MARK DAVIES

Professor of Linguistics
https://www.mark-davies.org
mark@mark-davies.org
Retired 2020

EDUCATION

- **1992**. PhD Iberoromance Philology and Linguistics, Univ. of Texas at Austin. Double specialization: Syntax and Historical Linguistics.

- **1989**. MA, Spanish Linguistics, Brigham Young Univ., Provo, Utah.

- **1986**. BA, Spanish and Linguistics (double major), Brigham Young Univ., Provo, Utah.

EMPLOYMENT

- **2006-2020**. Professor of Linguistics. Brigham Young University

- **2003-2006**. Associate Professor of Linguistics. Brigham Young University

- **1999-2003**. Associate Professor of Spanish Linguistics. Illinois State University

- **1992-1999**. Assistant Professor of Spanish Linguistics. Illinois State University

---

GRANTS

- **2015-2018**. National Endowment for the Humanities. (To create a two billion word extension to the Corpus del Español and a one billion word extension to the Corpus do Português)

- **2014-2016.** Korea Research Foundation. ("Synchronic and Diachronic Aspects of English What Constructions: A Corpus-Based Approach"). Co-PI with Jong-Bok Kim, Kyung Hee University.

- **2013-2015**. Arts and Humanities Research Council (AHRC), United Kingdom. (SAMUELS: Semantic Annotation and Mark-Up for Enhancing Lexical Searches). Sub-contract to University of Glasgow.

- **2012-2015**. National Science Foundation. ("A Linguistic Taxonomy of English Web Registers". Co-PI with Douglas Biber of Northern Arizona University)

- **2008-2010.** National Endowment for the Humanities. (To create a 400 million word, annotated, online corpus of historical American English, 1810s-2000s.)

- **2004-2006**. National Endowment for the Humanities. (To create a 45 million word, annotated, online corpus of Portuguese, 1300s-1900s. Co-PI with Michael Ferreira of Georgetown University.)

- **2002-2004**. National Science Foundation. (To research the "Multi-dimensional analysis of register variation in Spanish". Co-PI with Douglas Biber of Northern Arizona University.)

- **2001-2002**. National Endowment for the Humanities. (To create a 100 million word, annotated, online corpus of Spanish, 1200s-1900s)

PUBLICATIONS

Books and Monographs

1. **(2017)** *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. Second edition: revised and expanded. Routledge. (Co-authored with Kathy Hayward Davies)

2. **(2010)** *A Frequency Dictionary of American English: Word Sketches, Collocates, and Thematic Lists*. Routledge. (Co-authored with Dee Gardner.)

3. **(2009)** *Corpus linguistic applications: current studies, new directions*. Rodopi. (Co-editor, with Stefan Gries and Stefanie Wulff.)

4. **(2007)** *A Frequency Dictionary of Portuguese: Core Vocabulary for Learners*. Routledge. (Co-authored with Ana Maria Raposo Preto-Bay)

5. **(2005)** *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. Routledge.

6. **(2004)** *El uso del Corpus del Español y otros corpus para investigar la variación actual y los cambios históricos*. Tokyo: Univ. Sophia.

Journal articles and chapters in books

7. **(2023)** "Creating and Using "Virtual Corpora" to Extract and Analyse Domain-Specific Vocabulary at English-Corpora.org". In: J Pan and S Laviosa. *Corpora and Translation Education. New Frontiers in Translation Studies*. Springer: Singapore.

8. **(2021)** "The Coronavirus Corpus: Design, construction, and use". *International Journal of Corpus Linguistics*. 26(4): 583-98.

9. **(2021)** "Constitución de corpus crecientes del español". In Giovanni Parodi, Pascual Cantos, Chad Howe. *The Routledge Handbook of Spanish Corpus Linguistics*. (With Giovanni Parodi). 13-32.

10. **(2020)** "English what with absolute constructions: a Construction Grammar perspective." *English Language and Linguistics* 24(4): 637-666. (With Jong-Bok Kim)

11. **(2020)** "The TV and Movies corpora: design, construction, and use." *International Journal of Corpus Linguistics.* 26(1): 10-37.

12. **(2020)** "First Citations and First Occurrences: How Sensitive to Language Change Were Nineteenth- and Twentieth-Century American Dictionaries? *Dictionaries: Journal of the Dictionary Society of North America*. 41(2): 61-86. (With Don Chapman and Amanda Fronk)

13. **(2020)** The Internet-Based English Corpus iWeb and Its Application in Teaching and Learning English as a Second Language. *TEFLE*, (4), 73-81+12. (With PENG Xinjia).

14. **(2019)** "The advantages and challenges of 'big data': Insights from the 14 billion word iWeb corpus". *Linguistic Research* 36(1), 1-34. (With Jong-Bok Kim)

15. **(2019)** "The best of both worlds: Multi-billion word 'dynamic' corpora". In Piotr Bański, et a. *Proceedings of the Workshop on Challenges in the Management of Large Corpora* (CMLC-7) 2019. Mannheim: Leibniz-Institut fur Deutsche Sprache.

16. **(2019)** "If olive oil is made of olives, then what's baby oil made of? The shifting semantics of Noun+Noun sequences in American English." In J. Egbert & P. Baker (Eds.), *Using corpus methods to triangulate linguistic analysis*, New York: Routledge. 163-84. (With Jesse Egbert)

17. **(2019)** "Historical shifts with the into-causative construction in American English." *Linguistics* 57: 29-58. (With Jong-Bok Kim)

18. **(2018)** "Sorting them all out: Exploring the separable phrasal verbs of English." *System* 76: 197-209. (With Dee Gardner)

19. **(2018)** "Using (and useful) corpora for the study of the history of English". In *Teaching the History of the English Language*, eds. Chris Palmer and Colette Moore. MLA Options for Teaching Series.

20. **(2018)** "Corpus-based studies of lexical and semantic variation: The importance of both corpus size and corpus design." In *From data to evidence in English language research* (Digital Linguistics), eds. Suhr, Carla, Terttu Nevalainen and Irma Taavitsainen. Leiden: Brill. 34-55.

21. **(2018)** "Uso del Corpus del Español y los corpus relacionados para la lexicografía histórica española." *In Historia del léxico español y Humanidades digitales*. Eds. Alejandro Fajardo, et al. Berlin: Peter Lang. 49-76.

22. **(2017)** "Using Large Online Corpora to Examine Lexical, Semantic, and Cultural Variation in Different Dialects and Time Periods". *In Corpus-Based Sociolinguistics*, ed. Eric Friginal et al. London: Routledge. 19-82.

23. **(2016)** "The Effect of Representativeness and Size in Historical Corpora: An Empirical Study of Changes in Lexical Frequency." In *Studies in the History of the English Language VII: Generalizing vs. particularizing methodologies in historical linguistic analysis,* eds. Don Chapman, Colette Moore, and Miranda Wilcox. Berlin: De Gruyter / Mouton. 131-50. (With Don Chapman)

24. **(2016) "**The Into Causative Construction in English: A Construction-based Perspective." *English Language and Linguistics* 20 (1): 55-83. (With Jong-Bok Kim)

25. **(2016)** "A response to 'To what extent is the Academic Vocabulary List relevant to university student writing?". *English for Specific Purposes* 42: 62-68. (With Dee Gardner)

26. **(2015)** "Corpora: An Introduction". In *Cambridge Handbook of English Corpus Linguistics*, eds. Douglas Biber and Randi Reppen. Cambridge: Cambridge University Press. 11-31.

27. **(2015)** "A Corpus Linguistic Approach to Vocabulary Learning for University Students." In *ESL Readers and Writers in Higher Education: Understanding Challenges, Providing Support*, eds. Norm Evans, Neil Anderson, and William Eggington. London: Routledge. 180-197. (With Dee Gardner)

28. **(2015)** "Introducing the 1.9 Billion Word Global Web-Based English Corpus (GloWbE)." *21st Century Text*. (Peer-reviewed, online journal).

29. **(2015)** "Exploring the Composition of the Web: A Corpus-based Taxonomy of Web Registers". *Corpora* 10 (1): 11-45. (With Douglas Biber and Jesse Egbert)

30. **(2015)** "Expanding Horizons in the Study of World Englishes with the 1.9 Billion Word Global Web-Based English Corpus (GloWbE)." *English World-Wide* 36: 1-28. (With Robert Fuchs)

31. **(2015)** "The importance of robust corpora in providing more realistic descriptions of variation in English grammar". In *Linguistic Vanguard* (peer-reviewed online journal from Mouton de Gruyter)

32. **(2015)** "Developing a Bottom-up, User-based Method of Web Register Classification" in its current form for publication in *Journal of the Association for Information Science and Technology*. (With Douglas Biber and Jesse Egbert)

33. **(2014)** "Making Google Books n-grams useful for a wide range of research on language change". *International Journal of Corpus Linguistics* 19 (3): 401-16.

34. **(2014)** "Powerful (yet simple) comparisons of a wide range of phenomena in British and American English". *ICAME Journal* 38:35-56.

35. **(2014)** "Creating and Using the *Corpus do Português* and the *Frequency Dictionary of Portuguese*". In *Working with Portuguese Corpora*, eds. Tony Berber Sardinha and Telma Ferreira. Continuum Publishers. 89-110.

36. **(2014)** "Examining syntactic variation in English: the importance of corpus design and corpus size". *English Language and Linguistics* 19 (3): 1-35.

37. **(2013)** "Google Scholar vs. COCA: two very different approaches to examining academic English". *Journal of English for Academic Purposes* 12: 155-165.

38. **(2013)** "A New Academic Vocabulary List." In *Applied Linguistics* 35: 1-24. (With Dee Gardner)

39. **(2013)** "Establishing Corpora from Existing Data Sources". In *Data Collection in Sociolinguistics: Methods and Applications*, ed Christine Mallinson, et al. London: Routledge. 210-12.

40. **(2012)** "Expanding Horizons in Historical Linguistics with the 400 million word Corpus of Historical American English". *Corpora* 7: 121-57.

41. **(2012)** "Examining Recent Changes in English: Some Methodological Issues". In *The Oxford Handbook of the History of English*, eds. Terttu Nevalainen and Elizabeth Closs Traugott. Oxford: Oxford Univ. Press. 263-87.

42. **(2012)** "Recent shifts with three nonfinite verbal complements in English: Data from the 100 million word TIME Corpus (1920s-2000s)". In *Current Change in the English Verb Phrase*, ed. Bas Aarts, et al. Cambridge: Cambridge Univ. Press. 46-67.

43. **(2012)** "The 400 Million Word Corpus of Historical American English (1810-2009)". In *English Historical Linguistics 2010*, ed. Irén Hegedus, et al. Philadelphia: John Benjamins. 217-50.

44. **(2012)** "Looking at Recent Changes in English with the Corpus of Contemporary American English (COCA)". *21st Century Text*. (Peer-reviewed, online journal).

45. **(2012)** "Comparisons between Google Books and Google Books Corpus." *Computer-assisted Foreign Language Education*. 145:15-18. (With Xingfu Wang).

46. **(2011)** "Synchronic and Diachronic Uses of Corpora". In *Perspectives on Corpus Linguistics: Connections & Controversies*, eds. Vander Viana, Sonia Zyngier and Geoff Barnbrook. Philadelphia: John Benjamins. 63-80.

47. **(2011)** "Creating and Using the Frequency Dictionary of Contemporary American English: Word Sketches, Collocates, and Thematic Lists". In *Corpus-based studies in language use, language learning, and language documentation*, ed. John Newman, et al. Amsterdam: Rodopi. 283-97.

48. **(2011)** "The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English". *Literary and Linguistic Computing* 25: 447-65.

49. **(2010)** "More than a peephole: Using large and diverse online corpora". *International Journal of Corpus Linguistics* 15: 405-11.

50. **(2010)** "Semantically-based, learner-oriented queries with the 400+ million word Corpus of Contemporary American English". *Łódź Studies in Language*, ed. Stanislaw Goźdź-Roszkowski. Frankfurt: Peter Lang.

51. **(2010)** "Creating Useful Historical Corpora: A Comparison of CORDE, the Corpus del Español, and the Corpus do Português". In *Diacronía de las lenguas iberorromances: nuevas perspectivas desde la lingüística de corpus*, ed. Andrés Enrique-Arias. Frankfurt/Madrid: Vervuert/Iberoamericana. 137-66.

52. **(2010)** "What students need (and want): semantically-oriented queries in large online corpora". *SYNAPS* (Bergen) 24: 27-40.

53. **(2009)** "The 385+ Million Word Corpus of Contemporary American English (1990-2008+): Design, Architecture, and Linguistic Insights". *International Journal of Corpus Linguistics.* 14: 159-90.

54. **(2009)** "Relational databases as a robust architecture for the analysis of word frequency". In What's in a Wordlist?: In *Investigating Word Frequency and Keyword Extraction*, ed. Dawn Archer. London: Ashgate. 53-68.

55. **(2008)** "Spanish and Portuguese Corpus Linguistics". *Studies in Hispanic and Lusophone Linguistics*. 1:149-86.

56. **(2008)** "The corpus-based Frequency Dictionary of Portuguese: A new tool for learners and teachers." In *Proceedings of TALC 8: Teaching and Language Corpora*, ed. Ana Frankenberg-Garcia, et al. Lisbon. (Co-authored with Ana Maria Raposo Preto-Bay)

57. **(2008)** "The Corpus of Contemporary American English--a Useful Tool for English Teaching and Research". *Computer-Assisted Foreign Language Education in China*. 5:24-31 (Co-authored with Wang Xingfu and Liu Guohui).

58. **(2007)** "Pointing Out Frequent Phrasal Verbs: A Corpus-Based Analysis". *TESOL Quarterly* 41:339-59. (Co-authored with Dee Gardner)

59. **(2007)** "Semantically-based queries with a joint BNC/WordNet database". In *Corpus Linguistics Twenty-five Years On*, ed. Roberta Facchinetti. Amsterdam: Rodopi. 149-167.

60. **(2006)** "Towards the first comprehensive survey of register variation in Spanish". In *Corpus Linguistics Beyond the Word: Corpus Research from Phrase to Discourse*, ed. Eileen Fitzpatrick. Rodopi. 73-86.

61. **(2006)** "Vocabulary Coverage in Spanish Textbooks: How Representative is It?" In *Selected Proceedings from the Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, ed. Jacqueline Toribio. Cascadilla. 132-43. (Co-authored with Timothy L. Face). 132-43.

62. **(2006)** "Spoken and written register variation in Spanish: A Multi-dimensional Analysis." *Corpora* 1:1-37. (Co-authored with Doug Biber, James Jones, and Nicole Tracy-Ventura).

63. **(2005)** "The advantage of using relational databases for large corpora: speed, advanced queries, and unlimited annotation". *International Journal of Corpus Linguistics* 10: 301-28.

64. **(2005)** "On diachronic shifts with Spanish se: preliminary evidence from large electronic corpora." In *Romance Corpus Linguistics II: Corpora and Diachronic Linguistics*, ed. Claus Pusch, et al. Guntar Naar. 431-42.

65. **(2005)** "Vocabulary Range and Text Coverage: Insights from the Forthcoming Routledge Frequency Dictionary of Spanish". In *Selected Proceedings from the 7th Hispanic Linguistics Symposium*, ed. David Eddington. 106-15.

66. **(2005)** "Advanced research on syntactic and semantic change with the Corpus del Español". In *Romance Corpus Linguistics II: Corpora and Diachronic Linguistics*, ed. Claus Pusch, et al. Guntar Naar. 203-14.
Reprinted in: *Corpus Linguistics. Critical Concepts in Linguistics* (6 vols.). Ed. Teubert, Wolfgang & Ramesh Krishnamurthy. London: Routledge. 337-48 (Volume 5).

67. **(2004)** "Student use of large, annotated corpora to analyze syntactic variation". In *Corpora and Language Learners*, ed. Guy Aston, et al. Philadelphia: John Benjamins. 259-69.

68. **(2004)** "Student use of large corpora to investigate language change". In *Applied Corpus Linguistics: A Multidimensional Perspective*, ed. Thomas Upton, et al. Amsterdam: Rodopi. 207-22.

69. **(2003)** "Diachronic Shifts and Register Variation with the "Lexical Subject of Infinitive" Construction. (Para yo hacerlo)". In *Linguistic Theory and Language Development in Hispanic Languages*, ed. Silvina Montrul and Francisco Ordóñez. Somerville, MA: Cascadilla Press. 13-29.

70. **(2003)** "Annotation without lexicons: an alternative to the standard bootstrapping approach". In *Proceedings from Corpus Linguistics 2003*, ed. Paul Rayson, et al. 174-83.

71. **(2002)** "Un corpus anotado de 100.000.000 palabras del español histórico y moderno". *SEPLN 2002* (Sociedad Española para el Procesamiento del Lenguaje Natural). 21-27.

72. **(2002)** "'Esto es ligero de fazer: Object to Subject Raising in Medieval and Early Modern Spanish". In *Structure, Meaning, and Acquisition of Spanish*, ed. James F. Lee, et al. Somerville, MA: Cascadilla Press. 19-31.

73. **(2001)** "Creating and using multi-million word corpora from web-based newspapers". In *Corpus Linguistics in North America*, eds. Rita C. Simpson and John M. Swales. Ann Arbor: U Michigan P. 58-75.

74. **(2000)** "Using multi-million word corpora of historical and dialectal Spanish texts to teach advanced courses in Spanish linguistics". In *Rethinking Language Pedagogy from a Corpus Perspective*, eds. Lou Burnard and Tony McEnery. Frankfurt am Main; New York: P. Lang. 173-85.

75. **(2000)** "Syntactic Diffusion in Spanish and Portuguese Infinitival Complements". In New Approaches to Old Problems: In *Issues in Romance Historical Linguistics*, eds.Steven Dworkin and Dieter Wanner. Amsterdam; Philadelphia: John Benjamins. 109-27.

76. **(1999**) "The Historical Development of Subject Raising in Portuguese: A Corpus-Based Approach". *Neuphilologische Mitteilungen* 100:95-110.

77. **(1999)** "A Computer Corpus-Based Study of Subject Raising in Modern Portuguese". *Lingvisticae Investigationes* 21:379-400.

78. **(1998)** "The Evolution of Spanish Clitic Climbing: A Corpus-Based Approach." *Studia Neophilologica* 69:251-63.

79. **(1997)** "A Corpus-Based Approach to Diachronic Clitic Climbing in Portuguese." *Hispanic Journal* 17: 93-111.

80. **(1997)** "Using Large Computer-Based Corpora as a Philological Tool: An Analysis of Four Medieval Spanish Bibles." *Dactylus* 16: 70-92.

81. **(1997)** "The History of Subject Raising in Spanish". *Bulletin of Hispanic Studies* (Liverpool) 74: 399-411.

82. **(1997)** "A Corpus-Based Analysis of Subject Raising in Modern Spanish." *Hispanic Linguistics* 9: 33-63.

83. **(1996)** "The Diachronic Interplay of Finite and Nonfinite Verbal Complements in Spanish and Portuguese." *Bulletin of Hispanic Studies* (Glasgow) 73:137-58.

84. **(1996)** "The Diachronic Evolution of the Causative Construction in Portuguese." *Journal of Hispanic Philology* 17:261-92.

85. **(1995)** "The Evolution of Causative Constructions in Spanish and Portuguese." In *Current Research in Romance Linguistics*, ed. John Amastae, et al. Philadelphia: John Benjamins, 1995. 105-122.

86. **(1995)** "The Evolution of the Spanish Causative Construction." *Hispanic Review* 63:57-77.

87. **(1995)** "Analyzing Syntactic Variation with Computer-Based Corpora: The Case of Modern Spanish Clitic Climbing". *Hispania* 78:370-380.

88. **(1994)** "Parameters, Passives, and Parsing: Explaining Diachronic Shifts in Spanish and Portuguese". In *Variation and Linguistic Theory*, ed. K. Beals, et al. Chicago: CLS. Vol 2. 46-60.

89. **(1992)** "A Tentative Bibliography of Historical Spanish Syntax." *Hispanic Linguistics* 5:279-351.

Reviews

90. **(2009)** Review of Using Spanish Corpora (Giovanni Parodi). *Modern Language Journal*. 93: 467-68.

91. **(2009)** Review of The International Corpus of English – British Component (ICE-GB), the Diachronic Corpus of Present-day Spoken English (DCPSE), and ICECUP 3.1. *Language*. 85: 443-45.

92. **(2004)** Review of Léxico Hispanoamericano (Peter Boyd-Bowman, et al). *La Coronica: A Journal of Medieval Spanish Literature and Language* 33: 259-64.

93. **(2004)** Review of Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching (Sylvaine Granger, et al). *Modern Language Journal.* 88: 469-70.

94. **(2001)** "Review of Construcciones causativas en el español medieval (Milagros Alfonso Vega). *Revista Canadiense de Estudios Hispánicos* 25: 329-30.

95. **(1995)** "Omnipage and WordCruncher: Tools for Creating and Searching Digitized Text Corpora." *La Corónica* 23:111-115.

CONFERENCE PRESENTATIONS / WORKSHOPS.

1. **(2026)** "What does AI know about language change in English?" Studies in the History of the English Language (SHEL 14). Brigham Young University. [Keynote address]

2. **(2025)** "The relevance of large, structured corpora in the age of Large Language Models"**.** 16th International Conference on Corpus Linguistics. Univ Salamanca. [Keynote address]

3. **(2025)** "English-Corpora.org: Challenges, innovation, and sustainability". Corpus linguistics 2040: Which data, which methods, which models? Mannheim, Germany. [Keynote address]

4. **(2020)** "The expanded one billion word COCA corpus: new features for researchers, teachers, and learners". TALC20: Teaching and Learning Corpora. Univ Perpignan, France. [Keynote address]

5. **(2020)** ---------. 5th Asia Pacific Corpus Linguistics Conference. Yonsei Univ, Korea. [Keynote address] (Canceled because of COVID-19).

6. **(2019)** "Making very large corpora useful: Creating Virtual Corpora in the BYU corpora". 4th Corpus Linguistics in China Conference (CLIC 2019), Shanghai, China. [Keynote address]

7. **(2019)** "Using the new 14 billion word iWeb corpus for teaching and research". Corpus Linguistics 2017. Cardiff Univ, Wales.

8. **(2019)** "The best of both worlds: Multi-billion word "dynamic" corpora". Challenges in the Management of Large Corpora (CMLC). Cardiff, Wales.

9. **(2019)** "Using the 7.8 billion word NOW corpus to look at a wide range of recent changes in English". ICAME 40. Univ Neuchatel, Switzerland.

10. (**2018**) "iWeb: The 14 billion word, learner-oriented Web Corpus". 4th Corpus Linguistics in China Conference (CLIC 2018), Wuhan, China. [Keynote address]

11. **(2018)** "New corpus tools for research and teaching: The Corpus del Español and WordAndPhrase (Spanish), the NOW "monitor" corpora (English / Spanish), and the new

iWeb corpus". Congreso Internacional de Lingüística del Corpus (Asociación Española de Lingüística del Corpus). Caceres, Spain. [Keynote address]

12. **(2018)** "The 14 billion word iWeb corpus: a new tool for research, teaching, and learning". Korean Association for Corpus Linguistics. Seoul, South Korea. [Keynote address].

13. **(2018)** "Transparent Free Relatives: A Corpus-Based Perspective". American Association for Corpus Linguistics. Georgia State University. (With Jong-Bok Kim)

14. **(2018)** "Using the new 14 billion word iWeb corpus for teaching and research". American Association for Corpus Linguistics. Georgia State University.

15. **(2017)** Looking at variation and change in language and culture with large online corpora. University of Tübingen (Germany). [Invited lecture]

16. **(2017)** Looking at variation and change in language and culture with large online corpora. University of Freiburg (Germany). [Invited lecture]

17. **(2017)** Examining linguistic variation (historical, dialectal, and genre-based) with the BYU corpora (COCA, COHA, EEBO, GloWbE, and NOW). 7th Biennial Conference on the Linguistics of Contemporary English. Vigo, Spain. [Plenary speaker]

18. **(2017)** Why size alone is not enough: The importance of historical, genre-based, and dialectal variation in language. Congreso Internacional de Lingüística Computacional y de Corpus. Instituto Caro y Cuervo, Bogota, Colombia. [Keynote speaker]

19. **(2017)** The New 4 Billion Word NOW Corpus, with 4-5 Million Words of Data Added Every Day. Corpus Linguistics 2017. Univ Birmingham, UK.

20. **(2017)** Looking at variation and change in language and culture with large online corpora. Calvin College. [Invited lecture]

21. **(2016)** Creating and using domain-specific "virtual corpora" with the 1.9 billion word Wikipedia Corpus (and other BYU corpora). American Association of Corpus Linguistics (AACL). Iowa State Univ.

22. **(2016)** The new 3 billion word NOW corpus, with five million words of data added every day. American Association of Corpus Linguistics (AACL). Iowa State Univ.

23. **(2016)** Why size alone is not enough: The importance of historical, genre-based, and dialectal variation in language. Big data in linguistic studies. Kyung-Hee Univ. [Plenary speaker]

24. **(2016)** The new 2.9 billion word NOW Corpus: up-to-date as of … yesterday. The 20th Workshop on Linguistics and Language Processing. Kyung-Hee University. [Plenary speaker]

25. **(2016)** Creating domain-specific 'virtual corpora' with the 1.9 billion word Wikipedia Corpus (and other BYU corpora). ICAME 37. Chinese University of Hong Kong.

26. **(2016)** Exploring cultural and semantic differences between World Englishes with the GloWbE (and NOW) corpora. ICAME 37. Chinese University of Hong Kong.

27. **(2015)** "Why size alone is not enough: The importance of historical, genre-based, and dialectal variation in language". From Data to Evidence in English language research: Big Data, Rich Data and Uncharted Data". University of Helsinki. [Keynote speaker]

28. **(2015)** "The INTO CAUSATIVE Construction: Insights from Large Corpora". Symposium on Complementation in Honour of Professor Juhani Rudanko. Univ. Tampere, Finland. [Plenary speaker]

29. **(2015)** "Examining syntactic variation in English: the importance of corpus design and corpus size." English linguistics and corpora: research issues and language teaching innovations. Université Paris Est Créteil. [Keynote speaker]

30. **(2015)** "Advanced lexicographical research with the new Wikipedia Corpus". Dictionary Society of North America (DSNA). University of British Columbia.

31. **(2015)** "A large corpus-based study of the historical development of [Noun+Noun] sequences in American English". Studies in the History of the English Language (SHEL). University of British Columbia.

32. **(2014)** "The development of the "causative V-ing" construction in American English.. International Conference of English Historical Linguistics. University of Leuven, Belgium.

33. **(2014)** "Expanding Horizons in the Study of World Englishes with the 1.9 Billion Word Global Web-Based English Corpus (GloWbE)." American Association for Corpus Linguistics. Northern Arizona University.

34. **(2014)** "Determining the Technical Vocabulary of Academic English: A Corpus-Based Analysis." American Association for Corpus Linguistics. Northern Arizona University. (With Dee Gardner)

35. **(2014)** "Using large online corpora for research and teaching". The 17th Korea-Japan Workshop on Linguistics and Language Processing. Kyung Hee University. Seoul, Korea. [Keynote speaker]

36. **(2014)** Series of four invited workshops (total of 12 hours) dealing with lexicographical research, to graduate students of Guangdong University of Foreign Studies (GDUFS), Guangzhou, China.

37. **(2013)** "How can big data best be exploited for OED? Historical linguistics using large corpora". "The Oxford English Dictionary: defining the future" symposium. University of Oxford. [Invited]

38. **(2013)** "Looking at a wide range of linguistic changes with the 400 million word Corpus of Historical American English (1810-2009)". Workshop on Diachronic Corpora. University of Athens, Greece. [Invited]

39. **(2013)** "The 155 billion word Google Books corpus: Can it be used for serious research on diachronic syntax?" Studies in the History of the English Language. Brigham Young University.

40. **(2013)** "Examining syntactic variation in English: the importance of corpus design and corpus size". International Conference on English Linguistics (ELSOK**).** Seoul, South Korea. [Keynote speaker]

41. **(2013)** "Using Online Corpora to Examine Variation and Change in English". Workshop at the International Conference on English Linguistics (ELSOK**).** Seoul, South Korea. [Invited]

42. **(2013)** "Using COCA and WordAndPhrase-Academic for teaching and learning Academic English". Lithuanian Applied Linguistics Association. Vilnius, Lithuania. [Keynote speaker]

43. **(2013)** "Using an advanced interface for Google Books to examine a wide range of historical changes". International Conference on Historical Linguistics. Univ. Oslo, Norway.

44. **(2013)** "A Corpus-Based Study of Lexical Developments in Early and Late Modern English." Early Modern Digital Agendas workshop. Folger Library, Washington. D.C. [Invited]

45. **(2013)** "Comparing Google Books (Standard) and Google Books (BYU/Advanced)". 'What Can We Do With 500 Billion Words?' workshop. Univ. Indiana. [Invited]

46. **(2013)** "Finally: Powerful yet simple comparisons of a wide range of phenomena in American and British English." American Association for Corpus Linguistics (AACL). San Diego State Univ.

47. **(2013)** "The new Academic Vocabulary List" American Association for Corpus Linguistics (AACL). San Diego State Univ. (Presented with Dee Gardner)

48. **(2012)** Second Language Research Forum. Univ Pittsburgh / Carnegie Mellon Univ. [Invited workshop]

49. **(2012)** "Using the Corpus of Contemporary American English (COCA) to gain insights into academic writing". Third Tamkang International Conference on Second Language Writing. Tamkang University, Taiwan. [Keynote speaker]

50. **(2012)** "The 155 billion word Google Books corpus: Can it be used for serious research on diachronic syntax?" ICAME (International Computer Archive of Medieval and Modern English). Univ Leuven, Belgium.

51. **(2011)** "The 400 Million Word Corpus of Historical American English (1810-2009) and the 155 Billion Word Google Books (American English) Corpus: New Tools for Research on Late Modern English". Middle and Modern English Corpus Linguistics. Osaka, Japan. [Keynote speaker]

52. **(2011)** "A Comparison of Google Books and the Corpus of Historical American English (COHA): Researching Lexical, Morphological, Syntactic, and Semantic Changes in American English". American Association for Corpus Linguistics. Georgia State University.

53. **(2011)** Empirical Methods in Cognitive Linguistics Workshop (EMCL-5.5). Univ. Chicago. [Invited speaker]

54. **(2011)** "Change then and change now: Mapping linguistic changes in English with the Corpus of Historical American English and the Corpus of Contemporary American English". Exploring the Boundaries and Applications of Corpus Linguistics. Univ. Alabama. [Keynote speaker]

55. **(2011)** "On the use of phrasal verbs with up". Exploring the Boundaries and Applications of Corpus Linguistics. Univ. Alabama. (Presented with Betsy Snider)

56. **(2011)** "The development of come and go + ADJ (come true/ go crazy) in American English: looking at historical shifts in semantic prosody". Exploring the Boundaries and Applications of Corpus Linguistics. Univ. Alabama. (Presented with Cate Mumford)

57. **(2010)** "The 400 million word Corpus of Historical American English (1810-2009): A new tool for in-depth research on Late Modern English". Late Modern English Conference. Univ. Sheffield, England.

58. **(2010)** "Why corpus architecture matters (or: why a corpus is much more than just a balanced collection of texts)". Rede Corpus workshop. Univ. Santiago de Compostela, Spain. [Invited speaker]

59. **(2010)** "Using the Corpus of Contemporary American English and the Corpus of Historical American English to study historical change and current variation in English syntax". American Association of Teachers of English Grammar. Brigham Young University, Provo, UT. [Keynote speaker]

60. **(2010)** "The 400 million word Corpus of Historical American English (1810-2009): A new tool for in-depth research on Late Modern English". International Conference of English Historical Linguistics. Pecs, Hungary.

61. **(2010)** "The Frequency Dictionary of American English: Word Sketches, Collocates, and Thematic Lists". Second International Symposium on Lexicography and L2 Teaching & Learning. Chongqing, China. [Keynote speaker]

62. **(2009)** "Examining recent syntactic shifts with the Corpus of Contemporary American English". ICAME (International Computer Archive of Medieval and Modern English). Lancaster Univ, England.

63. **(2009)** "The Corpus of Contemporary American English (385+ Million Words, 1990-present)". Conference on "Practical Applications in Language and Computers (PALC)". Univ. Lodz, Poland. [Keynote speaker]

64. **(2009)** "Examining genre-based variation and recent historical shifts with the Corpus of Contemporary American English". Conference on "Design and Delivery of Online Corpora". Univ. Glasgow, Scotland. [Keynote speaker]

65. **(2009)** "The 385+ Million Word Corpus of Contemporary American English (1990-present): A new tool for examining language variation and change". Digital Humanities 2009. Univ. Maryland.

66. **(2009)** "Large text archives and structured corpora: Case studies from syntactic shifts in twentieth-century American English". 19th International Conference on Historical Linguistics. Nijmegen, Netherlands.

67. **(2009)** Workshop on Online Corpora. Univ. Bergen, Norway. [Invited presentation]

68. **(2009)** American Association for Corpus Linguistics. Univ. Alberta. [Keynote speaker] "The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English"

69. **(2008)** "The 360 million word "BYU Corpus of American English (1990-2007)". ICAME (International Computer Archive of Medieval and Modern English). Univ. Zurich.

70. **(2007)** "Investigating Recent Linguistic Shifts with a New 100+ Million Word Corpus of American English from the 1900s. SHEL 5 (Studies in the History of the English Language". Univ. Georgia.

71. **(2007)** "From the Corpus del español to the Corpus do portugues (and back again): Evolving architectures for historical corpora". Colloquium on Ibero-Romance Historical Corpora. Univ. Balearic Islands, Spain. [Keynote speaker]

72. **(2007)** "A 100+ Million Word Corpus of American Magazines, 1900-1999". ICAME (International Computer Archive of Medieval and Modern English). Stratford-upon-Avon.

73. **(2009)** Series of presentations on using online corpora. Univ. Murcia, Spain. [Invited presentation]

74. **(2006)** "A new, 37 million word, Web-based corpus of historical English". ICAME (International Computer Archive of Medieval and Modern English). Univ. of Helsinki.

75. **(2006)** "Competing architectures for large historical corpora". Workshop on Historical Text Mining (AHRC ICT Methods workshop). Univ. of Lancaster (England). [Keynote speaker]

76. **(2006)** "Towards a 250 Million Word Corpus of Historical English". Bringing Text Alive: The Future of Scholarship, Pedagogy, and Electronic Publication (Text Creation Partnership). U Michigan.

77. **(2006)** "Resolving Trade Name Legal Disputes through Corpus Research". American Association of Applied Corpus Linguistics. Northern Arizona U.

78. **(2006)** "Incorporating "meaning-based" searches into corpus architectures and interfaces". American Association of Applied Corpus Linguistics. Northern Arizona U.

79. **(2006)** "Size, speed, and annotation with historical corpora". Digital Historical Corpora - Architecture, Annotation, and Retrieval. Dagstuhl Int'l Conference and Research Center for Computer Science (#06491). [Invited speaker]

80. **(2005)** "Vocabulary Coverage in Spanish Textbooks: How Representative is It?" Conference on the Acquisition of Spanish and Portuguese as First and Second Languages. Penn State Univ.

81. **(2005)** "Relational databases as a robust architecture for the analysis of word frequency". AHRC ICT Methods Network: Expert Seminar on Linguistics: Word Frequency and Keyword Extraction. Univ. of Lancaster, England. [Invited speaker]

82. **(2005)** "A new interface for examining use of synonyms (and other related words) in the BNC." Corpus Linguistics 2005. Univ. of Birmingham, England.

83. **(2004)** "The frequency and distribution of [se] constructions in Spanish: a corpus and learner-based approach." 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages. Univ. of Minnesota. (With Timothy L. Face)

84. **(2004)** "Incorporating register variation into BNC queries: a relational database approach." Sixth International Conference on Teaching and Language Corpora. Granada, Spain.

85. **(2004)** "El uso del Corpus del Español y otros corpus para investigar la variación actual y los cambios históricos." Series of workshops presented at Sophia University (Tokyo, Japan). [Invited presentation]

86. **(2004)** "Creating and Using Corpora to Investigate Language Change and Variation." Department of Linguistics, Sophia University (Tokyo, Japan).

87. **(2004)** "El diseño y uso de los corpus grandes para investigar el cambio lingüístico y la variación actual." Kobe University, Japan.

88. **(2004)** "A joint BNC/WordNet database: the best of both worlds". The Fifth North American Symposium on Corpus Linguistics. Montclair State, NJ.

89. **(2004)** "A multi-dimensional analysis of register variation in Spanish.". The Fifth North American Symposium on Corpus Linguistics. Montclair State, NJ.

90. **(2004)** "A match made in corpus heaven: the BNC and WordNet in relational database form." 25th Conference of the International Computer Archive of Modern and Medieval English. Verona, Italy.

91. **(2004)** "The impact of phrasal forms in corpus-based vocabulary studies". AAAL 2004 (American Association of Applied Linguistics). Portland, OR.

92. **(2003)** "How much vocabulary is enough?: Insights from recent corpus-based studies". 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages. U New Mexico.

93. **(2003)** "A multidimensional analysis of register variation in Spanish". 7th Hispanic Linguistics Symposium. U New Mexico.

94. **(2003)** "Advanced research on syntactic and semantic change with the 100 million word, fully-annotated Corpus del Español.". 2nd Freiburg Workshop on Romance Corpus Linguistics. U Freiburg, Germany. September 2003.

95. **(2003)** "On the frequency, use, and omission of se: Evidence from the 100 million word Corpus del Español". 2nd Freiburg Workshop on Romance Corpus Linguistics. U Freiburg, Germany. September 2003.

96. **(2003)** "Annotation without lexicons". Corpus Linguistics 2003. Lancaster University, UK. March 2003.

97. **(2003)** "Relational n-gram databases as a basis for unlimited annotation on very large corpora". Workshop on the Shallow Processing of Large Corpora. Lancaster University, UK. March 2003.

98. **(2002)** "Using Relational Databases to Create Highly Searchable and Very Large Corpora". The Fourth North American Symposium on Corpus Linguistics. IUPUI, Indianapolis, IN.

99. **(2002)** "Student use of a 100 million word, fully annotated corpus of Spanish to model language variation and change". Fifth International Conference on Teaching and Language Corpora. Bertinoro, Italy.

100. **(2002)** " Un corpus anotado de 100.000.000 palabras del español histórico y moderno". SEPLN 2002 (Sociedad Española para el Procesamiento del Lenguaje Natural). (Univ. de Valladolid, Spain).

101. **(2002)** "Modeling Syntactic Change with the Fully Annotated, 100 Million Word 'Corpus del Español': Suppresion of se with Causative Verbs". Sixth Hispanic Linguistics Symposium. Univ. of Iowa.

102. **(2002)** "A Searchable, Fully-Annotated, 100 Million Word Corpus of Historical Spanish Texts". Kentucky Foreign Language Conference. Univ. of Kentucky.

103. **(2002)** "A 100 Million Word Corpus of Historical and Modern Spanish, Searchable by Grammatical Category, Lemma, and Related Words". XXX Romance Linguistics Symposium. Cambridge, England.

104. **(2002)** "How to Make Large Corpora both Fast and Highly Annotated". 5th Annual CLUK (Computational Linguistics in the UK) Research Colloquium. Leeds, England.

105. **(2001)** "Multimillion Word Online Corpora as a Tool for Language Learning". The Seventh Sloan-C International Conference on Online Learning: Emerging Standards of Excellence in Asynchronous Learning Networks. Orlando, FL.

106. **(2001)** "Virtually Unlimited Annotation on Very Large Corpora". IRCS Workshop on Linguistic Databases. Philadelphia, PA.

107. **(2001)** "Dialectal Variation and Diachronic Shifts with the "Preposition + Subject + Infinitive" Construction (para yo hacerlo)". Fifth Hispanic Linguistics Symposium. Univ. Illinois.

108. **(2001)** "Large Historical Corpora on the Web: Helping Students to Model Linguistic Change". The Third North American Symposium on Corpus Linguistics and Language Teaching. Boston, MA.

109. **(2000)** "Diachronic Shifts in Spanish Raising Constructions". 4th Hispanic Linguistics Symposium, Indiana University.

110. **(2000)** "Using Large Computer-Based Corpora as a Philological Tool:An Analysis of Five Medieval Spanish Bibles". 35th International Congress on Medieval Studies, Univ. Western Michigan, May 2000.

111. **(2000)** "Using Large Computer-Based Parallel Texts to Study Lexical (and Other) Changes from Old Spanish to Modern Spanish". Kentucky Foreign Language Conference, April 2000.

112. **(1999)** "Diachronic shifts in the interpretation of Spanish infinitival complements". Conference on Spanish Semantics and Pragmatics. Ohio State Univ.

113. **(1999)** "Languages as dialects and dialects as languages: explaining parallel syntactic shifts in Spanish and Portuguese". International Conference on Historical Linguistics (ICHL XIV). Univ. of British Columbia.

114. **(1999)** "Creating Multimillion Word Corpora from Web-based Newspapers". North American Symposium on Corpora in Linguistics and Language Teaching. Univ. of Michigan.

115. **(1999)** "Modeling syntactic change: evidence from computer-based studies of infinitival complements in Spanish and Portuguese". Linguistic Symposium on Romance Languages (LSRL 29). Univ. of Michigan.

116. **(1998)** "Using multimillion word corpora of historical and dialectal Spanish texts to teach 'Advanced Spanish Syntax'". *Teaching and Language Corpora 1998* conference. Oxford University, England.

117. **(1998)** "A Corpus-Based Analysis of Subject Raising in Historical and Modern Spanish". Univ. Texas-Austin Colloquium on Romance Linguistics. Austin, TX

118. **(1997)** "Using Large Computer-Based Corpora to Investigate Language Variation and Change". Deseret Language and Linguistics Symposium. Provo, UT.

119. **(1996)** "The Use of Large Computer-Based Corpus in Research and Teaching". Colloquium on Spanish Linguistics. Roanoke, VA.

120. **(1995)** "The Diachronic Evolution of Portuguese Clitic Climbing". Annual Meeting of the Modern Language Association (MLA), Chicago, IL.

121. **(1995)** "Exploring Foreign Language Resources on the Internet". A series of presentations and workshops given to Illinois K-12 teachers; organized under the auspices of the Illinois Council on the Teaching of Foreign Languages (ICTFL).

122. **(1994)** "A Corpus-Based Approach to Modern Spanish Clitic Climbing." American Association of the Teachers of Spanish and Portuguese Annual Meeting (AATSP). Philadelphia, PA.

123. **(1994)** "Parameters, Passives, and Parsing: Explaining Diachronic Shifts in Spanish and Portuguese". Parasession at the Chicago Linguistics Society (CLS). Chicago, IL.

124. **(1992)** "Explaining Diachronic Shifts in Spanish and Portuguese Causative Constructions". Linguistic Symposium on Romance Languages (LSRL XXII). El Paso, TX.

125. **(1991)** "A Diachronic Look at Infinitival Complements in Spanish and Portuguese." Modern Language Association (MLA). San Francisco, CA.

126. **(1991)** "Parameters in Diachronic Spanish and Portuguese Causative Constructions." South Central Modern Language Association (SCMLA). Dallas, TX.

127. **(1991)** "Parameters in the Development of Diachronic Infinitival Complements in Spanish and Portuguese." Language Association of the Southwest (LASSO). Austin, TX.

128. **(1991)** "Towards a Unified Account of Diachronic Spanish Clitic Placement." Linguistic Symposium on Romance Languages (LSRL XXI). Santa Barbara, CA.

129. **(1990)** "Functional-Typological Explanations for Diachronic Shifts in Spanish Clitic Placement." "Explanation in Historical Linguistics" conference. Univ. Wisconsin-Milwaukee

---

UNIVERSITY AWARDS

Brigham Young University (2003-present)

- 2017. Creative Works Award. (One awarded each year at BYU, for projects and resources that have had impact throughout the US and the world).

- 2015-2016. Karl G. Maeser Research and Creative Arts Award. (One of two awarded at BYU (from among 1,500 faculty members), in recognition of excellence in research)

- 2013-2017. Humanities Center fellow. (Once of six in the College of Humanities).

- 2010-2012. College professorship (One of five in the 200+ faculty College of Humanities).

- 2008. Barker Lectureship (One scholar in language/linguistics selected from among the faculty in the College of Humanities)

Illinois State University (1992-2003)

- 2001-02. One of three mentors for the Distance Learning Training Program.

- 2000. Grant from the "Extended University" to develop online course: "Variation in Spanish Syntax".

- 1999. Grant from the "Extended University" to develop online course: "History of the Spanish Language".

- 1997. Grant from the Provost's Office to develop online course: "Foreign Language Resources on the Internet".

- 1997. Teaching Initiative Award (One of seven awarded at the university)

- 1997. "Faculty Fellow" award from the Center for the Advancement of Teaching.

- 1995. Research Initiative Award (One of seven awarded at the university).

- 1996. Office of Instructional Technology Grant to develop a course to be taught entirely by means of the Internet

- 1996. "Best of Illinois" award from the 1200+ member Illinois Council on the Teaching of Foreign Languages (ICTFL) (One award given each year)

TEACHING

Brigham Young University (2003-present)

- Corpus Linguistics (Ling 485)

- Creating Corpora (Ling 595, with Earl Brown)

- Corpus-based Approach to the History of American English (ELang 495: Senior Seminar)

- Historical and Comparative Linguistics (Ling 450)

- Empirical Methods in English Linguistics (ELang 273)

- English Grammar (ELang 325)

- Visual Basic (CHum 287)

Graduate Courses (Illinois State University)

- History of the Spanish Language (421.15)

- Variation in Spanish Syntax (360.15)

- Foreign Language Resources on the Internet (389.41)

- Descriptive and Comparative Grammar [Functional/Typological Syntax] (485.11)

- Seminar in Spanish Linguistics [Approaches to Spanish Syntax] (425.15)
- Advanced Spanish Syntax (310.15)

Undergraduate Courses (Illinois State University)

- Advanced Spanish Grammar (213.15)
- Advanced Spanish Composition
- Advanced Spanish Conversation (214.15)
- Third Semester Spanish (115.15)
- First Semester Spanish (111.15)

Graduate Course (Visiting Professor, Brigham Young University, Summer 1995)

- Spanish Morphosyntax (625, Graduate Seminar)

CONSULTANCIES (LEGAL)

- 2004-05. Expert witness for Wilmer Cutler Pickering Hale and Dorr LLP (Washington DC) in a case involving the generic status of a product name in Latin America
- 2008. Expert witness for Wilmer Cutler Pickering Hale and Dorr LLP (Washington DC) in a case involving the generic status of a product name in Latin America (different lawsuit)