



2008

Barcelona

**XIII**

Proceedings of the  
**euralex**  
International  
Congress

**Elisenda Bernal  
Janet DeCesaris (eds.)**



INSTITUT UNIVERSITARI  
DE LINGÜÍSTICA APLICADA  
UNIVERSITAT POMPEU FABRA

sèrie activitats 20

EURALEX congresses are true to the word "international" in their title: the contributors to these proceedings come from all around the world, and the number of languages discussed is high. These proceedings contain papers written in six languages: Catalan, English, French, German, Italian, and Spanish. This significant presence of many languages seems only fitting for a European association focused on the study and representation of words, and signals that interest in researching and improving dictionaries is widespread.



[www.documentauniversitaria.com](http://www.documentauniversitaria.com)



[www.iula.upf.edu](http://www.iula.upf.edu)

Proceedings of the XIII EURALEX International Congress  
(Barcelona, 15-19 July 2008)



**Proceedings of the XIII EURALEX International Congress  
(Barcelona, 15-19 July 2008)**

Elisenda Bernal  
Janet DeCesaris  
(eds.)

INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA  
UNIVERSITAT POMPEU FABRA  
Barcelona 2008

Reservats tots els drets. El contingut d'aquesta obra està protegit per la Llei, que estableix penes de presó i/o multes, a més de les corresponents indemnitzacions per danys i perjudicis per a aquells que reproduïssin, plagiessin, distribuïssin o comunicuessin públicament, en la seva totalitat o en part, una obra literària, artística o científica, o la seva transformació, interpretació o execució artística fixada en qualsevol classe de suport o comunicada a través de qualsevol mitjà, sense la preceptiva autorització.

Responsables de l'edició: Elisenda Bernal, Janet DeCesaris  
Maquetació: Elisenda Bernal, Jesús Carrasco, Alba Milà, Emma Vila

Direcció de les Publicacions de l'IULA: Mercè Lorente  
Coordinació tècnica de les Publicacions de l'IULA: Gemma Martínez

© DOCUMENTA UNIVERSITARIA ®  
[www.documentauniversitaria.com](http://www.documentauniversitaria.com)  
© Edicions a Petició, SL  
[www.edicionsapeticio.com](http://www.edicionsapeticio.com)

per a  
Institut Universitari de Lingüística Aplicada

Primera edició: juny de 2008  
© els autors  
© Institut Universitari de Lingüística Aplicada  
Pl. de la Mercè, 12  
08002 Barcelona

Disseny del logo i del CD-ROM: Jesús Carrasco  
Disseny de la coberta: Cass  
Adaptació de la coberta: Documenta Universitaria  
Impressió: Gràfiques Trema (coberta) i Divermat (interior)  
Dipòsit legal: B-33.195-2008  
ISBN 13: 978-84-96742-67-3

## Contents

Presentació .....	23
Foreward .....	25
<i>Janet DeCesaris</i>	

### PLENARY LECTURES

Tradition and Innovation in Catalan Lexicography .....	35
<i>Joaquim Rafel</i>	
FrameNet Meets Construction Grammar .....	49
<i>Charles J. Fillmore</i>	
Sobre la discontinuidad en un diccionario histórico .....	70
<i>José Antonio Pascual</i>	
Lexical Patterns: from Hornby to Huston and Beyond .....	89
<i>Patrick Hanks</i>	
Twenty-five years of Dictionary Research: Taking Stock of Conferences and other Lexicographic Events since LEXeter'83 .....	131
<i>Reinhard R. K. Hartmann</i>	

### PAPERS

Book of Abstracts .....	151
<i>Section 1. Computational Lexicography</i>	
Approaches to Computational Lexicography for German Varieties .....	251
<i>Andrea Abel, Stefanie Anstein</i>	
AnCora-Verb: Two Large-scale Verbal Lexicons for Catalan and Spanish .....	261
<i>Juan Aparicio, Mariona Taulé, M. Antònia Martí</i>	
Multi-Level Reference Hierarchies in a Dictionary of Swahili .....	269
<i>Piotr Bański, Beata Wójtowicz</i>	

Multidimensional Ontologies: Integration of Frame Semantics and Ontological Semantics .....	277
<i>Guntis Bārzdīņš, Normunds Grūzītis, Gunta Nešpore, Baiba Saulīte, Ilze Auziņa, Kristīne Levāne-Petrova</i>	
The Structure of the Lexicon in the Task of Automatic Lexical Acquisition .....	285
<i>Núria Bel, Sergio Espeja, Montserrat Marimon</i>	
SAOL-PLUS—A New Swedish Electronic Dictionary .....	291
<i>Sture Berg, Louise Holmer, Anki Hult</i>	
Matching Verbo-nominal Constructions in FrameNet with Lexical Functions in MTT .....	297
<i>Myriam Bouveret, Charles J. Fillmore</i>	
Syntactic Behaviour and Semantic Kinship of Selected Danish Verbs .....	309
<i>Anna Braasch</i>	
An Author's Dictionary: The Case of Karel Čapek .....	323
<i>František Čermák</i>	
Aide à la construction de lexiques morphosyntaxiques .....	331
<i>Claude de Loupy, Sandra Gonçalves</i>	
Bottom-up Editing and More: The E-forum of <i>The English-Chinese Dictionary</i> .....	339
<i>Jun Ding</i>	
An Electronic Lexicon for Turkish Idiomatic Compounds Headed by Verbs .....	345
<i>Elif Eyigoz</i>	
Mordebe Admin—A Lexical Management System .....	351
<i>José Pedro Ferreira, Sílvia Barbosa, Maarten Janssen</i>	
Lexicon Creator: A Tool for Building Lexicons for Proofing Tools and Search Technologies .....	359
<i>Thierry Fontenelle, Nick Cipollone, Mike Daniels, Ian Johnson</i>	
Generation of Word Profiles on the Basis of a Large and Balanced German Corpus .....	371
<i>Alexander Geyken, Jörg Didakowski, Alexander Siebert</i>	

El <i>Diccionario de diccionarios do galego medieval</i> .....	385
<i>Ermesto González Seoane, María Álvarez de la Granja, Ana Isabel Boullón Agrelo, María Rodríguez Suárez, Damián Suárez Vázquez</i>	
Shimmering Lexical Sets .....	391
<i>Patrick Hanks, Elisabetta Ježek</i>	
The Use of Context Vectors for Word Sense Disambiguation within the ELDIT Dictionary .....	403
<i>Kateryna Ignatova, Andrea Abel</i>	
Meaningless Dictionaries .....	409
<i>Maarten Janssen</i>	
Software Demonstration: The TshwaneLex Electronic Dictionary System ....	421
<i>David Joffe, Malcolm MacLeod, Gilles-Maurice de Schryver</i>	
GDEX: Automatically Finding Good Dictionary Examples in a Corpus .....	425
<i>Adam Kilgarriff, Miloš Husák, Katy McAdam, Michael Rundell, Pavel Rychlý</i>	
Finding the Words Which are Most X .....	433
<i>Adam Kilgarriff, Pavel Rychlý</i>	
Corpus as a Means for Study of Lexical Usage Changes .....	437
<i>Michal Křen, Jaroslava Hlaváčová</i>	
Non-heads of Compounds as Valency Bearers: Extraction from Corpora, Classification and Implication for Dictionaries .....	449
<i>Ekaterina Lapshinova-Koltunski</i>	
The Lexicographic Portal of the IDS: Connecting Heterogeneous Lexicographic Resources by a Consistent Concept of Data Modelling .....	457
<i>Carolin Müller-Spitzer</i>	
Multilingual Open Domain Key-word Extractor Proto-type .....	463
<i>Alessandro Panunzi, Marco Fabbri, Massimo Moneglia</i>	
Refining and Exploiting the Structural Markup of the eWDG .....	469
<i>Thomas Schmidt, Alexander Geyken, Angelika Storrer</i>	
<i>An Anglo-Saxon Dictionary</i> and a Morphological Analyzer of Old English ..	483
<i>Ondrej Tichy, Jan Čermák</i>	

*Section 2. The Dictionary-Making Process*

El programa de ejemplificación en los diccionarios didácticos .....	489
<i>María Bargalló Escrivá</i>	
Sobre las construcciones pronominales y su tratamiento en algunos diccionarios monolingües de cuatro lenguas románicas .....	495
<i>Paz Battaner, Irene Renau</i>	
La distribució de la informació contextual en els elements estructurals d'un article de diccionari: col·locacions, restriccions lèxiques i definició .....	505
<i>Judit Feliu, Joan Soler</i>	
The Greek High School Dictionary: Description and issues .....	515
<i>Maria Gavrilidou, Voula Giouli, Penny Labropoulou</i>	
Desafíos de la definición .....	525
<i>Juan Gutiérrez Cuadrado</i>	
The Funny Mirror of Language: The Process of Reversing the English-Slovenian Dictionary to Build the Framework for Compiling the New Slovenian-English Dictionary .....	535
<i>Simon Krek, Mojca Šorli, Polonca Kocjančič</i>	
Making a thesaurus for learners of English .....	543
<i>Diana Lea</i>	
Structure de la définition lexicographique dans un dictionnaire d'apprentissage explicatif et combinatoire .....	551
<i>Jasmina Milićević</i>	
Frames and Semagrams. Meaning Description in the General Dutch Dictionary .....	561
<i>Fons Moerdijk</i>	
A Systematic Approach to the Selection of Neologisms for Inclusion in a Large Monolingual Dictionary .....	571
<i>Ruth O'Donovan, Mary O'Neill</i>	
Lexicographic Treatment of Italian Phrasal Verbs: a Corpus-based Approach ..	581
<i>Cristina Onesti</i>	
El <i>Dizionario Italiano Garzanti</i> en el marco de la Lexicografía italiana contemporánea .....	587
<i>Giuseppe Patota</i>	

Lemmatisierungspraxis und -problematik im Autorenwörterbuch am Beispiel des <i>Goethe-Wörterbuchs</i> .....	599
<i>Thomas Schares, Christiane Schlaps</i>	
Von der Markierung zur Beschreibung: Besonderheiten des (Wort-) Gebrauchs in <i>lexiko</i> .....	607
<i>Ulrich Schnörch</i>	
Requirements for the Design of Electronic Dictionaries and a Proposal for their Formalisation .....	617
<i>Dennis Spohr</i>	
Alphabetic Proportions in Estonian Monolingual and Bilingual Dictionaries ..	631
<i>Enn Veldi</i>	
<i>Section 3. Reports on Lexicographical and Lexicological Projects</i>	
<i>Dictionnaire de Néologismes du Portugais Brésilien (décennie de 90):</i> conception et processus d'élaboration .....	637
<i>Ieda Maria Alves</i>	
A Digital Dictionary of Catalan Derivational Affixes .....	643
<i>Elisenda Bernal, Janet DeCesaris</i>	
Recopilación y estructuración del vocabulario de especialidad en el <i>Nuevo Diccionario Histórico del Español</i> (RAE) .....	649
<i>José Carriazo Ruiz, Marta Gómez Martínez</i>	
<i>Portal de léxico hispánico: una herramienta para el estudio del léxico</i> .....	655
<i>Gloria Clavería, Marta Prat, Joan Torruella, Cristina Buenafuentes, Margarita Freixas, Carolina Julià, Mar Massanell, Laura Muñoz, Sonia Varela</i>	
ISO-Standards for Lexicography and Dictionary Publishing .....	663
<i>Marie-Jeanne Derouin, André Le Meur</i>	
Construir un diccionario de derivación del español en el siglo XXI. La arquitectura de la información al servicio de la lexicografía .....	669
<i>María Teresa Díaz García, Inmaculada Mas Álvarez</i>	
<i>El Diccionari de l'Institut d'Estudis Catalans (2007):</i> el tractament de la pronominalitat verbal .....	679
<i>Imma Fradera, Olga Fullana, Pere Montalat, Carolina Santamaria</i>	

Diskurswörterbuch —Zur Konzeption eines neuen Wörterbuchtyps .....	689
<i>Heidrun Kämper</i>	
Turning Roget's <i>Thesaurus</i> into a Czech Thesaurus .....	697
<i>Aleš Klégr</i>	
MEDLEX+: An Integrated Corpus-Lexicon Medical Workbench for Swedish .....	703
<i>Dimitrios Kokkinakis, Maria Toporowska Gronostaj</i>	
Semiotic Conceptualization of Human Body: Lexicographical or Database System Description .....	713
<i>Grigory E. Kreydlin</i>	
<i>Repertorio analitico dei dizionari bilingui francese / italiano</i> .....	717
<i>Jacqueline Lillo</i>	
Ein elektronisches Lexikon im OLIF-Format für die Erzählanalyse .....	729
<i>Marc Luder, Simon Clematide, Bernhard Distl</i>	
Introducing BAWE: A New Lexicographical Resource .....	737
<i>Hilary Nesi</i>	
Development of the Integrated Concordancer for the Corpus of the 17th to 19th Century Culinary Manuscripts .....	741
<i>Doo-hyun Paek, Kil-im Nam, Mi-hyang Lee, Eui-jeong Ahn, Hyeon-ju Song</i>	
Diccionario de los glifos maya con descripción visual estructural .....	747
<i>Obdulia Pichardo-Lagunas, Grigori Sidorov</i>	
Presentación del <i>Diccionario Coruña de la lengua española actual</i> .....	753
<i>José-Álvaro Porto Dapena, M.<sup>a</sup> Eugenia Conde Noguero, Félix Córdoba Rodríguez, M.<sup>a</sup> Montserrat Muriano Rodríguez</i>	
Pedagogical Criteria for Effective Foreign Language Learning: A New Dictionary Model .....	763
<i>Dídac Pujol, Joan Masnou, Montse Corrius</i>	
On The Lexis and Cloth and Clothing Project .....	771
<i>Stuart Nels Rutten</i>	
ISLEX—An Icelandic-Scandinavian Multilingual Online Dictionary .....	779
<i>Aldís Sigurðardóttir, Anna Hannesdóttir, Halldóra Jónsdóttir, Håkan Jansson, Lars Trap-Jensen, Þórdís Úlfarsdóttir</i>	

e-LIS: Electronic Bilingual Dictionary Italian Sign Language-Italian .....	791
<i>Chiara Vettori, Mauro Felice</i>	
Verbs of Science and the Learner's Dictionary .....	797
<i>Geoffrey Williams</i>	
<i>Section 4. Bilingual Lexicography</i>	
La définition dans les dictionnaires bilingues: problèmes de polysémie et d'équivalence interlangues .....	807
<i>Samia Bouchaddakh</i>	
La place de la morphologie constructionnelle dans les dictionnaires bilingues: étude de cas .....	813
<i>Bruno Cartoni</i>	
Friend or Enemy? A Case Study of Lexical Comparison between Italian, German and Japanese Bilingual Dictionaries .....	821
<i>Mauro Costantino</i>	
User-friendly Dictionaries for Zulu: An Exercise in Complexicography ...	827
<i>Gilles-Maurice de Schryver, Arnett Wilkes</i>	
Systemhaftigkeit in zweisprachiger Lexikographie: Zur Darstellung deutscher und russischer Possessivpronomen .....	837
<i>Dmitrij O. Dobrovol'skij, Artëm V. Šarandin</i>	
La equivalencia en los diccionarios bilingües: un enfoque semántico .....	843
<i>Juan Fernández Fernández</i>	
Le Casse-tête des dictionnaires bilingues pour traducteurs: le cas des dictionnaires arabes bilingues .....	855
<i>Lynne Franjé</i>	
What to Say about <i>mañana</i> , <i>totems</i> and <i>dragons</i> in a Bilingual Dictionary? The Case of Surrogate Equivalence .....	869
<i>Rufus Gouws, Danie J. Prinsloo</i>	
Du support d'information à l'outil lexicographique —la lexicographisation du guide touristique .....	879
<i>Patrick Leroyer</i>	
Lexical Entries and the Component of Pronunciation in Tshivenda Bilingual Dictionaries .....	887
<i>Munzhedzi James Mafela</i>	

L'accès aux Séquences Figées dans les dictionnaires électroniques bilingues Français-Italien .....	895
<i>Michaela Murano</i>	
Méthode sociolinguistique d'étiquetage du niveau de langue dans les dictionnaires bilingues (sur l'exemple d'un dictionnaire français-ukrainien) .....	903
<i>Natalya Schevchenko</i>	
On the Presentation of Onomastic Idioms in Bilingual English-Polish Dictionaries of Idioms .....	909
<i>Joanna Szerszunowicz</i>	
<i>Section 5. Lexicography for Specialised Languages—Terminology and Terminography</i>	
QRcep: A Term Variation and Context Explorer Incorporated in a Translation Aid System on the Web .....	915
<i>Takeshi Abekawa, Kyo Kageura</i>	
ECODE: A Pattern Based Approach for Definitional Knowledge Extraction .	923
<i>Rodrigo Alarcón Martínez, Gerardo Sierra Martínez, Carme Bach Martorell</i>	
Environmental Terminology in General Dictionaries .....	929
<i>Araceli Alonso Campo</i>	
Gestor de terminologia multilingüe d'accés lliure .....	937
<i>Jordi Bover Salvadó, Marta Grané Franch</i>	
TESAURVAI: Herramienta para la extracción, anotación y organización de términos .....	941
<i>Jesús Cardeñosa, Carolina Gallardo, Ángeles Maldonado-Martínez, Jorge Vergara</i>	
Filling the Gap: A Three-Language Philological Dictionary Based on Contexts from Authoritative Sources .....	947
<i>Laura Cignoni</i>	
<i>Risotto, spaghetti, vino: Ingredients for a Good Gastronomic Dictionary .</i>	957
<i>Elisa Corino</i>	
Léxico específico de la piel. Presentación de un proyecto terminográfico ....	965
<i>María García Antuña</i>	

Slovene Terminology Web Portal .....	971
<i>Vojko Gorjanc, Simon Krek, Špela Vintar</i>	
Prototypes and Discreteness in Terminology .....	979
<i>Pius ten Hacken</i>	
New Voices in Bilingual Russian Terminography with Special Reference to LSP Dictionaries .....	989
<i>Olga Karpova, Konstantin Averboukh</i>	
LSP Dictionaries and their Genuine Purpose: A Frame-based Example from MARCOCOSTA .....	997
<i>Pilar León Araúz, Pamela Faber, Chantal Pérez Hernández</i>	
Marqueurs définitionnels et marqueurs relationnels dans les définitions du DAAFAPS .....	1007
<i>Pierluigi Ligas</i>	
A Constructional Approach to Terminological Phrasemes .....	1015
<i>Silvia Montero Martínez</i>	
Bilingual Terminology Acquisition from Unrelated Corpora .....	1023
<i>Rogelio Nazar</i>	
El sistema métrico decimal en la lexicografía española del s. XIX .....	1031
<i>Luisa Pascual Fernández</i>	
An English-Polish Glossary of Lexicographical Terms: A Description of the Compilation Process .....	1041
<i>Mirosława Podhajecka, Monika Bielińska</i>	
Wissensdarstellung und Benutzerfreundlichkeit in einem zweisprachigen terminologischen Rechtswörterbuch: Der Fall Hochschulrechtes .....	1051
<i>Natascia Ralli, Tanja Wissik</i>	
Palabras y términos “lingüística y contextualmente determinados” .....	1057
<i>Gemma Sanz Espinar</i>	
Lexicographic Document Templates: Text Genre Conventions in Corporate Lexicography .....	1065
<i>Henrik Køhler Simonsen</i>	
Terminology Practice in a Non-standardized Environment: A Case Study .....	1073
<i>Elsabe Taljard</i>	

*Section 6. Historical and Scholarly Lexicography and Etymology*

La reforma pombalina de la enseñanza: de la <i>Prosodia</i> de Bento Pereira al <i>Parvum Lexicon</i> de Pedro da Fonseca .....	1081
<i>Ana Margarida Borges</i>	
Un diálogo implícito: la relación entre Joan Corominas y José Luis Pensado a través de su producción lexicográfica .....	1097
<i>Rosalía Cotelo García</i>	
Velázquez de la Cadena y la lexicografía bilingüe inglés/español .....	1105
<i>Cecilio Garriga, Raquel Gállego</i>	
Description of Loan Words in French School Dictionaries: Treatment of Words of Foreign Origin in <i>Dictionnaire Hachette junior</i> (2006) and <i>Le Robert junior illustré</i> (2005) .....	1115
<i>Nathalie Gasiglia</i>	
The Multiword Expressions in the TLIO: Lexicographic Practises, Data and Classification Criteria .....	1123
<i>Mariafrancesca Giuliani</i>	
GASTEREA: Digital Diachronic Thesaurus of Latin Food Words and their Heritage in European Languages .....	1139
<i>Alexandra Grigorieva, Svetlana Hautala, Natalia Romanova</i>	
The Role of <i>Foclóir Gaeilge-Béarla</i> Néill Uí Dhónaill in Irish Language Lexicography in the Twentieth Century .....	1149
<i>Liam MacAmhlaigh</i>	
Macro- and Microstructure Experiments in Minor Bilingual Dictionaries of XIX and XX Century .....	1155
<i>Carla Marello, Marco Tomatis</i>	
Le programme TLF-Étym: apports récents de l'étymologie comparée-reconstruction .....	1165
<i>Gilles Petrequin, Marta Monda Andronache</i>	
De la 1 <sup>re</sup> à la 2 <sup>e</sup> édition du <i>Dictionnaire de l'Académie française</i> : marques diastratiques et diaphasiques .....	1175
<i>Marie-Alix Poteaux, Louise Dagenais</i>	

Primer contacto de las lenguas española e inglesa: el <i>Sex linguarum Latinae, Teuthonice, Gallice, Hispanice, Italice, Anglice, dilucidissimus dictionarius</i> .....	1181
<i>M. Jesús Redondo Rodríguez</i>	
L'informatisation du FEW: attentes et modélisation .....	1189
<i>Pascale Renders, Christelle Nissille</i>	
El tratamiento de los números en el diccionario .....	1199
<i>Francesc Rodríguez Ortiz, Cecilio Garriga Escribano</i>	
Le DÉCT ( <i>Dictionnaire Électronique de Chrétien de Troyes</i> ): un modèle pour la lexicographie d'aujourd'hui? .....	1203
<i>Pierre Souvay Gilles, Pierre Kunstmann</i>	
<i>Vulgar and popular</i> in Johnson, Webster and the OED .....	1209
<i>Kate Wild</i>	

#### *Section 7. Dictionary Use*

Papel de los diccionarios de colocaciones en la enseñanza de español como L2 .....	1215
<i>Margarita Alonso Ramos</i>	
Frequency in Learners' Dictionaries .....	1231
<i>Paul Bogaards</i>	
United in Diversity: Dutch Historical Dictionaries Online .....	1237
<i>Katrien Depuydt, Jesse De Does</i>	
Noun and Verb Codes in Pedagogical Dictionaries of English: User-friendliness Revisited .....	1243
<i>Anna Dziemianko</i>	
Teaching the Systematic Dictionary Use as a strategy for Accuracy and Confidence Building .....	1251
<i>Penelope Kambaki-Vougioukli</i>	
Improving Dictionaries .....	1259
<i>Ari Kernerman</i>	
Teaching Dictionary-using Skills for Online Dictionaries—An Attempt at a Theoretical Framework for South Africa .....	1265
<i>Juliane Klein</i>	

Can Dictionary Skills Be Taught? The Effectiveness of Lexicographic Training for Primary-school-level Polish Learners of English: .....	1273
<i>Robert Lew, Katarzyna Galas</i>	
Looking Up “Hard Words” for a Production Test: A Comparative Study of the NOAD, MEDAL and MW Collegiate Dictionaries.....	1287
<i>Don R. McCreary</i>	
Giving Them What They Want: Search Strategies for Electronic Dictionaries .....	1295
<i>Michal Boleslav Měchura</i>	
Adverb Use in EFL Student Writing: From Learner Dictionary to Text Production .....	1301
<i>Gill Philip</i>	
The Electronic Dictionary in the Language Classroom: The Views of Language Learners and Teachers .....	1311
<i>James Ronald, Shinya Ozawa</i>	
Mother-tongue’s Little Helper (The Use of the Monolingual Dictionary of Slovenian in School) .....	1317
<i>Tadeja Rozman</i>	
<i>Section 8. Phraseology and Collocation</i>	
Las unidades fraseológicas eventivas en los diccionarios bilingües Español-Vasco .....	1325
<i>Axun Aierbe Mendizabal</i>	
<i>REDES. Diccionario combinatorio del español contemporáneo</i> .....	1333
<i>Nieves Almarza Acedo, Yolanda Lozano Ramírez de Arellano</i>	
Propuesta de anotación semántica para una base de datos paremiológica .....	1337
<i>Elena Alonso Pérez-Ávila</i>	
From Dictionary to Phrasebook? .....	1345
<i>Sylviane Granger, Magali Paquot</i>	
Collocational false friends: description and treatment in bilingual dictionaries .....	1357
<i>Ulrich Heid, Danie J. Prinsloo</i>	

Analysis of Collocations in Russian: Corpus vs. Dictionary .....	1365
<i>Maria Khokhlova</i>	
The Lemmatisation of Lexically Variable Idioms: The Case of Italian-English Dictionaries .....	1373
<i>Chris Mulhall</i>	
Proyecto para la redacción de un diccionario de locuciones del español ..	1379
<i>Inmaculada Penadés Martínez</i>	
A Comparative Analysis of Definitions of Phrasal Verbs in Monolingual General-purpose Dictionaries for Native Speakers of American and British English .....	1385
<i>Magdalena Perdek</i>	
Inclusión de los papeles semánticos de FrameNet en DiCE .....	1393
<i>Sabela Prieto González</i>	
Colocaciones léxicas en diccionarios generales monolingües del español .....	1401
<i>Laura Romero Aguilera</i>	
<i>Una bella esperienza, una buona prova. A corpus analysis of purely evaluative adjectives in Italian .....</i>	1409
<i>Irene Russo</i>	
From Subdomains and Parameters to Collocational Patterns—On the Analysis of Swedish Medical Collocations .....	1421
<i>Emma Sköldbberg, Maria Toporowska Gronostaj</i>	
Aspectos de fraseografía bilingüe español-alemán: la equivalencia frente a la definición .....	1433
<i>Aina Torrent-Lenzen</i>	
A Multilingual Electronic Database of Distributionally Idiosyncratic Lexical Items .....	1445
<i>Beata Trawiński, Jan-Philipp Soehn, Manfred Sailer, Frank Richter</i>	
For an Extended Definition of Lexical Collocations .....	1453
<i>Agnès Tutin</i>	
SciE-Lex: A Lexical Database of Collocations in Scientific English for Spanish Scientists .....	1461
<i>Isabel Verdaguer, Anna Poch, Natalia-Judith Laso, Eva Giménez</i>	

Database of Bavarian Dialects (DBÖ) Electronically Mapped ( <i>dbo@ema</i> ). A System for Archiving, Maintaining and Field Mapping of Heterogeneous Dialect Data .....	1467
<i>Günther Fliedl, Marcus Hassler, Christian Kop, Heinrich C. Mayr, Jürgen Vöhringer, Eveline Wandl-Vogt</i>	
Incomprehensible Languages in Idioms—Functional Equivalents in Bilingual Dictionaries .....	1473
<i>Monika Woźniak</i>	
<i>Section 9. Lexicological Issues of Lexicographical Relevance</i>	
Prepositions in Dictionaries for Foreign Learners: A Cognitive Linguistic Look .....	1477
<i>Arleta Adamska-Salaciak</i>	
Lexicographie historique, noms de métier, féminisation: quelle méthodologie? .....	1487
<i>Fabienne Baider</i>	
Scale-free Networks in Dictionaries .....	1499
<i>Ágota Fóris</i>	
La place du métalangage dans la définition lexicographique: l'exemple des définitions des mots syncatégorématiques dans le TLF .....	1505
<i>Paolo Frassi</i>	
Verbal Aspect and the Frame Elements in the FrameNet for Polish .....	1511
<i>Jadwiga Linde-Usiekniewicz, Magdalena Derwojedowa, Magdalena Zawistawska</i>	
Verbos que traban discurso: implicaciones lexicográficas para el DAELE .....	1519
<i>Carmen López Ferrero, Sergi Torner Castells</i>	
Verb Class-specific Criteria for the Differentiation of Senses in Dictionary Entries .....	1529
<i>Kristel Proost</i>	
Les dictionnaires québécois et le problème de la norme linguistique .....	1537
<i>Elmar Schafroth</i>	
On the Desirability of Using Labels in Dictionaries .....	1547
<i>Geart Van der Meer</i>	

*Section 10. Other Topics*

Definición lexicográfica y orden de la información de las palabras: el caso del euskera .....	1557
<i>Xabier Alberdi Larizgoitia, Julio García García de los Salmones, Iñaki Ugarteburu Gastañares</i>	
From Lexicographic Evidence to Lexicological Aspects: A Cognitive Linguistic Perspective on Phonaesthetic Intensifiers .....	1565
<i>Silvia Cacchiani</i>	
Dictionaries for University Students: A Real Deal or Merely a Marketing Ploy? .....	1575
<i>Iztok Kosem</i>	
El tratamiento lexicográfico de <i>de toute façon, de quelque façon</i> y <i>d'une certaine façon</i> en el DEC .....	1585
<i>Ana Llopis Cardona</i>	



## Presentació

La sèrie Activitats de les publicacions de l'Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra és una iniciativa que pretén la difusió de les ponències i seminaris celebrats i organitzats per l'IULA: jornades de treball, simposis, col·loquis, congressos i tota mena d'accions d'intercanvi i difusió.

Dins de l'àmbit general de la lingüística aplicada, els volums de la sèrie presenten les actes d'activitats de caràcter monogràfic, sobre els temes propis de la recerca del centre: lingüística de corpus, lingüística computacional, enginyeria lingüística, terminologia, morfologia, neologia, lexicografia, anàlisi del discurs general i especialitzat, variació lingüística o representació del coneixement. D'entre aquestes activitats, cal destacar per la seva periodicitat les Jornades de corpus lingüístics, iniciades el 1993; les diferents edicions del Simposi Internacional de Terminologia, iniciat el 1997; les Jornades de Lexicografia, iniciades el 1999; i el Seminari de Neologia, iniciat el 2000.

Tots els textos d'aquesta sèrie es publiquen en les llengües originals en què van ser presentats pels seus autors i han estat revisats per ells mateixos. L'edició de cada publicació està coordinada per algun dels professors que va formar part del comitè d'organització de l'activitat. L'edició és finançada amb l'ajut institucional que va donar suport a l'activitat, i les tasques d'edició i de coordinació de cada volum són realitzades directament per membres de l'Institut Universitari de Lingüística Aplicada.

Esperem que aquesta iniciativa, que va començar l'any 1996, continuï arribant als investigadors, professors i estudiants interessats en la lingüística aplicada i contribueixi a difondre en la nostra comunitat temàtiques encara poc consolidades.



## Foreward

The year 2008 marks the twenty-fifth anniversary of the founding of the European Association for Lexicography, EURALEX, in Exeter, England. In these twenty-five years the association has grown considerably, to become an active academic organisation with members around the world and a highly respected journal (the *International Journal of Lexicography*, published by Oxford University Press). Three years passed between the first meeting organised by Reinhard Hartmann in 1983 and second EURALEX international congress in Zurich in 1986, and since then the association has organised its international congress every two years. On behalf of the organising committee of the XIII EURALEX International Congress held July 15-19 at Universitat Pompeu Fabra in Barcelona, it is my pleasure to present this volume of proceedings, which includes the plenary papers, contributed papers, posters, and summaries of software demonstrations.

EURALEX congresses are true to the word “international” in their title: the contributors to these proceedings come from all around the world, and the number of languages discussed in the papers is high. These proceedings contain papers written in six different languages—Catalan, English, French, German, Italian, and Spanish. This significant presence of many languages seems only fitting for a European association focused on the study and representation of words, and signals that interest in researching and improving dictionaries is widespread.

The Institut Universitari de Lingüística Aplicada (IULA) of Universitat Pompeu Fabra is especially honoured to host the XIII EURALEX International Congress. This marks the first time the congress has been held in Catalonia, and the second time it has been held in Spain (the 1990 congress was organised by Manuel Alvar Ezquerro in Málaga). Our university proudly carries the name of Pompeu Fabra, the leading figure in the movement to create a literary standard for Catalan in the early 20th century. Pompeu Fabra is known in Catalan culture for his work to establish a single, unified orthography for all dialects of the language, for setting down the grammar that the Catalan Academy of Sciences, the Institut d’Estudis Catalans, would adopt for the standard language, and—importantly for us—for writing a dictionary. The *Diccionari general de la llengua catalana*, published in Barcelona in 1932, is fondly referred to as the ‘diccionari Fabra’ and was the dictionary of reference for standard Catalan for over sixty years. Since its creation, our university institute IULA has offered post-graduate training in lexicography, and dictionaries lie at the centre of our research group InfoLex; in addition to carrying out research, several of our members have extensive practical experience in writing Catalan and Spanish dictionaries. In this context, it seemed fitting for Universitat Pompeu Fabra—one of the few universities in the world named after a lexicographer, with a university institute committed to applied linguistics—to organise a congress celebrating 25 years of researching dictionaries.

This volume contains the five plenary lectures given at the congress, which we publish here in the same order as they were presented. All of these papers represent significant contributions by well-known figures in lexicography; here I shall only name the authors and briefly mention the topics they discuss because the papers speak for themselves. Joaquim Rafel i Fontanals of the Institut d'Estudis Catalans and the Universitat de Barcelona opened the congress with an excellent overview of Catalan lexicography, from the earliest bilingual glossaries to the Institut's current dictionary and corpus projects. Charles J. Fillmore, Professor Emeritus of the University of California at Berkeley, has long been interested in applying advances in grammatical analysis to dictionary representation and in this paper discusses the relationship of his FrameNet project to dictionaries. José Antonio Pascual of the Real Academia Española discusses several issues that commonly occur in historical lexicography in the context of the Academia's historical dictionary, which is undoubtedly the most ambitious dictionary project ongoing today in Spanish. For some years now, the Hornby Trust has generously contributed to EURALEX congresses by sponsoring a lecture in honour of A. S. Hornby, the pioneering figure in learner's dictionaries for non-native speakers. In the Hornby lecture, Patrick Hanks of Masaryk University discusses his recent research using large corpora to identify verb meanings and patterns to the tradition of learner's dictionaries. The closing lecture at the congress was given by R. R. K. Hartmann, whose initiative twenty-five years ago resulted in the creation of EURALEX and is one of the figures most responsible for establishing dictionary research as an academic discipline. The organising committee would like to extend its sincere thanks to all of the plenary speakers for setting the tone for this volume, which we believe represents a significant contribution to the literature on dictionary research.

Submissions to EURALEX 2008 belong to several categories: full papers, short papers, student papers, posters, and software demonstrations. The papers in these proceedings are organised into the following sections, which were the basis for the congress's call-for-papers:

Computational Lexicography and Lexicology

The Dictionary-Making Process

Reports on Lexicographical and Lexicological Projects

Bilingual Lexicography

Lexicography for Specialised Languages – Terminology and Terminography

Historical and Scholarly Lexicography and Etymology

Dictionary Use

Phraseology and Collocation

Lexicological Issues of Lexicographical Relevance

Other topics

The papers are organised alphabetically by first author within each section on the CD-ROM included with this volume.

This volume is a clear indication that the European Association for Lexicography has fully entered the digital age. The number of participants in EURALEX congresses has grown steadily in the past 25 years; over 350 people were expected to attend the 2008 congress. As EURALEX congresses have grown, so has the number of pages in the proceedings. The sheer size of this congress, coupled with the fact that the proceedings needed to be ready for distribution the first day of the congress, convinced us that we had to make maximum use of computerised resources. The EURALEX 2008 organising committee, in consultation with the EURALEX Executive Board, thus implemented several changes from past practice that merit mention here. First, the format of the proceedings has changed. The book you are reading contains the five plenary papers presented at the congress, the abstracts of all the papers, posters, and software demonstrations accepted for presentation, and an index to the accompanying CD-ROM that contains the bulk of contributions. This hybrid print/CD-ROM format offered two main advantages: lower production costs and increased flexibility in paper length, as most papers were no longer printed. In addition, we hoped that use of the digital format for most papers would reduce the number of errata.

A further change that proved important for the organising committee was the decision to use software to manage most aspects of paper submissions, congress registration, certificates, and mailing lists. The previous experience of colleagues at the Institut Universitari de Lingüística Aplicada with software written for our Institut's activities by Jesús Carrasco to manage the tasks associated with congress organisation convinced us that we needed to adopt those measures for EURALEX 2008. By using this computerised system, we were able to meet our goal of increasing the number of languages used in EURALEX congress organisation to five (Catalan, English, French, German, and Spanish, the languages of the congress website). The system also allowed us to efficiently track all aspects of the review process, the production of the proceedings, and congress registration.

The proceedings of EURALEX congresses have become important references to consult in dictionary research because of the consistently high quality of the papers submitted. On behalf of everyone associated with the organisation of EURALEX 2008 at Universitat Pompeu Fabra, I would like to thank all the contributors for submitting very interesting work, and for meeting the tight production schedule of these proceedings.

### ***Review procedure for EURALEX 2008***

The call-for-papers for EURALEX 2008 was extremely successful: the organisers had received over 300 submissions by the end of November, 2007, when the call was closed. All submissions were sent via e-mail to us and became part of a database. At the beginning of December, 2007, I assigned each submission to two referees. The referees received the anonymous abstracts, most of which were 6-8 doubled-spaced pages long, along with the corresponding evaluation forms. Once we received the completed evaluation forms, software compiled all the information. We were thus able to have two independent blind reviews for all 300 submissions in time

for the meeting of the programme committee in Barcelona at the end of January, 2008. Neither the identity of the reviewers nor that of the authors was revealed on the documentation given to the members of the programme committee to ensure anonymous evaluation. The members of the programme committee discussed each and every submission individually, and decided which papers would be accepted. The decisions were quite difficult and more than half the submissions had to be rejected because of time and space constraints at the congress.

The use of this double-blind peer review procedure has ensured the academic excellence of the papers, posters, and demonstrations presented at EURALEX 2008 and contained in these proceedings. I am pleased to thank the following people who participated in this review process:

Arleta Adamska-Sałaciak (Uniwersytet im. Adama Mickiewicza w Poznaniu)  
Toni Badia (Universitat Pompeu Fabra)  
Paz Battaner (Universitat Pompeu Fabra)  
Elisenda Bernal (Universitat Pompeu Fabra)  
Paul Bogaards (Universiteit Leiden and editor, *International Journal of Lexicography*)  
Anna Braasch (Center for Sprogteknologi, Københavns Universitet)  
František Cermák (Univerzita Karlova v Praze)  
Marie-Hélène Corréard, lexicographer and ex-President, EURALEX  
Gilles-Maurice de Schryver (Universiteit Gent)  
Janet DeCesaris (Universitat Pompeu Fabra)  
Anne Dykstra (Fryske Akademy)  
Ruth Fjeld (Institutt for lingvistiske og nordiske studier)  
Thierry Fontenelle (Microsoft Corporation)  
Cristina Gelpí (Universitat Pompeu Fabra)  
Rufus Gouws (Stellenbosch University)  
Pius ten Hacken (University of Wales, Swansea)  
Patrick Hanks (Masarykova Univerzita)  
Reinhard Hartmann (University of Exeter, emeritus)  
Ulrich Heid (Universität Stuttgart)  
Mercè Lorente (Universitat Pompeu Fabra)  
Henrik Lorentzen (Det Danske Sprog- og Litteraturselskab)  
Carla Marellò (Università di Torino)  
Don McCreary (University of Georgia)  
Rosamund Moon (University of Birmingham)  
José Ignacio Pérez Pascual (Universidade da Coruña)

Michael Rundell (Lexicography MasterClass)  
Joan Soler (Institut d'Estudis Catalans)  
Lars Trap-Jensen (Det Danske Sprog- og Litteraturselskab)  
Geart Van der Meer (Rijksuniversiteit Groningen)  
Serge Verlinde (Université catholique de Louvain)  
Leo Wanner (ICREA-Universitat Pompeu Fabra)  
Geoffrey Williams (Université de Bretagne-Sud)

The following people joined me on the EURALEX 2008 programme committee:

Anna Braasch, Københavns Universitet  
Anne Dykstra, Fryske Akademy  
Carla Marello, Università di Torino  
Joan Soler i Bou, lexicographer, Institut d'Estudis Catalans  
Leo Wanner, ICREA-Universitat Pompeu Fabra  
Geoffrey Williams, Université de Bretagne-Sud

Their knowledge of the field and their experience with the association and EURALEX congress organisation enabled us to put together what we trust you will find is a stimulating portrait of dictionary research as currently practiced.

### *Acknowledgements*

I would like to take this opportunity to acknowledge the various levels of support that our organising committee has received from Universitat Pompeu Fabra. I am pleased to acknowledge the enthusiastic support of the Rector of our university, Dr. Josep Joan Moreso. From the outset, our initiative to hold EURALEX 2008 at Universitat Pompeu Fabra was welcomed by both the former director of the Institut Universitari de Lingüística Aplicada (IULA), Dr. M<sup>a</sup> Teresa Turell, and by the IULA's current director, Dr. Mercè Lorente. In the fall of 2006, I called the first meeting of the EURALEX 2008 organising committee, and I am both pleased and proud to say that all members have actively participated in the many facets of the congress:

Araceli Alonso, Institut Universitari de Lingüística Aplicada, UPF  
Encarnación Atienza, Departament de Traducció i Interpretació, UPF  
Paz Battaner, Institut Universitari de Lingüística Aplicada, UPF  
Elisenda Bernal, Euralex 2008 Secretary, Institut Universitari de Lingüística Aplicada, UPF  
Cristina Gelpí, Institut Universitari de Lingüística Aplicada, UPF  
Carmen López, Departament de Traducció i Interpretació, UPF  
Guilhem Naro, Institut Universitari de Lingüística Aplicada, UPF  
Juan Manuel Pérez, Institut Universitari de Lingüística Aplicada, UPF

Irene Renau, Institut Universitari de Lingüística Aplicada, UPF

Joan Soler, Institut d'Estudis Catalans

Sergi Torner, Institut Universitari de Lingüística Aplicada, UPF

Maria Wirf, Departament de Traducció i Interpretació, UPF

As I mentioned above, we depended heavily on a computerised system to manage the review process and to register congress participants. Jesús Carrasco, our computer wizard at the Institut Universitari de Lingüística Aplicada, provided us with a superior tool that simplified our work. His attractive graphic designs, including the EURALEX 2008 logo and the multilingual website for the congress, were an added plus.

Congress organisation in a university setting involves a great deal of paperwork, and several members of the staff of the IULA—Vanessa Alonso, Jesús Carrasco, Sylvie Hochart, and Gemma Martínez—all contributed significantly to EURALEX 2008. Four undergraduate students of the Facultat de Traducció i Interpretació—Javier Gimeno, Albert Ibáñez, Alba Milà, and Emma Vila, as well as four of our graduate students at the IULA—Irene Renau and Alexandra Spalek also aided us in preparing congress materials, and Araceli Alonso and Juan Manuel Pérez aided us in editing the proceedings.

The organisation of EURALEX 2008 at Universitat Pompeu Fabra and the publication of these proceedings received financial support from the following sources, which we gratefully acknowledge:

Ministerio de Educación y Ciencia de España, *Acción complementaria no. HUM2007-30786-E*, awarded to IULA-UPF for the organisation of the XIII International Congress of the European Association for Lexicography – EURALEX.

Generalitat de Catalunya, Agència de Gestió d'Ajuts Universitaris i de Recerca, ARCS Grant no. ARCS2-00044, awarded to UPF for the organisation of the XIII International Congress of the European Association for Lexicography – EURALEX.

Ministerio de Educación y Ciencia de España, Proyecto HUM2006-07898 (*“Las categorías nombre y adjetivo en el Diccionario de aprendizaje del español como lengua extranjera”*, J. DeCesaris, Principal Researcher).

Ministerio de Educación y Ciencia de España, Proyecto HUM2006 –06982 (*“Las categorías verbo y adverbio en el Diccionario de aprendizaje del español como lengua extranjera”*, Paz Battaner, Principal Researcher).

The Hornby Educational Trust.

Institut d'Estudis Catalans.

Institut Ramon Llull, Generalitat de Catalunya.

Real Academia Española.

'la Caixa' Savings Bank.

Departament de Traducció i Filologia, Universitat Pompeu Fabra

Facultat de Traducció i Interpretació, Universitat Pompeu Fabra

Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra

I would also like to take this opportunity to acknowledge our sponsors from the publishing world, the names of which are listed separately at the conclusion of this foreward. Developments in publishing directly bear on dictionaries, and the willingness of publishers to sponsor and actively participate in EURALEX congresses is a clear sign that professional circles are interested in the association's activities and in the dictionary research being carried out in academic settings and elsewhere. Congress participants, in turn, greatly benefit from the book exhibits and demonstrations publishers present.

I am particularly indebted to three members of the organising committee whose work and dedication deserve special mention. Paz Battaner, Araceli Alonso, and Elisenda Bernal have accompanied me in the adventure of organising EURALEX 2008 from the very beginning. Elisenda especially is responsible for the timely production of these proceedings. It is a pleasure to thank them here.

Janet DeCesaris  
Chair, XIII EURALEX International Congress  
May, 2008

The Organising Committee of the XIII EURALEX International Congress is pleased to acknowledge the generous support of the following publishers:

Chambers Harrap Publishers Ltd.

Diccionarios SM

Enciclopèdia Catalana

HarperCollins Publishers

K Dictionaries

Langenscheidt KG

Larousse Editorial, S.L.

Sociedad General Española de Librería, S.A.

Oxford University Press

Walter de Gruyter GmbH & Co

## PLENARY LECTURES



# Tradition and Innovation in Catalan Lexicography\*

Joaquim Rafel i Fontanals

Universitat de Barcelona and Institut d'Estudis Catalans

The Organizing Committee of the XIII International Congress of the European Association for Lexicography has given me the honor of delivering the inaugural address. The organizers suggested to me that this occasion was an excellent opportunity to provide congress participants and readers of EURALEX volumes from around the world with a view of Catalan lexicography. The result is the following paper, which is divided into two parts. In the first part I provide a brief history of our lexicographic tradition; in the second, I discuss Catalan lexicography in the 20th century and concentrate on recent activities in the field of lexicography undertaken by the Institut d'Estudis Catalans. Our lexicographic tradition aims both to satisfy the specific needs of a society that uses Catalan as its language of communication and to adopt rigorous methodological criteria grounded in linguistic analysis. As I hope to show, our dictionaries are on a par with current lexicographic practise and incorporate several innovative characteristics.

## From the beginnings to the 19th century

In the early stages of the language, interest in lexis is no different in Catalan than it is in other European languages, especially Western European languages. Alongside medieval glosses in Latin, which were included to improve the comprehension of texts, glosses in Catalan appeared throughout the XII, XIII and XIV centuries. These are the first rudimentary manifestations of what would later become glossaries and bilingual dictionaries. These glosses, together with lexical equivalents in Catalan that appear in some grammars written to facilitate the study of Latin, are the first testimonies we have of interest in Catalan lexis, although these works are not truly lexicographical in nature (Colon & Soberanas 1985[1991]: 11-38).

Lexicographic works begin to appear in the 15th century in the form of bilingual dictionaries concentrating on Catalan and Latin. These dictionaries were designed for decoding or studying Latin. A significant exception worthy of note is the thematic Catalan-German/German-Catalan vocabulary published in 1502. This anonymous work is an adaptation of the Italian-German vocabulary compiled by Adamo di Roudila, which had been published some years earlier, in 1477 (Colon & Soberanas 1985 [1991]: 56-59). Besides this isolated case, the publication of bilingual dictionaries, first for Catalan and Latin and then for Catalan and Spanish,

---

\* I am grateful to the congress organisers for their kind invitation. The original paper was written and presented orally at the congress in Catalan, and I would like to take this opportunity to thank Janet DeCesaris for her English translation.

would continue through the mid 19th century. During this long period Catalan dictionaries are conceived of merely as collections of words with equivalents in other languages (in most cases, Latin and Spanish). Among the most noteworthy books were the *Liber elegantiarum* by Joan Esteve, published in 1489, which is the first collection of phrases in Catalan with the corresponding phrases in Latin, and the Catalan adaptation by Gabriel Busa of Antonio Nebrija's Latin-Spanish/Spanish-Latin dictionary. Busa's adaptation was first published in 1507 and would be reedited several times; it was to be a strong influence on other dictionaries. We should also mention the *Tesaurus Puerilis* by Onofre Pou, published in 1575; the *Fons verborum et phrasium*, by Antoni Font, published in 1637; the *Thesaurus verborum ac phrasium* by Pere Torra, published in 1640; and the *Gazophilacium catalano-latinum*, by Joan Lacavalleria, published in 1696. The situation of Catalan lexicography in the 18th century is, according to Colon & Soberanas, disheartening, to say the least; there is the *Diccionario valenciano-castellano* by Carles Ros, which appeared in 1764, as well as a few other minor works. In all of these dictionaries, however, Catalan is treated as a function of the other languages, without the autonomy that would make it the object of study itself with lexicographic criteria, i.e., from the standpoint of the meaning of lexical units and use in Catalan.

This was the situation of Catalan lexicography when European languages in more important cultures were already the object of autonomous, independent lexicographic works—monolingual dictionaries. It is useful to recall here the most representative works produced in Europe in the 16th, 17th, and 18th centuries. First, there was the *Vocabolario degli academici della Crusca*, written between 1591 and 1612, which was to become the model and stimulus for several dictionaries of reference in other languages; in French, there appear three dictionaries at the end of the 17th century that are testimonies to three different attitudes towards language and social position: the *Dictionnaire français* by Pierre Richelet (1680), the *Dictionnaire Universel* by Antoine Furetière (1690) and the *Dictionnaire de l'Académie Française* (1694) (Collinot-Mazière 1997: 25). In the case of Spanish, there was the *Tesoro de la Lengua Castellana o Española* by Sebastián de Covarrubias (1611) and later the first dictionary of the Spanish Academy (*Diccionario de la lengua castellana, en que se explica el verdadero sentido de las voces, su naturaleza y calidad, con las frases o modos de hablar, los proverbios o refranes y otras cosas convenientes al uso de la lengua*), commonly known as the *Diccionario de Autoridades*, which was published in six volumes between 1726 and 1739. Samuel Johnson published his *Dictionary of the English Language* a few years later, in 1755, and Johan Cristoph Adelung published his *Grammatisch-kritisches Wörterbuch der Hochdeutschen Mundarten* between 1774 and 1786.

We find nothing similar for Catalan during this period. A series of mainly socio-political factors was responsible for this situation. I shall not go into detail here, but suffice it to say that during this time (16th-18th centuries), the use of Catalan in writing and in the public arena declined in favour of Spanish. This process, by which Spanish supplanted Catalan in many contexts, reached its high point at the beginning of the 18th century with the *Nova Planta* ('New Design') decrees (1707

in Valencia, 1716 in Catalonia and 1717 in Majorca and Eivissa). In addition, as the Bourbon dynasty came to power in Spain, more laws were put into place to prevent Catalan from having any role in education or in public administration.

At the beginning of the 19th century, after a series of unsuccessful initiatives (especially, the dictionary projects of the Acadèmia de Bones Lletres de Barcelona ('The Academy of Letters of Barcelona'), the 2-volume *Diccionario catalan-castellano-latino* by Joaquim Esteve, Josep Bellvitges and Antoni Juglà is appeared (in 1803 and 1805). This dictionary, however, is not a monolingual dictionary along the lines of the European dictionaries mentioned above; rather, it is a trilingual dictionary with brief definitions and notes in Catalan for only a part of the entries. The fact that the title of the dictionary is in Spanish gives us an indication of the social status of Catalan at that time. Nevertheless, this dictionary represents the first and only successful effort in Catalan 17th century lexicography, despite the fact the dictionary was printed at the beginning of the 18th century (Colon & Soberanas 1985 [1991]: 135). This dictionary, even though it was not written by a language academy, was written by people who were members of the Acadèmia de Bones Lletres and who had belonged to the academic commissions that had (unsuccessfully) tried to produce a dictionary for the Acadèmia. The *Diccionario catalan-castellano-latino* has been described as the starting point for modern Catalan lexicography (Colon & Soberanas 1985 [1991]: 144) and we can consider it an indirect testimony to what might have been the first institutional or academic dictionary for the language.

We must look to the mid-19th century to find the first dictionary of Catalan with all the words defined in the same language as the headwords. The work by Pere Labèrnia, the *Diccionari de la llengua catalana, ab la correspondència castellana y llatina*, published in 1839-1840, is truly conceived of as a monolingual dictionary, although—as the title indicates—it also contains translation equivalents in Spanish and Latin. Pere Labèrnia was also a member of the Acadèmia de Bones Lletres and his dictionary has a certain institutional flavour to it; he dedicates the dictionary to the Acadèmia and asks the Academy to endorse it (see Rafel 2004: 108 for further discussion). The front page of the dictionary explicitly states that the dictionary is published under the auspices of the Acadèmia de Bones Lletres. It is important to stress that Labèrnia's work was the reference dictionary for Catalan for many years. Several subsequent editions appeared, and this dictionary was the basis for the encyclopaedic dictionary published by the Salvat publishing firm at the beginning of the 20th century (Colon & Soberanas 1985 [1991]: 155-156).

### **The first decades of the 20th century**

The beginning of the 20th century was witness to two events that were to have a significant impact on Catalan lexicography.

The first important event was an initiative by Antoni Maria Alcover, published in 1901 in the document *Lletra de convit* ('Invited letter'), to undertake a large dictionary project. This dictionary project, as described by Alcover, would include Old Catalan, modern Catalan (with forms from the oral and written language),

literary variants and all the dialectal forms found within the language's geographic area (Alcover 1911: 370).

The second event was the creation of the Institut d'Estudis Catalans, the Catalan Academy of Sciences, in 1907. The Institut is an academic, scientific and cultural institution for scientific research on all aspects of Catalan culture. The goals of the Institut, according to the by-laws, are as follows: a) to study the Catalan language, establish its standard grammar and vocabulary and ensure that the standardisation of the language is coherent in all geographic areas in which the language is spoken; b) to contribute to the planning, coordination, and dissemination of research in science and technology; and c) to support the progress and development of society with its activities and to provide advisory studies for public officials and other institutions, as may be necessary. The Institut is currently structured into five thematic sections: History-Archaeology, Biological Sciences, Science and Technology, Philology, and Philosophy and Social Sciences. Although the establishment of the standard language is the responsibility of the Institut as a whole, this specific task—in addition to other language-related work—is of special concern to the Philological Section, which was created in 1911.

What we know about the early days of the Philological Section's activity is that it assumed Alcover's 1901 project as its own. It also approved the preparation of a *Diccionari català provisional* ('Provisional Catalan Dictionary') based on the written language with little attention given to dialectal varieties; this dictionary was to include the most common words of the literary language. Seen in retrospect, we can understand these decisions as a sort of compromise, or as a way to satisfy two quite different positions on the language and the actions that were considered most necessary at that point in time. The two opposing positions were represented by Alcover (whose dictionary would include all variants of the language, including geographic and historical forms) and by Pompeu Fabra (whose dictionary would unify and, to some extent, purify the language) (Rafel 1996: 219-220 and ensuing discussion).

It would take us too far afield to discuss the various vicissitudes these two projects experienced over the years. I shall only mention here that Alcover left the Institut in 1918 because of disagreements with other members and took with him the materials gathered up to that time. Francesc de Borja Moll continued this project after Alcover's death in 1932, and the result is the *Diccionari català-valencià-balear*, which was published in 10 volumes between 1926 and 1962. This important work is considered to be one of the most complete collections of lexical items and expressions in the Romance languages, and is particularly rich in data from Old Catalan and Catalan dialects.

The second project, led by Pompeu Fabra, became part of the Institut's task of establishing the standard language. The first important tangible result was the *Diccionari ortogràfic* ('Spelling Dictionary'), which appeared in 1917 under the direction of Fabra. The publication of this work was extremely important: on the one hand, it represented the consolidation of the orthographic rules that had been

approved in 1913; on the other, it meant that the form of the words considered most necessary for communication had been agreed upon and as such responded to the Institut's mission of providing a standard for the language. The Philological Section continued working on the language's dictionary of reference, which was known as the *Diccionari de la llengua literària* ('Dictionary of the Literary Language'). Work on this dictionary, however, did not progress as quickly as expected, mostly because of the inherently slow nature of academic work. The dictionary had been begun (up to the letter C) and the first fascicles were about to be published when the dictatorship of Primo de Rivera (1923-1930) rose to power in Spain. The dictatorship was opposed to Catalan language and culture, and consequently work on the dictionary ceased. Given the need for a standard reference dictionary, Pompeu Fabra took on the project himself, with financial backing from private sources. He used and completed the materials that had been collected for the *Diccionari de la llengua literària*, and the result was the publication of the *Diccionari general de la llengua catalana* in 1932. The dictionary appeared at the beginning of Spain's republican period, at a time when society was restoring Catalan institutions and the language was beginning to return to its proper social status. In this context, Fabra's dictionary provided an essential service to Catalan society. The Philological Section corrected and added new entries to the dictionary for the 1954 edition, and several later editions of the dictionary were published, up through the 32nd edition (which appeared in 1994 (Mir & Solà: 11). Fabra's dictionary was the reference dictionary for the standard language until the Institut d'Estudis Catalans published an updated dictionary in 1995. Dictionaries that Fabra consulted and took as models include the *Diccionario de la lengua española* of the Spanish Royal Academy (1925), *Webster's New International Dictionary of the English Language* (1909), and especially the *Dictionnaire general de la langue française* by A. Hatzfeld, A. Darmesteter and A. Thomas (no date, published in fascicles between 1890 and 1900) (Colon 2007).

### **The second half of the 20th century**

It is well known that the results of the Spanish Civil War were extremely prejudicial for Catalan culture in general and for the Catalan language particularly. The Institut d'Estudis Catalans was officially disbanded, although some clandestine activity was carried out and even tolerated at certain points. This illegal status of the Institut began at the end of the Spanish Civil War in 1939 and continued to the beginning of the transition to democracy in 1976, when a royal decree returned the Institut to its original status.

During this time, as I mentioned above, the Philological Section revised and published the second edition of Fabra's dictionary, which would be reprinted several times after 1962. In the following years, the Philological Section limited itself to publishing short lists of words that were circulated informally, and these entries were incorporated in the dictionary as of the 4th edition (1966) and in an appendix to it as of the 5th edition (1968), which commemorated the centennial of Fabra's birth.

At the same time private publishing firms begin to publish bilingual dictionaries and occasionally a monolingual dictionary, which was always much indebted to Fabra's work. Of special note here is the publication in 1982 of the *Diccionari de la llengua catalana* by the firm Enciclopèdia Catalana. This dictionary includes all the entries in Fabra's dictionary, and adds a substantial number of new words that the company was able to obtain as it worked on the publication of the *Gran enciclopèdia catalana*, a 15-volume work published between 1962 and 1980. For several years, this initiative was particularly important in Catalonia because it assumed a leading role in academic circles, because the Institut had not yet returned to its normal research activities. In the 1980s, by which time Catalan had regained a public presence in the media, in education, in the public administration, etc., several dictionaries of varying types and length are published. Special mention should be given to the *Gran diccionari de la llengua catalana*, published in 1988 by Enciclopèdia Catalana, because it includes many more entries than the earlier dictionary. We must also note the many dictionaries published by the firm Edicions 62 in the year 2000 (*Gran diccionari 62 de la llengua catalana*, *Diccionari 62 manual de la llengua catalana*, *Diccionari 62 essencial de la llengua catalana* i *Petit diccionari 62 de la llengua catalana*). More details about these and other dictionaries of Catalan may be found in Haensch (1990) and, for those dictionaries published between 1940 and 1988, in Cabré & Lorente (1991).

None of these projects represented progress in methodological terms. In many ways, lexicography in Catalan was at a standstill. The dictionaries published by commercial companies during this period did not show much interest in incorporating methodological innovations. In fact, they did quite the opposite: they tended to adopt a conservative attitude, maintaining traditions acquired over the years. I can only agree with Colon & Soberanas when they somewhat discouragingly state that Catalan lexicography did not respond to the current demands of modern society.

This situation was all the more disheartening given that the 20th century was witness to many significant advances in linguistics, almost none of which had an impact on our lexicographic practise. Moreover, in the second half of the 20th century many researchers discussed lexicographic theory and methods that could be used in dictionary preparation, which resulted in a significant bibliography; practically none of these works had any influence on the dictionaries named above. The technological revolution, with its many advances in computerisation, permits a type of language processing that is directly applicable to lexicography, but in our context these advances had only a minor effect. Colon, as far back as 1977, noted the need for modernisation of Catalan lexicography: "el català no hauria de trobar-se absent de cap de les metodologies modernes" ('Catalan should not be without any modern technology') (Colon 1977: 16).

## Recent dictionary projects undertaken by the Institut d'Estudis Catalans

As the Institut d'Estudis Catalans slowly moved towards regaining its place as a research institution, in 1983 the Philological Section of the Institut d'Estudis Catalans debated the best way to organise its activities in the field of lexicography. It decided to undertake a major dictionary project employing modern lexicographic techniques and principles. This project would be called the *Diccionari del català contemporani*, and aimed both to incorporate the main advances in linguistic analysis and to apply the possibilities of modern technology. At the same time, while this project was in progress, the Philological Section would also produce a new edition of Pompeu Fabra's dictionary, which was still the reference for the standard language even though it was clearly out-of-date.

At this point, then, the Institut d'Estudis Catalans becomes very active in lexicography. The by-laws of the Institut state that its missions include (1) setting the standard language and (2) describing actual use of the language. In terms of the lexicon, the Institut addressed these missions by undertaking two dictionaries: the *Diccionari de la llengua catalana* ('Dictionary of the Catalan Language') as the reference for the standard language and the *Diccionari del català contemporani* ('Contemporary Catalan Dictionary') project as the description of the language's vocabulary based on actual usage.

### *The Diccionari de la llengua catalana ('Dictionary of the Catalan Language')*

Once institutional activity in Catalan was reinstated, the society needed an up-to-date dictionary for the standard language. The solution adopted by the Institut was to write a dictionary based on Pompeu Fabra's *Diccionari general de la llengua catalana* that would not change the basic criteria that Fabra had used; rather, the new dictionary would include new vocabulary and update the dictionary. Communicative needs had changed dramatically since 1932, and as a result a large number of new words and senses had come into the language. Many of these words and senses had been included in commercial dictionaries but were not present in the official dictionary of the standard language. This updated version of Fabra's work was published in 1995.

Following the publication of the *Diccionari de la llengua catalana*, the Philological Section decided that it needed to publish a shorter dictionary concentrating on commonly used vocabulary, a dictionary without so many technical words that could be used at schools and by people who did not need such a large work. It thus published the *Diccionari manual de la llengua catalana* ('Manual Dictionary of the Catalan Language') in 2001. This dictionary is not simply an abridged edition of the earlier dictionary, but rather includes some changes designed to make it easier to consult for the general public.

As soon as the first edition of the *Diccionari de la llengua catalana* appeared in 1995, work began on preparing the second edition, which was published in March, 2007. At the same time, a digital version of the dictionary was made available free

to the public on Internet. The digital edition allows many complex searches that are particularly useful for language professionals (teachers, language consultants, editors, etc.).

### *The Diccionari del català contemporani ('Contemporary Catalan Dictionary')*

The *Diccionari del català contemporani* is the name of the ongoing project at the Institut d'Estudis Catalans. The Philological Section of the Institut realised that it needed to start a new period in its history, and that its work needed to benefit from the scientific, methodological, and technological advances that had occurred during the long period of forced inactivity. The project is structured into two stages: the first stage concentrated on developing language resources, while the second concentrates on producing a descriptive dictionary, the *Diccionari descriptiu de la llengua catalana*, based on those resources.

#### *The First Stage: Creating language resources for Catalan*

The first stage of the project, which is now completed, included the development of the Computerised Text Corpus of Catalan (in Catalan, the *Corpus Textual Informatitzat de la Llengua Catalana* or CTILC) and the Lexicographic Database (*Base de Dades Lexicogràfica* or BDLex), which contains the text of the most important dictionaries published in the 19th and 20th centuries.

##### *a) Main features of the CTILC*

From a chronological point of view, the CTILC includes texts spanning the period 1832-1988, in other words, more than 150 years in the history of Catalan. The date 1832 marks the beginning of the cultural movement to recover the use of the literary language, and 1988 was the date of the most recent texts included in the corpus. In terms of text typology, the corpus includes both literary and non-literary texts. These two groups were then divided into several subgroups, which allowed us to make a balanced choice of texts for each group (literary and non-literary) and for each subgroup (for literary texts, the four traditional genres: narrative, poetry, theatre, and essay; for non-literary texts, ten thematic domains: philosophy; religion and theology; social sciences; the press; pure and natural sciences; applied sciences; art, leisure and sports; language and literature; history and geography; and, correspondence). The corpus contains 52 million words, 44% of which correspond to literary text and 56% to non-literary text. The 3,299 texts in the corpus are of many different lengths. One of our main concerns was to ensure that the corpus would be as representative as possible, in other words, we wanted the set of texts included to reflect the written language as best as possible. Our objective was to strike a balance among the various text types that had been produced between 1832 and 1988.

A significant feature of the CTILC is that the entire corpus is lemmatised, which means that we have analysed each token and classified it grammatically so that it could be linked to a base reference form. This allowed us to achieve two important goals: on the one hand, graphic forms corresponding to different grammatical forms (such as verb forms belonging to a single conjugation or inflected forms

belonging to different lemmas) were disambiguated; on the other, forms belonging to an inflectional series were all linked to the lemma in question. The CTILC can be used in many different ways precisely because the entire corpus is lemmatised: for example, by looking up one lemma we obtain information on all the inflected forms, including the entire range of spelling variants that appeared in texts from different periods; we can also obtain all the derived forms containing evaluative or intensifying affixes which, according to morphological criteria in Catalan (as in other Romance languages) do not produce new lemmata. The number of lemmata in the corpus is 149,185, corresponding to 678,386 grammatical forms and 51,253,680 tokens.

The CTILC was developed during the period 1985-1997, which makes it a pioneering project in southern Europe. Once it was completed, a tool specially designed to consult it to write a dictionary was developed, although obviously this tool can also be used for other types of studies based on empirical data. The CTILC is currently available for consultation at no cost via Internet on the website of the Institut d'Estudis Catalans, and the consultation procedure allows you to obtain a series of contexts for each lemma looked up. As part of the corpus project, between 1996 and 1998 the Institut published a frequency dictionary of Catalan (Rafel 1996-1998). This frequency dictionary includes all the lexical data and corpus statistics from the project, and was made available both in a 3-volume printed edition and on 2 CD-ROMs.

#### *b) Main features of the BDLex*

The Lexicographic Database (BDLex) contains the thirteen most important Catalan dictionaries from the 19th and 20th centuries. This database was created to facilitate quick, systematic access to the information contained in these dictionaries, and was conceived as a complementary resource within the overall project. The various elements that constitute the structure of the dictionaries included in the BDLex have been identified and labelled systematically, so that they can be consulted individually or collectively.

#### *The Second Stage: Writing a descriptive dictionary*

The goal of the second stage of the ongoing DDC project is to write a descriptive dictionary of Catalan based primarily on the analysis of the data in the CTILC. We understand a 'descriptive dictionary' to be a dictionary that defines the lexical units in the language based on actual usage, regardless of prescriptive criteria.

One of the reasons a language academy decided to undertake a descriptive dictionary of this nature is the belief that prescriptive norms will be better justified if the language as used is objectively studied; in the case of the Institut d'Estudis Catalans, the justification is twofold, in that the Institut's by-laws specify that the Institut is not only responsible for establishing what constitutes standard usage but also for furthering the scientific study of the language (to quote the by-laws, "ocupar-se de l'estudi de la llengua"). With a descriptive dictionary of these characteristics,

the Institut d'Estudis Catalans has initiated a project guided by principles that are generally accepted in modern lexicography, and aims to create a valuable tool to comply with its mission as the academic institution charged with updating standard Catalan.

### *General characteristics of the DDLC*

In an overview such as this, we are not able to detail all the structural features of the *Diccionari descriptiu de la llengua catalana* (DDLC); here discussion will be limited to the most general characteristics (see Rafel 2007 for more details).

The DDLC does not aim at being a theoretical dictionary, although it shares some of the properties associated with theoretical dictionaries. The DDLC does not attempt to serve a particular pragmatic purpose nor does it have a specific pedagogical goal. Rather, the information in the dictionary is the result of a specific research programme throughout which we have attempted to maintain the highest scientific standards. We specifically want to ensure that the information presented in the dictionary is expressed as precisely as possible, which implies a certain degree of formalism in the language used and in the layout of the dictionary. In this sense, the DDLC comes close to the concept of a 'purely language dictionary' or a 'linguistic dictionary' as proposed by Dirk Geeraerts (1985), as opposed what is generally called a 'language dictionary'. As such, the ideal users of the dictionary are language professionals. It is important to note, however, that this work is not only aimed at experts: rather, the dictionary strives to be useful for experts yet accessible to any educated person interested in language issues. In other words, the dictionary aims to combine the highest standards in explaining the information presented in the work and at the same time aims to present it in a clear, easily understood fashion.

As mentioned above, the DDLC strives to incorporate the methodological and technological advances of modern lexicography. Because it is a multifunctional database, it has the characteristics of an electronic dictionary that is not only useful for research but also as the basis for several types of works addressed to the public. The database structure allows us to make the dictionary available over the Internet, in addition to the traditional book format.

One of the most important characteristics of the DDLC is that is based entirely on the CTILC, that is, it is a strictly corpus-based dictionary. This is neither the time nor place to argue for the advantages of a corpus-based dictionary; I should note, however, that writing a corpus-based dictionary implies that the corpus not only determines which headwords are included, but also the meanings of the lexical units as actually used, as opposed to using a priori methods (based on consulting other dictionaries) or lexicographers' intuition. Using a corpus allows us to use statistical data to set the headword list and other aspects of the dictionary. In addition, it allows us to include illustrative examples taken from citations with specific references, so that we can locate a particular sense or meaning in time. Since a corpus was used as our reference, the DDLC not only includes frequency data but also complete and systematic information on the syntactic structures used with the words included in the dictionary.

Given the type of dictionary we are working on, the definitions in the DDLC do not include encyclopaedic information (based on the description of reality) and concentrate instead on linguistic information (description of the meaning, the lexical restrictions, and the syntactic properties associated with each word). The defining text differentiates between the elements strictly reflecting the conceptual meaning (elements that are intrinsic to the definition) and those that refer to conditions or selective restrictions (extrinsic elements) that derive from the argument structure of the lexical unit defined (Rafel 2006). These extrinsic elements are related to the notion of *entourage* as described by Rey-Debove (1971) or that of *contorno* developed by Seco (1979).

Each entry in the DDLC contains quantitative information on use as shown by the corpus. This information is shown in a simplified fashion, and is represented graphically (and not numerically); each entry belongs to one of the five groups we established. For the first three frequency groups, the DDLC also contains percentages of the morphological categories in the corpus.

Another characteristic that is specific to this work is the writing process. The DDLC is not being produced in strict alphabetical order. Starting from the basic alphabetical order, we write—together with the entry for the word corresponding to alphabetical order—the entries for other lexical units that are related in form or in meaning, regardless of where those entries will be in the dictionary. This means that each lexicographer is responsible for writing a series of headwords that are related to one another, and the lexicographers work in parallel. This procedure attempts to avoid to the extent possible the lack of structural and descriptive coherence that characterises most existing dictionaries.

### *Current status of the DDLC*

At the time of this writing, we have reached the midway point in the dictionary and have completed some 88,000 entries. As a result of the special procedure used to write the dictionary, approximately a third of the finished entries correspond to the first letters of the alphabet (A and B), and the remaining two-thirds belong to the other letters.

The DDLC was made available on-line in January, 2005. The digital edition of the dictionary can be consulted at no charge on the website of the Institut d'Estudis Catalans, and is continually being updated. The number of entries in the on-line edition increases as each group of entries is completed. Furthermore, any changes made to previously completed sections are incorporated into the dictionary automatically.

With this digital edition, the public can consult the DDLC using a dynamic tool that relates elements both within and across entries. The electronic edition also includes several hyperlinks to elements that are external to the dictionary.

\* \* \*

As I come to the conclusion of this paper, it becomes obvious that space limitations prevent me from discussing all aspects of the development of Catalan lexicography. Nevertheless, I hope to have shown that our lexicographic tradition is characterized by a strong spirit to overcome adversity. Despite the significant obstacles in its path, Catalan lexicography has made significant accomplishments and today is well on its way to incorporating the conceptual and methodological advances required by modern lexicography.

## References

- Alcover, A. M. (1911). "Crònica de la Secció Filològica de l'Institut d'Estudis Catalans". *Butlletí del Diccionari de la Llengua Catalana* VI (nº 20, August-September 1911). 368-372.
- Cabré, M. T.; Lorente, Mercè (1991). *Els diccionaris catalans de 1940 a 1988*. Barcelona: Publicacions de la Universitat de Barcelona.
- Collinot, A.; Mazière, F. (1997). *Un prêt à parler: le dictionnaire*. Paris: Presses Universitaires de France.
- Colon, G. (1977). "La lexicografia catalana: realitzacions i esperances". In Colon, G. (ed.). *Actes del quart Col·loqui Internacional de Llengua i Literatura Catalanes, Basilea, 22-27 de març de 1976*. Barcelona: Publicacions de l'Abadia de Montserrat. 11-35.
- Colon, G. (2007). "Introducció al Diccionari general de la llengua catalana". In Mir, J.; Solà, J. (dirs.) (2007). *Pompeu Fabra. Obres completes. Vol. 5, Diccionari general de la llengua catalana*. Barcelona: Institut d'Estudis Catalans. 15-40.
- Colon, G.; Soberanas, A.-J. (1985). *Panorama de la lexicografia catalana. De les glosses medievals a Pompeu Fabra*. Barcelona: Enciclopèdia Catalana. [2nd ed., 1991]
- Geeraerts, D. (1985). "Les données stéréotypiques, prototypiques et encyclopédiques dans le dictionnaire". *Cahiers de Lexicologie* 46 (1). 27-43.
- Haensch, G. (1990). "Katalanische Lexicographie". In Hausmann, F. et al. (eds.). *Wörterbücher. Ein internationales Handbuch zur Lexicographie. Dictionaries. An International Encyclopedia of Lexicography. Dictionnaires. Encyclopédie internationale de lexicographie*. Berlin - New York: Walter de Gruyter. II, 1770-1788.
- Mir, J.; Solà, J. (dirs.) (2007). *Pompeu Fabra. Obres completes. Vol. 5, Diccionari general de la llengua catalana*. Barcelona: Institut d'Estudis Catalans.
- Rafel, J. (1996). "El diccionari de l'Institut i el Diccionari Fabra". In *Estudis de Lingüística i Filologia oferts a Antoni M. Badia i Margarit*. Barcelona: Publicacions de l'Abadia de Montserrat i Departament de Filologia Catalana. 3, 217-269.
- Rafel, J. (dir.) (1996-1998). *Diccionari de freqüències* [3 volumes and 2 CD-ROMs]. Barcelona: Institut d'Estudis Catalans.
- Rafel, J. (2004). "La lexicografia institucional: el cas del català". In Battaner, P.; DeCesaris, J. (eds.). *De Lexicografia. Actes del I Symposium Internacional de Lexicografia (Barcelona, 16-18 de maig de 2002)*. Barcelona: Institut Universitari de Lingüística Aplicada - Universitat Pompeu Fabra. 103-122.

- Rafel, J. (2006). “Els elements extrínsecs en les definicions lexicogràfiques: teoria i aplicació”. In Bernal, E.; DeCesaris, J. (eds.). *Palabra por palabra. Estudios ofrecidos a Paz Battaner*. Barcelona: Institut Universitari de Lingüística Aplicada - Universitat Pompeu Fabra. 201-218.
- Rafel, J. (2007). “Prescripción y descripción en la actividad académica: el *Diccionari descriptiu de la llengua catalana*”. In Campos, M. et al. (eds.). *Reflexiones sobre el diccionario [Actas del I Congreso Internacional de Lexicografía Hispánica, A Coruña, setembre de 2004]*. A Coruña: Universidade da Coruña. 9-33.
- Rey-Debove, J. (1971). *Étude linguistique et sémiotique des dictionnaires français contemporains*. Paris-The Hague: Mouton.
- Seco, M. (1979). “El contorno en la definición lexicográfica”. In *Homenaje a Samuel Gili Gaya. In memoriam*. Barcelona: Biblograf. 183-191. Also in Seco, M. (1987). *Estudios de lexicografía española*. Madrid: Paraninfo. 35-45; 2<sup>nd</sup> ed. Madrid: Gredos. 47-58.



# Border Conflicts: FrameNet Meets Construction Grammar

Charles J. Fillmore

University of California, Berkeley &  
International Computer Science Institute, Berkeley

## 1. The problem

I count myself among the linguists who believe in a continuity between grammar and lexicon (Fillmore et al. 1988, Joshi 1985), and I entertain the common image that each lexical item carries with it instructions on how it fits into a larger semantic-syntactic structure, or, alternatively, on how semantic-syntactic structures are to be built around it. My remarks here specifically concern an ongoing effort to describe and to annotate instances of, non-core syntactic structures, and to see how the products of this work can be integrated with the existing lexical resource, called FrameNet (FN), which is a set of procedures, and a growing database for recording the meanings and the semantic and syntactic combinatorial properties of lexical units. The FrameNet project, which I have directed since 1997, has recently begun exploring ways of creating a *constructicon*, a record of English grammatical *constructions*, annotating sentences by noting which parts of them are licensed by which specific constructions.

The grammatical constructions that belong in the larger constructicon—that is, in a construction-based grammar—include those that cover the basic and familiar patterns of predication, modification, complementation, and determination, but the new project is concentrating on constructions that ordinary parsers are not likely to notice, or that grammar checkers are likely to question. Some of them involve purely grammatical patterns with no reference to any lexical items that participate in them, some involve descriptions of enhanced demands that certain lexical units make on their surroundings, and some are mixtures of the two.

## 2. The work, the product, and the limitations of FrameNet

Since many features of the new resource are modeled on FrameNet, I think it useful to review FN's goals and activities, and the features of its database (Baker et al. 2003, Fillmore et al. 2003). FrameNet research amounts to

1. describing *lexical units* (LUs) in terms of the *semantic frames* they evoke, and describing those frames (i.e., the situation types, etc., knowledge of which is necessary for interpreting utterances in the language),
2. defining the *frame elements* (FEs) of each frame that are essential for a full understanding of the associated situation type (the frame elements are the props, participants, situation features that need to be identified or taken for granted in sentences for which the frame is relevant),

3. extracting from a very large *corpus* example sentences which contain each LU targeted for analysis (FN has worked mainly with the British National Corpus),
4. selecting from the extracted sentences representative samples that cover the range of combinatorial possibilities, and preparing *annotations* of them as layered segmentation of the sentences, where the segments are labeled according to the FEs they express, as well as the basic syntactic properties of the phrases bearing the FE,
5. displaying the results in *lexical entries* which summarize the discovered *combinatorial affordances*, both semantic and syntactic, as *valence patterns*, and creating links from these patterns to the annotated sentences that evidence them, and
6. defining a network of *frame-to-frame relations* and the graphical means of displaying these, that will show how some frames depend on or are elaborations of other frames.

### 2.1. *The frames*

The frames developed in FrameNet are the conceptual structures against which the LUs in the FN lexicon are understood and defined (Fillmore 1982, Fillmore & Atkins 1992, 1994). These can be as general as the location of some entity in an enclosure, or as specific as interest on investment.

One FN frame that is simple enough to describe completely, and just complex enough to be interesting, is the so-called Revenge frame, the nature of which requires understanding a kind of history. In that history, one person (we call him the OFFENDER) did something to harm another person (what he did we call the OFFENSE and his victim we call the INJURED\_PARTY); reacting to that act, someone (the AVENGER, possibly the same individual as the INJURED\_PARTY) acts so as to do harm to the OFFENDER, and what he does we call the PUNISHMENT. Thus, we have the frame Revenge, and the frame elements AVENGER, OFFENDER, OFFENSE, INJURED\_PARTY, and PUNISHMENT. Other features of the Revenge frame include the fact that this kind of pay-back is independent of any judicial system. There is a very large set of verbs, adjectives and nouns that evoke this frame, by which we mean that when users of the language understand these words, their understanding includes all of the elements of that scenario. Among the verbs that evoke this frame are *avenge* and *revenge*, the nouns include *vengeance* and *retribution*, there are phrasal verbs like *pay back* and *get even*, adjectives like *vengeful* and *vindictive*, support constructions like *take revenge on*, *wreak vengeance on*, and *exact retribution against*, plus prepositional adverbials like *in retribution*, or *in revenge*.

FrameNet has developed descriptions of over 800 frames to date, and nobody is ready to estimate how many there are altogether. The list from the time of the last official release can be found at <http://framenet.icsi.berkeley.edu>.

## 2.2. *The frame elements*

The frame elements (FEs) are somewhat analogous to the *deep cases* of early Fillmore (Fillmore 1968, 1971), *thematic roles* in various generativist writings (Jackendoff 1990), *actants* and *circonstants* in the Tesnière tradition (Tesnière 1959). There are good reasons for not tying the frame elements into any of the familiar lists of semantic roles (agent, patient, theme, experiencer, instrument, etc.). Since annotators are asked to find expressors of frame elements in actual sentences, FE names that are memorable in respect to the frame itself will facilitate such identifications. Thus to take the case of the arguments of *replace* in a sentence like

[I] replaced [my stolen bicycle] [with a much cheaper one],

it makes more sense to refer to the phrases introducing the two bicycles as the OLD and the NEW than to try to figure out how well these roles can be accommodated in the “standard” lists. (The missing bicycle, in fact, is not a participant in the event described by the sentence but is a necessary element of its meaning.) The recognition of FE commonalities across frames is made possibly by the system of frame-to-frame relations.

We wanted to think of the frame elements as representing the kinds of information that could be expressed in the sentences and phrases in which the frame is “active”, and we wanted to be able to discover which parts of a sentence reveal information about which frame element. There is an important constraint on this task, distinguishing it from annotation practices that seek to learn everything about each event in a continuous text. Since the information we record is supposed to be relevant to the syntactic description of a given lexical unit, we require that the frame elements we attend to are *in grammatical construction with* the lexical unit being described. Annotators will ignore event-relevant information elsewhere in the text.

We make a distinction between *core* and *peripheral* FEs. The core FEs are those that are conceptually necessary in any realization of the frame by the nature of that frame; the peripheral frame elements are the adjuncts that fit the familiar description “time, place, and manner, etc.”, especially the “etc.” (the core/periphery distinction can vary across frames; for verbs like *reside*, *elapse*, and *behave*, the locative, temporal and manner components, respectively, are not peripheral). A characteristic of the peripheral FEs is that they have essentially the same meaning and the same syntactic marking wherever they appear; whatever distributional limitations they have are explained by the fact that frames about *happenings* can take time and place modification, frames about *intentional acts* can take instrument and purpose modification, and so on. A third kind of frame element is what we refer to as *extrathematic*: these are expressions (like benefactives or phrases like *in revenge* or *in return*) that have the effect of situating the event signaled by the target’s frame in some larger or coterminous situation.

The goal of FrameNet lexical descriptions is, for each frame-bearing word, to match the word’s semantic combinatorial requirements with the manner of their syntactic realization. Reversing the point of view, we seek to recognize in the syntactic nature

of the phrases around a given frame-bearing lexical unit, information about the participants in situation that is an instance of the frame. The resulting pairing of semantic and syntactic roles constitutes the *valence* description of the item.

### 2.3. *Example sentences*

The goal in providing examples was to have, for each lexical unit, a full set of illustrations of its basic combinatorial properties, and we preferred sentences whose content was clearly relevant to the meaning of the word being exhibited. If we were looking for an illustration of *knife*, we would prefer *the butcher sharpened his knife* than *the poet photographed a knife*. These example-selecting decisions were made in resistance to several kinds of pressure. Some members of the research community wanted to see sentences of the most frequent type; but for many verbs, the most frequent examples had mainly pronouns (*I risked it*). Some wanted us to include complex and distorted sentences as well as the simplest type; some wanted us to make sure we include creative uses of a word wherever we found them, scolding us for neglecting metaphor and other figurative uses: our view echoes that of Patrick Hanks (MS), namely, that we had the obligation to produce clear descriptions of the *norm*, leaving it to some auxiliary research to explore the ways in which speakers *exploit* the norm for creative expression. Where a metaphorical use was *lexicalized*, the LU resulting from that lexicalization was included in its appropriate frame.

### 2.4. *The annotation*

The original mission of FN was purely lexicographic: to annotate a variety of typical uses of each target LU and to seek to cover a wide range of relevant contexts for the LU (i.e., all of its *valence* possibilities and representative samples of its semantic collocates), and this meant creating a collection of sentences in which each was annotated with respect to one word in it. Thus a sentence like

She smiled when we told her that her daughter had been nominated to receive an important award.

might be annotated for the verb *smile* alone, as a member of the *Make\_faces* frame, where it belongs in the set *frown, grimace, grin, pout, scowl, smile, smirk*.

As the size of the lexicon increased, it became clear that there were sentences for which FN was prepared to describe many of the words in it, and ultimately we received a subcontract to look into the possibility of producing full text annotations. That meant annotating each word in the sentence—that is, each frame-evoking word—. For the above example, that would mean showing the frame structure of the words *smile, tell, daughter, nominate, receive, important* and *award*. For our purely lexicographic purposes, we would have no reason to annotate the word *told* in this sentence—we already have more than enough examples of the lemma—but it would have to be done here again in order to prepare the semantic structure of the sentence as a whole. Obviously this need increased our eagerness to find ways of automating parts of the annotation process.

FrameNet has to date annotated a growing number of texts, some of them viewable on the FN website. Most of them are only partially annotated, partly because they contain lexical material FN has not yet worked through, and partly because they contain meaningful grammatical patterns that FN annotation has not been prepared to capture.<sup>1</sup>

The annotations themselves are presented in layered stand-off representation in multiple layers. For lexicographic annotations, one layer identified the target LU and its frame; another represented the FEs in the phrases that serve as its valents; one indicated the phrase types of the constituents so identified; one indicated the grammatical function of each valent; and a few other layers were dedicated to special features associated with individual parts of speech. The FEs were annotated manually, the GF and the PT labels were attached automatically and checked manually. Annotations viewable on the FrameNet website show only the frame element labeling, as in Figure 1.

[<sub>Fluid</sub> The River Liffey] FLOWS <sup>Target</sup> [<sub>Source</sub> from west] [<sub>Goal</sub> to east] [<sub>Area</sub> through the center of the city] [<sub>Goal</sub> to Dublin Bay].

Figure 1: FE annotation of a sentence

Full text annotations consist of sets of layers, each corresponding to one target LU. It is virtually impossible to get a view of the full annotation of a long sentence, but there is some experimental work being done to derive dependency trees from these, with the nodes indicating lexical heads and their frames, the branches labeled according to the frame element represented by the dependent nodes.

One special feature of FN annotation is the recording of FEs that are conceptually present but syntactically missing. These are sorted into *constructional null*, such as the missing subject of an imperative sentence; *indefinite null*, such as the object of intransitivized *eat*, *sew*, *bake*, etc.; and *definite nulls (zero anaphora)*, entailing that the missing element has to be recoverable in the context, such as the missing object of *we won* (what is understood but unexpressed is the contest—not the prize), the missing preposition phrase in *she arrived* (where the destination has to be known) or *mine is similar* (where the unexpressed comparand has to be part of the conversation), and so on. The last of these plays an important role in construction annotation as well. Such information is associated with the annotation of the LU that licenses the omission.

---

<sup>1</sup> The texts—chosen because other researchers are examining them as well—were taken from the Wall Street Journal section of the Penn TreeBank, the Nuclear Text Initiative website, and a selection of Berlitz Travel Guides that have been made available to the American National Corpus.

## 2.6. The entries

Each LU is identified by lemma, part of speech, and frame name. The LUs were chosen because of their membership in one of the frames being covered by FrameNet, and what that means is that in many cases the most common use of a lemma is not to be found: FN researchers have not reached that frame yet. Almost all features of the lexical entry are produced automatically: handmade features include a simple definition.<sup>2</sup> For valence-bearing words, the entry contains a table showing the ways in which each frame element can match a phrase type, and a separate table showing the variety of ways in which combinations of FEs and PTs make up the valence exhibited by individual sentences. Viewers of the valence descriptions can toggle between core FEs only, or all FEs found in the sentences—core, peripheral, and extrathematic.

The entries for nouns that designate events or states of affairs also include information about the existence of *support verbs* and *support prepositions*; access to the sentences will reveal which FEs are represented among the arguments of the LU's verbal or prepositional support.<sup>3</sup>

## 2.7. Frame-to-frame relations

Since frames can differ from each other in granularity, and some frames are clearly related to other frames, it has proved necessary to create an ontology of frames, linked to each other by several kinds of relations. Figure 2 is a display of the frame relations centered on `Commercial_transaction`:

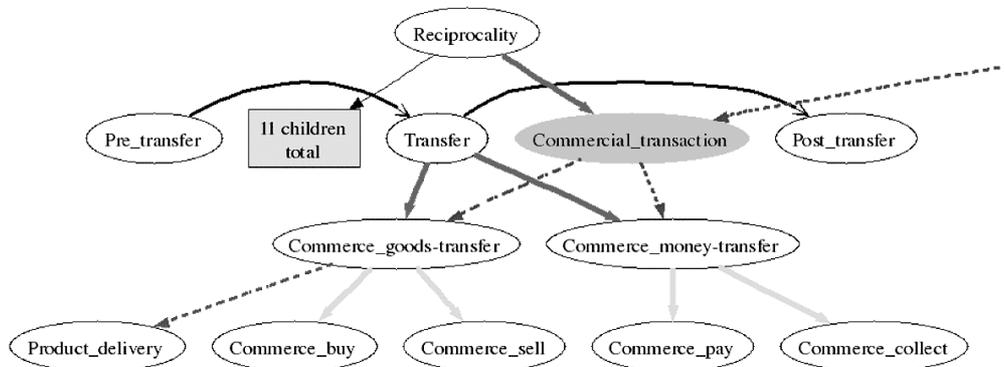


Figure 2: Frame-to-frame relations centered on `Commercial_Transaction`

Several different kinds of relations can be seen in this diagram. `Commercial_transaction` has two components (related to the mother node by a *Part\_of* relation

<sup>2</sup> The purpose of the definition is purely mnemonic, to aid the user in knowing which sense of a word is being analyzed in a given entry. Where appropriate the definitions were taken from the *Concise Oxford Dictionary 10*, with permission from Oxford University Press. Others were in-house.

<sup>3</sup> The current database shows no way of classifying support constructions along the line of the lexical functions of the MTT model of Igor Mel'čuk and his colleagues, though various researchers are seeking to derive such information automatically from the FN annotations. (Rambow et al., MS, Bouveret & Fillmore, MS)

as indicated by the broken line), and these are *Commerce\_goods\_transfer* and *Commerce\_money\_transfer*. Each of these is a type of (=has an *Inherits* relation to) the frame *Transfer*. The low frames *Commerce\_buy* and *Commerce\_sell* have separate *Perspective\_on* relations to *Commerce\_goods\_transfer*, and the frames *Commerce\_pay* and *Commerce\_collect* have *Perspective\_on* relations to *Commerce\_money\_transfer*. Thus, a commercial transaction is an instance of *Reciprocity*, involving two co-occurring reciprocal transfers, one of goods and one of money. Buying and Selling are perspective-varying instances of goods-transfer, differing from the point of view of the buyer and the seller; and similarly with paying and collecting (=charging) and their relation to money-transfer.

### 3. FrameNet treatment of multiwords so far

The construction-building work concerns itself with linguistic knowledge that goes beyond simple grammar and simple words, and hence it will include various kinds of idioms and other multiwords. There are many kinds of multiwords that already fall within the scope of FrameNet work.<sup>4</sup> Among the multiwords covered by current FrameNet<sup>5</sup> we find

1. phrasal verbs, with particles, which are simply treated as two-part verbs that take a specific particle as a syntactic valent; the particle is more or less motivated, but can't be understood as simply contributing its own meaning
  - a. Intransitive: *pick up* (increase), *take off* (start flying)
  - b. Transitive: *take up* (consider), *take off* (remove)
2. words with selected prepositional complements, listed with preposition, syntactically selects P-headed phrase
  - a. Verbs: *depend on*, *object to*, *cope with*
  - b. Adjectives: *fond of*, *proud of*, *interested in*
  - c. Nouns: *fondness for*, *pride in*, *interest in*
3. support constructions—syntactically separate, treated as evoking a frame linked to the noun rather than the verb
  - a. Verbal heads: *take comfort in*, *take pride in*, *put emphasis on*
  - b. Prepositional heads: *at risk*, *in danger*, *under arrest*
4. combinations—combining selected prepositional complement with particle or noun
  - a. *put up with* (tolerate), *break in on* (interrupt)
  - b. *take comfort in*, *place emphasis on*
  - c. *take into possession*, *take under consideration*

---

<sup>4</sup> Josef Ruppenhofer delivered a paper on this topic at an earlier Euralex meeting (Ruppenhofer et al. 2002).

<sup>5</sup> FN treatment of compound words has more or less awaited the capability of constructional annotation. In the current databases, there are compounds that are simply treated as single unanalyzed units, and there are others in which the head is a frame-bearing word and the modifier is labeled as an FE in the head's frame. FN has lacked the means of describing a compound word both as a unit on its own and as having an internal structure.

5. transparent nouns—the first noun in [N of N] structures signifying types, aggregates, portions, units, measures, epithets, etc.; the motivation for recording these is to be able to recognize selectional or collocational relations between the context and the *second* noun
  - a. *my gem of a wife, in a part of the room, on this part of the shelf, wreak this kind of havoc.*

#### 4. Full-text annotation and the confrontation with constructions

In carrying out full-text annotation the goal was to end up with structures which could be the basis of the semantic integration of the whole sentence. Working with one of those linguist-invented sentences like

The Secretary ordered the Committee to consider selling its holdings to the members

we should be able to identify straightforwardly the participants in the *ordering* event: *the Secretary* gave the order, *the Committee* received the order, and *to consider selling its holdings to the members*, specifies the order. For the verb *consider*, the entity that was to do the considering was *the Committee*, and *selling its holdings to the members* was to be the content of such considerations; and the three participants in the *selling* event are to be *the Committee* as seller, *the members* as buyer, and *the holdings* as the asset destined to change ownership. The words *Secretary*, *Committee* and *members* are all relational nouns used without any indication of what the other term of the relation is, and that's possible if that other entity is understood in the context. A simple frame-annotated dependency tree will fairly well capture the meaning of the whole, with word-frame pairs making up the node labels, the branches labeled according to the semantic role, and with the missing entities in the relational nouns marked with the possibility of indexing them to contextually given entities.

One doesn't have to look far to find sentences containing structures that do not lend themselves to such simple treatment. Here are the first three sentences of a leader from the Economist newspaper of June 17, 2007, with comments on those features that go beyond simple lexicon and simple grammar.

For all the disappointments, posterity will look more kindly on Tony Blair than Britons do today. Few Britons, it seems, will shed a tear when Tony Blair leaves the stage on June 27th after a decade as prime minister, as he finally announced this week he would do. Opinion polls have long suggested that he is unpopular.

1. *for all the disappointments*:  
*for all X* is a concessive structure with a meaning like “in spite of X”; seems to be restricted to definite objects; not best treated as a complex preposition
2. *look kindly on*:  
a phrasal verb with the meaning “judge positively”

3. *[posterity] will look more kindly on Tony Blair than [Britons] do [today]:*  
a comparative structure with a double-focus comparand—[Britons] [today], each accented, requiring the semantic unpacking of *posterity* as something like [the world] [in the future] (a contestable interpretation)
4. *few Britons:*  
not a vague indication of cardinality like *a few Britons*, semantically a negator (= “not many”), creating a *negative polarity* context (see item 6)
5. *it seems:*  
an epistemic parenthesis, bearing no structural relation to the rest of the sentence but limited in the positions that would welcome it
6. *shed a tear:*  
a VP collocation of the *minimizer* type, appropriate to the negative polarity context created by *few*; similar in this respect to *drink a drop, lift a finger, give a damn, eat a bite*
7. *leave the stage:*  
metaphor, referring here to leaving the PM-ship
8. *on June 17th:*  
use of the preposition *on* with day-level temporal units (cf. *in March, at noon, in the morning*)
9. *June 27th:*  
one of various ways of pairing a date with a month name
10. *as prime minister:*  
*as* selecting “role” name; requires context implying service in a role
11. *as he announced he would do:*  
relativizer *as* (consider replacing *as* with *which*)
12. *would do:*  
the form of VP ellipsis (including *do* after a modal) found in BrE missing or rare in AmE (*as he announced he would*)
13. *this week:*  
an expression in which the first element is taken from the list *this/next/last* and the second is a calendric unit name like *week, month, year*, but not *day*
14. *have long suggested:*  
the use of *long* in the meaning “for a long time” has numerous contextual constraints, difficult to pin down; here both (a) the position between *have* and the participle and (b) restriction to certain classes of verb meanings seem necessary (compare *I have long known that ...* with *\*I long knew that...* and *\*I have long lived in California.*)

## 5. Constructions and the new constructicon

Section 3 offered a number of ways in which the behavior of multiword expressions can be incorporated into the FN lexicon and into FN-style annotations, that is, where the information recorded is mainly limited to a small number of requirements

that lexical items impose on their immediate grammatical environment. Stepping outside of that is a definite new challenge.

### 5.1. *The annotation challenge*

How did FrameNet become concerned with such matters? First, with our efforts in full text annotation, we became interested in the possibilities of making better coverage of all of the linguistic properties of texts, not just those involving simple predicates and their valence structures. Second, it seems clear that while with support constructions we moved slightly beyond “standard” valence projections, the view of syntactic structure within which we explained the syntactic concomitants of lexical selection needs to be expanded. Third, the community in Berkeley that got started with FrameNet is also a community that has an interest in the broader theory of grammatical constructions. Fourth, and most importantly, it seemed likely that *the same data structure and annotation software devised for lexical annotation could be assigned to the treatment of constructions*.

In 2007 FrameNet received a small grant for doing exploratory research on designing a constructicon, an inventory of “minor” grammatical constructions, and to demonstrate a means of annotating instances of them. The parallels to ordinary FN lexical annotation were striking, as can be seen in Table 1.

Lexical FrameNet	Constructicon
Frame descriptions describe the frames and their components, set up FE names for annotation, and specify frame-to-frame relations; lexical entries are linked to frames, valence descriptions show combinatory possibilities, entries link valence patterns to sets of annotated sentences.	Constructicon entries describe the constructions and their components, set up construction elements (CEs, the syntactic elements that make up a construct), explain the semantic contribution of the construction, specify construction-to-construction relations, and link construction descriptions with annotated sentences that exhibit their type.
The FEs are given names according to their role in the frame, and provide labels for the phrases in the annotations that give information about the FE.	The CEs are named according to their function in the constructs, they provide the labels on words and phrases in annotated sentences.
The syntactic properties—grammatical functions and phrase types—are identified for all constituents that realize frame elements.	Phrase types are identified for constituents that serve as CEs in a construct; for constructions that are headed by lexical units, grammatical function labels will also be relevant.
Example sentences are selected that illustrate the use of the lexical units described.	Example sentences are selected and annotated for the ways they illustrate the use of the construction.
Annotations identify the LU, the FEs, and the GFs and PTs of the segments marked off.	Annotations contain labels for the CEs and identify, for lexically marked constructions, the relevant lexical material.

Lexical FrameNet	Constructicon
Valence patterns are identified, and linked to the annotations.	Varieties of construct patterns are identified and linked to the annotations.
Frame-to-frame relationships are documented and displayed in a separate resource.	Construction-to-construction relationships are identified and (will eventually be) displayed

Table 1: Lexical and Constructional Description and Annotation Compared

The questions to ask for setting up an annotation system for constructions include: What is the constituent (the *construct*) within which a construction operates? What needs to be tagged within a construct? What are the functions of the elements of the construction? What if anything reveals to the reader/listener that there's anything special about the sentence?

In FN lexicographic annotation, we describe a frame and its components or participants, we annotate sentences by identifying the target lexical item and bracketing off the valents and labeling them with frame element names. In constructional annotation, then, we should be able to describe a construction and name the parts of sentences that are the constituents of the constructs licensed by the construction, and then to bracket off those components and assign them labels assigned to the elements of the construction. One important difference is that often there is no *target* LU to link the construction to.

Figures 3 and 4 show the similarity of lexical and constructional annotations, as they appear in the annotation tool. The lexical example represents the clause *one of them accused Mr Wisson of kidnapping*; the constructional example represents the sentence *None of these arguments is notably strong, let alone conclusive*. The list of labels at the bottom of each is the list appropriate to a single level: the FE level in the lexical example, the CE level in the construction example.



Figure 3: Lexical annotation of the verb accuse in the Judgment\_Communication frame

Layer	None of these arguments is notably strong, let alone conclusive.
CE	Context_of_focus
CEE	First_conjunct
CstrPT	A J P
GF	
PT	A J P
GovX	
Sent	
Other	

Let.alone - CE   Let.alone - CEE   Let.alone - CstrPT   Let.alone - GF   Let.alone - PT   Let.alone - GovX   Let.alone - Sent   Let.alone - Other

CEE <FS>   Context\_of\_focus ...   First\_conjunct <F...>   Second\_conjunct ...   Trigger <FS>

Figure 4: Constructional annotation of a phrase built around the conjunction let alone

## 5.2. *The varieties of constructions needing annotation*

The assumption that it would be easy to adapt the FrameNet annotation tool to construction annotation turned out to be false. Essentially the first half of the year of this grant passed by before a proper annotation tool was ready. Finally, in the spring semester, there are two graduate students working on the project, Russell Lee-Goldman and Russell Rhodes, with strong backup by Michael Ellsworth and Project Manager Collin Baker. By the time of the Euralex meeting, I expect to be able to give a coherent report on our accomplishments and their significance. In the meantime, however, I offer some hastily gathered notes on the types of constructions we need to cover. In the final report almost all of the construction descriptions will include references to the relevant literature, omitted here with apologies, including names like Boas, Borsley, Croft, Goldberg, Jackendoff, Kay, Lakoff, Lambrecht, McCawley, Michaelis, O'Connor, Pullum, Pustejovsky, Sag, Wierzbicka, Zwicky.

### 5.2.1. *Lexical constructions*

For an important class of cases, the grammar allows words with one meaning to be paired with the *combinatory affordances* that are common to a semantically defined class of words (in the case of verbs, this amounts to valence patterns; for nouns, the difference between proper and common nouns, or that between count and non-count nouns; for adjectives the difference between scalar and non-scalar adjectives). The word *coercion* is sometimes used to cover such relationship.

We can distinguish the words that are “at home” with these affordances from the words that are their “guests”. There is an obvious problem for a corpus-based lexicon-building effort like FrameNet, since there is no automatic way of telling the difference: should the derived behavior of “frequent guests” be listed in the lexicon or merely recognized in context as an instance of the construction? It’s a problem for lexicography in general, since the decisions that need to be made one way or another are not always clearcut.

EXAMPLES include the phenomena in much of the literature on Argument Structure Constructions, especially in the work of Adele Goldberg. The meanings created by these constructions involve specified relations between the meaning of the “guest” and the semantic expectations of the “host” pattern: *slipping someone a banknote* is using a slipping action to give someone a banknote, *wriggling into the swimsuit* is “entering” the swimsuit (putting it on) with a wriggling motion; an event of *sneezing the napkin off the table* is one in which the air current created by a sneeze has motive force. With nouns, examples like *we had beaver for dinner* show the use of the name of an animal with the grammar of a mass noun, coercing a construal as the flesh of the animal prepared for human consumption.<sup>6</sup>

### 5.2.2. Verbs with contextual requirements outside of their phrasal projection

For the kinds of examples we have in mind under this category it should be possible simply to specify the greater context as part of the combinatory affordances—but there is no familiar formal way to do this within theories of valence. The most common cases are words that fit negative polarity contexts, contexts including negation straight on or other sources of general *irrealis* contexts, like questions, conditional clauses, and dozens of others (since we are mainly interested in identifying cases and annotating them, the kinds of careful formulation that a true grammar would need can be glossed over). Verbs that require contexts that involve both ability and negation allow various ways of expressing those contexts.

EXAMPLES include *can't stand*, *can't afford*, *can't tell*, *can't seem to...*, *can't help*. The contexts can be expressed in different ways: in *were you ever able to afford such luxuries?* the polarity is not triggered by a negative morpheme, and the ability is expressed by an adjective rather than a modal. In *it's too dark to tell what they're doing*, the semantics of “not + able” is entailed in the meaning of *too*. In the case of the verb *brook* a first impression might be that its required negation is “local”—i.e., in the determiner of the direct object—but the negation can be presented by an external negation with *any* replacing the *no* in the determiner position: *I will brook no interruption, I am too busy to brook any distraction*.

### 5.2.3. Templatic constructions

Some constructions seem to require a pattern of fixed positions with strict requirements on what can fill those positions: such is the case of the linguistic way of expressing proportions of the kind  $A:B=C:D$ ; it is sufficient to think of the sentences as providing ways of pronouncing the symbols in such a representation.

EXAMPLES are often found in lower-grades test questions: *Six is to three as four is to two; blood is to red as snow is to white*.<sup>7</sup>

---

<sup>6</sup> The construction does not merely convert the animal name into the name of a continuous substance. A sentence like the neighborhood fox likes beaver is not licensed by this construction.

<sup>7</sup> These sentences could be given a somewhat tortured parse, involving the extraposition of the *as*-phrase: if we think of *as four is to two* as identical to *what four is to two*, and as naming a particular relation, then we can see the pattern by putting things “back”: *Six is [what four is to two] to three*.

#### 5.2.4. A mere five dollars

There is a phrasing of numerical expressions that requires (a) the singular indefinite determiner, (b) an adjective that qualifies a number, and (c) a number, such that the combination demands a noun head that matches the number and can contradict the singularity of the article *a*. That is, for something like *a mere five dollars*, all three elements are required: *\*a five dollars* doesn't work, *\*mere five dollars* doesn't work, *\*a mere dollars* doesn't work. We see the construction as determining the prenominal phrase only: in the manner of an ordinary cardinal number, the noun can be deleted if its nature is understood in the context—as people or dollars, for example, in *a mere two million*.

EXAMPLES show adjectives with minimizing, neutral and maximizing senses: *a paltry twenty cents*, *an additional thirty pages*, *a whopping seven billion dollars*. An expression like *another \$200* is a disguised instance of this construction, where *an+other* is analogous to *an+additional*, and *\$200* is shown as *two-hundred +dollars*. The modifying adjectives that appear in constructs that instance this construction make up an interesting class.

#### 5.2.5. Presentative constructions

George Lakoff has discussed a family of constructions using *here* and *there* which have important communicative functions. Formally, they begin with *here* or *there*, they have a verb which most typically is *be*, *come*, *go*, *sit*, *stand*, or *lie*, with the restriction that if the subject is a pronoun it precedes the verb but if it is a lexical NP it follows the verb, and utterances of them have the function of announcing something about the appearance or presence of something. In the complete version, they include some kind of secondary predicate, that can be an adjective, a preposition phrase, a participial phrase, or a with(out) clause.

EXAMPLES include *here comes that old fool*; *there she stood, with her hands on her hips*; *here comes Billy, crawling on his hands and knees*; *here I am, ready to serve*.

#### 5.2.6. Wherewithal

There is a construction which uses the determiner *the* and a noun construed as naming a resource; it is followed by an indication of what the resource could be used for, expressed as an infinitival VP or a *for*-PP; and its governing context identifies someone as a POSSESSOR (or not) of a sufficient supply of the resources to carry out the purpose represented by the noun's complement. A parallel construction exists with the word *enough* in place of *the*. The name it's been given is due to the fact that the noun *wherewithal* occurs *only* in this construction!

EXAMPLES with physical resources include *I don't have the resources to landscape the garden*, *we lack the staff for such a project*, *who will provide me the wherewithal to accomplish this*, *they denied me the funds to complete the job*, *do we have the fuel to make it to the next town?* Nouns that designate spiritual resources that fit the same construction include *courage*, *spirit*, *will*, *guts*, *balls*, and several others. Arguments that this construction is needed include the observation that the combination of

the nominal and the complement cannot serve as a self-standing NP: *\*we spilled the fuel to make it to the next town*. The purpose complement can be omitted in contexts where it is understood: A sentence like *where did you find the cash?* can be an instance of this construction, addressed to someone who had just bought an expensive car, or it can be used simply to refer to some until-now misplaced amount of money. The existence of the *Wherewithal* construction explains that ambiguity.

### 5.2.7. Gapping and Right Node Raising

Some constructions are purely organizational, and have no lexical components beyond conjunctions or words that can function as conjunctions. Those referred to as Gapping and Right Node Raising (RNR) omit phrases whose meaning is shared against elements that are in focal contrast.

EXAMPLES of RNR include *John loves, but Mary hates, rock music*, where comma intonation separates the two truncated conjuncts from their common completion; gapping is seen when the shared element is between the focal elements: *John loves peaches and Mary apples*. Those are obviously made-up sentences, chosen for their brevity. An attested sentence that exemplifies both of these constructions simultaneously is *Bears have become largely, and pandas entirely, noncarnivorous*.

### 5.2.8. Let alone

*Let alone* is a conjunction whose combinatory potential and semantic-pragmatic interpretation are discussed in Fillmore-Kay-O'Connor 1988 and some discussions following that. Briefly, the pieces that are in focal contrast can be<sup>8</sup> assembled with their surrounding contexts to form two propositions, one of these propositions is responsive to the context (i.e., to some assumed or expressed *context proposition*), the other is strongly asserted by the speaker, and it contextually entails the first.

EXAMPLES include the sentence in Figure x, *None of the arguments is notably strong, let alone conclusive*. Numerous examples of multiple foci are found in the FKO article. *Let alone* sentences frequently exemplify RNR: *I wouldn't touch, let alone eat, anything that ugly* (Made-up sentence).

### 5.2.9. Verb one's way

A much-studied construction is a way of providing motion verbs by inserting a verb that indicates an action by which someone is able to move, or a path through which

---

<sup>8</sup> For example:

Context proposition spoken by interlocutor: *Can you give me a dollar?*

Direct response to the context proposition: *I won't give you a dollar.*

Response that strongly entails the context-relevant response: *I wouldn't lend my mother a nickel.*

Result: *I wouldn't lend my mother a nickel, let alone give you a dollar.*

Relevant scales for the triple contrasting foci: I'm more likely to lend money to someone than to give it away; I'd be more generous to my mother than to you; a dollar is a lot more than a nickel.

the mover moves, or an activity on the mover's part during which they moved. The structure is (a) verb plus (b) possessive pronoun coreferential to the moving entity plus (c) the word way: VERB one's WAY. The most neutral verb that is "at home" in this construction is make (*Let's start making our way home.*) The verb *wend* exists *only* in this construction.

EXAMPLES that show the variety include *She pushed her way through the crowd, the river winds its way through the prairie, we dined our way through the south of France.*

#### 5.2.10. *In one's own right*

A number of constructions depend on the extended reflexive possessive pronoun *one's own*: *he finally has a room of his own, you're on your own now*, but one we have examined is the adjunct *in one's own right*. A typical background assumption for its use is something like this: A is affiliated with B in some way (a relative, an assistant), B is already known for some property or accomplishment, the sentence asserts that same property or accomplishment of A, and the construction conveys the assumption that A's accomplishments are not due to the affiliation with B. The son of a poet can be *a fine poet in his own right*, the husband of a famous chemist can be *an accomplished chemist in his own right*. It would sound odd to say of the wife of right-wing radio commentator Rush Limbaugh that *she is a major intellectual in her own right*, without invoking a belief that Mr. Limbaugh is a major intellectual. (I don't even know if he's married—this is just an example).

#### 5.2.11. *Rate phrases*

The concept of rate is expressed in English with two adjacent NPs in which the first identifies a quantity of units of some type and the second introduces a unit of a different type across which the measurement applies, more or less as numerator to denominator. Typically the second NP is marked with *a* or *per*, but other types occur as well. These expressions express such notions as growth rate, frequency, fuel efficiency, speed, and the like.

EXAMPLES include *it grows four inches a day*, but also *four inches every three days; my Hummer gets seven miles a gallon; our committee meets twice a week; we were moving at 150 km per hour*. The type of rate can be calculated by comparing the two kinds of units, and can be supported by making note of aspects of the governing context, such as the items *grow, meet, gets, and at* of the examples.

#### 5.2.12. *Measurement phrases*

Some scalar adjectives, but not all, support measurement qualifiers that indicate a quantity of units used for values on the scale.

EXAMPLES include *five meters long/wide/tall/thick*, and *seventeen years old*. Weight and cost values are expressed verbally, with the verbs *weigh* and *cost*; there is no *\*twenty pounds heavy* or *\*twenty dollars expensive*. Comparative expressions, however, can have measured "gaps" across the board: *twenty pound heavier, twenty dollars cheaper, three years older*, etc.

### 5.2.13. *Deictically anchored calendar units*

The lexical set *this-next-last* occurs in several constructions dedicated to locating a reference time to the present moment—the *temporal deictic center*—with respect to *calendric* time periods like *week*, *month*, and *year*. *This* makes reference to the period containing “now”; *next* refers to the period following the period containing “now”; and *last* refers to the period preceding the period containing “now”. These patterns do not apply to days, however: at the day level the same functions are served by the lexical items *today*, *yesterday*, *tomorrow*.

EXAMPLES illustrating one of the constructions, simply identifying a period, are *next year*, *last month*, *this week*; a second construction uses these words to mark a recurring point or subdivision of a larger unit and locates the event within the lower unit with respect to whether the larger period is current, past, or future to “now”: *next Wednesday*, *last summer*, *this August*; the third construction uses *next* and *last* in a fixed pattern where the word is understood as picking up the immediately preceding mention of the time entity: *the week after next*, *the month before last*, and *the summer after next*, *the Christmas before last*.

### 5.2.14. The + *Adjective*

Expressions like *the rich* and *the poor* are usually thought of as showing these adjectives being “used as a noun”. Instead of attributing a part-of-speech change to the adjective, it would seem that a better analysis is that the combination THE + Adjective-Phrase behaves like a full NP. How else could we understand *the very rich*, *the very young*? Not as very modifying a noun, presumably. The constraints seem to be that the adjectives designate some categorizing property of humans; the resulting phrase is human, generic, and plural. Certain adjectives—*poor*, *rich*, *young*, *old*—are “frequent guests” of this construction, but the lexicographers’ decision to identify them as actual nouns in those contexts does not seem helpful.

### 5.2.15. *Adjective* + and + *Adjective*

These same adjectives can be used, in roughly the same meaning, when they surround and, as in *he was beloved of rich and poor alike*. In this case the definite article is not needed, but the conjunction is necessary: *\*he was beloved of poor* does not work.

### 5.2.16. Degree modifiers of adjectives

It’s difficult to decide how many constructions are needed for the intended family of constructions, perhaps several, with constructional inheritance connecting them. Some examples communicating sufficiency or excess have extraposable complements: *too* and *enough* go with an accompanying infinitival VP, *so* goes with a *that*-clause. Others question a scalar value posed in the context, require negative polarity, are accented, and do not have an extraposed complement.

EXAMPLES include *she’s not that young*, *you can’t be too hungry* or *you’d help us get dinner ready*, *you’re too young to understand*, *he’s so senile that he can’t follow the*

*conversation, I am hungry enough to eat a horse.* For *too* and *enough*, the complement can be omitted when the idea is contextually given: *she's too young, she's not old enough.*

#### 5.2.17. Adjective comparison

Comparison makes up a huge topic, that will not be conquered during the time of this pilot study, but they're included here because of some further constructions that will include them. The comparative markers also carry extrapositional complements: *more/er-* and *less* → *than*; [*not...*] *so* and *as* → *as*.

EXAMPLES include *She's much more intelligent than you said, are you as angry as you seem, it's less warm today than it was yesterday.*

#### 5.2.18. Comparative Negation with *no* rather than *not*

If I say that *you're not more qualified for the job than I am*, I could believe that we are both well qualified, and that I should certainly be included among the candidates. On the other hand, if I say that *you're no more qualified for the job than I am*, it's assumed that we're both barely qualified, and (say) I'm complaining that they had no right to give you the job. Using this construction seems to suggest that both of the things being compared are at the low end of the scale. *Your puppy is no bigger than a mouse!*

#### 5.2.19. NP-internal degree-modified adjectives

All of the adjective modifiers we've just reviewed can be used predicatively, but there is a construction that allows them to be used attributively, *but only in the case of a singular indefinite count noun*. Those that have extrapositional complements allow them to be extraposed after the noun. The adjectival part precedes the indefinite article. (Compare [*an*] [*intelligent*] man with [*too intelligent*] [*a*] man.) A variant of the construction has an intrusive *of* which sounds more natural in some contexts than others. We have nothing to say about that just now.

EXAMPLES include *you're too intelligent a man to act like that, that's much bigger of a house than we need, that's as sensible a solution as we can expect, is it really that big of a problem, that's no bigger a problem than others we had in the past, that's so big a problem that we'll never be able to deal with it, is this big enough of a box?* The limitation to indefinite singular count nouns is striking: *\*it's not that hot of soup, \*they're no older of people than my parents.*

#### 5.2.20. *One's every something*

I once proposed that a particular expression with *every* was dedicated to talk about indulgence fantasies, but have learned from corpus data that it is also frequent in paranoid talk.

EXAMPLES of the former kind include *we are here to meet your every need, you will obey my every command, my every dream has been fulfilled, I've satisfied my every*

*wish*; but the other kinds include *why are you dogging my every step, they watch my every move, he records my every gesture*. And there are neutral expressions as well, so it probably requires no more than a sense of extreme attentiveness. Whatever it is, the relationship between the POSSESSOR and the noun has to be agentive in some way—it cannot be one of simple possession: *\*they stole my every donut* doesn't seem to work.

#### 5.2.21. Plural-noun reciprocals as predicates

Some plural undetermined nominals can occur as predicates indicating a symmetrical social relation between two people. *We were best friends in high school* can be expressed from one member's point of view: *I was best friends with him in high school*. If the subject is singular, a *with* is needed to identify the other member of the relationship. This only works with nominals that indicate some kind of social relation that inherently is (like *cousin* or *friend*) or can be (like *brother* or *sister*) symmetrical: *we're siblings* can stand alone as a predicate, *we're sons* requires mention of the second term of the relationship, *\*I was foreigners with him in Japan* doesn't work: *foreigner* isn't a relation between two people

EXAMPLES include *we were colleagues in the post office, she is cousins with a very rich man*, and, from the web, *my theory is that Harry's mother is siblings with Voldemort*.

## 6. Opportunities for a construction-expanded FrameNet

The decision to enter constructional information and lexical information in the same database turns out to have many advantages. In particular, it's seldom necessary to worry about whether we're dealing with a lexical or a grammatical structure. Some products of a construction are simply lexical units in essentially every way, except in that they are “generated” rather than requiring individual listing in a dictionary's wordlist: this is true of the products of argument structure constructions as well as a number of derivational patterns, morphological or “zero” derivation. The lexicographer might now have a principled way of deciding whether a “frequent guest” deserves inclusion in the lexicon's standing wordlist. Some constructs behave like ordinary lexical items in their external environment, and can then be annotated as equivalent to single LUs in their own right: the reciprocal *best friends* can be annotated as an ordinary symmetric predicate of the kind that permits both joint and disjoint expression of the paired participants. The phrase *to push one's way* in its external syntax works just like an ordinary motion verb and acquires the valence expectations shared by ordinary motion verbs and can be annotated as such. Many of the constructions produce constituents that fit their environment in normal ways requiring nothing special: a rate expression classified as indicating FREQUENCY, or SPEED, or UNIT\_PRICE, or WAGES, can combine with whatever marking goes with the governing predicate and find its place in the annotations for that predicate. The zero anaphora facts that FrameNet has encountered in preparing lexical descriptions are similar to those that occur with constructions as well, and pose similar challenges to

theories of anaphora. Thus, to take a sentence like *otherwise most members wouldn't have the funds*, a search for cohesion with preceding texts would have to include the condition implied by *otherwise*, the organization presupposed by *members*, and the purpose-indicating complement of the Wherewithal construction that *the funds* are needed for.

Whether parsers can recognize (and interpret) instances of special constructions will remain to be seen. It's possible that a very large sample of construction-annotated texts could provide the learning corpus for statistics-based parsers. An apparent number agreement failure could lead to interpretations that permit such possibilities: *she is friends with the president, a mere twenty pages*. In many cases there are overt markers of a construction that could initiate specific steps to find the components (the phrase *let alone*). A comma before a conjunction in will trigger a search for discontinuities permitted by RNR and Gapping structures. And in some cases the failure to find, in the immediate context, a needed valent of a verb or head of a modifier should guide the search for explanations: the hanging *largely* in the sentence *bears have become largely and pandas entirely noncarnivorous* should serve as a clue.

## References

- Fillmore, C. J. (1982). "Frame semantics". In *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Co. 111-137.
- Fillmore, C. J.; Atkins, B. T. S. (1992). "Towards a frame-based organization of the lexicon: The semantics of RISK and its neighbors". In Lehrer, A.; Kittay, E. (eds.). *Frames, Fields, and Contrast: New Essays in Semantics and Lexical Organization*. Hillsdale: Lawrence Erlbaum Associates. 75-102.
- Fillmore, C. J.; Atkins, B. T. S. (1994). "Starting where the dictionaries stop: The challenge for computational lexicography". In Atkins, B. T. S.; Zampolli, A. (eds.). *Computational Approaches to the Lexicon*. Oxford: Oxford University Press. 349-393.
- Fillmore, C. J.; Kay, P.; O'Connor, M. C. (1988). "Regularity and idiomaticity in grammatical constructions". *Language* 64 (3). 501-538.
- Goldberg, A. (1995). *Constructions. A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Jackendoff, R. (1990). *Semantic Structures*. Cambridge: MIT Press.
- Joshi, A. K. (1985). "Tree-adjointing grammars: How much context sensitivity is required to provide reasonable structural descriptions?". In Dowty, D.; Karttunen, L.; Zwicky, A. (eds.). *Natural Language Parsing*. Cambridge: Cambridge University Press. 206-250.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.
- Tesnière, L. (1959). *Elements de syntaxe structurale*. Paris: Klincksieck.

# Sobre la discontinuidad de las palabras en un diccionario histórico originada por nuestros datos

José Antonio Pascual  
Real Academia Española

## *Abstract*

*In this paper I discuss an seemingly unimportant methodological issue, but one that is of major practical interest in determining how to present examples in an historical dictionary. The issue I will discuss is how to represent those words and senses that “disappear” from usage—and consequently from the historical dictionary—for a long period of time, only to reappear at a later point in history. I discuss several examples to show the implications this sort of development has for the history of these words and senses, which suggests that they should be represented in the Academy’s historical dictionary in a special way.*

Por *discontinuidad* me refiero a la condición de aquellas palabras cuyos ejemplos se interrumpen una o varias veces a lo largo de la historia, por un espacio considerable de tiempo. No se trata de un hecho fundamental metodológicamente hablando, pero sí resulta importante que lo tomemos en consideración para esas decisiones de sentido común que se han de afrontar en la confección de un diccionario histórico. Pues si este está planeado para ser construido primero y consultado después con medios informáticos y tiene además la pretensión de que la presentación de los hechos facilite las posibilidades de interpretarlos, hemos de tener claro cómo mostrar la biografía de estas voces. En el caso de los ejemplos que hemos de incluir en el *Nuevo diccionario histórico de la lengua española* (NDHE) partíamos de lo obvio: establecer un campo especial para la primera y última documentación de cada palabra y colocar luego entre esos dos extremos un par de citas significativas de cada época, actuando con el sentido común que nos lleva a fijarnos en el modo de proceder de otros diccionarios históricos. Esta representación produce, sin embargo, una falsa impresión de lo ocurrido en aquellos casos en que los datos con que contamos no son representativos. Ello nos ha llevado a atender a aquellos ejemplos en que se rompe la continuidad y se da un salto de un período histórico a otro, de un ámbito geográfico a otro, de un estilo a otro...

Ciertamente, cuando hablamos de discontinuidad solemos pensar en esos arcaísmos especiales que suponen una versión léxica de lo que se conoce por “discurso repetido”, que nos acerca más al juego que podemos hacer con las palabras que a la intención de aprovecharlas reciclándolas para su uso. Es lo que ocurre con *magín*, que tal y como ha mostrado Alberto Blecua<sup>1</sup>, se emplea en textos literarios desde

---

<sup>1</sup> Blecua, A. (2005). “Tres notas léxicas al episodio de la *Cueva de Montesinos*”. *BRAE* 85. 112-126.

finales del siglo XVI en situaciones cómicas y en boca de los rústicos; pero, más que un reflejo de los usos de este grupo social, se trataba de un mero remedo literario que acogió Cervantes y fue después ajeno a la literatura, hasta que revivió a finales del siglo XIX, gracias a la utilización que hizo de él Galdós, admirador del *Quijote*. De forma parecida actuó Mariano José Larra con *grida*, que había encontrado en la *Crónica* de don Álvaro de Luna y la empleó, dándole el sentido de “pregón”, en *El doncel de don Enrique el doliente*: “Al punto los jueces de campo mandaron al rey de armas y al faraute dar una grida o pregón, que ninguno fuese osado...”, “disponíanse los archeros a conducir a Elvira al suplicio, estaba ya en pie el impasible verdugo y repetía por tercera vez el rey de armas su grida...”.<sup>2</sup> Situación como la del problemático *hapax* cervantino *lercha*, que muchos pensamos que se trata de un error por otra voz,<sup>3</sup> que exhuma J. Sanchis Sinisterra en un texto lleno de guiños a Cervantes, en boca de un airado Chanfalla —nombre de claro cuño cervantino—: “¡Malditos sodomitas! ¡Debieran matarlos a todos y ensartarlos por las agallas, como sardinas en lercha!”.<sup>4</sup>

Eso mismo ocurre con *alpende*, voz fundamentalmente gallega y portuguesa, desde donde se extendió al andaluz fronterizo con Extremadura y Portugal, y al canario; aunque en la Edad Media es posible que se empleara también en leonés y llegara incluso al aragonés, por más que no haya encontrado en la actualidad rastros en estos dialectos hispánicos.<sup>5</sup> No sorprende, pues, encontrar ejemplos de la palabra en escritores gallegos como Emilia Pardo Bazán, Ramón María del Valle Inclán y Wenceslao Fernández Flórez, así como en Juan Ramón Jiménez, de Huelva, y en

---

<sup>2</sup> Vid. González-Zapatero, B. “La relación entre las formas verbales simples y analíticas en un diccionario histórico” (en prensa). Incluso aparece escrita esta voz —no sé si de ello es responsable el escritor o lo es el impresor— en la forma *grita*: “Dispuesta ya la liza en esta forma, que hemos procurado describir todo lo más fielmente que nos ha sido posible, mandaron los jueces al rey de armas y faraute dar una grita o pregón anunciando el combate, que iba a verificarse en comprobación del juicio de Dios a falta de otras pruebas, y mandando comparecer a las partes o a sus campeones”.

<sup>3</sup> “Oh encantadores aciagos y malintencionados, y quien os viera a todos ensartados por las agallas, como sardinas en lercha”, que F. Rico en su edición de Miguel de Cervantes, *Don Quijote de la Mancha*, Madrid: Alfaguara, 2007, p. 622, cambia en *percha*. Vid. Hernández, P. (2006). “Sardinas en leche”. *Pliegos Yuste*. 49-56.

<sup>4</sup> Sanchis Sinisterra, J. (1992). “El retablo de Eldorado”. En *Trilogía americana*. Madrid: El Público. 56.

<sup>5</sup> Vid. Corominas, J. (con la colaboración de J. A. Pascual) (1980-1990). *Diccionario crítico etimológico castellano e hispánico*. 6 vols. Madrid: Gredos, s. v. (se cita más adelante como DECH; se recurre también en alguna ocasión al DCEC, es decir, el *Diccionario crítico etimológico de la lengua castellana*, que es la primera edición de esta obra de J. Corominas, publicada en Madrid, por la Editorial Gredos, de los años 1955 a 1957) y Pascual, J. A.; Santiago, R. (2004). “Voces romances en la documentación latina leonesa de la Edad Media”. En *Escritos dedicados a José María Fernández Catón, II*, León: Centro de Estudios e Investigación San Isidoro. 107-112. p. 1097, 1098. Al dato que proporcionamos ahí puede añadirse el del *Fuero Juzgo*, citado en el *Corpus Histórico* de la Real Academia Española (CORDE), que se puede consultar “on line”.

Enrique Nácher, de Canarias. La emplea también Juan Benet<sup>6</sup>, posiblemente por haberla oído en sus estancias de trabajo en Galicia y en el extremo occidental de León; Rosalía Vázquez, de origen gallego y que vivió mucho en Galicia, utiliza *alpendes* en la traducción del libro de Ken Follet *Los pilares de la tierra* (“en alpendes a lo largo de los muros de la iglesia, podían verse [...] esculpiendo los bloques de piedra con cinceles de hierro...”)<sup>7</sup>, aunque eligiendo la variante preferida por la Academia.

Entra *alpende* en el *Diccionario* académico en 1884<sup>8</sup>, con el significado de “casilla o cobertizo que sirve para custodiar enseres de mina o de fundición”, que aparecía en una serie de diccionarios técnicos. A esta acepción se le añadió otra más en la edición del DRAE de 1925: “cubierta voladiza de cualquier edificio, y especialmente la sostenida por postes o columnas, a manera de pórtico”, que podría haber sido el origen de la acepción anterior. En la edición del DRAE de 1936<sup>9</sup> se introdujo la voz *alpendre* como propia de Galicia y distinta de la anterior, con varias acepciones, derivadas todas ellas del significado ‘cobertizo’, que se redujeron a esa única acepción en la edición de 1950<sup>10</sup>. A partir de la edición de 1956 del DRAE se remite de *alpendre* a *alpende*, con lo que desaparece esa acepción general de ‘cobertizo’, pero se recupera, al fin, en la edición de 1992 —sin ninguna precisión diatópica—, al unirse *alpendre* y *alpende* en el mismo artículo.

Esta voz, que un diccionario histórico debería representar, según las acepciones, con una marca diatópica y otra técnica, no fluye en esa amplia avenida que es la corriente del río de una lengua a la que, por comodidad y con buscada inexactitud, llamamos el léxico común. Por ello lo probable es que sea su aparición en la obra de Juan Benet —si no es en el diccionario académico— lo que explique que Félix de Azúa<sup>11</sup> nos haya dado la sorpresa de contemplar en un desolado mundo barcelonés unos “ruinosos alpendres de uralita”.

<sup>6</sup> “Pronto el agua comenzó a filtrarse a través del alpendre”. Benet, J. (1967<sup>1a</sup>, 1996). *Volverás a Región*. Barcelona: Destino. 124; “A la sombra del alpendre del cuartelillo”. Margerot Benet, J. B. (ed.) (1984). *Saúl ante Samuel*. Madrid: Cátedra. 338.

<sup>7</sup> Follet, K. (1996). *Los pilares de la tierra*. Barcelona: Plaza y Janés. 46.

<sup>8</sup> Real Academia Española, *Diccionario de la lengua española*. Se acude normalmente a la última edición, la 22<sup>a</sup>, Madrid: Espasa-Calpe 2001; si bien, en otros casos, como en el presente, se citan ediciones anteriores, proporcionando su la fecha. En el título de la obra aparece *castellana* en lugar de *española* desde la primera edición, hasta la de 1914.

<sup>9</sup> *Vid.* Campos M.; Pérez Pascual, J. I. (2007). “Armando Cotarelo Valledor y los galleguismos del DRAE, 1936”. En *Ex amicitia et admiratione. Homenaje a Ramón Santiago*. Madrid: Ed. del Orto. 199.

<sup>10</sup> Los datos dialectales procedentes de Galicia, Andalucía y Canarias dan cuenta de una “construcción de mala calidad, donde se guardan instrumentos de trabajo, trastos, etc. o se acogen animales”. *Vid.* las acepciones § 1, § 2, § 4 y § 6 del *Diccionario histórico de la lengua española* (1960-1994) de la Real Academia Española. Madrid. (DHLE), s. v.

<sup>11</sup> *El País*, 10 de julio de 2007, p. 13.

Vamos a introducirnos ya por aquellos casos en los que la discontinuidad está originada por la endeblez de nuestros materiales; situación que hemos de tener en cuenta en la presentación de los ejemplos de un diccionario histórico: para que puedan completarse en el futuro con nuevo datos o con hipótesis sobre el porqué de tales discontinuidades.

## 1. Discontinuidad entre los indicios de la existencia de un concepto y su expresión por medio de una palabra

Empezamos por señalar la distancia que suele mediar entre las informaciones preléxicas sobre algunas palabras, a las que a veces llegamos a acceder, y su primera documentación. Como las lenguas no disponen de un nombre para designar cada uno de los aspectos de la realidad,<sup>12</sup> existen situaciones en que podemos expresar un concepto, por más que no dispongamos de una etiqueta para él; y, sin embargo, a veces logramos allegar información indirectamente lexicográfica, que muestra las dudas, vacilaciones y perplejidades de los hablantes, previas a la aparición de la palabra que designa esa realidad. Es el caso de *esquí*, que surgió para nombrar un objeto que a lo largo del tiempo resultaba ajeno y distante para la mayor parte de los habitantes de la Península Ibérica, si bien a lo largo del siglo XX se fue haciendo, poco a poco, habitual. Empezó siendo un término propio del léxico deportivo, que es como la emplea a finales del siglo XIX Ricardo Becerro de Bengoa,<sup>13</sup> orientándonos por medio de un adjetivo sobre la realidad geográfica en la que los esquís son normales: “patinan con los *skiss* noruegos”; seis años después seguimos encontrando la misma referencia a Noruega: “*skis*, largos patines de uso muy frecuente en todos los países donde la nieve cubre el suelo durante casi todo el invierno. En Noruega están muy generalizados los *skis*”<sup>14</sup>; hemos de esperar dos años más para que un periodista los sitúe en otro país, Suecia, haciendo además una descripción somera del artilugio: “Los pasados días celebróse en Holsenkollen un concurso de *skis*; los Reyes de Suecia realzaron con su asistencia el acto, al cual asistió una inmensa multitud. Como es sabido los *skis* son unos patines de madera de más de un metro de longitud, de que se sirven los suecos con verdadera maestría”.<sup>15</sup> En ese momento, sin que fuera una realidad experimentada directamente por la mayor parte de los lectores de los periódicos y revistas citadas, se trataba, por lo menos, de algo conocido precisamente a través de ellos, que se etiquetaba con la palabra extranjera *ski*, con que se designaba el objeto en otras lenguas. Tanto *ski* como *skiador* aparecen en la *Enciclopedia Espasa*<sup>16</sup> con una definición y etimología que se mantienen en el

---

<sup>12</sup> Cruse, A. (2004). *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press. 127-128.

<sup>13</sup> *La Ilustración Española y Americana*, 22 de noviembre de 1898.

<sup>14</sup> *ABC*, 3 de febrero de 1904.

<sup>15</sup> *Blanco y Negro* (1906), p. 778.

<sup>16</sup> *Enciclopedia universal ilustrada europeo-americana*. Barcelona: Espasa, 1905-1903. T. 56, s. v.

*Diccionario de la lengua española* de Alemany y Bolufer de 1917.<sup>17</sup> En la lexicografía académica se registran al fin las variantes *esquí*, *esquiador* y *esquiar* en el *Diccionario Manual* de la Real Academia Española de 1927<sup>18</sup>, junto a *ski*, esta última como voz danesa; y de ahí pasan al DRAE en la edición de 1936.

Acabo de referirme a una realidad no excesivamente distante, al menos no tanto como lo había sido en la segunda mitad del siglo XVI, en que Antonio de Torquemada describe una forma sorprendente de esquís, demostrando que no disponía de una palabra española ni extranjera para designarlos:

Los que han de caminar a pie encima de los yelos, si quieren hazer con brevedad un camino, toman un madero rollizo de una madera muy fuerte, y por sola una parte es llano, sobre la cual asientan los pies, atando el pie siniestro al madero y llevan el derecho suelto, en el cual llevan un çapato hechizo, y a la punta con un hierro hecho de tal manera que, aunque den un gran golpe en el madero, ningún daño recibe el pie, porque da en hueco; y en las manos llevan unos bordones grandes, como medias lanzas, con tres puntas muy agudas al cabo, y proveyéndose de lo necessario para el camino, yendo uno solo o muchos en compañía, puesto cada uno encima de su palo, sacan el pie derecho atrás y danle un muy gran puntapié, y el palo rollizo comienza a resbalar por el yelo, con tan gran ligereza, que algunas vezes no para en tanto trecho como un grandíssimo tiro de ballesta, y aún más; y quando sienten que el madero va parando, dan con el bordón en el yelo, hincando las tres puntas en él, que de otra manera caerían, y tornando a componerse, vuelven a dar otro golpe; y así, en una hora, caminan tres y cuatro leguas.<sup>19</sup>

La descripción de esos curiosos artilugios para desplazarse por la nieve, cuya aparición nos asegura la inexistencia de la palabra, se distancia en cientos de años del momento en que el objeto contó con un nombre que lo designara.

En otros casos esa información nos ilustra sobre las dudas y vacilaciones que pueden existir antes de que los hablantes opten por un determinado neologismo, en los

---

<sup>17</sup> Alemany Bolufer, J. (1917). *Diccionario de la lengua española*. Barcelona: Ramón Sopena.

<sup>18</sup> Real Academia Española (1927). *Diccionario manual e ilustrado de la lengua española*. Madrid.

<sup>19</sup> De Torquemada, A. (1983). *Jardín de flores curiosas*. En Rodríguez Cacho, L. (ed.) (1994). *Obras Completas*. Madrid: Biblioteca Castro. Vol. I. 495-904. 861, 862. De esa descripción parece tomada la que se hace en *Los trabajos de Persiles y Sigismunda* (Romero Muñoz, C. (ed.); 1997; Madrid: Cátedra; 398-399), dato e idea que debo a la amabilidad de Rosa Navarro. En el texto citado de Torquemada (p. 862) se habla de otras realidades mejor conocidas en la época, describiendo lo que hoy llamamos *trineo*, que se designa como *tabladillo* (p. 862, 864) o los *patines* (“unos hierros llanos con unas puntas adelante, a que llaman patines, y con éstos resbalan por los yelos, de suerte que en poco tiempo hacen muy largo camino”). Más de cuarenta años después del libro de Antonio de Torquemada, Diego de Ufano en su *Tratado de artillería*, de 1613 (*apud* Blas, C. (2007). *Estudio léxico de los tratados de artillería españoles del siglo XVI* [tesis doctoral]. Salamanca. 706) cita *trineo* acompañado de un sinónimo derivado del arag. *eslizar* “deslizar”: “trineo o eslizo”. El verbo que A. de Torquemada empleaba para “deslizarse” era *deleznarse* (p. 862).

momentos mismos de su puesta en circulación; situación que muestra, no sin ironía, José Fernández Bremón al referirse al *cine* o los *rayos x*:

La Academia de la Lengua no quiere explicarse y decirnos de un modo oficial qué vocablo debe usarse para designar esas fotografías de movimiento que están hoy tan en boga. [...]. Por fortuna, ya no se llama solo cinematógrafo; otro le da el nombre de muovógrafo, más corto, pero difícil de pronunciar; algunos le suavizan denominándole movígrafo, y una chula le llamaba monisabio, dando a entender que se trataba de figuras que el vulgo llama monos y que éstos tienen un carácter científico y progresivo. Nosotros no podemos adoptar ninguno mientras la Academia no decida. Y no sería malo, ahora que se ha abierto una oficina donde por corto interés se explica y enseña la acción de los rayos X, que den también nombre adecuado a ese fenómeno, que permite ver el esqueleto de las personas vivas: nosotros, mientras no se halle nombre mejor, le llamaremos transparencia.<sup>20</sup>

Para este tipo de vacilaciones disponemos en la época moderna de buena información en el campo de la moda, de los deportes, de la técnica, etc., solo con atender a la prensa, donde aparecen cientos de observaciones como la siguiente sobre la voz *descalificación*:

En rigor podríamos traducir la palabra disqualification, empleada en una acepción misma tanto en el lenguaje hípico de Inglaterra como en el de Francia, por la castellana invalidación. Pero preferimos admitir aquel barbarismo, porque si bien esta palabra comprende el sentido absoluto de aquella, no precisa su significado para los sportsmen españoles en este caso determinado. Calificación se llama el conjunto de condiciones impuestas a un caballo, a un propietario o a un jockey para poder tomar parte legalmente en una carrera. La descalificación, es por consiguiente la pérdida o anulación de la calificación.<sup>21</sup>

Incluso hay que contar en esta etapa que hemos denominado preléxica con los casos en que un traductor abandona en su texto una palabra del original, bien porque no la entiende, bien porque deja para más adelante dar con la traducción adecuada. No obstante, son datos que, aunque debemos acogerlos, no sirven como antecedentes de la existencia de una voz, sino, por el contrario, como prueba de su inexistencia en un determinado momento, como ocurre con *altilobi* en el *Libro de los Gatos*<sup>22</sup> o con *frachaso* y *biçaro* en la traducción de la *Divina Commedia* del Marqués de Villena<sup>23</sup>, con las que se adaptan voces extranjeras que no se entienden —la francesa *antiloipe* y las italianas *fracasso* o *bizarro*— y que no se convierten, por tanto, en signos en la lengua de llegada.

---

<sup>20</sup> *La Ilustración Española y Americana*, 15 de noviembre de 1896.

<sup>21</sup> *El Campo: Agricultura, Jardinería, Sport*, 1 de enero de 1880.

<sup>22</sup> Pascual, J. A.; García, R. (2007). *Límites y horizontes en un diccionario histórico*. Salamanca: Ed. de la Diputación de Salamanca. 173.

<sup>23</sup> Pascual, J. A. (1974). *La traducción de la Divina Commedia atribuida a D. Enrique de Aragón. Estudio y edición del Infierno*. Salamanca: Universidad de Salamanca. 96, 97.

No sería necesario decir que para el mayor número de las palabras que se han de introducir en un diccionario histórico no contamos con este despertar a la vida como el de las que acabo de citar y no vamos a poder, por tanto, prepararnos para asistir a su primer vagido. Lo normal es que los elementos léxicos se nos aparezcan de repente en un texto, con más o menos vitalidad, dispuestos en cualquier caso a abrirse camino por el discurrir de la lengua.

## 2. La provisionalidad en la formulación de una hipótesis filológica

Creada una palabra o recibida en herencia, su futuro depende de múltiples factores. Llegamos así a ese momento en que lo incompleto de nuestros datos origina que algunas palabras presenten amplios espacios vacíos de ejemplos, en ese *continuum* que es su historia: por completos que sean los corpus con los que contamos, no por ello reflejan de un modo fiel la realidad histórica, ya que su información ni es exhaustiva ni, aunque lo fuera, podría asegurarnos que una voz no existiera por no aparecer en el corpus. Lo cual obliga a suplir estos huecos con hipótesis que permitan encontrarles un sentido.

Cuando, con los pocos ejemplos de *escollo* que había reunido, tuvo que dar Corominas una explicación sobre esta palabra hubo de orientarse por su ausencia en la *Celestina*, en el *Quijote* y en los diccionarios de Nebrija, Covarrubias, e incluso en el de Cristóbal de las Casas —que se sirve en cambio de *peñasco*<sup>24</sup>; era razonable que en esas condiciones concluyera en que se trataba de un italianismo tardío y literario, caracterizador del léxico del Barroco. No es una explicación desdeñable, pero requiere de algunos retoques, ahora que podemos situar la voz en obras literarias y de navegación del siglo XVI; todo lo cual permite rellenar el hiato que teníamos desde los primeros años del siglo XVI a los primeros del siglo XVIII y consiguientemente cambiar la idea que nos hacíamos de esta palabra en el DECH<sup>25</sup>, de forma que su paso a los diccionarios de *Autoridades* y de Terreros, no solo supone la recuperación de un término propio de literatura barroca, sino que enlaza también con su empleo en la jerga marítima en España y, sobre todo, en América, donde incluso había penetrado en la lengua común.<sup>26</sup>

No tenían nada que ver los medios de que disponía el sabio filólogo catalán con los que contamos ahora, como el *Corpus del Diccionario Histórico del Español* (CNDH), de 50 millones de ocurrencias, aunque todavía en fase experimental. Con estos materiales mejoran ostensiblemente las posibilidades que nos brindaba el DECH

---

<sup>24</sup> El *Nuevo Tesoro Lexicográfico del Español* (S. XIV-1726) (Nieto Jiménez, L.; Alvar Ezquerro, M.; 2007; Madrid: Arco Libros; 11 vols) confirma lo tardío de la introducción de *escollo* en los diccionarios españoles: no aparece hasta 1600, por más que de esa fecha a 1721, acojan la voz diecinueve diccionarios —de ellos varias ediciones del de Oudin—.

<sup>25</sup> DCEC, s. v. *muro*.

<sup>26</sup> Tomo esta explicación de M. J. Gomez Gonzalvo (2007), *El español americano del siglo XVIII en la obra de Abad y Lasierra* [tesis doctoral], Universidad de Zaragoza, p. 334, quien proporciona además un dato precioso sobre su uso en América, a través del conocimiento de esta voz que muestra Fr. Íñigo Abad Lasierra.

para entender mejor la historia en que algunas palabras se comportan —real o aparentemente— en su devenir como las aguas del Guadiana. Es el caso de *muralla*, préstamo que sustituye a *cerca* y *muro*.<sup>27</sup> Si al principio, siguiendo la opinión de Corominas<sup>28</sup>, me parecía que se trataba de un italianismo, llegué luego a pensar que posiblemente entrara en dos ocasiones en nuestra lengua y por dos conductos distintos: en el siglo XIV, como galicismo, a través del aragonés<sup>29</sup>, de donde se extendería su uso a algunos escritores del Cuatrocientos que tenían la vista puesta en los usos de Aragón<sup>30</sup>; desaparecería después y volvería a entrar a finales del XVI de la mano del italiano. Los datos que logré allegar del siglo XVI<sup>31</sup>, tan exiguos, me llevaron a esa idea; sin embargo, he de cambiarla de nuevo, gracias a la información del CNDH, en el que los nuevos ejemplos del siglo XVI no animan a pensar que se hubiese dado entonces una interrupción en el uso de esta palabra.

### 3. Las hipótesis de los propios hablantes

No son solo los filólogos quienes se ven obligados a hacerse una idea de las cosas, a partir de una información fragmentaria, sino que los propios hablantes actuamos del mismo modo, pues no disponemos, frente a lo que se cree, de una información completa de los usos de nuestra lengua. Lo mostraré, para empezar, con las perplejidades que pueden presentarse a un hablante actual, por medio de algunas palabras y acepciones en las que el español europeo y el americano presentan algunas sutiles, y no tan sutiles, diferencias.

---

<sup>27</sup> Son *cerca* y *muro* los vocablos que se emplean en documentos leoneses y castellanos medievales: *Fueros Leoneses*, Alfonso X, *Otras de Roma*, Juan de Mena, R. de Clavijo, *Embajada a Tamorlán* y luego en la traducción de 1515 de la *Divina commedia* de P. Fernández de Villegas, de 1515. *Cerca* y *muro* es lo que se registra incluso en escritores navarros y aragoneses, como García de Euguí, Carlos de Viana, Martínez de Ampíes.

<sup>28</sup> DCEC, s. v. *muro*.

<sup>29</sup> Esta voz, cuyo punto de partida debió ser el francés se registra pronto en francoprovenzal, en occitano y en catalán; en aragonés está ya en Fernández de Heredia. *Vid.* datos en Pascual, J. A. (1988). “Los aragonesismos en *La Visión Delectable* del Bachiller Alfonso de la Torre”. En Ariza, M.; Salvador, A.; Viudas, A. (eds.). *Actas del I Congreso Internacional de Historia de la Lengua Española*. Cáceres: Arco Libros. 647-676, p. 653.

<sup>30</sup> Pascual, J. A. *La traducción*. cit. p. 98, 99. cf. DECH, s. v. *muro*.

<sup>31</sup> Puedo añadir a los datos castellanos de *muralla* presentados en *La traducción*, cit. p. 98 y 99, y en *Los aragonesismos*, cit., p. 653: “en la muralla labrada se cría la culebra maldita”, Mac E. Barrich (ed.), 1976, *Tercera parte de la tragicomedia de Celestina* (1536), Philadelphia: University of Pennsylvania Press, acto xxiv, p. 239; en *La muy lamentable conquista y cruenta batalla de Rhodas*, traducida del latín por Christoval de Arcos (Valladolid: Juan de Villquirán, 1549) se alterna *muro* (f.º 28 vº a [numerado por error como 27]), y *muralla* (fols. 29 rº a, 34 vº b y 51 vº b [numerado por error como 41]); “el hazer el aproge, que llaman, que es allegarse a las murallas y fossos”, B. de Mendoza, 1594, *Theórica y práctica de guerra*, 94.

Ahí tenemos *amigable*<sup>32</sup>, que me sorprendía encontrarla en mi juventud en los libros editados en la Argentina. Pensaba que se trataba de una palabra recién creada, pues no sabía que había sido normal, frente a *amistoso*, en el español clásico. A lo largo de mi vida he ido viendo como esa voz se ha reintroducido en España, de donde había desaparecido en la práctica, e incluso ampliándose al ámbito específico de la informática (en que un *entorno amigable* no es conmutable por un *entorno amistoso*). Su recuperación aquí, a través de América, sería un caso claro de discontinuidad; aunque de discontinuidad relativa, pues no se había perdido en todos los niveles, sino que solo se había adelgazado su uso, tal y como nos permite comprobar el CNDH: así se registra, aparte de en un amplio número de escritores americanos, en Enrique Gil y Carrasco, Mesonero Romanos, Estébanez Calderón, Joaquín Costa, Emilia Pardo Bazán, Concha Espina y Pedro Laín Entralgo. La penetración de la voz en la lengua actual hace que se vea como un hecho normal —me refiero al uso lingüístico— el siguiente ejemplo: “Julio Martínez Santaolalla aparece en actitud amigable junto al sentado *Reichsführer* nazi [Himmler]”.<sup>33</sup>

Algo parecido ha ocurrido con *portar*, palabra que, simplificando los hechos, un mexicano puede utilizar en combinaciones que en la última edición del *DRAE*, de 2001, se consideran anticuadas: por ejemplo, con *un traje* o *un bolso*<sup>34</sup>, casos en los que un español se serviría de *llevar* o *traer* y reservaría el otro verbo para aludir a los objetos importantes que se muestran ostentosamente, preferentemente en situaciones formales —por ejemplo, *un féretro*, *una cruz* o *una bandera*, etc.—<sup>35</sup>

<sup>32</sup> Trato de esta palabra en “The Necessary Role of History in Dictionaries of Current Spanish”. En Gorrochategui, J. (ed.) (2003). *Basque and (Paleo)hispanic Studies in the Wake of Michelena’s Work*. Vitoria: Servicio Editorial de la Universidad del País Vasco. 83-108, p. 95.

<sup>33</sup> Pie de foto de *El País*, 6 de abril de 2008, “Cultura”, p. 46.

<sup>34</sup> Usos que encuentro en la traducción que J. C. Barrera hace de la novela de E. McBain, 1962, *Mírenlos, Muertos*, México: Diana, 29, 31, 41, 72.

<sup>35</sup> Como es la de un entierro en que los “familiares y amigos portaban ayer el féretro de Jaled Kelkal en Lyon” (pie de foto de *El País*, 7 de octubre de 1995, p. 8; aunque se trata de una crónica enviada desde Francia, en la que aparece un uso abusivo galicista de *portar*, como “La bombona de gas portaba las huellas digitales del joven”; el día 22 de octubre de 1995 vuelve a repetirse la foto y parcialmente el titular, en el que aparece de nuevo: “portaban su féretro”) o una procesión en la que se portan distintos objetos: “tres vecinos en cabeza, uno portando la cruz, otros dos sendos faroles de hierro, detrás el cura con la Custodia en la mano” (J. P. Aparicio, “Domingo”, *El País*, 16 de junio de 1996, p. 7) o una cabalgata en la que se porta una carroza: “Tres asociaciones de jóvenes y vecinos portarán las carrozas en la cabalgata de los Reyes Magos” (*Tribuna* de Salamanca, 22 de diciembre de 1996, p. 23) o un peregrino, o quien actúa como tal, que recoge “unas gotas de agua en una ampolla que portará en su peregrinación hacia Venecia” (*El País*, 11 de septiembre de 1996, p. 4) o pueden unos manifestantes portar “unos carteles” para pedir “el traslado de los presos al País Vasco” (*El País*, 18 de octubre de 1996, p. 17) o “porta[r] banderas con símbolos nazis” (I. Ferrer, *El País*, 12 de abril de 2000, p. 80).

o referido el verbo a las armas: *portar armas*<sup>36</sup>. Paralelamente *portador* se emplea en registros formales<sup>37</sup>, a pesar de que no cuenta para el registro coloquial con un sinónimo \**llevador*, derivado de *llevar*. El hecho es que se están acortando estas diferencias entre el español de uno y otro lado del Atlántico con respecto al uso de *portar* —a lo que, incidentalmente, puede haber ayudado en algunos casos el francés—. Resulta así normal encontrar en libros y periódicos publicados en España ejemplos en que *portar* se combina con *una chaqueta*<sup>38</sup>, *unos pendientes*<sup>39</sup>, *la documentación*<sup>40</sup>, *pasaportes*<sup>41</sup>, *billetes*<sup>42</sup>, *un teléfono celular*<sup>43</sup>, *un perro de peluche*<sup>44</sup>,

<sup>36</sup> Luis Sepúlveda, de origen chileno, lo emplea en su novela *Un viejo que leía novelas de amor*, Barcelona: Tusquets, 1995, p. 60. En cuanto a la prensa española, encuentro en ella: “el fugitivo no portaba ningún arma de fuego” (*El País*, 18 de octubre de 1995, p. 14), “tenía licencia para portar armas” (*El País*, 6 de junio de 1995, p. 3), “el derecho constitucional a portar armas” (*El País*, 13 de abril de 1996, p. 6), “la policía [británica] [...] cumplía su función sin portar más armas que la porra” (S. Juliá, *El País*, 2 de junio de 1996, p. 17), “portaba un revólver y varios cargadores” (*El País*, 19 de diciembre de 1996, p. 18), “portaba una pistola”, (telediario de la 1ª cadena de TVE, el 13 de julio de 1996, a las 15.15 horas), “un par de millones de personas [...] portando escopetas” (J. Araújo, *El País*, 16 de octubre de 1996, p. 32), “portaban cuatro minas magnéticas” (I. Ferrer, *El País*, 11 de julio de 2000, p. 72), “para impedir que Juan José le clavara el objeto punzante que portaba” (*El País*, 14 de octubre de 1998, p. 31).

<sup>37</sup> Aparte de aquella disparatada monserga que tantas veces oímos los españoles en el franquismo de que “el español es portador de valores eternos”, podemos leer que “nadie dice que las personas individuales son auténticas portadoras de derechos humanos” (F. Savater, cit. por J. Pérez Royo, *El País*, 17 de diciembre de 1998, p. 14); “Serán portadores de una modernidad que ya sabe que ahora se vende el cuerpo y no el alma” (J.-M. Ullán, *El País*, 6 de marzo de 1998, p. 36); “ha sido cortejado [...] por políticos conocidos, que le ha visto como portador de las ideas de cambio, orden y lucha contra la corrupción y la delincuencia” (P. Bonet, *El País*, 18 de octubre de 1996, p. 3); “Consciente de que Europa es un continente portador de civilización” (Preámbulo de Proyecto de Constitución europea, *El País*, 29 de mayo de 2003, p. 5). O para casos muy concretos: “un oficial de alto rango, portador de comprometedor información” (*El País*, 30 de octubre de 1996, p. 10), “el tabaco es portador de una droga, la nicotina” (X. Bru de Salas, *El País*, 18 de febrero de 1998, p. 12); “las partículas intermediarias de las fuerzas, como los fotones, portadores de la luz” (*El País*, 27 de septiembre de 2000, p. 36); “el sedicente prestigio de una cultura depende de la categoría social de sus portadores” (Adorno, Th. W. (1962). *Prismas*, trad. de M. Sacristán. Barcelona: Ariel. 35). Por no entrar en usos especializados como “cheque al portador”, “portador de una enfermedad” o “el artefacto, de fabricación casera explotó cuando un portador lo manipulaba” (*El País*, 28 de junio de 1997, p. 6).

<sup>38</sup> *El País*, 21 de octubre de 1995, p. 29.

<sup>39</sup> *El País*, 12 de agosto de 1996, p. 12.

<sup>40</sup> *El Mundo* 20 de diciembre de 1996, p. 28; *El País*, 31 de mayo de 1996, p. 15.

<sup>41</sup> J. Duva, *El País*, 15 de octubre de 1997, p. 25.

<sup>42</sup> *El País*, 27 de octubre de 1997, p. 30.

<sup>43</sup> *El País*, 31 de mayo de 1996, p. 47.

<sup>44</sup> *El Adelanto de Salamanca*, 25 de octubre de 1996, p. 53.

*una bolsa*<sup>45</sup>, *una maleta*<sup>46</sup>, *droga*<sup>47</sup>, *carne* o *masa muscular*<sup>48</sup> y, naturalmente, aparece en combinaciones más formales, como las que dan lugar a *portar un gen*<sup>49</sup>, y aun a *portar la culpa*<sup>50</sup>. Esta ampliación del uso que ha experimentado la palabra en España puede explicar ejemplos como los siguientes, no exentos de ironía: “Penetraba en el vestíbulo seguido de Domi, que portaba sus prendas como un escudero a su príncipe”<sup>51</sup>, “...creía haber sido atacado por un mosquito hembra [...] cuyo abdomen, en lugar de portar un agujón, llevaba la aguja con la que...”,<sup>52</sup> “el reconocimiento de la región genital, que se alumbraba con una bombilla para ver si portaba parásitos”.<sup>53</sup>

Con todo, en los ejemplos anteriores de *amigable* y *portar* se perciben algunas diferencias: mientras que a mediados del siglo XX yo entendía el significado de *amigable*, aunque no perteneciera ni siquiera a mi léxico pasivo, ello no me llevaba a emplearlo en una situación formal, sino a rechazarlo de plano; *portar*, en cambio, no lo hubiera evitado en tal situación. Eso mismo me ocurría con *liviano*<sup>54</sup>, del que, a diferencia de lo que ocurre con los hablantes argentinos, para muchos de los españoles pertenece al léxico pasivo, ya que el término usual es *ligero*; sin embargo en un estilo formal podemos acudir a este adjetivo, como hizo José Ortega y Gasset (“capricho liviano”)<sup>55</sup> o recientemente Antonio Muñoz Molina<sup>56</sup>. La situación es muy parecida a la de *angosto*, aunque con algunas diferencias que, como en el caso de *portar*, tienen que ver con las palabras con que se combina: ese joven español de mediados del siglo XX que era yo, a través de cuyos ojos estoy observando aquella realidad, hubiera disentido también de un argentino prefiriendo calificar unos escalones de *estrechos* y no de *angostos*, pero no hubiera dudado en acudir a *angosto* aplicándolo a un valle, a un paso (en el sentido de ‘lugar por el que se puede pasar’),

<sup>45</sup> *El País*, 5 de marzo de 1997, p. 21.

<sup>46</sup> Camilieri, A. (2000). *La voz del violín*, trad. de M<sup>a</sup>. A. Menini Pagès. Barcelona: Emecé. 135.

<sup>47</sup> J. M. Lázaro, *El País*, 15 de marzo de 1997, p. 28.

<sup>48</sup> “Domingo”, *El País*, 1 de agosto de 1999, p. 11.

<sup>49</sup> *El País*, 4 de febrero de 1999, p. 28.

<sup>50</sup> A ello se refiere un juez muy formal en su escritura, A. Ibáñez, *El País*, 8 de febrero de 1998, p. 22.

<sup>51</sup> Longares, M. (2002). *Romanticismo*. Madrid: Punto de Lectura. 174.

<sup>52</sup> J. J. Millás, *El País*, 12 de diciembre de 2003, p. 72.

<sup>53</sup> S. Sánchez Montero, “Domingo”, *El País*, 26 de diciembre de 1997, p. 19.

<sup>54</sup> Para su historia, relacionada con *ligero*, *vid.*, Eberentz, R. (1998). “Dos campos semánticos del español preclásico, *fácil* y *difícil*”. En Andrés Suárez, J.; López, L. (eds.). *Estudios de lingüística y filología españolas. Homenaje a Germán Colón*. Madrid: Gredos. 167-183, p. 170-172.

<sup>55</sup> Ortega y Gasset, J. (1998). *La rebelión de las masas* [1930]. Mermall, T. (ed.). Madrid: Castalia. 2<sup>a</sup> ed. 281.

<sup>56</sup> Muñoz Molina, A. (1997). *Plenilunio*. Madrid: Alfaguara, 1997, p. 367. Su uso va aumentando o esa es la sensación que yo tengo. Leo así “como yo era tan liviano, la señorita me tomó bajo los hombros y me atrajo”, B. Hrabal, 2003, *Yo que he servido al rey de Inglaterra*, trad. de M. Mlenjková y A. Ortiz, Barcelona: Destino, p. 19.

a una habitación<sup>57</sup>, en una situación parecida a la que señalaba antes sobre algunos usos de *portar* en la actualidad. No quiere esto decir que los demás compartieran esta apreciación mía de las cosas, pues me estoy situando ante la realidad como testigo de los hechos lingüísticos, con la misma ingenuidad con que debe actuar el lexicógrafo cuando examina las fichas sobre las que ha de construir la historia de una palabra. En el caso de *angosto* mi opinión se daba de bruces con el uso clásico, que, en cambio, mantenía doña Emilia Pardo Bazán al escribir, por ejemplo, “vestidito angosto”<sup>58</sup>; es un uso que mantiene también Antonio Muñoz Molina<sup>59</sup>, a menos que —no le es fácil al propio lexicógrafo saberlo— en vez de servirse de una palabra antigua esté eligiendo algo que ha aprendido de sus lecturas de textos americanos.

#### 4. La discontinuidad de este tipo en el léxico de la vida cotidiana del pasado

Los ejemplos del apartado anterior nos previenen sobre las dificultades que se le presentan al hablante para interpretar hechos que en apariencia resultarían fáciles de entender por la mera cercanía temporal que mantiene con ellos, que no nos ha bastado para saber por qué A. Muñoz Molina empleaba *angosto* en una novela suya. Por ello, las pretensiones del lexicógrafo han de reducirse muchas veces a dejarlo todo preparado para que sean después los filólogos quienes aborden los problemas en las mejores condiciones posibles. Vamos a mostrarlo a través del hiato que se da entre el registro de alguna voz en los documentos antiguos y los informes dialectales del siglo XX.

##### 4.1. Legua ‘duela’

La voz *legua* ‘duela’ —sin ninguna relación con la *legua* ‘medida de longitud’— aparece en un documento zamorano de 1276 (“XVI *leguas* de otra cuba e arcos pora estas *leguas*”<sup>60</sup>); se trata de un celtismo<sup>61</sup> propio del área occidental peninsular,

<sup>57</sup> “el ventilador no era suficiente para ventilar aquel cuarto angosto”, A. Tabucchi, 1995, *Sostiene Pereira*, trad. de C. Gumpert y X. González Rovira, Barcelona: Anagrama, p. 34. Recuerdo haber encontrado una “casa angosta” en las páginas 15 y 16 un libro en que José Moreno Villa recogía sus experiencias sobre Nueva York, cuya referencia he perdido.

<sup>58</sup> *España Moderna*, febrero de 1896.

<sup>59</sup> A. Muñoz Molina, *Op. cit.*, p. 416, y *passim*.

<sup>60</sup> *Tumbo Blanco de Zamora*. Me sirvo de la transcripción mecanografiada de este documento, hecha por J. L. Martín, que tan amablemente puso a mi disposición.

<sup>61</sup> El célt. \*LEUBA (J. Pokorny, 1959-1969, *Indogermanisches etymologisches Wörterbuch*, Bern: Francke, p. 690: *Leup-*, *leub-*, *leugh-*; y A. Walde, 1927, *Vergleichendes Wörterbuch der indogermanischen Sprachen*, publicado por J. Pokorny, Berlin, t. II, p. 417 ss.) no presenta problemas fonéticos ni semánticos: entre los significados que adquieren los resultados de estas raíces indoeuropeas están los de “corteza”, “cesta”, “maderamen”, “caparazón”, “recipiente de madera”, “cáscara”, “tabla”, “cráneo”, etc., que se avienen perfectamente con el significado ‘duela’ del término hispánico. Existe, por otro lado, un grupo de palabras románicas que podrían tener alguna relación con esta raíz, como algunas que aparecen en el *Französisches etymologisches Wörterbuch*, de W. v. Wartburg, Basel: Helbing & Lichtenhahn, t. V, p. 457 y ss., s. v. *lūpus*, y, sobre todo, en p. 370, s. v. \**liobba* prerr. “vaca”, que se extienden

leonesa, gallega<sup>62</sup> y portuguesa<sup>63</sup>. Llega al castellano, donde se registra en el becerro abulense de 1303<sup>64</sup>; luego en la traducción cuatrocentista del I canto de la *Commedia* se explica así el it. *lulla*: “es lo que llamamos legua de cuba o costera”<sup>65</sup>; y aparece en el *Libro de miseria de omne*:<sup>66</sup>

<sup>61</sup> [cont.] hasta Suiza y se prologan hasta Albania; incluso, en el ámbito hispánico una serie de voces que el DECH (s. v. *lobo*) relaciona con *loba*, con una explicación semántica que dista mucho de ser razonable: el ant. *loba* y murc. *lobada* “lomo entre surco y surco”, y hasta el ast. *llobacho* “madero o travesaño fuerte que une las tiernas del carro en su parte media” (Fernández González, A. (1959). *El habla y la cultura popular en Oseja de Sajambre*. Oviedo. p. 299; cfr. Penny, R. (1978). “Trenca del lobu”. En *Estudio estructural del habla de Tudanca*. Tübingen: Niemeyer. p. 155; y los datos que presenta M. Alvar, editor de G. Rohlf, 1979, *Estudios sobre el léxico románico*, Madrid: CSIC, p. 57, n. 81). A. Llorente Maldonado (“Correspondencias entre el léxico salmantino y el léxico de Aragón, Navarra y la Rioja”. En *Serta Philológica F. Lázaro Carreter* I. Madrid: Cátedra. 1983. 329-341, p. 332) acepta “su posible origen céltico”. Se fija ahí en un *leyua* “contraventanas exteriores” en Vera del Bidasoa, (en M. Alvar (dir.), A. Llorente, T. Buesa y E. Alvar (col.), 1979-1981, *Atlas Lingüístico y Etnográfico de Aragón Navarra y Rioja* (ALEARN), 6 vols., Madrid: CSIC, t. VI, mapa 808, Na 100), pues “tanto la duela como la contraventana son tablas o cosas hechas con una tabla o varias tablas”. No sabe si “tendrá que ver etimológicamente con *legua*” o si “se trata de una voz vasca de distinto origen; procede sencillamente de una confusión de los encuestadores del ALEARN que debieron señalar las contraventanas al preguntar a sus informantes y estos debieron creer que aquello por lo que les preguntaban eran las ventanas mismas, hecho justificable en una encuesta tan difícil como esta (vid. García Mouton, P. (1996). “Lenguas en contacto en Vera del Bidasoa”. *RDTP* 51. 209-219, p. 211). Este sentido “contraventanas” no aparece en el *Diccionario vasco-español-francés* de R. M. de Azkue, 2 vols., el t. I publicado en Bilbao; y el II en Bilbao y París: Paul Geuthner [Hay edición facsímil con introducción de L. Michelena, 1984, Bilbao: Euskaltzaindia]), ni lo conocen los vecinos de Vera a quienes he preguntado, ni lo da J. Caro Baroja (*La vida rural en Vera del Bidasoa*, Madrid, 1944, s. v, p.12), mientras que el sabio antropólogo señala que las ventanas se denominan allí *leyuak*, que es la pronunciación navarra y guipuzcoana del vasco *leihoa* (p. 13). En el propio ALEARN encontramos que a la “ventana pequeña” se la denomina *leyotikiya* (VI, mapa 807) y a la “ventana para dar luz al desván”: *leyua* y *arleyua* (VII, mapa 918) y R. M. Castañer Martín (*Estudio del léxico de la casa en Aragón, Navarra y Rioja*. Zaragoza: Diputación General de Aragón, 1990, p. 137) le da el significado de “ventana”.

<sup>62</sup> Mi maestro José Luis Pensado me proporcionó con toda amabilidad un dato de 1418, en el *Libro consistorio de Santiago*: “cuatro *levuas* de cerna”, con una leve diferencia fonética sin importancia, frente a la forma zamorana; el Padre Sarmiento es testigo del uso de *liobas* “duelas” en Lemos (DECH, s. v. *duela*, donde se busca una explicación muy forzada de esta forma; vid. Pensado, J. L. (1973). *Catálogo de voces y frases de la lengua gallega*, de Fr. Martín Sarmiento. Salamanca: Universidad de Salamanca. 347). En la actualidad es voz normal en gallego para “duela” (Lorenzo, X. (1982). *A terra*, Vigo: Galaxia. 101).

<sup>63</sup> *leivas* “duelas” se usa en el portugués minhoto (Tavares, D. A. (1944). *Esboço dum Vocabulario Agrícola Regional*. Lisboa. 473).

<sup>64</sup> Barrios, A. (1981). *Documentación medieval de la Catedral de Ávila*. Salamanca: Ediciones de la Universidad de Salamanca. 276, 289, 299.

<sup>65</sup> Penna, M. (1965). “Traducciones castellanas antiguas de la *Divina Comedia*”. *Revista de la Universidad de Madrid* 14. 126.

<sup>66</sup> Tesauro, P. (ed.) (1983). *Libro de miseria de omne*. Pisa: Giardini Ed. Estr. 425.

De los carpenteros falsos dezirvos he su afar  
 quando los quiebra la lecuá bien la saben rremendar  
 ca la cobren con el çello o la fazen aplanar  
 así que lo non entiende el que la quiere mercar.

Este último ejemplo tomado de un texto en el que se dan rasgos lingüísticos claramente aragoneses<sup>67</sup>, podría ser una pista de la extensión del celtismo a Aragón<sup>68</sup> en la Edad Media; a menos que lo tomemos como un rasgo leonés de la obra. A la primera posibilidad nos anima el derivado *leguado*, que existió no solo en el Centro y Occidente peninsulares (“siete cubas e dos leguados de otras cubas”, en 1415<sup>69</sup> y “arcos, cubas y leguados”, en 1620<sup>70</sup>), sino que llegó también a Aragón, donde lo ha registrado J. Terrado Pablo en 1407: “un *leguado* esbaratado de cuba”<sup>71</sup>; incidentalmente, aunque no es imposible un denominal en *-ado*, no debería dejarse de lado que se tratara de un sufijo átono de los estudiados por Ramón Menéndez Pidal<sup>72</sup>, que daría lugar a una forma como \**léguado* (cf. *lóbado*, *relóbado*, *nuégado*), de donde se explicaría fácilmente el cambio acentual, de un modo particular en los ejemplos aragoneses.

Este viejo vocablo, nacido en el ámbito del leonés y extendido por el castellano e incluso al aragonés, fue sustituido en el siglo XVI por *duela*<sup>73</sup>, aunque también por

<sup>67</sup> Como lo es ese *çello* “arco” que aparece ahí mismo. El DECH (s. v.) proporciona la primera documentación de esta voz en un inventario murciano de 1614. Está en el *Libro de las buenas andanças e fortunas que fizo Lope Garçia de Salazar*, edición de C. Villacorta [tesis doctoral], UPV, 2002, libro XX. *Arco* es normal en el ámbito leonés: aparece no solo en el documento zamorano citado más arriba, sino en una copia de otro de 1270 (Ruiz Asencio, J. M.; Martín Fuertes, J. A. (1994). *Colección documental del archivo de la catedral de León IX (1269-1300)*. León: Centro de Estudios e Investigación “San Isidoro”. 2295).

<sup>68</sup> Hay un par de *leguias* en un documento de Alfaro de 1289 (Menéndez Pidal, R. (1919). *Documentos lingüísticos de España, I: Reino de Castilla*. Madrid: Anejos de la RFE [hay reimpresión en Madrid: CSIC, 1966]. 168), que quizá pudiera tratarse, aunque parece improbable, de *leguas*. Cf. González Ollé, F. (1980). *Lengua y literatura españolas medievales*. Barcelona: Ariel. 514; y Alvar, M. (1976). *El dialecto riojano*. Madrid: Gredos. 70, n. 23b.

<sup>69</sup> “Inventario de los bienes de doña Leonor de la Vega”. En Pérez, R.; Calderón, M. (1983). *El Marqués de Santillana: biografía y documentación*. Santillana del Mar. 164.

<sup>70</sup> “Ordenanzas de Miranda del Castañar”. En Álvarez Villar, J. (1980). *La villa ciudad de Miranda del Castañar*. Salamanca: Centro de Estudios Salmantinos. 3ª ed. 123.

<sup>71</sup> Terrado Pablo, J. (1991). *La lengua de Teruel a fines de la Edad Media*. Teruel: Instituto de Estudios Turolenses. 277; se relaciona ahí este *leguado* con el ant. *legar* “liar”, lo que condiciona su definición: “liaza, conjunto de mimbres con que se forman aros para los toneles y cubas”.

<sup>72</sup> Menéndez Pidal, R. (1905). “Sufijos átonos en español”. En *Bausteine für romanischen Philologie*. Halle: Niemeyer. 386-400, p. 396. Vid. también Craddock, J. R. (1972). “Las categorías derivacionales de los sufijos átonos: *pícaro*, *páparo* y afines”. En *Studia Spanica in Honores R. Lapesa III*. Madrid: Gredos. 219-231.

<sup>73</sup> DECH, s. v.; si bien un Bernaldus *Doela* aparece como testigo en un documento leonés de 1228 en la *Colección diplomática del monasterio de Sahagún (857-1300), V (1200-1300)*. Fernández Flórez, J. A. (ed.) (1994). León: Centro de Estudios e Investigación San Isidoro. § 1646.

*bastos* (“siete bastos de cuba, tres de roble Y cuatro de pino”, “tres bastos de cubas roblizas” en documentos castellanos del siglo XIV<sup>74</sup>), *costera* (citado antes en la trad. italiana del canto I de la *Commedia*), *casco* (en un documento zamorano de 1446: “dos cascos de cubas”<sup>75</sup>) o *tabla* (que recuerdo haber encontrado en documentos antiguos de Valladolid).

Los datos con los que cuento para *legua* y *leguado* harían pensar que estas palabras desaparecieron de repente y que mucho después, también de la noche a la mañana, volvieron a salir a la superficie. J. de Lamano encontró *legua* en la Ribera salmantina<sup>76</sup> y en el ALCL<sup>77</sup> (II, mapa 342) aparece en varios pueblos de Salamanca (puntos 502, 602, 102; como *lengua* en el 202), en uno zamorano (punto 500; lo he oído también en el pueblo zamorano de Villalcampo), en dos de Palencia (puntos 101, 102: en ambos como *leba*), en uno de Valladolid (punto 600: como *leba*), en uno de Ávila (punto 400: como *lengua*), en dos de Segovia (puntos 300 y 301: en ambos como *lengua*; y en el punto 302 de esa misma provincia como *leguao*). No tengo noticia de que esta palabra se extienda en la actualidad al sur de Salamanca: no la registran ni M. A. Marcos<sup>78</sup> ni A. Viudas<sup>79</sup>, si bien A. Llorente la tenía por viva en el occidente extremeño<sup>80</sup>; tampoco lo encontraron en Andalucía los encuestadores del ALEA<sup>81</sup>. Aparece, en cambio, en el área aragonesa, donde registra *legua* A. Llorente Maldonado, en la zona limítrofe con Soria y Guadalajara y en Hecho<sup>82</sup>, así como *leguao* en tres pueblos situados al S. O. de Zaragoza<sup>83</sup>.

#### 4.2. Barajones

Puede la discontinuidad estar relacionada con el léxico particular de una realidad distante para una gran parte de los hablantes de una lengua, tal y como voy a ejemplificar con la voz *barajones*, voz que vuelve a situarse entre la vestimenta para afrontar la nieve. Media un largo espacio de tiempo entre sus primeras documentaciones y las modernas: se registra por primera vez, en la forma *baraliones*,

<sup>74</sup> Castro Toledo, J. (ed.) (1981). *Índices de documentos de la Colección Diplomática de Tordesillas*. Valladolid: Inst. Cult. Simancas. 160, 161, 179, 197

<sup>75</sup> Vaca Lorenzo, A. (1988). *Documentación medieval del archivo parroquial de Villalpando (Zamora)*. Salamanca: Ediciones de la Universidad de Salamanca. 186.

<sup>76</sup> de Lamano, J. (1915). *Dialecto vulgar salmantino*. Salamanca, s. v. [Reimpresión publicada en Salamanca: Ediciones de la Diputación de Salamanca, 1989].

<sup>77</sup> Alvar, M. (1999). *Atlas lingüístico de Castilla y León*. 3 vols. Valladolid: Junta de Castilla y León.

<sup>78</sup> Marcos, M. A. (1979). *El habla de Béjar*. Salamanca: Centro de Estudios Salmantinos.

<sup>79</sup> Viudas, A. (1980). *Diccionario Extremeño*, Cáceres: Universidad de Extremadura.

<sup>80</sup> Llorente Maldonado, A. *Op. cit.* 331.

<sup>81</sup> Alvar, M.; con la colaboración de A. Llorente Maldonado y G. Salvador (1961). *Atlas Lingüístico y Etnográfico de Andalucía*. T. I, mapa 214: *Duela, tabla y costilla*.

<sup>82</sup> Llorente Maldonado, A. *Op. cit.* pp. 331, 332, con datos tomados del ALEARN, II mapa 208, Z 503, Z 505, Z 506, Hu 102.

<sup>83</sup> *Id.*, *ibid.*: considera *leguao* “derivado evidente de *legua*”, pero lo explica —no logro saber por qué— por etimología popular.

en un texto bajolatino español de 1236<sup>84</sup>; luego en el *Cronicon Mundi*, en latín, de Lucas de Tuy (a1239) aparecen unos “rusticorum calciamenta que uulgariter incole auarcas et baraliones uocant”<sup>85</sup>, que en la traducción castellana de esa obra, de la 2ª mitad del siglo XV, son “abarcas y uarallones”<sup>86</sup>. Un larguísimo silencio separa estos datos de su empleo por J. M<sup>a</sup> de Pereda, quien escribe “Quitóse los barajones en un periquete”<sup>87</sup>, y las referencias modernas de su uso en Álava, Santander, Asturias y León<sup>88</sup>. Silencio que ni M. Fernández de Enciso, a principios del siglo XVI, ni A. de Torquemada, un poco después, supieron romper cuando, refiriéndose a esos artilugios que se emplean para andar por la nieve, no tuvieron más remedio que acudir a una voz de sentido muy general para designarlos; no porque no existiera el objeto en ese momento en nuestro país, sino porque estaría restringido a aquellos pocos y perdidos lugares en que había que caminar a menudo por la nieve:

[los albanos] En invierno ninguno puede subir a [los montes Cáucacos], y en verano suben pocas veces, y los que suben pónense en cierta forma unas tablas en los pies, sobre que van por encima de la nieve; y otro tanto hacen los iberos para subir en las sierras de Armenia.<sup>89</sup>

Estos ponen en los pies unas tablas anchas como un palmo, o poco más, y de las puntas sale un báculo encorvado para arriba que toman con las manos, y todo ello aforrado o cubierto de unas pieles de animales que llaman rangíferos, y con esto caminan de cierta forma encima de las nieves sin hundirse.<sup>90</sup>

### 4.3. Estoyo ~ estoxo

A la falta de datos del pasado puede añadirse el hecho de que los espacios lingüísticos no permanecen inalterables a lo largo de la historia, como vemos con la voz *estoyo* ‘estuche’ (y su variante *estojo*), palabra de transmisión popular procedente del latín *STUDIUM*, y documentada en los siglos XIV y XV. En la progresiva y permanente castellanización que sufrió el espacio leonés, *estoyo* se sustituyó por *estuche*; este último un cognado del dialecto vecino castellano, donde había penetrado no directamente desde el latín, sino a través del occitano. Aquel *estoyo* del leonés medieval que, atendiendo a nuestra documentación, no vuelve a aparecer en un larguísimo espacio de tiempo, reaparece en la actualidad solo en asturiano,

<sup>84</sup> DECH, s. v.

<sup>85</sup> Falque, E. (2003). “La inserción del romance en los textos históricos latinos medievales”. En Perdiguero, H. (ed.). *Lengua romance en textos latinos de la edad Media. sobre los orígenes del castellano escrito*. Burgos: Instituto Castellano de la Lengua. 71-79, p. 74.

<sup>86</sup> Puyol, J. (ed.) (1926). *Traducción de la Crónica de España*, por Lucas, obispo de Tuy. Madrid. 339.

<sup>87</sup> Bonet, L. (ed.) (2006). J. M<sup>a</sup> de Pereda, *Peñas arriba* [1895]. Barcelona: Galaxia Gutemberg. 197, 288.

<sup>88</sup> DECH, s. v. En el DHLE se da entrada al dato moderno, tomado de García Lomas, *abarajonar* “enredarse al andar con las abarcas llamadas barajones”.

<sup>89</sup> M. Fernández de Enciso, *Suma de Geografía*, Madrid, 1948, p. 113.

<sup>90</sup> A. de Torquemada. *Op. cit.*, p. 863.

como resto más fuerte del antiguo leonés, en las formas *estoyo* ~ *estoxo*, referidas a distintos tipos de cajoncitos.<sup>91</sup>

## 5. Los saltos en la documentación en los momentos finales en que desaparece una voz

Vamos a referirnos a una situación en parte diferente, originada por la merma en el uso de las palabras cuando estas están a punto de desaparecer.

### 5.1. Decibir

Es el caso de *decibir*, perteneciente sobre todo al léxico jurídico, propio del aragonés<sup>92</sup> y de su “área lingüística contigua”<sup>93</sup>, que aparece también al otro lado del castellano, en el occidente peninsular: en gallego<sup>94</sup> y en la que también podríamos llamar su “área lingüística contigua”, es decir el portugués<sup>95</sup> y posiblemente el leonés<sup>96</sup>. Si esta

<sup>91</sup> Canellada, M. J. (1944). *El bable de Cabranes*. Madrid: Anejo XXXI de la RFE, vocabulario, s. v. Menéndez García, M. (1965). *El Cuarto de los Valles*. Oviedo: Instituto de Estudios Asturianos. Vol. II, s. v., Cano, A. M. (1981). *El habla de Somiedo (Occidente de Asturias)*, separata de los n.ºs. 4 y 5 de *Verba*. Santiago de Compostela. 77; y de la misma autora *Vocabulario del bable de Somiedo*, Oviedo: Instituto de Estudios Asturianos, 1982, s. v. Existiendo en la actualidad *estojár* “esponjarse” y *estojado* “estar esponjado, gordo” en Cáceres y Salamanca, es de prever que *estojó* ~ *estoyo* hayan pervivido en el ámbito rural de estas provincias hasta tiempos recientes.

<sup>92</sup> A los datos que proporciono en mi trabajo “Los aragonesismos”, cit., pp. 671 y 672, añádase: Lagüens Gracia, V. (1992). *Léxico jurídico en documentos notariales aragoneses de la Edad media (siglos XIV y XV)*. Zaragoza: Diputación General de Aragón. 98.

<sup>93</sup> En el significado con que emplea esta etiqueta, para el léxico de un área nordoriental de la Península Ibérica, Joan Veny, 1991, “Huellas aragonesas en los dialectos catalanes medievales”, *Actas del congreso de lingüistas aragoneses*, Zaragoza, p. 84-102, y los trabajos allí citados de B. Pottier, G. Colón, P. Bec y J. A. Frago. Está en occitano (Levy, E. (1894-1924). *Provenzalisches Supplement-Wörterbuch*. 8 vols. Leipzig: O. R. Reisland. T. II, p. 25, s. v. *decebre, decebemen*) y catalán (Alcover, A.; Moll, F. de B. (1926-1968). *Diccionari català – valencià – balear*. 10 vols. Palma de Mallorca, s. v. *decebre*; DECH s. v. *concebir*; Corominas, J. con la colaboración de J. Gulsoy y M. Cahner (1981). *Diccionari etimològic i complementari de la llengua catalana*. Barcelona: Curial. T. II, s. v. *concebre*).

<sup>94</sup> En la documentación de Noia editada por M. del C. Barreiro (*A documentación notarial do concello de Noia (séculos XIV-XVI)*. *Lectura, edición e léxico*, [tesis doctoral], Santiago, 1994), se repite hasta la saciedad la fórmula: “constrengudo per força nen deçebudo per engano” (docs. § 8.8., 1381; § 9.6, 1385; § 16.5, 1397; § 17.6, 1397; § 19.5, 1398; § 20.5, 1403; § 21.5, 1403; § 25.5, 1409; § 29.14, 1415; § 34.6, 1422. En alguna ocasión el escribano se equivoca y, en lugar de *deçebudo per engano*, escribe *reçebido per engano* (§ 13.6, 1395). Vid. más datos en Pascual, J. A. “Los aragonesismos”, cit. p. 672, n. 106.

<sup>95</sup> Vid. Pascual, J. A. “Los aragonesismos”, cit., p. 673, n. 107.

<sup>96</sup> Además de los datos leoneses que muestro en “Los aragonesismos”, cit., p. 672, n. 106, *decibir* aparece en un sínodo leonés de 1318 (García y García, A. (1984). *Synodicon Hispanum, III: Astorga, León y Oviedo*. Madrid: BAC. 290). *Deçibimiento* está en textos que contiene rasgos leoneses, como el *Fuero Juzgo* (así en el código murciano, Perona, J. (ed.) (2002). *El Fuero Juzgo. Estudios críticos y transcripción*. Murcia: Región de Murcia, Consejería de Educación y Cultura. p. 189, lectura que presenta también el texto editado por la Real Academia Española del *Fuero Juzgo en latín y castellano cotejado con los más antiguos y preciosos códices*. Madrid: Ibarra, 1815. Vid. Fernández Llera, V. (1929). *Gramática y vocabulario del Fuero Juzgo*. Madrid, vocabulario, s. v.).

voz forma parte en la Edad Media del léxico culto de los dialectos hispánicos que rodeaban al castellano, es posible que existiera también en este último dialecto, por más que los datos con que cuento no sean realmente significativos: está en los textos de Berceo<sup>97</sup>, pero cuando el testimonio de una palabra marcada se reduce al empleo que hace de ella el poeta riojano, no es este un indicio decisivo de lo castellano; y no sirve, por otro lado, de apoyo a esa posibilidad su aparición en el *Alexandre*, pues si bien encontramos *decibido* en esta obra, eso ocurre en el ms. P, aragonés, es un caso el que el manuscrito O, leonés, lleva *destendido*<sup>98</sup>.

Termina esta palabra por esfumarse, pero en distintos momentos, según los dialectos: en aragonés lo utiliza aún a mediados del siglo XVI el tudelano Jerónimo de Arbolanche, aunque como arcaísmo, según Fernando González Ollé<sup>99</sup>. En el Occidente peninsular debió desaparecer o estar muy cerca de su desaparición ya en el siglo XV, pues en las copias del siglo XV de los documentos gallegos citados en nota 94 se sustituye *decebudo* por *induzido*<sup>100</sup>, igual que en el documento leonés de 1318 ya citado, al copiarse en el siglo XV, se cambia *descibir* en *rescibir*<sup>101</sup>. Si hubiéramos aceptado el testimonio de Berceo como castellano, en este dialecto habría decaído este verbo aun antes, pues en los manuscritos berceanos del siglo XIV —que modernizan y castellanizan su lengua— se evita *deçibido* y se adopta en su lugar *engañado*<sup>102</sup>.

Al no tener la seguridad de que fuera esta una palabra castellana en el siglo XIII no nos atrevemos a decidir si en este caso su sustitución en los códices de Berceo del siglo XIV supone una modernización o una castellanización; y, sin embargo, volvemos a encontrar *decibir* en el siglo XV, en escritores que escriben en una variedad que nadie dudaría en considerar netamente castellana, pero que se complacen de vez en cuando en introducir aragonesismos, como hacen el Marqués de Villena, el Marqués de Santillana<sup>103</sup>, Gómez Manrique, y Fernán Pérez de

<sup>97</sup> Solalinde, A. G. (ed.) (1922). de Berceo, G. *Milagros de Nuestra Señora*. Madrid: Espasa Calpe, estrofas, 15 c y 558 a; Dutton, B. (ed.) (1967). *Vida de San Millán de la Cogolla*. London: Tamesis Books, estrofa 111 a; *El duelo que fizo la Virgen María el día de la pasión de su fijo Jesucristo*. En Dutton, B. (ed.) (1975): de Berceo, G. *Obras completas*. T. III. London: Tamesis Books, estrofa 83 c.

<sup>98</sup> Se trata de la estrofa 2038 del *Alexandre*: “Átalus redor sí mandó fer un roído / cuidó que eran velas, fue Poro decebido / metió se en las naves del rey percibido, / ovo en poca d’ora [e]l [Idaspis] trocido”. Nelson, D. A. (1979). *Gonzalo de Berceo, El libro de Alexandre*. Madrid: Gredos. Vid. Keller, J. (1932). *Contribución al vocabulario del Poema de Alixandre*. Madrid, s. v.

<sup>99</sup> González Ollé, F. (ed.) (1969-1972). J de Arbolanche, *Las abidas*. 2 vols. Madrid: CSIC, glosario.

<sup>100</sup> Barreiro, M. del C. (1476, 1493, 1494). *A documentación*, cit., § 55.6, 1476; § 64.4, 1493; § 65.7, 1494.

<sup>101</sup> García y García, A. “Engaños que disen por rescibir las gentes”. *Synodicon*, cit., 290.

<sup>102</sup> Marden, C. (1928). *Cuatro poemas de Berceo*. Madrid. 29.

<sup>103</sup> Emplea la voz en la *Comedieta de Ponza*: “*desciben* las aves” (18 b ), lectura que da p. ej., el ms. 2763 de la Biblioteca de la Universidad de Salamanca, fº. 96 rº, aunque otras textos —entre ellos los que M. Durán toma como base de su edición de las *Poesías*, del Marqués de Santillana, Madrid: Castalia, 1975, p. 249— presentan la *lectio facilior* “*descienden* las aves”.

Guzmán (quien emplea el part. *decepta*)<sup>104</sup>. ¿Se trata de aragonesismos? ¿Estamos ante arcaísmos castellanos? ¿Podría tratarse de ambas cosas a la vez? Lo que es claro es que estas documentaciones que surgen ahora aisladas en siglo XV y tan distantes del *decebimiento* que aparece a principios del XVIII, en el “Suplemento” al *Tesoro* de Covarrubias<sup>105</sup>, explicado allí como “vocablo antiguo castellano”, son los estertores de una voz que lleva camino de desaparecer del español.

## 5.2. Caler

*Caler* tiene un comportamiento parecido a *decibir*, aunque mantenga algunas pequeñas diferencias con esta voz. En efecto, al igual que esta última —al menos en una de las interpretaciones que le dábamos— apareciendo en los dialectos hispánicos, se pierde su pista, primero en castellano, donde vuelve a dejarse ver en el siglo XV como un arcaísmo<sup>106</sup>; se mantuvo un poco más en leonés<sup>107</sup>; y ha llegado hasta la actualidad en aragonés<sup>108</sup>. Pero aparecen aquí problemas filológicos que complican las cosas, en esos momentos en que se percibe con claridad el abandono de la palabra: de los manuscritos en que se nos conserva la *Visión Deleytable*<sup>109</sup>, los más castellanizantes mantienen *caler* coincidiendo con el incunable tolosano de 1489 —muy aragonesizado—, mientras que un manuscrito aragonesizante como el ms. 3367 de la Biblioteca Nacional de Madrid cambia —aunque no sea la única disidencia de este tipo— *caler* en *cumplir*. Si no resulta fácil dar con la razón de esto, tampoco lo es valorar su reaparición en el *Viaje de Turquía*<sup>110</sup>, aunque sí luego en Lope de Vega, pues lo ha exhumado como un arcaísmo leonés<sup>111</sup>.

<sup>104</sup> Vid. Pascual, J. A. “Los aragonesismos”, cit., p. 673 y 674, n. 107.

<sup>105</sup> Arellano, I.; Zafra, R. (eds.) (2006). S. de Covarrubias. *Tesoro de la lengua castellana o española*. Madrid: Editorial Iberoamericana, s. v.

<sup>106</sup> Vid. Pascual, J. A. *La traducción*, cit., p. 133.

<sup>107</sup> Mejor, en los restos del leonés que reflejan los textos hipercharacterizados dialectalmente, como los de Juan del Enzina, vid. Pascual, J. A. “Los aragonesismos”, cit., pp. 657, 658 y cf. Pascual, J. A. (1993). “La edición crítica de los textos del Siglo de Oro: de nuevo sobre su modernización gráfica”. En García Martín, M. (ed.). *Estado actual de los estudios sobre el Siglo de Oro*. Salamanca: Ediciones Universidad de Salamanca. 37-57, p. 48.

<sup>108</sup> A los datos aragoneses y de escritores castellanos aragonesizantes del siglo XV que doy en *La traducción*, cit., p. 36 y en *Los aragonesismos*, cit., p. 133, añádase su aparición en el incunable navarro de Esteban de las *Regulae* de Masparrauta, de 1492: *caler*: “conuenir o caler o complir: oportet”, f<sup>o</sup>. 48.

<sup>109</sup> Vid. la apreciación que se hace sobre algunos códices de esta obra en J. A. Pascual, “Los aragonesismos”, cit., p. 648.

<sup>110</sup> “non cale irme a la mano”. García Salinero, F. (ed.) (1980). *Viaje de Turquía*. Madrid: Cátedra, p. 111.

<sup>111</sup> Zamora Vicente, A. (1983). “Sobre la fabla antigua de Lope de Vega”. En *Philologica hispaniensa in honorem Manuel Alvar* I. Madrid: Gredos. 645-649, p. 649.

## 6. Conclusión

Los huecos que se perciben en la historia de una palabra —igual que en la de una acepción— nos lleva a quienes estamos redactando la Planta del *Nuevo diccionario histórico del español* a organizar la presentación de los ejemplos del diccionario de una forma que permita ver, por un lado, la provisionalidad de nuestros datos, y sirva, por otro, para superar de momento esta provisionalidad con hipótesis razonables.

## Lexical Patterns: From Hornby to Hunston and Beyond

Patrick Hanks  
Masaryk University

*I start with a brief summary of A. S. Hornby's achievement in creating the Idiomatic and Syntactic English Dictionary (ISED 1942), a work which gradually mutated, through many editions, into the present Oxford Advanced Learner's Dictionary of Current English. Among Hornby's radical innovations was a focus on examining the patterned nature of language and presenting patterns of word use in a succinct form for assimilation by language learners. He saw that each verb is associated with a different set of syntactic patterns, and he was able to impose order on apparent chaos by picking out structural threads and establishing templates for pattern analyses. The 5th edition of OALD, edited by Jonathan Crowther (1995) and the 6th edition, edited by Sally Wehmeier (2000) were recensions of Hornby's work using corpus evidence.*

*Hornby and his mentor, H. E. Palmer, had an intuitive understanding of the patterned nature of language, but they lacked the evidence that was necessary for a detailed empirical study of the collocational patterns associated with different meanings of each word. This had to wait until the advent of very large corpora, inspired in particular by the vision and practice of J. M. Sinclair. As early as 1966, Sinclair predicted that patterns of lexis "would not yield to anything less than a very large computer". Much of his life's work was devoted to developing sound linguistic theory on the basis of empirical analysis of corpus evidence. His principles were taken up by subsequent linguists, for example Alan Partington, Michael Hoey, Susan Hunston, and Gill Francis.*

*In his 1987 paper entitled "The nature of the evidence", Sinclair stresses the importance of distinguishing significant collocations from random co-occurrences. The first attempt to undertake statistical analysis of collocations in a corpus was by Church and Hanks (1990), but it was not until Kilgarriff, Rychlý, and their colleagues developed the Word Sketch Engine (Kilgarriff et al. 2004) that a user-friendly tool was made widely available for people to see at a glance how the meanings of a semantically complex word are associated with and indeed activated by its collocates.*

*Modern corpus tools such as these bring us full circle, back to Hornby's original vision of patterns of word use and word meaning. It is now possible to examine that vision in the light of massive bodies of evidence. Not only does this lead inexorably to new theoretical insights into the nature of language, it also make it possible to develop new kinds of dictionaries for human learners and computational applications alike—dictionaries that focus rigorously on patterns of word use, rather than (say) on historical semantics and morphology.*

*In the second part of the lecture, I give a progress report on the corpus-driven Dictionary of English Verb Patterns currently being developed at the Masaryk University in Brno. I compare the Pattern Dictionary with Pattern Grammar and discuss some problems of lexicographical analysis, such as finding the right level of generalization for each element in a pattern. How is one sense to be distinguished from another? For a word with many patterns, are some patterns more important than others, and if so why? How are creative uses of a word distinguished from more mundane uses? What is the role, in pattern analysis, of an ontology?*

### **1. A. S. Hornby and English lexicography in the 1930s**

In 1923 a shy young man of 25 called Albert Sydney Hornby (known affectionately to his friends and colleagues as “Ash”), armed with a degree in English from University College, London, sailed to far-away Japan to start a career as a teacher of English. This event was to have far-reaching consequences for English lexicography. Hornby proved to be a gifted and skilled teacher, who had sound theoretical instincts and a motivation to explain the meaning and use of English words in terms that ordinary students could understand and assimilate. In 1931 he was invited by Harold E. Palmer, director of the Tokyo Institute for Research into English Teaching, to participate in a programme of vocabulary research. Five years later, in 1936, when Palmer left Japan, Hornby was appointed head of research at the Institute.

At the institute in Tokyo, Hornby compiled lists of important collocations in English, using his wide reading and his intuitions as a teacher of English. He worked with Palmer on English verb syntax and on vocabulary selection for learners at different levels. At least three of the insights of Palmer, Hornby, and their colleagues in the 1930s have provided a principled foundation for much subsequent work, including modern corpus-driven lexicography (which, it should be said, is still in its infancy today). These three principles may be summarized as follows:

1. Language in use is highly patterned. Each word is typically associated with only a small number of syntactic patterns.
2. Ordinary everyday communication consists of utterances based on patterns of usage built up around a small number of very frequent words, each of which is used in a comparatively small number of patterns or structures. At the same time, usage also encompasses a very large number of other possible and actual words and structures, some of which are used only very rarely.
3. The verb is the pivot of the clause. In the front matter of OALD, Hornby asserts: “Verb patterns are the most important”, and urges learners to “spend a few hours studying ... verb patterns”, as “the ordinary grammar-book and dictionary usually fail to supply adequate information on such points.”

Hornby and his colleagues were well aware that the English monolingual dictionaries available in the 1920s and 30s did not take account of these principles. In fact, those dictionaries were quite unsuitable for pedagogical purposes. The focus in those days—for example in Fowler’s brilliant *Concise Oxford Dictionary* (COD) of 1911, the best-selling English dictionary of its time, between the wars—was on historical philology. From its first publication, the full title of COD was *The Concise Oxford Dictionary of Current English*, but it was not until the 8th edition (1990, edited by Robert Allen) that COD really earned that subtitle, and even then it gave no account of structured patterns of usage. In the first seven editions of COD, prominence was given to word history and etymology, despite the subtitle. The oldest known meaning of each word was placed first, provided only that it was still current<sup>1</sup>. For example, the noun *carnation*, denoting a kind of sweet-smelling flower, was nested under the adjective *carnal*, meaning ‘of or pertaining to flesh’, because both were thought to be derived from Latin *carne* ‘flesh’. In those days, etymologists thought that the flower was so named because of its fleshy pink colour.<sup>2</sup> The first sense of the noun *camera* was given as ‘a small vaulted room’, not an apparatus for taking photographs, even though photography was already a well-established technology in Fowler’s day and the ‘small room’ sense of *camera* was already rare or obsolete. This quaint editorial policy, inherited from the great historical lexicographical enterprises of the 18th and 19th centuries, was and is associated with some curious value judgements about the nature of language and word meaning, for example:

1. that older meanings are somehow better than modern meanings,
2. that the language of our forebears is somehow better than our own, and
3. that “the language (whatever language it may be: English, French, Spanish, Catalan, Latin, Greek, or other) is going to the dogs”.

The belief that the language is going to the dogs has been around at least since the 5th century BC (it was satirized in ribald terms by Aristophanes), and it continues to be reflected in at least two of America’s best-selling dictionaries, published under the name of America’s first and most belligerent lexicographer, Noah Webster.

As a practical teacher of English, Hornby recognized that historical principles of lexicography are irrelevant to effective language learning and that learners need a dictionary offering practical rules and models of current usage on which to build their own competence. Unlike many of his contemporaries, he decided to do something about it. With the aid of two colleagues, Edward Gatenby and Harold Wakefield, he set about compiling an *Idiomatic and Syntactic English Dictionary*

---

<sup>1</sup> An exception to the historical order of senses in early editions of COD was that senses that had become totally obsolete were relegated to last position and labelled “Obs.”. This was no doubt the justification for the subtitle.

<sup>2</sup> Modern etymologists think that *carnation* is more probably an alteration, by folk etymology, of Arabic *qaranful* ‘clove or clove pink’, from Greek *karyophyllon*.

(ISED). This was completed in 1941 and published by Kaitakusha in 1942.<sup>3</sup> It was the first dictionary of English as a foreign language, initiating a genre that has evolved into a rich variety of present-day forms. In 1948, it was re-published unaltered by Oxford University Press under the new title *A Learner's Dictionary of English*. It became an international and perennial best-seller. In the third edition (1974), the name *Oxford* was added to the title.

One of the pleasures of preparing this lecture was revisiting the first edition of Hornby, Gatenby, and Wakefield and discovering for how fresh, clear, readable, and easy to understand that first edition was. In the second (1963) and subsequent editions, Hornby lost some of that freshness and ease of use, apparently under the influence of the then current 4th edition (1951) of the *Concise Oxford Dictionary*. He made the following changes among others:

- Putting in many thousands of additional entries and subentries, greatly increasing coverage.
- Nesting subentries under root words, e.g. he moved *blackbird* and *blackboard* and nested them under *black*.
- Using a swung dash to represent repetition of the headword within an entry, so that *blackbird* and *blackboard* are represented as *~bird* and *~board*.
- Rewriting definitions in a more formal style, apparently in pursuit of the principles of consistency and substitutability. Many of ISED's glossed examples became formal definitions. So, for example, ISED had glossed examples like this, one of several under *blame*:
  - o *He blamed his failure on the teacher [blamed the teacher for his failure]*  
(= he said that it was the teacher's fault)

In OALD2 this was swept away and subsumed with other examples under a general definition of the verb:

- o find fault with; fix the responsibility on (sb. or sth.) (*for* sth.): *Bad workmen often ~ their tools. He ~d the teacher for his failure.* (Colloq.)  
*He ~d the teacher for his failure.*

It is not clear that these 1963 changes were entirely beneficial. The increase in coverage undoubtedly gave the dictionary more potential usefulness as an aid for decoding tasks (i.e. for reading and understanding), but it reduced its usefulness as an encoding aid (for writing and speaking idiomatically) by making it harder for

---

<sup>3</sup> The first edition is still available on Kaitakusha's website, with a blurb written over 66 years ago: "This dictionary has been compiled to meet the needs of foreign students of English. It is called *Idiomatic and Syntactic* because the compilers have made it their aim to give as much useful information as possible concerning idioms and syntax. It is hoped that the dictionary will be of value to those who are learning English as a foreign language." The publisher might care to note in some future version of this blurb that, during the intervening two-thirds of a century, those hopes have been amply fulfilled.

learners to find what they are looking for. This adverse effect was compounded by the policy of nesting and the use of the swung dash. The swung dash undoubtedly saved some space, but it made words much harder to recognize and may well have baffled some learners. At any rate, thirty years later, in the 5th edition (1995), OALD abandoned the swung dash, and in the 6th edition (2000) the policy of nesting was likewise abandoned: *blackbird* and *blackboard*, along with thousands of other compounds, were restored to the headword status that Hornby and co. had originally given them.

In 1942, shortly after the outbreak of war and shortly before publication of their dictionary, the three lexicographers left Japan under a programme for the exchange of enemy nationals. Both Hornby and Gatenby went on to distinguished careers in the British Council.

In 1954, Hornby published *A Guide to Patterns and Usage in English*, a lexically based practical and partial grammar whose approach to analysis and tabular style of presentation reflected the methods employed in the *Advanced Learner's Dictionary*.

The *Advanced Learner's Dictionary* went through several editions under the editorship of Hornby and, subsequently, some other able lexicographers, including Tony Cowie and Jonathan Crowther. A much appreciated feature of the early editions was the guidance given on grammar and usage, on principles that had been devised by Hornby and Palmer. The 5th edition, edited by Jonathan Crowther (1995) and the 6th edition, edited by Sally Wehmeier (2000), were radical recensions of the work of Hornby and other previous editors in the light of corpus evidence. Hornby's name continues to grace the title page of the current (7th) edition of OALD, although his co-workers have been consigned to oblivion. Hornby himself acknowledged what most lexicographers know but the public perhaps do not, namely that lexicography is a team game. He was always careful to pay tribute to the work of his original collaborators (Gatenby and Wakefield) and their successors.

## 2. Clause roles, Hornby's Verb Patterns, and OALD

Hornby's recognition of the patterned nature of usage and the central importance for language learners of knowing the syntagmatics of verbs led him to formulate a summary of English verb patterns. He frequently drew attention to the danger for foreign learners of false analogy, for example, forming an ungrammatical sentence such as “\*I proposed him to come”, either by false analogy with a verb pattern in their own native language or by false analogy with well-formed grammatical sentences in English such as “I asked him to come” and “I told him to come”. It is important, therefore, for learners to know, not only what verbs means, but also how to use them. Hornby believed that this could be achieved if learners would spend a few hours memorizing the verb patterns, so that, for example, before using the verb *propose* in an essay, they could look it up in OALD, see that it is used in verb pattern 9, and thereby know that the correct idiomatic phraseology is “I proposed that he should come”, not “\*I proposed him to come”.

This was done systematically for each verb in the dictionary, by stating a pattern number in square brackets alongside definitions and examples, e.g. [VP9]. The pattern numbers refer to a look-up table in the front matter of the dictionary. In the 1974 edition of ALD, for example, VP9 is “S + *vt* + *that*-clause” (as in “*I proposed that he should come*”). This contrasts with VP10, “S + *vt* + dependent clause/question” (as in “*She asked whether he would come*”) and VP11, “S + *vt* + noun/pronoun+ *that*-clause” (as in “*I told him that he should come*”).

It would difficult to understate the importance of this insight from the point of view of lexical and grammatical theory. As we shall see, it plays a central role in corpus-based research into the relationship between meaning and use.

A word is needed here on clause roles. In 1963 Hornby did not consider the subject of the clause to be part of the pattern. In 1974, he and Tony Cowie did. This, in my opinion, was a step in the right direction, from which, unfortunately, EFL lexicographers have since backed away. Its importance only becomes apparent when an attempt is made to assign semantic values to clause roles, in order to distinguish one sense of a verb from another. The subject of a clause is part of its pattern. For the vast majority of clauses, the subject has the default semantic value [[Human]]. These constitute the unmarked cases. More interesting are the marked cases, where the semantic value of the subject is not [[Human]]. Examples are: *one administrative entity swallowing up another administrative entity* (distinct from *a human swallowing a physical object*), or *an ideology firing people with enthusiasm* (distinct from *a human firing people from their jobs*). The semantic relationship between a verb and the rest of the clause is a relationship among clause roles, not merely between the verb and various nouns, adjectives, or prepositions. This is a fine point, but failure to take note of it has sometimes led to confusion in lexical analysis.

The terminology of generative linguistics, which makes binary divisions and refers to the subject as the “external argument”, lumping adverbials and objects together as part of the “verb phrase” (which, in more esoteric terminology, is sometimes called the “inflection phrase”), is unhelpful in this respect. Empirical linguists such as Quirk and (with minor variations) Biber, Sinclair, Halliday and others, recognize five clause roles, in an analytic structure which has come to be known informally by the mnemonic SPOCA. Since grammar these days is a hotbed of terminological confusion, with considerable potential for misunderstanding, it is worth taking a few moments to summarize the five basic clause roles. They have a central part to play in verb pattern analysis. They are:

- |   |            |   |
|---|------------|---|
| S | Subject    | Obligatory in English, except in a) imperatives (e.g. “ <i>eat your greens!</i> ”) and b) elliptical clauses such as natural responses, e.g. “ <i>What was he doing?</i> ” — “ <i>eating an apple</i> ” |
| P | Predicator | The verb group, including auxiliary verbs and negatives, but not nouns or prepositional phrases. By some writers, V (verb) is used in place of P.   |

- O Object English clauses may have one, two, or no objects. In SPOCA, the object is a clause role in its own right, not part of the verb phrase, as it is in grammars based on predicate logic
- C Complement A clause role that is co-referential with either the subject or the object of the clause. Examples of subject complements are the adjective *happy* in *he seems happy* and the noun phrase *the pronunciation editor* in *Carolyn was the pronunciation editor*. Examples of object complements are as in *She made him happy; they appointed her pronunciation editor*.
- A Adverbial (sometimes called Adjunct). A clause may have any number of adverbials. A distinction is made between obligatory adverbials (for example, the locative adverbial *on the table* in *He put the cup on the table*) and optional adverbials (for example, the time adverbial in *He died in 1974*).

Translated into SPOCA, Hornby's 1963 patterns look like this:

1. S P O. *We lit a fire.*
2. S P {to/INF}. *He wants to go.*
3. S O {to/INF}. *They want him to go.*
4. S P O (to be) SC. *I consider it (to be) a shame.*
5. S P INF. *I made him do it. Will you help me carry this box?*
6. S P O {V-ing}. *He kept me waiting. I saw him running off.*
7. S P O OC (adj.). *Don't get your clothes dirty.*
8. S P O OC (noun). *They elected him president.*
9. S P O {V-en}. *He got the document printed. I have never heard Italian spoken.*
10. S P O A. *He took his hat off. Mr Smith showed me to the door.*
11. S P {(that) CL}. *I suppose (that) he will be late.*
12. S P O {(that) CL}. *I warned you that he would be late.*
13. S P {Wh- CL}. *I know why he did it.*
14. S P O {Wh- to/INF}. *We showed him how to do it. They told him when to start.*
15. S P {Wh- CL}. *I wonder what it is. I don't mind where we go.*
16. S P O {Wh- CL}. *Tell me what it is. Ask him where he put it.*
17. S P {V-ing}. (A) *They stopped talking.* [Compare *They stopped to talk*—different meaning] (B) *He began talking.* [Compare *He began to talk*—same meaning] (C) *It needs doing.* [Compare *it needs to be done*—passive meaning]

- 
18. S P O A. (A: with *to*, alternating with 19): *He gave some money to his wife.* (B: with *for*, alternating with 19): *He bought a watch for his wife.* (C, with other prepositions, not alternating with 19): *They criticized him for being late. He was throwing stones at a dog.*
  19. S P O O. (A): *He gave his wife some money.* (B): *He bought his wife a watch.* (C): *Lord, forgive us our sins. The rain lasted all day.*<sup>4</sup>
  20. S P C [C expressing duration, distance, price, or weight]. *It lasted all day. We walked (for) five miles. His car cost €12,000. It weighs five tons.*
  21. S P. *Birds fly. We all eat, breathe, drink, and die. The sun was shining.*
  22. S P C. *This is a book. This book is mine. The leaves have turned red.*
  23. S P A. *The sun rises in the east.*
  24. S P A. *He called on me.*
  25. S P {*to*/INF}. *We stopped to have a rest.*

Admiration for Hornby's insights into the nature of syntax and his organized presentation of pattern structures should not blind us to the fact that there are some problems with the way that he presented the data. In the first place, there are 25 verb patterns, which is a lot for a learner to memorize and know how to apply. This is made harder by a number of subtle semantically motivated subdivisions and by the fact that the clause roles are expressed in abstract terms, being referred to by numbered references to a look-up table, rather than by a phrase or name with mnemonic value.

A second problem is that not only did Hornby revise his patterns from time to time, but also he changed their order and numbering. It must have been hard for teachers and students who had taken the trouble to memorize the patterns of the 1963 edition to relate them to the new order and recognize that, say, VP11 of 1963 corresponded to VP9 of 1974. There is no obvious way of associating the term "VP9" or "VP11" with a *that*-clause. Conscientious teachers and learners must have spent many hours thumbing back to the front matter of the dictionary. Less conscientious users would have simply ignored them, thus failing to benefit from the important information about idiomatic phraseology which they encapsulated.

A third problem is that in Hornby's work there is no obvious motivation for the order of patterns as they are presented in the front matter of the dictionary—no differentiation, for example, between simple clause structures on the one hand and more complex structures involving subordinate clauses or infinitive forms on the

---

<sup>4</sup> Subclasses A and B of pattern 19 alternate with pattern 18 (SPOA)—e.g. *He gave some money to his wife; he bought a watch for his wife.* Subclass C does not so alternate. In the second example of C, the phrase *all day* is not really an object at all, but rather a time adverbial, even though it does not have a prepositional head.

other hand<sup>5</sup>. They are all jumbled together, higgledy-piggledy, and there is quite a lot of overlap. There are some over-subtle distinctions.

These are no doubt among the reasons why the verb patterns were eventually greatly simplified in more recent editions of the *Oxford Advanced Learner's Dictionary*. By the time of Crowther's 5th edition (2000), patterns were no longer identified by numbers but rather by abbreviated phrases with mnemonic value. Thus the 1974 "VP11", with its rather clumsy front-matter apparatus "S + *vt* + noun/pronoun+ *that*-clause", had become a simple mnemonic: "Vn (that)". Here, each element in the pattern name has mnemonic value: V means 'verb', n means 'noun', and '(that)' signifies a clausal complement introduced by the subordinating conjunction *that*. The conjunction is often omitted in informal English speech and writing, hence the brackets.

In OALD6, the patterns are set out more clearly in the front matter, but in a much reduced form. The emphasis is on streamlining the presentation for the user. The number of patterns has been reduced from 25 to 20, but there is little significant loss of information. Pattern numbers have been abandoned in favour of mnemonics, and the technical grammatical terminology has been reduced to a minimum. In the front matter, the summary of patterns is more carefully ordered. Patterns that take clauses are separated from the rest: the 20 patterns are summarized and organized under six subheadings, in order of gradually increasing complexity, with example sentences, as follows:

Intransitive verbs

[V]                    *A large dog appeared.*

[V + *adv/prep*]    *A group of swans floated by.*

Transitive verbs

[VN]                    *Jill's behaviour annoyed me.*

[VN + *adv/prep*]   *He kicked the ball into the net.*

Transitive verbs + two objects

[VNN]                  *I gave Sue a book for Christmas.*

<sup>5</sup> A difficulty related to this last point is that the learner had to deal with two types of grammatical element: the functional SPOCA elements that encode 'who is doing what to whom' in a clause, and formal elements such as *to/INF*. For example, Hornby's pattern 3 is "SVO *to/INF*". Some modern descriptive linguists would see this as involving two different subtypes of O. For instance, Francis et al. (1996) identify Hornby's pattern 3 as an SVOO pattern, with *to-infinitive* being regarded as an object on a different syntactic 'layer', thus:

	Verb group	Noun group	<i>to</i> -infinitive clause
Subject	Verb	Object	Object
My girlfriend	nagged	me	to cut my hair.

## Linking verbs

[V-ADJ]	<i>His voice sounds hoarse.</i>
[V-N]	<i>Elena became a doctor.</i>
[VN-ADJ]	<i>She considered herself lucky.</i>
[VN-N]	<i>They elected him president.</i>

## Verbs used with clauses or phrases

[V <b>that</b> ]	
[V ( <b>that</b> )]	<i>He said that he would prefer to walk.</i>
[VN <b>that</b> ]	
[VN ( <b>that</b> )]	<i>Can you remind me that I need to buy some milk?</i>
[V <b>wh-</b> ]	<i>I wonder what the job will be like.</i>
[VN <b>wh-</b> ]	<i>I asked him where the hall was.</i>
[V <b>to</b> ]	<i>The goldfish need to be fed.</i>
[VN <b>to</b> ]	<i>He was forced to leave the keys.</i>
[VN inf]	<i>Did you hear the phone ring?</i>
[V <b>-ing</b> ]	<i>She never stops talking.</i>
[VN <b>-ing</b> ]	<i>His comments set me thinking.</i>

## Verbs + direct speech

[V speech]	<i>“It’s snowing,” she said.</i>
[VN speech]	<i>“Tom’s coming to lunch,” she told him.</i>

In the dictionary itself, if a sense of a verb participates in more than one pattern, the patterns are stated alongside individual examples, rather than before the definition. Thus, sense 1 of *propose*, “to suggest a plan, an idea, etc., for people to think about and decide on”, is illustrated by no less than seven example sentences, showing participation in no less than five patterns, one of which records a British/American variation in the wording of the subjunctive in the *that* clause:

◇ [VN] *The government proposed changes to the voting system.* ◇ *What would you propose?*

◇ [V **that**] *She proposed that the book be banned.* ◇ (BrE also) *She proposed that the book should be banned.*

◇ [VN **that**] *It was proposed that the president be elected for a period of two years.*

◇ [V **-ing**] *He proposed changing the name of the company.*

◇ [VN **to** inf] *It was proposed to pay the money from public funds.*

Let us come back to Hornby's original point, namely that “\**I proposed him to come*” is a grammatical error. Is this contradicted by the third and fifth patterns in this entry? OALD6 tries to explain in a “help note”: “This pattern is only used in the passive.” Unfortunately, it does not say that this comment applies also to the “[VN **that**]” pattern. Moreover, “VN” is, as a matter of fact, never true of the verb *propose* in this sense. You can *propose a plan or idea*, but you cannot \**propose a person to do something* or \**propose a person that something*. The impersonal passive, with proleptic *it*, cannot be equated with an object of an equivalent active use. You cannot say “\**The government proposed him to pay the money from public funds*” or “\**We proposed the government to pay the money from public funds.*” Hornby was right all along; by oversimplifying the grammatical apparatus, his successors have got it wrong with regard to the impersonal passive, which must be recognized as a pattern in its own right, not treated as a transformation of an active.

I mention this small point to illustrate just how difficult it is to get the details of patterns of idiomatic usage right. Grammatical descriptions as subtle, detailed, and factually accurate as that of OALD6 can only be teased out, word by word, with results achieving reasonable accuracy, by painstaking analysis of large quantities of corpus data, supported by a reasonably sophisticated grammatical apparatus.

An excellent research topic for a Ph.D. dissertation would be a comparison of verbs and verb patterns in the pre-corpus 1963, 1974, and 1989 editions of Hornby's and Cowie's OALD with the corpus-based 5th edition, edited by Crowther and the 6th edition edited by Wehmeier. This would not merely be of historical interest. Because neither Hornby nor Cowie had a corpus—corpora had not been invented in their day—they were reliant to a large extent on their intuitions when describing patterns. No doubt these intuitions were excellent and finely tuned, as they were both experienced, insightful, and widely read teachers of English, but there are many places in the dictionary where corpus evidence have prompted their successors to revise their entries. A systematic comparison of the pre- and post-corpus editions, focusing in particular on the description of verb behaviour, would shed valuable light on the complementary roles of evidence and intuition in a succession of highly skilled lexicographical teams striving to achieve what is essentially the same goal, using very similar descriptive apparatus but very different kinds of source data.

### 3. Patterns in English Dictionaries since Hornby

It is instructive to compare the treatment of verb patterns in monolingual English dictionaries since Hornby. A certain amount of ambivalence on the part of lexicographers may be detected. Is it really the job of the dictionary to explain grammar in the tradition of Hornby, and if so, how should it be done? The general tendency in recent years has been for British dictionaries to focus on word meaning and word classes, to minimize the explicit grammatical apparatus and terminology, to describe collocates in terms of their word classes rather than their clause roles or semantic types, and to attempt to convey grammatical information by judicious selection (or in OALD's case, construction) of examples.

Let us first look in more detail at the entry for *propose* in OALD6 and consider some issues of lexicographic principle. Although detailed discussion of a single verb does not constitute a statistically valid sample for purposes of dictionary evaluation, it does raise some interesting theoretical and practical points, which have far-reaching consequences. The full entry in OALD6 is as follows:

**propose**, *verb*

[SUGGEST PLAN] **1** (*formal*) to suggest a plan, an idea, etc., for people to think about and decide on: [VN] *The government proposed changes to the voting system.* ◇ *What would you propose?* ◇ [V **that**] *She proposed that the book be banned.* ◇ (BrE also) *She proposed that the book should be banned.* ◇ [VN **that**] *It was proposed that the president be elected for a period of two years.* ◇ [V **-ing**] *He proposed changing the name of the company.* ◇ [VN **to inf**] *It was proposed to pay the money from public funds.* [HELP] This pattern is only used in the passive.

[INTEND] **2** to intend to do sth: [V **to inf**] *What do you propose to do now?* ◇ [V **-ing**] *How do you propose getting home?*

[MARRIAGE] **3** ~ (sth) (to sb) to ask sb to marry you: [V] *He was afraid that if he proposed she might refuse.* ◇ [VN] *He proposed marriage.*

[AT FORMAL MEETING] **4** [VN] ~ sth | sb for /as sth to suggest sth at a formal meeting and ask people to vote on it: *I propose Tom Ellis for chairman.* ◇ **to propose a motion** (= to be the main speaker in support of an idea at a formal debate)—compare OPPOSE, SECOND.

[SUGGEST EXPLANATION] **5** [VN] (*formal*) to suggest an explanation of something for people to consider [SYN] PROPOUND: *She proposed a solution to the mystery.*

[IDM] **propose a toast (to sb) | propose sb's health** to ask people to wish sb health, happiness, and success, by raising their glasses and drinking.

With the exception of the quibble about the impersonal passive, discussed above, this is lexicography of the highest order of delicacy, clarity, and accuracy—about as good as it gets, given existing assumptions about lexicographic principles. The definitions are clear, the grammatical description is well thought out, and the example sentences are (with minor exceptions) well constructed for the benefit of learners.<sup>6</sup>

---

<sup>6</sup> Example sentences in OALD6 and 7 are designed to illustrate patterns of normal usage. Generally, they are not actual quotations from texts in a corpus, but rather corpus-inspired constructs designed to illustrate linguistic competence. In this respect, the recent editions of OALD differ from other corpus-based dictionaries. I shall say no more about this controversial issue here.

It is with some diffidence, therefore, that I will now suggest a move towards new lexicographical principles. My suggestions are inspired by corpus analysis. I would like to believe that, if Hornby had had access to corpus evidence, he would have been sympathetic to these proposals. The aim is to take the best in traditional and current lexicographical practice and ask whether it could be better. I propose eight new principles.

1. *Avoid fine-grained semantic distinctions*

Computational linguists often assert that distinctions in dictionary definitions are “too fine-grained”. One motive in this complaint is that computational linguists want definitions to be mutually exclusive, but this is a mistake. It confuses natural language with predicate logic. There is much overlap everywhere in matters of word meaning. Nevertheless, it may be that, as lexicographers, we have something to learn from this more general complaint: it can also be read as a polite way of telling us that some dictionary entries are not merely too fine-grained but needlessly repetitious. In OALD’s entry for *propose*, is it really necessary to make a distinction between senses 1 and 5, for example? Proposing an idea and proposing an explanation are semantically very close and should perhaps not be distinct, since it is simply not possible to come up with two distinct lexical sets of direct objects: one of nouns that mean ‘idea’ and the other of nouns that mean ‘explanation’. Consider the phrase ‘propose a hypothesis’. Is this sense 1 or sense 5 of *propose*? Is a *hypothesis* an idea or an explanation? The answer, of course, is that it is both. Sets of direct objects often constitute a chain of overlapping Wittgensteinian family resemblances. It is hard to defend the idea that there are two different transitive senses of *propose* in this case. A few sharp slashes of Ockham’s razor (avoiding the needless replication of entities) are called for. (I hasten to add that OALD is not the only dictionary that makes this unnecessary distinction.)

2. *Do not confuse domain with meaning*

*Proposing a motion* (sense 4) is semantically identical to *proposing a plan for people to think about and decide on* (sense 1). They belong together. OALD6 rightly gives information about the domain in sense 4 (“*at a formal meeting*”), but this does not need to be dressed up as a semantic distinction. A domain-commented example at sense 1 would be clearer and more elegant.

3. *Take account of the semantic types of collocates*

In contrast to the previous point (2), sense 4 needs to be split. *Proposing someone for/as a specific role* does not fit well semantically with *proposing a motion*. Both uses of this verb are indeed typical of the domain of formal meetings and parliamentary procedure, but there the similarity ends. There is a semantic difference. The two meanings (*propose a motion* and *propose*

*a person for a role*) are distinguished by the semantic types [[Person]] and [[Proposal]], and the explanation will be clearer if they are kept separate. A person is not a proposal.

4. *Patterns should play a role in organizing the entry*

The preceding point implies that patterns which have different meanings should be treated as different senses even if they are in the same domain. Carrying this idea a step further, we see that sense 1 of *propose* in OALD6 is associated with no less than five patterns. This is not a problem in itself, but the question must be asked: are they all in the right place? Consider the problematic last pattern, “◇ [VN to inf] *It was proposed to pay the money from public funds.*” Semantically, this is on a borderline between the [SUGGEST PLAN] sense and the [INTEND] sense. As so often happens, there is no sharp boundary between the two senses. For lexicographical purposes, however, a clearer focus will be achieved if all the [to inf] patterns are grouped at the [INTEND] sense, i.e. sense 2, not sense 1. This will work well because a *to*-infinitive governed by the verb **propose** always signals intended action.

5. *Seek the right level of delicacy in pattern description*

In sense 3, the [MARRIAGE] sense, the essential point is that, if there is no direct object, i.e. if *propose* is intransitive, the normal meaning is “ask someone to marry you”, not “suggest a plan to them”. The intransitive pattern is the normal one, and for the sake of clarity of exposition, it should be separated out, not integrated with other possible ways of realizing the same meaning. Conversely, some patterns crop up, rather confusingly, in several different senses of: for example, there are four occurrences of the verb pattern “[VN]” in this entry: in senses 1, 3, 4, and 5. Do these really represent different senses? Would anything be lost if they were lumped together? If they really represent different senses, can they be differentiated according to the semantic type of the nouns involved?

6. *The right level of pattern delicacy implies sorting according to semantic type, not just word class*

The addition of a [VN] pattern to sense 3, reinforced by the rubric “~ (sth) (to sb)”, is less than helpful. It muddies the waters: if taken literally, it could be read, wrongly, as implying that a sentence such as *John proposed a swim (or a cycle ride) to Mary* means that he asked her to marry him! The right level of delicacy requires explicit mention of the noun *marriage*, not the indefinite pronoun *sth* (*something*). This apparently simple and obvious point has far-reaching consequences.

For the vast majority of transitive uses of *propose*, the preferred semantic type of the direct object is [[Event]] or [[Plan]]. Unfortunately, this is obscured by another common linguistic phenomenon, namely ellipsis. In examples such as ‘Local government officials were able to **propose**

new dual-carriageway trunk roads’, the underlying meaning is ‘Local government officials were able to *propose the construction of new dual-carriageway trunk roads*’, where the absent noun *construction* is of semantic type [[Event | Plan]].

7. *Seek the right level of delicacy in syntactic comments*

The HELP note at the last pattern of sense 1 is misleading because it is underrestricted. This pattern is normally found only in the impersonal passive. It would, for example, be stretching idiomaticity to say \**Money was proposed to be paid from public funds*. This is exactly the sort of invented borderline example—just about possible, but bizarre or abnormal and not supported by evidence—that has bedevilled armchair linguistics for the past half century and led to much pointless speculation about a sharp dividing line between syntactically well-formed and ill-formed sentences.

A similar point arises regarding the  $\diamond$  [VN **that**] pattern, also in sense 1. *It was proposed that the president be elected for a period of two years* is likewise an impersonal passive. It is not idiomatic to say \**The president was proposed that he be elected for two years*, still less the active, which I suppose would have to be something like \**They proposed the president that he be elected for two years*, which is gibberish.

8. *The right level of syntactic delicacy implies clause role description, not word classes*

The apparatus of OALD6’s verb patterns is beautifully clear and simple. So at first sight it may appear unnecessarily pedantic to insist that VN ought to be SPO. In this case, it makes little difference, but elsewhere use of word classes in place of clause roles has led to errors in analysis, as we shall see shortly. It is also important to distinguish patterns that are typically active from those that are typically passive.

Similar points could be made about other verbs, and not only in OALD but also in all the other leading learner’s dictionaries, for they are all *meaning-driven*. What I am proposing here, in a nutshell, is a dictionary that is not merely *corpus-driven*, but *pattern-driven*.

To show how this works, let me start by illustrating a possible new version of the OALD entry, taking account of the above points:

**propose**, *verb*

[SUGGEST PLAN] 1 to suggest a plan, an idea, or an explanation of something, for people to think about and decide on: [S P O] *The government proposed changes to the voting system.*  $\diamond$  [S P O] *What would you propose?*  $\diamond$  [S P **that**] *She proposed that the book be banned.*  $\diamond$  (BrE also) *She proposed that the book should be banned.*  $\diamond$  [ it be P (impersonal passive) **that**] *It was proposed that the president be elected for a period of two years.*  $\diamond$  [S P

**-ing**] *He proposed changing the name of the company.* ◇ [AT A FORMAL MEETING] **to propose a motion** (= to be the main speaker in support of an idea at a formal debate)—compare OPPOSE, SECOND.

[INTEND] **2** to intend to do sth: [S P **to** inf] *What do you propose to do now?* ◇ [S P **-ing**] *How do you propose getting home?* ◇ [it be P (impersonal passive) **to** inf] *It was proposed to pay the money from public funds.*

[MARRIAGE] **3** [S P] to ask sb to marry you: *He was afraid that if he proposed she might refuse.*

[SUGGEST FOR A ROLE] **4** [S P O **for /as** role] to suggest at a formal meeting that someone should be elected to a particular role: *I propose Tom Ellis for chairman.*

[CELEBRATE] **5** **propose a toast (to sb) | propose sb's health** to ask people to wish sb health, happiness, and success or celebrate their achievement, by raising their glasses and drinking.

The obvious differences are slight. The definitions and examples are mostly unchanged. The overall length is slightly shorter, even though the grammatical apparatus is slightly more elaborate. Small improvements in accuracy and conciseness can be achieved by applying the eight principles outlined above to the existing wording of a verb entry, even to a comparatively 'open' word like *propose*. Much greater improvements are achieved by application of these principles to words at the more 'idiomatic' end of the scale, such as *devour* and *scratch*. There is insufficient room to discuss these here in terms of OALD's entries for these words. Instead, I would ask readers to go straight to the pattern dictionary entries in section 6 of this paper, where a more radical mapping of meaning onto use is proposed, and to make their own comparisons. The entries in all current dictionaries, including OALD, are meaning-driven, i.e. they ask the question, "How many senses does each word have, and what is the definition of each sense?" The question addressed by the pattern dictionary (of which a sample is given in section 6) is, "How many patterns does each word participate in, and what is the sense of each pattern?"

Before moving on to that, however, I would like to comment briefly on grammar in some other dictionaries. The grammatical apparatus of the first edition of the *Longman Dictionary of Contemporary English* (LDOCE; 1978) was at least as elaborate as that of Hornby, but even more impenetrable for ordinary learners. By the corpus-based 3rd edition (1995), LDOCE had adopted a much simpler grammatical apparatus, joining the general trend of learners' dictionaries away from explicit grammar patterns. The only technical terms that it uses are Intransitive [I] and Transitive [T]. The third argument of a verb is described, systematically, as [+ adv/prep], thus:

**amble** *v.* [I always + adv/prep] to walk in a slow relaxed way: [+ **along/across etc**] *the old man came out and ambled over for a chat.*

The front matter of LDOCE3 comments: “You cannot simply say ‘he ambled’ without adding something like ‘along’ or ‘towards me’.”

The *Cambridge International Dictionary of English* (CIDE, 1995), which changed its name to *Cambridge Advanced Learner’s Dictionary* (CALD) for the second edition (2005)<sup>7</sup>, adopts a similar apparatus to that of OALD and LDOCE: minimal and economical, but sufficient.

Surprisingly, the grammatical apparatus of the *Macmillan English Dictionary for Advanced Learners* (MEDAL, 2002) is Spartan to the point of being misleading. This seems to be a deliberate policy, since the principals involved in creating MEDAL had all worked on other learners’ dictionaries, which have more sophisticated, though perhaps minimal, grammar patterns. Presumably, it was decided as a matter of policy that MEDAL should focus on meanings, examples, and collocations, not on grammatical abstractions.

Let me illustrate this with the MEDAL entry for *amble*. Like traditional dictionaries, MEDAL distinguishes transitive and intransitive subcategories of verb senses but, unlike LDOCE, OALD, and CIDE, it generally neglects or misstates the third argument, if there is one. So *amble* in MEDAL is described simply as *verb [I]*. This implies that sentences like *\*the old man ambled is a well-formed sentence of English*. It is not. You have to say where he ambled to—*along, out of the house, into the pub, or whatever*.

This is only one of several examples that could be mentioned. Cumulatively, they add up to a misleading account of verb grammar. It seems that MEDAL has allowed its desire to keep things simple for the learner to be carried to the point where the policy interferes with accurate reporting of the facts of the language.

Cobuild falls into a similar trap. In the second and subsequent editions, the grammar pattern for *amble* in this dictionary is correctly given as “V adv/prep”. Unfortunately, the Cobuild definer forgot to replicate the adv/prep in the definiendum (the first part of the full-sentence explanation). The Cobuild explanation reads: ‘When you *amble*, you walk slowly and in a relaxed manner’. This gives the same mistaken impression as MEDAL. There is a word missing. Cobuild’s full-sentence explanation should read, “When you *amble* somewhere, you walk there slowly and in a relaxed manner.”

In other cases, for example *put*, MEDAL hints at the obligatory third argument, which in this case is really an adverbial of location, but it does so (or tries to do so) only by mentioning specific prepositions, not the relevant clause role, e.g.

**put sth in/on/through/etc. sth**

**put sth into/over/out/etc. sth**

---

<sup>7</sup> Several dictionary publishers in Britain have got into the habit of changing the name of their dictionaries with new editions, even if there is comparatively little alteration. Conversely, when a successful dictionary is totally rewritten, so that it is, in fact, a completely different book, its former title may be retained—and the book may even be published under the name of a long-dead editor. No doubt these things are done for good marketing reasons, but they add to the already difficult complexities of giving accurate bibliographical details for lexicographical works.

This is inadequate, a) because it is verbose and b) because it fails to get the right level of generalization. To focus on specific prepositions is irrelevant, obscuring the equally important fact that the adverbial argument for this verb is often realized by other prepositional phrases and indeed as a single word, e.g. *put it here*, *He put the bin outside*.

However, it must be said that the MEDAL entry is not as inadequate in this respect as American dictionaries of English, for example Merriam Webster's Collegiate (MW), which focuses obsessively, repetitively, and often inaccurately on the transitive/intransitive distinction, while saying nothing at all about the third argument, seemingly being unaware of it. MW implies, for example, that *\*I put the cup* is a well-formed sentence of English. Of course, it is not. An adverbial of location is obligatory—you must say *where* you put it.

The root of this lexicographical problem, like many others in the grammatical apparatus of pre-corpus dictionaries, goes back 1,500 years. Latin grammarians such as Aelius Donatus and Priscian divided verb uses into transitive and intransitive, but they did not recognize adverbials as an essential part of clause structure. English grammarians of the 18th century, presumably under the impression that English is really Latin in disguise, did not recognize them either, and English dictionaries in the 19th and 20th centuries followed the 18th-century grammarians in this and other respects. Some current dictionaries have made no attempt to update their grammatical apparatus or to offer an adequate description of the syntagmatic patterns of word behaviour. Merriam Webster's Collegiate is in this tradition, not only failing to identify the third argument of verbs but also postulating nonexistent intransitive variants of transitive senses, for example:

**put**, *vi.* **1** to start in motion; GO, esp: leave in a hurry.

It is hard to know what to make of this. It seems to imply that *\*John put* and/or *\*the train put* are well-formed sentences of English, meaning “John (or the train) went, or left in a hurry”. But in fact, they mean no such thing: they are both meaningless and ungrammatical. And this definition cannot be an attempt to cover the nautical expression *put to sea*, for that is dealt with in a second sense:

**2** *of a ship*: to take a specified course: <put down the river>.

As with so many of Merriam Webster's minor definitions, in the absence of supporting evidence we must resign ourselves to a state of unresolved bafflement. Similar problems afflict the recording of other grammatical features in all American English dictionaries, for example phrasal verbs and determiners. They do not exist in Latin, so their existence is not explicitly recognized in American dictionaries. This is a shocking state of affairs. The corpus revolution and the grammatical analyses of Quirk and other empirically minded grammarians, which have led to so many improvements in British monolingual dictionaries, have up to now been passed by in American lexicography, suffering as it does under the stranglehold of a market leader that has made little or no investment in serious lexicographical research or innovation for over 40 years.

Let us return to our main theme, namely patterns in EFL dictionaries. It is a pleasure to report that, even though MEDAL does not account for adverbials correctly, it does a good job on phrasal verbs and determiners. MEDAL had the great advantage that the compilers were able to use a state-of-the-art tool for corpus analysis, the Sketch Engine (Kilgarriff et al, 2004), to help them select significant collocations and write definitions reflecting these. The dictionary is peppered with explicit reports on common collocations, e.g.

Words frequently used with *propose*:

*nouns*: change, idea, plan, reform, scheme, solution, theory

If you believe that learners of a language build their own competence analogically on the basis of salient examples, these lists of collocates must be of great benefit. We should bear in mind, however, Hornby's scepticism about the reliability of analogy as a learning tool. The debate about the relative merits of rule-based approaches and analogical approaches to language learning will no doubt continue to run and run for many decades to come.

Here is MEDAL's entry for *propose*:

**propose**

1. [T] *formal* to suggest a plan, idea, or action: *Einstein proposed his theory of general relativity in 1915.* ◇ *I propose going to an early film and having dinner afterwards.* ◇ + **that** *She proposed that we see a marriage guidance counsellor.*
2. [T] to make a formal suggestion in a meeting for people to think about and vote on:  
 ◇ **propose sb for sth** *I propose Sue Wilson for treasurer.* ◇ **propose doing sth** *France has proposed creating a rapid-reaction force to deal with the crisis.*  
**2a. propose a motion** to formally suggest an idea or plan at a meeting.
3. [I/T] to ask someone to get married to you: +**to** *He proposed to her in August.* ◇ **propose marriage** *He proposed marriage, but she refused.*
4. [T] If you propose to do something, you intend or plan to do it: *I propose to tell them the absolute truth.*

It can readily be seen that most of the information that is in OALD6 and in LDOCE is presented here in a similar order and in similar wording, though formatted slightly differently. At sense 3, MEDAL's explicit mention of *propose marriage* is rather more helpful than OALD's formulation, "*~ (sth) (to sb)*". It is arguable, however, that the grammatical label at MEDAL's sense 3 should simply be [I] and the lexically specific transitive alternation 'propose marriage' should be ignored, on the grounds that it is rare, used only for clarification, and covered by sense 1 anyway.

The notion that the infinitive in sense 4 represents a transitive [T] is debatable. It is arguably more helpful to learners to classify the to-infinitive as a clausal argument, and to reserve [T] for noun phrases. However, MEDAL is not alone in taking this view of infinitives: Francis et al. (1996), for example, takes a similar line.

The fourth dictionary I wish to mention in this context is Cobuild. Cobuild has always been a corpus-driven dictionary, so it was spared the expense of having to revise all of its entries in the light of corpus evidence. However, it has made up for this by adopting radically different policies with regard to grammatical description in different editions, and replacing all its examples from new corpus data anyway. The first edition (1987) offered a SPOCA-based description of the clause structure associated with each meaning of each verb. The grammatical apparatus of this first edition received mixed reviews. Admittedly, it was often cumbersome and hard to follow, and occasionally got things wrong or ventured into controversial territory. These may be among the reasons why Cobuild2 adopted a more minimalist, streamlined approach to grammatical description. The grammatical descriptions were moved to sit alongside examples rather than explanations, which yields a great improvement in clarity. However, in addition, the SPOCA-based terminology was abandoned, and the apparatus for grammatical description was reduced to a word-class based system similar to those adopted by later editions of LDOCE and OALD. I cannot help feeling that this move, shared by all EFL dictionaries, has been a case of throwing out the baby with the bathwater. You will see why, I hope, in the discussion of pattern grammar in section 4 below. Simplicity and clarity are great virtues, but not if they are bought at the expense of descriptive adequacy.

### **propose, proposes, proposing, proposed (COBUILD3)**

1. If you propose something such as a plan or idea, you suggest it for people to think about and decide upon:
 

<i>Britain is about to propose changes to European Community institutions.</i>	V n
<i>It was George who first proposed that we dry clothes in that locker.</i>	V that
2. If you propose to do something, you intend to do it.
 

<i>It's far from clear what action the government proposes to take</i>	V to-inf
<i>And where do you propose building such a huge thing?</i>	V -ing
3. If you propose a theory or an explanation, you state that it is possibly or probably true, because it fits in with the evidence that you have considered.
 

<i>This highlights a problem faced by people proposing theories of ball lightning.</i>	V n
<i>Newton proposed that heavenly and terrestrial motion could be unified with the idea of gravity.</i>	V that

4. If you propose a motion for debate, or a candidate for election, you begin the debate or the election procedure by formally stating your support for that motion or that candidate.  
*A delegate from Siberia proposed a resolution that he stand down as party chairman.* V that  
*I asked Robin Balfour and Derek Haig to propose and second me.* V n
5. If you propose a toast to someone or something, you ask people to drink a toast to them.  
*Usually the bride's father proposes a toast to the health of the bride and groom.* V n
6. If you propose to someone, or propose marriage to them, you ask them to marry you.  
*He had proposed to Isabel the day after taking his seat in Parliament.* V to n

A unique feature of Cobuild is that it systematically attempts to capture informally the collocational preferences of each sense of each word, by means of 'full-sentence definitions', of which the first part is usually the definiendum (the phrase or pattern that is to be defined), encoded within the definition. Cobuild is also "corpus-driven". Patterns are discovered through corpus analysis. It is, therefore, disappointing to have to note that, in terms of the distinction being made in this paper, the entry structure of Cobuild is meaning-driven rather than pattern-driven. *Proposing an idea* and *proposing a theory*, for example, are treated as separate senses, just as they are in other dictionaries. If the compilers had focused on patterns rather than senses, this dubious semantic distinction might have been treated as a single pattern. Like all other existing major dictionaries, Cobuild's starting point is a list of senses for each word, not a list of the patterns in which the word normally participates. It was also criticized for a tendency to verbosity. Critics have associated this tendency with the "full-sentence definitions". In my opinion, the criticism of verbosity was to some extent justified and indeed was addressed in the second edition. However, associating this with full-sentence definitions misses the point. Cobuild is the only serious attempt by any dictionary to systematically identify collocates by semantic type (as opposed to word class), in the definiendum. The impression of verbosity of the definitions results from two factors: firstly, a tendency not to know when to stop, as in the original definition 7 of *proportion* (below) and secondly the frequent attempts to deal with more than one pattern at the same time, as in sense 4 of *propose* (above).

In the first edition of Cobuild (1987), definition 7 of *proportion* read as follows:

If you say that something is big or small **in proportion to** something else, you mean that it is big or small when you compare it with the other thing or measure it against the other thing.

This is undeniably verbose. In the 2001 edition, it was reduced to:

If something is small or large **in proportion to** something else, it is small or large when compared with that thing.

This is a full-sentence definition, but not especially verbose.

MEDAL defines *proportion* (sense 2) as: “the correct, most useful, or most attractive relationship between two things”, and offers the phrase *in proportion to* with an example (“his head is large in proportion to his small frame”) but no definition. An undefined example may be the best strategy for such a phrase.

It is time to move on, but before we leave the question of whether any existing dictionaries are pattern-driven, there is one more dictionary to consider. It is not a learner’s dictionary but a dictionary aimed at native speakers. Somewhat surprisingly, it comes closest to showing how sense distinctions can be made on the basis of patterns. The *New Oxford Dictionary of English* (1998)<sup>8</sup> is (so far) the only dictionary of English aimed at native speakers that both takes corpus evidence seriously and incorporates grammatical descriptions in the Hornby tradition. The entry for *propose* is as follows:

### **propose**

► **verb 1.** [with obj.] put forward (an idea or plan) for consideration and discussion by others: *he proposed a nine-point peace plan* | [with clause] *I proposed that the government should retain a 51 per cent stake in the company.*

■ nominate (someone) for an elected office or as a member of a society: *Roy Thomson was proposed as chairman.*

■ put forward (a motion) to a legislature or committee: *the government put its slim majority to the test by proposing a vote of confidence.*

■ [with infinitive] intend to do something: *he proposed to attend the meeting.*

**2.** [no obj.] make an offer of marriage to someone: *I have already proposed to Sarah* | [with obj.] *one girl proposed marriage to him on the spot.*

In this entry, there is a clear attempt to associate sense distinctions with pattern distinctions, using SPOCA as a basis. Keen-eyed readers will no doubt notice that there is no mention of the expression *propose a toast*.

So far, I have discussed grammar patterns in relation to lexical definitions. But notice that the grammatical error mentioned by Hornby, *\*I proposed him to come*, is not a simple error of structural pattern. The structural pattern “S P O to/INF” (or, if you prefer, “V n to/INF”) is perfectly correct for *propose* in some contexts, e.g. *the Council proposed a plan to widen the road*. The error lies in the selection of a word of the wrong semantic type—the personal pronoun *him*, which has Semantic Type [[Human]] rather than [[Plan]]—in the object slot.

<sup>8</sup> Now marketed in a revised edition as the *Oxford Dictionary of English* (not to be confused with Oxford’s great historical work in lexicography, the 20-volume *Oxford English Dictionary*).

#### 4. Pattern grammar vs. pattern dictionary

There are two possible approaches to using a corpus to identify patterns in text: pattern grammar and pattern dictionary. Both have their merits; both have their shortcomings. The *Pattern Grammar* of Hunston and Francis (2000; H&F) is based on the grammatical apparatus of the second edition of the Cobuild dictionary. It is founded on corpus analysis (i.e. on real texts) and seeks empirically valid generalizations. The following remark (p. 83) is highly relevant:

One of the most important observations in a corpus-driven description of English is that patterns and meanings are connected.

On pages 199-207, H&F discuss grammatical patterns in a short text, reproduced below, which they refer to as “the Joseph Byers text”. In this section of my paper, I shall use this text and the H&F discussion of it to illustrate some of the differences between a pattern dictionary and a pattern grammar and to show how the two approaches are complementary. Here is the text:

Private Joseph Byers was the first Kitchener volunteer to be executed. He was 17 and under age when he enlisted in the 1st Royal Scots Fusiliers in November 1914, and was sent to France with two weeks training. By January 1915, his inexperience and the horrors he witnessed caused him to go absent without leave with another private, Andrew Evans. Byers pleaded guilty, believing that his candour would save him from the death sentence. Despite being under age, he was given no representation at his trial, and he and Evans faced a firing squad at Loche on February 6.

According to rumours, one of them did not die until the third volley, leading to speculation that the firing squad had fired wide to avoid killing the youth.

Table 1 (below) compares the verb patterns identified by H&F in this text with the relevant pattern of each verb in the Pattern Dictionary of the Corpus Pattern Analysis project (CPA, in progress). H&F limit themselves to expressing patterns all at more or less the same level of generalization, almost exclusively in terms of word classes (parts of speech), with the exception of certain prepositions. CPA, by contrast, devotes a great deal of attention to selecting the appropriate level of generalization to capture the meaning of the lexical pattern and to contrast it with other meanings activated by other patterns for the same verb. This necessitates a much richer grammatical apparatus, including identifying, among other things, the semantic type of the subject and, for each clause role, statistically significant collocates grouped by semantic type. Semantic types are identified in double square brackets and refer to a shallow ontology (see section 8 of this paper).

Verb	Pattern Grammar	Pattern Dictionary	Comments
execute	V n	[[Human 1]] execute [[Human 2]]	Semantic types distinguish this sense from others of the same verb, e.g. ‘execute an order’.
enlist	V in n	[[Human]] enlist [NO OBJ] {in [[Human Group = {Military}]}	CPA marks intransitive patterns explicitly. This pattern contrasts with patterns such as “[[Human]] enlist [[Assistance]]”.
send	<i>be</i> V-ed to n (passive of V n to n)	[[Human 1]] send [[Human 2]] [A[Direction]]	
witness	V n	[[Human]] witness [[Event]]	
cause	V n to-inf	[[Anything 1]] cause [[Anything 2]] {to/INF [V]}	In this pattern, semantics add nothing to the basic word-classes.
go	V adj	[[Human]] go [NO OBJ] {absent   AWOL}	Light verb (“delexical verb” in Sinclair’s terminology), with a Subject Complement. This small lexical set activates a particular meaning of <i>go</i> , contrasting with several other patterns of <i>go</i> having a Subject Complement, e.g. <i>go</i> { <i>mad</i>   <i>bananas</i> }.
plead	V adj	[[Human]] plead [NO OBJ] {guilty   {not guilty}}	The adj. in this pattern is a Subject Complement, populated by a lexical set of just two possible items, {guilty} and {not guilty}.
believe	V that	V that [[Human]] believe [NO OBJ] {(that) [CLAUSE]}	
save	V n <i>from</i> n	[[Anything]] save [[Entity]] {from [[Event = Bad]]}	Discussed in Church and Hanks (1990)—a paper not mentioned in Hunston and Francis’s bibliography.
give	be v-ed n (passive of V n n)	[[Human   Event]] give [[Entity 1 = Recipient]] [[Entity 2 = Benefit]]	See discussion of <i>lose</i> below.

Verb	Pattern Grammar	Pattern Dictionary	Comments
face	V n	[[Human]] face [[{Event   Possibility}= Bad]]	Contrasts with [[Entity]] face [A[Direction]]
die	V	[[Animate]] die [NO OBJ] ([A])	Even though an Adverbial is not an obligatory part of the structure of <i>die</i> (and indeed <i>die</i> is often cited as a “one-argument verb”), the norm for <i>die</i> is that it normally governs an <u>optional</u> Adverbial.
lead	V to n	[[Anything]] lead ([[Human]] {o [[Belief]])}	Contrasts with patterns such as [[Route]] lead {to [[Location]]}
fire	V adj	[[Human]] fire [NO OBJ] ([A[Direction]])	H&F fail to identify this pattern correctly. See discussion below.
avoid	V -ing	[[Human   Animal]] avoid [ING]	This pattern contrasts with the pattern [[Human]] avoid [[Event]], e.g. <i>he managed to avoid extradition</i> , where the [[Human]] is typically a Patient not an Agent.

Table 1: Comparison of Pattern Grammar with Pattern Dictionary

It should be mentioned here that the book in the Cobuild series on verb patterns—Francis et al. (1996), which was published four years before H&F—is the one that goes farthest in the delicacy of its grammatical apparatus for syntagmatic distinctions. For example, it groups noun arguments of verbs together according to broad semantic classes, where possible. Thus, under the pattern “V n for n” (Francis et al., 1996: p. 370), there is a meaning group “reward and punish”, which associates this sense of *execute* not only with a direct object and a prepositional phrase (V n for n), but also with the semantic value ‘human’ for both subject and object. It also associates this sense with a third argument—an adverbial governed by *for*—which encodes the thing that the person has done to warrant the reward or punishment. Other members of this meaning group are particular senses of *arrest*, *excuse*, *forgive*, *prosecute*, *punish*, *reward*, *pay back*, *sue*, and *thank* (e.g. *He told officers he wanted to pay them back for locking him up*).

This kind of classification would be sufficient to distinguish *executing a person for murder* from *executing a plan or order*. Unfortunately, however, Francis et al. do not specifically mention this second sense of *execute* under the relevant pattern, “V n”. Their book is a grammar, not a dictionary, so only the most frequent examples of each pattern are given. *Execute a plan* was not common enough to be selected

as a meaning group under “V n”. Francis et al. (1996) is now out of print. It was a pioneering effort in corpus-based grammar and should be revived, as it sketches out some important principles for lexical analysis, which deserve closer study.

Let us now consider the verb *fire* in the fragment ‘speculation *that the firing squad had fired wide*’. I shall go into some detail on this example. The meaning is clear, but how is it constructed? Does it represent a realization of a pattern, or is it anomalous? H&F say:

... we use [the word ‘pattern’] to indicate a sequence of elements that occur with a particular lexical item in this text, whether or not such a sequence is typical. For example, we show the verb *fire* ... with the pattern **V adj**, even though that pattern is productive, is not particularly frequent with this word, and does not distinguish this verb from others.

It seems odd to claim that *any* observed sequence of element can count as a “pattern”. In contrast to what H&F say here, the Pattern Dictionary classifies as patterns only those syntagmatic strings that can be shown, by analysis of corpus evidence, to be typical—i.e. conventional, recurrent chunks of meaningful linguistic behaviour. Classifying just any sequence of elements as a pattern, no matter how idiosyncratic it may be, would seem to defeat the purpose of pattern analysis, opening the floodgates to any observed sequence of elements, no matter how rare or bizarre. In Hanks (forthcoming) a fundamental distinction is made between normal patterns of word use and abnormal uses which deliberately exploit the normal patterns. The latter class includes not only creative metaphors, but also elliptical and anomalous arguments. An example of an anomalous argument of the verb *fire* is ‘stinking spray’ in 1, which exploits the pattern element [[Projectile]], which is populated canonically by words such as bullet, round, shell, rocket, missile, flare.

1. Anyone who has encountered a skunk will know that before it **fires** its stinking spray it issues clear warnings of its intentions.

Be all that as it may, it seems to me that in this particular case, ‘the firing squad fired wide’, there is a pattern, but H&F have failed to identify it correctly. This is because they do not acknowledge clause roles. The pattern in fact consists of a syntagmatic structure with semantic values, expressed as:

[[Human]] fire [NO OBJ] [A[Direction]]

The first hurdle for a lexical analyst in constructing this pattern and applying it to the verb *fire* is to recognize that there is an intransitive verb pattern and that this intransitive pattern is semantically linked to a transitive pattern, “[Human] fire [[Artifact = Firearm]]”. The second hurdle is to recognize that “[Human] fire [NO OBJ] ([A[Direction]])” is a pattern in which a firearm is implied by coercion, even though it is not mentioned explicitly. We can now explain the word *wide*. Contrary to what H&F say, this is in fact not an adjective at all, but an adverbial—a one-word lexical realization of [A[Direction]], a realization of a kind found with several other verbs, for example *aim wide*, *drop short*, *go home*. It answers the question, “Where

did they aim?” or “What did they aim at?” It belongs in the same clause role as *fired over their heads* and *fired into the crowd*.

This analysis goes beyond simple word classes. It introduces contrasts based on the semantic values of collocates, not just syntagmatic structures. In pursuance of this goal, let us ask the sort of question that is asked by the Berkeley FrameNet project, namely: what are the frame elements involved in the semantic frame of people using firearms? We can compile a list like this:

Agent – the person firing the gun

Instrument – the gun or other firearm used

Projectile – the bullet or shell that is fired from the firearm

Target – the thing aimed at or hit

In corpus analysis, frame elements like these are mapped onto idiomatic uses of the lexical items involved. The direct object of the verb *fire* (when the verb is transitive) can be either the Instrument (*fired a gun*) or the Projectile (*fired a shell*).

The difference between a pattern grammar and a pattern dictionary is that a pattern grammar seeks generalizations that affect very large numbers of lexical items, whereas a pattern dictionary looks at each lexical item individually and asks how many patterns it participates in—and what they mean. A pattern dictionary uses patterns to distinguish different meanings of a verb. To do this, it must introduce into the apparatus more delicate structural levels than mere word classes. The theoretical foundations for doing this can be traced back to Halliday (1966) and Sinclair (1966). It seems obvious enough that *firing a gun* activates a different meaning of the verb *fire* from that activated by *firing a person*, even though both these phrases have the structural pattern “V n”. The direct objects must be distinguished according to their semantic types: [[Firearm]] and [[Human]] respectively. Next, discovery procedures are needed to predict whether a lexical item that occurs as the direct object of *fire* is more likely to be a [[Firearm]] or a [[Human]]. To do this effectively, semantic values must be assigned to the arguments of patterns. These semantic values are encoded in a shallow ontology, which I will discuss in the next section.

## 6. Introducing semantic values of arguments into patterns

Consider for a moment the verb *enlist*. It has two senses. The H&F pattern grammar rightly shows that ‘enlist in the army’ (grammar pattern: V *in* n) and “enlist someone’s help” (grammar pattern: V n) have different meanings. Here, the pattern dictionary and the pattern grammar agree. But this is only the tip of the iceberg of verb meaning distinctions. Many competing meanings of verbs have precisely identical patterns in terms of the limited apparatus of grammatical analysis that H&F use, as we have seen. Meaning distinctions very often depend on a distinctive semantic type of one or more of the arguments. For example, *firing bullet from a*

*gun* and *firing a person from a job* can both be described as “V n from n”, using the terminology of pattern grammar. To get the meaning distinction, we need to invoke a more delicate analytic level than mere word classes. The two different meanings of *fire* are activated by differences of semantic type in the direct object slot, namely [[Human]] and [[Projectile]]. This distinction is confirmed by differences of semantic type in the prepositional object slot, namely [[Firearm]] and [[Activity]]. The majority of semantic distinctions for polysemous verbs are of this kind, not the ‘enlist’ kind.

Sometimes, it is the distinction in semantic type of a prepositional object that makes all the difference. For example, consider the verb *sail*. One common use of this verb is to *sail through* something. Here we have a verb + preposition. Is this sufficient evidence to decide the meaning of the verb *sail*? No! It is necessary also to know the semantic type of the prepositional object. Consider the following three examples:

1. In 1577 he set out in the Pelican (afterwards renamed the Golden Hind) for the river Plate, **sailed through** the Straits of Magellan, plundered Valparaíso, rounded the Cape of Good Hope, and completed the circumnavigation of the world .
2. Jeremy Irons, who **sails through** the role with charm and panache.
3. I’ve even heard 12-year-olds **sail through** this work [Samuel Barber’s violin concerto]

The meaning depends on the semantic type of the prepositional object: [[Location]] vs. [[Activity]]. It may be objected that acting roles and violin concertos are not activities. This is true, but irrelevant. It overlooks the fact that what is meant in 2 is the *acting* of the role and in 3 the *playing\_of* Barber’s violin concerto. Playing a concerto is, of course, an activity. These are examples of the kind of semantic coercion, a notion introduced by Pustejovsky (1995). Nouns like *role* and *work* are coerced by the verb+preposition combination *sail through* into having the semantic type of the activity most normally associated with them: acting and playing. This is how the prepositional objects of 2 and 3 activate the ‘accomplish with ease’ sense of *sail through*, while the prepositional object in 1 activates the sense ‘pass through in a boat’. In example 1, “through the Straits of Magellan” is just one of many adverbial of direction governed by the verb *sail* in this sense, while 2 and 3 are much more idiomatic constructions.

Introducing the semantic types of lexical items as an analytic level is necessary, but it unleashes a veritable hailstorm of problems for the lexical analyst. The only dictionary which has even attempted to capture relevant meaning-determining collocations at this level is Cobuild. Cobuild is defective in many ways, but it at least made a start on addressing the question of how word meaning is related to word use. Without corpus evidence and statistical measures of collocational salience, the question cannot be addressed seriously at all.

Patterns are useless if they do not have predictive power. Therefore, for each lexical item in each pattern of each verb, we need to ask. “How likely is it that we will see this word again, as a collocate of our target word, in a comparable expanse of text?” Answers can be computed statistically on the basis of large samples. *CPA (Corpus Pattern Analysis)* regards a pattern as a relationship between sets of collocates. A single word cannot be a pattern.

Verb patterns consist of two or more words in a syntagmatic structure. There is also a paradigmatic element in a verb pattern: the arguments consist of lexical sets of nouns or other words. Typically, these lexical sets are sets of synonyms. Adjective patterns are also syntagmatically structured: the adjective is either a modifier of a particular set of nouns or is related to a set of nouns and structures by a linking verb such as *be* or *seem*. The semantic analysis of adjectives is much like that of verbs in this respect.

Noun patterns, however, do not necessarily have a syntagmatic structure. Significant collocates can be in an unstructured relationship with one another and still function in the same way as structural patterns in assigning probabilities to the selection of a relevant meaning of a target word. To take a simple example, the noun *doctor* has at least two senses: 1) medical practitioner, and 2) bearer of an advanced academic degree. The first sense is much commoner, and is typically distinguished by collocation with any of a very large number of words such as *patient, dentist, surgeon, nurse, treat, symptom, or hospital*. If these words are found anywhere close to the target word (*doctor*), it is a fair bet that the medical sense is the one that should be selected. On the other hand, if *doctor* occurs near words such as *degree, philosophy, divinity, or letters*, the rarer academic sense is more likely to be the correct one.

I hasten to add that in the first sentence of the preceding paragraph emphasis must be placed on ‘not necessarily’. It is undeniable that many nouns, especially nouns that are derived from verbs, do have a syntagmatic structure. But unstructured collocation is a phenomenon more associated with nouns than with verbs.

Rather than prolonging the theoretical discussion, I will conclude this section by quoting some examples of entries from the Pattern Dictionary. I will not discuss the points raised by these entries in any great detail, as a whole workshop would be needed to do that properly. A fuller discussion of the aims of the Pattern Dictionary and a contrastive study with FrameNet and other work will be found in Hanks and Pustejovsky (2005).

Pattern elements vary greatly in scope: a pattern element may be any of the following:

- a) a whole phrase (e.g. an adverbial of direction)
- b) a cluster of nouns sharing the same semantic type or other attribute (e.g. [[Human]]), or
- c) an individual word (typically, individual words are pattern elements of idioms).

The patterns for each verb aim at being mutually exclusive: that is, if the nouns, adjectives and other words that realize each argument of a verb in an unseen clause are assigned to the right semantic type—i.e. the right place in the project’s shallow ontology—then the meaning of the clause as a whole can be identified with reasonable confidence. Meanings are expressed as **implicatures**.<sup>9</sup> Each implicature is ‘anchored’ to the corresponding pattern by replication of pattern elements in both pattern and implicature. Not all pattern elements are replicated in the implicature, however. Idioms, in particular, are very weakly anchored. The converse is also true: occasionally, a pattern element is found only in an implicature and not in the pattern itself. This happens when an argument is strongly implied by a verb even though it is not explicitly present in the clause structure.<sup>10</sup>

The first example is the verb *amble*, discussed above, which has only one sense and one pattern. There are no surprises here. The purpose of showing it is merely to begin to familiarize the reader with the metalanguage of the Pattern Dictionary.

### **amble**

1. PATTERN: [[Human | Animal]] amble [NO OBJ] [A[Direction]]  
PRIMARY IMPLICATURE: [[Human | Animal]] walks slowly and in a relaxed manner in a certain [[Direction]]  
COMMENT: [A[Direction]] is almost invariable present in the syntagmatics, although semantically it is unimportant, as the focus of this verb is on manner of motion, not on the direction of movement.  
EXAMPLE: Two sheep and a goat ambled up over the roof and grazed on its turf.

Notice that the primary implicature is anchored to the pattern by repetition in both places of as many clause roles as possible. In this way, a link is established between meaning and use.

The next example is the entry for the verb *devour*. This, too, is fairly straightforward. It has four lexico-semantic patterns, all of them realizations of the “V n” syntactic structure. Note that the basic sense is manner of eating, not just eating, and this has give rise to two other conventional patterns: a person devouring a book and one institution devouring another.

---

<sup>9</sup> The primary implicature is closest to a dictionary definition. But a verb pattern is a hook onto which any number of secondary implicatures can be hung.

<sup>10</sup> For example, the phrasal verb *bandy words around* implies an exchange of words between two people, acting alternately as audience and utterer, even though normally only one of them is explicitly realized in any given sentence.

### devour

1. PATTERN: [[Human 1 | Animal 1]] devour [[{Animal 2 = Food} | {Physical Object = Food}]]  
 PRIMARY IMPLICATURE: [[Human 1 | Animal 1]] hungrily eats [[Animal 2 = Food | {Physical Object = Food}]]  
 SECONDARY IMPLICATURE: [[Human 1 | Animal 1]] eats all of [[{Animal 2 | Physical Object} = Food]], so that nothing is left  
 EXAMPLE: *Prince Khalid Bin Sultan ... is said to have turned pale when Egyptian commandos **devoured** live chickens and rabbits in a show of bravado. | Here a swarm of common starfish are rapidly **devouring** the carcass of a fish.*  
 FREQUENCY: 58%
2. PATTERN: [[Human]] devour [[Document]]  
 PRIMARY IMPLICATURE: [[Human]] eagerly reads [[Document]]  
 EXAMPLE: *The author's explanation of why people **devour** books about the rich is appropriately cynical.*  
 FREQUENCY: 14%
3. PATTERN [[Human | Institution 1 | Abstract 1]] devour [[Institution 2 | Abstract 2]]  
 PRIMARY IMPLICATURE: [[Human | Institution 1 | Abstract 1]] takes over, uses, absorbs, and destroys [[Institution 2 | Abstract 2]]  
 EXAMPLE: *No peaceful international order is possible if larger states can **devour** their smaller neighbours.*  
 FREQUENCY: 24%

My final example is the verb *scratch*. This is more complex. 14 patterns may be distinguished. Some of the distinctions are quite fine-grained, but they are of vital importance in answering the question “Who did what to whom?”. No distinction is made between semantic and pragmatic implicatures, though secondary implicatures often express pragmatics. As far as CPA is concerned, they are all part of the conventional meaning of these expressions.

### scratch

1. PATTERN: [[Human | Physical Object 1]] scratch [[Physical Object 2]]  
 PRIMARY IMPLICATURE: [[Human | Physical Object 1]] marks and/or damages the surface of [[Physical Object 2]]  
 SECONDARY IMPLICATURE: Typically, if subject is [[Human]], [[Human]] does this by dragging a fingernail or other pointed object across the surface of [[Physical Object 2]]  
 EXAMPLE: *I remember my diamond ring **scratching** the table. | 'I'm sorry sir, but I'm afraid I've **scratched** your car a bit!'*  
 FREQUENCY: 19%

2. PATTERN: [[Human]] scratch [[Language | Picture]] {on [[Inanimate = Surface]]}  
 PRIMARY IMPLICATURE: [[Human]] writes or marks [[Language | Picture]] on [[Inanimate = Surface]] using a sharp edge or pointed object  
 EXAMPLES: *A Turkish schoolboy who had **scratched** the word ‘Marxism’ on his desk.* | *Names of infant Mulverins had recently been **scratched** on the wall.*  
 FREQUENCY: 9%
3. PATTERN: [[Human | Animal]] scratch [[Self | Body Part]]  
 PRIMARY IMPLICATURE: [[Human | Animal]] repeatedly drags one or more of his or her fingernails rapidly across [[Body Part]]  
 SECONDARY IMPLICATURE: typically, [[Human | Animal]] does this in order to relieve itching  
 EXAMPLE: *Without claws it is impossible for any cat to **scratch** itself efficiently.*  
 FREQUENCY: 16%
4. PATTERN: [[Human]] scratch {head}  
 PRIMARY IMPLICATURE: [[Human]] rubs his or her {head} with his or her fingernail(s)  
 SECONDARY IMPLICATURE: often a sign that [[Human]] is puzzled or bewildered  
 EXAMPLE: *He peered down at me and **scratched** his head as he replaced his cap* | *Having just struggled through a copy of the Maastricht Treaty I can only **scratch** my head that anyone would wish to sign it* [METAPHORICAL EXPLOITATION].  
 FREQUENCY: 14%
5. PATTERN: [[Human 1 | Animal 1]] scratch [[Human 2 | Animal 2]]  
 PRIMARY IMPLICATURE: [[Human 1 | Animal 1]] uses the fingernails or claws to inflict injury on [[Human 2 | Animal 2]]  
 EXAMPLES: *Mary was starting to pull her sister’s hair violently and **scratch** her face in anger.*  
 FREQUENCY: 9%
6. PATTERN: [[Inanimate]] scratch [[Human | Animal]]  
 PRIMARY IMPLICATURE: [[Inanimate]] accidentally inflicts a superficial wound on [[Human | Animal]]  
 EXAMPLE: *A nice old Burmese woman brought us limes—her old arms **scratched** by the thorns.*  
 FREQUENCY: 2%

7. PATTERN: [[Bird = Poultry]] scratch [NO OBJ] (around)  
 PRIMARY IMPLICATURE: [[Bird = Poultry]] drags its claws over the surface of the ground in quick, repeated movements  
 SECONDARY IMPLICATURE: typically, [[Bird = Poultry]] does this as part of searching for seeds or other food.  
 EXAMPLE: *A typical garden would contain fruit and vegetables, a few chickens to **scratch** around*  
 FREQUENCY: 3%
8. PATTERN: [[Human]] scratch [NO OBJ] {around | about} {for [[Entity = Benefit]]}  
 PRIMARY IMPLICATURE: [[Human]] tries to obtain [[Entity = Benefit]] in difficult circumstances  
 COMMENT: Phrasal verb.  
 EXAMPLE: *Worrying his head off, **scratching** about for the rent*  
 FREQUENCY: 4%
9. PATTERN: [[Human]] scratch {living}  
 PRIMARY IMPLICATURE: [[Human]] earns a very poor {living}  
 COMMENT: Idiom.  
 EXAMPLE: *destitute farmers trying to **scratch** a living from exhausted land.*  
 FREQUENCY: 6%
10. PATTERN: [[Human 1]] scratch {[[Human 2]]'s {back}}  
 PRIMARY IMPLICATURE: [[Human 1]] helps [[Human 2]] in some way  
 SECONDARY IMPLICATURE: usually as part of a reciprocal helping arrangement  
 COMMENT: Idiom.  
 EXAMPLE: *Here the guiding motto was: you **scratch** my back, and I'll scratch yours—a process to which Malinowski usually referred in more dignified language as 'reciprocity' or 'give and take'.*  
 FREQUENCY: 1%
11. PATTERN: [[Human | Institution]] scratch {surface (of [[Abstract = Topic]])}  
 PRIMARY IMPLICATURE: [[Human | Institution]] pays only very superficial attention to [[Abstract = Topic]]  
 COMMENT: Idiom.  
 EXAMPLE: *As a means of helping Africa's debt burden, ... it barely **scratches** the surface of the problem.*  
 FREQUENCY: 11%

12. PATTERN: [[Human 1]] scratch [[Entity]]  
 PRIMARY IMPLICATURE: [[Human 1]] looks below the obvious superficial appearance of something ...  
 SECONDARY IMPLICATURE: ... and finds that the reality is very different from the appearance.  
 COMMENT: Imperative. Idiom.  
 EXAMPLE: *Scratch any of us and you will find a small child.*  
 FREQUENCY: 2%
13. PATTERN: [[Human | Physical Object 1 | Process]] scratch [[Physical Object 2 | Stuff]] {away | off}  
 PRIMARY IMPLICATURE: [[Human | Physical Object 1 | Process]] removes [[Physical Object 2 | Stuff]] from a surface by scratching it  
 COMMENT: Phrasal verb.  
 EXAMPLE: *First he scratched away the plaster, then he tried to pull out the bricks.*  
 FREQUENCY: 2%
14. PATTERN: [Human]] scratch [[Language | Picture]] {out}  
 PRIMARY IMPLICATURE: [[Human]] deletes or removes [[Language | Picture]] from a document or picture  
 COMMENT: Phrasal verb.  
 EXAMPLE: *Some artists ... use 'body colour' occasionally, especially solid white to give that additional accent such as highlights and sparkles of light on water which sometimes give the same results as scratching out.*  
 FREQUENCY: 1%

## 7. An ontology of shimmering lexical sets

I am arguing here that, in order to understand how meaning in language works, it is necessary to start by analysing verbs in context, using a large corpus. The first step is to distinguish the normal, conventional uses of each verb from abnormal, unusual uses. Abnormal uses are set aside for later analysis. To find the conventional uses of verbs, we first identify the different structural patterns—relationships among clause roles—of the kind described by Hornby and his successors. Then each structural pattern is subdivided according to the semantic types of the words in the clause roles, insofar as these activate different meanings of the verb. This work is greatly facilitated by selection of statistically significant or 'salient' collocates in each clause role. There are now tools (in particular, the Sketch Engine) that make it possible to instantly identify salient collocates in different clause roles and other syntagmatic relationships for any content word in any corpus of any language. Grouping these significant collocates into clusters implies that the nouns (at least) must be grouped into an ontology according to their semantic type. How is this to be done?

An ontology or thesaurus is called for, in which words are organized in a semantic hierarchy: the sort of apparatus first implemented by Wilkins (1668) and subsequently by Roget (1853) and Miller and others (1995). Attempts to adopt existing ontologies for CPA proved unsatisfactory, so currently considerable effort is being put into building a shallow ontology that reflects how nouns are actually used in relation to verbs. A prototype of this ontology is outlined in Pustejovsky et al. (2004). The top type is called [[Anything]]. When this is used in a pattern, it means that absolutely any noun, without reference to its semantic classification, can be used in that particular clause role.

The top levels of the CPA Ontology, in a somewhat simplified and schematized version, look like this:

- Anything
  - Eventuality
    - Event
    - State
  - Entity
    - Physical Object
      - Inanimate
        - Artifact
      - Animate
        - Human
        - Animal
        - Plant
    - Abstract

It will be seen that:

The top type (Anything) is divided into Entities and Eventualities.

Eventualities are divided into Events and States.

Entities are divided into Physical Objects, Abstracts.

Physical Objects are divided into Animates and Inanimates.

... and so on. There are many more subdivisions and interrelationships.

The terms used in the ontology are not to be thought of as English words, but rather as addresses which will be populated with words. Interesting questions arise when we come to populate the addresses with actual words.

An ontology is usually considered to be an ordered set of hyponyms, synonyms, and co-hyponym, which are in a fixed relationship to one another because they share certain properties of meaning.

For example, a bird is a living creature or [[Animate]], so this word and its synonyms and hyponyms belong in the ontology somewhere under [[Animate]]. Synonyms of *bird* are very few: in fact, its only true synonym is the rather archaic word *fowl*. Hyponyms of *bird*, on the other hand, are plentiful. They include *sparrow*, *finch*,

*osprey*, *hawk*, *penguin*, and a very large number of other words. Already, we can sense trouble ahead, for whereas *sparrow*, *finch*, *osprey*, *hawk* all activate a particular sense of the verb *fly*, the noun *penguin* does not; it is more associated with the verbs *swim* and *waddle*. On the other hand, when we are analysing the verb *breed*, *penguin* re-joins the set of entities that breed (in the sense ‘have offspring’). Thus, in relation to different verbs, some members of a lexical set drop out, while, when we move on to a different verb, other members come in. In this sense, a lexical set may be said to “shimmer”. Its membership is not constant, but changeable. Nevertheless, lexical sets of nouns, in a hierarchically organized ontology, are necessary to pick out different meanings of verbs. The hierarchical organization is necessary because different verbs take arguments at different levels of generality.

One important variable that must be mentioned here is the tension between the principle of idiomaticity and the principle of openness. Sinclair (1991) identifies a tension between what he calls the *open-choice principle*:

a way of seeing language as the result of a very large number of complex choices. At each point where a unit is complete (a word or a phrase or a clause), a large range of choices opens up and the only restraint is grammaticalness

and the *idiom principle*:

Many choices within language have little or nothing to do with the world outside. ... A language user has available to him or her a large number of semi pre-constructed phrases that constitute single choices.

Consider the verb *abandon*. The vast majority of uses of this verb represent a simple transitive structure, i.e. S V O. The subject is normally [[Human]], but what about the direct object? You can abandon an activity, plan, or project—all words that belong in the [[Event]] hierarchy—or a refrigerator, a car, or a TV set—words which come under [[Physical Object]]. You can also abandon [[Human]]s, e.g. your friends or your wife and children—and you can abandon a [[Location]] such as a hilltop or a defensive position. You can also abandon something that are [[Abstract]] such as a scientific theory or an ideology. So *abandon* seems to be a good example of an open-choice verb. On the other hand, the implicatures of *abandoning one’s wife and children* are quite different from those of abandoning a hypothesis or a fortress. There is a general overall sense (‘go away from and no longer have anything to do with X’), but there are also a number of specific implicatures associated with different types of thing that are abandoned. So having grouped the direct objects according to their semantic types, the lexical analyst still has to decide whether to lump or split the senses of *abandon*, and if splitting, how delicate the splits should be. There is no simple right-or-wrong answer to this question: the decision must be motivated by the degree of delicacy required by the intended user or application.

Not only is there variation in co-hyponyms when an ontology is applied to real texts, but also there is also variation in focus. Quite often, a noun denoting a part or a property is used in alternation with a noun denoting a whole entity. Take the verb *calm* as an example. Typically, you calm an animate entity such as a person or

a horse, although in fact, only a subset of animate entities normally occurs as the direct object of *calm*. You do not, for example, normally talk about calming *insects* or *spiders*. However, the direct object slot is also very often used to focus on relevant properties of an entity. You can, without change of meaning, calm people's fears or anxieties. *Fear* and *anxiety* are not animate entities; they are properties of animate entities. They focus on the relevant property of the person or animal concerned; they do not activate a different sense.

With other verbs, the focus may be on parts of the whole. You can *repair* a house or a car, but you can also, in the same sense, repair the roof or windows of a house or car, or some other part such as the headlights of a car or the brickwork of a house.

Then there are words whose semantics cut across semantic classes, e.g. *pet*. Some but not all mammals are pets; some but not all birds are pets. Pet is a role assigned to individuals, not a semantic class within a scientific classification of the universe—and yet, nevertheless, there is a fairly distinctive class of pets, with distinctive properties.

Clustering of lexical items in verb arguments is an important (though up to now neglected) topic in lexical analysis. It needs to be matched with the traditional semantics of lexical sets, as found in thesauruses and ontologies. But, as we have seen, quite a sophisticated analytical apparatus will be required to group words into relevant sets, and not all decisions can be made by algorithm: some lexicographical judgement will always be called for.

## 8. Applications

There is neither sufficient time nor space here to engage in a full discussion of all the potential applications of the CPA Pattern Dictionary, but a brief sketch may help to set the project in perspective. It is not intended as a dictionary for foreign learners or, indeed, any ordinary every human user. The apparatus of brackets and implicatures, I am told, looks intimidating to ordinary users, although in fact it is really quite simple.

The main purpose of the project is to provide empirically well-founded links between word meaning and word use. To do this, it proceeds via patterns of use, which can be recognized explicitly and measured. It is, therefore, an infrastructure resource with a great many potential applications. These include:

- Computational natural language understanding systems
- Machine translation (associating meaning with patterns in two languages, rather than words)
- Anomaly detection—distinguishing unusual words and expressions from normal phraseology
- Semantic web—processing meaning in unstructured text
- Natural language generation—idiomatic phraseology

- In lexicography: future dictionaries with a much clearer focus on normal phraseology as well as meaning
- Pedagogical applications, including automatic error identification.

On the computational side, Rumshisky (forthcoming) reports on the use of pattern elements to identify automatically the correct meaning of polysemous verbs in free text. She deals with automatic identification of semantically diverse lexical sets that activate the same sense of the predicate. Semantically diverse nouns are grouped into lexical sets on the basis of association with sets of ‘selectionally equivalent’ verbs (i.e. verbs that share selectional preferences in a given argument position).

## 9. Conclusion

It is a truism that context determines meaning, but it is hard to decide what counts as a meaning and what counts as a context. This is an abiding problem for lexicographers, language learners, translators, and computational language processing alike. The problem of identifying context and meaning in unseen text is the theme of this paper.

A. S. Hornby pointed us in the right direction by drawing attention to the highly patterned nature of language in use and by constructing a framework of structural patterns to which different meanings and different idiomatic uses of each verb could be related. He successfully identified the clause structures involved, though these have subsequently been revised and streamlined by his successors. However, Hornby’s patterns took no account of the semantic types of the arguments of verbs. With the resources that were available during his lifetime, he was not able to go much further than analysis of clauses in terms of clause roles and part-of-speech classes, even if he had wanted to. Since then there have been some improvements in clarity, as well as some retrograde steps such as the substitution of analysis in terms of word classes for analysis in terms of clause roles.

In this paper, I have proposed a return to clause roles (rather than word classes) as an essential first step before proceeding to a more sophisticated analysis which systematically relates word meaning to word use. Future lexicography will, I predict, include projects that are pattern-driven rather than meaning-driven. It will include analysis of verb meaning in relation to the semantic types of clause roles, not merely structural patterns of word classes.

To see how uses of a particular verb in different patterns have different meanings, it is necessary to first find the verb, then correlate the grammatical patterns of the verb with its salient collocates, which are grouped together according to different aspects of their semantics. Current priorities for the Pattern Dictionary project include creation of an empirically well-founded ontology and a methodology for representing entities, their properties, and their parts, as meaning-determining collocates in different argument positions.

Identifying statistically significant collocations, grouping them into patterns, and building an ontology are future tasks for lexicographers, corpus analysts, and computational linguists, working together hand in hand.

## Acknowledgements

I would like to thank the Hornby Trust and the organizers of Euralex 2008 for the honour of inviting me to give this lecture. Thanks are also due to Gill Francis and Gilles-Maurice de Schryver for comments on earlier drafts of this paper.

The research was funded in part by the Academy of Sciences of the Czech Republic (project T100300419) and the Czech Ministry of Education (National Research Program II project 2C06009).

## References

### *Dictionaries*

- [CIDE1]. Procter, P. and others (1995). *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press.
- [CIDE2 (CALD)]. Woodford, K. et al. (2005). *Cambridge Advanced Learner's Dictionary* (= 2nd edition of CIDE). Cambridge: Cambridge University Press.
- [COBUILD1]. Sinclair, J. M.; Hanks, P. and others (1987). *Collins Cobuild English Language Dictionary*. London: HarperCollins.
- [COBUILD2, 3]. Sinclair, J. M.; Fox, G.; Francis, G. and others (<sup>2</sup>1995, <sup>3</sup>2001). *Collins Cobuild English Language Dictionary*. London: HarperCollins.
- [COD1]. Fowler, H. W.; Fowler, F. G. (1911). *Concise Oxford Dictionary of Current English*. Oxford: Oxford University Press.
- [COD8]. Allen, R. and others (1990). *Concise Oxford Dictionary of Current English, 8th Edition*. Oxford: Oxford University Press.
- [ISED]. Hornby, A. S.; Gatenby, E. W.; Wakefield, H. (1942). *Idiomatic and Syntactic English Dictionary*. Kaitakusha. Republished in 1948 by Oxford University Press as *A Learner's Dictionary of Current English*. See OALD.
- [LDOCE1]. Procter, P. and others (1978). *Longman Dictionary of Contemporary English*. Harlow: Longman.
- [LDOCE3]. Rundell, M. and others (1995). *Longman Dictionary of Contemporary English, 3rd edition*. Harlow: Longman
- [LDOCE4]. Bullon, S. and others (2003). *Longman Dictionary of Contemporary English, 'new edition'*. Harlow: Longman
- [MEDAL]. Rundell, M. (2002). *Macmillan English Dictionary for Advanced Learners*. London: Macmillan.
- [NODE (ODE)]. Hanks, P.; Pearsall, J. and others (1998). *New Oxford Dictionary of English*. Oxford: Oxford University Press. (2nd edition 2003 published as *Oxford Dictionary of English*.)

- [OALD2]. Hornby, A. S. and others (1962). *Oxford Advanced Learner's Dictionary of Current English, 2nd edition*. Oxford: Oxford University Press. (Second edition of ISED).
- [OALD3]. Hornby, A. S.; Cowie, A. and others (1974). *Oxford Advanced Learner's Dictionary of Current English, 3rd edition*. Oxford: Oxford University Press.
- [OALD4]. Cowie, A. and others (1974). *Oxford Advanced Learner's Dictionary of Current English, 4th edition*. Oxford: Oxford University Press.
- [OALD5]. Crowther, J. and others (1995). *Oxford Advanced Learner's Dictionary of Current English, 5th edition*. Oxford: Oxford University Press.
- [OALD6]. Wehmeier, S. (ed.) and others (2000). *Oxford Advanced Learner's Dictionary of Current English, 6th Edition*. Oxford: Oxford University Press.

### **Ontologies**

- Leibniz, G. W. (1704). *Table of Definitions*. Excerpts, selected and edited by Emily Rutherford. In Hanks, P. (ed.) (2008). *Lexicology: Critical Concepts in Linguistics*. Vol. 1. London: Routledge.
- Miller, G.; Fellbaum, C. and others (1995). *WordNet*. At <http://wordnet.princeton.edu>.
- Roget, P. M. (1852). *Roget's Thesaurus*. Many subsequent editions and imitations.
- Wilkins, J. (1668). *Essay Towards a Real Character and a Philosophical Language, part II: Universal Philosophy*. London: the Royal Society. Excerpts in Hanks, P. (ed.) (2008). *Lexicology: Critical Concepts in Linguistics*. Vol. 1. London: Routledge.

### **Grammars**

- Aelius Donatus. 4th century AD. *Ars grammatica*.
- Dixon, R. M. W. (1991). *A New Approach to English Grammar, on Semantic Principles*. Oxford: Oxford University Press.
- Francis, G.; Hunston, S.; Manning, E. (1996). *Collins Cobuild Grammar Patterns 1: Verbs*. London: HarperCollins.
- Hornby, A. S. (1954). *A Guide to Patterns and Usage in English*. Oxford: Oxford University Press.
- Hunston, S.; Francis, G. (2000). *Pattern Grammar*. Amsterdam: John Benjamins.
- Priscian. c. 500 AD. *Institutiones grammaticae*.
- Quirk, R.; Greenbaum, S.; Leech, G.; Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

### **Other literature**

- Baker, C. F.; Fillmore, C. J.; Cronin, B. (2003). "The Structure of the FrameNet Database". *International Journal of Lexicography* 16 (3). 281-296.
- Halliday, M. A. K. (1966). "Lexis as a linguistic level". In Bazell, C. E.; Catford, J. C.; Halliday, M. A. K.; Robins, R. H. (eds.). *In Memory of J. R. Firth*. London: Longman. 148-162.
- Hanks, P. (forthcoming). *Lexical Analysis: Norms and Exploitations*. MIT Press.

- Hanks, P.; Pustejovsky, J. (2005). "A pattern dictionary for natural language processing". *Revue Française de Linguistique Appliquée* 10 (2). 63-82.
- Miller, G.; Fellbaum, C. (1985-91). "Five papers on WordNet". At: <ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps>.
- Pustejovsky, J. 1995. *The Generative Lexicon*. Chicago: The MIT Press.
- Pustejovsky, J.; Hanks, P.; Rumshisky, A. (2004). "Automated Induction of Sense in Context". In *Proceedings of COLING 2004*, Geneva, Switzerland.
- Rumshisky, A. (forthcoming). "Resolving Polysemy in Verbs: Contextualized Distributional Approach to Argument Semantics". In Lenci, A. (ed.). *From Context to Meaning: Distributional Models of the Lexicon in Linguistics and Cognitive Science*. Special Issue of the *Italian Journal of Linguistics*.
- Sinclair, J. M. (1966). "Beginning the study of lexis". In Bazell, C. E.; Catford, J. C.; Halliday, M. A. K.; Robins, R. H. (eds.). *In Memory of J. R. Firth*. London: Longman. 148-162.
- Sinclair, J. M. (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

#### **Web sites**

- [CPA]. Corpus Pattern Analysis: <http://nlp.fi.muni.cz/projects/cpa/>.
- FrameNet: <http://framenet.icsi.berkeley.edu/>.



# Twenty-five Years of Dictionary Research: Taking Stock of Conferences and Other Lexicographic Events since LEXeter '83

Reinhard R. K. Hartmann

## Introduction

Events like this Congress are not only useful to all of us for the purpose of meeting people, exchanging information and learning new things about our subject, but occasionally they give us food for thought, and make us aware of the whole scene beyond our working context.

I have been asked to address this topic, as 25 years have passed since EURALEX was first established at a conference called LEXeter '83, which I had the pleasure of organising. I am not going to attempt to review the entire history and progress of (meta-)lexicography since then, but I will offer some personal reflections, first by emphasising the value of such conferences and then by branching out to a few wider issues, such as associations and their venues, lexicography centres and research networks, research perspectives and dictionary projects. Finally, I will consider some priorities, such as how to obtain a fuller picture of our discipline and its development, with special attention to Europe.

To help me stay on track and not to be distracted by too many irrelevancies, I have condensed the most important facts to be discussed in the form of six tables, which will give selected examples of some of the points that I am going to make, hopefully demonstrating what their implications might be for lexicography.

## Conferences

Conferences are important, in a number of ways. They allow us to share information, in terms of explaining what we are doing and of learning what others are doing, thus helping to reduce our relative isolation; they enable us to get to know each other better and to be inspired by other people, all of which can encourage us to promote change and to improve theory and practice all-round.

Place	Year	Topic(s)	Series
Copenhagen	1982 ff.	general lexicography	Symposium 1-13 [2007]
Exeter [et al.]	1983 [ff.]	general lexicography	Conference and other meetings → EURALEX Congress 1-13 [2008]
Heidelberg	1986 [ff.]	(German) historical lexicography	Konferenz → Heidelberg Lexikographisches Kolloquium 1-4 [2000] and other meetings
Balatonfüred [et al.]	1990 ff.	computational lexicography	Conference 1-7 [2003]
Jaén	1991 ff.	Hispanic lexicography	Seminario 1-6 [2003]
Paris / Cergy	1993 ff.	general / bilingual lexicography	Journée / Colloque 1-16 [2008]

Table 1: (Influential) Conferences

In the early 1980s time was obviously ripe for conferences, so we began to experience not only a rise in the number of individual meetings, but even of some that turned into ‘conference series’ (see Table 1). One of the first of these were the Copenhagen Symposia, which came about through the rather unusual interdisciplinary collaboration between the professors of German and English at Copenhagen University (Hyldgaard-Jensen & Zettersten 1983), one reason why interlingual topics have tended to dominate many of these meetings. Nine years after the first of a total of 13 of these Lexicography Symposia, the Nordic Association for Lexicography emerged (see below).

Just over a year after the Copenhagen Symposium No. 1 came LEXeter ’83, one of the most important events in my own personal experience (Hartmann 1984). I had invited six scholars as plenary speakers: Herbert Ernst Wiegand (from Heidelberg) and John Sinclair (from Birmingham) for specifying the contents and outer edges of our discipline, Ladislav Zgusta (from Illinois) and Tony Cowie (from Leeds) for illustrating progress in bilingual and EFL lexicography, respectively, and Frank Knowles (from Aston in Birmingham) and Juan Sager (from UMIST in Manchester) for covering aspects of computing and terminology work.

Among the various projects that started at LEXeter ’83 was that of the international encyclopedia of lexicography *Wörterbücher/Dictionaries/Dictionnaires (W/D/D)*, the double series *Lexicographica. Series Maior* (of which the Exeter proceedings became Volume No. 1) and *Lexicographica. International Annual for Lexicography*. The most important innovation was the establishment of a European association (see below).

Three years later came the conference on historical lexicography at Heidelberg (Wiegand 1987), where several of the important historical dictionary projects in Europe were represented. Other meetings followed at Heidelberg, such as the annual local ‘Lexikographisches Kolloquium’ (which constituted 4 more volumes in the *Lexicographica Series Maior*).

The three conferences at Copenhagen, Exeter and Heidelberg were not the only ones held at the time; there were at least 6 in Europe in the period from 1982 to 1983: 2 in West Germany, 1 in East Germany, 1 in Italy, 1 in Czechoslovakia and 1 in Yugoslavia; outside Europe, there were at least 4: 1 in the United States, 1 in Barbados, 1 in India and 1 in Australia. There were even other conference series beginning, even before then. One that had preceded the Heidelberg conference by nearly 30 years started with a colloquium at Strasbourg, held in 1957, 3 years before the launch of the *Trésor de la langue française* (TLF), which led to 2 more so-called 'Round-table' meetings on historical dictionaries, first at Firenze in 1971 and then in Leiden in 1977.

And remarkably, quite a few years later, this theme was taken up again, first by another impressive single conference at Heidelberg (Städtler 2003) and then by another conference series that started in Leicester in 2002 (Coleman & McDermott 2004) and reappeared at two-year intervals, first in Italy, then in the Netherlands and then in Canada, earlier this year.

In the 1990s, there were Hungarian, Spanish and French initiatives which are also listed in Table 1: the CompLex series of 7 conferences launched by Ferenc Kiefer at Balatonfüred, the 6 Seminarios at Jaén organised by Ignacio Ahumada Lara, and the 16 Journées/Colloquia at Paris and Cergy-Pontoise established by Jean Pruvost, which have now spilled over to Germany, Italy, Spain and Canada to form an influential multiple conference series.

## Associations

Table 2 documents some of the societies and associations that have been set up in the last 25 years, concentrating on the ones in Europe.

Associations	founded	Members	Meetings	Proceed.	By-products
EURALEX	1983	240	13 biennial congresses	13	<i>LexSM, Newsletter, IJL, Who's Who</i>
NFL	1991	250	10 biennial conferences	10	<i>NLO, LexicoNordica</i>
BLS	1995	41	5 biennial conferences	2?	<i>Leksikografski pregled</i>
EAFT	1996	90	3 biennial summits	?	List of organizations
AELex	2002	35	3 biennial congresses	3	<i>Revista de Lexicografía</i>
ATL	2004	28	4 annual conferences	4	<i>Terminology &amp; lexicography</i>

Table 2: (European) Associations

EURALEX has a most impressive record: 13 congresses, all held in different places in different countries, all with published proceedings, amounting to a total of over a thousand papers. Not only the conference papers were important, but also

other things called 'by-products' in Table 2, such as the EURALEX *Newsletter*, the *International Journal of Lexicography* (from 1988), the *Who's Who in Lexicography* (1996) and several EURALEX Seminars and Surveys, e.g. the meeting on learners' dictionaries arranged by Tony Cowie at Leeds (Cowie 1987) and several investigations of dictionary use carried out by Sue Atkins, Krista Varantola and Frank Knowles (Atkins 1998).

At the 3<sup>rd</sup> EURALEX Congress in Budapest, as part of a panel discussion, I presented a paper reviewing the progress made at 65 lexicography meetings, from the famous Bloomington IN conference in 1960 (Householder & Saporta 1962) to the BudaLEX '88 Congress (Magay & Zigány 1990), in terms of the topics discussed in a total of 1,317 papers. The slightly sceptical conclusion I came to then was (Hartmann 1990: 573) that "(C)onferences are no guarantee for reducing the barriers to communication: sometimes they can create new barriers", as it is often very difficult to see how the personal messages of individual papers fit into the respective overall programme, and how the overall contents of the proceedings progress forward to those of the next meeting.

In Table 2, five other associations (and their special features) are listed: the Nordisk Forening for Leksikografi (NFL), the Bulgarian Lexicographic Society (BLS), the European Association for Terminology (EAFT), the Association for Spanish Lexicography (AELEX), and the British Association for Terminology and Lexicography (ATL). I have been fortunate, indeed, to be associated with the foundation of EURALEX, but I have also enjoyed the honour of being present at the initiation of two other international associations *not* mentioned in Table 2: AFRILEX (at Stellenbosch in South Africa, in 1995) and ASIALEX (at Hong Kong, in 1997), each with their own succession of conferences, and in between, I have also attended three of the biennial meetings of the Dictionary Society of North America (DSNA, established 1977), but none of those of the Australasian Lexicography Association (AUSTRALEX, founded 1990).

Genres of publications other than conference proceedings have emerged one after the other, such as textbooks, bibliographies of the lexicographic literature, bibliographies of dictionaries, and even a bibliography of dictionary bibliographies (Cop 1990); and there are also some websites offering relevant information, but none of these make up for the deficit of inadequate bibliographical treatment of dictionaries, as I found out when I tried (in Hartmann 2006) to establish how many onomasiological dictionaries and thesauruses had been published for a range of about 20 European languages.

What I particularly like about the genre of *Festschrift* volumes is the fact that they often contain information on the dedicatees and their numerous contacts with colleagues and students at their own institutions (and elsewhere), details which may not be available from other sources. Many scholars with EURALEX connections have been honoured in this way, such as Arne Zettersten at Copenhagen, Herbert Ernst Wiegand at Heidelberg, John Sinclair at Birmingham, Juan Sager at Manchester, Martin Gellerstam at Göteborg, Olga Karpova at Ivanovo and Paz

Battaner at Barcelona. Yet another text genre, dictionaries of the terminology of lexicography such as the *DoL*, I can only mention in the references.

### Dictionary Research Centres

My next topic is Dictionary Research Centres (other names are used too, for these bodies, such as ‘institute’, ‘department’, ‘laboratory’ and ‘group’, as shown in Table 3). The DRC at Exeter is not listed, as it was moved to Birmingham in 2001, just before my retirement. Others include Aarhus in Denmark, Poznań in Poland, Barcelona in Spain, Cergy-Pontoise in France, and Göteborg in Sweden, each with their special projects, meetings, and publications.

Institutions	founded	Leaders	M.A./ M.Phil.	Ph.D.	Special features
Center for Lexikografi, HHS, Aarhus Universitet	1996	H. Bergenholtz	50	8	Danish LSP dictionaries
Department of Lexicology & Lexicography, AMU Poznań	1996	A. Adamska- Sałaciak	25	3	Bilingual dictionaries
Grup InfoLex, UPF, Barcelona	1996	P. Battaner, J. DeCesaris	9	6	Terminology & lexicography
DRC (& CCL), Birmingham University	2001	R. Moon	14 + 2	2	Corpus lexicography
Métadif, Université de Cergy-Pontoise	2002	J. Pruvost	30	15	French dictionaries, Journées
Lexikaliska Institutet, Göteborgs Universitet	2003	S.-G. Malmgren	4	4	Swedish dictionaries

Table 3: (Pioneering) Dictionary Research Centres

I have not included a few others, because they are not exclusively dedicated to lexicography (such as Copenhagen in Denmark, Erlangen-Nürnberg in Germany and Lyon 2 in France), or because their founders are in the process of retiring (Béjoint at Lyon and Hausmann at Erlangen), or because the staff specialising in dictionary projects are not marked out by such an institutional title (as at the Universities of Heidelberg in Germany and Oslo in Norway, the Vrije Universiteit Amsterdam in the Netherlands, the Istituto Linguistica Computazionale at Pisa in Italy or the Institut für Deutsche Sprache at Mannheim in Germany).

I cannot go into details about the difficulties of running a Dictionary Research Centre (such as the six listed in Table 3). The main problem seems to me to be the job of building bridges between theory and practice, between monolingual and bilingual lexicography, between historical and pedagogical dictionaries, between general and terminological lexicography, and between academic projects and commercial

interests. This means, in turn, having to keep in touch through so-called ‘networks’, a factor which has been acknowledged by a number of institutions.

One of the most important considerations is the training of future lexicographers, e.g. through M.A. and Ph.D. programmes. At Exeter, the first EURALEX Congress helped to set up not only the DRC, but also an M.A. programme partly supported by European funding, such as the *ERASMUS* project between 1990 and 1993, which brought together a consortium of universities interested in a new M.A. course in Lexicography, and the *Thematic Network Project in the Area of Languages* between 1996 and 1999, whose Sub-Project No. 9 was devoted to Dictionaries in Language Learning (Hartmann 1999). This helped to promote contacts between those who were involved, especially from places like Exeter in the United Kingdom, Lille in France, Amsterdam in the Netherlands, Gent in Belgium, Tampere in Finland, Aarhus in Denmark, Lisbon in Portugal, and Thessaloniki in Greece.

In the last few years I have built up a (‘LexiDiss’) database of over 1,000 dissertations at M.A. and Ph.D. levels from universities around the world. I have put down the respective figures for the six Dictionary Research Centres listed in Table 3, together with some of their ‘special features’, such as the kinds of reference works produced or debated there. However, it must be admitted that (a) such higher degree dissertations are often ignored, especially if they have not been published in book form, although they constitute one of the most important and informative ways in which original dictionary research can be carried out, and (b) in addition to such postgraduate research, undergraduate courses providing training on aspects of lexicographic practice are also essential (both were surveyed by Edward Gates in 1997).

### **Metalexigraphic frameworks**

To set the scene for the next section, we can use a paper by Franz Josef Hausmann (1989), where he put forward the argument that metalexigraphy, or the theory of lexicography, is much older than we might think, particularly if we take into account the fact that some of the relevant texts go back quite a long time, such as critical accounts of important dictionary projects (like Paulo Beni’s *Anti-Crusca* 1612), prefaces of dictionaries (like Johnson’s *Dictionary of the English Language* 1755), articles in encyclopedias (like D’Alembert’s on ‘dictionnaire’ in the French *Encyclopédie* 1754), and various monographs since the 1930s, which Hausmann lists in his bibliographical references.

Table 4 gives a brief summary of the six main components or branches or perspectives of dictionary research, together with some representative names (for more on these, with several sub-divisions or aspects of each, cf. Hartmann 2001: 41ff. or Hartmann 2003: 2-7).

Perspectives	Topics	Pioneers	Relevant Texts
Dictionary criticism	Evaluating quality	P. Beni (1612)	Wiegand (1998-2005)
Dictionary history	Tracing traditions	J. Murray (1900)	Katz (1998)
Dictionary typology	Classifying genres	L. V. Ščerba (1940)	Landau (2001)
Dictionary structure	Formatting information	J. Dubois (1962)	Bergenholtz & Tarp (1995)
Dictionary use	Observing reference acts	C. Barnhart (1962)	Lew (2004)
Dictionary IT	Applying computer aids	R. Busa (1971)	Pruvost (2000)

Table 4: (Metalexigraphic) Research Perspectives

There is space only for a few brief comments on each of these six perspectives:

The first is *dictionary criticism*, or the totality of efforts to evaluate the quality of dictionaries. In spite of its relatively long history (I have already mentioned Beni's reaction to the VAC 1612), this perspective is still a rather underdeveloped specialisation. One relevant contribution to this field consists of four critical volumes: the first two (Wiegand 1998, 2002) on the German learner's dictionaries published by Langenscheidt (LGW 1993) and De Gruyter (WDF 2000), and the second two (Wiegand 2003, 2005) on the German commercial dictionary published by Duden (DGW 1999). For both ventures, Wiegand had asked a total of 100 experts to provide critical comments on specific features of these 3 dictionaries under 29 chapter headings, such as grammar, semantics, usage labels and text structure, quite a unique effort which has not been attempted for many other languages (it is necessary, of course, to consider the relative status of languages such as German vis-à-vis a global language like English, on the one hand, and smaller languages, like Norwegian and Croatian, on the other).

The second perspective is *dictionary history*. The majority of such historical accounts deal with traditional dictionary genres like Johnson's DEL or Murray's OED, or dictionaries like those of the Italian or the French or the Spanish Academies, but there is still a need to broaden the treatment (as shown by Katz 1998) of 'reference sources' such as encyclopedias, books of quotations, almanacs, yearbooks, manuals, maps, biographies, bibliographies and even government documents.

The third perspective, *dictionary typology*, often follows the tendency of defining the dictionary as an account of the origin of a language's vocabulary, in the form of the so-called historical or etymological dictionary. We owe it to the Russian Lev Vladimirovič Ščerba (1940) that a wider range of reference works is being covered nowadays by the notion of 'dictionary'. What Ščerba did was to bring order into the classification of dictionary genres, in terms of six abstract dichotomies or binary oppositions, but he (and much more recently, scholars like Katz and Wiegand) have demonstrated that genres like the general commercial dictionary deserve as much

attention as the historical dictionary, that pedagogically-oriented dictionaries and LSP/terminological dictionaries are as important and useful as general-purpose dictionaries, and that bilingual dictionaries can be as essential as monolingual dictionaries. And the textbooks on lexicography, like Sidney Landau's (2001), are also beginning to recognise these realities.

The perspective of *dictionary structure* also needs much more attention. One pioneer in this field was the French lexicographer Jean Dubois who argued (back in 1962) that the dictionary could be approached as text or communicative discourse and, therefore, could be analysed and processed with the means of linguistic science. There have been several attempts since then to isolate the ways in which information is formatted in dictionaries, such as 'microstructure' (or entry design) and 'macrostructure' (or overall organisation) to describe the structural design and complexity of reference works.

In Article 36 of the encyclopedia *W/D/D*, Hausmann & Wiegand (1989) made several more distinctions which have been absorbed into the metalexigraphic literature, such as the textbook edited by Bergenholtz & Tarp (1995), so that today we have a whole hierarchy of notions, ranging from 'microstructure' (or entry formatting) to 'macrostructure' (or lemma-list) and 'megastructure' (or the combination of 'macrostructure' and 'frame structure', or 'outside matter'), and then on to 'mediostructure' (or cross-reference systems), 'distribution structure' (or relative stress on linguistic or encyclopedic information) and 'access structure' (or indexing). As Bergenholtz & Tarp and their five co-authors demonstrate in relation to LSP lexicography, a better understanding of structural features should also benefit the other perspectives of dictionary research.

The fifth perspective, *dictionary use* (or the *user perspective*, as it is often called nowadays), probably started with the famous paper by Clarence Barnhart delivered at the 1960 conference at Bloomington, Indiana, which suggested (Barnhart 1962: 161)

(that) it is the function of a popular dictionary to answer the questions that the user of the dictionary asks, and dictionaries on the commercial market will be successful in proportion to the extent to which they answer these questions of the buyer.

The implication of this is, firstly, that the complexities of dictionary structure often play a significant part, and secondly, that it is our duty to find out what structural and other problems the user has and whether (and how) his/her reference acts can be observed and improved. Several of my doctoral students have contributed answers to these questions, and there were also several meetings, such as the Exeter BAAL Seminar on the user perspective 1978 (Hartmann 1979), Tony Cowie's 1985 EURALEX Seminar on learner lexicography and the EURALEX Survey initiated by Sue Atkins. A good overview over the user perspective is provided by Robert Lew's book (2004) *Which Dictionary for Whom?*

Finally, there is the perspective of *dictionary IT* (and several other alternative names for applying electronic aids to lexicography). One pioneer in this area was Roberto Busa, an Italian Jesuit working on Classical texts (such as Thomas Aquinas), who, in his contribution to the *Encyclopedia of Library and Information Science* (Busa 1971) on the subject of ‘concordance-making’, acknowledged the benefits of computers for such text-processing techniques.

Among the increasing range of manuals on offer, Jean Pruvost’s book *Dictionnaires et nouvelles technologies* (2000) is a useful text (see below on the wider context of interdisciplinary collaboration).

### **Towards *reference science***

We turn next to the issue of where metalexicography is moving as a discipline. One notion that has intrigued me for quite a few years is the possibility of an overarching field which might be labelled ‘reference science’, defined by McArthur (1998: 218) as “the study of all aspects of organizing data, information, and knowledge in any format whatever, for any purpose whatever, using any materials whatever”, and identifying at least three sub-fields, *lexicography* (or dictionary-making in the narrow sense), *encyclopedics* (or the production of encyclopedias and other general reference works such as atlases, gazetteers and almanacs), and a third which does not have a name yet but covers tabulations (such as time-tables), directories (such as telephone books), catalogues and other compendia, for all of which the contribution of information technology is vital.

Sven Tarp (2007: 178) has recently expressed similar ideas, demanding that lexicography must, in order to meet the new challenges, “... project itself far beyond its traditional limits”, and suggesting the name ‘infology’ or ‘informology’ for what I would prefer to call ‘reference science’. As a result of such developments, we would have *reference professionals* producing *reference works* for people with *reference needs* and *reference skills*, just what is needed.

Table 5 is based on recent lecture (Hartmann forthcoming a). Arranged from bottom to top, from ‘basement’ to ‘roof’, it lists the requirements that must be met so that a subject field can be regarded as a scholarly discipline. The important question is: has lexicography (or reference science) ‘arrived’ in the academic community as a discipline?

Criteria	Examples
Institutions	Associations, academies, research centres, publishers ...
Modes of discourse	Textbooks, conference proceedings, journals, monographs, reference works ...
Methods	Data-collection (corpus work), surveying (observation), testing (experimentation) ...
Perspectives	Critical p., historical p., typological p., structural p., user p., IT perspective ...
Body of knowledge	Professional processes (compilation phases), text typology, text structure ...
Subject-matter	(a) Practice (fieldwork, description, presentation) (b) Theory (stock-taking, factor analysis, principles)

Table 5: (Chief) Criteria for Disciplinary Status

Our field certainly seems to meet all the specified criteria:

- *firstly*, it has a so-called ‘subject-matter’, both in terms of the practical activities and the theoretical principles that may underlie them,
- *secondly*, it has a ‘body of knowledge’ that emerges from the professional processes as well as the development of lexicographic traditions,
- *thirdly*, we have seen that there are a number of research ‘perspectives’ (which were listed in Table 4),
- *fourthly*, there are ‘methods’ to consider, not so much the working procedures used in dictionary compilation, but the methodological tools needed to investigate lexicographic facts, such as data collection, observation and testing, which differ from one perspective to another (cf. Hartmann forthcoming c),
- *fifthly*, there are appropriate ‘modes of discourse’, such as conference proceedings, textbooks, journals, monographs, festschriften and reference works,
- and *sixth* and lastly, there are the ‘institutions’ in which lexicography and dictionary research is carried out (such as the DRCs listed in Table 3).

Having decided what sort of discipline lexicography or metalexicography constitutes, we need to be aware of all the various ways in which it can or needs to cooperate with other disciplines. I find it useful to distinguish between ‘mother’ disciplines, ‘sister’ disciplines’, ‘daughter’ disciplines, and ‘data-supplying’ disciplines.

- *Mother disciplines* might include linguistics, semiotics, and ‘reference science’.
- *Sister disciplines* include lexicology, terminology, translation studies, language teaching, library science, and media studies.

- *Daughter disciplines* include indexing, word-processing, printing, and publishing.
- Some people would argue that IT could be considered a ‘mother discipline’, others might say that it is a ‘sister discipline’, still others might even suggest that it should be regarded as a ‘daughter discipline’.
- *Data-supplying* disciplines include those fields which supply the encyclopedic facts and technical terms used in specialised subjects, such as language studies, dialectology, art, philosophy, social sciences, medicine and many, many others.

All this requires the promotion of various forms of interdisciplinary collaboration, e.g. in joint research projects and even ‘research networks’, as mentioned above.

## Conclusion

I hope that I have managed to draw your attention to some pioneering conferences, individual scholars and collective associations. Then, widening the scope to the topic of dictionary research centres, I talked about what the perspectives of dictionary research are (or should be), and what the criteria and limits of disciplinary status are for lexicography, asking whether there might be something like a new ‘reference science’ emerging.

One conclusion that it would be fair to draw is that it is difficult to make wide generalisations, as the situational contexts in which lexicographic practices and theories are pursued do still vary quite a lot, by country, by language, by cultural tradition, by dictionary type, by educational institution, by publisher, and even by individual lexicographer (for more on desiderata in dictionary research, cf. Hartmann forthcoming c).

One other conclusion arises from this realisation, which I came to during the period since I was invited to address this meeting. For quite some time, I had been collecting lists, not only of conferences, associations and dictionary research centres, but also of dictionaries, lexicographers, dissertations and bibliographical references. During the last 12 months or so, I have started to combine all of these into what I now call an ‘International Directory of Lexicography Institutions’. I do not have time to go through Table 6 in detail, but I hope that you find this comparative if still incomplete extract useful, and that you will contact me and let me have details about your own institution(s).

Institutions	Internat.	DE	ES	FR	GB	IT
Associations	EURALEX EAFT (+6)	— DTT (+1)	AELex AETer (+2)	— SFT (+2)	— ATL (+4)	— AITerm (+1)
Academies, Language Boards etc.	Nordterm, InfoTerm (+6)	DFG. HAW, IdS (+ 6)	CSIC, RAE, IEC (+4)	CNRS, AF, ATILF (+5)	AHRC, OED, BNC (+5)	CNR, AdC, ILC (+6)
DRCs at Universities, Colleges	—	Erl-N'b, H'berg (+5/83)	UPF, UdC (+3/62)	Lyon2, Cergy (+4/79)	B'ham, G'gow (+4/117)	Torino, Milano (+2/46)
Publishers	—	Duden, L'sch (+?/79)	Gredos, VoxBi (?/62)	Larous, LeRob (+?/86)	Oxford, ChaHar (+?/137)	Zing, Zani (+?/70)
Dictionaries						
historical	—	DWB	DHLC	TLF	OED	VAC
general	—	DGW	DLE	DFC	ODE	DLI
learners' (lex.) term.	— NLO	LGW W/D/D	DSAL DLP	RJUN —	COBLD DoL	DAIC —
Total OCLC	...?	18,603	14,180	21,027	88,263	6,574
Periodicals	IJL, LexN, LexIA, Term (+3)	GLing, LES (+1)	RLex (+?)	CLex, ÉtLex (+2)	RefRev (+1)	SLI (+?)

Table 6: (International) Directory of Lexicography Institutions

There are two implications of all this: Should there perhaps be a section at the next EURALEX Congress devoted to 'reference science', and another one on the possible uses of an 'international directory'?

## Bibliographical references

### *Dictionaries and other reference works:*

[COBUILD]. *Collins COBUILD English Language Dictionary* ed. by J. Sinclair [et al.]. London & Glasgow: Collins, 1987. [2<sup>nd</sup> ed. 1995].

[DDEE]. *Dictionary of Dictionaries and Eminent Encyclopedias* comp. by T. Kabdebo & N. Armstrong. London: Bowker-Saur, 1997.

[DAF]. *Dictionnaire de l'Académie française*. Paris: J.B. Coignard, 1694.

[DEL]. *Dictionary of the English Language [in which the words are deduced from their origins ...]* comp. by S. Johnson. London: W. Strachan, 1755.

- [DLP]. *Diccionario de Lexicografía Práctica* comp. by J. Martínez de Sousa. Barcelona: VOX Bibliograf.
- [DGW]. *Duden. Das große Wörterbuch der deutschen Sprache in zehn Bänden* comp. by Wissenschaftlicher Rat der Dudenredaktion and ed. by G. Drosdowski. Mannheim: Bibliographisches Institut/Dudenverlag [1<sup>st</sup> ed. 6 volumes], 1976-1981, [3<sup>rd</sup> ed. 10 volumes, 1999].
- [DoL]. *Dictionary of Lexicography* comp. by R. R. K. Hartmann & G. James. London: Routledge, 1998/2001.
- [DSAL]. *Diccionario Salamanca de la lengua española*. Madrid: Universidad de Salamanca-Santillana, 1996.
- [ELIS]. *Encyclopedia of Library and Information Science* ed. by A. Kent & H. Lancour. New York: CRC Press [33 volumes], 1971.
- [ENC]. *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers* ed. by J. Le Rond d'Alembert & D. Diderot. Paris: Briasson et al., 1751-1772.
- [LDBT]. *Lexicography. A Dictionary of Basic Terminology* comp. by I. Burkhanov. Rzeszów: WSP, 1998.
- [LGW]. *Langenscheidts Großwörterbuch Deutsch als Fremdsprache* ed. D. Götz et al. Berlin & München: Langenscheidt, 1993.
- [NLO]. *Nordisk leksikografisk ordbok* comp. by H. Bergenholtz et al. (Skrifter utgitt av NFL 4). Oslo: Universitetsforlaget AS, 1997.
- [OED]. *Oxford English Dictionary* ed. J. Murray, Robert Burchfield, J. Simpson [et al.]. Oxford: Clarendon Press [1<sup>st</sup> ed. 10 volumes] 1884-1928, [12 volumes and supplement 1933, [4-volume supplement] 1972-1986, [2<sup>nd</sup> ed. 20 volumes] 1989, 3<sup>rd</sup> ed. in preparation.
- [TLF]. *Trésor de la langue française. Dictionnaire de la langue du XIX<sup>e</sup> et du XX<sup>e</sup> siècle* ed. by P. Imbs et al. Nancy: Gallimard [16 volumes] [from 1960, published from] 1971-1994 [electronic version: [www.inalf.fr/tlfi](http://www.inalf.fr/tlfi)].
- [VAC]. *Vocabolario degli Accademici della Crusca ...* Firenze: Accademia della Crusca 1612, 1623, 1691, 1738, 1863-1923.
- [WDF]. *[De Gruyter] Wörterbuch Deutsch als Fremdsprache* ed. by G. Kempcke. Berlin: W. de Gruyter, 2000.
- [W/D/D]. *Wörterbücher/Dictionaries/Dictionnaires. Ein internationales Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie internationale de lexicographie* (Handbücher zur Sprach- und Kommunikationswissenschaft Vol. 5.1, 5.2, 5.3). Ed. by F. J. Hausmann et al. Berlin: W. de Gruyter 1989-1991, Supplementary Vol. 5.4 *Dictionaries. An International Encyclopedia of Lexicography: Recent Developments with Special Focus on Computational Lexicography* ed. by R. Gouws et al. [in preparation].
- [WNT]. *Woordenboek der Nederlandsche Taal* ed. by M. de Vries et al. Leiden: Instituut voor Nederlandse Lexicologie & A.W. Sijthoff [40 volumes], 1864-1998.
- [WWL]. *Who's Who in Lexicography. An International Directory of EURALEX Members* comp. by S. McGill. Exeter: University Dictionary Research Centre, 1996.

### *Other literature*

- Atkins, B. T. S. (ed.) (1998). *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*. Tübingen: M. Niemeyer.
- Barnhart, C. L. (1962) "Problems in editing commercial monolingual dictionaries". In Householder, F. W.; Saporta, S. (eds.). *Problems in Lexicography* [Conference at Bloomington 1960]. Bloomington: Indiana University Press. 161-181
- Beni, P. (1612). *L'Anticrusca ovvero il paragone del italiana lingua* [Reprint ed. by G. Casagrande]. Firenze: Accademia della Crusca [2 volumes 1982, 1983].
- Bergenholtz, H.; Tarp, S. (eds.) (1995). *Manual of Specialised Lexicography*. Amsterdam: John Benjamins.
- Busa, R. (1971) "Concordances". *ELIS* 5. 592-604.
- Coleman J.; McDermott, A. (eds.) (2004). *Historical Dictionaries and Historical Dictionary Research* [International Conference, Leicester 2002]. Tübingen: M. Niemeyer.
- Cop, M. (1990). *Babel Unravelled. An Annotated World Bibliography of Dictionary Bibliographies, 1658-1988*. Tübingen: M. Niemeyer.
- Cowie, A. (ed.) (1987). *The Dictionary and the Language Learner* [EURALEX Seminar 1985]. Tübingen: M. Niemeyer.
- Dubois, J. (1962). "Recherches lexicographiques: esquisse d'un dictionnaire structural". *Études de linguistique appliquée* 1. 43-48
- Gates, J. E. (1997). "A survey of the teaching of lexicography, 1979-1995". *Dictionaries* 18. 66-91. [Reprinted in Hartmann, R. R. K. (ed.) (2003). *Lexicography: Critical Concepts*. London: Routledge. I, 124-147.]
- Hartmann, R. R. K. (ed.) (1979). *Dictionaries and Their Users. Papers from the 1978 B.A.A.L. Seminar on Lexicography*. Exeter: U. of Exeter Press.
- Hartmann, R. R. K. (ed.) (1983). *Lexicography: Principles and Practice*. London: Academic Press
- Hartmann, R. R. K. (ed.) (1984). *LEXeter '83 Proceedings* [EURALEX 1<sup>st</sup> Congress]. Tübingen: M. Niemeyer.
- Hartmann, R. R. K. (ed.) (1986). *The History of Lexicography. Papers from the DRC Seminar at Exeter, March 1986*. Amsterdam: John Benjamins.
- Hartmann, R. R. K. (1988) "The learner's dictionary: Traum oder Wirklichkeit?". In Hyldgaard-Jensen, K.; Zettersten, A. (eds.). *Symposium on Lexicography III* [Proceedings of Copenhagen 1986]. Tübingen: M. Niemeyer. 215-235.
- Hartmann, R. R. K. (1990). "A quarter of a century's lexicographical conferences (symposium)". In Magay, T.; Zigány, J. (eds.) (1990). *BudaLEX '88 Proceedings* [EURALEX 3<sup>rd</sup> Congress]. Budapest: Adadémiai Kiadó. 569-575
- Hartmann, R. R. K. (1994). "The bilingualised learner's dictionary. A transcontinental dialogue on a relatively new genre". In James, G. (ed.). *Meeting Points in Language Studies. A Festschrift for Ma Tailai. Working Papers*. Hong Kong: The Hong Kong University of Science and Technology Language Centre. 171-183

- Hartmann, R. R. K. (ed.) (1999). *Dictionaries in Language Learning. Recommendations, National Reports and Thematic Reports from the TNP Project 9: Dictionaries*. At: [www.fu-berlin.de/elc/TNPproducts/SP9dossier.doc](http://www.fu-berlin.de/elc/TNPproducts/SP9dossier.doc).
- Hartmann, R. R. K. (2001). *Teaching and Researching Lexicography*. London: Pearson Education
- Hartmann, R. R. K. (ed.) (2003). *Lexicography: Critical Concepts* [3 volumes]. London: Routledge.
- Hartmann, R. R. K. (2005). "Pure or hybrid? The development of mixed genres". *Facta Universitatis* 3. 193-208.
- Hartmann, R. R. K. (2006). "Onomasiological dictionaries in 20<sup>th</sup>-century Europe". *Lexicographica. International Annual for Lexicography* 21/2005. 6-19.
- Hartmann, R. R. K. (2007). *Interlingual Lexicography. Selected Essays on Translation Equivalence, Contrastive Linguistics and the Bilingual Dictionary*. Tübingen: M. Niemeyer
- Hartmann, R. R. K. (forthcoming a). "Promoting interdisciplinary collaboration between Lexicology, Lexicography, Terminology and Translation: Towards Reference Science?" [Plenary lecture at 2007 Palermo COFIN Conference]. *Lexicology, Lexicography and Domain-specific Languages* ed. by G. Iamartino (Lexicography worldwide). Monza: Polimetrica
- Hartmann, R. R. K. (forthcoming b). "Mixed dictionary genres" [Art. 12]. In *W/D/D Supplementary Volume 5.4* ed. by R. Gouws et al.
- Hartmann, R. R. K. (forthcoming c). "Aids in metalexigraphic research" [Art. 41]. In *W/D/D Supplementary Volume 5.4* ed. by R. Gouws et al.
- Hausmann, F. J. (1989). "Kleine Weltgeschichte der Metalexikographie". In Wiegand, H. E. (ed.). *Wörterbücher in der Diskussion [I]. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*. Tübingen: M. Niemeyer, 75-109
- Hausmann, F. J.; Wiegand, H. E. (1989). "Components parts and structures of general monolingual dictionaries. A survey" [Art. 36]. In Hausmann, F. J. et al. (eds.). *Wörterbücher/Dictionaries/Dictionnaires. Ein internationales Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie internationale de lexicographie*. Berlin: W. de Gruyter. 5.1. 328-360.
- Householder, F. W.; Saporta, S. eds. (1962). *Problems in Lexicography* [Conference at Bloomington 1960]. Bloomington: Indiana University Press.
- Hyldgaard-Jensen, K.; Zettersten, A. (eds.) (1983). *Proceedings of the Symposium on Lexicography [Copenhagen 1982]*. Hildesheim: G. Olms.
- James, G. (ed.) (1989). *Lexicographers and Their Works*. Exeter: University of Exeter Press.
- James, G. (ed.) (1994). *Meeting Points in Language Studies. A Festschrift for Ma Tailai. Working Papers*. Hong Kong: Language Centre, The Hong Kong University of Science and Technology.

- Kachru, B.; Kahane, H. (eds.) (1995). *Cultures, Ideologies, and the Dictionary. Studies in Honor of Ladislav Zgusta*. Tübingen: M. Niemeyer.
- Katz, B. (1998). *Cuneiform to Computer: A History of Reference Sources*. Lanham MD & London: The Scarecrow Press.
- Landau, S. I. (2001). *Dictionaries. The Art and Craft of Lexicography*. Cambridge: Cambridge University Press [2<sup>nd</sup> edition].
- Lew, R. (2004). *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-bilingual Dictionaries by Polish Learners of English*. Poznań: Motivex.
- Magay, T.; Zigány, J. (eds.) (1990). *BudaLEX '88 Proceedings [EURALEX 3<sup>rd</sup> Congress]*. Budapest: Adadémiai Kiadó.
- McArthur, T. (1998). "What then is reference science?". In McArthur, T. *Living Words. Language, Lexicography, and the Knowledge Revolution*. Exeter: University of Exeter Press. 215-222.
- Murray, J. A. H. (1900). "The evolution of English lexicography". *The Romanes Lecture*. Oxford: Clarendon Press [Reprints ed. by R. Burchfield (1993) in *International Journal of Lexicography* 6, 89-122, and by R. Hartmann (2003). Vol. I: 45-69].
- Pruvost, J. (2000). *Dictionnaires et nouvelles technologies (Écritures électroniques 1)*. Paris: Presses Universitaires de France
- Ščerba, L. V. (1940). "Opyt obshchei teorii leksikografii". In *Izvestiia Akademii Nauk SSSR* 3: 89-117. [Reprints 'Towards a general theory of lexicography' ed. by D. Farina (1995) in *International Journal of Lexicography* 8, 304-350, and by R. Hartmann (2003). Vol. III: 11-50.]
- Städtler, T. (ed.) (2003). *Wissenschaftliche Lexikographie im deutschsprachigen Raum*. [Conference 2001 sponsored by Heidelberger Akademie der Wissenschaften]. Heidelberg: Universitätsverlag Winter.
- Tarp, S. (2007). "Lexicography in the information age". *Lexikos* 17. 170-179.
- Tarp, S. (ed.) (1998). *Perspektiven der pädagogischen Lexikographie des Deutschen [I]. Untersuchungen anhand von 'Langenscheidts Großwörterbuch Deutsch als Fremdsprache'*. Tübingen: M. Niemeyer.
- Tarp, S. (ed.) (2002). *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des 'de Gruyter Wörterbuch Deutsch als Fremdsprache'*. Tübingen: M. Niemeyer.
- Tarp, S. (ed.) (2003). *Untersuchungen zur kommerziellen Lexikographie der deutschen Gegenwartssprache I. 'Duden. Das große Wörterbuch der deutschen Sprache in zehn Bänden'. Print- und CD-ROM-Version*. Tübingen: M. Niemeyer.
- Tarp, S. (ed.) (2005). *Untersuchungen zur kommerziellen Lexikographie der deutschen Gegenwartssprache II. 'Duden. Das große Wörterbuch der deutschen Sprache in zehn Bänden'. Print- und CD-ROM-Version*. Tübingen: M. Niemeyer.
- Wiegand, H. E. (ed.) (1987). *Theorie und Praxis des lexikographischen Prozesses bei historischen Wörterbüchern. Akten der Internationalen Fachkonferenz Heidelberg, 3.6.-5.6. 1986*. Tübingen: M. Niemeyer.

## Appendix: Chronology 1983-2008

- 1983 LEXeter '83 Conference at Exeter GB (EURALEX Congress 1)
- 1984 Foundation of Dictionary Research Centre (Exeter); *Lexicographica* Series Maior Vol. 1; Shanghai Association for Lexicography founded at Shanghai CN; Lexicographical Society of India founded at Mysore IN
- 1985 DSNB Biennial Meeting 5 at Ann Arbor MI; *Lexicographica International Annual* No. 1
- 1986 DRC Seminar on the History of Lexicography at Exeter GB; Conference on Historical Lexicography at Heidelberg DE; ZÜRILEX '86 at Zürich CH (EURALEX Congress 2)
- 1987 InterLex Course 1 at Exeter GB; Translation & Lexicography Colloquium at Innsbruck AT
- 1988 BudaLEX '88 at Budapest HU (EURALEX Congress 3); *International Journal of Lexicography* No. 1
- 1989 International Encyclopedia *Wörterbücher/Dictionaries/Dictionnaires* Vol. 1
- 1990 EURALEX Congress 4 at Benalmádena ES; Symposium on Lexicography 5 at Copenhagen DK; AUSTRALEX Biennial Meeting 1 at Sydney AU
- 1991 NFL Biennial Conference 1 at Oslo NO; M.A. in Lexicography starts at Exeter GB
- 1992 EURALEX Congress 5 at Tampere FI; Colloquium on Onomasiological Dictionaries at Essen DE
- 1993 Lexicographical Society of China Conference 1 at Guangzhou CN; La Journée des Dictionnaires 1 at Paris FR
- 1994 EURALEX Congress 6 at Amsterdam NL; JdD International Colloquium 1 at Cergy-Pontoise FR; LSC Symposium on Bilingual Lexicography 1 at Dalian CN
- 1995 DSNB Biennial Meeting 10 at Cleveland OH; Festschrift in Honor of Ladislav Zgusta (Kachru & Kahane); Summer School/Seminar in Lexicography 1 at Ivanovo RU
- 1996 EURALEX Congress 7 at Göteborg SE; *Who's Who in Lexicography* published at Exeter GB; AFRILEX 1 at Johannesburg ZA
- 1997 NLO Dictionary of Lexicography (Bergenholtz et al.); JdD 5 at Cergy-Pontoise FR; Dictionaries in Asia Conference (and ASIALEX founded) at Hong Kong CN
- 1998 EURALEX Congress 8 at Liège BE; *Dictionary of Lexicography* (Hartmann & James); JLB Colloquium 1 at Paris FR; AUSTRALEX 5 at Brisbane AU
- 1999 ASIALEX Conference 1 at Guangzhou CN; NLF 5 at Göteborg SE; CompLex Conference 1 at Balatonfüred HU; International Symposium on Linguistic and Specialist Dictionaries at Kuwait KW
- 2000 EURALEX Congress 9 at Stuttgart DE; EUROPHRAS Conference 1 at Uppsala SE; Symposium on Lexicography 10 at Copenhagen DK
- 2001 Dictionary Research Centre moved from Exeter to Birmingham GB; LSC 5 at Beijing CN; JACET 1 at Tokyo JP

- 2002 EURALEX Congress 10 at Copenhagen DK; JdD 10 at Cergy-Pontoise FR; International Conference on Historical Lexicography 1 at Leicester GB; KOREALex Conference 1 at Seoul KR
- 2003 LSC Symposium on Bilingual Lexicography 5 at Shanghai CN; Seminar in Lexicography 5 at Ivanovo RU
- 2004 EURALEX Congress 11 at Lorient FR; International Conference on Historical Lexicography 2 at Gargnano del Garda IT; AELex Conference 1 at La Coruña ES
- 2005 DSNB Biennial Meeting 15 at Boston MA; AFRILEX 10 at Bloemfontein ZA; JLB Colloquium 5 at Paris FR
- 2006 EURALEX Congress 12 at Torino IT; International Conference on Historical Lexicography 3 at Leiden NL; KOREALex Conference 10 at Seoul KR; AUSTRALEX 10 at Brisbane AU
- 2007 International Conference on Lexicology & Lexicography of Domain-specific Languages at Palermo IT; Colloque en l'honneur d'H. Béjoint at Lyon FR; ASIALEX Conference 5 at Chennai IN
- 2008 TLF 50th Anniversary Conference at Nancy FR; International Symposium on Dictionaries and Encyclopedias at Aarhus DK; International Conference on Historical Lexicography 4 at Edmonton CA; EURALEX Congress 13 at Barcelona ES

## PAPERS



## **Approaches to Computational Lexicography for German Varieties**

Abel, Andrea; Anstein, Stefanie

### 1. Computational Lexicography and Lexicology

Corpora built for linguistic varieties of a pluricentric language such as German are an indispensable resource for a detailed and systematic variety comparison and dictionary development. We present desiderata and suggestions as well as methods from computational linguistics to systematically apply variety corpora for the enrichment, i.e. confirmation, extension and generation, of lexical entries in distinctive variant dictionaries for German. Examples are those variant dictionaries developed by Ammon et al. (2004) and Abfalterer (2007), where we focus on the South Tyrolean German language. On the one hand, we conducted a systematic frequency analysis in newspaper variety corpora for approved lists of South Tyrolean special vocabulary in order to possibly refine corresponding dictionary entries with corpus evidence. On the other hand, we filtered the list of words of our South Tyrolean corpus which could not be lemmatised by a tool developed for the variety in Germany. After removing special vocabulary collected for the South Tyrolean variety in other projects—e.g. legal terms, the remaining list was manually checked for possible new variant dictionary entries, thus—as an innovative variety corpus lexicographic approach—also automatically filtering a huge amount of data to extract only relevant data to be investigated in detail. In addition, we semi-automatically extracted lexical cooccurrences of our two newspaper corpora and compared their frequencies—with the assumption that those cooccurrences are worth being more closely investigated that have high frequency in the South Tyrolean corpus and very low frequency in the corpus from Germany. With these three methods we were not only able to refine dictionary entries for South Tyrolean German, but also to add new ones. The findings on variants can be re-used for further corpus annotation resulting in again better resources for computational variant lexicography of the kind described, which is also to be extended to more complex linguistic levels.

## **AnCora-Verb: Two Large-scale Verbal Lexicons for Catalan and Spanish**

Aparicio, Juan; Taulé, Mariona; Martí, M. Antònia

### 1. Computational Lexicography and Lexicology

In this paper, *AnCora-Verb* is presented: two large-scale verbal lexicons used for the semantic annotation with arguments, thematic roles and semantic class of AnCora corpora (*AnCora-Cat* for Catalan and *AnCora-Esp* for Spanish). Each corpus

contains 500,000 words with a multilayer annotation in different linguistic fields—from morphology to pragmatics. AnCora-Verb lexicons focuses on syntactic functions, arguments and thematic roles of each verbal predicate taking into account the verbal semantic class and those alternations in diathesis where the predicate can participate. This paper concentrates on the definition and characterization of verb classes and the criteria followed in the assignment of a verb to a specific class.

## **Multi-level Reference Hierarchies in a Dictionary of Swahili**

Bański, Piotr; Wójtowicz, Beata

### 1. Computational Lexicography and Lexicology

This paper can be classified into at least two categories: Computational lexicography and Reports on lexicographical projects, bordering on yet another, the dictionary-making process. The context is a lexicographic project that creates an electronic, TEI XML-encoded Swahili- Polish learner dictionary—with a goal of 10 000 entries in the first stage. Here, we focus on one of the innovative features that we want to introduce in the dictionary, at a relatively small cost— due to the way the dictionary will be compiled out of a Swahili corpus: explicit visualization of derivational hierarchies—essentially a learner-oriented feature, but also serves as a basis for further lexicographic/lexicological applications. We primarily discuss our motivation for this idea and its XML implementation. Nevertheless, by the Conference date, we should also be able to present an actual visualization of it, going beyond a mere set of colourful hyperlinks, which is the way it is presented in our test dictionary—composed of 300 hundred selected illustrative entries, currently being expanded to 1500, for database testing.

## **Multidimensional Ontologies: Integration of Frame Semantics and Ontological Semantics**

Barzdīņš, Guntis; Grūzītis, Normunds; Nešpore, Gunta; Saulīte, Baiba; Auziņa, Ilze; Levāne-Petrova, Kristīne

### 1. Computational Lexicography and Lexicology

Today FrameNeta—a state-of-the-art implementation of frame semantics—provides one of the best insights into lexical semantics and their interaction with the syntactic structure of the sentence. The main limitation of the current implementation is the insufficient level of formalization of frame descriptions, making it unsuitable for automatic text annotation without human supervision. Meanwhile, FrameNet usability would greatly benefit from more rigorous formalization and the consequential possibility for automatic annotation. Previous attempts at formalization have focused on enforcing strict ontological control of the *semantic*

*types* for the frame fillers—despite their insignificant use—due to high ambiguity—in the actual FrameNet. We propose a different approach relying on representation of FrameNet as a 4D multidimensional ontology that allows capturing of the "precedent" knowledge encoded in the manually annotated texts, like FrameNet's full-text annotation reports. This allows both to re- create FrameNet ontology from semantically annotated texts, as well as to use this representation for semantic annotation of new texts. Further extensions of this approach with 5th dimension for anaphora annotation is discussed as an alternative for the informal semantic type mechanism of FrameNet.

## **The Structure of the Lexicon in the Task of Automatic Lexical Acquisition**

Bel, Núria; Espeja, Sergio; Marimon, Montserrat

### 1. Computational Lexicography and Lexicology

In the task of automatic lexical acquisition, i.e. the induction of lexical information from texts, there have been no attempts to exploit theoretically-based models of the structure of the lexicon. Works like those of Bybee (1988) and Langacker (1987) propose a highly structured lexicon where words are related paradigmatically by phonological similarity and where lexical features are an emergent characteristic of the resulting structure. If so, a machine learning algorithm such as a Decision Tree (DT, Quinlan, 1945) should be able to learn the correlation between particular lexical features and the formal characteristics of words. In our experiment, the machine learner should be able to find a correlation between characters that form the words used for training it and the nominal feature /mass/. The ability of the trained learner to predict correctly whether nouns that it has not been shown in the training phase are mass nouns or not is proof that such a correlation exists and that it can be considered an emergent feature of the paradigmatic relations that relate words in the lexicon. The obtained results prove that a structured lexicon can provide information on lexical features.

## **SAOL Plus—A New Swedish Electronic Dictionary**

Berg, Sture; Holmer, Louise; Hult, Anki

### 1. Computational Lexicography and Lexicology

In September 2007, a CD version of the Swedish Academy Glossary, *SAOL Plus*, was released. In *SAOL Plus*, all inflected forms are shown in full text and virtually every text fragment is searchable. Standard search functions include *Search lemma*, *Search inflected forms* and *Search article text*. Advanced searches can be made with the usual wild cards. It also has an advanced tool for fuzzy search based on pronunciation.

*SAOL Plus* can be an asset for the public as well as an efficient and functional utensil for linguists. Moreover, thanks to the fuzzy search, it is useful for people with reading and writing disorders, as well as secondary language users.

System requirements: Windows 98/NT/2000/XP/Vista.

## **Matching Verbo-nominal Constructions in FrameNet with Lexical Functions in MTT**

Bouveret, Myriam; Fillmore, Charles, J.

### 1. Computational Lexicography and Lexicology

*Matching verbo-nominal constructions in FrameNet with Lexical Functions in MTT* Multiword expressions are described in the FrameNet project as complex lexical units linked as wholes to their semantic frames, or as instances of special grammatical constructions. Amongst multi-word expressions, we explore in this paper verbo-nominal constructions in Framenet and more specifically support verb expressions. This category, described in FrameNet literature, has not yet received a systematic treatment in the project. We have looked at several problems regarding its actual classification and theoretical background. Instead of light verbs, or support verbs, we define the object of our study as Support Verb Constructions (SVC). We explore in this connection a well known model in lexicography, the Explanatory and Combinatorial Lexicology, part of the Meaning Text Theory to observe the feasibility of using Lexical Functions in FrameNet for the purpose of encoding SVC. The verbs have functions varying from that of light verbs, where they contribute only the ability to treat the noun's frame as a verbal—e.g., tense-bearing, entity—through a variety of aspect, perspective, and register values. These structures are treated in various ways in FrameNet and in the Explanatory and Combinatorial component of the Meaning-Text-Theory.

Our goal in the present research is to understand the nature of these differences, and to consider whether the results obtained in one of them can be aligned with or incorporated into the other. We approach the parameters of comparison of the two systems by means of three themes characterizing verbo- nominal expressions:

- i. lexicalized collocations of the verb and the nominal head of its syntactic dependent,
- ii. the manner in which the kind of situation—the semantic frame—evoked by the noun is given verbal expression,
- iii. and the manner in which the syntactic arguments of the verb are interpreted as matching the semantic roles associated with the noun's frame.

## **Syntactic Behaviour and Semantic Kinship of Selected Danish Verbs**

Braasch, Anna

### 1. Computational Lexicography and Lexicology

The paper discusses relationships between the syntactic behaviour and meaning of selected verbs, with the focus on exploiting observable syntactic similarities for uncovering of semantic kinship. The investigation is inspired by the demand in language technology for large-scale lexicons that combine morphological, syntactic and semantic descriptions of the lemmas. The development of such a lexical resource is rather demanding, therefore, an enhancement of existing resources with additional information types is a worthwhile task. The computational lexicon for *Danish SprogTeknologisk Ordbase* (STO) comprises a comprehensive syntactic layer which is assumed to be suitable for enhancement with semantic information. The theoretical background for the current approach is the consensus on obvious relationships between a syntactic behaviour and a particular sense of lemmas, as a surface complementation structure reflects the underlying semantic argument structure. The idea is to test the feasibility of deriving semantic information systematically from the syntactic structures encoded in syntactic patterns.

In the pilot project, a sub-set of trivalent verbs that share syntactic constructions are extracted from STO; the material consists of 216 verbs subcategorising for a direct object and a prepositional object covered by eight syntactic patterns. The examination takes a syntactically based grouping of these verbs as its starting point and focuses on defining lexical classes in terms of shared prevalent meaning components. These components form the basis of the semantic label assignment to the particular groups. The material provides 20 basic semantic groups, such as *force*, *urge*, *judge*, *consider*, *remove*, *cheat*, etc. that can be refined into sub-groups along further semantic features or generalized into classes—e.g. *communicate-persuade*, *cause-change-of*, according to different degrees of granularity required. The present classifications of the verbs are also examined in relation to Levin's English verb classes (1993).

Our findings suggest that it is feasible—though within recognized limits—to exploit systematically the formalised syntactic descriptions in meaning group prediction.

## **An Author's Dictionary: The Case of Karel Čapek**

Čermák, František

### 1. Computational Lexicography and Lexicology

After a brief reminder of the long tradition of manually-based author's dictionaries, the possibility of a dictionary based on a full corpus and verified in a number of

aspects against a large corpus has re-emerged. Specifically, the plan of Karel Čapek's dictionary and its realisation is being discussed and its final shape shown, having a number of new, hitherto unused features. The Dictionary, being in fact split into four separate ones, is accompanied by the full Čapek's corpus on a CD where a lot of additional information can be found.

### **Aide à la construction de lexiques morphosyntaxiques**

de Loupy, Claude; Gonçalves, Sandra

#### 1. Computational Lexicography and Lexicology

Morphosyntactic lexica are a very important resource for natural language processing. Many exist; some are freely available for research. But many organisms still produce lexica, even for languages with available resources. In this paper, we present some techniques that can be leveraged to produce lexica more efficiently. Firstly, the format of the lexicon is important. We use a very simple format based on the association of a lemma and a flexion rule, avoiding dozens of entries for a single lemma. Secondly, the linguist must describe some basic elements: the tag list, the tool words and the flexion rules. Thirdly, a specific guesser makes the completion of the lexicon easier. We describe two ways of adding entries to the lexicon using a guesser which associates a lemma and a flexion rule to a word, or a flexion rule to a lemma.

### **Bottom-up Editing and More: The E-forum of *The English-Chinese Dictionary***

Ding, Jun

#### 1. Computational Lexicography and Lexicology

“Computer assistance may enable the lexicographer to prepare and revise dictionaries more quickly”—Barbara Kipfer's prediction made 20 years ago has already become a reality in our age of advanced information technology. Yet how much more quickly can the revision of dictionaries be carried out today? The envelope is now being pushed by the editors of *The English-Chinese Dictionary* (Unabridged) (ECD) through *bottom-up editing*, a new form of online lexicography. Following the launch of the second revised edition of ECD (April 2007), an electronic forum was introduced, linked to the website of Shanghai Yiwen Press, the publisher of the dictionary. For the time being, this e-ECD-forum is attracting more and more of its users to take part in *bottom-up editing*, i.e., pointing out errors and other problems detected in the dictionary directly to its editors through the Net. Three editors including the editor-in-chief participate in the e-forum discussion on a daily basis. Once the problems identified by the users are checked and properly

edited by the editors, they will be listed in the e-newsletter linked to the e-forum. This paper will first explain the functioning of the e-ECD- forum and how such direct interaction between users and editors of ECD proves rewarding to both parties, and secondly, illustrate the mistakes and deficiencies published on the e-forum. Lastly it will explore the potential benefits and problems of online collaborative lexicography in the near future.

## **An Electronic Lexicon for Turkish Idiomatic Compounds Headed by Verbs**

Eyigoz, Elif

### 1. Computational Lexicography and Lexicology

Turkish is a very creative language in terms of idiomatic compounds headed by verbs. Although traditional dictionaries include such compounds, syntactic and morphological properties of compounds are left unrepresented. Moreover, although essential elements of idiomatic compounds can be represented in subcategorization frames that refer to the argument positions of the verbs, it has been observed that subcategorization frames are impractical and even inadequate for representing the argument structure of idiomatic compounds headed by verbs in Turkish. This paper presents a design for representing properties of Turkish idiomatic compounds in a machine readable dictionary, which has been showcased in a sample dictionary for 322 idiomatic compounds.

## **Mordebe Admin—A Lexical Management System**

Ferreira, José Pedro; Barbosa, Sílvia; Janssen, Maarten

### 1. Computational Lexicography and Lexicology

The *Portal da Língua Portuguesa* is a website containing information about the Portuguese language oriented towards the general public. The largest part of the information on the *Portal* is lexical information concerning formal characteristics of words, such as orthography, derivations, loanwords and gentiles. The lexical information comes from a lexical database called MorDebe—or more precisely, a network of lexical databases called the Open Source Lexical Information Network (OSLIN). This abstract shows the general set-up of and major functions of MorDebe Admin, which is the lexicon management system for OSLIN. MorDebe Admin provides an easy and secure way of updating and editing the content of the different databases of OSLIN. Furthermore, much of the data on the Portal are organised as mini-dictionaries and MorDebe Admin provides an integrated collection of tools dedicated to the maintenance of these mini-dictionaries, as well as a built-in neologism tracking system. The software demonstration will illustrate these

functions from a user perspective, and how easy it is to maintain the data behind the Portal.

## **Lexicon Creator: A Tool for Building Lexicons for Proofing Tools and Search Technologies**

Fontenelle, Thierry; Cipollone, Nick; Daniels, Mike; Johnson, Ian

### 1. Computational Lexicography and Lexicology

In this paper, we describe *Lexicon Creator*, a tool designed to help developers produce lexical data for its use in a variety of linguistic applications such as spell-checkers, word-breakers, thesauri, etc. The tool enables developers to work on existing wordlists derived either directly from corpora or from previously created wordlist data. The key feature of the tool is that it enables linguists to rapidly create the morphological rules that are necessary to generate all the inflected forms of a given item. In many languages, a given word may have many forms, each distinguished by different endings attached to the stem of the word. A language like English is rather simple, morphologically—the verb *walk only* has the following forms: *walk, walks, walked, walking*, while other languages may have a number of different forms for a word. Yet, it is essential to create lexicons that can recognize and generate all the inflected forms of a given word, especially for applications such as spell-checkers—where overgeneration should be avoided, thesauri, grammar checkers, morphological analyzers/generators, speech recognition, and handwriting recognizers. It would be extremely time-consuming to code each of these forms individually, so it is necessary to develop this data more efficiently. *Lexicon Creator* allows linguists to classify these variations of the same word into *templates*, or morphological classes, which allow the automatic generation of all valid forms of a word. Once the templates describing the aforementioned variations have been defined, the data-coding task consists of assigning an input word to the correct template and checking that the forms generated automatically are valid. The article will also focus on the additional types of linguistic information which can be attached to words, depending on the intended application that will use the resulting full-form lexicon.

## **Generation of Word Profiles on the Basis of a Large and Balanced German Corpus**

Geyken, Alexander; Didakowski, Jörg; Siebert, Alexander

### 1. Computational Lexicography and Lexicology

In this paper we present the DWDS word profile system, a unified approach to the extraction of collocations for German, based entirely on finite state transducers. The

system is intended as an additional informational source for the DWDS web-platform (*www.dwds.de*). The DWDS website—with 2.5 million page impressions per month—is a widely used internet platform that provides a word-information system based on a large monolingual German dictionary and the DWDS-Kerncorpus, a balanced corpus of German texts of the 20th century. The DWDS word profile consists of two parts: a language-specific part—which consists of a complete German morphology and an efficient syntax parser for German, and a language-independent part comprised of a database management system for collocations and a corpus query engine, together with a web interface. We have applied the DWDS word profile to a balanced German corpus of the 20th century and subsequently present some technicalities. Another experiment using the DWDS word profile in conjunction with a tabloid newspaper shows that there may be significant differences between corpora, underlining the importance of the corpus choice for language learning as well as for the construction of lexical resources. Future work will focus on language learning; in particular, we will use a simplified tag set and a more systematic description of the word profile differences between corpora. We also plan to create word profiles for the DWDS-extended corpus, a 2 billion token corpus.

### ***El Diccionario de diccionarios do galego medieval***

González Seoane, Ernesto; Álvarez de la Granja, María; Boullón Agrelo, Ana Isabel; Rodríguez Suárez, María; Suárez Vázquez, Damián

#### 1. Computational Lexicography and Lexicology

This paper wishes to present some of the most notable tools of the *Diccionario de diccionarios do galego medieval*. This recently-published work is an electronic multidictionary that includes fourteen glossaries and vocabularies born out of Galician or Galician-Portuguese texts or textual collections from the Middle Ages.

### **Shimmering Lexical Sets**

Hanks, Patrick; Ježek, Elisabetta

#### 1. Computational Lexicography and Lexicology

For natural language processing and other applications, it has long seemed desirable to group words together according to their essential semantic type—[[Human]], [[Animate]], [[Artefact]], [[Physical Object]], [[Event]], etc.—and to arrange them into a hierarchy. Vast lexical and conceptual ontologies such as WordNet and BSO have been built on this foundation. Examples such as *fire a* [[Human]] (=dismiss from employment vs. *fire a* [[Weapon]] (=cause to discharge a projectile) have led to the expectation that semantic types such as [[Weapon]] and [[Human]] can be used

systematically for word sense disambiguation. Unfortunately, this expectation is often unwarranted. For example, one *attends* an [[Event]]—a meeting, a lecture, a funeral, a coronation, etc., but there are many events—e.g. a thunderstorm, a suicide—that people do not *attend*, while some of the things that people do attend—e.g. a school, a church, a clinic—are not [[Event]]s, but rather [[Location]]s where specific events take place. The sense of *attend* is much the same in all these examples, unaffected by differences in the semantic type of the direct object. Nevertheless, the pattern [[Human]]*attend* [[Event]] is well established and intuitively canonical.

The CPA (Corpus Pattern Analysis) project at Masaryk University, Brno, provides two steps for dealing with this kind of inconvenient linguistic phenomenon:

1. Non-canonical lexical items are coerced into "honorary" membership of a lexical set in particular contexts, e.g. *school, church, clinic* are coerced into membership of the [[Event]] set in the context of *attend*, but not, for example, in the context of *arrange*.
2. The ontology is not a rigid yes/no structure, but a statistically based structure of shimmering lexical sets.

Thus, each canonical member of a lexical set is recorded with statistical contextual information, like this: [[Event]]: ... meeting. Thus, the semantic ontology is a shimmering hierarchy populated with words which come in and drop out according to context, and whose relative frequency in those contexts is measured. A shimmering ontology of this kind preserves, albeit in a weakened form, the predictive benefits of hierarchical conceptual organization, while maintaining the empirical validity of natural-language description.

## **The Use of Context Vectors for Word Sense Disambiguation within the ELDIT Dictionary**

Ignatova, Kateryna; Abel, Andrea

### **1. Computational Lexicography and Lexicology**

The aim of this paper is to tackle the problem of Word Sense Disambiguation (WSD) within the ELDIT system. ELDIT (*Elektronisches Lernwörterbuch Deutsch-Italienisch*) is an online dictionary of German and Italian, as well as a web-based language-learning system targeted at language learners at elementary and intermediate level. In ELDIT, each word is linked with the corresponding dictionary entry with a list of senses. Nevertheless, selecting the suitable sense of a polysemous word as well as choosing the appropriate homonym in the lookup process is not a trivial task, especially for language learners at elementary level. Therefore, it is

desirable to make the dictionary work easier by automatically selecting the right sense of a word in a given context, which is a Word Sense Disambiguation task. While WSD has been studied intensively in fields such as Information Retrieval (IR), Machine Translation (MT), Question Answering (QA), etc., we present a novel setting, in which WSD is performed within an integrated dictionary system. For performing WSD, we first utilize different kinds of knowledge contained in the ELDIT dictionary, namely part of speech information, morphological knowledge, collocation patterns, and various example sentences as the basis for the context vectors technique. Besides, when the ELDIT dictionary does not provide sufficient data for building a context vector for a word, we fall back upon the vast Internet knowledge. By combining all these sources of information, the implemented module is able to automatically choose the most appropriate meaning of a word in a particular context. It achieves an average precision of 96% for disambiguating Italian and 93% for disambiguating German homonyms. The results for polysemous words greatly depend on how distinct the senses are and how many senses a word has. The evaluation, however, has shown that the approach we apply always outperforms the baseline system—namely, a simplified Lesk algorithm—and gives quite promising results. In addition to that, we show that the data obtained during our work can be re-used in a number of interesting tasks to serve the further improvement of the ELDIT system.

## **Meaningless Dictionaries**

Janssen, Maarten

### 1. Computational Lexicography and Lexicology

The creation of word meaning is one of the most time consuming parts of creating a dictionary. Although it is commonly thought that providing definitions for words is the primary function of dictionaries, it is not the most frequent one. Most dictionaries are used for looking up much more basic information, such as to see whether a word exists or to see whether it is spelled correctly. Dictionaries are relatively good at providing complete definitions for individual words but are not necessarily well equipped for more basic tasks. For many of these smaller tasks, users would be better off using smaller databases—or dictionaries—that focus only on the information the user is looking for rather than searching in a general language dictionary. A dictionary that leaves out most of the details traditionally included in the lexical entry not only makes it easier for the user to find the information he is looking for but also allows the lexicographer to put more focus on the relevant data. It does this by focusing on a single type of information; it becomes more feasible to treat it completely, consistently and coherently for the entire lexicon. The Open Source Lexical Information Network—henceforth OSLIN—is an attempt to create

such single-task lexical resources. This paper explains both the advantages and problems of such an approach.

### **Software Demonstration: The *TshwaneLex Electronic Dictionary System***

Joffe, David; MacLeod, Malcolm; de Schryver, Gilles-Maurice

#### 1. Computational Lexicography and Lexicology

In this presentation, use of the *TshwaneLex Electronic Dictionary* software module will be demonstrated. This module provides a complete, customisable solution for the publication of a dictionary electronically—as a CD-ROM, or for sale or download on the Internet, or both. The "base package" provides all the functionality of a modern, user-friendly Electronic Dictionary, and can be fully customised for the desired 'look and feel', branding, dictionary content and language(s) of a publisher's product. This provides a highly cost-effective solution for creating a professional Electronic Dictionary product, obviating the need for the kind of expensive custom-developed solutions that have traditionally been required. The system can be used for the immediate publication of dictionaries already in TshwaneLex, or in a comparable structured format—such as XML.

### **GDEX: Automatically Finding Good Dictionary Examples in a Corpus**

Kilgarriff, Adam; Husak, Milos; McAdam, Katy; Rundell, Michael; Rychlý, Pavel

#### 1. Computational Lexicography and Lexicology

Users appreciate examples. If a dictionary entry includes contextualized examples of the different senses a word may have, then the user generally gets what they want in a quick and straightforward way. Thus, there are grounds for including lots of examples and contexts. Producing good examples, however, can be labour-intensive, thus, expensive. We automatically found good candidate sentences in a corpus, with which lexicographers could work. The technology used to add examples to an online version of a leading dictionary: we describe and evaluate the project. We consider a range of other ways in which the finding of good examples can bridge the gap between corpora, dictionaries, and language learning.

## **Finding the Words Which are Most X**

Kilgarriff, Adam; Rychlý, Pavel

### 1. Computational Lexicography and Lexicology

Which English words are most distinctive of American English? Which Spanish verbs have a strong tendency to occur in the gerund? Which English nouns are most often used in the plural? All these questions can be answered in quite a straightforward way with a suitable corpus with appropriate markup. The task usually takes a moderate amount of programming. We present a tool which means that it is easy to produce lists of this kind—and many others—which needs no further programming. The work takes place in the framework of a leading corpus query tool.

## **Corpus as a Means for Study of Lexical Usage Changes**

Křen, Michal; Hlaváčová, Jaroslava

### 1. Computational Lexicography and Lexicology

The paper presents a corpus-based method for obtaining ranked wordlists that can characterise lexical usage changes. The method is evaluated on two 100-million representatively balanced corpora of contemporary written Czech that cover two consecutive time periods. Despite similar overall design of the corpora, lexical frequencies have to be first normalised in order to achieve comparability. Furthermore, dispersion information is used to reduce the number of domain-specific items, as their frequencies highly depend on inclusion of particular texts into the corpus. Statistical significance measures are finally used for evaluation of frequency differences between individual items in both corpora.

It is demonstrated that the method ranks the resulting wordlists appropriately and several limitations of the approach are also discussed. Influence of corpora composition cannot be completely obliterated and comparability of the corpora is shown to play a key role. Therefore, although highly-ranked items are often found to be related to changes of language usage, their relevance should be cautiously interpreted. In addition to several general language words, the real examples of lexical variation are found to be limited mostly to temporary topics of public discourse or items reflecting recent technological development, thus sketching an overall picture of lifestyle changes.

## **Non-heads of Compounds as Valency Bearers: Extraction from Corpora, Classification and Implication for Dictionaries**

Lapshinova-Koltunski, Ekaterina

### 1. Computational Lexicography and Lexicology

This paper describes an approach to the classification of nominal compounds based on their subcategorisation. German compound noun predicates, such as *Grundproblem*, *Beweislast* and *Schlussfolgerung* subcategorizing for a subordinate clause are semi-automatically extracted from text corpora and classified according to which of their components, the head or the non-head, is the valency bearer. In over 40% of cases the subcategorisation of compounds is not determined by their heads. This kind of information should be included in subcategorisation lexicons as well as dictionaries for human users. We show that our semi-automatic approach can be applied in natural language processing, especially in lexicon and dictionary creation.

## **The Lexicographic Portal of the IDS: Connecting Heterogeneous Lexicographic Resources by a Consistent Concept of Data Modelling**

Müller-Spitzer, Carolin

### 1. Computational Lexicography and Lexicology

The *Online-Wortschatz-Informationssystem Deutsch* (OWID; Online Vocabulary Information System German) of the Institut für Deutsche Sprache (IDS; German Language Institute) in Mannheim is a lexicographic Internet portal for various electronic dictionary resources that are being compiled at the IDS. It is an explicit goal of OWID, not to present a random collection of unrelated reference works but to build a network of actually related lexicographic products. Hence, the core of the project is the design of an innovative concept of data modelling and structuring. The goal of this granular data modelling is to allow flexible access of each individual lexicographic resource as well as access across diverse dictionary resources. At the same time, fine-grained interconnectedness of all resources should be made possible. Every lexicographic resource within OWID—*alexiko*, *Neologismenwörterbuch*, *Wortverbindungen online*, *Schulddiskurs im ersten Nachkriegsjahrzehnt*—accomplishes this requirement with regard to data modelling and structuring. The paper explains the underlying consistent concept of the data modelling for the overall heterogeneous lexicographical resources. Also it is shown, how the modelling potential has been converting into the Internet presence of OWID.

## **Multilingual Open Domain Key-word Extractor Proto-type**

Panunzi, Alessandro; Fabbri, Marco; Moneglia, Massimo

### 1. Computational Lexicography and Lexicology

Automatic Keyword extraction is now a mature language technology. It enables the annotation of large amount of documents for content-gathering, indexing, searching and for its identification, in general. The reliability of results when processing documents in a multilingual environment, however, is still a challenge, particularly when documents are not limited to one specific semantic domain. The use of multi-term descriptors seems to be a good mean to identify the content. According to our previous evaluations (Panunzi et al. 2006a, 2006b), the availability of multi-term keywords increases the performance with respect to mono-term keywords of 100% relative factor. The LABLITA tool presented in this demo works now in a multilingual environment, as well. The demo calculates on the fly the number of mono-term and multiword keywords of parallel documents in English, Italian, German, French and Spanish, and will allow the audience to judge: a) the enhancement bared by multiword keywords for the identification of content; and b) the comparability of performance obtained by the tool processing different languages.

## **Refining and Exploiting the Structural Markup of the eWDG**

Schmidt, Thomas; Geyken, Alexander; Storrer, Angelika

### 1. Computational Lexicography and Lexicology

In this paper, we describe a semi-automated approach to refine the dictionary-entry structure of the digital version of the *Wörterbuch der deutschen Gegenwartssprache* (WDG, en.: Dictionary of Present-day German), a dictionary compiled and published between 1952 and 1977 by the Deutsche Akademie der Wissenschaften that comprises six volumes with over 4,500 pages containing more than 120,000 headwords. We discuss the benefits of such a refinement in the context of the dictionary project *Digitales Wörterbuch der deutschen Sprache* (DWDS, en.: Digital Dictionary of the German language). In the current phase of the DWDS project, we aim to integrate multiple dictionary and corpus resources in German language into a digital lexical system (DLS). In this context, we plan to expand the current DWDS interface with several special purpose components, which are adaptive in the sense that they offer specialized data views and search mechanisms for different dictionary functions—e.g. text comprehension, text production—and different user groups—e.g. journalists, translators, linguistic researchers, computational linguists. One prerequisite for generating such data views is the selective access to the lexical items in the article structure of the dictionaries which are the object of study. For this

purpose, the representation of the eWDG has to be refined. The focus of this paper is on the semi-automated approach used to transform eWDG into a refined version in which the main structural units can be explicitly accessed. We will show how this refinement opens new and flexible ways of visualizing and querying the lexicographic content of the refined version in the context of the DLS project.

### ***An Anglo-Saxon Dictionary and a Morphological Analyzer of Old English***

Tichy, Ondrej; Čermák, Jan

#### 1. Computational Lexicography and Lexicology

The main stages in the project of the digitization of the *Anglo-Saxon Dictionary* by J. Bosworth and T.N. Toller are described and the value of the resulting data is considered. The paper suggests that the dictionary data need to be structurally tagged if we are to further benefit from the project beyond the current dictionary application. It is also noted that the re-tagging process can be partially automatized, but that it will have its complications due to the ambiguity of typographical tagging currently included in the data. An outline of the development of an Old English morphological analyzer, now in its early stages, is offered using the valuable digitized data of the Dictionary and drawing on a model of a functional Czech morphological analyzer. Envisaged problems, such as the building of stem- and affix-lexicons, Old English vowel variation and stem-final variation, are discussed and several solutions are proposed. The paper also proposes and accounts for some divergence from the model of the Czech analyzer reflecting differences between Czech and Old English morphology and slight differences in the final uses of the Modern Czech and Old English analyzers. Finally, the analyzer's future use, both as a part of the dictionary and as a stand-alone tool for parsing the corpora, for connecting the lexicon entries with text, etc., is suggested and some possibilities of future improvements, e.g. a word-formation or a syntactic analyzer, are indicated.

### **El programa de ejemplificación en los diccionarios didácticos**

Bargalló Escrivá, María

#### 2. The Dictionary-Making Process

In taking into account the value users place on exemplification, we intend, in this work, to show how the information contained in the microstructure is interlinked with that offered in the examples provided in Spanish didactic dictionaries. Additionally, in the terminology used by Rey-Debove (2005), we will observe the interrelation between the information programme and the exemplification programme.

In order to centre our discussion upon these questions, we will try to show to what extent redundant grammatical information is used between the descriptive and illustrative parts, or the complementarity between both these parts. We will analyze the various ways in which this relationship is manifested in order to posit generalizations about the exemplification programme in the dictionaries in question.

To conclude, we will show that, from our point of view, little attention is given to the project as a whole, since a study of some of the questions linked to grammatical information reveals that there are no unified criteria about how the relationship between the information programme and the exemplification programme should develop.

### **Sobre las construcciones pronominales y su tratamiento en algunos diccionarios monolingües de cuatro lenguas románicas**

Battaner, Paz; Renau, Irene

#### **2. The Dictionary-Making Process**

Some non-native student errors in Spanish show the difficulty of pronominal constructions in Spanish language, which is an aspect that has been analysed under all possible approaches in the grammatical bibliography, but has received little attention from a lexicographical point of view. Our aim in this paper is to propose proper treatment of this issue in the *Spanish Learner's Dictionary for Foreign Speakers* (DAELE). We review the advances in grammatical analyses of these constructions for Spanish, and later we observe the treatment they have received in several dictionaries of other Romance languages, to decide which parameters we will take into account for verb entries presenting these constructions. Basing ourselves on the grammatical studies, we establish a classification of ten types of pronominal uses, grouped according to if they are: A) uses deriving from the grammar, in which the pronoun represents an argument of the verb; B) constructions in some languages, such as Spanish and Catalan, that admit what appear to be reflexive pronouns which in fact do not represent arguments of the verb; and C) alternations with or without the pronoun that may or may not display a change in meaning. With this classification in mind, and although the description for Spanish cannot be generalized to all Romance languages, we review some general monolingual and new learner's dictionaries for foreign speakers. For the selection of verbs consulted, we take the examples from French provided by Fontenelle (2004). The treatment of the pronominal uses is analyzed for their appearance in headwords, in a specific sense or subsense, in the examples, and in the observations or remarks. There is some variation across dictionaries, and also within a single dictionary, and that some uses are rarely or irregularly included. Our initial conclusion is that in the DAELE we

should adopt some solutions that do not introduce more variation and should strive to simplify the analysis.

### **La distribució de la informació contextual en els elements estructurals d'un article de diccionari: col·locacions, restriccions lèxiques i definició**

Feliu, Judit; Soler, Joan

#### 2. The Dictionary-Making Process

The main goal of this paper is to discuss the need of improving the *encoding dimension* of general monolingual dictionaries by considering the treatment of non-phraseological lexical combinations—particularly collocations. This will be attained through an analysis of the notion of collocation from a lexicographic point of view taking into account both empirical and theoretical approaches to the general problem of lexical combinations. Moreover, lexical information extracted from corpus must be analysed and included in general monolingual dictionaries bearing in mind that it can be distributed in different elements of an entry— definition, example, etc.— depending on each dictionary structure. Thus, authors will shed some light on the lexicographic task in order to determine whether the co-occurrence of two lexical items must be retrieved and, if so, how and where this lexical and semantic information should be organised. In this sense, it will be demonstrated that the definition and the example fields are not enough to retain and reflect the real use of collocations extracted from corpus. Some guidelines will be provided in order to help the lexicographer in the dictionary-making task concerning the distribution of the extracted information in the collocation field and also in the semantic-restriction pattern and its corresponding definition.

### **The Greek High School Dictionary: Description and issues**

Gavriliidou, Maria; Giouli, Voula; Labropoulou, Penny

#### 2. The Dictionary-Making Process

This paper reports on the compilation of a monolingual Greek pedagogical dictionary targeted at young native language learners, namely secondary education students, aged between 12 and 15. The dictionary, which is in printed form, has been designed to be used in the classroom as a supporting tool for language learning, but also as reference work tailored to meet students' needs for language understanding and production both at school and in everyday activities outside school. To this end, considerations on user-friendliness have been accounted for, and the design and implementation of the dictionary content have built primarily on the needs and requirements of schoolchildren pertaining to the specific age group. The dictionary

comprises 15,000 lemmas covering general language vocabulary along with terms belonging to subjects taught at the specific level of education. Information that is central to the pedagogical targets of language learning has been encoded for each lemma, i.e., part of speech, morphology—difficult inflectional forms, domain, register, definitions, usage examples, etc. Finally, useful comments focus on interesting aspects of certain words' semantics, usage, register etc. The central feature of the dictionary is the headword organization which employs systematically word formation criteria: derivatives by suffixation are organized in word-families, while prefixes are included in the dictionary as independent headwords accompanied with lists of derivatives or compounds on the basis of derivational and semantic criteria. The paper presents the framework of the project and its specifications, discusses the main methodological principles that underlie its construction and elaborates on the dictionary description, the main problems faced and the solutions adopted in the process of its compilation.

### **Desafíos de la definición**

Gutiérrez Cuadrado, Juan

#### 2. The Dictionary-Making Process

This paper reviews the definition of several words in current Spanish dictionaries: *mono* 'monkey', *simio* 'ape', *primate* 'primate', *orangután* 'orang-utan', *gibón* 'gibbon', *chimpancé* 'chimpanzee', *gorila* 'gorilla'. The analysis of the definitions attempts to make it clear that there are some serious inconsistencies in the encyclopaedic information of the dictionaries. It will also be shown that the substitution principle of the definiendum for the *definiens* is problematic in the sense that its conditions of usage are not well defined. Thus, we find that Spanish dictionaries must incorporate the encyclopaedic information in a specific way. On the other hand, our criticism reflects that the substitution principle should be reformulated. The analyses of the questions being examined here reveals that a theoretical and methodological debate is needed in Spanish lexicography. This paper attempts to demonstrate that some of the objections to the definitions of current Spanish dictionaries are due to lack of criticism in the Peninsula regarding the problems related to the encyclopaedic information of the definitions. It would also be advisable to review the questions related to the substitution principle.

## **The Funny Mirror of Language: The Process of Reversing the English-Slovenian Dictionary to Build the Framework for Compiling the New Slovenian-English Dictionary**

Krek, Simon; Šorli, Mojca; Kocjančič, Polonca

### **2. The Dictionary-Making Process**

The article describes the process of reversing the English-Slovenian dictionary database in XML format to create the framework for compiling the Slovenian-English dictionary. The aim was to maximize the abundance of information in an extensive dictionary database with a complex and detailed structure. The process involved lemmatization and POS-tagging of both source and target languages, construction of routines to form the preliminary list of possible headwords and their translation equivalents, as well as routines which enabled the grouping of numerous dictionary examples available in the original dictionary under the appropriate translation equivalent. The result is the reversed dictionary database in XML format with the DTD and XSL file to control the layout for viewing the database in Internet browsers or other XML-aware—dictionary—editors. The article presents the process of reversing the dictionary and the features of the final database. It also reflects on the linguistic issues concerning the fact that the database represents only the mirror image of the English-Slovenian contrastive relation and argues that the contrastively undistorted lexical information from a monolingual Slovenian reference corpus has to be taken into consideration when compiling the new Slovenian-English dictionary.

## **Making a Thesaurus for Learners of English**

Lea, Diana

### **2. The Dictionary-Making Process**

This paper explains the principles and methodology behind the selection and presentation of synonyms in the *Oxford Learner's Thesaurus—a dictionary of synonyms* (April 2008). The needs of learners when consulting a thesaurus are different from those of native speakers: so different, in fact, that they need a completely different kind of thesaurus to consult. Native speakers have a large bank of language stored in their brains; the thesaurus, for them, is simply a means of accessing this information. It reminds them of words that they already know but cannot bring to mind. For language learners, the traditional thesaurus contains far too many words, and not nearly enough information about any of them. They need a thesaurus that will not only enable them to access information, but will also teach them things they did not know before. The first task was to decide which words to

include. A conceptual framework was established, dividing the language into areas of thought and experience. Words under each heading were sorted into groups of near-synonyms. A system of frequency counting was used to order the synonyms and eliminate the less frequent. The resulting entry list was checked against a core vocabulary for learners. Then the entries had to be written. Here, the list of synonyms—forming pretty much a complete entry in a traditional thesaurus—was just the beginning. Each synonym was defined and exemplified. Careful thought was given to register, usage and collocation. Notes contrast the meaning and usage of pairs or groups of words that are particularly hard to tease apart. The aim of the learner's thesaurus is to expand the learner's word bank. It both adds words to the bank, words that the learner did not even know before, and helps learners choose more effectively between words that they have met before, where their knowledge of the exact meaning and usage of the words was previously incomplete.

## **Structure de la définition lexicographique dans un dictionnaire d'apprentissage explicatif et combinatoire**

Milićević, Jasmina

### 2. The Dictionary-Making Process

The paper focuses on the construction of lexicographic definitions for an electronic dictionary targeting intermediate-to-advanced learners of French as a second language. It proposes a learner-friendly adaptation of definition formats developed for a theoretical lexicon of a particular type—*Explanatory-Combinatorial Dictionary* [= ECD] *of Contemporary French*. It is argued and, hopefully, demonstrated that it is possible to construct lexicographic definitions that are both theoretically sound and palatable for language learners. The work on the adaptation of existing definition formats has shed light on these definitions themselves—which in some cases needed to be modified—thus demonstrating, once again, the interdependence of applied and fundamental research. After presenting the basic structure of an ECD-style definition, the paper details the types of modifications needed to make it learner-friendly. These range from 'superficial' modifications, aimed at a better readability of the definition—typography, indentation, color, etc., to substantial ones—simplification of the vocabulary and syntax of the defining language, omission of the definition components deemed non-essential in a learner's dictionary, changing the word-sense division within a polysemic word in case the level of detail seems too high for our purposes, etc. The proposed approach is illustrated with the example of evaluation verbs, such as *approuver* '(to) approve', *désapprouver* '(to) disapprove', *blamer* '(to) blame', *critiquer* '(to) criticize', etc., for which a definition template and definitions themselves are given.

## **Frames and Semagrams. Meaning Description in the General Dutch Dictionary**

Moerdijk, Fons

### 2. The Dictionary-Making Process

This paper discusses the semagram, an innovation in the way of describing meaning in lexicography, as used in the *Algemeen Nederlands Woordenboek* (General Dutch Dictionary). A semagram is the representation of knowledge associated with a word in a frame of *slots* and *fillers*. *Slots* are conceptual structure elements which characterise the properties and relations of the semantic class of a word—e.g. colour, smell, taste, composition, components, preparation for the class of beverages. The abstract meaning frame for such a semantic class is called *type template*. After a motivation for the use of frames in lexicography we reveal how semantic classes are determined and how type templates are composed. We illustrate this with the type template of the animal names and show how the semagram of *cow* is based upon it. We conclude by summing up the main advantages of the use of semagrams.

## **A Systematic Approach to the Selection of Neologisms for Inclusion in a Large Monolingual Dictionary**

O'Donovan, Ruth; O'Neill, Mary

### 2. The Dictionary-Making Process

For each new edition of *The Chambers Dictionary*, around 1,000 new words are selected by Chambers' lexicographers for inclusion. In preparing the latest edition, we seized the opportunity to use new corpus and database technology to improve neologism detection and selection. Our resources included the large, recently built Chambers Harrap International Corpus (CHIC), our automated word-tracking system, the databases developed for our new words monitoring programmes and a new tool for ranking words by corpus frequency. We report on the results of our work in this area: a systematic approach to neologism detection and investigation that complements the expertise of lexicographers.

## **Lexicographic Treatment of Italian Phrasal Verbs: a Corpus-based Approach**

Onesti, Cristina

### 2. The Dictionary-Making Process

Italian phrasal verbs—or *verbi sintagmatici*— have seen a growth of interest in the Italian linguistic panorama. However, a more systematic analysis is needed to clarify the theoretical status of these verb-particle constructions and to improve their

lexicographic treatment, which is still inconsistent, as shown by an overall comparison between the major Italian monolingual dictionaries. Difficulties are related with both an unclear semantic classification of them and the lack of frequency and productivity data about their formation. Following the classification of Masini (2005) in terms of intensification, direction, metaphoric and actional meaning, the present paper carries out a case study aiming at frequency data in phrasal verbs with *via*, in particular about the presence of an Aktionsart contribution of the particle. The meaning of accomplishment seems to be clearly traceable in the newsgroup messages analyzed (corpus NUNC-It), although very few dictionaries record it. Further corpus data should offer schemes of regularity in the usage of other verb-particle constructions making their lexicographic treatment more effective.

## **Il *Dizionario Garzanti* nel quadro della lessicografia italiana contemporanea**

Patota, Giuseppe

### **2. The Dictionary-Making Process**

Over the last fifteen years, Italian lexicography has achieved significant results in the field of historical as well as general dictionaries, thanks to the publication of outstanding works which provide the users with a wide range of information, both in paper and digital form, namely: phonemic transcriptions, etymological indications, the dating of the entries with the original occurrence, use frequency, phraseology, synonyms and antonyms, polyrematic units, grammatical notes. As they are specifically geared to provide surveys of written and spoken Italian based on real evidence, these works have come to be crucial methodological tools for knowledge not merely linguistic. The *Dizionario Italiano Garzanti* is to be placed within the framework of this 'new' lexicography. The dictionary, which from the beginning has featured clarity, accessibility, and comprehensiveness, has profoundly changed its scope over time, and its 2008 version can be regarded as the achievement of a long-standing commitment to language evolution. My paper will address the most significant transformations of the *Dizionario* by accounting for its developments in the context of the most recent history of Italian lexicography.

## **Lemmatisierungspraxis und -problematik im Autorenwörterbuch am Beispiel des *Goethe-Wörterbuchs***

Schares, Thomas; Schlaps, Christiane

### **2. The Dictionary-Making Process**

The macrostructure of an author's dictionary as determined by (the rules of) lemmatization so far has attracted little attention in practical and theoretical lexicography, with most publications on the topic covering small-sized dictionaries that represent only specific segments of a writer's works and vocabulary. The *Goethe Dictionary* (GWb), in contrast, endeavors to treat the complete vocabulary used by Johann Wolfgang Goethe in his prolific literary, scientific, philosophical, historiographic etc. writings as well as in his letters and, to some extent, his conversations. The word list contains over 90.000 headwords, thus making the project the largest dictionary on an author's idiolect worldwide. The form and placement of lemmas in the GWb is determined by a set of rules that over the decades of work on this dictionary have been collected in a style manual but which are far from comprehensive or even static. Examples from the published parts of the GWb will demonstrate a number of decisions that typically arise in author's dictionaries and the way they are solved in the GWb, including questions of an author's idiolectal aberration from the orthographic norm, the frequency of hapax legomena, and morphological idiosyncrasies that in the GWb led to specialized sublemmas. In our paper, we would like to emphasize the need for research, both practical and theoretical, into the special problems of lemma presentation in author-centred lexicography and its methodological foundation as we believe that to show the full range of an individual's lexicon will not only contribute to the field of metalexigraphy in particular but will also help to gain valuable insights into the lexicological, semantic, grammatical, and, incidentally, historical dimensions of language in general.

## **Von der Markierung zur Beschreibung: Besonderheiten des (Wort-) Gebrauchs in *elexiko***

Schnörch, Ulrich

### **2. The Dictionary-Making Process**

*Elexiko* is a lexicological-lexicographic, corpus-guided German Internet reference work (cf. [www.elexiko.de](http://www.elexiko.de)). Compared to printed dictionaries, in *elexiko*, restrictions on space disappear. Specific comments on the use of a word do not need to be given in traditional abbreviated forms, like the so-called field labels or usage. In this paper, I will show its advantages for the description of the particular pragmatic

characteristics of a word: I will argue that traditional labelling such as *formal*, *informal*, *institutional*, etc. cannot account for the comprehensive pragmatic dimension of a word and that these are not transparent, particularly for non-native speakers of German. The main focus of the paper will be on an alternative approach to this dictionary information—as suggested by *ellexiko*. I will demonstrate how narrative, descriptive and user friendly notes can be formulated for the explanation of the discursive contextual embedding or tendencies of evaluative use. I will outline how lexicographers can derive such information from language data in an underlying corpus which was designed and compiled for specific lexicographic purposes. Both, the theoretical- conceptual ideas and their lexicographic realisation in *ellexiko* will be explained and illustrated with the help of relevant dictionary entries.

## **Requirements for the Design of Electronic Dictionaries and a Proposal for their Formalisation**

Spohr, Dennis

### 2. The Dictionary-Making Process

We discuss recent analyses of the requirements for the design of electronic dictionaries, building primarily on the accounts by de Schryver (2003), Chiari (2006), Heid (2006) and Tarp (2008). These requirements suggest a richer formalization of dictionary models than is usually the case in traditional database and plain XML-based approaches, and we therefore argue in favour of a formalisation of these requirements in the framework of a strongly typed formalism. The discussion focuses on users' needs, needs of specific applications of Natural Language Processing, and multifunctionality—in the sense suggested by Gouws (2006) and Heid/Gouws (2006). We further point out the benefits of a richer formalization of dictionary models that goes beyond the traditional view on lexical resources, and strengthens our claim by providing evidence from related work on lexicon modelling in OWL DL (Burchardt et al., 2008).

## **Alphabetic Proportions in Estonian Monolingual and Bilingual Dictionaries**

Veldi, Enn

### 2. The Dictionary-Making Process

The paper discusses alphabetic proportions in Estonian general monolingual and bilingual dictionaries with Estonian on the left-hand side. As no data about the Estonian alphabetic proportions were available, the alphabetic proportions were calculated on the basis of the corpus-based *Frequency Dictionary of Standard*

*Estonian* (Kaalep and Muischnek 2002). The findings were then used to configure the Estonian ruler in TshwaneLex dictionary compilation software—alphabetic proportions for English, Afrikaans and several Afrikan languages as well as an excellent background to this problem can be found in De Schryver 2005. Subsequently, the established proportions were used as a yardstick for comparing three monolingual and six bilingual dictionaries. The six bilingual dictionaries included four general Estonian-English dictionaries—one of them not completed as of yet but revealing potential problems—and two school dictionaries—Estonian-German and Estonian-Russian. The findings show that while alphabetic proportions have generally been followed quite successfully, some Estonian dictionaries show a tendency to be skewed—some dictionaries become more thorough towards the end of the alphabet while others show the opposite trend. The problem is more challenging for those dictionary projects that require decades for completion and where the dictionary is published in fascicles—in the Estonian lexicographic practice one has in mind the *Explanatory Dictionary of Standard Estonian*, which started publication in 1988. There can also naturally be instances where single alphabetic stretches may reveal perceived overtreatment or undertreatment. The paper also argues whether the alphabetic proportions of certain letters can vary to some extent depending on the selection of words listed under them—e.g. inclusion of large numbers of foreign and learned words beginning in *a*, *b*, *d*, *g*, *f* in Estonian dictionaries can increase the proportions of these letters as they are less typical of native words. The paper ends with the firm conviction that in recent years it has become much easier to control the progress of a lexicographic project.

### ***Dictionnaire de Néologismes du Portugais Brésilien (décennie de 90): conception et processus d'élaboration***

Alves, Ieda Maria

#### 3. Reports on Lexicographical and Lexicological Projects

*The Brazilian Portuguese Neologisms Dictionary* (1990s), expected to be completed in early 2008, was created to present neologisms from contemporary Brazilian press from January 1993 to December 2000. The dictionary corpus has been taken from the *Contemporary Brazilian Portuguese Neologism Database*, a tool aimed at collecting and investigating neologisms from contemporary Brazilian press—newspapers *Folha de S. Paulo* and *O Globo* and magazines *IstoÉ* and *Veja*—since January 1993. Samples were randomly taken from these sources: the newspaper *O Globo*, from the first Sunday of the month; the magazine *IstoÉ*, from the second week of the month; the newspaper *Folha de S. Paulo*, from the third Sunday of the month; and the magazine *Veja*, from the last week of the month. In the investigated period, 13,500 neologistic-lexical units were collected. The collected data shows that

with 30%, the prefixed formations are the most frequent in the group of neologistic words. The other formations correspond to subordination composition (19%), borrowings (17%), syntagmatic formations (13%), suffixed formations (8%), coordination composition (5%), semantic neology (4%), blending (2%) and other formations (2%). These results are reflected in the dictionary's macrostructure, which will present about 3,000 entries according to the frequency proportion, e.g. the prefixed formations will correspond to 30% of the entries. Each article demonstrates necessarily: lexical units, grammatical references, definitions, context, context references, linguistic notes—types of formation, attested years and eventually label, variant, acronym or abbreviated form, encyclopaedic notes and attestation in dictionaries after 2000. For example: *jogador-chave* sm *Jogador que se destaca em uma equipe. O relatório que Matsunaga entregou ao técnico Akira Nishino aponta o meia-atacante Juninho como o do Brasil. (FSP, 21-07-96) Composição por subordinação. Atestado (4) em 1994, 1996, 1998.*

### **A Digital Dictionary of Catalan Derivational Affixes**

Bernal, Elisenda; DeCesaris, Janet

#### 3. Reports on Lexicographical and Lexicological Projects

This paper presents the digital dictionary of Catalan derivational affixes, awarded with the Laurence Urdang Award 2005. The aim of this dictionary is to provide a tool for lexicographers that will help them systematize the representation of the language's morphology in dictionaries, and provide an in-depth description of Catalan derivational affixes in the form of a dictionary that should be of interest both to linguists and language professionals and those seeking a model for similar projects dealing with other languages.

### **Recopilación y estructuración del vocabulario de especialidad en el *Nuevo Diccionario Histórico del Español* (RAE)**

Carriazo Ruiz, José Ramón; Gómez Martínez, Marta

#### 3. Reports on Lexicographical and Lexicological Projects

The *New Historical Dictionary of the Spanish Language* is the latest work the Spanish Royal Academy has decided to undertake. In this lexical compilation, scientific and technical vocabulary will be taken into account and included, regardless of what other previous dictionaries, such as general language dictionaries, have done so far. For this purpose, a team of lexicographers in Cilengua (La Rioja) is studying the method to select, extract and tag the vocabulary used in specialized areas of knowledge. This presentation explains the different steps that will be followed in order to include the specialized terms and their history in the dictionary, such as the

establishment of a representative corpus, the selection of terms or the introduction of subject matter labels.

### ***Portal de léxico hispánico: una herramienta para el estudio del léxico***

Clavería, Gloria; Prat, Marta; Torruella, Joan; Buenafuentes, Cristina; Freixas, Margarita; Julià, Carolina; Massanell, Mar; Muñoz, Laura; Varela, Sonia

#### 3. Reports on Lexicographical and Lexicological Projects

The *Portal de léxico hispánico* (Hispanic vocabulary portal) is the result of the research project “Banco de datos diacrónico e hispánico: morfología léxica, sintaxis, etimología y documentación”, carried out in the last few years thanks to funding by the Ministerio de Educación y Ciencia (Ref N°: HUM2005-082149-C02-01) and the Generalitat de Catalunya (Ref N°: SGR2005-00568). This website brings together scientific information about the vocabulary of the Ibero-Romance languages and their dialectal variations. The objective of this site is twofold: on the one hand, it aims to provide a tool for all Internet users to obtain both diachronic and synchronic information about Hispanic vocabulary, and on the other hand, it aims to provide a useful tool for the work on the *Nuevo Diccionario Histórico* of the Real Academia Española. The *Portal de léxico hispánico* fundamentally brings together information of bibliographic, lexicologic and documentary nature relating them to a wide range of data bases. It contains data from the digital version of the *Diccionario crítico etimológico castellano e hispánico* by J. Coromines and J. A. Pascual (Madrid: Gredos, 1980-1991) together with data from other sources. The consultation interface of the portal allows bibliographic, diachronic, diastatic, diatopic, etymologic, graphic, phonetic-phonological, morphosyntactic and semantic information to be searched for specific words, as well as their documentation in ancient and modern texts. In the communication, detailed information will be available about the origin, aims, characteristics and present state of the *Portal de léxico hispánico*. There will also be a demonstration of the consultation interface available on the Internet.

### **ISO-Standards for Lexicography and Dictionary Publishing**

Derouin, Marie-Jeanne; Le Meur, André

#### 3. Reports on Lexicographical and Lexicological Projects

Many things have changed in the field of dictionary production during these last ten years. With the introduction of digital support and networking, the lifespan of dictionaries has been considerably extended. The dictionary manuscript has become a unique data source that can be re-used and manipulated time and again by

numerous in-house and external experts. The traditional relationship between author, publisher and user has now been expanded to include other partners, such as data-providers—either publishers or institutions or industry-partners, software developers, and language-tool providers. All these dictionary experts need a basic common language to optimize their work flow and to be able to co-operate in developing new products while avoiding time-consuming and expensive data manipulations. Dictionary users also need to receive more reliable information about new lexicographic products. In this paper we will first of all present the ISO standardization for Lexicography which takes these new market needs into account, and then go on to describe the two new standards: *Presentation/Representation of entries in dictionaries* which was published in March 2007 and *Lexicographical production and marketing: Concepts and vocabulary* which was launched in the summer of 2007. In conclusion, we will outline the benefits of standardization for the dictionary publishing industry.

### **Construir un diccionario de derivación del español en el siglo XXI. La arquitectura de la información al servicio de la lexicografía**

Díaz García, María Teresa; Mas Álvarez, Inmaculada

#### **3. Reports on Lexicographical and Lexicological Projects**

In the age of communication, new technological advances are made every day in the field of information and linguistics. *Deriv@* is a system that makes the most of linguistic data. All data managed by this application are connected with derivational morphology, since derived words constitute the starting point from which relationships with other linguistic elements are established. *Deriv@* is a linguistic database with two models of representation: one for all Spanish words created by means of derivation; the other for the corresponding Latin words. We offer a grammatically-analysed corpus, both synchronically and diachronically, to make customised queries according to the user's interests, and also to create a dictionary of derived words. *Deriv@* allows and facilitates non-restricted access to all the information available in the two databases. This model is valid for any form derived from Spanish or from other Romance languages. It offers users the possibility of using it real time, thus letting them interact with the system and contribute to its improvement.

## **El Diccionari de l'Institut d'Estudis Catalans (2007): el tractament de la pronominalitat verbal**

Fradera, Imma; Fullana, Olga; Montalat, Pere; Santamaria, Carolina

### 3. Reports on Lexicographical and Lexicological Projects

The aim of this paper is to explain the work that has been carried out in the Oficines Lexicogràfiques to regularize the treatment of verbal pronominality in the *Diccionari de la llengua catalana* of the Institut d'Estudis Catalans (DIEC) and to show the result that the second edition of this dictionary, published in April of 2007, reflects. It is well known that the DIEC is not a dictionary started from scratch, but it is based on the *Diccionari general de la llengua catalana* of Pompeu Fabra (DGLC). The first and second editions of the DIEC represent an updating of the DGLC, not only regarding nomenclature and definitions but also regarding the revision and the systematization of some lexicographic treatments, as in the case of the treatment of verbal pronominality. Our exposition is divided into five sections:

1. First, we will briefly present the treatment of verbal pronominality in the DGLC and in the first edition of the DIEC (DIEC1).
2. Then, we are going to analyze which type of linguistic constructions is considered pronominal in these two dictionaries.
3. After that, we are going to state the theoretical frame that allowed us to establish the bases on which the lexicographic criteria are built.
4. Fourth, we are going to explain the lexicographic criterion that governs the treatment of verbal pronominality in the second edition of the DIEC (DIEC2) and some results of the application of this criterion will be shown.
5. Finally, we will conclude our exposition by presenting the reach of this application.

## **Diskurswörterbuch—Zur Konzeption eines neuen Wörterbuchtyps**

Kämper, Heidrun

### 3. Reports on Lexicographical and Lexicological Projects

After a brief discussion on the term *discourse*, discourse will be related to the tasks of a discourse dictionary. The paper goes on developing the subject of discourse lexicography, which is a lexicographic presentation of discourse vocabulary, of the net of its semantic relations, and of the societal and historical circumstances of the usage people have made of it. This background will be useful for the presentation of two types of discourse dictionaries. On the one hand, they are based on the same primary conception. On the other hand, they are adapted to the respective discourse

constellations. The first example is the result of a project on the early post-war period and presents the already-existing discourse dictionary of this project. The content of this dictionary is the vocabulary of three different groups, which participate in one discourse and specifically represent its main item. Since this dictionary also exists in electronic version, this concept will be proved by examples taken out of this version. The second example refers to a project running on the 1967/68 protest period. The vocabulary of this discourse makes up a set of several single discourse items, while these items constitute the leading subject of the discourse of 1967/68: *democracy*. Thus, the task of the lexicographic description of a complex discourse like this is not at least: to assign the discourse vocabulary to the single discourses and to describe the different usages relating to these single discourses. The paper ends with a draft of a lexicographic program based on the type *discourse dictionary*.

## Turning Roget's Thesaurus into a Czech Thesaurus

Klégr, Aleš

### 3. Reports on Lexicographical and Lexicological Projects

Turning *Roget's Thesaurus* into a Czech Thesaurus in a report on how a thesaurus of the Czech language was compiled on the basis of *Roget's Thesaurus*, the following issues are covered:

1. Reasons for undertaking the thesaurus project—to redress the unbalance between the semasiological and onomasiological description of Czech by compiling a counterpart to the two large alphabetical dictionaries of Czech;
2. Strategy and philosophy, and the choice of the source text—combination of translation and original compilation; decision to use an available and well-proven model, a shorter version of *Roget's Thesaurus*, to resolve the issue of a classificatory system and format;
3. Phase one: a project grant—awarded by Charles University for a three-year project, Computerized Thesaurus of the Czech Language, resulting in a preliminary translated version of the Czech thesaurus and the publication of a sample volume as an output;
4. Phase two: expanded version for publication—moving from translation to original compilation for greater autonomy of the Czech thesaurus and expanding the average of 80 items per entry to 300 using Czech sources; specific rules required for entry structure, the type and order of subentries, etc, to ensure the uniform format of the entries;

5. Compiling the index—to achieve the standard index-length equal to that of the dictionary text, a procedure combining manual and mechanical shortening was devised to abridge the dictionary text;
6. Conclusion. Compilation of a thesaurus via translation from another language is a possible procedure. Supplementing translation with original compilation based on target-language resources is nevertheless recommended if a truly national thesaurus is to result.

### **MEDLEX+: An Integrated Corpus-Lexicon Medical Workbench for Swedish**

Kokkinakis, Dimitrios; Toporowska Gronostaj, Maria

#### 3. Reports on Lexicographical and Lexicological Projects

This paper reports on the work carried out developing MEDLEX+, a medical corpus-lexicon workbench for Swedish. This project, which is still under active development, has been going on for some years now within the Department of Swedish language at Göteborg University. At the moment, the workbench incorporates:

- i. an annotated collection of medical texts—including 20 million tokens and 45,000 documents,
- ii. a number of language processing software programs, including tools for collocation extraction, compound segmentation and thesaurus-based semantic annotation, and
- iii. a lexical database of medical terms—containing 5,000 medical entries. MEDLEX+ is a multifunctional lexical resource due to a structural design and content which can be easily queried. The medical workbench is intended to support lexicographers compiling lexicons and also lexicon users more or less initiated in the medical domain. MEDLEX+ can also assist researchers working on either lexical semantics or natural language processing (NLP) applications with focus on medical language. The linguistically and semantically annotated medical texts in combination with a set of smart queries turn the corpora into a rich repository of semasiological and onomasiological knowledge about medical terms and their linguistic, lexical and pragmatic properties. These properties are recorded in the lexical database with a cognitive profile. The MEDLEX+ workbench seems to offer a constructive help in many different lexical tasks.

## **Semiotic Conceptualization of Human Body: Lexicographical or Database System Description?**

Kreydlin, Grigory E.

### 3. Reports on Lexicographical and Lexicological Projects

The paper focuses on some possible means of lexicographical description of the human body both in natural languages and in nonverbal semiotic systems. The Russian language and the Russian body language present significant material for constructing semiotic representation of the human body and its parts. Two basic modes for such a representation—explanatory dictionaries and database systems—are discussed in detail. It is argued that database systems provide, on the one hand, more explicit and rigorous format for the comparative analysis of gestures, postures, mimics and other nonverbal signs, and natural language expressions, on the other hand, than explanatory language and gesture dictionaries.

## ***Repertorio analitico dei dizionari bilingui francese-italiano***

Lillo, Jacqueline

### 3. Reports on Lexicographical and Lexicological Projects

This research has aimed at analytically listing bilingual French-Italian, Italian-French dictionaries available in public and private libraries. The team of about thirty researchers has visited almost 400 libraries in France and Italy and also in Netherlands, Spain and Great-Britain. 800 different editions have been found from the first in 1583 to 2000—conventional date. An analytical description has been provided for each of them. It gives general information on their author, title, printing city, publishers, volume measures, typology, etc., and more specific information on the metalexical languages, the paratext—introduction, illustrations, etc., the nomenclature and the microstructure itself—phonetics, etymology, descriptive glosses, labels, examples, etc. All the information, registered in a data base, allow us to present a pretty realistic view on bilingual French-Italian and Italian-French lexicography from the very beginning. Various figures are included in this article to show: the number of dictionaries per century, the most productive authors—over 15 editions, the production per author—it is very interesting to see that almost half of the authors have published only one dictionary, the places of publication per century and all together. This bibliography—*Quattro secoli di lessicografia franco-italiana 1583-2000. Repertorio analitico di dizionari bilingui*—is published by Peter Lang.

## **Ein elektronisches Lexikon im OLIF-Format für die Erzählanalyse**

Luder, Marc; Clematide, Simon; Distl, Bernhard

### 3. Reports on Lexicographical and Lexicological Projects

We present the JAKOB lexicon, a semantically rich German lexical resource, and its migration to the OLIF format (Open Lexicon Interchange Format). This lexicon is part of a web-based text and narrative analysis application. The JAKOB narrative analysis is a qualitative research tool to systematically analyze patient's narratives. It conceptualizes narratives as dramaturgically-constructed linguistic productions and interprets them with regard to the unconscious conflicts of the narrator contained there in. In this process, narratives are extracted from transcripts, then a linguistic analysis is performed, and after that the vocabulary is encoded according to predetermined psychological conceptual categories incorporated in the JAKOB lexicon. The need for the proper treatment of multi-word units in the JAKOB project made OLIF a reasonable target format. OLIF is word-sense oriented and allows a broad linguistic description—syntactical, morphological, and semantic—for each lexical entry. The OLIF data categories and attributes are well defined in the case of German but it turned out that the data-category labels in OLIF aren't specified very clearly sometimes. In addition to that, there are few resources that prove their practical use. In a corporate project, the lexicon was half-automatically reassessed and finally migrated. OLIF is an open XML-based standard for structuring lexical data and provides a rich choice of linguistic categories and predefined values. Multi-word entries represent an essential improvement for the JAKOB application. The narrative texts represent spoken language; therefore the utterances aren't well formed, in most cases, and not eligible for a standard syntactic analysis. We use a construction-grammar approach to gather the sense of multi-word expressions in the text and to match them to lexicon entries with their corresponding conceptual categories. We use multi-word entries as containers for constructions—form-meaning units—like idioms and collocations. Further investigations will show to which extent more general constructions can be lexicalized. Our project goal is to improve precision in coding the JAKOB narratives. We decided to create an OLIF database, using the XML schema as the basis for the database structure. Thus, import and export of OLIF data is straight-forward. The implementation is object-oriented and solely based on open source software using PHP / MySQL.

## **GEST 2.0: A Gestionary of Emblems for Cross-Cultural Communication & Media Accessibility**

Mesa Lao, Bartolomé; Bartoll Teixidor, Eduard

### **3. Reports on Lexicographical and Lexicological Projects**

Nowadays, media accessibility is gaining relevance and it is becoming socially more and more important in almost all areas of our daily lives. Among the different types of media products, the most common ones are those combining the audio and the visual channel. In such products, semantic content can be transmitted by using the audio channel as well as the visual, or both at the same time. From this perspective, an important semantic aspect that can be conveyed by visual means is body language. Body language is a broad term for forms of communication using body movements or gestures instead of—or in addition to—sounds, verbal language, or other forms of communication. Due to the relatively high degree of information contained within human gestures, it seems to be necessary to open new fields of research based on this paralanguage. The idea behind this study of body language is to present the structure of a multicultural ‘gestionary’, a multimedia dictionary in progress of culturally coded gestures for audiovisual translators. The study of semantic aspects of culture-based gestures should prove useful for audio describers when dealing with meaning, context of use and verbal formulation of such gestures. For instance, compared to the field of second- language learning, this topic has not kept pace with the level of interest in the area of Audiovisual Translation. The creation of a new multimedia dictionary of gestures reflects our interest in putting together in a single project three complementary fields:

1. The creation of new tools for audiovisual translators.
2. The possibilities of web 2.0 technologies to develop socially generated projects.
3. The need to find new ways to go further in media accessibility.

## **Introducing BAWE: A New Lexicographical Resource**

Nesi, Hilary

### **3. Reports on Lexicographical and Lexicological Projects**

This paper reports on the compilation of the British Academic Written English (BAWE) corpus, a collection of almost 3000 proficient student assignments produced at three representative universities in the UK. BAWE was designed to fill a gap in current corpus resources by complementing other writing collections which represent expertly written academic text—such as the TOEFL 2000 Spoken and

Written Academic Language Corpus, or non-expert and non-discipline specific student writing—such as the Louvain Corpus of Native English Essays, and the Cambridge Syndicate Examination corpus. Prior to the development of BAWE the few small corpora of writing produced by university students within their disciplines had either been compiled for individual scholarly purposes, or were in the form of inadequately documented and unannotated ‘essay banks’ for student use. The BAWE corpus, in contrast, is a large, formally compiled collection of assignments at four levels of study, from first year undergraduate to masters level, accompanied by detailed contextual information. Thirteen broad macrogenres have been identified in the corpus, including the *essay*, of course, and writing generically similar to the published research article, but also including other types of writing, neglected in the literature, which reflect the purpose of university level study. The full corpus will be freely available to researchers from January 2008, and it is foreseen that it will provide a—currently unique—resource for designers of dictionaries for advanced learners, particularly those learners studying at university level in the medium of English.

### **Development of the Integrated Concordancer for the Corpus of the 17th to 19th Century Culinary Manuscripts**

Paek, Doo-hyun; Nam, Kil-im; Lee, Mi-hyang; Ahn, Eui-jeong; Song, Hyeon-ju

#### **3. Reports on Lexicographical and Lexicological Projects**

The aim of this project is to develop the Integrated Concordancer for food-related terms used in Korean culinary manuscripts from the 17th to the 19th century. The Integrated Concordancer may be utilized by Korean linguists who wish to make use of culinary manuscripts as research materials for the history of the Korean language. Additionally, it might be useful for culinary scholars of traditional foods, and also for the general public. The tasks of the current project are twofold. The task is, firstly, to construct a corpus by collecting hand-written culinary manuscripts written between the 17th century and the 19th, and develop a web-based search engine. Secondly, to extract headwords of everyday words from the corpus of the 17th to 19th century manuscripts and compile a source book for traditional culinary terms by making and utilizing concordance data by frequency, part of speech, and semantic pattern. The current project is a two-year project—starting in August 2007, ending in July 2009—which will eventually become available to the public.

## **Diccionario de los glifos maya con descripción visual estructural**

Pichardo-Lagunas, Obdulia; Sidorov, Grigori;

### 3. Reports on Lexicographical and Lexicological Projects

The deciphering of Mayan script is an intricate but interesting problem. During years, the community of Mayan researchers was not open to the usage of computer tools. Still, the progress of the computer science and the current state of Mayan research proves the necessity of this type of software. We present the project related to the development of Mayan script database, which is the first necessary step in development of computer representation of Mayan script data. The database contains several tables and allows for various queries. The main idea of the project is the development of the system that would allow managing Mayan script data for specialists as well as for persons without any previous knowledge of Maya. This includes structural visual description of glyph images, expert system facilities, and, in future, calculation of glyphs similarity and development of digital corpus for analysis of similarity of the contexts on the fly. Another possible direction of further investigations is confirmation of deciphering results using large corpus data.

## **Presentación del *Diccionario Coruña de la lengua española actual***

Porto Dapena, José-Álvaro; Conde Noguerol, Eugenia; Córdoba Rodríguez, Félix; Muriano Rodríguez, Montserrat

### 3. Reports on Lexicographical and Lexicological Projects

The *Diccionario Coruña de la lengua española actual* has been a work in progress since the year 2000. It is a monolingual dictionary of current standard Spanish language that covers both European and American Spanish. One of the main features of our dictionary is the possibility of two access methods—alphabetically/semasiologically and onomasiologically. Our work begins with an alternative scheme of the structure of a lexical semantic field. This field guides the whole process involving the compilation of the semasiological section. It is useful, for instance, to separate real meanings—i.e. invariant or paradigmatic meanings, but not senses—although different senses will be present in the dictionary under its relating meaning. We understand that there are different meanings in a word when it belongs to different lexical paradigms. The verb *componer* in *Albéniz compuso Iberia* belongs to the field of the verb *crear* ‘to create’, but in *El relojero compuso el reloj* the verb means *arreglar* ‘to repair’. The paradigmatic section will become a structural dictionary of the Spanish language. We are not trying to create a thesaurus, but rather to describe the structure of Spanish vocabulary applying the linguistic criteria of structural semantics. This structure will be a set of trees—one for each field—showing the semantic relations: synonyms, hyponyms, hypernyms,

meronyms, etc., as well as relations like causativity. Every meaning in the alphabetical section is linked with these trees.

### **Pedagogical Criteria for Effective Foreign Language Learning: A New Dictionary Model**

Pujol, Dídac; Masnou, Joan; Corrius, Montse

#### 3. Reports on Lexicographical and Lexicological Projects

This paper presents the pedagogical criteria used in the making of the *Easy English Dictionary with a Catalan-English Vocabulary* (EED), a new dictionary model for lower intermediate learners of English as a foreign language. The dictionary described renders an account of the philosophy and the results of a specific lexicographical project centred on English as the L2 and Catalan as the L1. The pedagogical criteria on which the EED is based are: structural criteria, linguistic criteria, cultural criteria and illustration criteria. The paper examines the treatment that each of these four aspects has received in different types of dictionary and, after pointing out their weaknesses and limitations, proposes a new dictionary model that seeks to promote a more effective learning of foreign languages. The most innovative aspect of the EED concerns its structure: the EED is a bilingualized dictionary—i.e. it combines the advantages of both monolingual and bilingual dictionaries, but unlike in classical, immediate bilingualized dictionaries, in the new—deferred—dictionary model the L1 translation does not minimize the L2 definition. The EED also takes advantage of the L1 language and culture, something which the vast majority of dictionaries for L2 learning do not do: the new model uses L2 words similar to L1 ones as well as cultural referents familiar to the L2 learner. Finally, the new dictionary model presented in this paper considers illustrations as an important means of contextualization and linguistic production.

### **On The Lexis of Cloth and Clothing Project**

Rutten, Stuart Nels

#### 3. Reports on Lexicographical and Lexicological Projects

A proposal for discussing the goals, limits, benefits and problems of creating a multilingual dictionary, using the web-based Lexis of Cloth and Clothing Project as a basis for consideration. Using PowerPoint slides and examples from ongoing work, the presentation will demonstrate both the methods in use for the dictionary and will raise questions regarding lexical practice when developing dictionaries for describing the lexis of multilingual communities.

### **ISLEX—An Icelandic-Scandinavian Multilingual Online Dictionary**

Sigurðardóttir, Aldis; Hannesdóttir, Anna; Jónsdóttir, Halldóra; Jansson, Håkan; Trap-Jensen, Lars; Úlfarsdóttir, Þórdís

#### 3. Reports on Lexicographical and Lexicological Projects

This paper presents ISLEX, an inter-Nordic project based in Reykjavík, Iceland, with partners in Gothenburg, Bergen and Copenhagen. The aim of the project is to develop an online dictionary site with Icelandic as the source language and the three Scandinavian languages— Swedish, Norwegian (with two official standards) and Danish—as the target languages. The dictionary is planned to contain 50,000 lemmas, with a development period of six years. In 2011, or possibly sooner, the site will be publicly available on the Internet, free of charge. In this article, the main features of the project are presented with particular emphasis on database design, editorial principles and priorities.

### **e-LIS: Electronic Bilingual Dictionary Italian Sign Language-Italian**

Vettori, Chiara; Felice, Mauro

#### 3. Reports on Lexicographical and Lexicological Projects

This paper presents the design of e-LIS (Electronic Bilingual Dictionary Italian Sign Language (LIS) - Italian), an ongoing research project at the European Academy of Bolzano started in 2004. It is the first attempt to build a sign-language reference dictionary that contains definitions and examples in the sign language itself and which offers a search engine that guides the user in the process of reconstructing and retrieving the sign he is looking for. Thus, not only can users go from Italian to LIS, but also from LIS to Italian.

### **Verbs of Science and the Learner's Dictionary**

Williams, Geoffrey

#### 3. Reports on Lexicographical and Lexicological Projects

This paper looks at how the verbs of science are displayed in the OALD and then compares them to a specialised corpus. The individual entries will be studied to see whether the scientific aspect is signalled, how the definitions are structured, and if implicit information is carried in the examples. The examples are analysed using Halliday's Systemic and Functional Grammar (SFG). The results in each section are compared with usage cases found in a subcorpus from the British National Corpus and in a specialised corpus.

## **La définition dans les dictionnaires bilingues: problèmes de polysémie et d'équivalence interlangues**

Bouchaddakh, Samia

### 4. Bilingual Lexicography

In this paper, we intend to tackle the issue of the definition in the so-called “encoding bilingual dictionaries” or “active bilingual dictionaries”. We focus more specifically on French-Arabic dictionaries. Our main objective is to demonstrate the interest of definition, precisely the one based on the principles of the Explanatory and Combinatorial Lexicology, for both the lexicographer and the user of the active bilingual dictionary. This kind of definition allows us to identify the internal structure of lexical meaning, to select the best equivalent and to make explicit the relations of polysemy and equivalence between the two languages.

## **La place de la morphologie constructionnelle dans les dictionnaires bilingues: étude de cas**

Cartoni, Bruno

### 4. Bilingual Lexicography

In this study, we questioned the role and the place of constructional morphology in bilingual dictionaries. As it is the case with monolingual dictionaries, it is becoming increasingly common to find morphological elements in bilingual dictionaries. Presenting such elements in a bilingual context, however, raises questions as to representations of meaning, translation selection and choice of examples. We show a comparison of the treatment of productive prefixes of Italian in 5 bilingual dictionaries—French was the target language in three of them, English in the other two. This comparison shows the differences in the treatment of prefixation and its coverage. First, we notice that not all the prefixes are represented in these dictionaries, and that some lexical elements are labelled as prefixes even if this status is contested. In terms of treatment, we examine in particular aspects of polysemy and of multiple translations. Regarding the polysemy of some prefixes, many dictionaries simply avoid marking the difference, while others add specific sense indicators. The most frequently used method for presenting the translation of prefixes, however, is through examples. In analysing these examples, we notice the inadequacy of certain translation equivalents, especially from a productive point of view. There is a noticeable absence of any information on the productivity of these constructional elements. From the perspective of understanding neologisms and sometimes even producing them, this lack of precision is regrettable.

## **Friend or Enemy? A Case Study of Lexical Comparison between Italian, German and Japanese Bilingual Dictionaries**

Costantino, Mauro

### **4. Bilingual Lexicography**

This work aims to demonstrate with a case study of bilingual Italian-German, Italian- Japanese, and German-Japanese dictionaries, exactly to what extent lexical drifting can manifest itself. Through a few examples, the study will raise issues of gender and cultural problems in bilingual dictionaries, and the peculiar case of a distant language like Japanese in comparison to German and Italian. The importance of considering both cultural and background knowledge when both building and consulting bilingual dictionaries will be stressed, in order for one not to obtain an outcome contrary to expectation. Finally, suggestions about how to overcome these problems will be made, with consideration given to the difficulty of dealing with strongly culturally-bound terms and their meanings.

## **User-friendly Dictionaries for Zulu: An Exercise in Complexicography**

de Schryver, Gilles-Maurice; Wilkes, Arnett

### **4. Bilingual Lexicography**

In this paper the main features of Bantu lexicography are analysed through several case studies of Zulu dictionary features. Examples from both existing dictionaries as well as a forthcoming reference work are used in the analysis, which develops from verbs and nouns, gradually including more word classes, and ending with a detailed study of possessive pronouns. The latter serves as one example of the complex mappings that occur in the creation of bilingual dictionaries where the two languages involved have very different grammatical structures. In this case, one concept—that of a possessor and its possession—has only a few members in English, but hundreds in Zulu. It is shown how one can deal with such a mass of data in a structured, systematic and linguistically-sound way, all the while aiming to produce a user-friendly end product. All the members of this single concept are collectively referred to as a paradigm, and it is indicated that some members are homonymous with members of other paradigms—a fact which exponentially complicates the dictionary treatment. Several suggestions are made for the lexicographic treatment of conjunctively written Bantu languages, and all the claims, as well as all the data, are based on facts derived from a large general- language Zulu corpus.

## **Systemhaftigkeit in zweisprachiger Lexikographie: Zur Darstellung deutscher und russischer Possessivpronomen**

Dobrovolskij, Dmitrij; Sarandin, Artem

### 4. Bilingual Lexicography

In this paper, we discuss systematic approaches to the lexicon representation in bilingual dictionaries. For empirical data we take how possessive pronouns are treated in the *New Comprehensive German-Russian Dictionary (NCGRD)*. Every type of pronouns builds a closed class of words, and its members have to be described in the same terms and in the same format. In the ideal case, each deviation from lexicographic uniformity must be understood as a signal that a given word displays unique linguistic features as compared to other members of the same class.

An additional difficulty is that the German and the Russian pronoun systems are arranged according to non-identical principles. Therefore, every German possessive pronoun can be translated into Russian not only by its "regular" equivalent (*мой, твой, его* etc.), but also by the pronoun *свой*, under certain syntactic and discursive conditions. These conditions have to be explicitly stated in the entry. There are also contexts where a given German possessive pronoun has to be omitted in the Russian translation, cf. *hast du (dir) deine Hände schon gewaschen? – ты уже вымыл руки? hast du schon mit deiner Mutter gesprochen? – ты уже поговорил с матерью?* This phenomenon clearly depends on the semantic class of the noun modified by the pronoun in question, i.e. it is rule-governed; these rules need to be declared in the lexicographic description.

The pronoun *mein* in certain contexts must be translated not by *мой ты* but by *наш* 'our'; cf. *meine Fakultät – наш факультет*. The reason is that the Russian word *мой* 'my', being different from the German word *mein*, denotes exclusive possession, i.e. *мой* means 'mine and only mine'.

All these cases will be illustrated by entries from NCGRD. They demonstrate far-reaching uniformity, so that every deviation from the uniform format—as compared to both various possessive pronouns in German and their Russian equivalents—is a meaningful constituent of their lexicographic representation.

## **La equivalencia en los diccionarios bilingües: un enfoque semántico**

Fernández Fernández, Juan

### 4. Bilingual Lexicography

In this paper, we present a proposal to analyze bilingual dictionaries' lexical equivalents. Lexical correspondence between two languages can be analyzed

adopting the point of view of different linguistic domains—e.g. pragmatics or syntax. Our approach is based upon semantics. The aim of this paper is to find out ways of discovering conceptual differences between bilingual dictionaries' equivalents which are supposed to have the same meaning. For this reason, we put into practice a semantic analysis based on semantic decomposition of monolingual dictionaries' definitions of the equivalents given by bilingual dictionaries. We make use of the lexicographical definition, the semantic metalanguage—especially, the metalanguage which consists in the up-to-date Wierzbicka's semantic primes, and corpus linguistics. These conceptual analysis' results are shown by means of conceptual trees which relate the different concepts which are part of the given equivalents' definitions. Thanks to the Wierzbicka's semantic primes, which have proven to be universals, we can follow an objective reasoning behind the differences and similarities between proposed lexical correspondences by these dictionaries. As a result, we can gain valuable conceptual knowledge that will sensitize us to languages' lexical variety and richness. This is something that cannot be easily shown in commercial bilingual dictionaries by reasons of time and pressures of the publishing market. Thus, our proposal is a small contribution to reflect on these dictionaries' functions regarding their users—language learners or language professionals. A further step is to provide an alternative to their common conceptual structure, in order to compare the lexicon of two languages in a more reasonable way, or in accordance to language's anisomorphism.

### **Le Casse-tête des dictionnaires bilingues pour traducteurs: le cas des dictionnaires arabes bilingues**

Franjié, Lynne

#### 4. Bilingual Lexicography

Translators have long called for translators' bilingual dictionaries that would include ready- to-use equivalents. In order to make them, lexicographers must ponder on the bilingual dictionary as a translator's working tool, hence not only from a lexicographic point of view, but also from a translation one. Studying the translations included in dictionaries—such as Arabic bilingual ones—shows that they are problematic as they are *out-of-context* and *transitional*. The issue becomes even more complex when the entries at stake are culture-related, for it is common knowledge that 'shared culture'—for example, that related to social and religious realities—often varies from one culture to another, as it is the case between Arab and French cultures for instance. Semantic voids are numerous in these cases and the lexicographer finds himself compelled to make difficult choices. It is these choices that this paper means to examine by analysing cultural entries in largely-used Arabic-French and Arabic-English bilingual dictionaries. Studying these cultural

entries entails determining the types of translations included in the dictionaries. One could then conclude that Arabic bilingual dictionaries are in fact meant for translators although they are deficient in some ways. One solution would be to include authentic functional translations, namely by using parallel corpora from which translations can be extracted, thus enriching dictionary entries.

### **What to Say about *mañana*, *totems* and *dragons* in a Bilingual Dictionary? The Case of Surrogate Equivalence**

Gouws, Rufus; Prinsloo, Danie J.

#### **4. Bilingual Lexicography**

There are frequent instances in any given language pair where a suitable translation equivalent is not available to be treated as source and target language in a bilingual dictionary. This is known as zero equivalence and can be regarded as the most complex type of equivalence to be dealt with in a bilingual dictionary. This paper will focus on the various ways in which lexicographers of different dictionaries deal with the lack of equivalence and the subsequent use of surrogate equivalents. There are a number of strategies that the lexicographer can use when dealing with instances of zero equivalence, e.g. the use of glosses, paraphrases, illustrations and even text boxes with lexicographic comments. This paper suggests different types of surrogate equivalents based on user needs, and it will be done in accordance with the relevant dictionary functions, i.e. the cognitive function and the communicative functions of text reception, text production and translation. A linguistic gap can be identified when the speakers of both languages are familiar with a certain concept but when one language does not have a word to refer to it, whereas the other language does have such a word. A referential gap can be postulated when a lexical item from language A has no translation equivalent in language B. This would be because the speakers of language B do not know the referent of the lexical item from language A. Acknowledging different degrees of complexity in the relation of surrogate equivalence leads to a tiered view of the concept. The first level in the hierarchy provides for linguistic gaps where a mere gloss or brief paraphrase of meaning will suffice. More complicated are the gaps where the surrogate equivalent also has to provide grammatical guidance. The top tier in the hierarchy provides for referential gaps where taboo, culture-specific or sensitive values have to be expressed.

## **Du support d'information à l'outil lexicographique: la lexicographisation du guide touristique**

Leroyer, Patrick

### **4. Bilingual Lexicography**

The development of lexicographic products for tourists is one of the most productive lexicographic activities in the world, with the publication of paper and online bilingual travel dictionaries, phrase books, and tourist guides often containing a dictionary component. Additionally, software companies propose multilingual, downloadable dictionary solutions that can be printed on demand or consulted via a PDA or a WAP phone. There are two explanations to this lexicographic infatuation: the huge expansion of tourism world-wide and the extensive communicational and knowledge-oriented informational needs of tourists. However, metalexicography has shown very little promise to this field of lexicography, and has solely dealt with the communicative needs of tourists. In this contribution, I will outline a new lexicographic method that can be used to satisfy the aforementioned needs of tourists—also, namely lexicographisation—which is the lexicographic transformation of tourist guides performed to ensure fast and easy access to user and situationally-adapted information.

## **Lexical Entries and the Component of Pronunciation in Tshivenda Bilingual Dictionaries**

Mafela, Munzhedzi James

### **4. Bilingual Lexicography**

*Lexical Entries and the Component of Pronunciation in Tshivenda Bilingual Dictionaries* Pronunciation is defined by Allen (1990) as the way in which a word is pronounced, especially, with reference to a standard. It involves a set of symbols, each of which always represents the same sound. Languages pronounce orthographic symbols differently. In some languages, orthographic symbols written identically are pronounced differently. Tshivenda is characterized by orthographic symbols which are written identically, but can be pronounced differently. These are orthographic symbols such as *tsh*, *ts*, *tsw* and *pf*. The same orthographic symbol can be pronounced as an aspirated sound or ejected sound. For example, the orthographic symbol *ts* can be pronounced as [tsʰ] in *tsika* (to press down) or [tsh] in *tsimbi* (metal). Poulos (1990) says that the actual pronunciation is determined by the words in which the orthographic symbols are used. Definitions of headwords in a dictionary consist of many components, for example word category, morphology, pronunciation, etymology, meaning and illustrative examples. The pronunciation

element becomes a necessity for bilingual dictionaries because the addressees of these dictionaries may be learners of a foreign language. “Pronunciation is, after all, the integral part of the lexical item”, as Sobkowiak put it in 2003. Giving the meanings of words is often thought to be the main purpose of a dictionary. It should be also noticed, however, that “the dictionary also contains other areas of information useful to the user” (Underhill 1980). Knowledge about pronunciation helps in checking any spelling the user is not sure of. Almost all Tshivenda dictionaries are bilingual and are therefore learner’s dictionaries. The compilers of the dictionaries did not include the component of pronunciation in the definition of lexical items. Therefore, learners of Tshivenda find it difficult to pronounce orthographic symbols which denote more than one phonetic sound. This presentation seeks to highlight the lack of pronunciation component in Tshivenda bilingual dictionaries and its effects on learners of the language. Three Tshivenda bilingual dictionaries will be used to illustrate some points in this regard.

### **L'accès aux Séquences Figées dans les dictionnaires électroniques bilingues Français – Italien**

Murano, Michela

#### **4. Bilingual Lexicography**

This paper presents the results of our research on a group of electronic French-Italian, Italian-French dictionaries on CD-Rom—DIF, Boch, Garzanti Clic, and Garzanti interattivo. We examine whether the characteristics of the electronic support can influence the access to the fixed sequences. This work deals particularly with the importance of diversified typography and new types of complex search—e.g. full-text search—which are now available for dictionary users.

### **Méthode sociolinguistique d'étiquetage du niveau de langue dans les dictionnaires bilingues (sur l'exemple d'un dictionnaire français-ukrainien)**

Shevchenko, Natalya

#### **4. Bilingual Lexicography**

This article describes a new sociolinguistic method in the labelling of unconventional units in bilingual dictionaries. This study was undertaken as part of the preparation of a French- Ukrainian dictionary of unconventional language.

## **On the Presentation of Onomastic Idioms in Bilingual English-Polish Dictionaries of Idioms**

Szerszunowicz, Joanna

### **4. Bilingual Lexicography**

The paper discusses the lexicographic description of onomastic idioms in contemporary English-Polish dictionaries of idioms, with a special focus on the cultural character of the onymic component. Onymic idioms are distinguished as a group of particular interest for lexicographers, since onyms tend to be culture-bound elements of international, national or local character. Thirteen English-Polish dictionaries of idioms have been analyzed so that the presentation of onymic idiomatic expressions in such lexicographic works could be discussed. The macro- and micro-structures of such dictionaries are analyzed in order to identify the problem areas in the bilingual description of onomastic idioms. From the cultural-linguistic point of view, two methods of presenting onomastic idioms are observed in the dictionaries, i.e. the inclusion of cultural information regarding the onym or the exclusion of such information. In the case of the inclusion of cultural information, the lack of consistency is common in one dictionary, i.e. some of the onyms are commented on, while others are not described at all. Since onyms tend to be culturally-specific components of idioms, cultural information is essential to ensure a proper understanding of the idiom. The problem of insufficient lexicographic description of such fossilized phrases is presented in order to draw attention to the need for the creation of an onomastic idiom dictionary, enabling both users and advanced learners (of English) to have an insight both into the language and the culture. Bearing in mind that idioms undergo various modifications when used in particular contexts, such an approach to describing onomastic units of idiomatic character renders it possible for the user to acquire a proper command of idioms containing onomastic components.

## **QRcep: A Term Variation and Context Explorer Incorporated in a Translation Aid System on the Web**

Abekawa, Takeshi; Kageura, Kyo

### **5. Lexicography for Specialised Languages - Terminology and Terminography**

In this paper we describe the method of exploring term variations and the contexts in which terms occur using the Web, to help English-to-Japanese translators working online. Many English-Japanese terminological dictionaries are available in electronic form, but most of them do not provide rich examples of terminological use including variations. This is a problem for translators, who may not have

sufficient knowledge on the use of terms in a specific subject they are translating. In order to augment this information gap, we have developed a system that explores actual use of terms using the information on the Web.

The system proceeds as follows:

1. when an electronic text in source language (English) is given, the system automatically looks up entries in terminological dictionaries including their variations, using the variation expansion rules;
2. map the English entry to the Japanese translations;
3. expand variations of Japanese terms on the basis of Japanese variation rules;
4. search the Web and provide actual use of the term including variations within the actual context. For variation expansion, we are using Fastr Platform and defining corresponding rules for English and Japanese variations. The system is incorporated into the system that helps online volunteer translators and augments the terminology look-up functions.

### **ECODE: A Pattern Based Approach for Definitional Knowledge Extraction**

Alarcón Martínez, Rodrigo; Sierra Martínez, Gerardo; Bach Martorell, Carme

#### 5. Lexicography for Specialised Languages - Terminology and Terminography

In this paper we present a pattern-based approach to the automatic extraction of definitional knowledge from specialised Spanish texts. Our methodology is based on the search of definitional verbal patterns to extract definitional contexts related to different kinds of definitions: analytic, extensional, functional and synonymic. This system could be a helpful tool in the process of elaborating specialised dictionaries, glossaries and ontologies.

### **Environmental Terminology in General Dictionaries**

Alonso Campo, Araceli

#### 5. Lexicography for Specialised Languages - Terminology and Terminography

This paper discusses how some specific Environment-related terms commonly used in general discourse have been represented in monolingual and learner's dictionaries in Spanish. Our discussion falls within wider research on the characterization of Environment-related lexical units and the relationship between specific domain terminology and lexicographic representation. We briefly compare the information provided in general language dictionaries of Spanish with that found in other

lexicographical traditions—for instance, in the English tradition— and find a lack of precision in lexicographic practice in relation to Environment-related terms. We outline some guidelines for improved representation of these units.

## **Gestor de terminologia multilingüe d'accés lliure**

Bover Salvadó, Jordi; Grané Franch, Marta

### 5. Lexicography for Specialised Languages - Terminology and Terminography

Attending the demand of several sectors asking for a terminology management tool suitable for specific or personal use, TERMCAT has developed a free-access terminology manager, available at our website ([www.termcat.cat](http://www.termcat.cat)). The tool, addressed to anyone interested in carrying out terminographic work, enables the management of any multilingual project that involves the compilation of terms in different languages and the systematization of concepts from different fields of knowledge. We would like to underline that every user would be able to customize the terminology management tool according to the project features and their personal needs, in order to speed up the process of data creation and modification. The most relevant contributions of the TERMCAT free-access terminology management tool are the following:

- Organizing the information in conceptual files.
- Including or deleting denominations, definitions, notes, contexts, observations and their attributes—grammatical category, range, linguistic hierarchy, source—in n languages.
- Consulting and modifying the properties of a dictionary—name, description, languages, ordering.
- Creating, maintaining and consulting the concept structure of a dictionary.
- Organizing the files in thematic or alphabetical order, and according to language.
- Allowing search based on the combination of a wide range of criteria: denominations, definitions and notes; and also according to language, field structure, hierarchy, grammar category or source.
- Consulting the alphabetical or thematic index of a dictionary.
- Consulting the files by edition mode or consultation mode.
- Importing and exporting a selection of files in several formats.

## **TESAURVAI: Extraction, Annotation and Term Organization Tool**

Cardeñosa, Jesús; Gallardo, Carolina; Maldonado-Martínez, Ángeles; Vergara, Jorge

### 5. Lexicography for Specialised Languages - Terminology and Terminography

TESAURVAI is a tool for extracting, annotating and organizing terms from a collection of digital documents. The main contribution of TESAURVAI is the unification of a term extractor and a thesauri builder in the same tool. The term extractor identifies terms, words and phrases in the input digital texts that are transferred to the thesaurus builder. TESAURVAI follows the international standards for the construction and management of thesauri, and it provides the following facilities: on the one hand, it is a tool to create thesaurus from scratch, allowing for the extraction, creation, edition and annotation of terms, as well as providing a user-friendly interface for establishing relations between terms and performing basic or advanced searches of terms. On the other, it is a tool to manage several thesauri and to import and export existent thesauri from text or XML files. Finally, TESAURVAI can build alphabetical, hierarchical and permuted indexes to be printed or exported as reports. TESAURVAI has been developed in Java and requires an external database to store the user's thesauri. The tool is compatible with any database manager provided with a Java Database Connectivity (JDBC) file, such as MySQL or Postgres. This tool has been developed within the framework of the PATRILEX (HUM2005-07260/FILO) project, sponsored by the Spanish Minister of Education. Currently, TESAURVAI is in a provisional version. A new version of the tool, which will be accessible on the Internet, will be available in July 2008.

## **Filling the Gap: A Three-Language Philological Dictionary Based on Contexts from Authoritative Sources**

Cignoni, Laura

### 5. Lexicography for Specialised Languages - Terminology and Terminography

This paper describes the methodology adopted for the creation of a multilingual—English- Italian-French—philological dictionary, designed to meet as far as possible the requirements of users in the field of philology, who need to use specific terms in a language other than their own. The project is addressed to graduate and post-graduate students, tutors and scholars, translators and interpreters for whom a glossary of specialised terms relative to a given universe of discourse is essential. The dictionary defines a variety of terms associated with philology and extends to other closely connected disciplines such as textual criticism, codicology, palaeography, epigraphy, papyrology, genetic criticism, etc. The three-language glossary is arranged in conventional form—each lexical entry listed alphabetically—and the

English terms are followed by their equivalents in Italian and French. The Greek or Latin words from which many philological words have derived and frequently used to refer to a particular concept or phenomenon are also included. The project involves recording not just the term but also a brief contextualized definition in each language, accurately quoting the source, certified and scientifically reliable, from which the information was drawn. Alongside these definitions, a number of other contextualizations appear, also derived from authoritative sources, and different types of illustrations—e.g. manuscripts, stems, images of people and places—relative to the terminology are provided. The textual data and images will be included into an application (Alpha version) of the PINAKES project, released in March 2007, which is able to deal with different types of information—text, scientific objects, tables or graphics. This ongoing dictionary project, at present covering a total of around 1000 words, is constantly enriched with new entries, definitions and contextualizations in the different languages.

### ***Risotto, spaghetti, vino: Ingredients for a Good Gastronomic Dictionary***

Corino, Elisa

#### 5. Lexicography for Specialised Languages - Terminology and Terminography

Gastronomy is commonly recognized as a basic “ingredient” of culture and tradition. As a central axis of various cultural components, food is a common denominator connecting both the Fine Arts and Science, as well as History, Anthropology, Sociology, etc. Additionally, gastronomy has become one of the world’s most important professions and continues its ascent. Italy has recently witnessed this growth as being the origin of the Slow Food movement and the hosting of the first University of Gastronomic Sciences, where specific lecture courses expressly focus on culinary jargon and on the linguistic, typological and historical analysis of menus, recipe books and recipes alike. The increasing need for thorough glossaries and dictionaries devoted to detailed studies of the subject is apparent. This paper is meant to deal with the vocabulary connected to food in its broad sense, and will attempt to provide a cross section of the lexicographical state of the art and propose a possible original source to be held up as a model for gastronomy dictionaries: Newsgroup corpora on cooking. The *Langenscheidt Praxiswörterbuch Gastronomie Italienisch* (2005) will be investigated as an example of an exhaustive dictionary: its word list compared with the 500 most frequent occurrences of nouns, adjectives and verbs in the NUNC-cooking (Newsgroup UseNet Corpora), amongst both its Italian and German versions. Finally, a case study on adjectives describing wine is presented to suggest new entries for a wine glossary.

## **Léxico específico de la piel. Presentación de un proyecto terminográfico**

García Antuña, María

5. Lexicography for Specialised Languages - Terminology and Terminography

This project is framed within the project I+D of the MEC Linguistic strategies applied to social communication: study of communicative necessities and design of materials in the social environment of medicine, administration and business, the head of which is Doctor Miguel Casas Gómez. For the realization of this project we have had the collaboration of the business world, according to an agreement with the Association of Andalusian Leather Manufacturers (Asociación de Empresas Andaluzas de la Piel, EMPIEL) and the Technological Center of Leather. A specific agreement and a contract of service supply is in the process of being signed, all with the support and unfailing advice of the Oficina de Transferencia de Resultados de Investigación (OTRI) of the Vicerectorate of Research, Technological Development and Innovation at the University of Cádiz. The main objective of this project is the management of a base of terminological facts of leather work that permits the development of effective translation tools and the regulation of the specific language in these fields. The introduction of this lexicon is important from the formative point of view, since a lexicographic work does not exist today in which the specific terminology of leatherwork can be investigated. This lack of specific terminology is an obstacle for the communication of knowledge among the professionals involved. Furthermore, this work will complete other advancement objectives, since it is effective and useful to describe the characteristics of a product, to create a positive attitude towards the sector in the customers. This will compensate the introduction into the market of low quality products from countries like China or India, and will contribute to convincing the customers of the superiority of the offer versus that of the competitor. Furthermore, this terminological project can contribute significantly to the regulation of commercial and technical language within the sector.

## **Slovene Terminology Web Portal**

Gorjanc, Vojko; Krek, Simon; Vintar, Špela

5. Lexicography for Specialised Languages - Terminology and Terminography

Work in the field of terminology is extensively supported worldwide as it enhances the transfer of science and technology. In Slovenia, there is a series of terminology-related activities running, and a significant number of terminology dictionaries and terminological data exist, but they are methodologically heterogeneous and often unavailable for public use. Traditionally, terminology work in Slovenia is closely connected with other activities in the field of lexicology and lexicography, especially

regarding the methodological approach to the compilation of dictionaries of specialised languages. Therefore, terminology work is mostly regarded only as a process involving the compilation of dictionaries. The paper presents the Slovene Terminology Web Portal project. The main objective of the project is to develop the Slovenian terminology portal to offer basic information on the principles of terminological work and to present a terminological database in a unified format. In the core of the presentation, there is a process of conversion of different types of existing terminology data from different sources into XML format with a simple DTD/schema and from there to unified TBX database. Simultaneously, the feasibility of linking textual resources and the extraction of term candidates with the terminological database is also shortly presented.

## **Prototypes and Discreteness in Terminology**

ten Hacken, Pius

### 5. Lexicography for Specialised Languages - Terminology and Terminography

Characterizing the nature of terms in their opposition to general language words is one of the tasks of a theory of terminology. It determines the selection of entries for a terminological dictionary. This task is by no means straightforward, because terms seem to have different properties depending on the field that is studied. This is illustrated by a brief discussion of examples: terms in mathematical linguistics, traffic law, piano manufacturing, and non-terms in the reporting of general experiences. Two properties can be derived from these discussions as candidates for the delimitation of terms from general words. Firstly, the degree of specialization. This property distinguishes specialized expressions in mathematical linguistics and in piano manufacturing from non-specialized expressions in traffic law and reporting general experiences. Secondly, the lack of a prototype. In mathematical linguistics and in traffic law, the definition of terms concentrates on the boundaries of the concept. In piano manufacturing and in reporting general experiences, concepts have a prototype and fuzzy boundaries. Defining the word *term* as a disjunction of the two properties implies that it is a less coherent concept than general language word, because it is only the complement of the latter. When the two properties are considered in isolation, it can be shown that the degree of specialization is a gradual property whereas the lack of a prototype is an absolute property. Whether or not we choose to use the name *term* for it, the latter property identifies a concept that is ontologically different from general vocabulary. I will reserve the name *term* for concepts that do not involve prototypes and call the professional expressions in piano manufacturing specialized vocabulary. By focusing on the boundary instead of the prototype, a terminological definition creates an abstract object for which there is no equivalent in general language words. Whereas general language words only

exist in the competence of the speakers, the abstract object associated with a term can exist independently of the knowledge of individual speakers. There are interesting parallels between the nature of these abstract objects and the nature of a piece of music. The creation of such an object on the basis of general language words can proceed by the selection of properties or the choice of a specific boundary on a scale.

### **New Voices in Bilingual Russian Terminography with Special Reference to LSP Dictionaries**

Karpova, Olga; Averboukh, Konstantin

#### 5. Lexicography for Specialised Languages - Terminology and Terminography

The article is devoted to the general review of modern bilingual LSP dictionaries in Russia. Main trends in current Russian bilingual terminography are distinguished through criticism of new types of LSP dictionaries of different subject areas with special reference to linguistic and encyclopedic reference books. Evolution of lexicographic description of different special domains in English-Russian and Russian-English terminography is being traced from humanitarian subject fields—economics and finance, business, law, mass-media and public relations, social work, immigration policy and the like—and natural sciences—biology, botany, physics, zoology—to technical disciplines—aviation, electronics, civil and nuclear engineering, etc.—and other subject areas—agriculture, architecture, philosophy, statistics, etc. Special attention is given to the analysis of Russian-English polytechnic dictionaries published in a new millennium showing the latest changes in Russian terminological vocabulary connected with borrowings of new terms and whole terminological systems—computers and new information technologies, logistics. Current developments and perspectives in progress in Russian bilingual terminology will be mentioned in the presentation.

### **LSP Dictionaries and Their Genuine Purpose: A Frame-based Example from MARCOCOSTA**

León Araúz, Pilar; Faber, Pamela; Pérez Hernández, Chantal

#### 5. Lexicography for Specialised Languages - Terminology and Terminography

A dictionary is written and designed for a specific addressee (*user group*). Primary considerations in this respect are users' profiles and the special needs of the user group (Bergenholtz and Nielsen 2006). User needs are inevitably linked to the knowledge level of potential readers, who have a situational context and engage in activities, which can be facilitated by lexicographic data. Such information significantly affects both the micro and macrostructural design of the lexical

resource, and is directly related to Wiegand's conception of genuine purpose (Wiegand 1998:52). These theoretical parameters dealing with users' profiles, users' needs and use situation should necessarily be reflected in the way information is packed in lexicographical entries, i.e. in the way definitions are organized and structured. This article examines how LSP dictionaries deal with this issue. The example chosen is the term *aquifer*. After a brief overview of how this term appears in current dictionaries, we show how it is represented in MarcoCosta, a frame-based lexical resource that facilitates the acquisition of specialized knowledge.

### **Marqueurs définitionnels et marqueurs relationnels dans les définitions du DAAFAPS**

Ligas, Pierluigi

#### 5. Lexicography for Specialised Languages - Terminology and Terminography

This paper analyzes relational and definitional markers and their function in meronymic, derivational and approximate definitions of nouns as they appear in the *Dictionnaire alphabétique et analogique du français des activités physiques et sportives*, currently under preparation. It is argued that definitional markers are semantically weak lexical substitutes with a metonymic or meronymic character, placed at the beginning of the definition and belonging to the same grammatical category as the defined lexical item. It is also argued that relational markers are words or groups of words whose function in discourse is to establish logical, spatial or temporal relations between two or more elements and which thus contribute to organize the definitional sentence and to illustrate the concept denoted by the lexical item. As mentioned supra, we have decided to exclude hyperonymic definitions—since they do not start with definitional or relational markers—and to concentrate on three types of definitions: meronymic—based on the relation between a whole and its parts, derivational—based on the relation between root and affixes, and approximate—that make use of markers such as *sorte de*, *espèce de*. We will analyze a corpus of such definitions and try to establish how these markers contribute to the fulfillment of the definition's role, by following mainly R. Martin's, E. Wüster's, J. Rey-Debove's, A. Auger's, A. Condamines's, E. Martin's definitional theories.

### **A Constructional Approach to Terminological Phrasemes**

Montero Martínez, Silvia

#### 5. Lexicography for Specialised Languages - Terminology and Terminography

Specialized discourse shows regularities in the lexical and syntactic patterning of terminological units. This fact, evidenced by corpus-based analysis, has spurred a number of studies on polilexical terminological units. In spite of the available

linguistic data, however, the systematic management of these units in specialized lexicography is still lacking. Apart from a few exceptions, terminological products, especially dictionaries, are inconsistent with their treatment of these units. Such arbitrary approaches are worthless within the context of the newer terminological knowledge bases. In this paper, we describe how the Lexical Grammar Model can offer an in-depth, principled description of such units. Meaning and grammar are seen as interdependent and complementary layers. So, the basic unit of grammar is a form- meaning pairing or construction that can be described as a conventionalized combination of form and meaning. In this vein, the lexical profile of a specialized concept is composed of constructions, which reflect its collocational patterns both at a lexical and a syntactic level. Thus, we use the umbrella term terminological phraseme (Meyer and Mackintosh 1994) to include entrenched, conventional combinations of linguistic units in the form of complex nominals and predicate-argument structures. These units are conceived as constructions codifying conceptual, experiential and syntactic information concerning the lexical concepts of a cognitive frame. Consequently, the frame is the element which constrains the potential relations holding between the lexical concepts, and the construals that the frame allows are only a subset of the construals allowed by the argument-taking heads. The basic qualia structure and the domain-specific relations account for such combinations and for the inheritance phenomenon. In sum, we present a theoretical and methodological approach that accounts for the lexical profiles of concepts in a consistent way, including the description of conceptual relations as well as the terms' combinatorial potential.

## **Bilingual Terminology Acquisition from Unrelated Corpora**

Nazar, Rogelio

### 5. Lexicography for Specialised Languages - Terminology and Terminography

This paper presents a simple yet effective technique for the extraction of term equivalents in different languages. In general, techniques for bilingual lexicon extraction have been related to the elaboration of parallel corpora and have yielded accurate results. However, parallel corpora of different domains and languages are not easy to compile. Because of this, some authors have explored techniques to extract a bilingual lexicon from nonparallel but comparable corpora, which are pairs of texts that are not exactly translations of each other but that roughly "talk about the same things". This paper describes an algorithm that performs bilingual terminology extraction without the need of large amounts of data; dealing with infrequent units; needing not the corpora to be comparable nor other resources like an initial bilingual lexicon to use as *seed words*. In spite of its simplicity, the results of this algorithm are comparable to those of the state of the art techniques, however it

supersedes them considering that it offers a domain and language independent method specially suitable for the extraction of specialized terminology, which is the most dynamic part of the lexicon and the most difficult to acquire.

## **El sistema métrico decimal en la lexicografía española del s. XIX**

Pascual Fernández, Luisa

### 5. Lexicography for Specialised Languages - Terminology and Terminography

The metric system is one of the clearest examples of the universal acceptance of scientific and technological vocabulary in nineteenth-century language. Its introduction into Spanish language coincided in time with its introduction in the other European languages. This vocabulary, however, has not always been rigorously included in dictionaries, as shown by dictionaries of the nineteenth-century. For this reason, I have decided to study the inclusion of the vocabulary related with the metric system in Spanish nineteenth-century dictionaries, this century being particularly interesting as far as the history of science and lexicography is concerned. The analysis is structured into two main parts. The first part is committed to the study of the already-mentioned nomenclature of the eleventh edition (1869), the twelfth edition (1884), and the thirteenth edition (1899) of the *Diccionario de la Real Academia Española de la Lengua*. The second part is devoted to the analysis of how metric vocabulary is incorporated in non-academic dictionaries—including the *Nuevo Tesoro Lexicográfico de la Lengua Española*. These parts are complemented with the comparison of that Spanish vocabulary with its French, English and Italian counterpart. In this sense, our focus lies in the European perspective. The conclusion of the research provides wide information about the first instance of metric vocabulary in Spanish within the European context and, consequently, we hope to shed some light on the way the analyzed dictionaries influenced each other. We hope to also be able to conclude the position of the Spanish Academy regarding this kind of vocabulary.

## **An English-Polish Glossary of Lexicographical Terms: A Description of the Compilation Process**

Podhajecka, Mirosława; Bielińska, Monika

### 5. Lexicography for Specialised Languages - Terminology and Terminography

In the present paper we describe the consecutive phases in the compilation of an English-Polish glossary of lexicographical terms, which is part of a larger dictionary project—still in the making. In doing so, we address some of the issues that made the compilation procedure methodologically difficult. On theoretical grounds, the main dilemma was whether lexicographical—i.e. mainly descriptive—or terminological/

terminographical—i.e. mainly prescriptive—principles should be followed, inasmuch as they result in different coverage, organisation and description of data. The most pertinent practical problem that we faced was, on the one hand, the variability of terms in English lexicographical discourse and, on the other one, the incompatibility of English and Polish terminological frameworks. It was therefore envisaged that, for the glossary to be used successfully in text reception, allowing alternative terms and determining various levels of equivalence between interlingual terms would be a necessity. The issues discussed here have been illustrated with selected English-Polish contrastive material.

### **Wissensdarstellung und Benutzerfreundlichkeit in einem zweisprachigen terminologischen Rechtswörterbuch: Der Fall Hochschulrecht**

Ralli, Natascia; Wissik, Tanja

#### 5. Lexicography for Specialised Languages - Terminology and Terminography

This paper presents the Italian-German Terminological Dictionary for University law in Italy and Austria. In particular, we will describe the microstructure of the dictionary and the typology of the given information with regard to the needs of the target group. The dictionary was produced and printed in 2007 by the Institute for Specialised Communication and Multilingualism of the European Academy of Bolzano on behalf of the Department for the Right to Education, University and Scientific Research of the Autonomous Province of Bolzano/Bozen-South Tyrol. The aim of this work is to compare the Italian and Austrian terminology of university law as well as to record their most recent changes and developments.

### **Palabras y términos “lingüística y contextualmente determinados”**

Sanz Espinar, Gemma

#### 5. Lexicography for Specialised Languages - Terminology and Terminography

Our first concern is the specificity of the terminology of the human and social sciences, which are said to be *language-dependent* or *culture-dependent*. However, we will consider the creation and the use of terms—as of all words—*language and context-dependent*. This assessment is contrary to traditional terminology theory, which considers terms to be univocal relations between concepts and their designation—contrary to words, and also considers terms from pure sciences and technics as language-independent and culture-independent. We will analyze what *language and context-dependency* mean for terms and no terms, for terms of more technical or positive sciences, as well as for terms of the human or social sciences. From a pragmatic point of view, the creation and use of any word or term is

supposed to be influenced by the context in which this word or term was created/used, so that they are linked, to some extent, to the creator—author-dependency—and his language— language-dependency, the culture—culture-dependency, the place—geography-dependency, the historical period—history-dependency, and the communicative aim—dependency on the communicative aim, which includes the type of circumstance or the person the speaker talks to. This process means that for translation and for terminographic purposes we will find some specificities in these cases, but we can formulate strategies to cope with them.

## **Lexicographic Document Templates: Text Genre Conventions in Corporate Lexicography**

Simonsen, Henrik Køhler

### 5. Lexicography for Specialised Languages - Terminology and Terminography

Communicators do not only need conventional lexicographic data offering assistance on the lexical and syntactic levels. Communicators also need information on text genre conventions offering assistance on the textual level. What good is it to correct terminology and equivalent constructions on the lexical and syntactic levels, if the text produced by the communicator does not adhere to the text genre conventions of the text genre in question? This challenge is addressed in this paper, which discusses a theoretical and practical proposal for converging text genre conventions and conventional lexicographic data in a corporate lexicographic reference work. The theoretical solution proposed is based on an eclectic convergence of theoretical considerations on corporate lexicography, internet lexicography and genre analysis. This theoretical triangulation has resulted in what is referred to as lexicographic document templates. Lexicographic document templates are necessary in a corporate context, because they supply the user with the "missing link" in effective corporate communication. The concept of lexicographic document templates has been incorporated in the search facility of the corporate lexicographic reference work, which means that the user performs a combined search for a lemma and a text type in the look-up process. In addition to the conventional lexicographic data offered in the lexicographic article, the user now also has access to lexicographic document templates offering a typical example of the text genre in question and also essential text genre- related information such as information on the communicative purpose, move structure and rhetorical features of the text genre. The concept of lexicographic document templates has been implemented in practice in the corporate lexicographic reference work ZooLex at Copenhagen Zoo, and experts as well as non-experts appreciate the added value of information on text genre conventions in connection with conventional lexicographic data.

## **Terminology Practice in a Non-standardized Environment: A Case Study**

Taljard, Elsabe

### 5. Lexicography for Specialised Languages - Terminology and Terminography

Terminology as independent discipline, as well as its practical applications is not yet well established for the South African Bantu languages. The aim of this paper is to illustrate some strategies that are currently employed to ensure sound terminology practice in a non- standardized environment, and at the same time contribute to terminology and language standardization of Northern Sotho, a language of lesser diffusion spoken by approximately 4 million people in the Republic of South Africa. Within the South African context, standardization of terminology needs to contribute to the elevation of the status of a previously disadvantaged language to that of fully-fledged official language. In the case of Northern Sotho, apart from its direct impact on terminological development, any form of terminological activity therefore must contribute to terminological standardization, and within the broader sociolinguistic context, to language standardization, since Northern Sotho has not yet been fully standardized. This paper presents the results of a case study based on the compilation of a quadrilingual explanatory LSP dictionary for chemistry in order to illustrate that sound terminology practice is indeed possible in an environment where the terminological infrastructure is not ideal, and that it can contribute not only to terminology development and standardization as such, but also over a wider spectrum to standardization of an as yet only partially standardized language.

## **La reforma pombalina de la enseñanza: de la *Prosodia* de Bento Pereira al *Parvum Lexicon* de Pedro da Fonseca**

Borges, Ana Margarida

### 6. Historical and Scholarly Lexicography and Etymology

The end of Jesuits' control over education in Portugal in 1759 and the consequent remodelling of Portuguese education led to new and promising procedures at social and state levels such as dismissal and nomination of teachers, syllabus planning, elaboration and fiscalization of didactic material. In fact, it is in this context of education reforms and of modification of the structure and customs of the Portuguese society that Marquês de Pombal forbids Bento Pereira's *Prosódias*, a group of dictionaries that supported the teaching of Latin and Portuguese. Therefore, the need to compose urgently a new dictionary that might answer the needs of Pombal's aims in relation to education and that could at the same time fulfil the capacities of school usage, emerges.

This new dictionary would be Pedro José da Fonseca's *Parvum Lexicon Latinum*, which would be concluded and published three years later, in 1762, under royal order. The simple idea of the usefulness of a little dictionary that made the learning of Latin and Portuguese easier represented the beginning of the modernization of the bilingual Latin-Portuguese lexicography that would, later on, allow for the improvement of the techniques used in the making of dictionaries.

The aim of this investigation is to establish a link between Bento Pereira's *Prosodia* and Pedro da Fonseca's *Parvum Lexicon*, by pointing out the main innovations in lexicography present in the nomenclature and in the structure of the articles.

### **Un diálogo implícito: la relación entre Joan Corominas y José Luis Pensado a través de su producción lexicográfica**

Cotelo García, Rosalía

#### 6. Historical and Scholarly Lexicography and Etymology

Our paper is part of a broader research into the profound change that transformed the *Diccionario Crítico Etimológico de la Lengua Castellana* (1954) by Joan Corominas, into the *Diccionario Crítico Etimológico Castellano e Hispánico* (1980) by Joan Corominas and José Antonio Pascual, the latter being a considerably more comprehensive and extensive edition. Our proposal stresses the importance of the implicit dialogue that Joan Corominas and Jose Luis Pensado kept through their lexicographic works. This dialogue would substantially improve the *Diccionario Crítico Etimológico de la Lengua Castellana* (1954). Thus, not only did Pensado include comments on this latter dictionary, but numerous corrections as well, in the Prologue of *Catálogo de Voces y Frases Gallegas* (1973), which he edited. Corominas assessed them, accepting most of the corrections and he introduced them in his new dictionary, the *Diccionario Crítico Etimológico Castellano e Hispánico*(1980). This huge lexicographic work arises our interest since most of its macro- and microstructural enlargement is based on a massive inclusion of galician entries— thanks to the editorial work of Pensado, actually. In consequence, this presentation seeks, firstly, to reflect the importance and consequences of this fruitful dialogue and, secondly, to vindicate the figure of Jose Luis Pensado in the *Diccionario Crítico Etimológico Castellano e Hispánico*, as well as Corominas' appreciation and recognition of his philological authority and erudition. Finally, we expect to highlight the undeniable productivity of scientific dialogue in the field of lexicography, since, as in any specialized area, it plays an essential role in the advance of modern research.

## **Velázquez de la Cadena y la lexicografía bilingüe inglés / español**

Garriga, Cecilio; Gállego, Raquel

### 6. Historical and Scholarly Lexicography and Etymology

Mariano Fernández de la Cadena (Mexico City 1778 – New York 1860), professor at Columbia University, was the author of a bilingual English-Spanish dictionary of great prestige, which is subsequently reissued and revised even today. However, the dictionary is a relatively unknown work, primarily because bilingual dictionaries have not been the focal point of much attention by researchers, and secondly, because the dictionary's sphere of influence has been centered in the United States. If we survey the primary literature, we can see that the work is referred to inaccurately. At first glance it is clear that this is an innovative dictionary that is clearly rooted in the Spanish lexicographical tradition that emerged in the mid-19th century at a significant point in the revival of Spanish lexicography. Likewise, as a bilingual dictionary that aims to meet the needs of American students, it manages to escape the asphyxiating dominion exercised by the Royal Academy in the field of Spanish lexicography. In this study, the characteristics of *A Pronouncing Dictionary of the Spanish and English Languages* are examined in detail in terms of both their macrostructure and microstructure, and special attention is paid to how the sciences technical fields are treated lexically, as they constitute one of the realms most sensitive to revival during this period. All of this is duly contextualised within contemporary lexicographical trends.

## **Description of Loan Words in French School Dictionaries: Treatment of Words of Foreign Origin in *Dictionnaire Hachette junior* (2006) and *Le Robert junior illustré* (2005)**

Gasiglia, Nathalie

### 6. Historical and Scholarly Lexicography and Etymology

French children learn to use dictionaries at the very beginning of their schooling. Between the ages of eight and twelve, they have access to general-purpose dictionaries which may deal with certain loan words. Our study analyses borrowings which are dealt with in a selection of this type of dictionaries: two French general dictionaries for cycle 3 which have substantial etymological content—*Dictionnaire Hachette junior* (2006) and *Robert junior illustré* (2005). The four leading general children's dictionaries for eight- to twelve-year-olds note between 116 and 619 borrowings from a selection of 4 to 52 languages. Like the dictionaries for cycle 2—students between the ages of five and eight, they may provide information about the phonographic features of the borrowings indicated as such, but as cycle 3

children are supposed to read alone and be at an age when the thirst for new knowledge is very strong, it is logical that dictionaries designed for them should offer more substantial entries in terms of the nature and relative systematization of the information they provide. According to each dictionary's individual structure this information might be presented in a single entry zone—as in *Larousse junior* (2003)—or three zones—in *Robert junior illustré* and *Dictionnaire Hachette junior*. Like the number of zones used, associated information types also vary: the information given is most often metalinguistic—phonographic, lexical, morphological, etymological, etc.—and sometimes cultural. In this analysis of the treatment of anglicisms in French dictionaries for eight- to twelve-year-olds, I propose to build a typology of etymology associated information and to examine how and where this information is given in *Dictionnaire Hachette junior* and *Robert junior illustré*, the two dictionaries which have a consistent etymological approach: 619 loan words identified in the first one and 495 in the second one.

### **Le polirematiche nel TLIO: pratiche lessicografiche, dati e criteri di classificazione**

Giuliani, Mariafrancesca

#### 6. Historical and Scholarly Lexicography and Etymology

This paper describes the data, the methodological problems and the directions, as well as the classification criteria involved in the lexicographic treatment of multiword expressions in the TLIO (*Tesoro della Lingua Italiana delle Origini*, cf. [www.oivi.cnr.it](http://www.oivi.cnr.it)). It is focused the importance of choosing to record and include multiword expressions into the microstructure of the entries, in order to show the semantic and syntactical interconnection binding free, recurrent and fixed combinations in the net of uses involving each data-base item. Particularly I describe and discuss the three level classification—collocations, idioms, phrases—used to arrange the data-base cooccurrences showing features of frequency or idiosyncratic semantic-syntactical structure. Some attention is paid to the definition of the idiomatic field drawn in the editing of a corpus based historical dictionary, often grounded on the decoding activities connected with the lexicographic description; finally I stress the contribution that linguistics and lexicography could get out of the collection and study of a high number of particular form- meaning pairs selected from historical documentation, especially if compared with similar modern lexical corpora.

## **Digital Diachronic Thesaurus of Latin Food Words and their Heritage in European Languages**

Grigorieva, Alexandra; Hautala, Svetlana; Romanova, Natalia

6. Historical and Scholarly Lexicography and Etymology

Our international lexicographical project is set to assemble all classical Latin culinary words of surviving texts to create the first *Digital Diachronic Thesaurus of Latin Food Words and their Heritage in European Languages* in history, an interactive searchable database structure of culinary contexts for exploring the history of the culinary words of antiquity and their reception from the Middle Ages and Renaissance up to the present. Each classical Latin word describing food is to be supported by its etymology when possible—with relevant quotations in Ancient Greek for Ancient Greek loanwords—and, if that particular word has progeny in other old and modern, especially Romance European languages, it will be provided with links to derivatives and appropriate contexts. We will also strive to cover the majority of Medieval Latin food words in the same way during the second stage of the project.

The project is in its initial stage now but we would like to show our colleagues the preliminary digital structure of the Thesaurus that allows the display of the historical chains of shifting lexical forms—including dialectal when possible—and their meanings. Every food context in the chain would be provided with a translation and a short commentary in English and in Italian describing its historical, anthropological, cultural and culinary peculiarities. This frontier, interdisciplinary project covers the whole history of European languages and literatures. It should be able to bring to light the varied typology of European culinary vocabulary—something nobody has done before—and, at the same time, help to preserve the rich culinary heritage and diversity of European countries. We hope it becomes an invaluable tool for many Classical, Medieval and Renaissance scholars and other researchers engaged in Language, History, Food Studies and so on.

## **The Role of *Foclóir Gaeilge-Béarla* Néill Uí Dhónaill in Irish Language Lexicography in the Twentieth Century**

Mac Amhlaigh, Liam

6. Historical and Scholarly Lexicography and Etymology

This paper sets out to chronicle the compilation and usage of the *Foclóir Gaeilge-Béarla*—or Irish-English Dictionary—by Niall Ó Dónaill, Tomás de Bhaldraithe and the lexicography team in An Gúm working on behalf of the Department of Education in the Republic of Ireland. As the primary modern dictionary of its time, its effect on the teaching and usage of the Irish language in the last quarter of the

twentieth century is profound. This is especially the case in light of the fact that no update or amendment to it has ever been seen fit to be produced. Unlike the forthcoming English-Irish dictionary in motion under the auspices of Foras na Gaeilge—the government body responsible under Irish law for the promotion of the Irish language and Irish language organizations—and Lexicography MasterClass, there is no likelihood of any new Irish-English dictionary being produced in the near future. The evolution of the dictionary began as a development from the publication of the English-Irish Dictionary of Tomás de Bhaldraithe in 1959 when an equivalent resource for language users was desired from the opposite perspective—that of the Irish language user looking for the appropriate and most up-to-date English idiom for the words sought. The paper analyses the strengths and weaknesses of the dictionary together with the reasons that necessitated the production of the dictionary as it was. The paper represents a flavour of the ongoing research in the area of Irish language lexicography of the twentieth century, utilising, among other sources, the papers of Tomás de Bhaldraithe situated in University College Dublin's Cártlann na gCanúintí—Irish language dialect archive, the papers of Muiris Ó Droighneáin, one of Ireland's foremost grammatical consultants and the papers and archive of An Gúm, the Irish language publishing wing of the Department of Education.

### **Macro- and Microstructure Experiments in Minor Bilingual Dictionaries of XIX and XX Century**

Marello, Carla; Tomatis, Marco

#### 6. Historical and Scholarly Lexicography and Etymology

Two bilingual English and French and English and German dictionaries and two multilingual dictionaries dealing with English, French, German and Italian with a peculiar macro- and microstructure will be considered in order to highlight their efforts to spare space and to help foreign learners of such languages. The first dictionary—Williams Smith, *A French Dictionary*, on a plan entirely new (1814)—tried to help English learners to reproduce the pronunciation of French words, the second—A.F. Inglott Bey, *A dictionary of English Homonyms* pronouncing and explanatory translated into Italian and French (1899)—arranged homonyms in three languages and explained them, the third—*Neues Universal- Wörtbuch der deutschen, englischen, französischen und italienischen Sprache* (1856)—insisted on comparison among languages and the fourth—Max Bellows' *Dictionary of German and English English and German* (1912)—tried to have both sections English-German and German-English on the same page and to exploit different types to distinguish parts of speech plus masculine, feminine and neuter gender. The paper will explore suggestions for more innovative format in electronic bilingual

dictionaries of the XXI century, since electronic dictionaries on Cd-Rom developed the search window, but did not venture to reinvent the electronic microstructure profile. During the XIX and XX century printers and lexicographers reflected upon improvements in printing layout above all when they aimed to meet the claims of middle-class buyers, asking for effective, pocket-size, not too expensive lexicographic tools.

### **Le programme TLF-Étym: apports récents de l'étymologie comparée-reconstruction**

Petrequin, Gilles; Monda Andronache, Marta

#### 6. Historical and Scholarly Lexicography and Etymology

Our topic is the French hereditary vocabulary considered in a new theoretical approach and focuses on the theme “Historical and learned Lexicography and Etymology”. From the point of view of the classical etymology any lexeme must find its origin in a graphic form. Therefore, there is nowadays a strong consensus among the specialists of the Romanic studies to revise this “classical” and philological lexicographic practice that puts the graphic form at the centre of the theory, which generates basic contradictions. Recognizing the oral form of a hereditary lexeme and renouncing the “graphic centrist” conception in the treatment of the hereditary vocabulary appears as an obvious necessity in the daily practice of the lexicographer. Moving away the “classical” method of the Romanic etymology, we propose to apply the system of the historical and comparative grammar to the field of the French etymology to reconstruct, by comparing different oral forms from Romanic languages, the oral form of the proto-language. Our submission presents three examples of etymological notes/headwords on the hereditary vocabulary developed and published by the program TLF-Étym of the linguistic laboratory ATILF (Analyse et traitement informatisé de la langue française; CNRS/Nancy-Université, France). These examples will allow us to demonstrate to what extent the practice of the etymology of the French hereditary word pool depends on the progress of the Roman etymology, with which it should go hand in hand from now on.

### **De la 1re à la 2e édition du *Dictionnaire de l'Académie française*: marques diastématiques et diaphasiques**

Pouteaux, Marie-Alix; Dagenais, Louise

#### 6. Historical and Scholarly Lexicography and Etymology

In the history of French dictionaries, the second edition (1718) of the *French Academy's Dictionnaire* (hereafter ACA2) has generally been perceived as a bare alphabetical re-arrangement of the first edition, published in 1694 (ACA1), in which

lexical entries were morphosemantically grouped under their primary root word. However, ACA2's preface and title (*Nouveau dictionnaire*) suggests that it underwent a more important revision than what has been believed. This research brings to light the significant progress which ACA2 represents in comparison with ACA1. In the first part, the various aspects of the dictionary microstructure of the letter *l* headwords are compared with each other. The second part is devoted to the analysis of the sociolinguistic marking on the basis of the diastratic and diaphasic usage marks, i.e. *bas*, *populaire*, *peuple* and *familier*. The results that arise from this study are, firstly, that 57% of the lexical units from the *l* corpus that are common to both editions are reworked in ACA2 and, secondly, the study shows that 47 to 83% of the lexical units tagged *bas*, *populaire*, *peuple* and/or *familier* were not included in ACA1. We then proceed to demonstrate to what extent the French Academy 1718 *Nouveau Dictionnaire* constitutes a new edition and not just an alphabetic reprint of the first edition.

**Primer contacto de las lenguas española e inglesa: el *Sex linguarum Latinae, Teuthonice, Gallice, Hispanice, Italice, Anglice, dilucidissimus dictionarius***

Redondo Rodríguez, M. Jesús

6. Historical and Scholarly Lexicography and Etymology

*Introito e porta de quele che voleno imparare e comprender todescho a latino, cioe italiano* [...] is a short and concise anonymous vocabulary published in Venice by Adamo di Rovila on August 12th, 1477. I am going to talk about its first edition in six languages inheritor of the original, *Sex linguarum Latinae, Theutonice, Gallice, Hispanice, Italice, Anglice, dilucidissimus dictionarius*, published in 1537 by John Renys and printed by James Nicolson in Southwarke. Some years before this edition, Spanish is included in the vocabulary in five languages by Garonum together with Latin, Italian, French and German in 1526.

In 1534 John Steelsius published one edition of the *Introito* in five languages, but German was changed for Flemish. Information was reorganized, thus changing the structure of the book. In addition to that, some words of the nomenclature were changed for other lexical elections. The Steelsius innovation was quickly spreaded in the English book market. John Renys adapted that polyglot text to satisfy British needs. He added the English language and maintained the practical, pragmatic spirit. The similarities between these catalogues can be seen through its structure, form, and content. The language layout is practically the same, and English is found in final position. Two versions use *teutonic* as an adjective to designate the Flemish language. The only exceptions to this "lexical coincidence" are some variants that

seem to be reading mistakes, transcription errors, or misprints. Those entries that were eliminated in 1534 are not present in 1537. Spanish and English words are the same. Both of them choose Flemish and not German, a fact that turns them out to be the two unique examples known, and preserved in five and six languages of Flemish language in Europe. This is the only *Vochabuolista* printing which took place in England, and the first to include a linguistic combination of English, Flemish and Spanish. Besides, it is also the only version in six languages.

### **L'informatisation du FEW: des attentes cohérentes d'une communauté scientifique à une modélisation au plus près du programme étymologique wartburgien**

Renders, Pascale; Nissille, Christelle

6. Historical and Scholarly Lexicography and Etymology

The computerisation project of the *Französisches Etymologisches Wörterbuch* aims at taking this reference dictionary on Romance linguistics out of its current state of under- utilization resulting from the complexity of its structures. In view of this, we have submitted on September 2007 a questionnaire designed to better understand the wishes and common practices of FEW users. In the present paper, we examine the first results of the survey and its impact on the future electronic modelisation of the FEW by focusing on cross-searching fields. Implicit information, such as dates, regionalisms or suffixes cannot be found automatically, except by using external tools, that must therefore be taken into account in the computerisation. The solutions considered here do not mean that "classic reading" will become obsolete. Nevertheless, we hope that we will be able to give users new ways to get to the dictionary thus allowing a new and more efficient use of the FEW...

### **El tratamiento de los números en el diccionario**

Rodríguez Ortiz, Francesc; Garriga Escribano, Cecilio

6. Historical and Scholarly Lexicography and Etymology

The definition of numbers involves a complexity that is often overlooked in dictionaries. Their double nature—both grammatical and semantic—means that on one hand they constitute part of a formal language—ex. *arithmetic*, and on the other, their morphological behaviour means they are no different from any other types of words. This complexity is accentuated by the way they can be considered either nouns or adjectives. On the other hand, words like *one*, *two*, *three*, etc. present the distinction that they can be written in two different ways: *one* / 1, *two* / 2, etc., i.e. either using a linguistic or an arithmetical sign. Additionally, their different forms, dependent on whether they are cardinals, ordinals, fractions, multiplicative,

distributive, collective, etc., involve differentiated lexical forms. Numbers also frequently possess figurative meanings or appear in an abundance of set phrases. Dictionaries have dealt with this problem in a variety of different ways. If we look at the design of Spanish dictionaries since the 18th century, we find a certain amount of vacillation that persists until the matter was more firmly established in the 20th century. Nor were there any major differences between dictionaries in different languages. This study presents the state of the question on the basis of an examination of popular Spanish general dictionaries, and proposes certain principles that could improve the coherence of dictionaries in the way they deal with this class of problem.

### **Aspectos gramaticales en la macro y microestructura de un diccionario bilingüe novohispano**

Romero Rangel, Laura; Mora-Bustos, Armando

#### 6. Historical and Scholarly Lexicography and Etymology

The purpose of this paper is to expose and explain certain grammatical aspects in one of the most important bilingual dictionaries in La Nueva España on sixteenth century: the *Vocabulario castellano y mexicano y mexicano y castellano* (1571), elaborated by prior fray Alonso de Molina. Currently in bilingual lexicography, there are different criteria to codify syntactic or grammatical information in lexicographical theory. Yet in the middle of the 16<sup>th</sup> century, these criteria did not exist. It was the linguistic sensibility of Molina that became the only way to express the grammatical functions of the lexical and phraseological units. Our intention with this presentation is to demonstrate the way that Molina, although not a lexicographer, was able to codify the following information:

1. *marcas gramaticales* —grammatical markers, for example: *mejor nombre comparativo*, *mejor adverbio comparativo* or *he adverbio para demostrar*, in Spanish entry, and *agora tiempo presente*, *axca*, *axcan*. *Aduer*, in nahuatl part;
2. *contornos sintácticos*—syntactic plan/outline, such as: *Abituar a alguno en agluna cosa*, *Cutir una vasija con otra*, or *Chamuscarse algo*; and
3. *ematización de compuestos y frases*, for example: *Abrego viento*, *Higas dar*, *Hambrear auer hambre*, *Hambre hauer o tener hambre de cualquier cosa*, *Harona bestia*, *Haldas poner en cinta* or *Achacoso ser*.

This is only a small sample of how to lexicographically handle special types of syntactic information.

## **Le DÉCT (*Dictionnaire Électronique de Chrétien de Troyes*): un modèle pour la lexicographie d'aujourd'hui?**

Souvay Gilles, Pierre Kunstmann

6. Historical and Scholarly Lexicography and Etymology

The DECT is an example of today's lexicographic practice. Its realization is completely computerized from the input to the on-lining. It calls on modern concepts of data encoding (XML) and diffusion—free access on the Web. The DECT is not just a dictionary searchable from the entries. It is in fact a real lexicographic tool made up of an annotated textual base—lemma and part of speech—with the manuscript's image, and the lexicon resulting from the texts analysis. It can be consulted in a traditional way—display of a page, of a verse, of an article...—or through specialized search forms, for instance, it is possible to look for co-occurring words in the texts—lemma *aimer* before an adverb, or to make a multi-criteria query in the lexicon—search for a word in a verb's definition. Moreover, it is always possible for the user to go from the lexicon to the texts and vice versa. The on-line base can be accessed at <http://www.atilf.fr/dect>. (French and English). The DECT's computerized component is built on a platform developed at the ATILF for historical linguistics projects. The same tools allow the consultation of other lexicographic projects, about ten instances. The DECT contributed, for a large part, to the platform development and constitutes, for it, the most successful instance.

## **Vulgar and Popular in Johnson, Webster and the OED**

Wild, Kate

6. Historical and Scholarly Lexicography and Etymology

The use of restrictive labels is one of the most subjective features of modern lexicography, and several studies have shown that dictionaries do not always agree in their application of, for example, *colloquial* and *informal*. Labels are also a problematic feature of pre- 20th century dictionaries, which did not provide lists or explanations of the labels they used. The purpose of this paper is to analyse the development of two labels—*vulgar* and *popular*—in Johnson's (1755) *A Dictionary of the English Language*, Webster's (1828) *An American Dictionary of the English Language*, and the first edition of the *Oxford English Dictionary* (1884-1933)—in order to consider how their meanings and connotations have changed, and what their use can tell us about the relative prescriptivism of the three dictionaries.

## **Papel de los diccionarios de colocaciones en la enseñanza de español como L2**

Alonso Ramos, Margarita

### 7. Dictionary Use

It is generally acknowledged within the Spanish as second language (SSL) community that collocations need to be taught and that collocation dictionaries are useful. Nevertheless, no one has carried out yet any experimental study to investigate what kind of collocation information must be included into a dictionary and how to encode it for a user to take full advantage of it. We describe the results obtained from a small experiment in the use of collocation dictionaries in the teaching of SSL. More precisely, the goal of this experiment is to verify whether the inclusion of semantic and syntactic information on collocations into the dictionary as well as examples of usage could correlate with a better performance on the part of learners.

This is namely the premise underlying the *Diccionario de colocaciones del español* (DiCE). DiCE is based on the Explanatory and Combinatorial Lexicology (Mel'cuk et al. 1995), where collocations are assigned semantic labels and syntactic tags – lexical functions. In order to weigh up how useful this information is, we had to compare the DiCE with another dictionary which did not include this information: the only dictionary that has been published in Spanish which deals with collocations is the *Diccionario combinatorio práctico* (DCP, Bosque 2006).

The experiment was conducted on 25 learners of Spanish and 5 native speakers. Its goal was to evaluate whether the users of the dictionaries had better results with the dictionary that included semantic and syntactic information of each collocation. Since we needed to know their previous knowledge, we decided to organize the test according to three different criteria:

1. without any collocation dictionary;
2. with the DCP, and
3. with the DiCE.

On the one hand, the results of the experiment are positive but, on the other, worrying. Positive because they confirm our premise: in general, students perform better when the dictionary includes semantic and syntactic information on collocations, and worrying because they show that in some cases, the performance of the students decreases when they use the dictionaries mentioned above. Further, more extensive studies are needed to investigate this phenomenon.

## **Frequency in Learners' Dictionaries**

Bogaards, Paul

### 7. Dictionary Use

The learners' dictionaries that exist for English all contain a restricted number of items. The vocabulary that is described in these dictionaries is selected on the basis of frequency of appearance in English. A far more limited number of items are marked as the most important ones, as these that all students should know at some time, because they constitute the lexical core of the language. The marking of high frequency is done in different ways in the five learners' dictionaries. The data provided are not always very useful and are sometimes inconsistent from dictionary to dictionary. An analysis is made of some samples taken from the five learners' dictionaries of English and the relevance of different types of frequency information is discussed.

## **United in Diversity: Dutch Historical Dictionaries Online**

Depuydt, Katrien; De Does, Jesse

### 7. Dictionary Use

The *Integrated Language Database of Dutch* (ILD) is a project of the Institute for Dutch Lexicology in Leiden, which integrates corpora, computational lexica and dictionaries describing the Dutch language from ca. 500 until the present. In 2007, the dictionary component was released, already containing two major historical dictionaries of Dutch, the *Woordenboek der Nederlandsche Taal* (WNT, Dictionary of the Dutch Language, 1500- 1976) and the *Vroegmiddelnederlands Woordenboek* (VMNW, Dictionary of Early Middle Dutch, 1200-1300). When, by 2009, the *Middelnederlandsch Woordenboek* (MNW, Dictionary of Middle Dutch, ~1250 – 1550) and the *Oudnederlands Woordenboek* ('ONW', Dictionary of Old Dutch, a current project at INL, to be finished in 2008, ca. 500–1200) will have been added, researchers of Dutch will have access to dictionaries covering the complete history of the Dutch language. The choice of a single application, integrating the dictionaries so that a user might query one or more dictionaries simultaneously, was a logical step because of the complementary nature of the dictionaries. The challenge was not only providing the user with optimal access to the dictionary information, but also doing so without compromising the uniqueness of each individual dictionary. We sketch the principles underlying the application.

## **Noun and Verb Codes in Pedagogical Dictionaries of English: User-friendliness Revisited**

Dziemianko, Anna

### 7. Dictionary Use

The aim of the present paper is to assess the user-friendliness of noun and verb coding systems in pedagogical dictionaries of English, measured by the frequency with which relevant information properly used in a productive task is located in codes. The influence of the following independent variables on the user-friendliness of codes is studied: the degree of syntactic congruity between Polish lexical items and English headwords, the form of codes, the grammatical category of headwords and the level of dictionary users' proficiency in English. To investigate the influence of the form of codes on their user-friendliness, codes in noun and verb entries were divided into mainstream—referring to formal categories, transparent and prevalent in pedagogical dictionaries, and alternative—which, used very sparingly in today's dictionaries, include reference to sentence functions—verbs—or many quite opaque symbols—nouns. Conclusions are drawn on the basis of an experiment in which almost 900 Polish subjects, advanced and intermediate in English, were involved in a translation task in which they had to use English noun and verb entries compiled for the purpose of the study. The results show that differences in grammar between Polish and English did not affect the consultation of either noun or verb codes. Strangely enough, alternative, and seemingly more demanding codes were strongly favored by the intermediate subjects, and—in the case of verbs—also the advanced ones. The part of speech played a very significant role at the higher level of proficiency, but was not important for the reference to codes by the less advanced. Finally, the higher level of proficiency in English made the subjects appreciate codes more fully, which may be seen as an argument for maintaining the over 70-year tradition of encoding syntactic information in pedagogical dictionaries of English.

## **Teaching the Systematic Dictionary Use as a Strategy for Accuracy and Confidence Building**

Kambaki-Vougioukli, Penelope

### 7. Dictionary Use

This is a longitudinal study, which started in 2004 and ended in 2005. There participated sixteen high-school pupils—same number of boys and girls, aged 13-15, of similar socioeconomic background, whose MT is Turkish but living in Thrace, Greece and attending Greek State Schools rather than minority Public Schools. The fact is that we expected to have more subjects but, unfortunately, we had to exclude a

lot of pupils due to a number of reasons such as differences in the socioeconomic level of the families, gender availability—having more male than female pupils, negative attitude towards the research, etc. What we are investigating is whether and to what extent the systematic use of both monolingual English dictionaries and bilingual Greek-English and English-Greek dictionaries could possibly result in a better reading comprehension and, in the long run, in an improvement and enrichment of their English vocabulary and, to a lesser extent, in Greek. Our aim is to reinforce their general linguistic competence and performance but also their strategic competence by encouraging them to use dictionaries when working at home, too. Furthermore, we are measuring their confidence levels before and after using dictionaries, at certain intervals over the whole period of the experiment. All the participants were given individualised instruction on dictionary use in pair and group work at certain intervals over the whole period of the experiment, too. It is important to notice that we are not really evaluating "certain" dictionaries, it is rather unrealistic as their resources are rather poor; nevertheless, we are trying to exploit what we really have at our disposal, that specific time. The results justified our expectations as most the students that collaborated seem to be very comfortable with dictionary use and confident with the information they expect to find there.

## **Improving Dictionaries**

Kernerman, Ari

### 7. Dictionary Use

Although printed dictionaries have reached a high level of sophistication, there is still much to be improved in order to enhance their usefulness. Prefaces, especially in learners' dictionaries, are not written for users or actual learners, but rather for their teachers, for other lexicographers or for reviewers. For example, Prefaces in learners' dictionaries explain such things as the use of word corpora, the character of the dictionary, the philosophy behind the dictionary, how the dictionary was written, what is different in each particular edition, etc. Interesting, but not helpful information for users. Though intended to be used universally, these dictionaries are culturally biased. Their British culture is irrelevant to the billion learners of English who live in non-English-speaking countries, and need locally or neutrally-oriented dictionaries to help them to communicate with people in other non-English-speaking countries. And the one-size-fits-all principle of monolingual learners' dictionaries does not replace the need to provide mother tongue translation. Many publishers keep adding information to the new editions, much of which is not helpful, reduces the dictionary's efficiency, and does not increase the user's knowledge. On top of that, the absence of a system of lexicography standards makes it difficult for users to refer to more than one dictionary. Giving preference to

corpora-determined frequency over the didactic value of presenting basic meanings first is a step backward, not forward. Besides, too much space is unnecessarily devoted to familiar words, at the expense of less familiar words. These, and other deficiencies of our modern dictionaries—including bilingual, native speakers' and specialized dictionaries—are discussed, with suggestions for rectifying them.

## **Teaching Dictionary-using Skills for Online Dictionaries—An Attempt at a Theoretical Framework for South Africa**

Klein, Juliane

### 7. Dictionary Use

The aim of this paper is to illustrate a theoretical approach to teach dictionary-using skills in South Africa. As the focus is on online dictionaries, only dictionary-using skills will be discussed. Teaching dictionary-using skills in a linguistically heterogeneous society, which has not yet developed a fully functional dictionary culture for all languages, is a difficult task. Not only must the different languages—e.g. conjunctively written languages and disjunctively written languages—be taken into account, but also the different user groups ranging from pupils/university students to ordinary people who want to use a dictionary have to be considered. Although the dictionary users are not a homogeneous group, the aim of teaching dictionary-using skills is the same for all groups: achieving a confident and successful use of dictionaries in the short term and creating a fully developed dictionary culture that includes all the languages which are official in South Africa in the long term. The teaching of dictionary- using skills could be divided into four stages:

1. teaching about dictionaries,
2. teaching basic skills to access dictionaries
3. teaching look-up strategies,
4. teaching strategies to decode the information found in the definition given by the dictionary.

Dictionary-using skills should be taught as early as possible in schools, and this teaching should be continued throughout the whole education process, i.e. it should not be taught as a single module, but rather as language methodology. In tertiary education institutions, dictionary-using skills could be integrated into academic literacy modules or taught in separate short language modules. Teaching dictionary-using skills to everybody else will be more difficult, as those who have finished their formal education cannot be reached as easily as pupils or university students. This group will mainly be taught through the dictionaries themselves. Teaching

dictionary-using skills to people through dictionaries implies that the dictionaries must be self-explanatory, which implies that the user interface and all instructions should be available in all the languages that the dictionary covers and not only in English. In addition to that, the dictionary should ideally be accompanied by a user manual in all languages the dictionary covers.

### **Can Dictionary Skills Be Taught? The Effectiveness of Lexicographic Training for Primary-school-level Polish Learners of English**

Lew, Robert; Galas, Katarzyna

#### 7. Dictionary Use

In the present paper we examine the question of whether dictionary reference skills can be taught effectively in the classroom. To this end, we test the reference skills of a group of Polish primary-school students attending English classes twice: prior to and following a 12-session specially-designed training program. Despite the subjects' high confidence in their reference skills reported in the accompanying questionnaire, they performed rather poorly on the pre-test. Following a training program, the performance improves substantially and significantly more than in a matched control group. We conclude that a dictionary skills training program may be effective in teaching language learners at this level to use dictionaries more effectively, though different skills benefit to different degrees.

### **Bringing Bilingual Dictionaries in from the Cold: Challenging Negative Perceptions and Practices in English Language Teaching**

Mandalios, Jane

#### 7. Dictionary Use

This paper deals with English language teaching (ELT) and learning. It considers the presenter's research into the use of bilingual dictionaries in those cases where English is the second/foreign language. The research was carried out amongst non-native speaker students and teachers, and also amongst teachers who were native speakers of English in an English-medium university in the United Arab Emirates. It showed that, in contrast to the teaching of other foreign languages, bilingual dictionaries are generally negatively viewed by ELT theorists and teachers. Yet, after a careful scrutiny of both the lexicography and literature used in ELT you realize that such a view is based on unsubstantiated opinions or questionable research. The study also shows that bilingual dictionaries are almost unanimously considered helpful by learners, yet their preferences are usually ignored or discouraged by teachers, many of whom do not speak the first language of their students, and who feel pressurized to follow the English-only approach that has dominated ELT for the

last 40 years (Phillipson 1992; Auberbach 1993). The students in the study exhibited poor dictionary skills, and little understanding of how efficient the use of bilingual dictionaries could be to enrich their receptive and productive vocabulary skills. A small action research component of the study indicated that these skills can be greatly improved by structured bilingual dictionary instruction. The presenter proposes that the findings of the study constitute evidence of a serious imbalance of power within ELT which can be defined as *pedagogic imperialism*, and calls for a critical reappraisal of both the role of bilingual dictionaries and the use of the native language when teaching English. Closer ties need to be established between the fields of lexicography and ELT, particularly in contexts where the theory and practice of teaching is dominated by native speakers who do not speak the first language of the learners.

### **Looking Up “Hard Words” for a Production Test: A Comparative Study of the NOAD, MEDAL, AHD, and MW Collegiate Dictionaries**

McCreary, Don R.

#### 7. Dictionary Use

We test this hypothesis: The *New Oxford American Dictionary* (NOAD), MW, AHD, and MEDAL equally meet the needs of American college students when they look up a *hard word*. On a production task, writing the word in an appropriate sentence, NOAD users scored much higher than the other three groups on every *hard word*, with only one exception per user. The *Macmillan English Dictionary for Advanced Learners* (MEDAL) users scored higher than the users of the *Merriam Webster’s Collegiate Dictionary*, 11th Edition (MW) or users of the *American Heritage Dictionary*, 2nd Edition (AHD), another collegiate desk dictionary. NOAD has several advantages over the other collegiate dictionaries, including microstructure and vocabulary coverage. Unfortunately, overall coverage of *hard words* is problematic in MEDAL, since it is intended for non-natives. MW users were hampered by their tendency to choose the first sense in the entry, which is the oldest historical sense in MW. This also applies to AHD. This suggests that American college students might consider buying NOAD for its usability and its vocabulary coverage.

## **Giving Them What They Want: Search Strategies for Electronic Dictionaries**

Mechura, Michal Boleslav

### 7. Dictionary Use

This paper deals with how humans search electronic dictionaries. It raises the point that users often make dictionary searches with misspellings, with inflected words copied and pasted from elsewhere, with complete sentences or fragments thereof, and with other kinds of low- quality input, and suggests methods for dealing with such phenomena in a preemptive manner. The issues addressed include searching with inflections, dealing with multi-word items, misspelling detection and text normalization. Additionally, the value of log files is emphasized as a source of information on user behaviour.

## **Adverb Use in EFL Student Writing: From Learner Dictionary to Text Production**

Philip, Gill

### 7. Dictionary Use

Adverbs, especially those occurring in *adverb+adjective* collocations, play a central role in the language that advanced learners are expected to produce in their argumentative writing. Submodifying adverbs of degree such as *closely*, *deeply*, *strongly* and *widely*, however, have been identified as being problematic for learners of English: Italian learners over-use *very* and *really* to the virtual exclusion of any other adverb (Philip 2007). This situation is due in part to the EFL curriculum, but monolingual and bilingual learner's dictionaries appear to do little to address the issue. This presentation examines the way in which lexical adverbs of degree are treated in the five major English dictionaries for advanced learners (CALD, COBUILD, LDOCE, MED and OALD). It also evaluates the way these same forms are treated in four bilingual dictionaries specifically aimed at Italian learners of English (Longman, Oxford Study, Rizzoli-Larousse, and Oxford-Paravia). The analysis reveals that these dictionaries do little or nothing to help students expand their working knowledge of adverbs of degree. In general, the presentation of lexical adverbs is regarded to be subservient to the adjectives from which they are derived. The information boxes which most modern learner's dictionaries include seem to focus on elementary matters of grammar and word choice rather than on the collocation of these polysemous, metaphorically- motivated language items. The presentation concludes by suggesting some ways in which monolingual and bilingual

learners' dictionaries might modify their treatment of lexical adverbs in order to enable students to identify and use alternatives to *very*, *really* and *a lot*.

### **The Electronic Dictionary in the Language Classroom: The Views of Language Learners and Teachers**

Ronald, James; Ozawa, Shinya

#### 7. Dictionary Use

The pocket electronic dictionary (PED) has the potential to be a powerful language learning tool. At the same time, it may be seen as an obstacle to communication, a waste of classroom time, and a source of conflict between foreign-language learners and the teachers. This presentation will report an in-depth survey of three sets of people influenced by the widespread presence and use of the PED in the classroom: foreign-language students, teachers who share the native language of the students, and teachers who are native speakers of the target language. The survey, which takes into account the beliefs, attitudes, and expectations of Japanese learners of English and of their teachers regarding the PED, revealed important differences in their opinions about how and when the dictionary should be used, in the effect of dictionary use on foreign language vocabulary development, and regarding users' needs for training or guidance in the use of electronic dictionaries. The presentation will also recommend means by which understanding of these differing perspectives may help both language learners and teachers make the most of the potential of the electronic dictionary.

### **Mother-tongue's Little Helper (The Use of the Monolingual Dictionary of Slovenian in School)**

Rozman, Tadeja

#### 7. Dictionary Use

The first part of the paper brings a brief overview of how Slovenian curricula and school books for primary and secondary schools incorporate the use of monolingual dictionaries in teaching of Slovenian as the mother tongue. In Slovenia there is no dictionary designed for the school population, with the exception of a few smaller lexicographic teaching materials for the youngest group of primary school children (6-9 years). The *Dictionary of Standard Slovenian* (*Slovar slovenskega knjižnega jezika*, hereafter DSS), currently the only existing general monolingual dictionary of Slovenian, is used in schools. Published in five volumes between 1970 and 1991, DSS is outdated in many aspects and from the viewpoint of use in schools it is even more problematical since it is not suited to the needs and language ability of this particular target audience. Especially problematical are the comprehensibility of the defining

language and the comprehensibility of the complicated labelling system, employed to convey the paradigmatic properties of entries and limitations regarding their use.

The second part of the paper presents the empirical research which had two aims: on the one hand, to determine how the school population understands various definition types from DSS and how comprehensible the set of its labels is, and, on the other hand, to test several versions of various microstructural elements which should be suited to this particular target audience. The research covers three groups of school population at three different stages of cognitive and language development (10-12, 13-15 and 16-18 years), and the results bring useful information regarding a possible concept of a school dictionary of Slovenian.

## **Las unidades fraseológicas eventivas en los diccionarios bilingües Español-Vasco**

Aierbe Mendizabal, Axun

### 8. Phraseology and Collocation

For decades, the written Basque language has become more and more common in many fields, such as literature, education, mass media, administration, science, etc. However, most of the texts written in Basque are translated from Spanish. Because of this, there are several bilingual Spanish-Basque dictionaries used by translators, writers, researchers and editors.

Bilingual Spanish-Basque dictionaries provide a lot of information on equivalents for phraseological units and collocations. Many of the phraseological units that appear in dictionaries are eventive phraseological units, and both of them can be common and specialized in nature. Bevilacqua (2001: 124) proposes that eventive specialised phraseological units are based on an event head—event noun, event verb or eventive past participial adjective—and at least one terminological unit. According to this author, the main phraseological unit is originally based on an event verb. Taking this into account, we have researched bilingual Spanish-Basque dictionaries in order to explore what the basic structure of those eventive phraseological units is like in Spanish, as well as one of their equivalents in Basque. For example, the dictionaries Zehazki and Elhuyar compile two structures for *transformación de energía* > *energia bihurtzea* and *transformar la energía* > *energia eraldatu* or *energia transformatu*; it is not common, however, to compile two or more Spanish equivalent phraseological structures. Usually, Spanish-Basque dictionaries compile one structure for each phraseological unit, which is mostly based on an event noun (*contaminación de las costas* > *kostaldearen kutsadura*; *contaminación de las aguas* > *urak kutsatzea*). While Spanish phraseological units are based on an event noun, Basque equivalent structures may be based on an

equivalent event noun or on an equivalent event verb. This paper will focus on equivalents, phraseological units and the variability of their basic structures depending on the language.

***REDES. Diccionario combinatorio del español contemporáneo***

Almarza Acedo, Nieves; Lozano Ramírez de Arellano, Yolanda

8. Phraseology and Collocation

*REDES. Diccionario combinatorio del español contemporáneo* is an attempt to reflect on the lexical restrictions and to analyse the structure of the language. Through different examples we want to go deeply especially into the utility of this dictionary for any speaker of Spanish and mainly for the students of the language. This dictionary demonstrates that it is necessary to know how the words are combined to express ourselves with accuracy.

**Propuesta de anotación semántica para una base de datos paremiológica**

Alonso Pérez-Ávila, Elena

8. Phraseology and Collocation

An electronic tool, such as an on line multilingual paremiological database, that would enable researchers or translators to search paremiological units of many languages and manipulate the information stored about them more easily, would greatly benefit the field of paremiology.

This paper deals with how the information provided by SpanishWordNet and MultiWordnet may be used to tag semantically paremiological units in Spanish and Italian within the database. In other words, tags related to the WordNet ontology are attached to each term in the proverb in order to provide more information about the domain that the proverb belongs to. We propose this annotation as a methodology to classify paremiological units that can be shared by different linguistic communities since it is based on an already widely used lexical resource developed in many languages: WordNet.

Unfortunately, it is not possible to tag the proverb as a whole unit due to its particular features: the meaning of the whole proverb cannot be easily derived from the meanings of its separate components. At the moment we are trying to supply the database with as much information as possible on the semantics of the components of the proverbs and the relations, such as hyponymy, that those parts present with the rest of the lexicon.

## **From Dictionary to Phrasebook?**

Granger, Sylviane; Paquot, Magali

### 8. Phraseology and Collocation

Language is characterized by a large number of conventionalized phrases which, unlike idioms, are largely regular, both semantically and syntactically. Biber et al. (1999) call these phrases *lexical bundles* and highlight the key role they play both in spoken and written discourse. In spite of their high frequency, these types of phrase have not yet received the place they deserve in dictionaries. In this article, we describe how they are integrated into monolingual learners' dictionaries of English and English-French bilingual dictionaries. The description shows that the presentation of these phrases is largely based on intuition and fails to reflect authentic usage as attested by corpus investigation. We make a plea for a more rigorous—corpus-based—integration of these phrases and illustrate our approach with a fully corpus-based section devoted to English for Academic Purposes (EAP), functions that has been integrated as a middle section in the new edition of the *Macmillan English Dictionary for Advanced Learners*.

## **Collocational False Friends: Description and Treatment in Bilingual Dictionaries**

Heid, Ulrich; Prinsloo, Danie J.

### 8. Phraseology and Collocation

Our starting point is that of translation equivalents: true friends in their use as individual lexical items often become false friends in collocations. It is the duty of the lexicographer to guide the user, especially in learners' dictionaries aimed at productive—encoding—use, in forming correct collocations and in warning the user of false friend cases. Our arguments are based on evidence from large newspaper corpora as well as on internet research. We will present several lexicographic presentation devices from printed dictionaries that allow lexicographers to warn users about false friend collocations. The study will be limited to false friend relations in general bilingual dictionaries, mainly for German, Dutch and Afrikaans. The compilation of dictionaries for false friends lies beyond the scope of this paper. We adopt a lexicographic notion of collocation, here, as used for example by the *Oxford Collocation Dictionary for Students of English* (2002). We use Hausmann's (2004) terms—base and collocate—to denote the elements of collocations. Klégr (2006) transfers the notion of false friends from single words to collocations and classifies the relevant cases according to categories known from translation theory. We propose the following simple arrangement of false friend collocations, inspired by the concept's basic principles:

1. word combination— lexical (co-)selection: if true friend single word equivalents exist in a language pair, we consider collocations as false friends where the cooccurrence of the two single word true friends is impossible in a given language;
2. morphosyntactic preferences: if true friend single word equivalents exist in a language pair, we consider collocations false friends where the languages differ with respect to morphosyntactic preferences, individual readings being equivalent;
3. differences with respect to usage domains.

### **Analysis of Collocations in Russian: Corpus vs Dictionary**

Khokhlova, Maria

#### 8. Phraseology and Collocation

The paper discusses the results of an experiment in collocation extraction in a corpus of Russian texts. The data obtained is compared to the data given for set expressions in modern Russian dictionaries in order to analyze from the standpoint of traditional lexicography what kind of phrases can be received by such an approach. The paper also explores the role of statistical measures for extracting collocations in Russian..

### **The Lemmatisation of Lexically Variable Idioms: The Case of Italian-English Dictionaries**

Mulhall, Chris

#### 8. Phraseology and Collocation

The choice of a suitable point of entry for an idiomatic expression is one of the most complex tasks a lexicographer faces throughout the compilation of a dictionary. This is further exacerbated by the possibility of lexical variation in certain expressions. This paper analyses twenty idioms with variable verbs (ten English / ten Italian) and twenty idioms with variable nouns (ten English / ten Italian) across six bilingual Italian-English dictionaries, *Il Ragazzini* (ZIR) (2006), *Hoepli Grande Dizionario di Inglese* (HGDI) (2003), *Collins Sansoni Italian Dictionary* (CSID) (2003), *Oxford-Paravia Italian Dictionary* (OXID) (2001), *Il Sansoni Inglese* (ISI) (2006) and *Hazon Garzanti Inglese* (HGI) (2006). The analysis highlights a number of problems in the treatment of lexically variable idioms. Firstly, bilingual Italian-English dictionaries do not have a definitive approach to dealing with the problem of lexical variation. Secondly, the consistency and comprehensiveness in the coverage of lexical alternatives varies significantly both within and across the Italian-English and English-Italian sections of dictionaries. The totality of such differences suggests that

a more systematic approach is required in order to achieve a greater consistency in the recording of the variable constituents of idioms.

### **Proyecto para la redacción de un diccionario de locuciones del español**

Penadés Martínez, Inmaculada

#### **8. Phraseology and Collocation**

The idea of creating a dictionary of Spanish idioms originates in the verification that currently there is no dictionary that solely includes these kinds of phraseology units, in contrast to other publications that compile other types of complex units, such as popular sayings. There are other reasons for the convenience of the project, more concretely, the deficient lexicographic treatment given to the assigning of grammar marks up to this day. This deficiency becomes apparent also in the assigning of syntagmatic combinatory, as well as the diastatic and diaphasic markings for idioms in Spanish phraseology dictionaries. The aforementioned dictionary will be onomasiological, semasiological, and will include a synonym and antonym thesaurus.

### **A Comparative Analysis of Definitions of Phrasal Verbs in Monolingual General-purpose Dictionaries for Native Speakers of American and British English**

Perdek, Magdalena

#### **8. Phraseology and Collocation**

This paper is an attempt to analyze the definitions of phrasal verbs in monolingual general- purpose dictionaries for native speakers of English. Four dictionaries from Great Britain and four from the USA published in the last decade provide material for the study which includes a total of 100 phrasal verbs. Bearing in mind the specific semantic load of phrasal verbs, their limitation as to the choice of objects as well as the fact that they are commonly used, this study aims at finding whether there exist significant differences in describing phrasal verbs on both sides of the Atlantic. Three aspects are analyzed in particular: word choice with emphasis on the occurrence of difficult, very formal and rarely used words; precision in rendering the meaning, and inclusion of objects typical of a given sense of a phrasal verb. The analysis reveals that there are certain areas of correlation but also points of differences, not only between the two lexicographic traditions but within each of them separately.

## **Inclusión de los papeles semánticos de FrameNet en DiCE**

Prieto González, Sabela

### 8. Phraseology and Collocation

The aim of our project is to enrich the actantial information of the *Diccionario de Colocaciones del Español* (DiCE) with labels about semantic roles. Since there are other projects which follow this line of research—such as FrameNet, we decided to include the existing information in the DiCE. Although our database focuses on collocations, it also identifies their actants and it compiles a wide-ranging corpus of predicates with their arguments. Therefore, the entry for each lexical unit contains the propositional form or argumental structure where a semantic description of the actants is given. For instance, in the entry for the noun *ira* we find: *ira de individuo X contra individuo Y a causa del hecho Z*. In this way, actants are also described semantically: *ira* is felt by a person against another person because of something. We are trying to add more semantical information to DiCE, by linking the actants of each lemma with the core elements of the relating frame, the same as FrameNet does. Taking up again the same example, the noun *anger* is set into the frame «Emotion\_directed» and it presents four core elements: *experiencer*, *expressor*, *stimulus*, *topic*. These nuclear elements can be related to the actants that appear in the DiCE, giving the dictionary some detailed semantic information, regarding not only the lemma but also the elements relating to this lemma. This process of connection will allow us to label the corpus compiled in the DiCE to make the most of the data.

## **Colocaciones léxicas en diccionarios generales monolingües del español**

Romero Aguilera, Laura

### 8. Phraseology and Collocation

The purpose of this paper is to describe the way some of the Spanish general monolingual dictionaries published during the last twelve years have dealt with lexical collocations, that is, those combinations of words that present certain combinatorial restrictions in the norm, basically semantic restrictions, imposed by usage (Corpas 1996). These have been the analyzed dictionaries: *Diccionario Salamanca de la lengua española*, directed by Juan Gutiérrez (1996); *Diccionario del español actual*, by Manuel Seco, Olimpia Andrés y Gabino Ramos (1999); RAE's *Diccionario de la Lengua Española* (2001); and *Gran diccionario de uso del español actual. Basado en el Corpus Cumbre*, directed by Aquilino Sánchez (2001). We have based our research on a corpus of 52 lexical collocations, which has been built on the analysis of the subentries starting with *b* in the chosen dictionaries. After that, we

have looked up the entries corresponding to each element that constitutes the collocation, in order to know if these dictionaries account for those same combinations in other parts of the lexicographical article. The analysis of the lexicographic information has focused on our aspects: a) the preliminary pages of each dictionary; b) the position of collocations in the lexicographic article; c) the inclusion of these units in a given article; and d) the grammatical category.

### ***Una bella esperienza, una buona prova. A corpus analysis of purely evaluative adjectives in Italian***

Russo, Irene

#### 8. Phraseology and Collocation

It is questionable how much pragmatic information should be included in a dictionary entry. In a native-speaker's dictionary such information is considered unnecessary, but nevertheless, a certain amount of it could be included as multiword expressions—fixed and semi-fixed—that are regarded as holistic units rather than compositional strings. In this work a corpus analysis of two purely evaluative adjectives in Italian—*bello*, *buono*—will shed light on substitutability among them in noun phrases. Mutual Information (Church & Hanks 1990) as a measure to compare and contrast the distribution of words in context highlights nouns for which *bello* and *buono* are interchangeable in NPs. We propose to manage adjectival polysemy clustering word senses according to similar evaluative functions. A dictionary entry for *bello* can be partially structured on the base of its strong similarity with *buono* in NPs contexts: *bello* and *buono* usages are informed by evaluative attitudes displayed by speakers.

### **From Subdomains and Parameters to Collocational Patterns—On the Analysis of Swedish Medical Collocations**

Sköldberg, Emma; Toporowska Gronostaj, Maria

#### 8. Phraseology and Collocation

This paper presents a study on Swedish collocations in an electronic medical lexicon, currently under construction at the University of Gothenburg, Department of Swedish Language. There are two strands discussed in the paper. The first one is about a knowledge-based, onomasiological, approach to detecting and analysing medical collocations and their patterning. The second one deals with the representation of these collocations in both a general lexicon module and a collocational lexicon module. In the latter module, there are some advanced search options made available which enable selective access to the content of the lexicon. It is assumed that the onomasiological approach to the analysis of medical collocations

complements the semasiological one and that the fusion of the two paves the way for a more consistent and exhaustive description of medical collocations and their patterns.

### **Aspectos de fraseografía bilingüe español-alemán: la equivalencia frente a la definición**

Torrent-Lenzen, Aina

8. Phraseology and Collocation

Aspects of bilingual phraseography Spanish-German—phraseological equivalence versus definition. In my paper I would like to discuss some problematic aspects of Spanish-German phraseography which frequently arise for both phraseologists and dictionary users. The document will focus, among other things, on the problems conveyed by the phraseological equivalents, the treatment that the contextual and partial phraseological equivalents should receive, and in some cases, the benefit of introducing definitions. In addition to this, it will also deal with how some verbal phraseological units of the German language should be mentioned if they constitute equivalents to units in Spanish. The practical experience that allows me to perform the analysis of the above-mentioned questions is the *Spanish-German Dictionary on Idioms*, currently being compiled by our team in association with the University of Applied Sciences in Cologne (Fachhochschule Köln).

### **A Multilingual Electronic Database of Distributionally Idiosyncratic Items**

Trawiński, Beata; Soehn, Jan-Philipp; Sailer, Manfred; Richter, Frank

8. Phraseology and Collocation

We present a multilingual electronic database of lexical items with idiosyncratic occurrence patterns. Currently, our database consists of: (1) a collection of 444 bound words in German; (2) a collection of 77 bound words in English; (3) a collection of 58 negative polarity items in Romanian; (4) a collection of 84 negative polarity items in German; and (5) a collection of 52 positive polarity items in German. Our database is encoded in XML and is available via the Internet, offering dynamic and flexible access.

## For an Extended Definition of Lexical Collocations

Tutin, Agnès

### 8. Phraseology and Collocation

Restricted lexical collocations have now been studied and encoded in dictionaries for over twenty years, and stable definitions have been provided for this notion by numerous scholars working on collocations (e.g. Hausmann 1989, Mel'čuk 1998, Heid 1994). They are roughly defined as recurrent combinations of two linguistic elements which have a syntactic relationship. One of the elements of the collocation, called *base*, keeps its usual meaning— autosemantic words (Hausmann 2004)— while the other, the *collocate*, is dependent on the other—synsemantic words—and usually has a less transparent meaning. Even though such a definition is nevertheless operational for a large number of lexical associations, it raises several problems. The first problem has to do with the binary status of the collocation and the unequal status of the two parts of the collocation, which has been questioned by several linguists (inter alia Siepmann 2006, Bartsch 2004) who suggest expanding the definition to associations of three or more elements. A second problem concerns the grammatical status of the collocations. Should functional words—and to what extent—be included in the definition of collocation? For example, in expressions such as *for fear of*, the whole combination can be analysed as a preposition, and not as a phrase contrary to prototypical collocations such as *pay attention*—verb phrase, *major problem*—noun phrase, *seriously injured*—adjective phrase. However, *fear in for fear of* can be considered as relatively transparent, and according to us, it should be considered a collocation. In this paper, we study these two issues in detail and call for an extended typology of restricted collocations. We examine the lexicographical consequences of such an extended definition.

## SciE-Lex: A Lexical Database of Collocations in Scientific English for Spanish Scientists

Verdaguer, Isabel; Poch, Anna; Laso, Natalia Judith; Giménez, Eva

### 8. Phraseology and Collocation

As a result of the widespread use of English in science and scholarship, there is an increasing need of reference tools which provide accurate information to non-native—especially junior—researchers on the correct use of lexico-grammatical patterns of non-technical words when writing their scientific papers in English and on the conventionalized phraseological characteristics of the genre. Our aim is to present SciE-Lex, a lexical database which provides information to help Spanish researchers to write research papers in English accurately. Whereas there are specialized monolingual and bilingual dictionaries with specific terminological

information, there is a shortage of reference tools supplying information on the correct use of syntactic and collocational patterns of non-technical words in the scientific register and on the conventionalized phraseological characteristics of the genre. Based on the analysis of a 3+ million word corpus of scientific English, in its first stage, SciE-Lex displays information on: word class, morphological variants, equivalent(s) in Spanish, patterns of occurrence, list of collocations, examples of real use, and notes to clarify usage. In a second stage we plan to include *lexical bundles*, compositional recurrent sequences of words, since several studies have confirmed the difficulties that learners have with them. Further research will provide SciE-Lex with information about the distribution of lexical bundles across the different sections and/or moves of the academic research article as well as their function in discourse.

### **Database of Bavarian Dialects (DBÖ) Electronically Mapped (*dbo@ema*). A System for Archiving, Maintaining and Field Mapping of Heterogeneous Dialect Data for the Compilation of Dialect Lexicons**

Wandl-Vogt, Eveline; Kop, Christian; Fliedl, Günther; Nickel, Jost; Scholz, Johannes

#### 8. Phraseology and Collocation

*dbo@ema* is a system for the archiving, handling and mapping of heterogeneous dialect data for dialect dictionaries. Within this software presentation:

- a. the users should get known to the general project aims of *dbo@ema*, that are:
  - development of a webbased, interactive data base
  - development of a webbased, interactive tool to map dialect data and background information of a dialect dictionary
  - development of a specific, free font for the phonetic transcription of dialect data in digital surroundings (further information see <http://www.wboe.at>)
- b. the users should get known to special tools of the software developed to compiling a dialect dictionary
- c. the users should get known to how geoinformation aids the compilation of a dialect dictionary
- d. the users should get known to the project *Wörterbuch der bairischen Mundarten in Österreich* (WBÖ) (Dictionary of Bavarian dialects in Austria) and

- e. the project *Datenbank der bairischen Mundarten in Österreich* (DBÖ) (Data base of bavarian dialects in Austria) that are both mother-projects to the project *dbo@ema*.

## **Incomprehensible Languages in Idioms: Functional Equivalents and Bilingual Dictionaries**

Woźniak, Monika

### 8. Phraseology and Collocation

Phraseology is a source of interesting information on the speakers' world view and different fixed expressions are used in different languages when one does not understand the message. Lack of understanding of what is said or written is often associated with the inability to comprehend the language, which is proved by the use of idiomatic expressions containing names of different foreign languages considered to be particularly difficult in a given society. In this paper several bilingual dictionaries are consulted in order to: 1) find equivalents of some expressions of that kind in English, Polish and Spanish; 2) review their lexicographical treatment; and 3) see how the recorded parallels correspond with the functional view of idiom equivalents proposed by Dobrovol'skij (2000a, b).

## **Prepositions in Dictionaries for Foreign Learners: A Cognitive Linguistic Look**

Adamska-Sałaciak, Arleta

### 9. Lexicological Issues of Lexicographical Relevance

The paper is an attempt to look at the problems faced by lexicographers compiling prepositional entries in dictionaries for foreign learners, and to suggest ways in which these problems could be alleviated. After discussing some of the reasons why prepositions are difficult to deal with in a dictionary, and reporting on the results of metalexigraphic studies examining the treatment of prepositions in monolingual English learners' dictionaries and in three bilingual English-Polish dictionaries, Cognitive Linguistics is suggested as a source of important insights which could be of assistance in solving practical lexicographic problems. Among those insights are: the idea that the linguistic structuring of space functions as a mental template for other domains; recognition of the polysemic sense network of prepositional meanings; preference for *principled polysemy* over earlier *unrestricted polysemy* approaches; introduction of rigid criteria for the recognition of separate senses; recognition of the fact that the overwhelming majority of spatial senses of prepositions are related through metonymy. Drawing on the cognitive linguistic analyses of the semantics of English prepositions offered by Tyler and Evans (2003),

some practical recommendations are made regarding ways in which prepositional entries in dictionaries for foreign learners could be made more informative and useful. These include a considerable reduction of the number of senses and examples of usage, an introduction of semantic ‘profiles’ at the beginning of entries, and supplementing verbal illustrations with simple graphics, highlighting the salient meanings of particular prepositions, the links between different senses, and the differences between semantically close and therefore frequently confused items.

### **Lexicographie historique, noms de métier, féminisation: quelle méthodologie?**

Baider, Fabienne

#### 9. Lexicological Issues of Lexicographical Relevance

This article investigates the way trade names in the feminine form are presented in the French etymological and historical lexicographic discourse. Several French decrees in 1986, 1994 and 2000 were issued to promote use of the feminine form of trade names in reference to women. Working from two different corpuses—one before the feminization policy and one after, the analysis establishes whether progress had been made in such usage: feminine forms have increased in the past 30 years, even though their presentation remains incomplete and sometimes even marginal. However, the study of the presence or absence of these feminine forms could provide insight into what the linguistic function of gender is for various lexicographers. For some, a different gender and a different form of a trade name—ex. *boulangier* and *boulangère*—do not justify the inclusion of the feminine form, since they are derived morphologically and semantically from the masculine word, even though this case is not necessarily true. On the semantic level, this reasoning presupposes that grammatical gender does not fulfill any relevant function for nouns denoting animates. If it is impossible to conclude that these different lexicographical discursive practices support an asymmetrical representation of the sexes because of their different treatment of grammatical gender, it is nevertheless certain that such reasoning deprives all feminine forms from etymological information, hence truncating the history of words.

### **Scale-free Networks in Dictionaries**

Fóris, Ágota

#### 9. Lexicological Issues of Lexicographical Relevance

The aim of this paper is to show, through the application of the mathematical model of scale-free networks, how the scale-free network of language is represented in the information contained in dictionaries. Research conducted in the last few decades

has proven that every phenomenon of nature and society—the relations of so many various systems—is organised into a complex system of networks. Research has also proven that complex networks can be analysed with the help of a common network model, and that the application of this network theory allows us to discover features of the analysed system that are not observable by other methods. After the discovery of the significance of networks, broad experimental and theoretical studies were launched to reveal the nature of networks and to apply the findings, and research on scale-free networks is the most outstanding among these. If we accept that the three components of the terminological unit may be modelled with the scale-free terminological network model (Fóris 2007), and that the language network is made up of at least these three networks, then we may suppose that dictionaries select and present various parts of this complex network from different approaches. The lexicographers' task—to put it simply—is to collect, record and make the data necessary for language use easily accessible. In order to meet this aim, dictionaries need to follow the three-sided structure of language networks. The various types of dictionaries compiled for different purposes developed a practical structure that reflects the structure of the language network. In the paper, I briefly touch upon the main characteristics of the scale-free network model that can be widely applied in linguistic research, and point out the lexicographic aspects of the model. Based on the network model we can draw conclusions concerning the practical structure of dictionaries. I demonstrate that the complex scale-free network structure of language containing three sub-systems enables us to use the language quickly and completely. I also illustrate and support the features of the language network model and its application with figures.

### **La place du métalangage dans la définition lexicographique: l'exemple des définitions des mots syncatégorématiques dans le TLF**

Frassi, Paolo

#### 9. Lexicological Issues of Lexicographical Relevance

The studies on lexicographic definitions connected with the French tradition take charge eminently of typology and leave aside the question of metalanguage. So, in lexicography, the metalinguistic definition is often considered in the typological frame. This is because the above-mentioned studies are mostly based upon definitions either of nouns or verbs. In my presentation I shall attempt to demonstrate, from defining statements of the syncategorematic words drawn from the *Trésor de la langue française*, that the metalinguistic definition is indeed a category of the definitions but that, when compared to the other categories, it requires a different criteria of analysis, due to its nature. In order to do this, I shall present, first, the different nature of this issue from a typological approach on one

side and a metalinguistic approach on the other. I shall expose, then, the main typological studies—in particular the unpublished document which is stored in the archives of the Laboratory ATILF ["Pour un nouveau cahier de normes...", 1979] as well as Martin (1983) and Rey-Debove (1998)—in which the question of the metalanguage is dealt with inside and following the example of typology to demonstrate that, if a definition such as *aiguillette—nom populaire de l'orphie*—is metalinguistic and a definition such as *chaise—siège à dossier sans bras*—is periphrastic, *nom* et *siège* are both hyperonyms, so that the typological criteria are not enough to distinguish between metalinguistic and periphrastic definition. Thus, I will establish, in accordance with Rey-Debove (1997), in which the definition is considered from a metalinguistic point of view—according to the syntactic relation between a lexical entry and its lexicographical definition, the principles which govern the metalinguistic analysis. The results will lead to three different categories of metalinguistic definitions of the syncategorematic words:

1. the definition refers to both infralinguistic and extralinguistic reality—in this case two sub-categories are possible:
  - a. the hyperonym refers to the infralinguistic reality while the specific semes refer to the extralinguistic reality;
  - b. the hyperonym refers to the infralinguistic reality while the specific semes, among which there is at least an autonym with "schize" (cf. Rey-Debove 1997: 116-118), refer to the extralinguistic reality;
2. the definition refers to the only infralinguistic reality;
3. the definition refers to the only extralinguistic reality.

### **Verbal Aspect and the Frame Elements in the FrameNet for Polish**

Linde-Usiekniewicz, Jadwiga; Derwojedowa, Magdalena; Zawisławska, Magdalena

#### **9. Lexicological Issues of Lexicographical Relevance**

This paper deals with theoretical and practical problems involved when describing a language from the morphological aspect within the FrameNet. In terms of aspect and in lexicographical description of the Polish language, there is a tendency to treat pairs where the aspectual distinction is marked by suffix as a single lexical unit. Where the aspectual distinction is marked by a prefix, pairs represent different units, e.g. *kaszlnac* (pf. *to give a cough*) -*kaslac* (impf. *to cough repeatedly*) vs. *pisac* (impf. *to write, to be writing*) -*napisac* (pf. *to have written*). More complex sense relations between perfective and imperfective verbs complicate matters even more. In addition, aspectual pairs differ in terms of what constitute their core frame elements.

Many perfectives differ from their imperfective counterparts since they transform temporal quantification from a non-core to a core element of the frame, e.g. *Przesiedział w bibliotece dwie godziny, studiując rekopisy*. (He sat for two (solid) hours in the library, poring over the manuscripts) vs. *Siedział w bibliotece (przez) dwie godziny, studiując rekopisy* (He sat in the library for two hours, pouring over the manuscripts). Because of this, in the Polish version of FrameNet, each member of an aspectual pair will be initially given a separate description. Once the respective frames and frame elements for each perfective and imperfective member of an aspectual pair are established independently, the two putative frames will be compared in order to see if they can be conflated into a single frame.

## Verbos que traban discurso: implicaciones lexicográficas para el DAELE

López Ferrero, Carmen; Torner Castells, Sergi

### 9. Lexicological Issues of Lexicographical Relevance

Our work falls within the framework of the Project for the Elaboration of a Dictionary for Learning Spanish as a Foreign Language—*Diccionario de aprendizaje del español como lengua extranjera*, ref. HUM2006-06982, in progress at the Universitat Pompeu Fabra. In particular we analyse the syntactic and discursive behaviour of five semantic classes (as set by Bosque 2004), since they amount to clusters of verbs which share both the same meaning in context as a grammatical behaviour is similar so it seems and there is a high frequency of use in each type of verb of the syntactic structures and patterns. These semantic classes are the following:

1. Verbs for introducing, unaccusative verbs of existence and apparition: *ocurrir, suceder, existir, aparecer, resultar*, etc. (cfr. Bosque y Demonte 1999);
2. Metalinguistic verbs or verbs expressing ways of talking: *decir, afirmar, asegurar, explicar, referir*, etc.;
3. Verbs that convey to what extent the information they introduce is relevant: *destacar, detallar, especificar, mostrar, sobresalir*, etc.;
4. Verbs of comparison and contrast: *comparar, contrastar, distinguir, diferenciar, oponer*, etc.;
5. Cause-consequence verbs: *causar, concluir, confirmar, conseguir, depender*, etc.

All of them are verbs that have been defined by text linguistics as explicit marks of textual connection in several lexicological works. The purpose of our analysis is to

define the syntactic patterns and the discursive values of these groups of verbs that have such a close meaning. The constructions and combinations akin to them where the different types of verb intervene have been described in detail (Bosque y Demonte 1999 and Bosque 2004, to mention two recent works); these descriptions should be completed with the information that shows quite specifically the shared meaning and the specific syntactic meaning of two verbs of each of the semantic classes considered. This information may be systematized to carry out the lexicographical description so that it may contribute to avoiding the mistakes foreign students learning Spanish may make when using units that are semantically similar.

### **Verb Class-specific Criteria for the Differentiation of Senses in Dictionary Entries**

Proost, Kristel

#### 9. Lexicological Issues of Lexicographical Relevance

This contribution deals with the representation of verbs with multiple meanings or senses in general monolingual dictionaries. Criteria for differentiating senses in dictionary entries have traditionally been formulated with respect to the vocabulary in general. This paper argues that, while some criteria do indeed apply to the entire lexicon, many of them are relevant only to specific semantic classes. This will be demonstrated considering two selected verb classes: *speech-act verbs* and *perception verbs*. Like verbs of other classes, *speech-act verbs* and *perception verbs* may be ambiguous in different but recurrent ways. Since recurrent patterns of ambiguity are always typical of particular semantic classes, class-specific semantic criteria are formulated to decide whether a particular ambiguous speech act or perception verb should be treated as being polysemous or homonymous in dictionary entries. In addition to these class-specific semantic criteria, the semantic-syntactic criterion of identity or difference of argument structure is suggested for the lexicographical representation of verbs which may not be considered to be polysemous or homonymous on the basis of semantic criteria alone. According to the suggested argument-structure criterion, these verbs should be treated as polysemous when their senses correlate with identical argument structures and as homonymous when their senses correlate with different argument structures properties. As opposed to the semantic criteria suggested, the semantic-syntactic criterion of identity vs. difference of argument structure applies to verbs of different semantic classes. However, as will be illustrated by the discussion of the different senses of *smell*, it may sometimes force us to treat different but related senses as corresponding to two distinct lexical items. In order to solve this problem, the criteria suggested are supplemented by a preference rule stating that semantic criteria apply prior to the semantic-syntactic criterion of identity vs. difference of argument structure...

## Les dictionnaires québécois et le problème de la norme linguistique

Schafroth, Elmar

### 9. Lexicological Issues of Lexicographical Relevance

This paper deals with dictionaries of French in Quebec and the problem of language norm. In 2008, Quebec will celebrate its 400th anniversary. The publishing of the first *Dictionary of Standard French in Quebec* (Dictionnaire FRANQUS)—*Français Québécois Usage Standard*, announced as an online version for autumn 2008 and supposedly available in its printed version in 2009, will mark a new and important step in the history of Canadian French lexicography. It will be the fifth dictionary of French published within the last 20 years in Quebec, each of them conveying its own normative point of view. The article deals with these four dictionaries: the *Dictionnaire du français Plus à l'usage des francophones d'Amérique* (DFP), 1988; the *Multidictionnaire de la langue française* (MLT), 4th edition 2003; the *Dictionnaire québécois d'aujourd'hui* (DQA), 1992/1993; the *Dictionnaire québécois-français. Pour mieux se comprendre entre francophones* (DQF), 1999.

After discussing the problem of linguistic norm in general and then, especially with regard to Quebec, each of the four dictionaries will be analyzed according to a set of criteria in order to reveal the items indicating normativity. As a matter of fact, there are different types of normativity, such as the maximum orientation towards the standard of European French or the adherence to a more “Quebecist” attitude—legitimizing a Quebec variety of French. The criteria are:

- the dictionaries' prefaces and introduction
- their labels indicating the value or the “correctness” of a word or a meaning
- any normative comment—the lexicographical description of English loan words—anglicisms being one of the major problems of language planning in Quebec.

## On Connotation, Denotation and All That, or: Why a *Nigger* Is Not a ‘Black Person’

Van der Meer, Geart

### 9. Lexicological Issues of Lexicographical Relevance

In my paper I intend to demonstrate that it is, in the case of monolingual dictionaries, preferable to incorporate usage labels like *formal*, *vulgar* etc. in the sense definitions themselves instead of making them almost invisible by hiding them in the margins of the entries. I will also argue that it should be attempted to make clear exactly why a lexical item is said to be e.g. *humorous*.

## Definición lexicográfica y orden de la información y de las palabras: el caso del euskera

Alberdi Larizgoitia, Xabier; García García de los Salmones, Julio; Ugarteburu Gastañares, Iñaki

### 10. Other topics

The aim of this paper is to show how every language is determined by its syntactic structure when arranging information and words in lexicographic definitions. In the first two sections we analyze the difficulties we find in Basque with hyperonymic definitions of nouns: the informative structure. This kind of definition imposes goes from the general—hyperonym—to the particula— specific characteristics; but this order contrasts with the expansion of nominal nucleus modifiers towards the left, often found in Basque. Bearing in mind that Basque history regarding monolingual lexicography is very recent, in the third section we discuss and evaluate the solution to the aforementioned problem that Ibon Sarasola provides in his *Euskal Hiztegia* (Basque Dictionary, 1996). Basically, the solution resorts to appositional structures that get us closer to an analytical and more communicative model of definition: this way, the informative nucleus—hyperonym—precedes the specifications—modifiers in apposition. Finally, we extract the following conclusions of general—lexicography—and individual nature—Basque lexicography:

Every language is determined by its syntactic structure regarding the model of hyperonymic definition and, because of that, it must look for its own syntactic-discursive strategies.

1. In the Basque dictionary *Euskal Hiztegia*, Sarasola specifies some syntactic-discursive strategies that have been proven adequate for the definition, which converge nicely with written tradition: one of the keys for these strategies is a moderate use of appositional structures to add specifications to the informative nucleus— hyperonym.
2. The conclusions we have arrived at are also valid for the terminographic definition written in Basque, which similarly cannot differ in excess from the paradigm *hyperonym (informative nucleus) + specific characteristics*.

## **From Lexicographic Evidence to Lexicological Aspects: A Cognitive Linguistic Perspective on Phonaesthetic Intensifiers**

Cacchiani, Silvia

10. Other topics

Depending on source domain, pattern of intensification and extent of grammaticalization, intensifiers may differ in a number of ways: degree (Paradis 2000, 2003) and degree and polarity sensitivity (Klein 1998); semantic prosody (Bublitz 1998); genre and register restrictions (Paradis 2000, 2003, Ito and Tagliamonte 2003), type and degree of expressivity, extent to which they can take part in reinforcing, aggravating or mitigating the underlying speech act, and, of course, collocational profile (Cacchiani 2005).

In the light of this, it is the purpose of this paper to show how lexicographic data can provide evidence in favour of adopting a cognitive-linguistic perspective on the process of loosening and meaning recreation which characterizes the development of intensifiers from other categories. Specifically, using data from the *Oxford English Dictionary*, I shall investigate the nature and use of phonaesthetic intensifiers (e.g. *howlingly*), within the framework of Ruiz de Mendoza's (1998ff) Combined Input Hypothesis. As will be seen, this helps shed light on the pattern of intensification (Lorenz 2002, Cacchiani 2005) at play while acknowledging the role played by contextual and encyclopaedic knowledge. Using the Combined Input Hypothesis, therefore, offers considerable lexicological insights while providing reasonable motivations for the polysemous nature of phonaesthetic intensifiers, and also accounting for discourse- pragmatic restrictions on their use. As such, it might integrate pragmatic, lexicographic and grammaticalization approaches to the study of intensifiers and, second, to the inclusion and representation of non-grammaticalized, peripheral intensifiers in advanced learner's dictionaries and, most importantly, bilingual dictionaries, which do not always include entries and subentries for phonaesthetic intensifiers.

## **Dictionaries for University Students: A Real Deal or Merely a Marketing Ploy?**

Kosem, Iztok

10. Other topics

Universities in English-speaking countries have experienced a sharp rise in the number of students in the past decades. One of the biggest problems students face is learning how to communicate in academic English, a language they have not experienced before. One of the tools that students often use to tackle language-

related problems during their study is a dictionary. There are many dictionaries on the market, but only a few claim to be designed specifically for university students. This paper takes a closer look at these few dictionaries, and attempts to identify their unique features by comparing them with general dictionaries. The analysis reveals that the only real difference lies in the additional material—e.g. sections on academic writing, and not in the dictionary macrostructure or microstructure itself. The second part of the paper focuses on some of the features that dictionaries for university students share with general dictionaries, such as being based on corpus data, and discusses why many of these features cannot actually be acknowledged as student-friendly. The final remarks point out that publishers, researchers, and lexicographers need to acknowledge that students are a specific group of dictionary users—users that need help, not only with regard to general language, but also with academic language.

### **El tratamiento lexicográfico de *de toute façon*, *de quelque façon* y *d'une certaine façon* en el DEC**

Llopis Cardona, Ana

10. Other topics

In this paper, I will examine the entries of the discourse markers *de quelque façon*, *d'une certaine façon* and *de toute façon* of the *Dictionnaire explicatif et combinatoire du français contemporain* (V.III) that Igor Mel'čuk led in the University of Montréal. I will propose to apply the existing format of the *lexies non descriptives* found in the IV volume to the definition of these discourse markers.

I will start with a linguistic description of these discourse markers in different levels. I will also explain the main aspects about macrostructure: *de quelque façon* and *d'une certaine façon* are recorded in different entries, but the first will take sends to the second, so there is only one definition for both markers. On the other hand, a *semantic building* can be established between *d'une certaine façon* and *de toute façon*.

Afterwards, I will analyze the aspects related to microstructure, namely: definition, lexical function and examples. The existing definition is a synonymous expression which is different from the typical definition of the DEC not only for the *lexies descriptives* but also for the *lexies non descriptives* found in the IV volume. With regard to the lexical function, several kinds of synonymy and antonymy are pointed out in the entries *d'une certaine façon* and *de toute façon*. The given expressions given of the lexical functions provide copious material to explore the differences between these discourse markers.

In conclusion, I note that the DEC is more appropriate for units that work in sentences, not in the discourse level; because the theoretical framework of the DEC was built to get the replacement of words and idiomatic expressions, but this system cannot deal with discourse markers. In addition, this lexicographical treatment doesn't include any pragmatic or communicative features, crucial to achieve a good description of these units.