PaCor 2016
Parallel Corpora: Creation and Applications

CONFERENCE PROGRAM

December 1-3, 2016, Santiago de Compostela, Spain
Practical information

Venue
University of Santiago de Compostela, Faculty of Philology, Avenida de Castelao s/n, 15782 Santiago.
Conference rooms:
   Salón de Grados: Ground floor
   B03 and B13: 1st Floor

Internet access
Two wireless networks are available:
1) eduroam If your university takes part in the eduroam initiative, you can register for the university’s network via eduroam with the user credentials of your home university.
2) Inside the Faculty, there is another internet access available for conference attendees: the network pacor2016
   Login: pacor2016 Password: RT%23rP

Coffee breaks and catering
Coffee breaks with snacks and drinks are served on Thursday and Friday in the Cafeteria, on the ground floor. During the short breaks, there will be water and fruit outside of the conference rooms.
No official lunches are provided. However, lunch facilities will be available in the Cafeteria. In addition, there are some restaurants very close to the Faculty building, all of which offer a daily menu set between 7 and 14 €. Some of them are indicated on the Google map of the Symposium webpage.

Social program

Conference Dinner on Friday
Restaurante San Francisco, Hotel Monumento.
Campillo de San Francisco nº 3 15705 Santiago de Compostela
Location indicated on the Symposium website.

Guide on Saturday: (Please sign up at the reception)
Meeting point: 10.30 at Plaza do Obradoiro in front of the Palacio de Raxoi (directly opposite of the Cathedral). See the Symposium website.
Welcome Address

As Organizing Committee, we are pleased to cordially welcome you to PaCor2016. We are delighted that colleagues from different countries and colleagues from other Spanish Universities have accepted our invitation. With your contributions, you all have made PaCor2016 possible. Parallel corpora are valuable linguistic resources for many different research fields and practical applications. The trigger of this Symposium and of our approach to the parallel corpora was our research on the expression of the spatial relations in German and in Spanish, an area our group has been working on for years. The available parallel corpora including German and Spanish proved to be not especially productive for our research. Hence, our approach to parallel corpora has its roots in the field of contrastive linguistics research.
Likewise, given the wide range of applications of parallel corpora, we assume that each of you entered the world of parallel corpora with different backgrounds, from different research fields and with different purposes: natural language processing, SMT, translation, applied linguistics, lexicography, language learning, among others. However, what is certainly common to all of us is that we have to manage a number of major challenges when building parallel corpora. Indeed, the main objective of this Symposium is to help us meet these challenges by bringing together different research perspectives. Another goal is to provide a platform for presenting existing parallel corpora and running projects including Spanish. In this regard, we are pleasantly surprised by the number and quality of these initiatives.
The symposium program reflects this plurality of interests and perspectives. In accordance with the main themes of the Symposium, our keynote speakers will lecture on the retrieval of parallel texts, on corpus management and exploitation and on applications of parallel corpora in different fields such as applied linguistics and translation studies. The presentations of the Symposium cover different steps to build parallel corpora and a wide variety of their applications. Looking at the program conference and the wide-range of abstracts, we are sure that they will provide all of us with a rich debate, interesting discussions and fruitful exchanges of views from a range of perspectives.
We invite you to take advantage of the opportunity for professional exchange and personal meeting offered by the Symposium. Like pilgrims have done for ages, we hope that you will leave Santiago inspired and that you may take home with you new insights, and the memories you have made of thought provoking discussions and exchanges.
Finally, we would like to thank you warmly for coming and we wish you successful and inspiring days in Santiago.

The Organizing Team
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30-09:00</td>
<td>Registration</td>
</tr>
<tr>
<td>09:00-09:30</td>
<td>Welcome Address: Mª JOSÉ LÓPEZ COUSO (Dean of the Faculty), IRENE DOVAL (Chair)</td>
</tr>
<tr>
<td>09:30-10:30</td>
<td>MARTIN VOLK (U. Zürich) Innovations in parallel corpus alignment and retrieval</td>
</tr>
<tr>
<td>11:00-11:30</td>
<td>COFFEE BREAK</td>
</tr>
<tr>
<td>11:30-12:00</td>
<td>B13 LIESBETH AUGUSTINUS (U. Leuven) Using parallel treebanks for comparative syntax</td>
</tr>
<tr>
<td></td>
<td>B03 PABLO GAMALLO (U. Santiago de Compostela) Strategies to build high quality bilingual lexicons from comparable corpora</td>
</tr>
<tr>
<td></td>
<td>B03 ADRIANO FERRARESI, SILVIA BERNARDINI (U. Bolonia) Building and aligning a tri-directional intermodal corpus: the case of EPTIC</td>
</tr>
<tr>
<td></td>
<td>B03 LIDUN HAREIDE (Møreforsking Volda) Comparable parallel corpora</td>
</tr>
<tr>
<td>12:15-12:45</td>
<td>BREAK</td>
</tr>
<tr>
<td>12:45-13:15</td>
<td>LUNCH</td>
</tr>
</tbody>
</table>

**Friday, December 2, 2016**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00-09:30</td>
<td>Registration</td>
</tr>
<tr>
<td>09:30-10:30</td>
<td>Welcome Address: Mª JOSÉ LÓPEZ COUSO (Dean of the Faculty), IRENE DOVAL (Chair)</td>
</tr>
<tr>
<td>11:00-11:30</td>
<td>COFFEE BREAK</td>
</tr>
<tr>
<td>11:30-12:00</td>
<td>B03 PETR ČERMÁK (Charles U.) InterCorp, corpus paralelo de 40 idiomas</td>
</tr>
<tr>
<td></td>
<td>B03 GIORGINA CERUTTI BENÍTEZ (U. Geneva) Evaluating fourth-generation concordancers to meet legal language research needs: a case study in legal translation studies</td>
</tr>
<tr>
<td>12:15-12:45</td>
<td>BREAK</td>
</tr>
<tr>
<td>12:45-13:15</td>
<td>LUNCH</td>
</tr>
<tr>
<td>Time</td>
<td>Room</td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
</tr>
<tr>
<td>15:30-16:30</td>
<td>Salón de Grados</td>
</tr>
<tr>
<td>16:30-17:00</td>
<td><strong>COFFEE BREAK</strong></td>
</tr>
<tr>
<td>17:00-17:30</td>
<td>B13</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>17:30-18:00</td>
<td>B03</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>17:30-18:00</td>
<td>B13</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>17:30-18:00</td>
<td>B03</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>18:15-18:45</td>
<td><strong>BREAK</strong></td>
</tr>
<tr>
<td>18:15-18:45</td>
<td>B13</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>18:15-18:45</td>
<td>B03</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>18:45-19:15</td>
<td>B13</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Thursday, December 1, 2016**
<table>
<thead>
<tr>
<th>Time</th>
<th>Speaker(s)</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:30-10:30</td>
<td>JOSÉ JOÃO ALMEIDA (U. Minho)</td>
<td>Exploring the world for parallel texts</td>
</tr>
<tr>
<td>11:00-11:30</td>
<td>XAVIER GÓMEZ GUINOVART (U. Vigo)</td>
<td>Del corpus CLUVI al corpus SensoGal: la incorporación de la semántica al estudio de los corpus paralelos</td>
</tr>
<tr>
<td>11:30-12:00</td>
<td>SHUYUAN CAO, IRIA DA CUNHA, MIKEL IRUSKIETA (U. Pompeu Fabra. UNED, U. Basque Country)</td>
<td>A Spanish-Chinese parallel corpus for natural language processing purposes</td>
</tr>
<tr>
<td>12:15-12:45</td>
<td>ANA ELVIRA OJANGUREN LÓPEZ (U. La Rioja)</td>
<td>Gathering information from a relational database to annotate and tag a parallel corpus of Old English</td>
</tr>
<tr>
<td>12:45-13:15</td>
<td>JAVIER MARTÍN ARISTA (U. La Rioja)</td>
<td>The design of a parallel corpus of Old English on a lexical database</td>
</tr>
<tr>
<td></td>
<td>PABLO RUANO SAN SEGUNDO (U. Extremadura)</td>
<td>Parallel corpora and corpus literary translation studies: exploring the translation of Charles Dickens’s novels into Spanish</td>
</tr>
<tr>
<td></td>
<td>MAR SOLIÑO PAZÓ (U. Salamanca)</td>
<td>Corpus PaGeS: Un corpus paralelo bilingüe para el estudio de las colocaciones. Un estudio piloto en la didáctica de las lenguas extranjeras</td>
</tr>
<tr>
<td>Time</td>
<td>Room B13</td>
<td>Room B03</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>15:30-16:30</td>
<td><strong>JOSEP MARCO BORILLO (U. Jaume I)</strong>&lt;br&gt;Living with parallel corpora: Potential and limitations for their use in translation research and translator education</td>
<td></td>
</tr>
<tr>
<td>17:00-17:30</td>
<td><strong>MARÍA TERESA SÁNCHEZ NIETO (U. Valladolid)</strong>&lt;br&gt;The concept of secondary sample corpus and its relevance for the work with online searchable parallel corpora</td>
<td><strong>PATRICIA SOTELO DIOS (U. Vigo)</strong>&lt;br&gt;Translation-oriented annotation of a multimedia parallel corpus of subtitles</td>
</tr>
<tr>
<td>17:30-18:00</td>
<td><strong>MERCÈ LORENTE, AMOR MONTANÉ, JORGE VIVALDI, LLUÍS DE Yzaguirre (U. Pompeu Fabra)</strong>&lt;br&gt;Parallel corpora of new Catalan Civil Code: a terminology harmonization project</td>
<td><strong>BÁRBARA MARTÍNEZ VILINSKY (U. Jaume I)</strong>&lt;br&gt;The compilation of a comparable and parallel corpus of crime novels to study unique items in translation</td>
</tr>
<tr>
<td>18:15-18:45</td>
<td><strong>TERESA MOLÉS-CASES, ULRIKE OSTER (U. Jaume I)</strong>&lt;br&gt;Presentation and compilation of a section of the COVALT Corpus (EN, FR, DE &gt; ES)</td>
<td><strong>MARCOS GARCIA, MARCOS GARCÍA-SALIDO, MARGARITA ÁLONSO RAMOS (U. da Coruña)</strong>&lt;br&gt;Discovering bilingual collocations in parallel corpora: a first attempt at using distributional semantics</td>
</tr>
<tr>
<td>18:45-19:15</td>
<td><strong>IRENE DOVAL, SANTIAGO FERNÁNDEZ, TOMÁS JIMÉNEZ, ELSA LISTE, BÁRBARA LÜBKE (U. Santiago de Compostela)</strong>&lt;br&gt;Corpus PaGeS: presentations of a bilingual corpus for multiple purposes</td>
<td><strong>CLARA INÉS LÓPEZ RODRÍGUEZ (U. Granada)</strong>&lt;br&gt;Combinatory potential of verbs in English and Spanish medical texts: finding “abnormal clinical findings” in parallel corpora</td>
</tr>
</tbody>
</table>
Keynote Speakers

**MARTIN VOLK** (Thursday 1, 9:30, Salón de Grados)
University of Zürich, Switzerland

**Innovations in parallel corpus alignment and retrieval**
Parallel corpora are valuable resources for language and translation studies. Large parallel corpora like Europarl or Opensubtitles are now available for research. They are most useful when they are aligned across languages down to the word level.

In this talk I will introduce and classify various search tools over parallel corpora.

I will argue that automatic word alignment provides added value for the linguistic exploration and data-driven investigation of large parallel corpora. I will explain how word alignment works and how it can be integrated in corpus query systems. As a case in point I will demonstrate our search system Multilingwis. I will present a number of examples on how to use parallel corpora for different linguistic studies such as computing synonyms, finding multiword expressions, and detecting new translation variants.

** ROSA RABÁDÁN ** (Thursday 1, 15:30, Salón de Grados)
University of León, Spain

**Parallel corpora: usefulness, usability and applications**
Parallel corpora are a key element in cross-linguistic studies. Designing a parallel corpus is a costly and time-intensive activity that comprises collecting, annotating, aligning and deciding upon an efficient browser. Current corpora design decisions are generally guided by the use the corpus is going to have, i.e. dictionary-making, grammatical contrast, statistical machine translation, unveiling translation regularities, etc. However, a review of a number of parallel corpora featuring Spanish and their uses suggests other criteria such as language combination(s) and directionality, among others, may also affect their usefulness.

The usefulness of parallel corpora will increase exponentially when data so obtained are combined with those yielded by comparable and/or monolingual corpora. Techniques such as Chesterman’s verification of target language fit can render valuable insights into learner’s interlanguage or discrepancies in translational obligatory adjustments.

After a brief consideration of a number of possible applications of parallel corpora illustrative corpus-based case studies of computer-aids for technical and professional writing and text evaluation will be presented.
JOSÉ JOÃO ALMEIDA (Friday 2, 9:30, Salón de Grados)
University of Minho, Portugal

**Exploring the world for parallel texts**

There is a huge amount of translated documents (bitexts) in the web and other sources. Unfortunately find them is not an easy task. In this talk we will discuss several techniques for mining bitexts from a variety of sources:

* bitext candidates from the web (based on patterns, name & URL similarities, topology patterns)
* bitext candidates from large document collections
* bitexts from multilingual documents (ex bilingual abstract collections)
* bi-books and bi-manuals
* subtitles (films, documentaries, etc)
* simple small bilingual documents (ex bilingual instructions, menus)
* translation memories, software messages.

We will also focus on some problematic situations, noisy parallel corpora, quality issues and measures.

* bitexts that were created by automatic machine translation
* partial translations
* untranslated parts
* comparable texts
* alphabetic ordered lists and dictionaries (difficult to align)
* alignment metrics and editions

Whenever pertinent, we will also discuss the linguistic and terminological richness of the bitexts.
Living with parallel corpora: potential and limitations for their use in translation research and translator education

It would seem at first sight that parallel corpora are close natural allies both of the translation researcher and the translator educator. And surely they are, but they have been rather slow in finding their way into the researcher’s (and the educator’s) toolkit. This is partly due to the very visible difficulties of compiling and preprocessing a parallel corpus; but there are other reasons as well. Baker’s influential articles of the mid-1990s pointed the way towards descriptive translation research based on comparable corpora, where source texts played virtually no role. In that kind of methodological framework, parallel corpora were not strictly necessary. But, as time went by, they increasingly came to be seen as an essential tool in corpus-based translation research, as translation decisions are often directly traceable to the source text profile.

Parallel corpora can be used in translation research in at least two ways: as the main source of data or as a supplement to data retrieved from a comparable corpus, enabling triangulation. In the former they may throw light on contrastive aspects or on translator techniques and methods, whereas in the latter they will tend to be searched for explanatory purposes, i.e. to account for differences perceived between the two components of a comparable corpus.

These two possible uses of parallel corpora will be illustrated by means of several studies carried out within the COVALT Group, at Universitat Jaume I, on such aspects as phraseology, connectives and collocation. In translator education, parallel corpora (like indeed any other kind of corpus) may be used both as documentation resources and as sources for materials to be designed by the trainer and then implemented in the classroom. But perhaps more ambitious schemes might be envisaged (in which corpora would play an important role in shaping the students’ overall view of the task of translating) if parallel corpora as a resource were more generally available.
Abstracts

(Sorted alphabetically by the name of the first author)

LIESBETH AUGUSTINUS
University of Leuven, Belgium

Using parallel treebanks for comparative syntax

In contrastive linguistics and translation studies it is common to use parallel corpora, see amongst others Johansson (2007). For studies in comparative syntax, however, we need syntactically annotated parallel corpora: “Exploring grammatical phenomena in a multilingual corpus is a difficult and time-consuming task involving manual intervention (...). In general, there is a lack of large multilingual corpora with advanced syntactic annotation. Developing such data is an important task for the future" (Johansson 2007: 37). In the years since Johansson wrote this, some syntactically annotated parallel corpora or "parallel treebanks" have been created, e.g. the SMULTRON parallel treebank (Volk et al. 2010), and the parallel treebanks included in the INESS platform (http://clarino.uib.no/iness). In contrast to (flat) parallel corpora, parallel treebanks are only available for a limited number of languages, and they are often small in size. Moreover, one typically needs sophisticated tools and methods to query such treebanks. Some users are deterred by this, which means that the potential of parallel treebanks will not be realized.

In order to make parallel treebanks accessible to non-technical users, Poly-GrETEL (http://gretel.ccl.kuleuven.be/poly-gretel) has been developed. It is an online tool which enables syntactic querying in parallel treebanks. Similar to the monolingual GrETEL environment (http://gretel.ccl.kuleuven.be/gretel-2.0), Poly-GrETEL allows users to query parallel treebanks by means of natural language examples instead of a formal query. This approach is called "example-based querying", which is a stepwise query procedure to create a search instruction with limited knowledge of the treebank annotations and the exact layout of the syntax trees. As an alternative one can skip this procedure and use the XPath query language to search the treebanks.

Currently Poly-GrETEL provides access to the Europarl parallel treebank for Dutch and English (Koehn 2005), which is automatically parsed and aligned on sentence and node level. Soon the German-Dutch parallel treebank will be added as well.

By combining the example-based query functionality with node alignments, Poly-GrETEL limits the need for users to be familiar with the query language and the structure of the trees in the source and target language. In this way, the tool facilitates the use of parallel treebanks for comparative linguistics and translation studies.

In addition to a demonstration of Poly-GrETEL, a number of methodological aspects regarding the use of automatically created parallel treebanks for comparative syntactic research will be discussed.

References:


**SHUYUAN CAO, IRIA DA CUNHA, MIKEL IRUSKIETA**
Pompeu Fabra University, National Distance Education University, University of the Basque Country, Spain

**A Spanish-Chinese parallel corpus for natural language processing purposes**

"Corpus-based discourse analysis for Natural Language Processing (NLP) is becoming more and more popular. In several NLP domains discourse information is being used, such as automatic summarization, machine translation (MT) and machine translation evaluation, information extraction, etc. Therefore, the aim of this work is to introduce a new Spanish-Chinese parallel corpus annotated with discourse information, which could help for the development of NLP applications concerning this language pair, such as MT or bilingual summarization.

The theoretical framework of this work is the Rhetorical Structure Theory (RST) (Mann and Thompson, 1988). RST addresses coherence by means of relations between text spans in a tree-like structure. Texts are segmented into discourse units and coherence relations are established between these discourse units. As these relations are recursive, they form text structures linking units or groups of units (spans). The units can be Nuclei or Satellites; satellites offer additional thematic information about nuclei. Relations can be nucleus-satellite (e.g. ANTITHESIS, CAUSE and EVIDENCE) and multinuclear (e.g. CONJUNCTION, JOINT, and SEQUENCE). This parallel corpus has been annotated manually with RSTTool (O'Donnell, 2000) and follows the annotation method by Iruskieta, da Cunha and Taboada (2015), in order to annotate and create a gold standard corpus under RST. Three kinds of information for Spanish and Chinese in the corpus will be provided to NLP researchers: (a) discourse segmentation, (b) central unit (CU) and (c) discourse relations.

Firstly, segmentation is the first step of discourse analysis and is useful for different NLP tasks, such as, evaluation of automatic segmentation systems, development of discourse parsers and text summarization. In our corpus, the users can compare the discourse similarities and differences between the parallel Spanish-Chinese texts by counting the quantity of EDUs, the number of discourse markers (DM), the position of DMs, etc.

Secondly, the CU is the key information or the main topic of a text, which can be applied to automatic summarization, development of intelligent systems and sentiment analysis. Genre, domain and discourse structure determine the position of the CU in a text; thus, by consulting the CU of the texts in the corpus, users can know how to organize the information of texts in different genres and domains. A good translation of the main topic or CU is also fundamental for a MT system.

Thirdly, discourse relations show the coherence of a language and are useful for several NLP tasks, such as, discourse parsing, information extraction, automatic summarization and evaluation of MT. Discourse structure can be used to detect discourse similarities and differences between Spanish-Chinese regarding discourse relations, type of relations and used translation strategies, among other issues."

---

**PETR ČERMÁK**
Charles University, Chzech Republic

**InterCorp, corpus paralelo de 40 idiomas**

La ponencia presenta InterCorp (www.korpus.cz/intercorp), el corpus paralelo online de 40 idiomas diferentes creado en la Facultad de Letras de la Universidad Carolina de Praga. Su novena versión, puesta en funcionamiento
en septiembre de 2016, cuenta en total con 1.647 millones de palabras; el componente español, el tercero más grande, consta de 105 millones. La lengua pivote es el checo: todos los textos existen en checo y paralelamente en una o más lenguas. De ello se desprende que existen no solo los 39 subcorpus creados por el checo y por cada lengua participante (p.ej.: subcorpus checo-español, checo-francés, etc.), sino también otros subcorpus que combinan dos o más idiomas diferentes del checo (p.ej.: subcorpus español-inglés). El corpus consta de dos grandes partes: la primera, denominada núcleo, la constituyen textos literarios, checos o de otros idiomas, y sus traducciones checas o a otros idiomas; la segunda, denominada colecciones, consta de textos administrativos, periodísticos y jurídicos, que mayoritariamente provienen de los repositores de textos de las instituciones de la Unión Europea. El ponente, que es el responsable del componente español del proyecto, no solo describirá las características básicas del corpus, sino que prestará igualmente atención a las diferentes maneras de utilizarlo. Comentará, además, algunos aspectos técnicos del funcionamiento del corpus, así como dos instrumentos específicos creados por el equipo de autores de InterCorp: el alineador de textos InterText y la interfaz KonText. Estos instrumentos son open-source: el equipo del InterCorp los ofrece a los interesados e incluso proporciona soporte técnico para su instalación y funcionamiento; es asimismo capaz de añadir funciones específicas según las necesidades de los usuarios.

GIORGINA CERUTTI BENÍTEZ  
University of Geneva, Switzerland

Evaluating fourth-generation concordancers to meet legal language research needs: a case study in legal translation studies

The advantages of using parallel corpora for research in applied translation studies is undeniable. This is also true in the field of Legal Translation Studies, where the potential of corpus-based studies as a methodology for researching legal language and translation (e.g. Gozdz-Roszkowski & Pontrandolfo, 2015; Biel, 2016), and as a tool in translator training (e.g. Monzó, 2008; Biel, 2010) has been demonstrated. However, there are major limitations regarding the use of legal parallel corpora, which concern their availability, as well as the tools that can be used to analyse them.

It is nowadays generally acknowledged that the tools used to investigate corpora have a considerable impact on the results obtained, and in the way research questions are addressed and expanded. Nonetheless, in many corpus-based studies, insufficient efforts have been devoted to establish the theoretical link between research questions, the specificities of the language for special purposes (LSP) analysed, and the corpus tools used to explore these questions.

In this context, our paper will highlight methodological considerations for the selection of concordancers in order to analyse parallel legal corpora. The results of a comparative analysis of currently available fourth-generation concordancers for parallel corpora –CPQweb, SketchEngine and ParaVoz– will be presented with a view to verifying their suitability for a specific research project according to its stated and implied needs. These tools have been tested in accordance with the relevant guidelines and norms for software evaluation. The results and conclusions drawn from this analysis can be extrapolated to other studies on the analysis of parallel specialized corpora.
References
Biel, Ł. (2016): Mixed corpus design for researching the Eurolect: a genre-based comparable-parallel corpus in the PL EUROLECT project Mieszana struktura korpusu do badania eurolektu–gatunkowy korpus porównawczo-równoległy w ramach projektu PL EUROLECT.

IRENE DOVAL, SANTIAGO FERNÁNDEZ, TOMÁS JIMÉNEZ, ELSA LISTE, BÁRBARA LÜBKE
University of Santiago de Compostela, University of La Rioja, Spain

Corpus PaGes: A bilingual parallel corpus for multiple purposes
The aim of this paper is to present the bilingual parallel corpus PaGeS, which is currently being built by the research group Spatiales from the University of Santiago de Compostela. The corpus was originally intended for linguistic research purposes, more specifically for the contrastive analysis of the expression of spatial relations in German and Spanish.

The Corpus PaGeS currently amounts to nearly 11 million tokens and consists of original texts in German and Spanish (about 5 million tokens) and their correspondent translation into the other language (about 5.200.000 tokens), as well as German and Spanish translations from third languages (about 550.000 tokens).

Our contribution will introduce the main characteristics of our corpus, focusing on its design and compilation process. Thus, the different criteria used for the selection of the texts and their preparation for further processing will be explained. Then, the process of sentence alignment with LF-Aligner and the steps developed for the manual review of the alignment will be described. Furthermore, the linguistic annotation of the texts (with Treetager for German and Freeling for Spanish), the indexing process at Solr search engine, and the development of querying tools with java technologies will be detailed.

Though the results so far are reasonably acceptable, our contribution will also point to some improvements needed as for word alignment and some aspects of searching procedures. Finally, we will briefly discuss other possible applications of our corpus, and show how it could be used in translation research or language teaching research.

ADRIANO FERRARESI, SILVIA BERNARDINI
The University of Bolonia, Italy

Building and aligning a tri-directional intermodal corpus: the case of EPTIC
Gile (2004: 10) suggests that translation and interpreting "share epistemological, methodological, institutional and wider sociological concerns” and that "[i]t, therefore, makes much sense for both disciplines to work together in
spite of the differences”. The intermodal corpus approach, pioneered by the Shlesinger (e.g. 1998), provides an ideal research framework to investigate commonalities and differences between interpreting and translation.

Initiated as an offshoot of EPIC (Russo et al 2012), and originally including only Italian and English speeches from 2004 and their interpreted and translated versions (Bernardini et al. 2016), the European Parliament Translation and Interpreting Corpus (EPTIC) is now growing through a community effort. The current version, developed jointly by the University of Bologna, UCLouvain and the University of Belgrade, includes three bidirectional sets of corpora: English <-> Italian, English <-> French, and French <-> Italian, such that each source language corpus is aligned to interpretations and translations into two target languages. EPTIC includes 64 speeches in English, 65 in French and 68 in Italian delivered in January 2011. The overall size of the trilingual corpus, which consists of 18 multi-aligned sub-corpora, is approximately 350,000 words.

In this contribution we describe the major methodological and technical challenges involved in the construction of this multilingual intermodal corpus. In particular, we focus on the processes and tools used for a) text alignment at sentence level (taking into account the large number of bi-directional alignment directions), b) annotation with linguistic and contextual information, and c) indexing for consultation by the public through the NoSketchEngine platform (Rychlý 2007). Finally, we report on recent attempts to also perform text-to-video alignment, allowing corpus users to access time-aligned multimedia files of the original and interpreted EP speeches from concordances.

References

PABLO GAMALLO
University of Santiago de Compostela, Spain

**Strategies to build high quality bilingual lexicons from comparable corpora**

This work outlines two strategies to automatically build bilingual dictionaries: One is based on a pivot language and existing bilingual dictionaries. The other relies on string similarity and cognate extraction. Both strategies have in common the use of translation equivalents extracted from comparable corpora to filter out odd bilingual pairs and validate the correct ones. The correctness of the entries validated with comparable corpora is reasonable, close to that achieved by using parallel corpora. The paper reports several case studies describing how to build new high-
quality bilingual lexicons with more than 90% precision. This outperforms state-of-the-art systems on bilingual extraction from comparable corpora, whose best scores hardly reach 70 or 80% accuracy.

ÁLVARO GARCÍA ARAGÓN, CELIA RICO PÉREZ
University of Europe, Spain

Multilingual parallel corpora in migration: the case of IN.MIGRA2-CM

IN.MIGRA2-CM (H2015/HUM3404) is an interdisciplinary project funded by the Autonomous Community of Madrid (Spain) and the European Social Fund, and carried out by the joint efforts of Universidad de Alcalá de Henares (leading institution), Universidad Nebrija and Universidad Europea de Madrid (collaborating institutions). This project has the aim of seeking and carrying out an analysis of the migrant population in Madrid from different, yet complementary, perspectives: sociolinguistics, language learning, discourse analysis and translation. With this premise, at the department of Translation at Universidad Europea, we have set the task of creating multilingual resources for translation in the field of migration: translation memories, terminology databases, machine translation and reference materials.

The foundation for all these resources remains in the creation of a multilingual parallel corpus (English, French and Spanish) to be ultimately applied as the baseline for feeding translation memories and machine translation engines. In this connection, worked conducted so far concentrates in building an ad hoc parallel multilingual corpus (Castillo, 2009; Corpas, 2001) that meets the specific needs of future users (translators). Thus, creating an ad hoc corpus in the area of humanitarian aid, and more specifically in migration, with texts available on the Internet, is of special interest in view of the scarcity of materials available to the translator (Rico, 2015).

The present communication will give specific details of the process of INMIGRA’s corpus compilation, its design, storage and alignment. It is important to mention that this is an ongoing project due to end by December 2018, so results are still in the works. Nevertheless, preliminary data will be put forward as an illustration of how INMIGRA addresses the challenge of creating specialized corpora for the purposes of machine translation.

References
MARÇOS GARCIA, MARÇOS GARCÍA-SALIDO, MARGARITA ALONSO RAMOS
University of A Coruña, Spain

Discovering bilingual collocations in parallel corpora: a first attempt at using distributional semantics
Parallel corpora have been widely used both in corpus linguistics and in natural language processing to extract linguistic equivalences between two or more languages. One of these resources is OpenSubtitles2016, a large repository of Movie and TV subtitles which includes aligned sentences for more than 60 linguistic varieties. Using the mentioned dataset, this paper explores the automatic extraction of bilingual collocations between Spanish and Portuguese. First, we extract monolingual collocations (e.g., the Spanish “alcanzar (la) madurez” – “reach (the) maturity”) using dependency parsing and standard association metrics, such as mutual information or t-score. Then, for the base of each collocation in the source language (e.g., 'madurez' in “alcanzar (la) madurez”), we search for translation equivalents in the collocations of the target language. If a translation is found (e.g., 'maturidade', in Portuguese), we also search for candidates of the collocate. Finally, if both elements appear in a previously extracted collocation of the target language, we select it as a translation equivalent of the source collocation (“alcanzar (la) madurez”, “atingir (a) maturidade”). The search of the translation equivalents is performed using a bilingual model of distributional semantics, which has been previously trained on an automatic aligned version of the same Spanish-Portuguese parallel corpus. Preliminary results in verb-noun combinations (where the noun is the base, acting as a direct object) show that the proposed strategy discovers high-quality collocations in parallel corpora.

PARIJAT GHOSHAL, XI RAO
University of Zürich, Switzerland

Normalization of shorthand forms in French text messages using word embedding and machine translation
This paper focuses on the normalisation of abbreviations and shorthand forms used in French text messages. In order to find the abbreviated forms, we aligned normalised and non-normalised French text messages and built a parallel corpus. Abbreviations and shorthand forms pose a challenge during the normalisation process, as they mostly cannot be resolved by typical spell checkers and dictionary lookups. Hence, we propose a comparison of two popular approaches for text normalisation, namely distributed representation of words that tend to appear in a similar context and character-based machine translation. Taking contextual information into consideration, Multivec is applied to retrieve the candidates for those "noisy" forms. As a more traditional approach, we use character-based machine translation to map the canonical and "noisy" forms.
We compare our results and observe the efficacy of our models while normalising deletions, substitutions, repetitions, swaps, and insertions, made to canonical forms. Our work is the first paper that uses Multivec as a tool and the Belgian corpus collected under the SMS4Science Project to normalise non-standard French. We compare our results to other popular approaches, such as word/phrase-based machine translation. Due to the unsupervised nature of learning in this project, our system is highly flexible and can be easily adapted to text normalisation in other domains, and it can be further improved with more training materials.

XAVIER GÓMEZ GUINOVART
University of Vigo, Spain

Del Corpus CLUVI al Corpus SensoGal: la incorporación de la semántica al estudio de los corpus paralelos

En esta exposición presentaremos una perspectiva histórica de la investigación sobre corpus paralelos llevada a lo largo de más de una década en el Seminario de Lingüística Informática de la Universidade de Vigo, centrándonos en algunas de las líneas de trabajo más fructíferas surgidas del desarrollo y explotación del Corpus CLUVI y en las investigaciones más recientes alrededor de la incorporación de los avances en semántica léxica computacional al diseño y estudio de los corpus paralelos.

LIDUN HAREIDE
Møreforsking Volda, Norway

Comparable parallel corpora

The aim of this paper is to present the concept of comparable parallel corpora (Hareide and Hofland 2012). Comparable parallel corpora are two or more translation-corpora specifically created to be comparable, and to allow for the testing of the same hypothesis on at least two different language pairs. Another benefit of comparable parallel corpora is that they allow for the testing of complex hypotheses like Halverson’s Gravitational Pull hypothesis (Halverson 2003, 2007, 2009, 2010). The two parallel corpora contain original text in one language and the translations of these texts into one other language, and they are matched as far as possible in terms of sampling frame; that is text-type, subject matter and communicative function as well as time-frame and the […] variety sampled” (Hareide 2014:25, based on Altenberg & Granger’s definition of comparable corpora (2002: 7–8)).

In this paper will briefly present the 4.1 million word Norwegian-Spanish Parallel Corpus (NSPC) (Hareide and Hofland 2012). Then I will outline how the the NSPC was sampled to be both a freestanding parallel unidirectional translation corpus of published contemporary Norwegian original texts and their published translations into Spanish, as well as designed to be comparable to the English-Spanish P-ACTRES corpus compiled at the University of León. I use Altenberg and Granger’s definition of comparable corpora: “original texts in each language, matched as far as possible in terms of text type, subject matter and communicative function” (Altenberg & Granger 2002: 7–8). In addition the NSPC matches the time frame of the P-ACTRES, where 98 per cent of the contents date from 2000 onwards (Izquierdo et al. 2008b: 35) as well as the Spanish variety sampled.
I will then outline the measures taken to ensure that the NSPC is freely accessible and freely downloadable. Finally I will describe how data from the NSPC, the P-ACTRES and the CREA was used to successfully test the enigmatic Gravitational Pull Hypothesis (Halverson 2003, 2007, 2009, 2010).

It is my claim that a corpus is only as good as its sampling frame. Here I stand on the shoulders of giants, as already in 1998, Tymockzo called for caution when designing corpora, pointing out that corpora are “products of human sensibility, connected with human interests and self-interests. All the more reason, therefore, to consider the objects of study, the data gathered in the databases, and the parameters defining the corpora themselves very carefully” (Tymockzo 1998: 3).

Also in 2005 Rabadán called for studies in other languages than English, and for transparency in methodology so studies can be replicated with other data sets (Rabadán 2005: 165–166). Even today most parallel corpora are compiled to solve specific research questions. This often means that the research done on these corpora is hard to replicate on other language pairs, and therefore in a strictly scientific sense the value of these studies is limited.
Digital corpora annotated with complex linguistic information (i.e. semantic, pragmatic and discourse features) are fundamental for training and testing algorithms in the NLP community, and they are essential as gold standards for testing the performance of human language technology. In the Corpus Linguistics community, the annotation of texts adds value to a corpus in terms of reusability, stability and reproducibility, and corpus annotation has a tremendous potential as a topic of methodological cutting-edge research both for theoretical and applied corpus studies (Lavid 2012, Lavid et al. 2013). In this paper I describe recent work on the development of a bilingual (English-Spanish) corpus consisting of original comparable and parallel texts from a variety of genres and annotated with complex linguistic features such as modality and evidentiality (Lavid et al. 2016), metadiscourse markers (Lavid and Moratón 2016) and thematisation (Lavid and Moratón 2015), as carried out within the framework of the MULTINOT project (Lavid et al. 2015). The annotation of these complex features in bilingual parallel texts poses important challenges for the researcher at the different stages of the corpus development, from the preprocessing phases to the manual annotation phase in its different substeps (development of annotation schemes, agreement studies, annotation of a larger sample). But the creation of this type of multifunctional resource allows the investigation of complex linguistic research questions which cannot be addressed on the basis of raw corpora or even with the help of an automatic part-of-speech tagging system. In the presentation I address some of the possible investigations which can be carried out using this multifunctional resource and report on the results of recent work in the area of discourse markers.

References
ELSAS LISTE LAMAS
University of Santiago de Compostela, Spain

Sensory paths in German and Spanish: a contrastive study based on the PaGeS corpus

This contribution examines and contrasts the occurrences of Sensory Paths (cf. Talmy 2000) with verbs of visual (1), auditory (2) and olfactory (3) perception, and Manner of speaking verbs (4) in German and Spanish. For this purpose, it relies on the occurrences of this type of verbs in the PaGeS corpus, in both original and translated texts:

(1) Margarida blickte zum Mond hinauf […]
Margarida miró la luna
(2) […] und lauschte in die Weite der Schneefläche hinaus […]
[…] escuchando a lo lejos en la llanura nevada que se extendía en todas direcciones […]
(3) Selbst durch seine Haut schien es [das Kind] hindurchzuriechen, in sein Innerstes hinein.
(4) Le pareció incluso que le olfateaba hasta atravesarle la piel para oler sus entrañas.
(5) […] im ersten Moment rief ich doch tatsächlich in das Dunkel des nächtlichen Kiosks hinein aus […]
(6) […] en un primer momento llegó a exclamar en la oscuridad nocturna del quiosco […]

The contrastive study of Sensory Paths has heretofore received little scholarly attention. One of the reasons for this gap may be the lack of large parallel corpora available. In fact, the studies carried out to date (e.g. Slobin 2009; Doval/Lübke 2013; Cifuentes-Férez 2014) are based on small and non-parallel corpora. Thus, the present study aims to delve deeper into the differences observed both in the frequency and the complexity of Fictive Paths in typologically different languages, such as German and Spanish. Moreover, it discusses the suitability of the PaGeS corpus for a study of this kind and suggest some aspects that could facilitate the retrieval of occurrences of Sensory Paths.

References:
CLARA INÉS LÓPEZ RODRÍGUEZ
University of Granada, Spain

Combinatory potential of verbs in English and Spanish medical texts: finding ‘abnormal clinical findings’ in parallel corpora

In medical texts, the adequate selection of verbs facilitates the comprehension of conceptual relations between the most prototypical semantic categories in Medicine. This lexical choice poses a challenge for translators and technical writers alike. However, verbs have not generally received enough attention in Terminology and Specialized Translation despite the fact that the meaning of a verb constrains the semantic roles and categories of the arguments that can combine with it (Buendía 2013).

The aim of this paper is to explore corpus query language options to retrieve verbs that usually modify noun phrases belonging to the conceptual category “abnormal clinical findings” (SnoMed controlled vocabulary) in medical texts in English and Spanish. We analyse two parallel corpora in English and Spanish with the help of Sketch Engine corpus query system. The first corpus includes popular science texts published in Scientific American between 2009-2012, and their translation into Spanish issued by the Spanish journal Investigación y Ciencia (López Rodríguez 2006a), while the second consists of the pairing OPUS2 Spanish – OPUS2 English corpus, based on European Union institutional texts and available from Sketch Engine.

The methodology and corpus query language used are being relevant in the research and development project “Combinatory Lexis in Medicine: Cognition, Text and Context” (FFI2014-51899-R) since this approach highlights the combinatorial potential of frequent verbs and terms used in the medical domain. The results will feed VariMed, a medical vocabulary aimed at highlighting the phenomenon of terminological variation in Medicine (Tercedor, Prieto y López 2013, López Rodríguez 2016b).

References


Parallel corpora of new Catalan Civil Code: a terminology harmonization project

Parallel corpora are usually linked to translation projects, especially with the help of computer-assisted translation systems. So, over the past few years, multilingual parallel corpora are generated regularly by linguistic services. But, in other cases, parallel corpora should be compiled by means of alignment tools applied to pre-existed translated texts. One of these situations is the functional translation of laws enacted by a national parliament for foreign parliaments. The functional approach to translation suggests that the target text was intended for a different purpose from that of the source text (Nord, 2005: 5). This translation activity usually increases when lawmakers work on news subjects, as new rights, new families, new environment policies, etc.

The aim of this paper is to present a collaborative project between the Catalan Parliament, the Institut d’Estudis Catalans and the Institut de Lingüística Aplicada (IULA) of the Universitat Pompeu Fabra. The Parliament of Catalonia has enacted, from 2002, six books of Civil Code, which updates our separate traditional civil law created during the Middle Age and survived until now. All the laws enacted by Catalan Parliament are always translated to Spanish by means of a Machine Translation (MT) system. More and more, some foreign parliaments ask for English functional translations of legal texts. These Catalan-English translations are not systematic nor automatic nor made by the same translator.

The Catalan Parliament and the Institut d’Estudis Catalans were interested to get a good English version of Catalan Civil Code, which could be published and disseminated. So it became necessary to review and harmonize all target texts, which are made in different time and by different translators. In order to review these multilingual texts, our project included a phase of text parallelization, in pairs, by means of ALINEA (De Yzaguirre et al. 2000), an alignment system developed in the Institut de Lingüística Aplicada (IULA) that can achieve word-to-word results between Romance languages for legal corpora analysis. ALINEA must be updated with improvements for two aspects: technical processing and Catalan-English alignment.

The second phase of project consisted in identifying terminology and phraseology problems of translations: too much lexical variation, lack of semantic precision, unsuitable lexical combinations and ambiguity, among others. In fact, with parallelization texts, a beginning translation review project became the definitive terminology harmonization project.

In this paper, real technical features of ALINEA will be presented as well as several samples of Catalan-English and Catalan-Spanish aligned texts from Catalan Civil Code. Furthermore, we would like to demonstrate how word-to-word parallelization makes easier terminology analysis, even for creating terminology and specialized phraseology resources.
JAVIER MARTÍN ARISTA
University of La Rioja, Spain

The design of a parallel corpus of Old English on a lexical database

This paper presents the initial design of a parallel corpus Old English-English. Such a corpus can be implemented on a relational lexical database with Filemaker software. To begin with, the main historical corpora of Old English are reviewed in order to justify the relevance of the undertaking. Then, the methodological issue is tackled of designing a corpus in which the Old English part is complete but the English part requires the ad hoc translation of some texts. The main questions addressed in the discussion include text selection and representativeness, coding, segmentation by word and fragment, annotation, tagging and alignment. These aspects are discussed with respect to the Old English part of the parallel corpus. The questions of annotation and tagging are raised with respect to the information that should be provided as well as the sources and procedures for automatisation. In this respect, the conclusion is drawn that, at word level, the minimal information available from tagging should include lemma, lexical class, gloss and morphological analysis; while annotation at word level should comprise, at least, the secondary sources. Finally, the sources of such information can be found in the relational database of Old English The Grid (www.nerthusproject.com), which consists of a lemmatiser, a database of morphological and lexicological analysis, a fragment concordance, a word concordance, a reference list of strong verbs and an index of secondary sources.

BÁRBARA MARTÍNEZ VILINSKY
Jaume I University, Spain

The compilation of a comparable and parallel corpus of crime novels to study unique items in translation

In this presentation, the stages followed for the creation and analysis of a comparable and parallel corpus of crime novels written in Spanish and translated from English into Spanish are described. This corpus was created as part of the work carried out for completing my doctoral dissertation. The starting point of this work was Tirkkonen-Condit’s (2004) Unique Items Hypothesis, which suggests that elements which are specific to a certain language and lack a direct or easily identifiable triggering element in the source language tend to appear less frequently in texts translated to that language than in texts written originally in it. In order to operationalise Tirkkonen-Condit’s hypothesis we first had to find a set of elements which could be considered typical of Spanish and unique in the sense that they lack a prima facie equivalent in English. We selected a group of verbal periphrases, demonstratives and appreciative suffixes.

The next fundamental step was the compilation a comparable corpus of contemporary crime novels made up of texts written originally in Peninsular Spanish and translations from American English into the same variety of Spanish, which permitted us to compare the frequencies of use of these elements in translated and non-translated Spanish, in a specific genre, language variety and time frame. We complemented this with a parallel component of the Spanish novels’ source texts in English, which allowed us to extract parallel concordances in order to find out to what extent the source text influences the translator’s linguistic decisions.
The compilation of this corpus consisted of six different stages: text selection, digitization, aligning of each text and its translation, segmentation of the aligned texts, tagging and indexing of the final corpus with the aid of IMS Open Corpus Workbench. Searches on the corpus were performed through its query processor, CQP. Throughout the compilation process and analysis of this corpus we tried different software and we encountered several problems and challenges, information that may be of use to researchers who are beginning the process of creating their own corpora.

INMACULADA MAS ÁLVAREZ, XANDRA SANTOS PALMOU
University of Santiago de Compostela, University of Salamanca, Spain

Corpus paralelos como herramientas lexicográficas. Notas a partir de la revisión de diccionarios en línea

En el momento actual la conexión entre diccionario y corpus es ya un hecho, gracias al desarrollo de herramientas que permiten mostrar en un instante los vocablos de una lengua en uso. Contamos, pues, con recursos multilingües mixtos que ofrecen en paralelo las concordancias de una forma lingüística en dos lenguas determinadas, además de información lexicográfica más o menos amplia o rigurosa. Parece claro que estamos en un camino de fusión frutífera y más que rentable, en especial para el trabajo de traducción, pero reconozcamos también que se trata de una vía aún en proceso, necesitada de análisis, evaluación y reflexión.

Tanto para la investigación léxica multilingüe, como para sus diversas aplicaciones, la elaboración e implementación de corpus paralelos se revela una tarea imprescindible. Necesaria, pero no suficiente, pues por el momento todo proceso automatizado ve limitada su fiabilidad. Así pues, resulta también esencial considerar en qué medida un diccionario está determinado por los corpus en que se basa o más bien son los corpus los que se pone a disposición del diccionario. Por supuesto, el formato de presentación en pantalla lleva asociada una serie de consideraciones no exentas de interés, como es el caso de la visualización de los datos.

En esta contribución se parte de unos pocos lemas del español para explorar algunos diccionarios multilingües que ofrecen ejemplarios paralelos extraídos de corpus diversos. A partir de las correspondencias que ofrecen en alemán y en portugués se presenta un conjunto de consideraciones encaminadas a analizar, evaluar y reflexionar sobre corpus paralelos y lexicografía.

Diccionarios: bab.la Diccionario, dict.cc Online-Wörterbuch, linguatools Kontext-Wörterbuch, Linguee Wörterbuch, MyMemory Translated.net, Reverso Context

Referencias
Mikhail Mikhailov
University of Tampere, Finland

PEST – Parallel Electronic corpus of State Treaties

It is needless to say how important state treaties are for establishing and regulating relations between countries. Bi- and multilateral treaties are signed after long negotiations and many compromises. All language versions of a treaty are usually explicitly pronounced authentic and equally authoritative. However, from the linguist’s point of view, texts in different languages cannot be identical. It is interesting to study the kind of differences that can be found in different language versions of international treaties. This can be done by the quantitative analysis of data from parallel corpora.

PEST is a parallel corpus of state treaties, which is being compiled at the University of Tampere. The corpus will consist of several subcorpora, each representing a different pair of states: Russia vs. Finland, Finland vs. Sweden, Sweden vs. Russia. There will also be a subcorpus of international conventions in all three languages plus English. The work began in 2015 and will be completed in 2017.

PEST is a full-text corpus, however, the annexes of the documents are not included. It is a diachronic corpus with texts starting from 1917 (the year of the revolution in Russia and the independence of Finland) and reaching to the present day. Naturally, many of the treaties are no longer valid; a fact which is irrelevant from the point of view of linguistics, translation studies and language technologies.

All versions of the documents are being aligned at the sentence level and are morphologically and syntactically annotated. The aligning and annotation is performed with open-source software (LF Aligner, Omorfi, etc.). For querying the corpus, the research group uses its own online corpus search engine TextHammer, which is designed especially for working with parallel corpora. The search engine has many functions, e.g. parallel concordances, frequency lists, n-grams, collocations, etc., and more are to come.

Currently, the Russian-Finnish subcorpus of PEST is in the final compilation phase, with 220 pairs of texts, and the compilation of the Swedish-Finnish subcorpus has begun. A possible drawback is that the corpus is not going to be big: the size is limited by the actual number of existing treaties. The topics are also the same: borders, trade, fishing, environment, etc.

Preliminary studies utilizing the Russian-Finnish data reveal that there are more documents which are likely to be drafted in Russian and translated into Finnish than vice versa. The Russian documents are also more emotional than the Finnish documents. The use of verbal tenses is different, as well, e.g. the simple past tense is fairly untypical in the Finnish texts, and even if occurrences of the Finnish perfect are added to these, there are still twice as many past tenses in the Russian texts. This all means that different language versions of the same treaty may in fact present the subject from slightly different angles.
**Teresa Molés-Cases, Ulrike Oster**
Jaume I University, Spain

**Presentation and compilation of a section of the COVALT corpus (EN, FR, DE > ES)**

The aim of our contribution is to present a section of the Corpus Valenciana de Literatura Traduïda (COVALT), created by the research group Covalt, of the Department of Translation and Communication at the Universitat Jaume I. This group has been working since 2001 on the compilation of the multilingual COVALT corpus, a four-million-word corpus, still under construction, made up of narrative works originally written in English, French and German and their Catalan translations published in the autonomous community of Valencia between 1980 and the present (cf. Guzman 2013). Since the members of the Covalt group are interested in translation research, and more specifically in the investigation of translated Catalan and Spanish, this corpus has recently been extended to include translations into Spanish published in Spain (cf. Oster & Molés-Cases 2015).

The section that we will introduce in our presentation is the multilingual corpus translated into Spanish, which amounts to about 3.5 million words. Our intervention will focus on explaining the process of its compilation in detail: aligning source and target texts with Déjà Vu, tagging grammatical category and lemma with TreeTagger, indexing in IMS Corpus Workbench and searching with Corpus Query Processor (CQP Web) (cf. Molés-Cases 2016). Although the task of creating an aligned, tagged and indexed corpus is time-consuming and a long-term project, its main advantage is that it provides a great array of search possibilities, such as complex sequences. In our intervention we will illustrate some of the most interesting ones.

Nowadays, the Covalt group is also working on the compilation of two comparable corpora of Catalan and Spanish literary texts. This will enable us to compare original and translated Catalan and Spanish and to study translation universals from a twofold perspective: not only through parallel corpora, but also through comparisons with comparable corpora (cf. Marco under revision).

**References**


Marco (under revision): Connectives as indicators of explicitation in literary translation: A study based on a comparable and parallel corpus. Target.


**Ana Elvira Ojanguren López**
University of La Rioja, Spain

**Gathering information from a relational database to annotate and tag a parallel corpus of Old English**

The aim of this paper is to select the information available from a relational database that is relevant for the annotation and tagging of a parallel corpus Old English-English. The presentation focuses on the relational
database The Grid, which represents a development of the lexical database Nerthus (www.nerthusproject.com), with the important difference that Nerthus is a type database with approximately thirty thousand files whereas The Grid constitutes a token database with around three million files. From the qualitative point of view, Nerthus consists of one block of information whereas The Grid comprises six blocks of related information, including an unlemmatized part, a lemmatized part, a fragment concordance, a word concordance, a reference list of strong verbs and an index of secondary sources. The fields in these databases are described and assessed as to their relevance for the annotation and tagging of the Old English texts of a parallel corpus Old English-English. The presentation shows that, in The Grid, the data have been organized by means of an index and two concordances, while the interpretation is provided by the morphological and lexical database Nerthus, the list of strong verbs and the database of indexation of secondary sources Freya. The list of strong verbs is a guide to headword spelling choice, given that lexical derivation is based on strong verbs. The conclusion is reached that the most relevant information for the annotation of the Old English texts in the corpus can be retrieved from the database of secondary sources Freya, while the lexical database Nerthus and the lemmatizer Norna can be used for both annotation and tagging.

SARA ROMÁN GALDRÁN
Jaume I University, Spain

Diseño y compilación de un corpus para el análisis de la comprensibilidad de consentimientos informados

El presente artículo explora el diseño y la compilación de un corpus textual electrónico, como herramienta para estudiar la comprensibilidad de consentimientos informados (CI) de ensayos clínicos.

En primer lugar, se determinarán los criterios que debe cumplir nuestro corpus para ser representativo, de calidad y servir a su objetivo. A grandes rasgos, el corpus será comparable y paralelo, ya que estará compuesto por tres subcorpus formados, respectivamente, por CI originales en español, CI originales en inglés y las traducciones en español de estos últimos. Debido a las características de los subcorpus, será, además, un corpus bilingüe y especializado. Se utilizarán textos completos, en lugar de fragmentos, para evitar diferencias de representatividad entre diversas partes del documento. Por último, a fin de determinar el número de documentos que formarán parte del corpus, uno de los criterios de diseño más debatibles, se evaluará el uso de la aplicación informática ReCor, que estima el umbral mínimo de representatividad de un corpus en función de su tamaño. Una vez perfilado el diseño del corpus, nos centraremos en su compilación y en los desafíos que esto conlleva. Pondremos particular énfasis en la escasa disponibilidad de consentimientos informados de ensayos clínicos, frente a otros tipos de consentimientos informados, a los que se puede acceder fácilmente desde la página web de centros sanitarios y hospitales. Además, destacaremos la dificultad de acceder a documentos originales en inglés junto con sus traducciones para compilar corpus paralelos especializados.

Este artículo pretende abrir el diálogo sobre posibles opciones para superar estos retos, como el uso de fuentes institucionales y la posibilidad de solicitar personalmente consentimientos informados a centros patrocinadores de ensayos clínicos, en este caso particular.
Parallel corpora and corpus literary Translation Studies: exploring the translation of Charles Dickens’s novels into Spanish

Over the last two decades, the use of corpus material and methodologies “has been gaining momentum in the rapidly evolving field of Translation Studies [...] and has played a crucial role in the development of new lines of scientific inquiry for the study of translational texts” (Ji and Oakes 2012: 177-178). This new area of research has been labeled corpus translation studies (henceforth CTS) and has adopted and developed “the methodologies of corpus linguistics to analyze translation and translating for descriptive and applied purposes” (Laviosa 2013: 228). This increasing use of corpus material and corpus methodologies in translation studies has recently reached literary translation studies too. The conflation of CTS and emerging disciplines such as corpus stylistics has made possible the analysis of literary translations using corpus methods.

In this regard, this presentation is meant to contribute another example to this branch of CTS. More specifically, I will analyze speech verbs in Charles Dickens’s Hard Times and their translation into Spanish in four different versions with the help of an aligned parallel corpus. Apart from their basic function of introducing speech, these verbs can also contribute to characterization. The regular occurrence of a particular speech verb to report the direct speech of a particular character helps to create a fictional personality. Given the important role they may play, the rendering of such verbs in four Spanish versions of this novel is assessed. To do so, a corpus-based methodology has been employed. Specifically, the presentation will focus on the the creation of the aligned parallel corpus containing the English text and the four Spanish versions was carried out. Aspects such as the digitalisation of texts, copyright and fair use issues –texts were scanned and processed with an OCR (Optical Character Recognition) software– or the concordancing software used to retrieve speech verbs from the original novel will be touched upon. Finally, evidence will provided that none of the four translations entirely preserves the characterizing value of the verbs, which may affect the way readers form impressions of characters in their minds. The use of this corpus methodology is thus seen to contribute to the field of literary translation studies.

References
especially in those settings where researchers do not build their own corpora, but use those that can be searched online. To that end, in section 2 a succinct typology of online searchable corpora for the study of language-pair related or translation-related phenomena will be drawn. In section 3, the concept of secondary sample corpus will be put forward, which, in turn, will shed new light on the status of web-based parallel (and parallelized) corpora. Section 4 will be devoted to general methodological requirements and to requirements of the quantitative approach as the main — but not only — approach shaping the research work in the fields of Corpus-based Descriptive Translation Studies and Corpus-based Contrastive Linguistics; the focus will lie on issues of replicability, comparability, availability and usability. In section 5, some brief conclusions will be drawn.

HUGO SANJURJO-GONZÁLEZ, MARLÉN IZQUIERDO
University of León, University of the Basque Country, Spain

The ACTRES Parallel Corpus (P-ACTRES 2.0)

The ACTRES Parallel Corpus (P-ACTRES 2.0) is a bidirectional English-Spanish corpus developed by ACTRES research group. P-ACTRES 2.0 contains over 4 million words both directions. From original English texts to their Spanish translations, the former P-ACTRES 1.0 (Izquierdo, Hofland & Reigem, 2008), is about 2.5 million, and from original Spanish texts to their English translations, 1.5 million words. P-ACTRES includes 5 subcorpora: books-fiction, books-nonfiction, newspaper articles, magazine articles and miscellaneous.

P-ACTRES 2.0 allows users to carry out corpus-based linguistic and textual contrast as well as Translation Studies projects either independently or jointly. It has proved to be a useful tool for studies at both lexico-grammatical and rhetorical level.

Technically, P-ACTRES 2.0 was designed by Knut Holland (University of Bergen). It employs Corpus Workbench (CWB) (Stefan & Hardie, 2011) for managing and querying and Treetagger for POS annotation (Schmid, 1995). Textual pairs were aligned employing TCA2 (Hofland & Johansson, 1998) and subsequently checked manually. P-ACTRES 2.0 is an upgrade from P-ACTRES 1.0 developed by Hugo Sanjurjo-González (University of León) in collaboration with Knut Hofland to house new repositories. It has been designed using HTML5 (“HTML5”, 2014) technology as well as JavaScript and Perl script files in order to show results of corpus queries and related statistics.

Further work on the P-ACTRES 2.0 features new annotation layers, supporting of n-grams queries and more sophisticated statistics based on R programming language (“R Core Team”, 2013).

References


The compilation and annotation of a parallel and multilingual corpus for analysing literary translations from German into Basque/Spanish

From the 1980s on, considerable efforts have been made to create different types of Basque corpora. However, the number of parallel and multilingual corpora is not very large. In my case, in order to be able to systematically analyse the translation of German literary texts translated into Basque, it was necessary to create a corpus from the ground up. Since some of the texts that were included in the mentioned corpus have been translated indirectly (mostly through the Spanish translation), intermediary versions were included in the case of those indirect translations. This means that sometimes up to three texts had to be aligned simultaneously. A tool called TraceAligner, which has been developed within the TRALIMA/ITZULIK research group, was used to align the bitexts and the tritexts of the corpus. Thus, it is a tool that serves to create parallel and multilingual corpora and to conduct queries in it.

The aim of the present paper will be to describe the design and compilation of the corpus using TraceAligner 3.0. Then, it will be presented how the corpus has been lemmatised and annotated at part-of-speech level. Finally, as part of the translation analysis, the process of extracting potential multi-word expressions will be shown.

Corpus PaGeS: Un corpus paralelo bilingüe para el estudio de las colocaciones. Un estudio piloto en la didáctica de las lenguas extranjeras

En esta contribución se intenta ofrecer un análisis de las colocaciones de la lengua alemana y su aprendizaje por sujetos aprendientes de lengua española como L1. El estudio se centra en el manejo del PaGeS-Corpus (creado por el grupo de investigación SpatiALES - Spatial Relations German/Spanish - de la Universidad de Santiago de Compostela) dentro del aula de alemán como lengua extranjera.

El objetivo de esta investigación es doble: por un lado, el análisis cuantitativo y cualitativo de las colocaciones con datos arrojados de este corpus bilingüe y, por otro lado, la descripción de la lengua en virtud de las frecuencias de uso de las colocaciones con SP en N., A., D. y/o G. o SP en forma de Grupo Prep. en alemán y su correspondiente traducción al español.

A través de esta herramienta se pretende constatar lo siguiente: cuáles son los conocimientos generales de nuestros alumnos de alemán a la hora de manejar este corpus y qué conclusiones de aprendizaje pueden sacar en
virtud de su manejo, una descripción precisa de las colocaciones en un uso auténtico, cómo puede ser empleado, cuál puede ser el papel del corpus en la investigación de una interlengua a la hora del proceso de aprendizaje y/o adquisición de la LE y cómo nuestros alumnos de alemán como lengua extranjera pueden evitar ciertos errores lingüísticos.

Referencias bibliográficas:


PATRICIA SOTETO DIOS
University of Vigo, Spain

Translation-oriented annotation of a multimedia parallel corpus of subtitles

This paper presents a pilot project that involves annotating a parallel corpus of subtitles with translation-related information. The Veiga corpus of English-Galician film subtitling, which can be accessed at http://sli.uvigo.gal/CLUVI/vmm_en.html, already contains two levels of annotation: on the one hand, the omission, addition and reordering of translation units, and on the other hand the in-cue and out-cue times and line breaks in the subtitles. All of the above mentioned aspects are tagged according to the XML CLUVI specification for parallel corpora, and a new set of tags is now being defined to annotate certain elements that are particularly present in film dialogue, as well as the translation techniques used to transfer these and other elements from English into Galician. The aim of the project is to enrich the corpus with translation-related data so that users can easily search for potentially problematic issues in audiovisual translation and observe the specific techniques used by the translator to render them in the subtitles.

A first pilot experiment is now being carried out with one of the films. In particular, the items currently being considered for annotation are the following: linguistic variation (dialect, slang, taboo language), idioms, culture-specific references, named entities, and orality features such as paralinguistic elements, discourse markers, interjections, vocatives, false starts or repetitions. As for translation techniques, they include amplification, reduction, neutralization, generalization, particularization, transposition, variation, modulation, adaptation, and equivalence. Although a number of studies have already empirically investigated the translation of some of these issues in a variety of parallel subtitles corpora, with a special focus on the transfer of humor, cultural references, taboo language and linguistic variation, none of these corpora have yet been processed and made available to the public. In this respect, the Veiga’s new level of annotation could potentially enhance the value of the corpus, which would provide a freely searchable database that could be used for translation teaching and research.
IRENE SZUMLAWSKI MORODO
Complutense University of Madrid, Spain

Corpus PaGeS: un corpus bilingüe para el análisis de las relaciones espaciales en español y alemán. El caso de WOLLEN con complemento direccional

El uso de los verbos modales con un complemento direccional pero sin el infinitivo de un verbo de desplazamiento es un caso especial dentro de las diferencias, de sobra conocidas, entre las lenguas españolas y alemana en cuanto a la expresión de la direccionalidad de un movimiento. Expresiones alemanas como “Ich will nach Hause” o “Sie wollen doch sicher zu ihr” deben modificarse necesariamente al traducirlas al español. Partiendo del verbo modal wollen se quiere mostrar en esta comunicación cómo el Corpus PaGeS posibilita un análisis contrastivo, en el que se pueden estudiar estas estructuras tanto desde la perspectiva de la traducción de textos del alemán al español como de las versiones alemanas de textos españoles.

Bibliografía


<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almeida, Jose Joao</td>
<td><a href="mailto:JJ@di.uminho.pt">JJ@di.uminho.pt</a></td>
</tr>
<tr>
<td>Alonso-Ramos, Margarita</td>
<td><a href="mailto:lxalonso@udc.es">lxalonso@udc.es</a></td>
</tr>
<tr>
<td>Althoff, Roswitha</td>
<td><a href="mailto:roswithaalthoff@edu.xunta.es">roswithaalthoff@edu.xunta.es</a></td>
</tr>
<tr>
<td>Amido Lozano, María Teresa</td>
<td><a href="mailto:t.amido@mx.uni-saarland.de">t.amido@mx.uni-saarland.de</a></td>
</tr>
<tr>
<td>Augustinus, Liesbeth</td>
<td><a href="mailto:liesbeth.augustinus@kuleuven.be">liesbeth.augustinus@kuleuven.be</a></td>
</tr>
<tr>
<td>Balla, Peter</td>
<td><a href="mailto:balla.peter9@upcmail.hu">balla.peter9@upcmail.hu</a></td>
</tr>
<tr>
<td>Balla, Peterne</td>
<td><a href="mailto:balla.peter9@upcmail.hu">balla.peter9@upcmail.hu</a></td>
</tr>
<tr>
<td>Bernardini, Silvia</td>
<td><a href="mailto:silvia.bernardini@unibo.it">silvia.bernardini@unibo.it</a></td>
</tr>
<tr>
<td>Brand, Christiane</td>
<td><a href="mailto:christiane.brand@th-koeln.de">christiane.brand@th-koeln.de</a></td>
</tr>
<tr>
<td>Brumme, Jenny</td>
<td><a href="mailto:jenny.brumme@upf.edu">jenny.brumme@upf.edu</a></td>
</tr>
<tr>
<td>Cao, Shuyuan</td>
<td><a href="mailto:shuyuan.cao@hotmail.com">shuyuan.cao@hotmail.com</a></td>
</tr>
<tr>
<td>Čermák, Petr</td>
<td><a href="mailto:Petr.Cermak@ff.cuni.cz">Petr.Cermak@ff.cuni.cz</a></td>
</tr>
<tr>
<td>Cerutti, Giorgina</td>
<td><a href="mailto:Giorgina.Cerutti@unige.ch">Giorgina.Cerutti@unige.ch</a></td>
</tr>
<tr>
<td>Catacata, Chris</td>
<td><a href="mailto:chris.e.catacata@gmail.com">chris.e.catacata@gmail.com</a></td>
</tr>
<tr>
<td>Doval, Irene</td>
<td><a href="mailto:i.doval@usc.es">i.doval@usc.es</a></td>
</tr>
<tr>
<td>Fernández Lanza, Santiago</td>
<td><a href="mailto:santiago.fernandez@unir.net">santiago.fernandez@unir.net</a></td>
</tr>
<tr>
<td>Ferraresi, Adriano</td>
<td><a href="mailto:adriano.ferraresi@unibo.it">adriano.ferraresi@unibo.it</a></td>
</tr>
<tr>
<td>Gamallo, Pablo</td>
<td><a href="mailto:pablo.gamallo@usc.es">pablo.gamallo@usc.es</a></td>
</tr>
<tr>
<td>García Aragón, Álvaro</td>
<td><a href="mailto:alvaro.garcia@universidadeuropa.eu">alvaro.garcia@universidadeuropa.eu</a></td>
</tr>
<tr>
<td>García Salido, Marcos</td>
<td><a href="mailto:marcos.garcias@udc.es">marcos.garcias@udc.es</a></td>
</tr>
<tr>
<td>Garcia, Marcos</td>
<td><a href="mailto:marcos.garcia.gonzalez@udc.gal">marcos.garcia.gonzalez@udc.gal</a></td>
</tr>
<tr>
<td>Ghoshal, Parijat</td>
<td><a href="mailto:parijat.ghoshal@uzh.ch">parijat.ghoshal@uzh.ch</a></td>
</tr>
<tr>
<td>Gómez Guinovart, Xavier</td>
<td><a href="mailto:xgg@uvigo.es">xgg@uvigo.es</a></td>
</tr>
<tr>
<td>González Pérez, Aarón</td>
<td><a href="mailto:aarongonzalez274@gmail.com">aarongonzalez274@gmail.com</a></td>
</tr>
<tr>
<td>Hareide, Lidun</td>
<td><a href="mailto:Lidun.hareide@moreforsk.no">Lidun.hareide@moreforsk.no</a></td>
</tr>
<tr>
<td>Hourani Martín, Dunia</td>
<td><a href="mailto:dunia.hourani_martin@uni-leipzig.de">dunia.hourani_martin@uni-leipzig.de</a></td>
</tr>
<tr>
<td>Izquierdo, Marlén</td>
<td><a href="mailto:marlen.izquierdo@ehu.eus">marlen.izquierdo@ehu.eus</a></td>
</tr>
<tr>
<td>Jiménez Juliá, Tomás</td>
<td><a href="mailto:tomaseduardo.jimenez@usc.es">tomaseduardo.jimenez@usc.es</a></td>
</tr>
<tr>
<td>Kaiser, Katharina</td>
<td><a href="mailto:katharina.kaiser@uni-konstanz.de">katharina.kaiser@uni-konstanz.de</a></td>
</tr>
<tr>
<td>Lavid-López, Julia</td>
<td><a href="mailto:julavid@filol.ucm.es">julavid@filol.ucm.es</a></td>
</tr>
<tr>
<td>Name</td>
<td>Email</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Liste Lamas, Elsa</td>
<td><a href="mailto:elsa.liste@usc.es">elsa.liste@usc.es</a></td>
</tr>
<tr>
<td>Lopez Rodríguez, Clara Ines</td>
<td><a href="mailto:clarailr@ugr.es">clarailr@ugr.es</a></td>
</tr>
<tr>
<td>Lorente Casafont, Mercè</td>
<td><a href="mailto:merce.lorente@upf.edu">merce.lorente@upf.edu</a></td>
</tr>
<tr>
<td>Lübke, Barbara</td>
<td><a href="mailto:barbara.lubke@usc.es">barbara.lubke@usc.es</a></td>
</tr>
<tr>
<td>Marco, Josep</td>
<td><a href="mailto:jmarco@trad.uji.es">jmarco@trad.uji.es</a></td>
</tr>
<tr>
<td>Martín Arista, Javier</td>
<td><a href="mailto:javier.martin@unirioja.es">javier.martin@unirioja.es</a></td>
</tr>
<tr>
<td>Martínez Vilinsky, Bárbara</td>
<td><a href="mailto:llenas@uji.es">llenas@uji.es</a></td>
</tr>
<tr>
<td>Mårzhäuser, Christina</td>
<td><a href="mailto:Christina.Maerzhaeuser@romanistik.uni-muenchen.de">Christina.Maerzhaeuser@romanistik.uni-muenchen.de</a></td>
</tr>
<tr>
<td>Mas Álvarez, Inmaculada</td>
<td><a href="mailto:inmaculada.mas@usc.es">inmaculada.mas@usc.es</a></td>
</tr>
<tr>
<td>Mikhailov, Mikhail</td>
<td><a href="mailto:mikhail.mikhailov@staff.uta.fi">mikhail.mikhailov@staff.uta.fi</a></td>
</tr>
<tr>
<td>Molés-Cases, Teresa</td>
<td><a href="mailto:tmoles@uji.es">tmoles@uji.es</a></td>
</tr>
<tr>
<td>Morys, Olha</td>
<td><a href="mailto:morisolga4@gmail.com">morisolga4@gmail.com</a></td>
</tr>
<tr>
<td>Müller Alonso, Medea</td>
<td><a href="mailto:medea_mma@hotmail.com">medea_mma@hotmail.com</a></td>
</tr>
<tr>
<td>Ojanguren López, Ana Elvira</td>
<td><a href="mailto:anojangu@unirioja.es">anojangu@unirioja.es</a></td>
</tr>
<tr>
<td>Oster, Ulrike</td>
<td><a href="mailto:oster@uji.es">oster@uji.es</a></td>
</tr>
<tr>
<td>Rabadán Álvarez, Rosa</td>
<td><a href="mailto:rosa.rabadan@unileon.es">rosa.rabadan@unileon.es</a></td>
</tr>
<tr>
<td>Rao, Xi</td>
<td><a href="mailto:xi.rao@uzh.ch">xi.rao@uzh.ch</a></td>
</tr>
<tr>
<td>Román Galdrán, Sara</td>
<td><a href="mailto:al352289@uji.es">al352289@uji.es</a></td>
</tr>
<tr>
<td>Ruano San Segundo, Pablo</td>
<td><a href="mailto:pauruano@unex.es">pauruano@unex.es</a></td>
</tr>
<tr>
<td>Sánchez Nieto, María Teresa</td>
<td><a href="mailto:maysn@lia.uva.es">maysn@lia.uva.es</a></td>
</tr>
<tr>
<td>Sanjurjo-González, Hugo</td>
<td><a href="mailto:hsang@unileon.es">hsang@unileon.es</a></td>
</tr>
<tr>
<td>Santos Palmou, Xandra</td>
<td><a href="mailto:xandrasantospalmou@gmail.com">xandrasantospalmou@gmail.com</a></td>
</tr>
<tr>
<td>Sanz Villar, Zuriñe</td>
<td><a href="mailto:zurine.sanz@ehu.eus">zurine.sanz@ehu.eus</a></td>
</tr>
<tr>
<td>Soliño Pazó, María Mar</td>
<td><a href="mailto:solino@usal.es">solino@usal.es</a></td>
</tr>
<tr>
<td>Sotelo Dios, Patricia</td>
<td><a href="mailto:psotelod@uvigo.es">psotelod@uvigo.es</a></td>
</tr>
<tr>
<td>Szumlakowski Morodo, Irene</td>
<td><a href="mailto:ireneszm@ucm.es">ireneszm@ucm.es</a></td>
</tr>
<tr>
<td>Tabares Plasencia, Encarnación</td>
<td><a href="mailto:tabares@uni-leipzig.de">tabares@uni-leipzig.de</a></td>
</tr>
<tr>
<td>Volk, Martin</td>
<td><a href="mailto:volk@cl.uzh.ch">volk@cl.uzh.ch</a></td>
</tr>
<tr>
<td>Yzaguirre Maura, Lluís de</td>
<td><a href="mailto:lluis.deyzaguirre@upf.edu">lluis.deyzaguirre@upf.edu</a></td>
</tr>
<tr>
<td>Zavala-Rojas, Diana</td>
<td><a href="mailto:diana.zavala@upf.edu">diana.zavala@upf.edu</a></td>
</tr>
</tbody>
</table>
Scientific Committee
José João Almeida, Univ. do Minho
Petr Cermák, Charles Univ.
Teresa Fanego Lema, USC
Pablo Gamallo, USC
Nader Haghani, Univ. of Tehran
Wang Lixun, Univ. of Hong Kong
Rosa Rabadán, Univ. of Leon
Guillermo Rojo, USC
Alberto Simões, Univ. do Minho
Martin Volk, Univ. of Zürich

Organizing Committee
Irene Doval, USC (chair)
Mª Teresa Sánchez Nieto, Univ. Valladolid (secretary)
Santiago Fernández Lanza, UNIR
Tomás Jiménez Juliá, USC
Elsa Liste Lamas, USC
Barbara Lübke, USC

Organization
Research Group SpatIAEs (I. Doval)
Dept. of English and German Philology
Faculty of Philology, USC
Web: www.usc.es/congresos/pacor
Email: pacor2016@usc.es