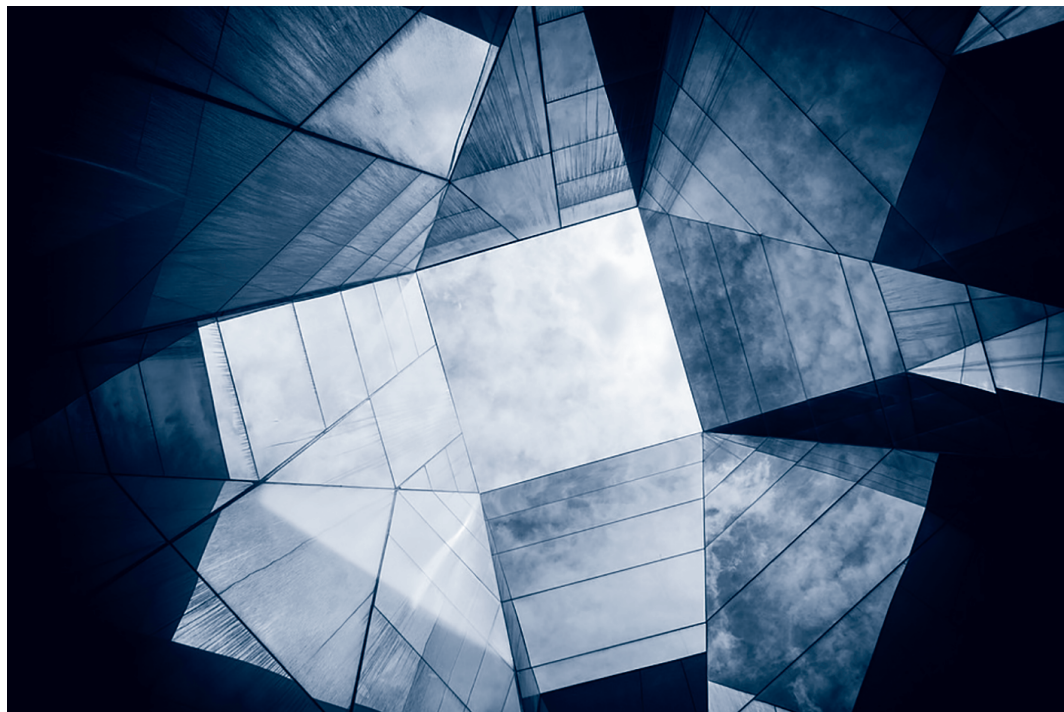


Multiperspectives in analysis and corpus design



Editores:

Miguel Fuster-Márquez

Carmen Gregori-Signes

José Santaemilia Ruiz

EDITORIAL COMARES



Interlingua

Miguel Fuster-Márquez
Carmen Gregori-Signes
José Santaemilia Ruiz
(eds.)

Multiperspectives in analysis and corpus design

Granada, 2020

Colección indexada en la MLA International Bibliography desde 2005

EDITORIAL COMARES

INTERLINGUA

249

Colección fundada por:
EMILIO ORTEGA ARJONILLA
PEDRO SAN GINÉS AGUILAR

Comité Científico (Asesor):

ESPERANZA ALARCÓN NAVÍO Universidad de Granada	MARIA JOÃO MARÇALO Universidade de Évora
JESÚS BAIGORRI JALÓN Universidad de Salamanca	HUGO MARQUANT Institut Libre Marie Haps, Bruxelles
CHRISTIAN BALLIU ISTI, Bruxelles	FRANCISCO MATTE BON UNINT, Roma
LORENZO BLINI UNINT, Roma	JOSÉ MANUEL MUÑOZ MUÑOZ Universidad de Córdoba
ANABEL BORJA ALBÍ Universitat Jaume I de Castellón	FERNANDO NAVARRO DOMÍNGUEZ Universidad de Alicante
NICOLÁS A. CAMPOS PLAZA Universidad de Murcia	NOBEL A. PERDU HONEYMAN Universidad de Almería
MIGUEL A. CANDEL MORA Universidad Politécnica de Valencia	MOISÉS PONCE DE LEÓN IGLESIAS Université de Rennes 2 – Haute Bretagne
ÁNGELA COLLADOS AÍS Universidad de Granada	BERNARD THIRY Institut Libre Marie Haps, Bruxelles
ELENA ECHEVERRÍA PEREDA Universidad de Málaga	FERNANDO TODA IGLESIA Universidad de Salamanca
PILAR ELENA GARCÍA Universidad de Salamanca	ARLETTE VÉGLIA Universidad Autónoma de Madrid
FRANCISCO J. GARCÍA MARCOS Universidad de Almería	CHELO VARGAS SIERRA Universidad de Alicante
CATALINA JIMÉNEZ HURTADO Universidad de Granada	MERCEDES VELLA RAMÍREZ Universidad de Córdoba
ÓSCAR JIMÉNEZ SERRANO Universidad de Granada	ÁFRICA VIDAL CLARAMONTE Universidad de Salamanca
HELENA LOZANO Università di Trieste	GERD WOTJAK Universidad de Leipzig
JUAN DE DIOS LUQUE DURÁN Universidad de Granada	

ENVÍO DE PROPUESTAS DE PUBLICACIÓN:

Las propuestas de publicación han de ser remitidas (en archivo adjunto, con formato PDF) a alguna de las siguientes direcciones electrónicas: anablen.martinez@uco.es, psgines@ugr.es

Antes de aceptar una obra para su publicación en la colección INTERLINGUA, esta habrá de ser sometida a una revisión anónima por pares. Para llevarla a cabo se contará, inicialmente, con los miembros del comité científico asesor. En casos justificados, se acudirá a otros especialistas de reconocido prestigio en la materia objeto de consideración.

Los autores conocerán el resultado de la evaluación previa en un plazo no superior a 60 días. Una vez aceptada la obra para su publicación en INTERLINGUA (o integradas las modificaciones que se hiciesen constar en el resultado de la evaluación), habrán de dirigirse a la Editorial Comares para iniciar el proceso de edición.

© Los autores

© Editorial Comares, S.L.
Polígono Juncaril
C/ Baza, parcela 208
18220 Albolote (Granada)
Tf.: 958 465 382

<https://www.comares.com> • E-mail: libreriacomares@comares.com
facebook.com/Comares • twitter.com/comareseditor • instagram.com/editorialcomares

ISBN: 978-84-1369-009-4 • Depósito legal: Gr. 737/2020

Fotocomposición, impresión y encuadernación: COMARES

Table of Contents

LIST OF CONTRIBUTORS	vii
INTRODUCTION	xi
MIGUEL FUSTER-MÁRQUEZ	
CARMEN GREGORI-SIGNES	
JOSÉ SANTAEMILIA RUIZ	
ADVANCE-FEE SCAMS: A CORPUS AND GENRE ANALYSIS	1
ISMAEL ARINAS PELLÓN	
PATRIZIA ANESA	
EL SABOR DE LAS MANZANAS: ANÁLISIS CONTRASTIVO (ESPAÑOL-INGLÉS) DE LA TERMINOLOGÍA OBJETIVA REFERIDA A LA EXPERIENCIA SENSORIAL DEL GUSTO.	15
LEONOR PÉREZ RUIZ	
MARÍA-TERESA ORTEGO ANTÓN	
TWO EXAMPLE MARKERS IN AND BEYOND EXEMPLIFICATION: DIALECTAL, REGISTER AND PRAGMATIC CONSIDERATIONS IN THE 21ST CENTURY	33
PAULA RODRÍGUEZ-ABRUÑEIRAS	
PROBABILISTIC GRAMMARS ACROSS REGISTERS: PRONOMINAL SUBJECT EXPRES- SION IN SOME VARIETIES OF ENGLISH	47
IVÁN TAMAREDO	
SEMANTIC FRAMES IN SCIE-LEX	61
ISABEL VERDAGUER	
EMILIA CASTAÑO	
NATALIA JUDITH LASO	

ACCESIBILIDAD, TRADUCCIÓN AUDIOVISUAL Y NORMAS EN LA SUBTITULACIÓN ONLINE: EMPAC (EUROPARTV MULTIMEDIA PARALLEL CORPUS)	73
IRIS SERRAT ROOZEN	
FINT-ESP: A CORPUS OF FINANCIAL REPORTS IN SPANISH.	89
ANTONIO MORENO-SANDOVAL	
ANA GISBERT	
HELENA MONTORO	
SPELLING NORMALISATION AND POS-TAGGING OF HISTORICAL CORPORA: THE CASE OF GUL, MS HUNTER 135 (FF. 34R–121V)	103
JESÚS ROMERO-BARRANCO	
ANNOTATING FACTUALITY IN THE TAGFACT CORPUS	115
GLÒRIA VÁZQUEZ GARCÍA	
ANA FERNÁNDEZ-MONTRAVETA	

list of contributors

PATRIZIA ANESA is a Senior Researcher in English Language and Linguistics at the University of Bergamo and she holds a PhD in English Studies from the University of Verona. Her research interests lie mostly in the area of specialised discourse, with particular reference to the investigation of knowledge asymmetries in professional communications. She has been a member of several national and international projects in the field of ESP. Email: patrizia.anesa@unibg.it

ISMAEL ARINAS PELLÓN is a lecturer at the Universidad Politécnica de Madrid (UPM). He teaches Communication for Specific Purposes to industrial engineering undergraduates. He is currently adapting materials for these courses. As a result of this syllabus adaptation, he is researching how technology is described in different contexts (patents, text-books, journal articles, marketing materials). Additionally, he is analysing persuasion strategies in different genres combining genre analysis, Maton's Legitimation Code Theory, social psychology, and phenomenology with corpus linguistics. He is a member of an Innovation on Education group known as Comm&Learn as well as of the Global Engineers Language Skills (GELS) international network. Email: ismael.arinas@upm.es

EMILIA CASTAÑO is an adjunct lecturer in the department of Modern Languages and Literatures and English Studies at the University of Barcelona. She holds a PhD in English Philology from the University of Barcelona and her research interests include corpus linguistics and the application of cognitive semantics, frame semantics, image schemas and conceptual metaphor theories, to discourse analysis. Email: e.castano@ub.edu

ANA FERNÁNDEZ-MONTRAVETA holds a PhD in Linguistics and Communication and an MA in Computational Linguistics. She is currently an associate professor in the English and German Department at the Universitat Autònoma de Barcelona and leader of the Inter-University Research Group on Language Applications (GRIAL). Her research focuses on English-Spanish contrastive analysis, more specifically on the field of constructions. In the last few years, she has carried out research on aspectuality, modality and factuality, always at sentence level. Especially relevant is her participation in the creation of various resources, among which there is the Spanish WordNet 3.0, highly useful for translation between Spanish and English, and several corpora annotated with linguistic information. She co-authored the books *Clasi-*

ficación verbal. Alternancias de diátesis and Las construcciones con 'se' en español, and has more than fifty publications. She has led various projects related to corpus compilation and annotation, such as SenSem. She currently participates in the TAGFACT project. Email: Ana.Fernandez@uab.cat

ANA GISBERT is Associate Professor in the Department of Accounting in the Faculty of Economics at Madrid Autonomous University. She was a Predoctoral Fellow at Lancaster University within the context of the HARMONIA European project on Accounting Harmonisation and Standardisation in Europe. She has published papers in the areas of international accounting, corporate governance, audit oversight and earnings management. Her current research interests are focused on the analysis of financial reporting narratives. Email: ana.gisbert@uam.es

NATALIA J. LASO is a Serra Hunter fellow in English Linguistics at the University of Barcelona. She holds a PhD in English Philology from the University of Barcelona and is also a member of the GRELIC-Lexicology and Corpus Linguistics Research Group. Her research is focused on two main areas: a) science writing and the main challenges that NNES writers face when writing their research in English; and b) the use of corpora in the linguistics classroom. She has co-edited the volume *Biomedical English: a corpus-based approach* (by Isabel Verdaguer, Natalia J. Laso & Danica Salazar. Eds.), published by John Benjamins Publishing. Email: njlaso@ub.edu

HELENA MONTORO ZAMORANO is a computational linguist, graduate in Translation and Interpreting (2017) and Master's degree in Spanish Language Research (2019) from the Universidad Autónoma de Madrid. She currently works at the Instituto de Ingeniería del Conocimiento, in the Social Business Analytics department. She participates in projects and research on Natural Language Processing, carrying out tasks -such as creation of computational grammars, elaboration of taxonomies and dictionaries, compilation and annotation of corpus or design of chatbots- related to the development of linguistic resources for the detection of sentiment in social media and the training of machine learning models. Email: helena.montoro@iic.uam.es

ANTONIO MORENO-SANDOVAL (BA 1986, MA 1988, PhD 1991, Universidad Autónoma de Madrid, UAM) is Associate Professor of Linguistics and Director of the Computational Linguistics Lab at UAM, and Director of the UAM-IIC Chair in Computational Linguistics. He is a former Fulbright postdoc scholar at the Computer Science Dept., New York University (1991-1992) and a former DAAD scholar at Augsburg Universität (1998). His training in Computational Linguistics began as a research assistant in the Eurotra Machine Translation Project (EU FP-2) and then at IBM Scientific Centre in Madrid (1989-1990). He was the principal researcher of the Spanish team in the C-ORAL-ROM Project (EU FP-5). Since 2010 he is Senior Researcher at the Instituto de Ingeniería del Conocimiento (IIC-UAM) in the Social Business Analytics group. Moreno-Sandoval has supervised 12 theses to completion. He is author or co-author of 5 books and over 80 scientific papers. Email: antonio.msandoval@uam.es

MARÍA-TERESA ORTEGO ANTÓN es Profesora Ayudante Doctor en el Departamento de Lengua Española de la Universidad de Valladolid y miembro del CITTAC de la mencionada universidad, así como del grupo interuniversitario ACTRES de la Universidad de León. Colabora con el grupo LEXYTRAD de la Universidad de Málaga y con el OLST de la Universidad de Montreal. Imparte docencia en la Facultad de Traducción e Interpretación de la Universidad

de Valladolid, en asignaturas relacionadas con la interpretación, la traducción especializada y las tecnologías de la traducción en la Facultad de Traducción e Interpretación. Licenciada en Traducción e Interpretación en 2007, obtuvo el grado de Doctor en Traducción y Comunicación Intercultural en noviembre de 2012 gracias a la concesión de una Ayuda FPI. Entre sus publicaciones destacan artículos en revista, capítulos de libro y libros sobre lexicografía bilingüe, terminología, tecnologías de traducción y lingüística de corpus aplicada a los Estudios de Traducción. Ha presentado ponencias en varios congresos internacionales. Además, es miembro de varios proyectos de investigación nacionales e internacionales sobre traducción especializada e interpretación. Ha trabajado como traductora e intérprete profesional durante más de diez años. Email: mariateresa.ortego@uva.es

LEONOR PÉREZ RUIZ es Profesora Titular de Universidad en el Departamento de Filología Inglesa de la Universidad de Valladolid y miembro del CITTAC (Centro de Investigación en Terminología bilingüe, Traducción especializada y Análisis Contrastivo). Su destino docente está en la Facultad de Filosofía y Letras donde imparte docencia en el Grado de Estudios Ingleses. Sus áreas de especialización son las lenguas con fines específicos, lenguas con fines académicos, la lingüística del corpus y la traducción. Ha sido investigadora principal en varios proyectos y también ha participado en proyectos europeos. Ha publicado numerosos artículos en revistas especializadas, capítulos de libros y monografías. También ha presentado trabajos en Congresos Nacionales e Internacionales. Email: lperezru@fyl.uva.es

PAULA RODRÍGUEZ-ABRUÑEIRAS holds a European Doctorate in Historical Linguistics from the University of Santiago de Compostela (January 2015). From August 2013 to May 2015, she was an assistant of Spanish at the Department of Spanish and Portuguese of the University of Wisconsin-Milwaukee (USA), where she also completed a two-year M.A. Degree in Spanish. Between September 2015 and January 2016, Paula held a visiting lecturer position at the Department of English, French and German of the University of Vigo, and in February 2016 she joined the Department of English and German at the Universitat de València, on a position as Lecturer in English. She has also carried out research at both national (University of Vigo, University of Santiago) and international (University of Helsinki, University of Manchester, University of Freiburg, University of Wisconsin-Milwaukee) universities, and at the British Library too. She is currently a member of GENTEXT and IULMA, and an affiliated member of LVTC (<http://view0.webs.uvigo.es/team/paula-rodriguez-abruñeiras>). Her research interests are historical linguistics, linguistic variation, new Englishes, corpus linguistics, critical discourse analysis and gender and sexual (in)equality issues. Email: paula.rodriguez@uv.es

JESÚS ROMERO-BARRANCO studied English Language and Literature at the University of Málaga, where he earned two MA Degrees, one in advanced English studies (2013) and the other in teaching English as a foreign language (2014). In 2017, he received his PhD with a thesis entitled “Early Modern English Scientific Text Types: Edition and Assessment of Linguistic Complexity of the Texts in MS Hunter 135 (ff. 34r–121v)”. He is a member of the Department of English and German (Universidad de Granada) and has been a visiting researcher at the Department of English Language (University of Glasgow) and the Department of the History of English (Adam Mickiewicz University, Poznań). Among his research interests, Dr. Romero is not only interested in Historical Linguistics, Palaeography and Codicology, but also in morphosyntactic variation in Present-day English and its varieties around the world. Dr Romero-Barranco has

published in journals such as *Atlantis*, *Studia Neophilologica* or *English World-wide*, among others. Email: jesusromero@ugr.es

IRIS SERRAT ROOZEN holds a PhD in Applied Languages, Literature and Translation by the Universitat Jaume I. She is currently a professor at the Catholic University of Valencia San Vicente Mártir. Her main lines of research include audiovisual translation, accessibility, and second language learning and teaching. Email: iris.serrat@uv.es

IVÁN TAMAREDO holds a BA in English Language and Literature (June 2012), and an MA (June 2013) and a PhD (November 2018) in English Studies. Formerly a pre- and postdoctoral researcher at the University of Santiago de Compostela under funding from the Spanish Ministry of Economy and Competitiveness (grant no. BES-2015-071233), Iván is currently a Substitute Lecturer in English at the University of Vigo. His research interests include varieties of English, probabilistic variation, language processing, and linguistic complexity. He has presented papers at several major international conferences (*ChangE* 2015 at Helsinki, and *ICAME* 36, 37, and 39 at Trier, Hong Kong, and Tampere, respectively) and published articles in peer-reviewed journals such as *English World-Wide*, *English Language and Linguistics*, *Atlantis*, and *ICAME Journal*. Email: ivan.tamaredo.meira@uvigo.es

GLORIA VÁZQUEZ is a senior lecturer in the area of General Linguistics at the University of Lleida and a member of the Interuniversity Research Group on Language Applications (GRIAL). At the beginning of her career, she was interested in the study of verbal semantic classes and their degree of internal homogeneity in relation to syntactic behavior, which is why she studied verbal syntactic patterns together with their constructional meaning. Her contributions to the description of Spanish verbal periphrasis stand out, and in recent years, she has started a new line of research related to the degree of factuality expressed in sentences. She has more than fifty publications, among which 2 books, *Verbal classification. Diathesis alternations* and *Constructions with “se” in Spanish*, stand out. Other important publications are two chapters in two manuals: one on corpus linguistics and another on Spanish lexicography. In these areas, she has also participated in the creation of some large-scale resources and collaborated in several funded research projects. She is currently the main investigator of the TAGFACT project. Email: gvazquez@dal.udl.cat

ISABEL VERDAGUER CLAVERA is Professor of English at the University of Barcelona. Her research interests and publications include the history of translation, contrastive linguistics, corpus linguistics, cognitive linguistics and lexicology. She coordinates the GRELIC-Lexicology and Corpus Linguistics Research Group and has coedited *Biomedical English: a corpus-based approach* with Laso and Salazar (John Benjamins, 2013). Her recent work includes “Identifying verb collocational patterns in a specialized medical journal corpus: a pedagogical approach to phraseology” (with Noguchi) 2018, in *RESLA*, 31, 2; and “Semantic frames and semantic networks in the Health Science Corpus” in *ELiEs* (forthcoming). Email: i.verdaguer@ub.edu

Introduction

In this volume, the readers of *Multiperspectives in Analysis and corpus design* will find nine selected peer reviewed and original contributions which deal with key aspects in recent trends in corpus linguistics, such as the developments in corpus design, compilation procedures and annotation, and the different analytical perspectives in which corpus techniques have become a core empirical methodology, either in isolation, or combined with other approaches that help reinforce arguments. It will be found that, in most of the articles, the authors themselves have compiled their own study corpus. Consequently, as it is customary in Corpus Linguistics research, a justification of the compilation procedure (e.g. sampling parameters or representativeness) is part and parcel of the discussion. The research areas to which corpus linguistics has been successfully applied in this volume include historical linguistics, linguistic variation, discourse analysis, computational linguistics and translation.

The first paper by Arinas and Anesa, "Advance-Fee Scams: A corpus and genre analysis", provides a detailed analysis of online scams. The authors state that online scams have "become a global criminal phenomenon which, worryingly, is on the increase". Their corpus contains over 500 fraudulent email texts. Their study aims to identify the persuasive strategies used by scammers to manipulate their victims by generating their confidence to deceive them. A keyword analysis is applied as a crucial analytical technique followed by a qualitative analysis. The analytical frameworks applied include Bhatia's notions of promotional letter move structure, Gao & Gao's representation of knowledge involved in fraud detection and prevention, Kahnenab's model of judgment and choice according to preferences and attitudes, and Fischer, Lea and Evans' persuasion strategies. An important conclusion is that scammers "use a strong narrative within

¹ All references to authors or titles mentioned in this introduction will be located in the bibliographical references provided by each contributor to this volume.

a recognizable genre”, creating the “illusion of intimacy, sincerity, and urgency”, to persuade and tease their victims.

Pérez and Ortego carry out a contrastive terminological analysis around food entities, in particular, apples: “El sabor de las manzanas: Análisis contrastivo (español-inglés) de la terminología objetiva referida a la experiencia sensorial del gusto”. In their paper, the authors of this second article perform a contrastive Spanish-English discourse study of terms used to name the gustatory perception. For that purpose, they make use of a comparable bilingual study corpus based on fact sheets on apples gathered from websites of Spanish and British food companies. They look into meaning choices through frequent collocates, and co-textual intensifying and mitigation strategies. They find that in their description of apple taste, these fact sheets typically appeal to sensory qualities such as sweetness or acidity, the evocation of different foods and beverages, the aroma and the touch. In their conclusion, they highlight that both languages use a wide variety of terminology to describe, with great accuracy, the different taste sensations that the consumption of an apple causes on the palate. They hope that this study can be useful for translators, who would need to use the appropriate terminology and discourse features when promoting online products.

The third article, by Rodríguez-Abruñeiras, examines the use of *for example* and *for instance* in Present-Day English in two corpora, the *British English 2006* and *American English 2006*, each of these containing one million words. The author follows Eggs and McElholm’s typology of exemplification, selection and argumentation, undertaking a corpus-based study of two mainstream varieties of English, namely British and American English, in the early twenty-first century. Among other aspects, her analysis reveals that both markers show quite similar overall distributions in the two general corpora, and that the argumentative function prevails over others in both mainstream varieties of English. Regarding text types, the results indicate that fiction has very few examples of these markers, when compared with informative prose, which is due, according to Rodríguez-Abruñeiras, to their differing rhetoric and communicative purposes. Thus, she argues that “fiction is creative and is not expected to be accurate except as a reflection of human experiences, whereas text types, such as science or law (which are argumentative by nature), are characterised “by the use of more accurate and concise kind of language”. Rodríguez-Abruñeiras claims that both markers “may also bring about various pragmatic nuances, such as focus or mitigation”, most particularly in scientific prose. In her conclusion, she highlights that there are no significant differences between British English and American English. In both varieties there is “clear preference for *for example* over *for instance*” and a similar use of the three functions she examined in this paper.

Tamaredo’s paper “Probabilistic grammars across registers: Pronominal subject expression in some varieties of English”, fourth in this monograph, takes a dialectal and variationist/sociolinguistic perspective. Tamaredo focuses on the internal and external constraints which determine the choice between overt and pronominal subjects among speakers of three varieties of World Englishes with different regional and cultural back-

grounds, namely British English, Indian English, and Singapore English. This research was carried out by exploring the corresponding three national corpora, containing one million words each, within the *International Corpus of English*. The author adopts the double approach of probabilistic grammar framework, which assumes that grammatical knowledge is at least partially experience- and a usage-based approach. The author takes into account various internal language-constraints as predictors of subject pronoun omission, and language-external factors (variety, mode of production, and level of formality). He makes use of VADIS variationist modelling, in order to examine probabilistic differences. Tamaredo's findings seem to demonstrate that, among the external factors, the mode of production and the level of formality outweigh variety to account for variation between omitted and overt subject pronouns in the three varieties. As for the internal factors, the author concludes that coordination and clause position are the most crucial constraints. As in many cases of morphosyntactic variation in PDE, Tamaredo finds that register-related factors appear "to override variety-specific patterns".

"Semantic frames in *SciE-Lex*" is the title of the fifth paper, by Verdaguer, Castaño and Laso. The paper focuses on the analysis of two English verbs, *to block* and *to inhibit*, very frequently used in biomedical discourse. The article discusses a new development of *SciE-Lex*, a lexical database of biomedical English, a development from their *Health Science Corpus*, a four-million-word corpus of biomedical English. In this new development, the authors applied Fillmore's frame semantics' approach and used the well-known database *Framenet*, which classifies individual words on the basis of their semantic frames. The aim of the *SciE-Lex* project is to help the Spanish biomedical community to publish papers in English which conform "to the conventions of scientific discourse". At present, *SciE-Lex* contains phonological, morphological, syntactic, semantic, collocational and phraseological information on highly frequent biomedical English words. The authors analyse the frame-based information for both verbs, which enables them to observe the syntactic and semantic patterns that both of them share. This approach makes it possible to integrate lexical meanings into a higher level of organization, as well as compare and contrast their preferred meanings, collocational and syntactic patterns against those in general discourse. Since the meanings of words in domain-specific texts tend to be more specific than in general use, the authors claim that their idiosyncrasies should be considered. The authors conclude that "the application of Frame Semantics to the semantic and syntactic description of the verbs *to block* and *to inhibit* in their specialised corpus bears witness to the existence of "remarkable differences between general and biomedical English".

Subtitling norms and accessibility in audiovisual translation is the focus of the sixth contribution in this monograph. Serrat's paper, "Accesibilidad, traducción audiovisual y normas en la subtitulación *online*: EMPAC (EuroparlTV) Multimedia Parallel Corpus", discusses issues of accessibility on the audiovisual content in subtitling on the online television channel EuroparlTV. Her analysis is based on EMPAC, a Spanish and English parallel (English/Spanish) corpus of 5 million words she has compiled for the years 2009-2017. In this contribution, Serrat discusses the norms which underlie the online

subtitles generated by translation professionals to determine whether they are aligned with the commitment to equal access to information for all European citizens. She homes in on the following relevant features: reading speed, pauses between subtitles, and characters per line, as well as (in)adequate segmentation of syntactic patterns. Her conclusions indicate that the lack of commonly accepted standards for online subtitling leads to malpractice and dubious quality and illegible subtitles. In her view, these are barriers for effective communication in this institutional EU context.

Moreno Sandoval, Gisbert and Montoro's "A corpus of financial reports in Spanish" provides an overview of the different steps taken to retrieve texts of annual reports in a corpus of contemporary Spanish financial narratives (FinT-esp) a corpus compiled by their research team. This is a contribution to the development of existing methodologies in computational linguistics and, more specifically, aims to help finance and accounting fields in the processing, classification and analysis of large amounts of financial narratives provided by PDF files. A great deal of the article is dedicated to the description of the contents of FinT-esp and the methodology followed for the compilation of the corpus. The authors use an adaptation and modification of the CFIE-FRSE tool developed by Mahmoud El Haj (Lancaster University), an application that detects the structure of annual reports and allows them to extract their contents at section level. Once this was done, the texts were processed with computational tools such as POS taggers or parsers. For example, they highlight that to help prospective users in their searches "[t]he texts are indexed with Elastic Search, which favours "a fast and easy search as if we were using a Google-like tool". To illustrate the usefulness of this specialised corpus, Moreno-Sandoval, Gisbert and Montoro discuss three lexical features: modality (to detect possible linguistic biases), keyword analysis (to identify domain specific vocabulary) and polarity (by applying a computational tool of sentiment analysis). Regarding possible applications, the authors state that it is already being used for terminology extraction and, claim that FinT-esp could become "a valuable source of data to facilitate both the development of new techniques in computational linguistics and the promotion of interdisciplinary studies between accounting and linguistic academics".

Romero-Barranco shows his interest in the compilation and annotation of historical corpora in his paper "Spelling normalisation and POS-tagging of historical corpora: The case of GUL, MS Hunter 135 (ff. 34r-121)". In this eighth contribution, Romero-Barranco brings in decisions or suggestions involving the process of compiling and annotating an early Modern English linguistic corpus. As a source of evidence, the author examines the criteria used to compile and annotate the Glasgow, University Library, MS Hunter 135 (ff. 34r-121v), a medical volume written in the first half of the sixteenth century. Given that early Modern English features spelling variation, it is his view that diachronicians would need to normalise the orthography if they wish to enhance the performance of an automatic POS-tagger as would be done with a contemporary text. The spelling normalisation is carried out with *WARD* (Rayson, Archer and Smith 2005). One of its most crucial features is that *WARD*, the variant detector tool, not only normalises spelling variants, but also adds a tag showing where this original

spelling is kept. Romero-Barranco also gives an overview of progress in automatic part-of-speech tagging. Most particularly he looks at what has been accomplished by developments of the CLAWS (Constituent Likelihood Automatic Word-tagging System) POS (Part of Speech) tagger. Both VARD and CLAWS taggers have been applied to this early Modern English document. The study shows that thanks to spelling normalisation, part-of-speech annotation with CLAWS increases its accuracy approximately by 15%. Particularly, CLAWS' accuracy is enhanced from 82.9% and 83.9% to 96.8% and 97.2% in the surgical text and the medical recipes. Nevertheless, Romero-Barranco points out that historical linguists who work with earlier texts need to make decisions about problematic issues such as how to deal with archaic words, genitives or compounds, among others. Additionally, he claims that automatic normalisations must be checked and manual amendments would be required.

In the last contribution to this volume, "Annotating factuality in the TAGFACT", Vázquez and Fernández-Montraveta focus on the complexity of applying automatic corpus annotation to do research on factuality and evaluation. The authors describe and give a rationale for the labels used for the annotation scheme of the corpus TAGFACT. The chapter takes into account, discusses and illustrates with numerous examples four aspects: eventual types (dynamic and non-dynamic situations) and the writer's commitment to the certainty of the assertion, polarity and time. Their analysis aims to capture "the degree of commitment of an author regarding the certainty of the facts he or she is narrating". The authors highlight three innovations which result from their approach. First, it helps distinguish dynamic from non-dynamic situations. Second, it allows the identification and annotation of absolute truths. Thirdly, it helps distinguish prototypical from eventual properties. Vázquez and Fernández-Montraveta claim that their system is the only one at present which is based exclusively on linguistic knowledge, for Spanish. Even though this model has been created for the annotation of a Spanish corpus, they claim that it may be applicable to other languages.

We must thank the Valencian Government (Generalitat Valenciana) for kindly giving financial support to this publication (project code 31/1769321).

MIGUEL FUSTER-MÁRQUEZ
CARMEN GREGORI-SIGNES
JOSÉ SANTAEMILIA RUIZ

El sabor de las manzanas: Análisis contrastivo (español-inglés) de la terminología objetiva referida a la experiencia sensorial del gusto

LEONOR PÉREZ RUIZ
MARÍA-TERESA ORTEGO ANTÓN
CITTAC, Universidad de Valladolid

Resumen: Las fichas descriptivas de las manzanas en las páginas web de empresas de fruta fresca incluyen información detallada respecto a sus características sensoriales. En este trabajo, a partir de un análisis contrastivo español-inglés de las referencias utilizadas para denominar los aspectos gustativos, así como sus colocaciones más frecuentes, basándonos en los datos de un corpus *ad hoc* comparable bilingüe, hemos detectado que la descripción del gusto hace referencia a cuatro aspectos fundamentales: al nivel de dulzor o acidez, a la evocación a distintos alimentos y bebidas, al perfume y al tacto. Además, dicha descripción va frecuentemente acompañada de amplificadores o mitigadores léxicos para expresar con precisión la mezcla de matices.

Palabras clave: manzana, características sensoriales, gusto, español, inglés, corpus.

1. INTRODUCCIÓN

En la actualidad, debido al cambio del paradigma de negocio tradicional, el uso generalizado de Internet por parte de las empresas para la venta de sus productos es una práctica habitual en todo el mundo. Las empresas de fruta fresca no son ajenas a esta realidad y la mayoría de ellas disponen de un portal Web, a través del cual dan a conocer las variedades que producen y venden no solo en España, sino también en el exterior. De hecho, en 2018 las frutas constituyeron el grupo agroalimentario con mayor saldo positivo en las exportaciones al exterior y los cítricos y las hortalizas están entre los productos más exportados en nuestro país en 2019 (MAPA, 2019: 9).

Para poder juzgar, con el fin de describir e intentar vender, si una pieza de fruta es de calidad, es necesario considerar las sensaciones que provoca a través de los sentidos. Por lo general, al valorar una manzana se comienza apreciando la sensación que produce al comerla, su nota hedónica. A continuación, se estima su cualidad, si tiene una textura, aroma o sabor determinados y, en tercer lugar, se considera la intensidad del producto (Alonso Gaité, 2011: 39). Estos aspectos, junto con otros también importantes para definir la calidad sensorial de una manzana (tamaño, color, brillo u olor), contribuyen a describir estas características visuales, táctiles, olfativas y gustativas.

En este sentido, los textos con información relativa a los alimentos alternan su enfoque entre la descripción objetiva de los mismos y la narración de las experiencias individuales de los consumidores a través de sus sentidos al ingerirlos. Estas descripciones, por tanto, pueden englobar no solo terminología objetiva, sino también evaluativa (Dubois, 2017: 10).

Es amplia la bibliografía existente dedicada al estudio y análisis de estas experiencias sensoriales polifacéticas desde distintas perspectivas. El tema se ha abordado desde campos tan variados como la horticultura (Daillant-Spinnler *et al.*, 1996; Stone & Sidel, 2004), el marketing (Krishna, 2012), o incluso en áreas del ámbito de las humanidades como la psicología (Abdi, 2002), la filosofía (Clark, 1993) o la antropología (Classen, 1997; Howes & Pink, 2010). También se ha estudiado el lenguaje sensorial con una perspectiva antropológica y cultural (Riley & Paugh, 2018) o desde la semántica (Diederich, 2015). Sin embargo, se le ha prestado menos atención desde la lingüística y, en concreto, desde el análisis del discurso contrastivo.

Así pues, en este trabajo pretendemos identificar y analizar la terminología utilizada para hacer referencia a la experiencia sensorial del gusto a la hora de describir las propiedades objetivas de las manzanas en las fichas descriptivas de dicho producto. Para extraer y analizar contrastivamente dicha terminología en las lenguas española e inglesa utilizaremos una metodología basada en la compilación y explotación de un corpus electrónico *ad hoc* comparable compuesto por fichas descriptivas de manzanas publicadas en las páginas web de industrias frutícolas españolas, por ejemplo, Villaveza, Frutas Erruz S.L., Finca Señorío de Rioja, La Rinconada, Fruit Nature, Cofrubi, Finca el Castro, Fruijet S.L., Santama, Viveros Requinoa, o Expofrut, así como de empresas frutícolas anglosajonas, tales como Growers Direct, USApple, Showaltres Orchard, Fishkill Farms, Wilkens Fruit & Fir Farm, Indian Ladder Farms o Fly Creek Cider Mill & Orchard¹, entre otras. En concreto, ante la variedad de terminología y fraseología utilizada, limitaremos nuestro estudio a la terminología de carácter neutro e imparcial utilizada para describir las sensaciones experimentadas a través del gusto, descartando aquellas que se refieren a los atributos sensoriales percibidos, sujetos al juicio individual del consumidor, que serán objeto de otro estudio posterior. Para este fin, realizaremos un análisis comparativo español-inglés del léxico que se emplea para hacer referencia a los aspectos gustativos y de sus colocaciones más frecuentes. En primer lugar, abordaremos las sensaciones gustativas y su lenguaje. Posteriormente, describiremos la metodología empleada para analizar si la descripción del gusto, que hace referencia a cuatro aspectos fundamentales: al nivel de dulzor o acidez, a la evocación a distintos alimentos y bebidas, al perfume y al tacto, se produce de forma semejante en las lenguas española e inglesa o, por el contrario, se utilizan parámetros diferentes en español y en inglés. A continuación, presentaremos los resultados clasificándolos según los cuatro aspectos mencionados. Dichos resultados nos permitirán vislumbrar

¹ El listado de los sitios web consultados para este trabajo aparece al final del mismo.

posibles patrones de comportamiento diferenciados entre lenguas a la hora de expresar con precisión la mezcla de matices gustativos. Por último, con los resultados obtenidos extraeremos las conclusiones.

2. LAS SENSACIONES GUSTATIVAS Y SU LENGUAJE

Como señalan Swahn *et al.* (2010: 592): “as consumers become more demanding about the quality of the apples they buy, it becomes more important to understand the sensory perception of apples in order to improve the marketing process for grocery retail stores”. Las fichas descriptivas de las manzanas en los portales web de las empresas hortofrutícolas incluyen información detallada del producto y de las percepciones sensoriales que este provoca. Para tener éxito con estas plataformas de venta *online*, y considerando las especiales características de un negocio en el que los clientes no solo están interesados en el precio sino también en la calidad del producto, es fundamental el dominio de estrategias lingüísticas y retóricas eficaces para la comercialización del producto.

Para describir estas propiedades y percepciones características, es necesario considerar los órganos sensoriales responsables de las mismas. El campo de las sensaciones gustativas resulta bastante complejo pues el sabor deriva de una variada combinación de percepciones sensoriales (Diederich, 2015: 36). Tampoco existe un acuerdo claro sobre qué órganos influyen en la percepción del sabor. Los expertos suelen coincidir en que son la nariz y la boca los responsables de producir esas impresiones (Bayarri & Costell, 2010: 46), aunque también se señala que el tacto contribuye de alguna manera a las mismas (Costa *et al.*, 2011). Así, al definir el sabor, se habla más propiamente de percepción bucal, englobando el gusto y las impresiones olfativas y táctiles, unidas al gusto, que se mezclan con frecuencia por la vía retronasal.

Las descripciones de las características gustativas de las manzanas con fines comerciales incluyen una terminología claramente expositiva, esto es, precisa y objetiva. Además, también consta de elementos discursivos con una marcada función persuasiva, semánticamente subjetivos, especialmente característicos de los textos con fines comerciales (Stehlík, 2009). Asimismo, esta comunicación se realiza entre experto y consumidor lego y, como en muchos lenguajes de especialidad, “existe un subcódigo, parcialmente coincidente con el subcódigo de la lengua común, caracterizado en virtud de unas peculiaridades ‘especiales’” (Cabré, 1993: 123). Por otro lado, como señala Sinclair (1991: 170), las palabras tienden a aparecer en colocaciones léxicas determinadas y, de esta forma, adquirir su significado a partir de ese contexto creado.

En este sentido, podemos constatar que las palabras no aparecen en el texto de manera aleatoria, sino que se agrupan en torno a elecciones lingüísticas determinadas (Sinclair, 1991: 10). Así, por ejemplo, el sustantivo inglés “*flesh*”, en el lenguaje sensorial de las frutas, suele con frecuencia ir unido a elecciones lingüísticas como *crisp*, *juicy*, *rich*, *firm*, *crunchy* o *tender* (Seppä *et al.*, 2012: 48; Corollaro *et al.*, 2013:114). El sustantivo español “carne” se asocia habitualmente con consistente, firme, densa, jugosa, crujiente o tierna (Alonso Gaité, 2011). Por otro lado, el lenguaje utilizado para describir el gusto varía dependiendo de la función que se persigue:

Tasters' language is made up of precise terms for concrete sensations: sweetness, acidity, bitterness, the smell of ethyl acetate, for example; and for more subtle sensations it consists of imprecise, but conventional terms which attempt to define a balance of flavours. In the first case, the word fits the perception well and is intelligible. In the second case, by trying to clarify the blurred image of their sensations, tasters are led to juggle with words (Peynaud, 1987: 180).

Por tanto, una vez definidas las principales características del lenguaje que se utiliza para hacer referencia a las características gustativas de las manzanas, procedemos a describir la metodología de análisis para identificar y analizar la terminología objetiva utilizada en las fichas descriptivas de las manzanas.

3. METODOLOGÍA

En este trabajo vamos a emplear una metodología eminentemente descriptiva que combina datos cualitativos y cuantitativos que se desprenden de la explotación de un corpus *ad hoc* comparable bilingüe. Esto nos servirá para obtener una primera aproximación a los términos y los patrones léxicos (Sinclair, 1991, 1996; Tognini Bonelli, 2001). En una segunda fase será necesario el contraste de los resultados con un mayor volumen de datos ampliando el corpus a otros productos frutícolas y el análisis a otras características sensoriales.

En primer lugar, hemos compilado un corpus electrónico *ad hoc* siguiendo la metodología descrita en Seghiri (2017) y Ortego Antón (2019: 77-90). En su diseño, la selección de textos para cada uno de los subcorpus se efectuó teniendo en cuenta los siguientes criterios: disponibilidad, tamaño, representatividad, equilibrio y perfil de usuario. Así pues, dicho corpus se caracteriza por ser comparable, en inglés y español, compuesto de fichas descriptivas de manzanas, de aproximadamente 60.000 palabras, 30.000 en cada lengua. En cuanto a la disponibilidad, los textos proceden de Internet, principalmente de explotaciones y empresas dedicadas a la producción y comercialización de este producto. Nuestro corpus está compuesto por textos de empresas hortofrutícolas que van desde pymes hasta grandes empresas y, por tanto, que se dedican al comercio mayorista y minorista. Todos los textos corresponden a una comunicación entre experto y no experto (consumidor).

En cuanto al tamaño, cada subcorpus incluye aproximadamente 200 textos. La procedencia de los textos redactados originalmente en inglés es principalmente de EE.UU. y Gran Bretaña, mientras que los textos en español son en su mayoría de España, aunque también hay textos de México, Argentina, Chile, Uruguay, Costa Rica y Colombia. El equilibrio se consigue combinando dos factores, a saber, la fecha, dado que los textos proceden de 2017 y 2019, y la procedencia, se han incluido más de 46 empresas.

En relación con la representatividad, que puede ser de dos tipos: cualitativa y cuantitativa, nuestro corpus cumple con ambas. Así, se ha llevado a cabo un riguroso proceso de compilación, atendiendo a los postulados de Seghiri (2017), puesto que se ha compilado seleccionando textos de un mismo género textual en dos lenguas: inglés

y español. Para comprobar si el corpus es representativo cualitativamente utilizamos el programa ReCor, diseñado por Seghiri (2006, 2015) y Corpas Pastor y Seghiri (2009, 2010). La representatividad cuantitativa se define como “la terminología básica empleada en este género” (Seghiri, 2017: 50). ReCor analiza cada subcorpus y genera dos gráficos de representatividad. Los datos que ofrece ReCor muestran que el corpus es representativo a partir de 200 textos y 30000 palabras o *tokens*.

Una vez compilado y descrito el corpus, para analizarlo se ha recurrido a la herramienta de software AntConc 3.5.7 (Anthony, 2018) que nos ha ayudado a extraer e identificar la terminología, así como a detectar los patrones de uso más frecuentes en este género textual, en español y en inglés.

Para describir el gusto utilizaremos cuatro categorías o marcos cognitivos (Salazar & Orozco, 2011: 265): *taste*, *sabor*, *flavor* (también se ha considerado la grafía del inglés británico *flavour*) y *aroma*, y sus colocaciones, junto con aquellas unidades léxicas que ayudan a matizar la descripción del gusto y que aparecen asociadas a estas denominaciones (Renouf & Sinclair, 1991: 128-143).

4. ANÁLISIS Y RESULTADOS

Tras identificar mediante el programa AntConc 3.5.7 (Anthony, 2018) las ocurrencias de los cuatro marcos cognitivos seleccionados (*taste*, *flavor/flavour*, *aroma* y *sabor*), y las colocaciones con otras unidades léxicas, siguiendo la metodología anteriormente indicada, hemos analizado aquellos términos que Peynaud (1987: 180) identifica como “precise terms for concrete sensations”.

Detectamos que la descripción del gusto hace referencia a cuatro aspectos fundamentales: en un primer grupo identificamos una amplia referencia a los sabores básicos percibidos por el sentido del gusto; un segundo grupo con gran presencia en el corpus evoca a varios tipos de alimentos y bebidas que la ingesta de estas manzanas produce; el tercer grupo hace alusión a la reminiscencia o recuerdo de un perfume o aroma característicos; y un último grupo menciona sensaciones asociadas con el tacto, con la textura de carne y piel, o a la impresión que produce el morder una manzana. Procedemos a analizar en detalle estos resultados.

4.1. Sabores básicos

Se incluyen habitualmente dentro de este grupo la dulzura, la amargura, la acidez y la salinidad. En el caso de las descripciones de las manzanas, se hace referencia principalmente al dulce y ácido. Comenzamos describiendo el atributo dulce, cuyas unidades léxicas asociadas tienen el mayor número de ocurrencias en el corpus.

4.1.1. Dulce

El atributo dulzor es uno de los sabores básicos de las manzanas (Swahn *et al.*, 2010: 596) por lo que el término dulce, sus derivados y otras unidades léxicas que también describen este atributo, aparecen con frecuencia en ambos corpus analizados.

En la Tabla 1 aparecen las unidades léxicas relativas a esta sensación de dulzor, asociadas a cada uno de los marcos analizados, ordenadas por el número de ocurrencias, a excepción de *aroma*, que no coloca con ninguna unidad léxica que haga referencia al atributo de dulce.

Marco	Colocación	Frecuencia absoluta
Sabor	dulce	54
	azucarado	2
	dulzón	1
Taste	<i>sweet</i>	32
	<i>honey</i>	5
	<i>sugary</i>	1
	<i>sweetness</i>	1
Flavor	<i>sweet</i>	81
	<i>honey</i>	10
	<i>honeyed</i>	4
	<i>sweetness</i>	3
	<i>sugars</i>	2
	<i>bubble gum</i>	1
	<i>candied</i>	1
	<i>cherry candy</i>	1
	<i>pear-drop</i>	1
	<i>sugary</i>	1
	<i>sweeter</i>	1
Flavour	<i>sweet</i>	8
	<i>saccharine</i>	1
	<i>sweetness</i>	1

Tabla 1. Unidades léxicas utilizadas para hacer referencia al sabor dulce

En estas descripciones se utilizan unidades léxicas específicas que hacen referencia al dulzor (*sweet*, *sweetness*, dulce, dulzura, dulzor), además de otras voces que, igualmente, evocan esta sensación como *sugary*, *sugars*, *saccharine* o azucarado.

En ocasiones se alude a ciertas golosinas, utilizando este símil para describir de forma precisa y original este tipo de sabor. Estos ejemplos solo aparecen en el subcorpus en inglés, resultando de gran creatividad el recurrir al chicle, a las gominolas o a los caramelos para matizar estas evocaciones, como se puede apreciar en los ejemplos extraídos del corpus que se presentan a continuación:

- 1) *A mild, sweet and aromatic **candied** flavor, sometimes reminiscent of bubble gum.*
- 2) *Anise, almond, **bubble gum**, spice, or **cherry candy**.*
- 3) *A drab appearance which belies a unique **peardrop** favor.*
- 4) *Some tasters detect a "**pear-drop**" flavor.*

También la miel es otra mención típica en la descripción del dulzor. En este caso sí se incluyen estas referencias en ambos corpus como vía de identificar este sabor:

- 5) *Crisp apple with a sweet, rich **honey** flavor.*
- 6) *A floral aroma, and slight **honey** flavor.*
- 7) *Crisp and bursting with a juicy, **honeeyed** flavor.*
- 8) *Sweet-sharp **honeeyed** flavor.*
- 9) Dulce como la **miel**.

Es común en ambos subcorpus, en la descripción de sabores básicos, el hacer uso de intensificadores para expresar con precisión la mezcla de matices. Se busca reflejar de la forma más rigurosa posible el sabor descrito. En la Tabla 2 incluimos una clasificación de estas unidades léxicas con las ocurrencias presentes en el corpus entre paréntesis, asociadas a los marcos analizados, y en una gradación de dulce intenso hasta no dulce. A simple vista se observa un abanico mucho más amplio, rico y descriptivo de apariciones en el subcorpus en inglés que en el subcorpus en español.

Intensidad	Español	Inglés
dulce intenso	muy dulce (14) / intensamente dulce (1)	<i>the sweetest (3)/ super sweet (4) / intense sweet flavor (2) / very sweet (12)/ is almost entirely sugary (1) / pronounced sweet (1)</i>
dulce	azucarado (2)	<i>sugary (8) / syrup(2) / saccharine (1) / saccharine juice (1)</i>
azucarado	dulzón (1)	<i>wonderful sweet (1)</i>
dulce miel		<i>sweet honey(1) / honey sweet (4) / honey-like (1) / honey (10) / honeeyed (4) / distinct honeeyed (1) / with some hints of honey (1) /slightly honey(1)</i>
golosinas		<i>candied (1) / pear-drop (1)/ cherry candy (1)/ bubble gum (2)</i>
otros sabores		<i>flavour sweet (1) / spicy sweet (1)</i>
bastante dulce		<i>sweet to very sweet (1) / predominantly sweet (1)</i>
dulce	dulce (157)	<i>sweet (300) / sweetness (17)</i>
dulce equilibrado		<i>balanced sweet (5) / pleasant sweet (2)</i>
ligeramente dulce	dulce pero no demasiado remarcado (1)/ un poco dulce (1)/ nada ácido (1)	<i>mildly sweet (7) / slightly sweet (3)/ nearly sweet (1) / light sweetness (1)/ not too sweet (1)</i>
ligeros toques ácidos	dulce con toques ácidos (1)	<i>sweet almost tart (1)</i>
dulce-ácido equilibrado		<i>balanced tart and sweet (1)/ balanced sweet-tart (3)/ balanced sweet and tart (2)/ sweet-tart balance (2) / a balance of acid and sweetness (1) / a balance of sugars and acids (1)</i>

dulce-ácido	dulce-ácido (1) / dulce/agrio (4) / entre muy dulce y ácido (1) / dulce y ácido (3)	<i>tart to sweet (1) / tangy sweet (3) / sweet sharp(1) / honey-tart (1) / sweet to tart (1) / sweet-tart (2) / tart balanced with sweet (1)</i>
no dulce		<i>almost savory (1)</i>

Tabla 2. Colocaciones de los términos relacionados con el dulzor

4.1.2. Ácidos

Junto con el léxico que hace referencia al sabor dulce, se han identificado unidades léxicas relativas a la descripción de la percepción de acidez (Tabla 3). No hemos detectado unidades léxicas asociadas con el término *aroma*.

Marco	Colocación	Frecuencia absoluta	
Sabor	ácido	32	
	acidez	4	
	agrio	4	
	acidulado	3	
	amargo	2	
Taste	<i>tart</i>	11	
	<i>tartness</i>	2	
	<i>acidity</i>	2	
Flavor	<i>sub acid</i>	40	
	<i>tart</i>	39	
	<i>sour</i>	11	
	<i>acid</i>	6	
	<i>acidity</i>	4	
	<i>sharp</i>	5	
	<i>brisk</i>	3	
	<i>crisp</i>	2	
	<i>tangy</i>	2	
	<i>acids</i>	1	
	<i>acidulous</i>	1	
	<i>savory</i>	1	
	Flavour	<i>tangy</i>	2
		<i>acidity</i>	1
<i>sub acid</i>		1	
<i>tart</i>		1	
<i>tartness</i>		1	

Tabla 3. Unidades léxicas utilizadas para hacer referencia al sabor ácido

Aunque la terminología utilizada en español para describir la percepción de acidez es variada (ácido, acidez, agrio, amargo, acidulado), la lengua inglesa, a la vista de los datos obtenidos del análisis del subcorpus correspondiente, contiene un vocabulario mucho más amplio (*tart, tartness, acid, acidity, tangy, sour, sharp, crisp, brisk*).

Como ocurría con el atributo dulzor, en las fichas descriptivas de las manzanas se hace hincapié en el grado de acidez que caracteriza a cada variedad. Así, estas categorizaciones aluden a un sabor poco o nada ácido (*intense sweet flavour with no tartness or acidity*; dulce poco ácida; sabor dulce con un leve toque ácido; dulce nada ácido; sabor dulce y baja acidez), bastante equilibrado (*taste is a balanced tart-and-sweet; a balanced sweet-tart taste; not too sweet, not too tart, just right*; equilibrio entre acidez y dulzor), o más intenso (*balanced sweet and tart enlivened by acidity*; sabor dulce con toques ácidos; sabor azucarado con un agradable punto ácido).

Por lo que respecta a los amplificadores o mitigadores léxicos que se utilizan para matizar de manera precisa la intensidad de la percepción, en la Tabla 4 se presentan ordenados por intensidad.

Intensidad	Español	Inglés
no ácido	nada ácido (1)	<i>with no acidity (1)</i>
baja acidez	dulcemente ácido (1) una chispa de acidez (1) leve toque ácido (1) con un toque ácido (1) punto ácido ligeramente ácido con un ligero toque ácido (1) suave toque ácido (1) poco ácido (3) de baja acidez (3)	<i>slightly sub acid (1)</i> <i>mildly sub acid (2)</i> <i>slightly acidulous (1)</i> <i>mild acidity (1)</i> <i>somewhat tart (3)</i> <i>with very little trace of acidity (1)</i> <i>not too tart (1)</i> <i>a mild tart (2)</i> <i>mild-tart (1)</i> <i>mildly-tart (3)</i> <i>almost tart (1)</i> <i>slightly tart (1)</i>
dulce-ácido	entre dulce y un poco ácido (1) dulce/agrio (4) dulce-ácido (1) entre muy dulce y ácido (1)	<i>sub acid to sweet (3)</i> <i>honey-tart (1)</i> <i>tangy-sweet (2)</i> <i>sweet-tart (11)</i> <i>tart and sweet (2)</i> <i>tart-sweet balance of flavor (1)</i>
algo ácido		<i>tart to sub acid (1)</i> <i>enlivened by acidity (2)</i> <i>tangy-tart (2)</i>
muy ácido	particularmente ácido (1) muy ácido (1) intensamente ácido (1)	<i>very tart (2)</i>

Tabla 4. Colocaciones de los términos relacionados con la acidez

A la vista de los datos expuestos en la Tabla 4, podemos afirmar que la creatividad expresiva del atributo acidez es más equilibrada en ambos idiomas que en el caso del dulzor.

En relación con otros sabores básicos, hemos incluido también aquellas descripciones de sabores que resaltan el carácter refrescante de la manzana (Tabla 5). En el corpus solo aparecen ejemplos de esta categoría en inglés y asociados a los marcos *taste* y *flavor*. El hecho de que este tipo de terminología sea característica únicamente en el subcorpus en inglés nos lleva a destacar, de nuevo, la existencia de una mayor precisión descriptiva en lengua inglesa.

Marco	Colocación	Frecuencia absoluta
Taste	<i>juicy</i>	4
	<i>dry</i>	1
	<i>effervescent</i>	1
	<i>Juice</i>	1
	<i>refreshing</i>	1
	<i>sprightly</i>	1
Flavor	<i>juicy</i>	5
	<i>sprightly</i>	5
	<i>astringent</i>	1
	<i>effervescent</i>	1
	<i>refreshing</i>	1
	<i>snappy</i>	1

Tabla 5. Unidades léxicas utilizadas para hacer referencia al sabor refrescante

4.2. Alimentos y bebidas

Las fichas descriptivas de las manzanas recurren a la mención a distintos alimentos o bebidas para describir las propiedades gustativas de esta fruta. Un capítulo importante tiene que ver con la referencia a otras frutas frescas (Tabla 6). En el subcorpus en inglés observamos abundantes ejemplos que identifican el sabor de la manzana con el de otros tipos de fruta, como la piña, el plátano, el plátano verde, la pera, etc. Sin embargo, en el subcorpus en español no hay referencias a frutas concretas, a excepción del caso, un tanto sorprendente, de la *berry*, en vez de utilizar el término español baya. En ambas lenguas, además, hay alusiones a la fruta en general, con términos como *fruity*, o a un determinado tipo, fruta tropical. También las manzanas despiertan recuerdos a frutos secos. En inglés se habla de la almendra y la nuez (*almond*, *almond-like*, *nutty*) y en español se nombra la nuez (anuezado, nuez).

Por otro lado, en menor medida, se describe cierto sabor que evoca al de determinadas plantas aromáticas o especies, ya sea mediante una referencia general (*spicy*, *spice*, especiado, especias) o mencionando variedades concretas (*fennel*, *anise*).

Además, en ambos corpus se alude a cómo el sabor de ciertas variedades de manzana recuerda al de distintas bebidas. En inglés se recurre al vino (*vinous*, *winey*,

wine-like) y la sidra (*cider*), mientras que en español se hace referencia al moscatel, un vino muy aromático.

Marco	Colocación	Frecuencia absoluta
Sabor	anuezado	2
	especias	2
	<i>berry</i>	1
	cítrico	1
	especiado	1
	moscatel	1
	tropical	1
Aroma	anuezado	2
	especiado	1
	nuez	1
Flavor	<i>spicy</i>	17
	<i>spice</i>	6
	<i>anise</i>	3
	<i>nutty</i>	3
	<i>cidery</i>	2
	<i>pineapple-like</i>	2
	<i>quince</i>	2
	<i>vinous</i>	2
	<i>almond</i>	1
	<i>almond-like</i>	1
	<i>banana</i>	1
	<i>fennel</i>	1
	<i>fruity</i>	1
	<i>lemony</i>	1
	<i>pear-like</i>	1
<i>wine-like</i>	1	
Flavour	<i>melon</i>	1
	<i>orange</i>	1
	<i>pear</i>	1
	<i>spicy</i>	1

Tabla 6. Unidades léxicas utilizadas para hacer referencia a alimentos y bebidas

Al describir las sensaciones y asociaciones a distintos alimentos y bebidas, las fichas descriptivas no solo hacen referencia a estos, sino que también se matizan utilizando diversos procedimientos de intensificación, cuantificadores o modificadores, como se muestra en la Tabla 7.

Intensidad	Español	Inglés
menos intenso	ligeramente tropical (1) algo anuezado (2) con notas de avellana (1) con notas de especias dulces (1) con un toque de especias (1) con un toque cítrico (1) recuerda a un toque de moscatel (1) un toque de canela (1) un toque mentolado (1)	<i>melon undertones (2)</i> <i>with some hints of green banana (1)</i> <i>slight quince (2)</i> <i>has hints of pear, orange, melon (1)</i> <i>slight anise (5)</i> <i>slight crabapple flavor (2)</i> <i>enjoy pear and melon tones when eaten fresh (1)</i> <i>a mild pear-like flavor (2)</i> <i>a hint of pear (1)</i> <i>hints of pear and strawberry (1)</i> <i>trace of pear (1)</i> <i>hint of spice, pear, and cherry (1)</i> <i>hints of pear, orange, melon (1)</i> <i>with spice notes (1)</i>
similar	semejante al de la nuez (1) como el de una Berry (2)	<i>pineapple taste (2)</i> <i>pineapply taste (1)</i> <i>a flavor comparison to fennel or anise (1)</i> <i>spicy-sweet (1)</i> <i>noted for its spicy flavor (1)</i> <i>similar to winesap flavor and texture (1)</i>
más intenso		<i>high quince (1)</i> <i>more lemony than (1)</i>

Tabla 7. Colocaciones de los términos relacionados con alimentos y bebidas

Una vez más observamos, como en los casos del dulzor y de la acidez, una mayor variedad fraseológica en el subcorpus en inglés que en el subcorpus en español.

4.3. Perfume

El tercer punto de nuestra clasificación se refiere a la categoría del perfume. Las características gustativas de algunas variedades de manzanas incluyen reminiscencias a ciertos olores. Indiscutiblemente el olor y el sabor trabajan de forma conjunta, ya que el sabor de la comida se detecta principalmente en la nariz, y no en la lengua. En el corpus analizado hemos identificado algunos términos que aluden a esta perspectiva aromática, asociados a los términos *sabor*, *taste* y *flavor*, que se presentan en la Tabla 8.

Marco	Colocación	Frecuencia absoluta
Sabor	perfume	5
	aromático	3
	perfumado	1
Taste	<i>aromatic</i>	2
	<i>floral</i>	1
	<i>perfumed</i>	1

Flavor	<i>aromatic</i>	22
	<i>balsamic</i>	1
	<i>perfumed</i>	1

Tabla 8. Unidades léxicas utilizadas para hacer referencia al perfume

Las unidades léxicas registradas en la Tabla 8 mencionan de manera global el perfume como característico de esa variedad de manzana, sin especificar un tipo concreto. Solo hemos identificado un caso en el que se señala que el perfume es floral, término también bastante genérico.

Estas unidades léxicas frecuentemente van acompañadas, como en los casos anteriores, de algún elemento enfatizador que señala la mayor o menor intensidad del perfume descrito, como se observa en la Tabla 9.

Intensidad	Español	Inglés
menos intenso	perfumado suave (1) poco perfumada (1) ligeramente perfumada (1) algo perfumada (1) su perfume ligero (1)	<i>almost perfumed taste</i> (1) <i>mildly aromatic</i> (1) <i>slightly perfumed</i> (1) <i>floral notes</i> (1)
neutro	aromático (2) perfumado (1) perfumada (8) un perfume muy característico (1)	<i>with a distinctive balsamic flavor</i> (1)
más intenso	muy aromático (1) muy aromática (14) muy perfumada (1)	<i>strongly aromatic</i> (1) <i>highly aromatic</i> (8) <i>very aromatic</i> (6) <i>with a rich aromatic sub acid flavor</i> (1) <i>rich aromatics</i> (1)

Tabla 9. Colocaciones de los términos relacionados con el perfume

4.4. Tacto

El último aspecto relativo a la descripción de las sensaciones concretas y objetivas que una persona puede experimentar al comer una manzana tiene que ver con el tacto. A excepción de *aroma*, en el corpus aparecen unidades léxicas asociadas con todos los marcos analizados, como se indica en la Tabla 10.

Marco	Colocación	Frecuencia absoluta
Sabor	<i>crujiente</i>	3
	<i>crudo</i>	1
	<i>cremoso</i>	1
Taste	<i>buttery</i>	3
	<i>crisp</i>	2
	<i>tender</i>	1

Flavor	<i>dense</i>	3
	<i>crunchy</i>	1
Flavour	<i>crisp</i>	1

Tabla 10. Unidades léxicas utilizadas para hacer referencia al tacto

Habitualmente las manzanas son valoradas por los consumidores en cuanto a si son crujientes, crocantes, densas, firmes, consistentes, finas, etc. Sin embargo, en nuestro corpus, naturalmente, los términos referidos a la textura asociados al sabor son mucho más limitados. Aun así, es interesante ver que aparecen determinados ejemplos asociados al sentido del gusto.

5. CONCLUSIONES

El fin de este trabajo ha sido la identificación y análisis de la terminología objetiva más común utilizada en el ámbito de la descripción con fines comerciales de las manzanas por parte de empresas mayoristas y minoristas dedicadas a su producción y venta. Nuestro estudio se ha centrado en las referencias a la experiencia sensorial del gusto, aspecto que no ha sido abordado con anterioridad en el campo de la Lingüística de Corpus.

Tras el análisis del corpus de fichas descriptivas de manzanas, en inglés y español, en general, ambos idiomas se caracterizan por contar con una amplia variedad terminológica para describir con gran precisión las distintas sensaciones gustativas que la ingesta de una manzana provoca en el paladar. Además, este discurso utiliza una terminología profusa para reflejar sensaciones concretas y objetivas.

Puesto que en las percepciones gustativas se mezclan diversas experiencias sensoriales, su descripción es rica en unidades léxicas específicas y descriptivas referidas no solo a determinados sabores, sino también a la evocación de otros alimentos, bebidas, perfumes y sensaciones táctiles. Todas estas descripciones con frecuencia se presentan acompañadas de amplificadores o mitigadores léxicos que contribuyen a cuantificar su intensidad.

En cuanto a las limitaciones de nuestro estudio, esta investigación, por razones de espacio se ha limitado al estudio y análisis de las unidades léxicas que aparecen asociadas a los cuatro marcos descritos: sabor, *taste*, *flavor*, *flavour* y *aroma*. Por ello, en el futuro pretendemos ampliar el estudio a nuevas asociaciones y colocaciones pues, con toda probabilidad, se identificarían un número más amplio de unidades léxicas referidas a la descripción del gusto.

En consecuencia, este estudio es tan solo un primer paso para llevar a cabo un análisis exhaustivo del lenguaje sensorial de las manzanas, que ha de resultar muy útil para concienciar a las empresas respecto al deterioro de la imagen corporativa que puede implicar para estas si al traducir del español al inglés su página web no se utiliza la terminología adecuada. Así, coincidimos con Ortego-Antón y Fernández Nistal (2020: 46) en la necesidad de colaborar con las empresas del sector de la agroalimentación

para que doten a sus páginas web de visibilidad internacional en la web del conocimiento y las páginas traducidas se adecuen a las convenciones propias del discurso de comercialización y descripción de las frutas típicas de la lengua inglesa para lograr un lenguaje que resulte atractivo al lector y también sirva para calificar adecuadamente los productos analizados.

Por último, las conclusiones derivadas de este estudio y de esta línea de trabajo han de tener una gran aplicabilidad para el campo de la horticultura y, de forma más general, en la comercialización de la producción de la industria agroalimentaria. Un estudio de estas características resulta de gran ayuda en la producción de textos con fines comerciales pues, dada la complejidad la descripción del gusto, los términos utilizados para describir estas sensaciones deben ser seleccionados con gran precisión.

BIBLIOGRAFÍA

- ABDI, H. (2002). What can cognitive psychology and sensory evaluation learn from each other? *Food Quality and Preference*, 13(7-8), 445-451.
- ALONSO GAITE, María Asunción. (2011) *Caracterización sensorial y físico-química de manzanas Reineta y pera Conferencia, figuras de calidad en Castilla y León*. Tesis Doctoral. León: Universidad de León.
- ANTHONY, Laurence. (2018) *AntConc (Versión 3.5.7.) [Computer Software]*. Tokyo, Japan: Waseda University.
- BAYARRI, Sara & Elvira COSTELL. (2010) "Sensory evaluation of fruit and vegetable flavors". En Yiu H. HUY (ed) *Handbook of fruit and vegetable flavors*. New Jersey: John Wiley & Sons, 45-58.
- CABRÉ, María Teresa. (1993) *La terminología: Teoría, metodología y aplicaciones*. Barcelona: Antártida.
- CLARK, Austen. (1993) *Sensory qualities*. Oxford: Oxford University Press.
- CLASSEN, Constance. (1997) "Foundations for an Anthropology of the Senses". *International Social Science Journal*, 49(3): 401-12.
- COROLLARO, María Laura, Isabella ENDRIZZI, Anna BERTOLINI, Eugenio APREA, María Luisa DEMATTÈ, Fabrizio COSTA, Franco BIASIOLI & Flavia GASPERI. (2013) "Sensory profiling of apple: Methodological aspects, cultivar characterisation and postharvest changes". *Postharvest Biology and Technology*, 77: 111-120.
- CORPAS PASTOR, Gloria & Miriam SEGHIRI. (2009) "Virtual Corpora as Documentation Resources: Translating Travel Insurance Documents (English-Spanish)". En Allison BEEBY, Patricia RODRÍGUEZ INÉS & Pilar SÁNCHEZ-GIJÓN (eds) *Corpus Use and Translating*. Amsterdam / Philadelphia: John Benjamins, 75-107.
- (2010) "Size Matters: A Quantitative Approach to Corpus Representativeness". En Rosa RABADÁN *et al.* (eds) *Lengua, traducción, recepción: en honor de Julio César Santoyo*. León: Universidad de León, Área de Publicaciones, 111-145.
- COSTA, Fabrizio, Luca CAPPELLIN, Sara LONGHI, Walter GUERRA, Pierluigi MAGNAGO & DUILIO PORRO. (2011) "Assessment of apple fruit texture by a combined acoustic-mechanical profiling strategy". *Postharvest Biology and Technology*, 61(1): 21-28.
- DAILLANT-SPINLER, Béatrice, Halliday MACFIE, Pamela BEYTS & Duncan HEDDERLEY. (1996) "Relationships between perceived sensory properties and major preference directions of 12 varieties of apples from the southern hemisphere". *Food quality and preference*, 7(2): 113-126.

- DIEDERICH, Catherine. (2015) *Sensory Adjectives in the Discourse of Food: A frame-semantic approach to language and perception* (Vol. 16). Amsterdam/Philadelphia: John Benjamins.
- DUBOIS, Danielè. (2017). "How words for sensory experiences become terms". *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 23(1): 9-37.
- HOWES, David & Sarah PINK. (2010) "The Future of Sensory Anthropology / the Anthropology of the Senses". *Social Anthropology* 18(3): 331-40.
- KRISHNA, Aradhna. (2012) "An integrative review of sensory marketing: Engaging the senses to affect perception, judgment and behavior". *Journal of Consumer Psychology*, 22(3): 332-351.
- MAPA. *Informe anual de comercio exterior agroalimentario, pesquero y forestal*. Madrid: Ministerio de Agricultura, Pesca y Medio Ambiente.
- ORTEGO ANTÓN, María-Teresa. (2019) *La terminología del sector agroalimentario (español-inglés) en los estudios contrastivos y de traducción especializada basados en corpus: los embutidos*. Berlín: Peter Lang.
- ORTEGO ANTÓN, María-Teresa & Purificación FERNÁNDEZ NISTAL. (2020). "Estudio contrastivo de la terminología de embutidos en inglés y en español con ParaConc y tlCorpus a partir del corpus paralelo P-GEFEM y del comparable C-GEFEM". En Miriam SEGHIRI (ed) *La lingüística de corpus aplicada al desarrollo de la competencia tecnológica en los estudios de traducción e interpretación y la enseñanza de segundas lenguas*. Berna: Peter Lang, 23-48.
- PEYNAUD, Émile. (1987) *The Taste of Wine*. London: Macdonald Orbis.
- RENOUF, Antoinette & John SINCLAIR. (1991) "Collocational frameworks in English". En Karin AIJMER & Bengt ALTENBERG (eds) *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman, 128-143.
- RILEY, Kathleen & Amy PAUGH. (2018) *Food and Language: Discourses and Foodways Across Cultures*. London: Routledge.
- SALAZAR, Nora & Guadalupe OROZCO. (2011) "El aroma de la manzana". *Interciencia*, 36(4): 265-271.
- SEGHIRI, Miriam. (2006) *Compilación de un corpus trilingüe de seguros turísticos (español-inglés-italiano): aspectos de evaluación, catalogación, diseño y representatividad*. Tesis doctoral. Málaga: Universidad de Málaga. Disponible en <http://hdl.handle.net/10630/2715> (Fecha de consulta: 05/12/2019).
- (2015) "Determinación de la representatividad cuantitativa de un corpus *ad hoc* bilingüe (inglés-español) de manuales de instrucciones generales de lectores electrónicos. En María Teresa SÁNCHEZ NIETO (ed) *Corpus-based Translation and Interpreting Studies: From description to application / Estudios traductológicos basados en corpus: de la descripción a la aplicación*. Berlin: Frank und Timme, 125-146.
- (2017) "Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/ español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores". *Babel*, 63 (1): 43-64.
- SEPPÄ, Lalia, Jenna RAILIO, Riikka MONONEN, Risto TAHVONEN & Hely TUORILA. (2012) "From profiles to practice: Communicating the sensory characteristics of apples to the wider audience through simplified descriptive profiles". *LWT-Food Science and Technology*, 47(1): 46-55.
- SINCLAIR, John. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- (1996). "The Search for Units of Meaning", *Textus*, IX (1): 75-106.
- STEHLÍK, Petr. (2009) Algunas particularidades del lenguaje publicitario en los anuncios de vinos. En *Vino jako multikulturní fenomén*. Olomouc: Filozofická fakulta Univerzity Palackého v Olomouci, 222-227.

- STONE, Herbert & Joel SIDEL. (2004) *Sensory evaluation practices*. California, US: Elsevier.
- SWAHN, Johan, Asa ÖSTRÖM, Ulf LARSSON & Inga-Britt GUSTAFSSON. (2010) "Sensory and semantic language model for red apples", *Journal of sensory studies*, 25(4): 591-615.
- TOGNINI-BONELLI, Elena. (2001) "Corpus Linguistics at Work". *Computational Linguistics*, 28(4): 583-583.

SITIOS WEB

<http://www.frutadetoro.com/>, <http://frutaserruz.com/frutas/>, <http://fincasenoriode-rioja.com/>, <http://www.frutaslarinconada.com>, <http://www.fruitnaturesat.com>, <http://www.cofrubi.com>, <http://www.fincaelcastro.es>, <http://fruijet.com/>, <http://www.santama.com.co>, <http://viverosrequinoa.cl>, <http://expofrut.com.ar>, <http://www.fruit-importer.co.uk/our-fruit>, <http://usapple.org/all-about-apples/apple-varieties/>, <http://www.showaltersorchard.com/the-orchard/>, <http://www.fishkillfarms.com/learn/all-about-apples/#>, <http://wilkensfarm.com/apple-info/>, <http://www.indianladderfarms.com>, <https://www.flycreekcidermill.com> (Fecha de consulta: 30/11/2019).

colección:
INTERLINGUA

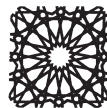
249

Dirigida por:

Ana Belén Martínez López y Pedro San Ginés Aguilar

The readers of *Multiperspectives in Analysis and Corpus Design* will find in this volume nine selected peer reviewed and original contributions which deal with key aspects in recent trends in corpus linguistics, such as the developments in corpus design, compilation procedures, and annotation. All the contributions in this book use corpus techniques as a core empirical methodology, either in isolation, or combined with other approaches. Furthermore, in most of the articles, the authors themselves have compiled their own study corpus. Consequently, as it is customary in Corpus Linguistics research, a justification of the compilation procedure (e.g. sampling parameters or representativeness) is part and parcel of the discussion. The research areas to which corpus linguistics has been successfully applied in this volume include historical linguistics, linguistic variation, discourse analysis, computational linguistics and translation.

We remain indebted to the Valencian Government (Generalitat Valenciana) for kindly giving financial support for this publication (project code 31/1769321).



COMARES
editorial

