears to be the result of a task effect.

A final verb we would like to discuss is STRUGGLE which appeared in the top-10 lists for 'V *with* n' and 'V *for* n' in ICLE Turkish but not in those for the other corpora we included in our analysis. The comparatively high frequencies of both '*struggle with*' (12 times in eight learner texts) and '*struggle for*' (nine times in nine texts) again appear to be the result of L1 Turkish crosslinguistic transfer. A common Turkish translation

equivalent of '*struggle with*' is '*noun+-le/-la mücadele et-*' ('*noun+with struggle practice*'). The verb phrase '*mücadele et-*' conveys the literal meaning of '*doing or practicing struggle*'. Among 1,488 instances of *mücadele* in the academic subsection of the TNC, we found 111 examples of the construction. Among the most frequent collocates of *mücadele* in the TNC we also found the Turkish translation equivalent of *for* (*için*). A search for the construction '*noun için mücadele et-*' ('*noun+for struggle do/practice*') retrieved 56 results in the academic subsection of the TNC. While this constitutes evidence of collocational transfer from the learners' L1, we need to keep in mind that, given the typological differences between the two languages, none of the 19 English VACs included in our study have clear one-to-one correspondences in Turkish, which means that our TNC analysis is not exhaustive, despite its level of detail.

5. Conclusion and outlook

Following recent research that looks at learner knowledge of verb constructions from a usage-based perspective and responding to the scarcity of research on patterns in L1 Turkish learner English, the goal of this study was to gain a better understanding of what Turkish learners know about a selection of frequent VACs in English and whether/how their production of these VACs is affected by L2 usage and L1 transfer. The research questions we addressed were:

RQ 1: How productive and how predictable are selected VACs in Turkish learner English compared to German and Spanish learner English?

RQ 2: In terms of dominant verb-VAC associations, do selected VACs in Turkish learner English differ from those in German and Spanish learner English? If so, in what ways?

RQ 3: Is the distribution of verbs in a set of high-frequency VACs in Turkish learner English influenced by English usage? Is this potential influence of usage stronger for Turkish than for German and Spanish learners?

RQ 4: Are there any noticeable effects of the first language on Turkish learners' use of English VACs?

To address RQ 1, we determined type-token ratios (TTRs) and normalized entropy values (H_{norm}) for all verbs used in 19 different VACs in ICLE Turkish and compared them against TTR and H_{norm} values for the same VACs in ICLE German and ICLE Spanish (using the BNC as a reference dataset). We found that type and token frequencies vary considerably across VACs in ICLE Turkish and that only a small subset

of the 19 constructions provide high enough token numbers for systematic quantitative analyses. Of those high-frequency VACs, 'V *for* n' and 'V *with* n' have particularly high TTR values which means that they are more productive in the Turkish learner data than other VACs (e.g. 'V *about* n'). The German and Spanish learner data show similar trends, although with a tendency for ICLE Spanish VACs to be more productive than VACs in ICLE Turkish and ICLE German. The comparisons of entropy scores across datasets revealed that verb distributions in VACs in Turkish learner writing tend to be considerably less predictable (and more even) than in L1 usage, reflected by higher H_{norm} values in ICLE Turkish than in the BNC. The entropy analysis also indicated that for a range of VACs Turkish learners also show lower predictability (and less Zipfian verb-in-VAC distributions) than their Spanish and German peers.

In response to RQ 2, we compared dominant verb-VAC associations for four high-frequency VACs between ICLE Turkish, ICLE German, and ICLE Spanish (with the BNC used for reference purposes). The focal VACs were 'V *about* n', 'V *for* n', 'V *in* n', and 'V *with* n'. While we observed some overlap in the most commonly used verbs in these VACs across datasets (e.g. DEAL, AGREE, LIVE, and BE in 'V *with* n'), we also noted several marked differences between learner production and native usage data, as well as verb choices that appear to be L1 Turkish specific. Examples include associations between 'V *about* n' and the verbs MENTION and DISCUSS, 'V *for* n' and the verbs STUDY, STRUGGLE, and DO, 'V *in* n' and the verb CHEAT, and 'V *with* n' and the verbs STRUGGLE, CHAT, and COMPETE. Results from a correlation analysis which allowed us to systematically compare how similar or different verb distributions in a VAC are between two learner groups (e.g. L1 Turkish vs. L1 German) allowed us to further address RQ2. It showed that for 'V *for* n' and 'V *in* n' Turkish learners' verb distributions are closer to those of Spanish learners whereas when it comes to 'V *about* n' and 'V *with* n' Turkish learners correlate more with German than Spanish learners in their verb choices. Overall, we did not observe any clear language typology related patterns (between Turkish and Spanish as verb-framed languages on the one hand and German as a satellite-framed language on the other) but instead fairly high correlation values across all three L1 groups for the four focus VACs.

Correlation values and scatterplots that visualize the overlap in verb preferences between two datasets also enabled us to address RQ 3. We found that L1 Turkish learners are sensitive to verb-in-VAC distributions in native usage and do not randomly pick verbs as slot fillers in a construction. This is particularly true for 'V *about* n' and 'V *for* n'. While the scatterplots show which of the verbs that are frequent in a VAC in usage also occur in the same VAC in ICLE Turkish, they also highlight verbs that are comparatively more frequent in Turkish learner production than in usage. Through concordance searches in the Turkish National Corpus we were able to gather evidence

to confirm that a likely reason for verb-VAC association patterns that deviated from native usage was crosslinguistic transfer from the learners' L1. This allowed us to respond to RQ 4 with an affirmative 'yes'.

To summarize our findings, we can say that the intermediate-to-advanced Turkish learners who contributed writing samples to ICLE have strong verb-constructional knowledge that is influenced in systematic ways both by native English (L2) usage and by typological and collocational patterns in Turkish (L1). Our findings only partially confirm typological similarities between Turkish and Spanish (both being verb-framed) and point to a more complex picture in which additional morphological aspects (not just the way in which path and manner of motion are expressed) play a role, and which affects L1 Turkish learner production of VACs. Verb framing alone does not explain the patterns of verb-in-VAC usage of this learner group; other lexicogrammatical patterns in Turkish as well as learners' creative semantic expansions of VACs (with near-synonyms of core verbs) also seem to play an important role. These findings support previous research on not only the expression of motion by Turkish learners (Babanoğlu, 2018; Demirtaş, 2010; Duruk, 2016; Durun, 2015; D. Yilmaz, 2018; İşler, 2014), but also on the challenges Turkish learners face in using English prepositions appropriately (Çabuk, 2009; Özışık, 2014), a problem they share with speakers of English as Lingua Franca (Cogo & Dewey, 2010; Seidlhofer, 2004). In our opinion, the creative use of verbs that are semantically related to central verbs in a VAC but not necessarily idiomatic choices, provides further evidence for the discursive hybridity (Mauranen, Perez-Llantada & Swales, 2010; Pérez-Llantada, 2014) of second language writing in which the 'constructicons' of a speaker's first and second language intersect.

Despite the insights it provided, our study has a number of remaining limitations that ought to be addressed in future research on usage-based SLA. The only source of L1 Turkish learner data available to us was ICLE Turkish, which consists exclusively of argumentative essays by intermediate and advanced college level EFL learners and is fairly small by today's standards (we noted in Section 4.1 that only four of our 19 VACs were well attested in ICLE Turkish). It would be valuable to carry out a similar analysis of VACs in corpora of spoken Turkish learner English, larger corpora of written Turkish learner English (ideally capturing a variety of text types), and in longitudinal or cross-sectional corpora that capture Turkish learner English (spoken and written) produced at different proficiency levels. This would allow us to describe not just what intermediate and advanced learners know about VACs but also how this knowledge develops over time. Given the scarcity of existing studies, we would also like to see more research that examines additional constructions in language produced by Turkish learners and learners of other non-European L1s. Our data analysis was limited to a set of 19 VACs of the 'verb plus preposition plus noun (phrase)' type

so our findings may not be representative of Turkish learners' VAC knowledge in general. Expanding this line of research to other types of constructions would hence be extremely valuable. Since some of our previous research on L1 German and L1 Spanish learner knowledge of VACs has highlighted the benefits of combining corpus data with psycholinguistic evidence (Ellis, Römer & O'Donnell, 2016; Römer, Roberson, O'Donnell & Ellis, 2014; Römer, Skalicky & Ellis, 2018), we think that adding an experimental component would strengthen studies on L1 Turkish learner VACs. Finally, while our concordance searches in a Turkish reference corpus (the TNC) led to important insights that helped with the interpretation of some of our findings, it was not exhaustive in that it did not include all possible translation equivalents of English VACs. Future work on learner constructions would benefit from taking a more systematic contrastive approach including analyses of the learner's L1, their L2, and their interlanguage (Gilquin, 2001; Granger, 1996), as it will help uncover sources of marked VAC usage by learners. We look forward to seeing future research that addresses some of these open tasks, all of which we believe will help us to even better understand how second language learners acquire English verb-argument constructions.

References

Aksan, Y., & Aksan, M. (2009) *Building a national corpus of Turkish: Design and implementation* (Vol. 3, pp. 299-310). Working Papers in corpus-based linguistics and language education.

Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U. U., Yilmazer, H. & Kurtoglu, Ö., Atasoy, G., Öz, S., Yıldız, İ. (2012) Construction of the Turkish National Corpus (TNC). *In Proceedings of the 8th International Conference on Language Resources and Evaluation* (pp. 3223-3227) (LREC 2012). Turkey: Istanbul. http://www.lrec-conf.org/proceedings/lrec2012/papers.html.

Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T., & Fujii, M. (2007) Language-specific and universal influences in children's syntactic packaging of manner and path: A comparison of English, Japanese, and Turkish. *Cognition, 102*(1), 16-48.

Beavers, J., Levin, B., & Tham, S. W. (2010) The typology of motion expressions revisited. *Journal of Linguistics, 46*(2), 331-377.

Burnard, L. (2007) *Reference Guide for the British National Corpus* (XML edition). Retrieved from: http://www.natcorp.ox.ac.uk/docs/URG.

Çabuk, S. (2009) The Use of Prepositions in Second Language Acquisition Process (Unpublished Master's Dissertation, Middle East Technical University).

Cadierno, T. (2008) Learning to talk about motion in a foreign language. In P. Robinson & N.C. Ellis (Eds.) *Handbook of cognitive linguistics and second language acquisition* (pp. 239-275). New York: Routledge.

_____ . (2013) Thinking for speaking in second language acquisition. In C. A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*. Malden, MA: Wiley-Blackwell. Accessed 11 May 2018 at http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal1213/full.

Cogo, A., & Dewey, M. (2012) *Analysing English as a lingua franca: A corpus-driven investigation.* Bloomsbury Publishing.

Durrant, P. (2013) Formulaicity in an agglutinating language: The case of Turkish. *Corpus Linguistics and Linguistic Theory, 9*(1), 1-38.

Duruk, E. (2016) Turkish EFL Learners' Naming Motion Events in English. *International Journal of English Language Teaching 4*(1). 47-55.

Durun, E. (2015) Turkish EFL Learners Naming Motion Events in English. *International Journal of English Language Teaching. 4* (1), 47-55.

Eeg-Olofsson, M., & Altenberg, B. (1994) Discontinuous recurrent word combinations in the London-Lund Corpus. In U. Fries, G. Tottie, & P. Schneider (Eds.), *Creating and Using English Language Corpora: Papers from the Fourteenth International Conference on English Language Research on Computerized Corpora*. Amsterdam: Rodopi, 63-77.

Ellis, N. C., & Ferreira-Junior, F. (2009) Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics, 7*, 188-221.

Ellis, N. C. & O'Donnell, M. B. (2014) Construction learning as category learning: A cognitive analysis. In T. Herbst, H.J. Schmid, & S. Faulhaber (Eds.), *Constructions Collocations Patterns.* Berlin: De Gruyter Mouton, 71-97.

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2014) Second language processing of verb-argument constructions is sensitive to form, function, frequency, contingency, and prototypicality. *Linguistic Approaches to Bilingualism, 4*(4), 405-431.

Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016) *Usage-based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar.* Malden, MA: Wiley.

Eskildsen, S. W., & Cadierno, T. (2007) Are recurring multi-word expressions really syntactic freezes? Second language acquisition from the perspective of usage-based linguistics. In M. Nenonen & Niemi, S. (Eds.). *Collocations and Idioms 1. Papers*

from the First Nordic Conference on Syntactic Freezes (Vol. 41). Joensuu, Finland: Joensuu University Press, 86-99.

Francis, G., Hunston, S., & Manning, E. (1996) *Grammar patterns 1: Verbs*. London: HarperCollins.

Gass, S., & Selinker, L. Eds. (1983) *Language transfer in language learning*. Rowley, MA: Newbury House.

Gilquin, G. (2001) The Integrated Contrastive Model Spicing up your data. *Languages in Contrast 3* (1), 95-123.

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. London: Routledge.

Goldberg, A. E. (1995) *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.

_____. (2003) *Constructions: a new theoretical approach to language. Trends in Cognitive Sciences*, 7(5), 219-224.

_____ . (2006) *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, A. E., & Casenhiser, D. 2008. *Construction learning and second language acquisition*. In P. Robinson & N.C. Ellis (Eds.), *A handbook of cognitive linguistics and SLA* . Mahwah, NJ: Erlbaum. 197-215.

Granger, S. (1996) From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Ajmer, B. Altenberg, & M. Johansson (Eds.), *Languages in contrast. Text-based cross-linguistic studies*. Lund, Sweden: Lund University Press. 37-51.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009) *International Corpus of Learner English*. Louvain, Belgium: Presses Universitaires de Louvain.

Gries, S. T. & Ellis, N. C. (2015) Statistical measures for usage-based linguistics. *Language Learning, 65*(S1), 228-255.

Hilpert, M. 2014. *Construction grammar and its application to English*. Edinburgh, UK: Edinburgh University Press.

İşler, Z. N. (2014) *EFL Learner' Use of Path Elements in Motion Event Expressions: A Study on Turkish University Students* (Unpublished Doctoral Dissertation). Middle East Technical University, Ankara, Turkey.

Jarvis, S. (2011) Conceptual transfer: Crosslinguistic effects in categorization and construal. *Bilingualism: Language and Cognition* 14(Special Issue). 1-8.

____ . 2013. Crosslinguistic influence and multilingualism. In Carol A. Chapelle (ed.), *The encyclopedia of applied linguistics*. Malden, MA: Wiley-Blackwell. Accessed 11 May 2018 at http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0291/full.

Jarvis, S., & Pavlenko, A. (2008) *Crosslinguistic influence in language and cognition*. New York: Routledge.

Kumar, U., Kumar, V., & Kapur, J. N. (1986) Normalized measures of entropy. *International Journal of General Systems, 12*(1), 55-69.

Mauranen, A., Perez-Llantada, C, & Swales, J. M. (2010) Academic Englishes: A standardized knowledge? In A. Kirkpatrick (Ed.), *The Routledge Handbook of World Englishes*. London: Routledge. 634-652.

Pérez-Llantada, C. (2014) Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, *14*, 84-94.

Özdemir, M. 2011). Yabanci dil olarak Türkçe öğretiminde edatlar (Unpublished Master's Dissertation). Istanbul University, Istanbul, Turkey.

Özışık, C. (2014) Identifying preposition errors of Turkish EFL students. *ELT Research Journal*, *3*(2), 59-69.

Özyürek, A., & Kita, S. 1999. Expressing manner and path in English and Turkish: Differences in speech, gesture, and conceptualization. In M. Hahn, & S.C. Stoness (Eds.), *Twenty-first Annual Conference of the Cognitive Science Society*. New Jersey: Lawrence Erlbaum Associates. 507-512.

Özyürek, A., & Ozcaliskan, S. (2000) How do children learn to conflate manner and path in their speech and gestures? Differences in English and Turkish. In E.V. Clark (Ed), *The Proceedings of the 30th Stanford Child Language Research Forum*. CSLI Publications. 77-85.

Özyürek, A., Kita, S., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2008) Development of Cross-Linguistic Variation in Speech and Gesture: Motion Events in English and Turkish. *Developmental Psychology*, *44*(4), 1040-1054.

R Development Core Team. (2012) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Römer, U., & Garner, J. R. Under review. The development of verb constructions in spoken learner English: Tracing effects of usage and proficiency.

Römer, U., O'Donnell, M. B., & Ellis, N. C. (2014) Second language learner knowledge of verb-argument constructions: Effects of language transfer and typology. *The Modern Language Journal, 98*(4), 952-975.

Römer, U., Roberson, A., O'Donnell, M. B., & Ellis, N. C. (2014). Linking learner corpus and experimental data in studying second language learners' knowledge of verb-argument constructions. *ICAME Journal, 38*(1), 59-79.

Römer, U., Skalicky, S., & Ellis, N. C. (2018) Verb-argument constructions in advanced L2 English learner production: Insights from corpora and verbal fluency tasks. *Corpus Linguistics and Linguistic Theory*. DOI: https://doi.org/10.1515/cllt-2016-0055.

Seidlhofer, B. (2004) 10. Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics, 24*, 209-239.

Slobin, D. I. (2004) The many ways to search for a frog: Linguistic Typology and the expression of motion events. In S. Strömqvist & L. Verhoeven (Eds.), *Relating events in narrative: Vol. 2. Typological and contextual perspectives.* Mahwah, NJ: Lawrence Erlbaum Associates. 219-257.

Słodowicz, S. (2007) Complement Control in Turkish. *ZAS Papers in Linguistics, 47*, 125-157.

Talmy, L. (2000) *Toward a cognitive semantics: Typology and process in concept structuring.* Cambridge, MA: MIT Press.

Toplu, A. B. (2011) *Linguistic expression and conceptual representation of motion events in Turkish, English and French: An experimental study.* (Unpublished PhD Dissertation). Middle East Technical University, Ankara.

Dubbing attitudes through tonal patterns: when tones speak louder than words

Sofía Sánchez-Mompeán
Universidad de Murcia, Spain
sofia.sanchez@um.es

Abstract

This article explores the conveyance of attitudinal content through intonation in dubbed dialogue and presents the findings from an empirical corpus-based analysis. Research-wise, intonation is hitherto an understudied topic in Audiovisual Translation and has generally taken a back seat in dubbing literature. However, its communicative value and attitudinal function in oral discourse cannot be overlooked when interpreting and producing dubbed speech. The possibility of associating a particular tonal pattern with specific attitudes has enabled the comparison between a number of English original and Spanish dubbed intonation phrases via a speech analysis software. The results obtained provide empirical data on the dubbing of the attitudinal content under analysis and account for the main trends that could negatively affect both the quality of the final outcome and the way the dubbed text is received by the target audience.

Keywords: dubbing, intonation, tonal pattern, attitude, attitudinal content.

Resumen

El presente trabajo explora el contenido actitudinal que se transmite entonativamente en el diálogo doblado y presenta los resultados de un estudio empírico basado en un corpus. La entonación es a día de hoy un campo de estudio minoritario en Traducción Audiovisual y suele ocupar una posición secundaria en la literatura sobre doblaje. Sin embargo, su valor comunicativo y su función actitudinal en el discurso oral ponen de relieve su importancia a la hora de interpretar y emitir el diálogo doblado. La posibilidad de asociar un patrón tonal determinado con una serie de actitudes ha permitido la comparación entre varias frases entonativas en inglés y en

español mediante un programa de análisis de voz. Los resultados obtenidos muestran datos empíricos sobre el doblaje del contenido actitudinal analizado y destacan los factores principales que podrían afectar a la calidad del producto final y a la forma en la que la audiencia meta recibe el texto doblado.

Palabras clave: doblaje, entonación, tono, actitud, contenido actitudinal.

1. Introduction

All languages are melodic in nature as a result of their prosodic variation. As put by Nilsenová & Swerts (2012: 78), "there are no languages in which utterances are produced at a constant tempo, in a straight monotone and with no variation in loudness or voice quality". One of the key ingredients for the production of speech melody is intonation, defined by Crystal (1985: 162) as "the distinctive use of patterns of pitch, or melody". This suprasegmental trait, however, cannot simply be described in terms of melodic patterns. Intonation is part and parcel of oral communication and provides a sound basis for coherent conversation, adding meaning to the words uttered by the speakers. As such, it can be considered a powerful source of information that unveils the addresser's inner feelings and implies the linguistic interpretation of pitch contours on the part of the listener (Cantero, 2002).

Despite its importance in oral discourse, intonation has barely been explored within the remit of Audiovisual Translation (AVT) and, more specifically, within the context of dubbing. One of the reasons behind this lack of popularity amongst AVT scholars might be the difficulty to disentangle the intonational system of a language. In fact, as put forward by Hirst & Di Cristo (1998: 1), "intonation is paradoxically at the same time one of the most universal and one of the most language-specific features of human language". When the attitudinal function of intonation comes into play, such remark becomes even more obvious insofar as different pitch contours can convey different attitudinal meanings and different languages can convey the same attitudes by resorting to different pitch contours (O'Connor & Arnold, 1973).

This paper gives intonation the attention it merits by bringing its significant role in dubbing to the fore. Focusing on the interplay between the interpretation and production of intonation in original and dubbed dialogues, the present research deals with the (un)successful conveyance of attitudes through tonal patterns in the Spanish dubbed version of one of the most popular American sitcoms on TV: *How I met your mother* (CBS, 2005-2014). To the best of the researcher's knowledge, this approach has not hitherto been subject of any empirical study in the area of AVT and Linguistics.

The article starts with a theoretical overview of intonation from a purely functional perspective. The focus is then placed on the attitudinal function of intonation to show how attitudes and tones can converge in speech in search of a successful outcome in dubbing. The next section is devoted to the methods and the bilingual corpus used in the study. An empirical corpus-based analysis will enable the quantitative and qualitative comparison between the source and target utterances under examination in terms of intonation and attitudinal content. This is followed by the discussion of the results obtained in the comparative analysis as well as by the most recurring trends observed in the conveyance of attitudes in the dubbed text. The conclusion emphasizes the relevant role of tonal patterns, which are not simply valuable but fundamental to interpret and reproduce the full richness of the source dialogue.

2. The functions of intonation

The functions that intonation can perform in a language are manifold. Wells (2006) argues that this suprasegmental trait can serve a total of six functions in speech, namely attitudinal, grammatical, accentual (or focusing), discourse (or cohesive), psychological and indexical. The first function —of special interest for the purpose of this article— is related to the attitude intended by the speaker with the use of a particular tone. According to Monroy (2005: 3), sentences consist of a logical side and an affective side that coexist in varying degree in a way that "no utterance in a language is entirely logical nor purely emotional". The prominence given to each side will depend on the speakers' intention when uttering their sentences. The second function of intonation, originally upheld by linguists such as Pike (1945) and Halliday (1967), refers to the link between intonation and grammar when it comes to the distribution of information within the utterance or the disambiguation of grammatically similar structures. Intonation can also be used as a focusing device (the accentual function) by the addresser "for contrastive or emphatic purposes" (Chun, 2002: 59). Thanks to this function, speakers can bring some parts of the utterance into or out of focus and indicate what information is new or what information has already been given in the dialogue. The discourse function, which owes its origins to Brazil (1975), is related to the achievement of coherence and cohesion beyond the sentence level. In this sense, intonation organises the discourse, regulates turn-taking and marks boundaries throughout sequences of tone units. The fifth function proposed by Wells (2006: 12) is the psychological function, necessary to arrange speech in different units "that are easy to perceive, memorize and perform" by speakers. Finally, the indexical function alludes to the crucial role of intonation to exhibit distinctive traits of personal and social identity amongst the interlocutors.

While the number and type of functions tend to vary slightly from author to author, a vast majority regard the expression of attitudinal meaning as the primary

function of intonation in oral discourse (Crystal, 1969; O'Connor & Arnold, 1973; Couper-Kuhlen, 1986; Monroy, 2005; Tench, 2011). This is but a reflection of the inextricable connection between attitudes and tones in speech. The sounds we use, as well as the words we use, are not chosen in an arbitrary way (O'Connor & Arnold, 1973). We follow the patterns we learned as children and resort to them depending on our communicative purposes. Since the number of tones we can use in a language is limited, the kinds of attitudes we can assign to a particular pitch pattern will also be restricted. The possibility of building links between a specific pitch contour and a limited set of attitudes or intentions (Mompeán & Monroy, 2010) allows us to ascribe a particular tone to a particular attitudinal content and a particular attitude to a particular tonal pattern. A wh-question, for instance, can be related to a businesslike attitude if uttered with a default falling tone, whereas it can sound encouraging and kind with the use of a rising tone. Similarly, a statement can convey a sense of detachment and reservation when uttered with a low fall but is bound to sound more casual and involved with the use of a high fall (O'Connor & Arnold, 1973).

This correlation is paramount from the point of view of dubbing. Since the tonal patterns used by the original actors will determine to a great extent the attitude they adopt towards that particular situation, voice talents need to transmit the same attitudinal content in the dubbed version to make the new dialogue "not simply a 'correct' or 'precise' one but, equally important, a *plausible* one" (Mateo, 2014: 130). This is not to say that they are expected to imitate the tones of the on-screen characters, but rather that they are expected to convey the same attitudes by resorting to the repertoire of tones offered by their own language.

Although it goes without saying that intonation is not the only available means to express attitudes, and other prosodic and paralinguistic features as well as kinesics and lexical or grammatical choice can provide important clues about the actors' attitudinal intention, the focus of analysis in the present study is placed on intonation. One of the reasons for this is Monroy's (2005: 2) remark that, amongst all the possibilities for expressing attitude, "from a purely linguistic standpoint, intonation seems to be the basic means to convey attitudinal meaning". A further reason for the choice of intonation as the mainstay of this study is O'Connor & Arnold's (1973: 5) idea that what allows actors to give different interpretations of a role previously played by another actor in a film or theatre play is essentially intonation, insofar as "it does provide important information which is not contained in any of the other features of utterances". The specificity of intonation and the possibility of establishing meaningful ties between tonal patterns and attitudes make this an eligible research topic in dubbing.

3. Dubbing attitudinal content

According to Gottlieb (2005), dubbing is a type of isosemiotic translation of a polysemiotic text. This means that the source and the target texts make use of the same semiotic channel, in this case the verbal auditory channel (i.e., oral speech is rendered as oral speech). However, synchronization between the original and the dubbed versions is not strictly limited to the verbal auditory channel. The nonverbal auditory channel, in which intonation is embedded, and the nonverbal visual channel also need to be taken into consideration in order to achieve a harmonious result. Even though the study of the nonverbal component of speech has taken a back seat in AVT research, especially as compared to the attention devoted to the verbal component, nonverbal features such as intonation occupy a fundamental place in dubbing. Intonation needs to be regarded as an intrinsic part of the audiovisual construct, one that contributes to understanding how characters behave and communicate with each other and that can even modify the semantic content of the utterance (Sánchez-Mompeán, 2016).

One of the tasks of voice talents is precisely to interpret the character's words within the context they have been uttered in. In Mateo's (2014) view, one of the cornerstones for an effective communication is primarily based on the interpretation of the speaker's meaning, which is very often transmitted by intonational features. In order that the original message is interpreted correctly, Cuevas Alonso (2017) emphasizes the need to understand how intonation works and what meaning this suprasegmental feature adds to the whole utterance. He argues that intonation can help dubbing actors opt for a particular interpretation while ruling out other interpretations of the same message. Intonation thus acts as a valuable tool for guiding dubbing practitioners through their task and should be mastered to comprehend and exploit its expressive richness. Voice talents are expected to grasp the attitude intended by the character and, under the supervision of the dubbing director (Whitman-Linsen, 1992; Agost, 1999), reproduce the same attitudinal content in the dubbed version.

The following example (from Crystal, 1975: 3) illustrates how intonation can modulate the attitudinal meaning of the sentence and influence its oral delivery into Spanish. In this case, by modifying the pitch direction (from fall to rise) in a wh-question such as "What are you doing?" the speaker's attitude changes. In order to reflect the intended attitudinal content in Spanish the dubbing actor will have to interpret correctly the attitudinal meanings added by intonation and convey the same attitude in the target language. From a translational viewpoint, the written form of the question would remain the same in both examples, the difference lying in the oral version delivered by the voice talent.

1) a. What are you ↘**doing**? (serious and abrupt)

b. What are you ↗**doing**? (friendly and interested)

Despite the fact that contextual factors as well as facial expression and body language can often help dubbing actors to infer the attitude adopted by the original character, intonation is sometimes the only available tool that leads to the correct answer. In general terms, the unsuccessful rendering of the attitude intended by the speaker might bring about an unnatural rendition, a shift in characterisation or in the comic purpose of the text (Hidalgo Navarro, 2011) and, ultimately, an unfavourable reaction from the target audience. Therefore, the attitudinal function of this suprasegmental feature as well as its essential role in spoken discourse cannot be underestimated or taken for granted in dubbing from both an academic and a professional perspective.

4. Methods

The study proposed here intends to offer a better insight into the dubbing of attitudes through intonation. The aim is to analyse the (un)successful rendering of attitudes in a Spanish dubbed corpus by comparing the tonal patterns used in both the original and dubbed versions. The following research questions will be addressed:

1. Is the attitudinal content attached to the on-screen actors' intonation reflected successfully in the dubbed version?

2. Are any regular patterns or prevailing trends observed in the conveyance of attitudes in the dubbed text?

3. In what ways can unsuccessful renderings affect the target audience's comprehension?

For the purpose of this study, attention will only be directed to those oral extracts where attitudes are predominantly transmitted and reflected through tonal variation. For instance, attitudinal content exclusively conveyed by body language or by other paralinguistic features will be discarded.

Before moving on to the procedure and the corpus used in the analysis, it is worth noting that the aim of this paper is not to judge the dubbing actors' job but to evaluate whether the attitudinal content transmitted intonationally in the source text is rendered in the target version (un)successfully. Although it is the oral discourse as dubbed by the voice talents of the sitcom the one to be assessed here, the dubbing actor is obviously regarded as one more agent of the dubbing process, one who

collaborates with other professionals to make an audiovisual product available in a different language to a different public (Chaume, 2012; Cuevas Alonso, 2017).

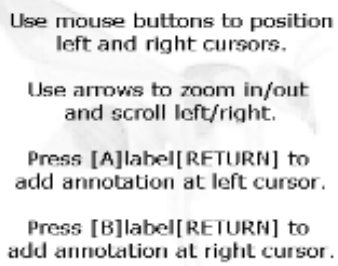
4.1. Design

To measure the (un)successful rendering of attitudes through tonal patterns in the dubbed text, an inventory of nine tones, namely low fall, high fall, low rise, high rise, low level, mid level, high level, fall-rise and rise-fall, was selected. This repertoire of tones suffices to account for the usage and variability of patterns in both languages (O'Connor & Arnold, 1973; Monroy, 2002). The occurrence of these nine tones was examined in four utterance types: statements, questions, exclamations and commands. Drawing on O'Connor & Arnold (1973), Brown et al. (1980) and Cruttenden (1997), who state that some attitudes are more likely to occur with some types of tonal patterns, it was possible to establish a direct correlation between the tone used in a particular utterance type and the attitudinal content related to that tone. This study adheres to the works by numerous experts in the field of intonation in both English (Crystal, 1969; O'Connor & Arnold, 1973; Collins & Mees, 2003; Wells, 2006; Tench, 2011) and Spanish (García-Lecumberri, 1995; 2003; Gutiérrez Díez, 1995; 2005; Monroy, 2002; 2005) to determine which attitudes can be related to a specific tonal pattern. This research is thus underpinned by existing studies that help envisage a taxonomy able to associate a tonal pattern with a repertoire of attitudinal meanings.

4.2. Procedure

The procedure followed in the corpus analysis can be summarised in several consecutive steps. The first stage consisted of the manual transcription of the original and dubbed scripts and the division of the dialogue into intonation phrases (i.e., a portion of spoken material with its own intonational pattern). The second step was to cut the original and dubbed intonation phrases into individual utterances with an online audio cutter and to arrange the samples per utterance type. Then source text and target text fragments were introduced into the speech analysis software SFS/WASP (Waveforms Annotations Spectrograms and Pitch), a very useful program developed by Mark Huckvale from the University College London to obtain the exact pitch contour of a recorded utterance. For illustrative purposes, a SFS/WASP screenshot is provided in Figure 1 below.

Figure 1. SFS/WASP screenshot



The next step of the process included an auditory and visual inspection of contours conducted by the researcher of this study in order to identify the prenuclear and nuclear patterns of each intonation phrase. Taking the aforementioned studies as a starting point, the tonal patterns observed per utterance type in both English and Spanish were associated with one or several attitudinal meanings. Then, the tonal patterns used by the original actors were compared to their dubbed counterparts as far as their attitudinal content was concerned. When the tonal patterns used in Spanish managed to convey the same attitude as the English utterance, the dubbed sentence was judged as a successful rendering. When the tonal patterns adopted by the voice talent failed to reflect the same attitude as the original sentence, the dubbed utterance was deemed as an unsuccessful rendering. The results obtained from the comparative analysis were then quantified and divided into two different groups: one including all those instances that successfully reflected the original attitudinal content and one gathering those cases that failed to transmit the attitude conveyed intonationally by the original character. Finally, data were explored from a qualitative point of view in order to identify regularities or prevailing trends in the conveyance of attitudes in the dubbed text and to draw conclusions on how unsuccessful renderings might affect the target viewers' comprehension.

4.3. Corpus

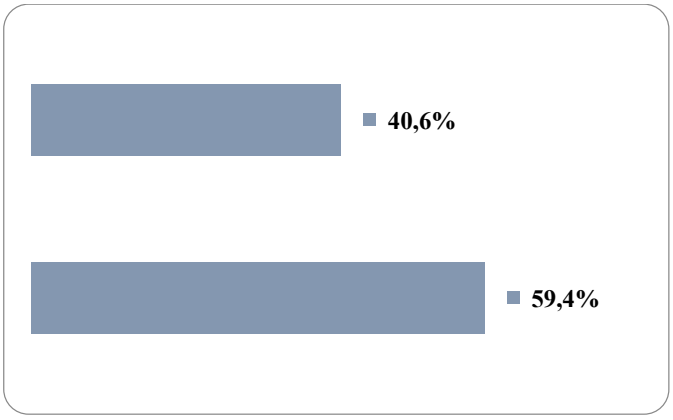
The popular sitcom *How I met your mother* (ST) and its dubbed version into Castilian Spanish (TT) was selected as the corpus of this study. The bilingual parallel corpus under examination comprised a total of 720 utterances extracted from 12 episodes (6 in English and 6 in Spanish) that featured around 270 minutes of fictional dialogue. Only speech produced by the five main characters of the sitcom, namely Ted Mosby, Barney Stinson, Marshall Eriksen, Robin Scherbatsky and Lily Aldrin, was singled out and analysed in this study to ensure consistency. Although, as shown by Sánchez-Mompeán (2016), intonation can also exert a substantial impact on the way

oral discourse is translated into Spanish, the focus of attention in this paper is on the oral rendition of the discourse as delivered by the dubbing actors of the sitcom. As for the genre selected for the present research, the situation comedy was mainly chosen for its proneness to mirror colloquial and spontaneous conversation (Romero-Fresco, 2009; Baños, 2014), its expressive possibilities (Savorelli, 2010) and its comic purpose. As argued by Zabalbeascoa (2001), intonation is indeed a strategic device to reflect colloquialism, to provide the oral text with more expressiveness and to trigger humorous situations.

5. Results and discussion

The total number of intonation phrases analysed was 720 in both languages. After examining the tonal patterns of the utterances under study in the parallel corpus, the comparative analysis revealed a high percentage of dubbed utterances that failed to reflect the attitudinal content transmitted intonationally by the original actor. As shown in Figure 2 below, 214 out of 360 dubbed utterances (59.4%) did not convey the attitude intended by the original speaker, whereas this attitudinal content was successfully delivered in 146 dubbed utterances (40.6%). The findings show that the dubbing actor very often resorts to a tonal pattern that fails to reproduce the attitude intended in the original utterance, maybe because the wrong attitude is inferred from the character's intonation.

Figure 2. Results obtained from the comparative analysis



A notable difference has also been found regarding the (un)successful rendering of attitudes per utterance type. The analysis has shown that the attitudinal content conveyed in the form of exclamatory sentences tends to be delivered successfully in the

dubbed version. A possible explanation for this finding could be the expressive load characterising these types of utterances, which might make it easier for voice talents to grasp and reproduce the attitude transmitted intonationally by the on-screen actor.

Due to space constraints, the focus will be placed here on those dubbed utterances that failed to reflect the original attitudinal content (59.4% of the corpus analysed). After analysing and comparing the bilingual corpus under study, four main patterns were detected in the conveyance of attitudes in the dubbed dialogue: (a) the reduction of the emotional involvement, interest and surprising load, (b) the mitigation of sarcasm and an air of detachment, (c) the use of a more categorical attitude when conveying a questioning tone or a reserving judgment, and (d) the rendition of confirmatory-seeking questions as information-seeking questions. These four trends will be exemplified for the sake of clarification. (The audio-visual extracts under analysis in the present paper are available upon request.)

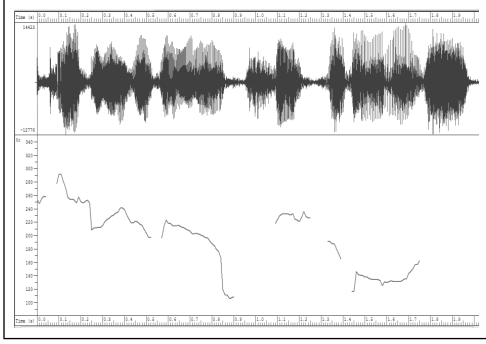
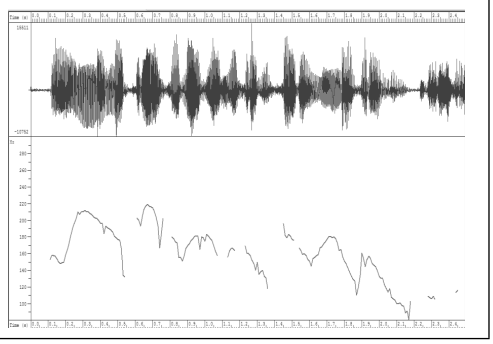
The three examples given below illustrate the first trend found in the analysis. The high falling tone has been replaced either by a level or a low pitch-range, thus reducing the emotional involvement as well as the interest and surprising load in the dubbed version.

In the first example, Ted resorts to a high fall with a twofold purpose: on the one hand, to offer his opinion about the fact that Marshall has accepted to work for a profit-seeking company and, on the other hand, to show interest in his friend's success. Broadly speaking, the fall is the commonest tone for declarative sentences and tends to imply definiteness and completeness. Yet, the use of this tone along with the use of a particular pitch-range and a particular head can add several attitudinal meanings to the sentence (Monroy, 2005). The high fall on the nucleus and the stepping downward head (Monroy, 2002) introduce a note of mild surprise on the part of the speaker and an underlying thought (Tench, 2011), in this case: "I am happy for you, but I think it is ironic that you now take the type of job you always complained about". This tonal pattern thus makes Ted sound mildly surprised, involved and even a bit sarcastic (Tench, 2011). The attitude implied intonationally by Ted triggers the answer given by Marshall, who immediately justifies himself for having accepted that job by admitting that "it's just an internship to make a little money".

The attitude implied by Ted's intonation should be easily inferred from the English audience but might not be so obvious in the Spanish version due to the tone adopted by the dubbing actor. A sustained high head and a low falling contour mitigates the involvement and the surprised attitude attached to the character's intonation and leaves out the sarcastic load. Instead, the Spanish utterance sounds more detached and unemotional, as if Ted's intention was just to express his opinion or belief (Quilis,

1993). The comparison between the tonal pattern used in the original (a) and in the dubbed (b) sentences is shown in the following SFS/WASP screenshots.

Example 1 (season 1 – episode 17)

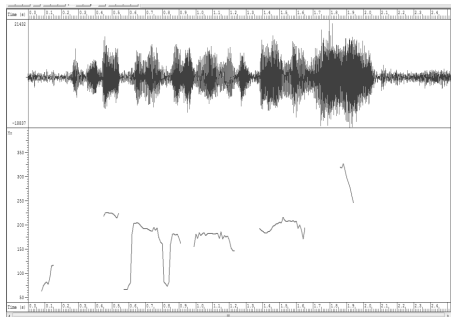
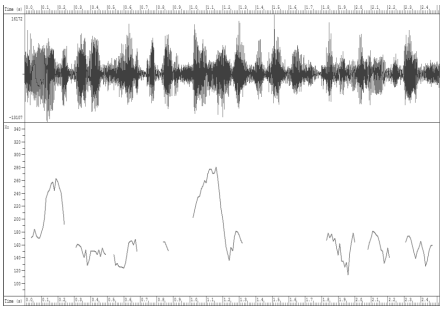
(a) I can't believe you're going all ↘**cor**porate on us.	(b) Aún no me creo que vayas a trabajar en una gran em↘**pre**sa.
	

*Syllable in bold representing the nucleus. **Arrows representing pitch direction (↘ high fall and ↘ low fall)

In the second example, the emotional involvement and excitement of Marshall at having almost reached 200,000 miles driving his old car is inferred from the use of a high head and a high fall (O'Connor & Arnold, 1973) and reinforced by an increase in the character's loudness. Both intonation and loudness play a paramount role to reflect the attitude of the character towards that particular situation. However, the Spanish utterance seems to hint at a different attitudinal meaning. In the dubbed version, the nuclear tone falls by default on the last word of the sentence, given that, as noted by Gutiérrez Díez (2005: 132), the rule in Spanish is to accent the last lexical item "regardless of the nature of such word". This shift in the focus placement along with the use of a (mid) level contour in dubbing make Marshall still sound anxious but much more relaxed and less emotionally involved than in the English version.

Moreover, the addition of laughter at the end of the dubbed sentence, absent in the source text, contributes to mitigating the original strength of feeling (Tench, 2011), which leads to the anxiety brought about in the subsequent shots because the engine stops working and the car fails to hit 200,000 miles. As far as loudness is concerned, whereas the English text (a) increases in intensity up to 85 dB (decibels), it ranges between 67-70 dB in Spanish (b) and hardly rises when uttering the nucleus.

Example 2 (season 2 – episode 17)

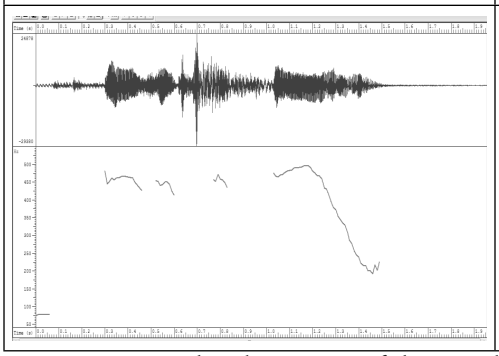
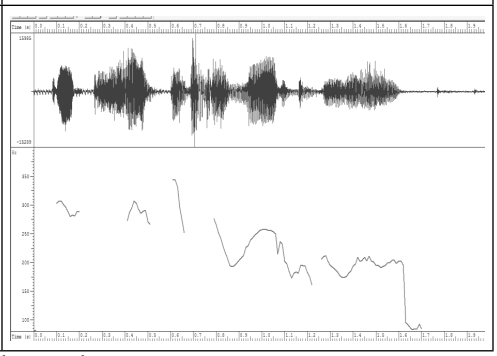
(a) I can't believe this moment is finally ↘**here**.	(b) No me puedo creer que por fin haya llegado este mo→**men**to.
	

* → representing a mid level tone.

A change in the attitude intended by the original character is also notable in Example 3. In this case, Lily's intonation is characterised by a sustained high head and a high fall, a tonal pattern that tends to reflect a deep interest on the part of the speaker and introduces a note of concern and agitation (O'Connor & Arnold, 1973). Lily's attitude is also reinforced here with the elongation of the first syllable of the nucleus (da**:**y). If we analyse this utterance by taking into consideration the visual code as well as the situational context, there is no doubt that Lily's intonation in the original version is determined by Marshall's sad face when he arrives home from his new job. The visuals thus lead Lily (and the audience) to expect a negative comment to her question before listening to Marshall's answer.

In this particular example, the dubbed version could have made use of the same contour as the English question, insofar as this pattern is also the best choice to express the emotional involvement of the speaker in Spanish (Navarro Tomás, 1944). However, the level tone coupled with a low key reduce Lily's involvement and interest in the target dialogue (Mompeán & Monroy, 2010). The attitude of concern and agitation transmitted by the original character in English has been replaced by a more enthusiastic question in Spanish that seems to ignore Marshall's sad face.

Example 3 (season 1 – episode 17)

(a) How was your first ↘**da:**y?	(b) ¿Qué tal tu primer →**dí**a?
	

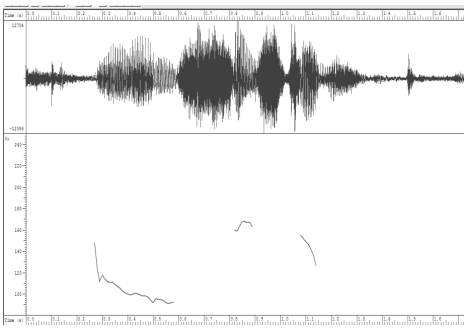
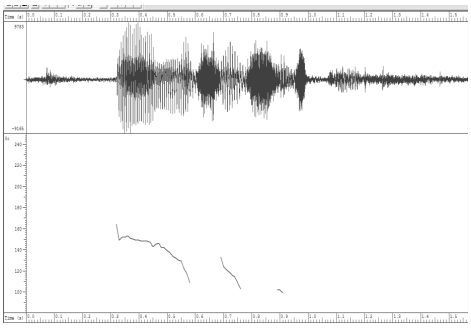
* **:** representing the elongation of the vowel sound.

The second trend detected in the corpus is related to the mitigation of sarcasm and an air of detachment, as illustrated in the following two examples. This trend seems to be highly associated with the replacement of the rise-fall adopted in the original utterance by a low fall in the dubbed version. In fact, the use of the rise-fall in English merits attention due to its implicative load in speech. As described by Collins & Mees (2003: 129), this tone, which is more frequently employed in English than in Spanish (Monroy, 2005), tends to introduce "an implication of an additional but unspoken message". This implication can be introduced in the form of irony or veiled sarcasm (Cruttenden, 2008), as is the case in Example 4.

In this scene, Ted and Marshall take Marshall's car to a local garage to get it fixed. Since Lily wants them to help with the paper folding she is preparing for her wedding with Marshall, she convinces Robin and Barney to go and meet their friends there while they wait for the car to be repaired. Even though Barney goes to the garage against his will, when they arrive he tells Ted and Marshall that the three of them insisted on joining them. The rise-fall adopted to utter the nucleus makes the audience aware that he is clearly being sarcastic and that he was compelled by the girls, thus creating a hilarious situation.

In the Spanish version of the utterance, the falling head and the low falling tone adopted by the dubbing actor successfully transmit the lack of excitement and of enthusiasm contained in the character's words. However, this tone also reduces the sarcastic implication in the dubbed text. As a result, the attitude intended by Barney in the original dialogue sounds less humorous in the dubbed version and the perception gained by the target audience is likely to vary.

Example 4 (season 2 – episode 17)

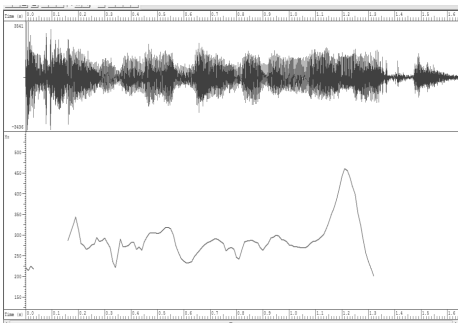
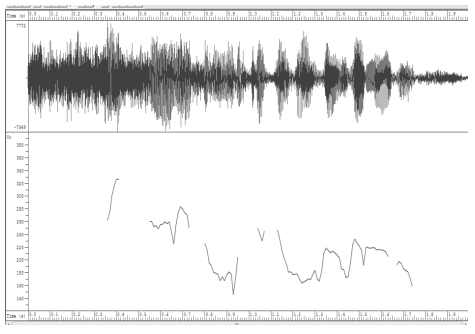
(a) We in^**sis**ted.	(b) Hemos insis↘**ti**do.
	

* ^ representing a rise-fall.

Another case of rise-falling tone is featured in Example 5. The rise-fall preceded by a sustained high head can express feelings of surprise and awe (Halliday, 1970) and even a hint of accusation (O'Connor & Arnold, 1973). Therefore, the use of this tone in the English utterance carries several implications from an attitudinal standpoint. In this scene, Ted tells Robin that he is in love with her and that he is moving to Chicago to put some distance between them. In Robin's use of intonation it is possible to infer that she is highly impressed by Ted's words (Crystal, 1975; Cruttenden, 2008) and even sounds somewhat accusatory (O'Connor & Arnold, 1973). The audience is bound to perceive from her use of intonation that she is not trying to persuade Ted to stay in New York but disagrees with his decision to leave.

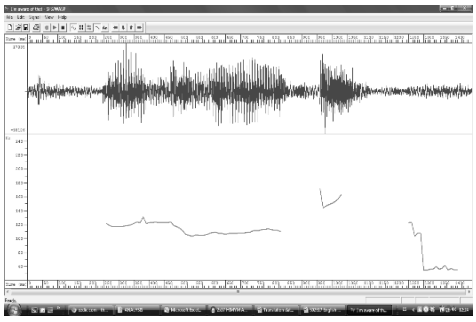
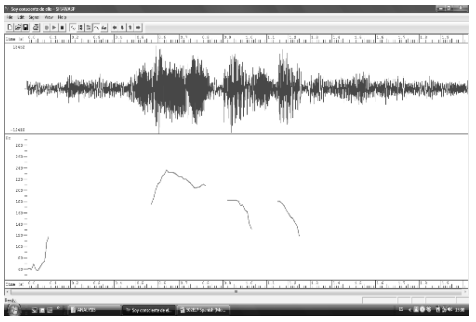
The use of a low fall in Spanish makes the character sound more detached and unsympathetic (Monroy, 2005), as if she was just presenting a fact rather than expressing her inner feelings. This pattern could thus make it difficult for the target audience to interpret the attitude intended by the original actor successfully.

Example 5 (season 9 – episode 17)

(a) I can't believe you're leaving New ↘**York**!	(b) ¡No puedo creer que te vayas de Nueva ↘**York**!
	

The use of a more categorical attitude when conveying a questioning tone or a reserving judgment has also been detected in the corpus under study. In Example 6 below, the low rise at the end of the English utterance preceded by a sustained low head implies a condescending attitude and introduces a reserving judgement (O'Connor & Arnold, 1973). This pattern can also be used to sound warm and gentle and to downgrade the impact of the meaning underlying the speaker's comment (Collins & Mees, 2003). The attitudinal content transmitted by this tonal pattern is perfectly suitable within the situational context of the statement. In this scene, Marshall is forced by his two elder brothers to drive to the coffee shop completely naked. When he arrives there, the drive-through attendant asks Marshall if he realised that he is not wearing any clothes. Although Marshall feels terribly embarrassed, he reserves his real feelings and tries to sound calm and gentle to defuse that tense situation. If the English utterance is compared to its Spanish counterpart, we find that the descending head and the low falling tone adopted by the dubbing actor sound more categorical and less reserved in Spanish than in English, thus creating a different impression about Marshall's attitude in this scene.

Example 6 (season 2 – episode 17)

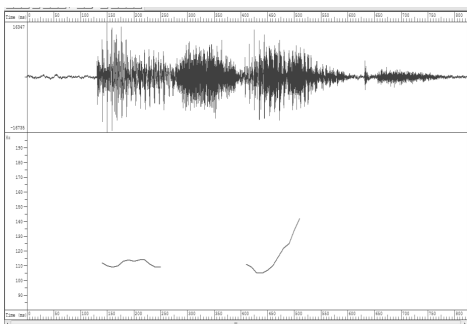
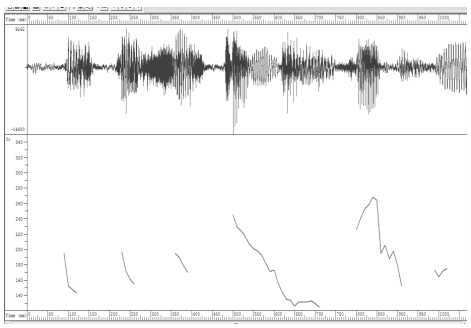
(a) I'm aware of ↗**that**.	(b) Soy consciente de ↘**e**llo.
	

* ↗ representing a low rise.

One of the most remarkable instances representing a shift in the questioning tone of the original utterance is shown in Example 7. In this scene, Ted tells Marshall that Robin kissed him before she took her plane to Russia. According to Wells (2006), yes-no questions generally take a rise and indicate that the speaker is asking for information and expects a more elaborated reply from the hearer (Monroy, 2012) at the same time as showing interest in the listener's answer. In Spanish, however, the questioning tone has been replaced by a more confident and somewhat insistent attitude with the use of a falling tone (Crystal, 1975). This pattern gives the impression that Ted is completely sure about Marshall's viewpoint and that there is no need to ask about his opinion. As a result, Marshall's answer to Ted's question in the original dialogue ("No, I'm not shocked") is to be perceived by the target audience as a contradiction to Ted's assertion in the dubbed version.

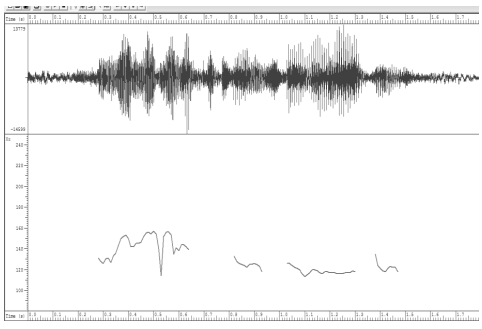
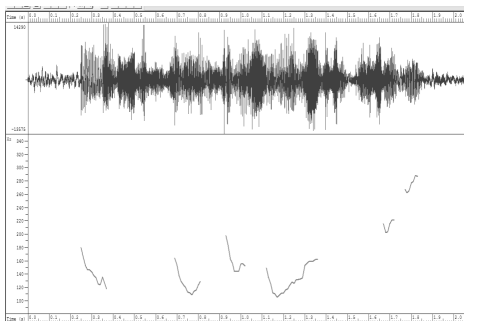
It is difficult to determine whether this change in intonation and, by extension, in the illocutionary force of the utterance under analysis took place in the pre-synchronisation (by the translator or the dialogue writer) or in the post-synchronisation (by the dubbing director or the dubbing actor) stage of the dubbing process. Whatever the case may be, there is no doubt that how characters say what they say exerts a direct impact on attitudinal meaning and should be interpreted and delivered correctly to avoid a potential change in the attitude intended by the original speaker.

Example 7 (season 7 – episode 17)

(a) Are you ↗**shocked**?	(b) Estás escandali↘**za**do.
	

Finally, the rendition of confirmatory-seeking questions as information-seeking questions has been observed in the corpus. In Example 8, given the absence of the subject-verb inversion, intonation holds the key to determining whether the character wants to ask a question or state a fact. In this scene, the speaker, Barney, puts in a request for confirmation by drawing upon a level tone. The level tone accounts here for the questioning nature of the English utterance and the confirmatory-seeking intention of Barney's words. Yet, the high rise in Spanish tinges the interrogative sentence with a note of surprise (García-Lecumberri, 2003) and turns the original confirmatory yes-no question into an information-seeking question (Estebas-Vilaplana & Prieto, 2009). The result is a change in the attitude intended by the original speaker.

Example 8 (season 3 – episode 17)

(a) You're really gonna pretend this never →**hap**pened?	(b) ¿Vas a actuar como si no hubiese pa↗**sa**do?
	

6. Conclusions

Intonation has not been a recurrent research topic amongst dubbing scholars nor has it received the attention it merits as far as training and teaching are concerned. The great relevance that this suprasegmental trait bears in oral discourse and its leading role in the conveyance of attitudinal meaning pave the way for a thorough analysis of how intonation is interpreted and produced in dubbed speech. The focus has been placed here on the (un)successful rendering of attitudes through tonal patterns in Spanish dubbing. The strong correlation that exists between the tone used and the attitudinal content intended by the speaker has invited comparison between the source and target dialogues to explore whether the English actors and the Spanish dubbing actors have adopted the same attitudes by resorting to the tonal repertoire offered in their respective languages.

The results of the comparative analysis show that the attitudinal content intended in the original version was successfully rendered in 40.6% of the cases, whereas 59.4% dubbed utterances failed to reflect the attitude transmitted intonationally by the on-screen characters. The trends observed in the conveyance of attitudes in the dubbed text are as follows: (a) the reduction of the emotional involvement, interest and surprising load, (b) the mitigation of sarcasm and an air of detachment, (c) the use of a more categorical attitude when conveying a questioning tone or a reserving judgment, and (d) the rendition of confirmatory-seeking questions as information-seeking questions. Most unsuccessful renderings are thus related to the use of a tonal pattern that does not reflect the attitude intended in the original utterance and brings about a different interpretation in the target language. As a result, the perception gained by the audience in Spanish and their comprehension can be affected, even more if there is an apparent contradiction between what the characters are saying and what their attitude towards that situation is —or between how characters say what they say and what is shown on screen— or if the intended comic purpose is reduced or completely lost in the dubbed version.

We agree with Mateo (2014) that the capital importance of intonation for both the interpretation and the conveyance of attitudinal meaning stresses the need to incorporate its study into the field of dubbing from both a theoretical and a practical point of view. In fact, the lack of research in this area along with the lack of specific training for dubbing professionals to learn about the singularity of intonation and its relevance in oral speech could have a detrimental effect on the final version delivered in Spanish. Baños (2014: 409) wrote that in the case of translation "the translator takes the role of the scriptwriter, and should thus master the linguistic features available in the target language". If we extrapolate this sentence to the context of dubbing, the voice talent would take the role of the original actor in the Spanish version and should

thus master the tonal patterns available in the target language. It is worth noting, however, that the paramount role played by intonation in dubbing should not be overlooked by any of the practitioners taking part in the process.

Amongst the future avenues of research to expand this topic, reception studies, which have witnessed an upsurge in popularity in recent years, will prove especially useful to achieve more quality in the final outcome and to cast new light on the audience's perception and comprehension when watching an audiovisual product dubbed into Spanish. With a view to teaching future professionals, there is little doubt that intonation should occupy a more prominent place in dubbing courses, and specific training should be given in this area.

References

Agost, R. (1999) *Traducción y doblaje. Palabras, voces e imágenes*. Barcelona: Ariel.

Baños, R. (2014) Orality markers in Spanish native and dubbed sitcoms: Pretended spontaneity and prefabricated orality. *Meta* 59(2): 406-435.

Brazil, D. (1975) *Discourse Intonation: Discourse Analysis Monographs no. 1.* Birmingham: English Language Research, University of Birmingham.

Brown, G., Currie, K.L. & J. Kenworthy. (1980) *Questions of intonation.* Baltimore: University Park Press.

Cantero, F.J. (2002) *Teoría y análisis de la entonación*. Barcelona: Edicions de la Universitat de Barcelona.

Chaume, F. (2012) *Audiovisual Translation: Dubbing.* Manchester: St. Jerome Publishing.

Chun, D. (2002) *Discourse Intonation in L2: From Theory and Research to Practice*. Amsterdam/Philadelphia: John Benjamins.

Collins, B., and I.M. Mees. (2003) *Practical Phonetics and Phonology*. New York: Routledge.

Couper-Kuhlen, E. (1986) *An introduction to English prosody*. London: Edward Arnold.

Cruttenden, A. (1997) *Intonation* (2nd ed). Cambridge: Cambridge University Press.

––––. (2008) *Gimson's Pronunciation of English* (7th ed). London: Hodder Education.

Crystal, D. (1969) *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

–––. (1975) *The English Tone of Voice.* London: Edward Arnold.

–––. (1985). *A dictionary of linguistics and phonetics* (2nd ed). Oxford: Blackwell Publishing.

Cuevas Alonso, M. (2017). Multimodalidad, comunicación y doblaje. La entonación. In X. Montero Domínguez (ed) 2017 *El Doblaje. Nuevas vías de investigación*. Granada: Editorial Comares.

Estebas-Vilaplana, E. & P. Prieto. (2009) La notación prosódica en español. Una revisión del Sp_ToBI. *Estudios de Fonética Experimental* 18: 263-283.

García-Lecumberri, M.L. (1995) *Intonational Signalling of Information Structure in English and Spanish*. Bilbao: UPV/EHV.

–––. (2003) Análisis por configuraciones: La escuela británica. In P. Prieto (coord) *Teorías de la entonación*. Barcelona: Ariel.

Gottlieb, H. (2005) Multidimensional Translation: Semantics turned Semiotics. *Proceedings of the Marie Curie Euroconferences MuTra: Challenges of Multidimensional Translation.* Saarbrücken, Germany, 2005. http://www.euroconferences.info/proceedings/2005_Proceedings/2005_Gottlieb_Henrik.pdf. 7 March 2018.

Gutiérrez Díez, F. (1995) *La función demarcativa de la entonación en inglés, castellano y catalán.* Murcia: Servicio de Publicaciones Universidad de Murcia.

–––. (2005) Intonation focus in the interlanguage of a group of Spanish learners of English. *Revista Alicantina de Estudios Ingleses* 18: 129-147.

Halliday, M.A.K. (1967) *Intonation and grammar in British English.* The Hague: Mouton.

–––. (1970) *A Course in Spoken English: Intonation*. Oxford: Oxford University Press.

Hidalgo Navarro, A. (2011) Humor, prosodia e intensificación pragmática en la conversación coloquial española. *Verba* 38: 271-292.

Hirst, D. & A. Di Cristo. (1998) A survey of intonation systems. In D. Hirst and A. Di Cristo (eds) *Intonation systems: A survey of twenty languages.* Cambridge: Cambridge University Press.

Mateo, M. (2014) Exploring pragmatics and phonetics for successful translation. *VIAL* 11: 111-136.

Mompeán González, J.A. & R. Monroy Casas. (2010) Nuclear tones in English: The case of household consumer products TV commercials. *ReVEL* 8(15): 229-241. http://www.revel.inf.br/files/artigos/revel_15_nuclear_tones_in_english.pdf. 22 March 2018.

Monroy Casas, R. (2002) El sistema entonativo del murciano coloquial. Aspectos comunicativos y actitudinales. *Estudios Filológicos* 37: 77-101.

–––. (2005) Spanish and English intonation patterns. A perceptual approach to attitudinal meaning. *Pragmatics & Beyond New Series* 140: 307-324.

–––. (2012) *La pronunciación simplificada del inglés británico.* Murcia: Servicio de publicaciones de la Universidad de Murcia.

Navarro Tomás, T. (1944) *Manual de entonación española*. New York: Hispanic Institute in the United States.

Nilsenová, M. & M. Swerts. (2012) Prosodic adaptation in language learning. In J. Romero-Trillo (ed) *Pragmatics and prosody in English language teaching.* Dordrecht/ Heidelberg/New York/London: Springer.

O'Connor, J.D. & G.F. Arnold. (1973) *Intonation of Colloquial English* (2nd ed). London: Longman.

Pike, K.L. (1945) *The intonation of American English*. Ann Arbor: University of Michigan Press.

Quilis, A. (1993) *Tratado de fonología y fonética españolas.* Madrid: Gredos.

Romero-Fresco, P. (2009) Naturalness in the Spanish dubbing language: A case of not-so-close *Friends*. *Meta* 54(1): 49-72.

Sánchez-Mompeán, S. (2016) 'It's not what they said; it's how they said it': A corpus-based study on the translation of intonation for dubbing. In A.M. Rojo López and N. Campos Plaza (eds) *Interdisciplinarity in Translation Studies. Theoretical Models, Creative Approaches and Applied Methods*. Bern: Peter Lang.

Savorelli, A. (2010) *Beyond Sitcom. New Directions in American Television Comedy*. North Carolina: McFarland & Company Inc.

Tench, P. (2011) *Transcribing the sounds of English: A phonetics workbook for words and discourse.* Cambridge: Cambridge University Press.

Wells, J. (2006) *English intonation: An introduction.* Cambridge: Cambridge University Press.

Whitman-Linsen, C. (1992) *Through the dubbing glass. The synchronization of American Motion Pictures into German, French and Spanish.* Frankfurt am Main: Peter Lang.

Zabalbeascoa, P. (2001) La traducción del humor en textos audiovisuales. In M. Duro (coord) *La traducción para el doblaje y la subtitulación*. Madrid: Ediciones Cátedra.