

Asociación Española de Lingüística de Corpus



aelinco



Boletín Electrónico AELINCO

Núm 7. Año 2020

[www.aelinco.es](http://www.aelinco.es)



# aelinco

Boletín Electrónico AELINCO

- 3 Mensaje del Presidente de AELINCO
- 4 Interview with Professor Tony McEnery
- 8 Corpus Linguistics, News Media and Social Good
- 10 Junta Directiva de AELINCO y Directores de Paneles
- 12 Memoria del XI Congreso Internacional de Lingüística de Corpus. CILC2019 València
- 15 XII Congreso Internacional de Lingüística del Corpus. CILC2020 Bèrgamo
- 17 Avance del XIII Congreso CILC2021 Oviedo
- 18 Revista Research in Corpus Linguistic (RiCL)
- 20 Rincón del Socio

Edita: Asociación Española de Lingüística del Corpus (AELINCO)

Director del Boletín: Miguel Fuster Márquez

Información de contacto:

Departamento de Filología Inglesa y Alemana

Facultad de Filología, Traducción y Comunicación

Avda Blasco Ibáñez 32

46010 Universitat de València

Tel: -34 963864262

info@aelinco.es

www.aelinco.es

© Asociación Española de Lingüística del Corpus, 2013

ISSN: 2340-7735

# Mensaje del Presidente de Aelinco

---

## ***Estimados socios y socias:***

Un año más, me corresponde dirigirme a vosotros y vosotras, en nombre de la Junta Directiva de la Asociación Española de Lingüística de Corpus, para extenderos un cordial saludo.

El congreso de València, que fue organizado con esmero y dedicación, discurrió con toda normalidad en un marco universitario absolutamente exquisito. Se renueva el éxito alcanzado por ediciones anteriores de los congresos de AELINCO. Estamos seguros de que el próximo congreso en Bérgamo, ya más cercano, será igualmente muy exitoso.

Todos los socios y socias somos ya conscientes de que AELINCO se ha convertido en una asociación madura no solo en cuanto a su edad (llegamos al congreso número doce en 2020 en Bérgamo, Italia) sino, muy especialmente, en cuanto a su calidad científica, fruto del compromiso de sus socios con la investigación y el desarrollo. Desde la Junta Directiva queremos agradecer a quienes hacéis que los congresos de AELINCO sean posibles –es decir, organizadores, directores y directoras de panel, ponentes y asistentes– por vuestro compromiso con AELINCO y deseamos animaros a que continuéis con vuestros esfuerzos por hacer que AELINCO siga creciendo y mejorando.

Creo sinceramente que me corresponde también, en estos momentos, lanzar la mirada hacia atrás y hacer patente mi reconocimiento al trabajo de presidentes anteriores con sus Juntas Directivas y directores de panel. Así, muy especialmente quiero destacar el trabajo brillante de Marisa Carrió como Tesorera. Durante sus respectivos mandatos, ha sido ejemplo de seriedad, responsabilidad y compromiso a la hora de conducir AELINCO. También se hace necesario reconocer el sólido trabajo de los editores de *Research in Corpus Linguistics (RiCL)* junto con sus equipos.

Finalmente, como Junta Directiva quedamos abiertos a cuantas iniciativas queráis trasladarnos con el fin de que AELINCO siga siendo una asociación dinámica y robusta.

Como siempre, quedo a vuestra disposición para lo que estiméis. Sin otro particular, aprovecho para saludaros muy cordialmente.

**Pascual Cantos Gómez**  
**Presidente de AELINCO**

# Interview with Professor **Tony McEnery** Lancaster University

Interviewed by: Alfonso Sánchez-Moya (RCC Postdoctoral Fellow at Harvard University)

---

***As you know, CILC 2019 has paid particular attention to exploring the connections between discourse analysis and corpus linguistics. Considering the fast and prolific development of linguistic research that integrates both discourse analysis and corpus linguistics, do you think it is possible to go back to times where discourse analysis was not that influenced by corpus linguistics research?***

Lots of researchers do still take non corpus-based approaches to discourse analysis, and I would even say it is appropriate to do so on occasion. If you are making general claims, of course you need general support. If your interest, on the other hand, is on a small set of documents, and you wish to look at that small set of documents, I would always question the wisdom of applying technology to that study, when

in fact you can do a detailed analysis of a relatively small document set using human expertise. So, for instance, if somebody is interested in just the

writings of a certain politician and a human expert analysis can read and comprehend that document set, they should get on and do it. It may be still the case that they find some concordancing technology helpful to allow for a quick scan, accurate word counts, etc; but I still believe that the human analyst is pretty good. In fact, machines are deficient in many respects on occasion, and human expert analysis needs to intersect with corpus analysis in order (a) to make sense of it, (b) to direct it, and also, importantly, to give it some justification, on 'why is this an interesting thing to look at'. Therefore, I am very permissive of people taking a range of

approaches as long as they think of the appropriacy of what they are doing. If they want to make large scale claims about large text collections, sitting down and trying to read it is a tremendously bad approach to it if they want to do something else in their life time; and harnessing computational technology can be very helpful to their purpose.

***Both in your plenary talk for CILC 2019 and in your most recent published articles you clearly demonstrate how corpus linguistics practice can be effectively used to investigate areas that may not seem related to linguistics per se, such as hydrology or the representation of key social actors such as homosexual men or prostitutes in the 17th century. In your opinion, how does corpus linguistics contribute to a better understanding of social issues of this sort in today's society?***

***“Linguistics is almost like a meta-discipline. Language intersects with, and erupts into, a whole range of human experience.”***

I should preface my comment by saying that I have a very broad conception of linguistics. Some colleagues don't. And they

are quite entitled to their view and I hope they think I am entitled to mine. For me, linguistics is almost like a meta-discipline. Language intersects with, and erupts into, a whole range of human experience. If you talk to sociologists, or look at the sociologist literature, you will see from time to time something that looks like discourse analysis. The same is true for psychology, because of the cognitive connections between the brain and language. Also, historians analyse texts routinely. In fact, much of a historian's job is textual analysis. The same applies to literature scholars. Now, my view is that linguistics is distinct from these disciplines in that

our principal focus is on the study of language. But, at the same time, I appreciate that I need to stretch out across the range of disciplines where my focus on language could be of assistance.

Let me provide an example. If I see a sociologist, a historian or even a hydrologist wanting to do something that looks like discourse analysis or text linguistics, but I see them using techniques that linguists would no longer use for the type of question they are investigating, then I think it is beholden to me to actually try to help them. For example, looking at a billion words of textual material in their archives when they are available electronically is something we can help them with. We can have techniques, we have frameworks and theories that they may want to draw upon when looking at that. However, I am not a historian, I am not a sociologist, I am not a psychologist, and I am not a literature researcher. However, I will intersect with their interests because their interests are textually-focused very often. In fact, I have worked very closely with historians and we have managed to create very interesting collaborations, since I can bring some apparatus from linguistics or the corpus-based approach to the study of these texts. I can also bring something like social actor theory to the study of those texts. Historians can bring their expertise in, for example, their understanding of the established historiography. This is how you get really interesting interactions. Research meetings of this sort are also very interesting: stitching both approaches together, reflecting on the method and also synthesising the findings and getting deeper insights from both approaches.

***You have also done applied work with a range of organisations in the public sector (such as the Environment Agency and the Home Office). Likewise, you have taken corpus linguistics to the private sector (British Telecom, IBM, Nokia). Based on these experiences, how can knowledge and practice in corpus linguistics be transferred to contexts outside academia?***

I guess the answer here is not so different to the example I gave with historians. Again, my principal focus is upon language. From time to time, language erupts into a whole range of areas. For example, I gave a talk at the Home Office in the UK on a particular topic. In the UK, each large Government Department has an academic attached to them, called the Chief Scientific Advisor. So the Chief Scientific Advisor of the Home Office came along and before I talked, turned to the audience and said “he is now going to talk to us about language”. Interestingly, almost everything that we look at in the Home Office at some point depends on an understanding of language, be it a written evidence, hate speech or whatever. The Home Office is indeed very interested in language on occasion. So that was a great opening for me because I had somebody there who understood that they were looking at say, for example, crime. But crime as intersecting with language evidence or even the commission of linguistic offence, hate speech, etc.

So, again, it is very similar to what I mentioned before. If somebody comes to me with a question that is principally rooted in language, I will comment on that. If they ask me anything about sentencing policy I’d say: ‘I am sorry I can’t really at the moment see the linguistic dimension of this, but if you could try to explain it to me; then I will look at it’. Our relevance is that we study something which manifests itself over and over again in a wide range of policy areas or in a wide range of products and you carry on like that, because of course it is the principal means by which humans communicate with one another.

***Based on your pioneering and central role in the field of corpus linguistics, have you perceived a pattern or an evolution in research in corpus linguistics over the last decade?***

At different levels, yes. At the conceptual level, I think some arguments that were active earlier have been pretty much put to bed. In fact, interestingly at

this conference, some of the references in the papers came to the same conclusion. For example, the wicked nature of annotation (or not) was quite a lively debate for a long time. I think that debate was useful to have in some ways, but also relatively non-sensical because I think I always make the point that just because we are not explicit about our linguistic analysis doesn't mean to say that they happen from the point of view of the scientific method or whether that's hiding our analysis by not annotating. As far as I am concerned, it was always good scientific practice to clearly show your annotations and share them with people (even if they throw them away!). But also, crucially, they can disagree with them. If I didn't share them with them, how could they disagree with them? Therefore, for me, that debate was always a question of good scientific practice rather a question of utility. I don't think anybody now is rallying against annotation, so that's one debate that has been put to bed.

Similarly, the corpus-based / corpus-driven argument had an extreme formulation, which I think people have sensibly moved away from. If you look at the writings of Karl Popper, he says that it is impossible to do science without conjecture. Science starts with conjecture. And conjectures are essentially introspections of some sort. That is to say, you normally gather data because you have a hypothesis you would like to explore, so the idea that the data describes itself is sort of extreme fantasy, I would say. And there would always be at least some element of human intervention in the selection of the questions, the selection of the data, in the formation of hypotheses. I think that view has pretty much prevailed. I don't think that there are people arguing for this extreme version of corpus-driven. Although I suppose that some people looking at machine learning may say 'that's corpus driven!'. Well, no! You are still choosing to expose the data to the algorithm so it is your own choice of data. These

algorithms are not wondering around at all. Even if it is a human-manufactured algorithm, the conjectures about language are therefore actually encapsulated within it. You can hide conjecture away inside an algorithm. But I would say that it is almost impossible to escape conjecture.

***In your view, which are the main opportunities and the most important challenges that both corpus linguistics and corpus linguists are likely to face in the near future?***

I will start with opportunities just to be optimistic. The principal opportunity is actually helping other researchers whose principal focus is not the study of language, such as sociologists or public policy makers, to embrace and use what we would do when we approach language. As a consequence of that, we have the opportunity to make a difference in the 'real' world. If we can, for instance, have a look at language around migration, then we can have an impact. I remember when we published our work on refugees and asylum seekers and how negatively they were represented in the UK press. This work had an impact, since people started to object to these uses of language around migrants. This even had an impact on how the BBC started to treat these words, and when David Cameron used some of these negative expressions, he was attacked in the media for using some of these dehumanising

metaphors and apologised and changed his position. Therefore, we can have an impact if we find the right people to work with. So that's another example

of how we should always talk to researchers in other areas, since joint efforts are more likely to trigger change in public discourse. This is for me an opportunity.

Now, challenges. I think one of the biggest challenges that corpus linguistics faces at the moment (or actually just linguistics) is a belief that large-scale computational approaches (or AI approaches) to the study of language can just solve things. Just

give answers. And you can push the linguist on a side. This is a challenge in a whole range of ways. The first challenge is that you get nuclear-headed research done by people who have never really met linguists or discourse analysts. This ends up with ignorance as the basis of discovery. And the great thing about ignorance is that everything is new. But we are researchers, and we shouldn't be making discoveries based on profound ignorance. So that's one level of the challenge: the fact that they have the symbolic capital or the faced credibility to actually use ignorance as the basis of discovery.

Of course, there is another level to it, which I have already alluded to. This happens when people don't sufficiently investigate their findings or may not consider what linguists may be able to offer to them and come up with apparently quite sophisticated approaches to discourse, like topic modelling, which appear to give wonderful results and appear to be relatively objective, applicable to a whole range of purposes. But when linguists come along and really start to try and use these things, they fall apart in their hands. Especially when they are exposed to lose the critical reasoning about language that linguists would bring along. There is where the challenge lies, not in people generating the packages that do these analyses, it is in making the voice heard. There's a wonderful computer scientist who has done this work, has all the symbolic capital, the fancy mathematics, etc. but then you come along and say, actually, the discovery made here is very opaque and subjective, they are randomly selecting some topics, they are imposing analyses without reading them. This is a Bayesian approach which would actually give a slightly different response each time you run it... Then, it is getting that sort of critical perspective that linguists can bring to the use of these packages against these

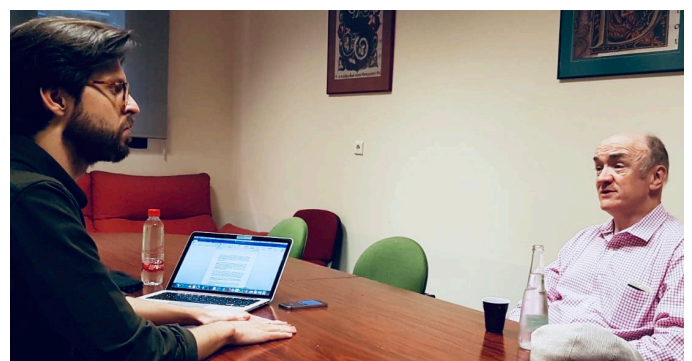
*“ The first challenge is that you get nuclear-headed research done by people who have never really met linguists or discourse analysts. This ends up with ignorance as the basis of discovery. And the great thing about ignorance is that everything is new. But we are researchers, and we shouldn't be making discoveries based on profound ignorance. ”*

highly credible actors with lots of symbolic capital that becomes the challenge. You know: it looks great, they've got big data, they can do all sorts of maths, they give apparently very clear answers, and you come along spoiling the party saying “what about the discovery procedure?”. It can be hard to have your voice heard in that context.

But it is actually even more important that we do have our voice heard in that context, even though it can be very tough. But well, I have had a career of very

tough. When I started out, nobody wanted to hear about corpora. It wasn't the right way of doing linguistics. The right way of doing linguistics was 'this way': close reading of individual texts, application of some theories to your analysis

or introspection, thinking about your sentences, working with exotic logics in order to work out the structure of sentences or the way in which meaning worked, etc. But then, people started saying that maybe looking at usage data is the way to go here and maybe frequencies are quite important. It is difficult to explain to people now just how unpopular that view was. Unpopular to the extent that it wasn't opposed, but simply ignored. That's how unpopular it was, you didn't even have to argue against it, you could simply ignore it and making your voice heard in that context was actually quite difficult.



# News media and social good

Maite Taboada ([mtaboada@sfu.ca](mailto:mtaboada@sfu.ca)). Simon Fraser University

---

Much of the research in corpus linguistics can be characterized as work that promotes social good or shines a light onto the lack thereof. This is particularly salient for researchers that work on news media, news values and the portrayal of certain social groups in media. Examples that come to mind in this area include: the GENTEXT research group at the University of Valencia on intimate partner violence in the news; the CASS Centre at Lancaster studying representations of immigrants, trans people and other under- and misrepresented groups; and basically all of the research in Critical Discourse Analysis.

I propose that we can label this type of corpus linguistics inquiry as research for good, part of the broad ‘data for good’ movement, in which researchers across a range of disciplines strive to contribute to a more equal society, more participatory democracy and a more accurate understanding of the problems facing our communities. Since accurate understanding and characterization of problems starts with how we conceive of and talk about those problems, linguists have a very important role to play in these endeavours.

My own research has evolved in this direction, in projects that combine analyses of news, misinformation, online news comments and gender representation in what I hope is a small contribution towards better discussions about the news and better gender representation in the news. In the Discourse Processing Lab, we are currently involved in three projects that touch upon those aspects:

- A study of online news comments, with the goal of building a moderation platform.
- Methods to identify whether a news article is an instance of misinformation (so-called ‘fake news’).
- The Gender Gap Tracker, which monitors the proportion of women and men quoted in news stories in mainstream Canadian media.

## Online news comments

News comments, those posted at the bottom of a news article online, can provide a venue for open discussion of ideas and engagement with other points of view. They are also, sadly, places of toxicity, abuse and the propagation of extreme ideologies.

Our project on online news comments seeks to automatically identify constructive comments, that is, comments that contribute to a dialogue about the article, providing opinion, but supporting it with evidence or reasons for that opinion. This is in contrast to toxic comments, those that are abusive and discourage an open exchange of ideas.

In order to better understand the nature of comments, we have built a corpus of opinion articles and their respective comments. The SFU Opinion and Comments Corpus (SOCC) contains all the opinion pieces published in the Canadian newspaper *The Globe and Mail* in the five-year period between 2012 and 2016, a total of 10,339 articles and 663,173 comments. Our current focus is the study of constructiveness and evaluation in the comments. To that end, we have annotated a subset of the large corpus (1,043 comments) with four layers of annotations: constructiveness, toxicity, negation and Appraisal. The corpus is freely available for non-commercial use. We have also built a prototype of an automatic moderation platform, which identifies the most constructive comments, to highlight or promote them, thus ‘drowning’ out the toxic and abusive.

The corpus can also be used to study many other aspects of journalistic discourse and online comments in general. I also think it can be used to study evaluative language, comparing the evaluative language found in the opinion articles and in the comments.

- Full description of the corpus and structure:  
<https://github.com/sfu-discourse-lab/SOCC>
- Direct link to the data:  
<https://researchdata.sfu.ca/islandora/object/islandora:9109>
- Moderation demo:  
<http://moderation.research.sfu.ca/>
- Paper: Kolhatkar, V.,H. Wu, L. Cavasso, E. Francis, K. Shukla and M. Taboada (to appear) [The SFU Opinion and Comments Corpus: A corpus for the analysis of online news comments](#). *Corpus Pragmatics*.

### Detecting fake news and misinformation

There is growing interest in the problem of misinformation, especially in online sources. This is an active area of research, with efforts aimed at identifying 'fake news' and curtailing their spread, often by tracking the sources of the articles and the people that distribute them, whether unwittingly or not. From the point of view of discourse analysis and corpus linguistics, an interesting question is whether the language of fake news is different than the language of fact-based news.

This project uses machine learning methods to distinguish fake from real. Machine learning methods learn to classify text by using features found in large corpora. The most important challenge we have faced so far is to find true fake news, so to speak. As language researchers, it is not our role to determine what is fake. Thus, we have turned to fact-checking websites and downloaded articles from those with labels assigned by experts.

We have published a paper on the corpus that we have collected with this method and started work on the classifier. A preliminary version of the classification system is available as a demo.

- Corpus: MisInfoText, a dataset of fact-checked news articles,  
<https://github.com/sfu-discourse-lab/MisInfoText>

- Fake news classification demo:  
<http://fakenews.research.sfu.ca/>
- Paper: Asr, F. T. and M. Taboada (2019) [Big data and quality data for fake news and misinformation detection](#). *Big Data & Society*. January-June 2019: 1-14.

### The Gender Gap Tracker

The [Gender Gap Tracker](#), developed within our lab and in collaboration with the non-profit organization [Informed Opinions](#), monitors the proportion of women and men quoted in news stories in mainstream English-language Canadian media. The dashboard (<https://gendergaptracker.informedopinions.org/>) has been live since October 2018 and is updated with new articles and statistics every 24 hours.

Our goal in this project is to bring awareness of the skewed proportions of men and women quoted. In the past months, the Tracker has stubbornly stayed at around 27% women quoted, clearly far from representing the 50% of women in the world. The causes are many, starting with the disproportionate number of men in positions of power, which are then, naturally, quoted and featured in news articles. Changing such numbers requires profound social change. Informed Opinions aims at specific change in one area: the proportion of experts quoted. While journalists do not have a choice in terms of gender when they need to quote the Prime Minister, they certainly have a choice when they approach an expert for an informed opinion on what the Prime Minister said. Informed Opinions is showing the statistics to journalists and newsrooms across the country, at the same time training female experts, especially at universities, to engage with the media. The statistics that the Gender Gap Tracker provides are a powerful motivator, for journalists, for producers and for female experts.

The projects I have described are large-scale and make use of big data. Of equal importance are qualitative analyses, studies employing corpus-assisted discourse analysis and studies with a range of data sizes and methodologies. I believe we can all make a contribution in corpus linguistics for social good.

# Junta Directiva de Aelinco y Directores de Paneles

---

La Junta Directiva de AELINCO la componen un Presidente, un Vicepresidente, un Secretario, un Tesorero y un Vocal:

## **PRESIDENTE**

*Pascual Cantos Gómez*

Universidad de Murcia

Facultad de Letras

30071 Murcia

Tel.: +34 868 88 4365

e-mail: [pcantos@um.es](mailto:pcantos@um.es)

## **VICEPRESIDENTE Y VOCAL**

*Miguel Fuster Márquez*

Universitat de València

Facultat de Filologia,

Traducció i Comunicació

46010 València

Tel.: +34 96 398 34 23

e-mail: [miguel.fuster@uv.es](mailto:miguel.fuster@uv.es)

## **SECRETARIO**

*Antonio Moreno Ortiz*

Universidad de Málaga

Facultad de Filosofía y Letras - 29071 Málaga

Tel.: +34 95 213 6670 | e-mail: [amo@uma.es](mailto:amo@uma.es)

## **TESORERA**

*Natalia Judith Laso Martín*

Universitat de Barcelona - 08007 Barcelona

e-mail: [njlaso@ub.edu](mailto:njlaso@ub.edu)

## **VOCAL**

*Giovanni Garofalo*

Università degli Studi di Bergamo

Piazza Rosate 2, Ufficio 208, 24129 Bérgamo (Italia)

e-mail: [giovanni.garofalo@unibg.it](mailto:giovanni.garofalo@unibg.it)

La Asociación Española de Lingüística de Corpus cuenta con un amplio abanico de especialidades, organizado en paneles. Cada uno de los paneles está dirigido por un socio, responsable de la evaluación y selección de las propuestas remitidas a los congresos de AELINCO. Los paneles con que cuenta actualmente AELINCO son nueve:

1. Diseño, elaboración y tipología de corpus
2. Discurso, análisis literario y corpus
3. Estudios gramaticales basados en corpus
4. Lexicología y lexicografía basadas en corpus
5. Corpus, estudios contrastivos y traducción
6. Variación y cambio lingüístico basados en corpus
7. Lingüística computacional basada en corpus
8. Corpus, adquisición y enseñanza de lenguas
9. Usos y aplicaciones específicas de la lingüística de corpus

El/la Director/a de Panel es elegido en la Asamblea Ordinaria de AELINCO y tiene una duración de tres años, renovable por otros tres. Es condición necesaria para ser director/a de panel ser socio de AELINCO. No es un cargo retribuido, pero se facilita su asistencia a los congresos anuales. Su elección responde a criterios científicos en el área que dirige. Sus obligaciones son:

- Gestionar las propuestas de comunicaciones, mesas redondas y pósters relativas a su panel en los congresos anuales responsabilizándose de su evaluación.
- Promover y dinamizar la participación de socios y no socios en los congresos anuales u otros actos científicos promovidos por AELINCO y relativos a su panel.

Cuando proceda la renovación de director/a de panel, las propuestas deberán presentarse a la Junta Directiva, por el interesado o por un mínimo de cinco socios de AELINCO, con dos semanas de antelación respecto al congreso anual de AELINCO. En caso necesario, las propuestas podrán hacerse también oralmente durante la Asamblea Ordinaria anual.

## Paneles y Directores

### **Diseño, elaboración y tipología de corpus**

*Begoña Crespo García*  
Universidade da Coruña  
Departamento de Filoloxía Inglesa  
Facultade de Filoloxía  
Campus da Zapateira, s/n  
15071 A Coruña (España)  
e-mail: [bcrespo@udc.es](mailto:bcrespo@udc.es)

### **Discurso, análisis literario y corpus**

*Gustavo A. Rodríguez Martín*  
Universidad de Extremadura  
Departamento de Filología Inglesa  
10003 Cáceres (España)  
e-mail: [garoma@unex.es](mailto:garoma@unex.es)

### **Estrudios gramaticales basados en corpus**

*Paula Rodríguez Puente*  
Universidad de Oviedo  
Departamento de Filología Inglesa, Francesa y Alemana  
Oviedo (España)  
e-mail: [rodriguezppaula@uniovi.es](mailto:rodriguezppaula@uniovi.es)

### **Lexicología y lexicografía basadas en corpus**

*Moisés Almela Sánchez*  
Universidad de Murcia  
30071 Murcia (España)  
Tel.: +34 868 88 3212  
e-mail: [moisesal@um.es](mailto:moisesal@um.es)

### **Corpus, estudios contrastivos y traducción**

*Sergio Maruenda Bataller*  
Universitat de València  
Departamento de Filología Inglesa y Alemana (IULMA)  
Facultad de Filología, Traducción i Comunicació  
Valencia (España)  
e-mail: [sergio.maruenda@uv.es](mailto:sergio.maruenda@uv.es)

### **Variación y cambio lingüístico basado en corpus**

*Carlos Prado Alonso*  
Departamento de Inglés, Francés y Alemán  
Universidad de Oviedo  
Campus de El Milán  
E-33011 Oviedo  
e-mail: [pradocarlos@uniovi.es](mailto:pradocarlos@uniovi.es)

### **Lingüística computacional basada en corpus**

*Chantal Pérez Hernández*  
Universidad de Málaga  
Departamento de Filología Inglesa, Francesa y Alemana  
Málaga (España)  
e-mail: [mph@uma.es](mailto:mph@uma.es)

### **Corpus, adquisición y enseñanza de lenguas**

*Begoña Clavel Arroitia*  
Universitat de València-IULMA  
46010 Valencia (España)  
Tel.: +34 963864262  
e-mail: [begona.clavel@uv.es](mailto:begona.clavel@uv.es)

### **Fines específicos y lingüística de corpus**

*Eva Mestre Mestre*  
Departamento de Lingüística Aplicada  
Escuela Politécnica Superior de Gandía  
Universitat Politècnica de València  
E-47730 Gandía  
e-mail: [eva.mestre@upvnet.upv.es](mailto:eva.mestre@upvnet.upv.es)

# CILC19 València



El Congreso Internacional de Lingüística de Corpus (CILC 2019) en su edición XI se celebró con gran número de participantes, y como estaba previsto y publicado en nuestros sitios web, entre los días 15 y 17 mayo de 2019 en la Universitat de València. El comité organizador estuvo compuesto por: Miguel Fuster-Márquez, Carmen Gregori-Signes, José Santaemilia Ruiz, y Paula Rodríguez-Abruñeiras. Todos ellos miembros de IULMA-UV y del departamento de Filología Anglesa y Alemania de la Universitat de València.

La edición CILC XI llevaba como título temático estelar *Análisis del Discurso desde la Óptica del Corpus - Corpus Approaches to Discourse Analysis*, una temática de especial interés en cuanto a la investigación de los organizadores se refiere y atendiendo al interés que despierta en un buen número de miembros de IULMA-UV. Destacamos a este respecto el gran nivel de las cuatro conferencias plenarias:

1. Maite Taboada (Universidad Simon Fraser). *Fantastic online comments and how to find them.*
2. Tony McEnery (Lancaster University). *Using Corpus Data to Explore Discourses of Climate in the Past.*

3. Alan Partington (Universidad de Bolonia). *That Other Kind of Cohesion. A CaDS Outline of Evaluative Cohesion.*
4. Nuria Lorenzo-Dus (Swansea University). *Developing Resilience against Online Grooming: A Corpus and Discourse Studies Approach.*

Igualmente, CILC 2019 contó con una mesa redonda de gran nivel titulada: *Analogía y tipología de corpus lingüísticos*. El Dr Pascual Cantos (Universidad de Murcia) actuó como moderador de la misma, y en la misma participaron los Dres. Antonio Briz (Universitat de València) y Giovanni Parodi (Universidad Católica de Valparaíso).

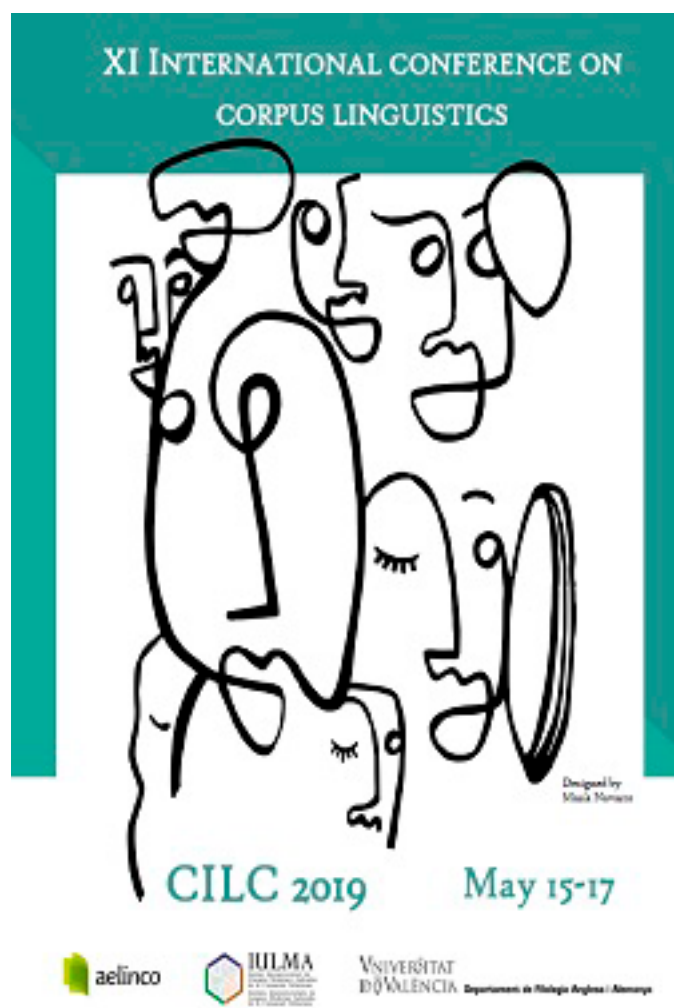
Como es habitual, la información y difusión internacional del Congreso se canalizó inicialmente a través de la plataforma de AELINCO: <http://www.aelinco.es/en>. Además se creó un sitio web específico para este congreso, con la colaboración de ADEIT: <https://congresos.adeituv.es/cilc2019/ficha.es.html>, encargada de facilitar toda la información del congreso y gestiones propias de una secretaría técnica (programas académico y social, libro de resúmenes, así como de llevar a cabo la inscripción de participantes, y asumir un gran número de gestiones).

### Comité científico

- Moisés Almela Sánchez (Universidad de Murcia, España)
- Paul Baker (Lancaster University, Reino Unido)
- Monika Bednarek (University of Sidney, Australia)
- Begoña Clavel Arroitia (Universitat de València-IULMA, España)
- Begoña Crespo García (Universidad de A Coruña, España)
- Sergio Maruenda Bataller (Universitat de València-IULMA, España)
- Tony McEnery (Lancaster University, Reino Unido)
- Eva Mestre Mestre (Universitat Politècnica de València, España)
- Nuria Lorenzo (Swansea University, Reino Unido)
- Alan Partington (Università di Bologna, Italia)
- Chantal Pérez Hernández (Universidad de Málaga, España)
- Carlos Prado Alonso (Universidad de Oviedo, España)
- Gustavo A. Rodríguez Martín (Universidad de Extremadura, España)
- Paula Rodríguez Puente (Universidad de Oviedo, España)
- Maite Taboada (Simon Fraser University, Canadá)

Siguiendo la costumbre establecida en las reuniones científicas de AELINCO, el congreso CILCXI recogió propuestas en torno a sus nueve paneles temáticos. Tras el proceso de revisión, fueron admitidas 150 contribuciones. A continuación, detallamos los paneles, sus directores y el número de contribuciones presentadas en cada uno de ellos:

1. Diseño, elaboración y tipología de corpus (Begoña Crespo): 20
2. Discurso, análisis literario y corpus (Gustavo A. Rodríguez Martín): 29
3. Estudios gramaticales basados en corpus (Paula Rodríguez Puente): 9
4. Lexicología y lexicografía basadas en corpus (Moisés Almela Sánchez): 11
5. Corpus, estudios contrastivos y traducción (Sergio Maruenda Bataller): 16
6. Variación y cambio lingüístico basados en corpus (Carlos Prado Alonso): 20
7. Lingüística computacional basada en corpus (Chantal Pérez Hernández): 12
8. 8. Corpus, adquisición y enseñanza de lengua (Begoña Clavel Arroitia): 17
9. 9. Usos y aplicaciones específicas de la lingüística de corpus (Eva Mestre Mestre): 16



Aunque la mayor parte de los participantes proceden de universidades españolas (69%), la apuesta internacional queda reflejada tanto en el programa como en el libro de resúmenes de CILC 2019. En ellos se constata el importante número de participantes de diferentes nacionalidades (31%).

Está previsto que durante este año 2020 vean la luz dos publicaciones como resultado del congreso. Una de ellas será publicada en la editorial *Peter Lang*, colección *Linguistic Insights*, y la otra en la editorial *Comares*, las cuales contendrán una selección de las aportaciones de mayor relieve y originalidad.

Por último, desearíamos agradecer en estas líneas el asesoramiento y apoyo proporcionados por la Junta Directiva de AELINCO, las directoras y directores de panel y también todas aquellas instituciones que han contribuido para hacer que CILC2019 fuera un éxito: el Departamento de Filología Anglesa y Alemana (UV), IULMA-UV, la Facultat de Filologia, Traducció i Comunicació (UV) así como la Generalitat València. Detrás de todas estas instituciones hay personas con nombres y apellidos que han puesto su particular empeño en hacer que CILC2019 sea recordado como un congreso memorable.

**Países participantes:**

País	Participantes
España	127
Bélgica	2
Brasil	4
Chile	1
China	1
R. Checa	5
Finlandia	1
Francia	4
Alemania	2
Grecia	1
Hungría	1
Irlanda	2
Italia	12
Polonia	4
Rusia	3
Suiza	1
Turquía	1
Reino Unido	9
EEUU	2
Venezuela	1

XII Congreso Internacional de Lingüística del Corpus

# CILC20 B érgamo

21-23 de mayo de 2020



La Asociación Española de Lingüística de Corpus (AELINCO) y el Comité Organizador de esta edición de su congreso anual tienen el gusto de invitarles a la *XII Conferencia Internacional de Lingüística de Corpus* (CILC2020), que tendrá lugar del 21 al 23 de mayo de 2020 en el *Dipartimento di Lingue, Culture e Letterature Straniere* de la Universidad de B érgamo (Italia).

Las dos temáticas vertebradoras de la conferencia serán el **Análisis del Discurso de Especialidad y los Estudios de Traducción desde la óptica del Corpus** (*“Corpus-Assisted Specialized Discourse and Translation Studies”*) y la reflexión propuesta se centrará en el rastreo, en los múltiples niveles del discurso especializado, de mecanismos metadiscursivos, patrones persuasivos, ideológicos o evaluativos que orientan la producción textual de múltiples comunidades de práctica (en términos de construcción de la identidad grupal o del género, de la inclusión o del rechazo de la alteridad, etc.), pese a la pretendida objetividad y ausencia de emotividad de la comunicación especializada.

Asimismo, se dará cabida a las aplicaciones de la lingüística de corpus dirigidas a investigar y aclarar el proceso traductor a través de corpus paralelos y/o comparables de textos especializados, con el propósito de definir un cuadro teórico-metodológico más preciso para la investigación traductológica y de arrojar luz sobre el posible papel desempeñado por la traducción en los cambios sociales en curso.

## PONENTES PLENARIOS

Han confirmado su participación los siguientes conferenciantes invitados:

- Mona Baker (University of Manchester)  
<https://www.research.manchester.ac.uk/portal/mona.baker.html>
- Paul Baker (University of Lancaster)  
<https://www.lancaster.ac.uk/linguistics/about-us/people/paul-baker>
- Sara Laviosa (Università degli Studi di Bari Aldo Moro)  
<https://www.uniba.it/docenti/laviosa-sara>

- Antonio Moreno Ortiz (Universidad de Málaga)  
<https://www.uma.es/departamento-de-filologia-inglesa-francesa-y-alemana/info/117649/antonio-moreno-ortiz/>

Asimismo, en el marco del congreso se desarrollará una mesa redonda moderada por Miguel Fuster Márquez (IULMA-UV) cuyo tema será: “Evaluation and Ideology in (specialised) discourse: a CL/computer-assisted approach”, en la que participarán:

- María Luisa Carrió Pastor (Universitat Politècnica de València)  
<http://www.upv.es/entidades/DLA/info/884030normalc.html>
- Tony McEnery (University of Lancaster)  
<https://www.lancaster.ac.uk/linguistics/about-us/people/tony-mcenery>
- Alan Partington (Università degli Studi di Bologna)  
<https://www.unibo.it/sitoweb/alanscott.partington>

El Comité Organizador de CILC2020 anima a todo aquel que desee participar a enviarnos una propuesta de presentación oral (20 minutos), póster, taller o seminario, siempre que la misma tenga relación con el lema de esta edición o con cualquiera de los nueve paneles temáticos de AELINCO:

1. Diseño, compilación y tipos de corpus
2. Discurso, análisis literario y corpus
3. Estudios gramaticales basados en corpus
4. Lexicología y lexicografía basadas en corpus
5. Corpus, estudios contrastivos y traducción
6. Corpus y variación lingüística
7. Lingüística computacional basada en corpus
8. Los corpus y la adquisición y enseñanza del lenguaje
9. Usos específicos de la lingüística de corpus.

Cada participante podrá enviar un máximo de dos propuestas, en **español, inglés o italiano**, hasta la fecha límite de presentación, el **10 de febrero de 2020**. Las instrucciones específicas y la hoja de estilo deben consultarse en este enlace: <http://www.aelinco.es/en/cilc-instructions-style-sheet>.

Las propuestas serán enviadas desde la plataforma EasyChair (<https://easychair.org/cfp/CILC2020>), al panel elegido para su revisión. Cada propuesta solo se puede enviar a un panel. La notificación de aceptación se efectuará el **30 de marzo de 2020**.

Una selección de los trabajos de investigación presentados en el CILC2020 será publicada en un volumen de la serie *Linguistic Insights* de la editorial Peter Lang, mientras que las mejores ponencias presentadas por jóvenes investigadores se publicarán en un volumen de la serie *Cerlis Series* (<http://dinamico.unibg.it/cerlis/page.aspx?p=257>).

Si tienen preguntas, sugerencias o desean realizar alguna consulta, pueden contactarnos en: [cilc2020@unibg.it](mailto:cilc2020@unibg.it).

Para información más detallada sobre la organización de CILC2020, visiten nuestro sitio web: <https://www.unibg.it/cilc2020>.

Pueden enviar esta convocatoria de propuestas a cualquier colega interesado.

Confiamos en una nutrida asistencia, les esperamos en Bèrgamo.

**El comité organizador de CILC2020**

Giovanni Garofalo  
Luisa Chierichetti  
Stefania Maci  
Andrés Ortega Garrido

# CILC2021 Oviedo

La decimotercera edición del Congreso Internacional de Lingüística de Corpus (CILC2021) de AELINCO tendrá lugar en Oviedo en el mes de mayo de 2021. El comité organizador lo componen varios profesores e investigadores del Departamento de Filología Inglesa, Francesa y Alemana de la Universidad de Oviedo (Uniovi). Las sesiones se celebrarán en el Edificio Departamental de la Facultad de Filosofía y Letras, situado en el Campus de Humanidades. El Edificio Departamental, inaugurado en el año 1905, es uno de los seis edificios en los que se divide la facultad y contiene espacios dotados de recursos tecnológicos actualizados para la celebración del evento, además de una amplia capacidad para asistentes.

Como es sabido, la mayoría de estudios lingüísticos basados en, o derivados de, corpus tienen como fundamento, implícita o explícitamente, la noción de género, así como otros conceptos relacionados tales como el de registro o dominio. La importancia de estos conceptos en la Lingüística de Corpus ha sido puesta de manifiesto en repetidas ocasiones en los estudios de Douglas Biber y otros investigadores. Por este motivo, el foco temático de esta edición es la Lingüística de Corpus en la era digital, con especial referencia a su aplicación en el análisis de géneros, registros y dominios. Además de este tema principal, en el congreso se prevé también que tengan cabida temáticas tales como la compilación de corpus y el análisis estadístico de datos lingüísticos. Dichas temáticas se tratarían en forma de talleres o seminarios que se impartirían durante el evento.

El comité organizador de CILC 2021 está convencido de que las comunicaciones que se presentarán en

este evento científico permitirán profundizar en el análisis de la temática principal del congreso. Para ello, esperamos poder contar con un amplio número de comunicaciones. Además, cuatro conferenciantes plenarios de reconocido prestigio internacional en el ámbito de la Lingüística de Corpus han confirmado ya su participación en el congreso:

- Teresa Fanego  
(Universidad de Santiago de Compostela, España)
- Gaëtanelle Gilquin  
(Universidad Católica de Lovaina, Bélgica)
- Jukka Tyrkkö  
(Universidad Linneo, Suecia)
- Roberto Valdeón  
(Universidad de Oviedo, España)



Universidad de Oviedo

# Revista Research in Corpus Linguistics



*Research in Corpus Linguistics (RiCL)* es la revista científica oficial de AELINCO, que hasta el año 2018 estuvo editada por Javier Pérez Guerra (Universidad de Vigo) y María José López Couso (Universidad de Santiago de Compostela) y que, desde 2019, coordinan Paula Rodríguez Puente y Carlos Prado-Alonso, ambos de la Universidad de Oviedo. Se trata de una revista indexada, de libre acceso, y con una periodicidad anual hasta 2019. A partir de 2020 *RiCL* publicará dos números al año. Todos los trabajos presentados para su publicación están sometidos a un proceso de revisión ciega por pares, tanto externo como por el propio Comité Científico. El pasado mes de noviembre se publicó el número 7 en la plataforma informática OJS, actualizada a la nueva versión de este software, plenamente compatible con todo tipo de dispositivos informáticos. Todos los contenidos de la revista están en libre acceso, disponibles a través del enlace <http://ricl.aelinco.es>. Coincidiendo con la redacción de esta nota, se está procediendo a la revisión de los trabajos recibidos para su publicación en el primer número

***Animamos a todos los socios y socias de AELINCO a que envíen sus trabajos para su consideración en RiCL, ya sea para el próximo número o cualquiera de los siguientes***

del volumen 8 de 2020.

*RiCL* forma parte de la Open Access Scholarly Publishers Association (OASPA), lo cual, entre otros beneficios, permite que las publicaciones de *RiCL* dispongan de DOI. Igualmente, *RiCL* está incluida en reconocidas bases de datos e indexación, tales como el directorio ERIH Plus (European Reference Index for the Humanities and Social Sciences), DOAJ (Directory of Open Access Journal), Dialnet, Latindex, Linguistic Bibliography Online, MIAR, Ulrich's web y WorldCat, lo cual le otorga una amplia visibilidad internacional.

Animamos a todos los socios y socias de AELINCO

a que envíen sus trabajos para su consideración en *RiCL*, ya sea para el próximo número o cualquiera de los siguientes. *RiCL* recibe propuestas de artículos durante todo el año. Adjuntamos el 'open call for papers' de *RiCL*, al cual se da difusión igualmente a través de los canales habituales. La recepción de originales se efectúa únicamente en línea a través de la web de la revista.

*Research in Corpus Linguistics (RiCL*, ISSN 2243-4712), the official journal of the Spanish Association for Corpus Linguistics AELINCO, is a scholarly peer-reviewed international scientific journal published twice a year, aiming at the publication of contributions which contain empirical analyses of data from different languages and from different theoretical perspectives and frameworks, with the goal of improving our knowledge about the linguistic theoretical background of a language, a language family or any type of cross-linguistic phenomena/constructions/assumptions.

*RiCL* invites previously unpublished submissions in three main forms:

1. **Research articles** reporting on research based on or derived from linguistic corpora (between 6,000 and 10,000 words).
2. **Reports on corpus construction, annotation, the development and application of corpus tools, software, etc.** (between 3,000 and 5,000 words).
3. **Book reviews in the field of Corpus Linguistics** (between 1,500 and 3,000 words).

*Aiming at the publication of contributions which contain empirical analyses of data from different languages and from different theoretical perspectives and frameworks*

The journal also publishes special issues on specific topics, with leading specialists in the field of corpus linguistics as guest editors. Potential editors are

welcome to send their proposals to the journal's e-mail: [ojs@aelinco.es](mailto:ojs@aelinco.es)

Specific areas of interest include corpus design, compilation and typology; discourse, literary analysis

and corpora; corpus-based grammatical studies; corpus-based lexicology and lexicography; corpora, contrastive studies and translation; corpus and linguistic variation; corpus-based computational linguistics; corpora, language acquisition and teaching; and special uses of corpus linguistics.

**Submissions** are received online throughout the whole year.

**Editors of RiCL:**

Paula Rodríguez-Puente and Carlos Prado-Alonso  
(University of Oviedo)

**RiCL** 

---

 **Research in  
Corpus Linguistics** 

---

# Rincón del Socio

---



## **Corpus of History English Texts (CHET)**

*Moskowich, Isabel | Sánchez Barreiro, Estefanía | Lareo, Inés | Lojo Sandino, Paula*

The Corpus of History English Texts (CHET), part of the Coruña Corpus of English Scientific Writing, came out in March 2019. Compiled by members of the MuStE group (University of A Coruña) and as is the case with its twin corpora, it contains samples of texts produced between 1700 and 1900 and intends to reflect the tendencies of academic and scientific writing present in late Modern English. Also, as in previous occasions the corpus is accompanied by a book (*Writing History in Late Modern English*, Amsterdam: Benjamins, <https://benjamins.com/catalog/z.225>) containing methodological accounts as well as pilot studies. It is available online and open access at <https://ruc.udc.es/dspace/handle/2183/21849>. The previous sections of the Coruña Corpus and its software have been made publicly available as well at.