

Literatur zur Korpuslinguistik

In diesem Bereich ist eine thematisch gegliederte Liste mit hilfreicher Literatur zum Thema Korpora und Korpuslinguistik zu finden.

- [Korpora und Korpuslinguistik: Überblick, Einführungen, Methodik](#)
- [Publikationen zu bekannten Korpora](#)
- [Korpora und gesprochene Sprache](#)
- [Korpora und Varietätenlinguistik](#)
- [Korpora und Soziolinguistik](#)
- [Korpora und Morphologie](#)
- [Korpora und Syntax](#)
- [Korpora und Sprachwandel](#)
- [Korpora und Internetkommunikation](#)
- [Korpora und kognitive Linguistik](#)
- [Statistische Analysemethoden für die Korpuslinguistik](#)

Korpora und Korpuslinguistik: Überblick, Einführungen, Methodik

Bilger, Mireille (1996): „Corpus de Portugais et d'Espagnol“. In: Recherches sur le français parlé 1/2, 124-130.

Chiari, Isabella (2007): Introduzione alla linguistica computazionale. Bari: Laterza (Percorsi, 99).

Dietrich, Wolf / Hoinkes, Ulrich / Roviró, Bàrbara / Warnecke, Matthias (eds.) (2006): Lexikalische Semantik und Korpuslinguistik. Tübingen: Narr (Tübinger Beiträge zur Linguistik, 490).

Freddi, Maria (2014): Linguistica dei corpora. Roma: Carocci (Bussole, 490).

Gerstenberg, Annette (2013): Arbeitstechniken für Romanisten. Eine Anleitung für den Bereich Linguistik. Berlin / Boston: De Gruyter [speziell Kap. 6: „Benutzung von Korpora und Datenbanken“, 7: „Sprachaufnahmen“, 8: „Die Erstellung eines eigenen Arbeitskorpus“].

Habert, Benoît (2009): Construire des bases de données pour le français. Paris: Ophrys (Collection L'essentiel français).

HSK 29.1 = Lüdeling, Anke / Kytö, Merja (eds.) (2008): Corpus Linguistics. An International Handbook. Berlin / New York: De Gruyter.

Kilgariff, Adam (2012): „Getting to know your corpus“. In: Petr Sojka, Ales Horák, Ivan Kopecek, Karel Pala (eds.): Text, Speech and Dialogue. 15th International Conference, TSD 2012, Brno, Czech Republic, September 3-7, 2012, Proceedings. Berlin / Heidelberg: Springer (Lecture Notes in Computer Science, 7499), 3–15.

Krefeld, Thomas / Lücke, Stephan / Von Ehrlich, Isabel (2014): „Digitalianistica. Die italienische Philologie unterwegs in die digital humanities“. In: Italienisch 72, 52-70. [Download](#)

Lemnitzer, Lothar / Zinsmeister, Heike (eds.) (2006): Korpuslinguistik. Eine Einführung. Tübingen: Narr.

O'Keefe, Anne / McCarthy, Michael (eds.) (2010): The Routledge handbook of corpus linguistics. New York: Routledge.

Pusch, Claus D. (2007): „Les corpus de linguistique romane en pays germanophones. Bilan et perspectives“. In: Revue Française de Linguistique Appliquée 12/1, 111-124.

Pusch, Claus D. (2014): „Les corpus romans contemporains“. In: Klump, A. / Kramer, J. / Willems, A. (eds.): Manuel des langues romanes. Berlin: De Gruyter (Manuals of Romance Linguistics, 1), 173-195.

Schmidt, Thomas / Wörner, Kai (eds.) (2012): Multilingual corpora and multilingual corpus analysis. Amsterdam / Philadelphia: Benjamins (Hamburg studies on multilingualism, 14).

Williams, Geoffrey (ed.) (2005): La linguistique de corpus. Rennes: Presses universitaires de Rennes.

Perkuhn, Rainer / Keibel, Holger / Kupietz, Marc (2012): Korpuslinguistik. Paderborn: Fink (LIBAC, 3433).

Publikationen zu bekannten Korpora

Al, Bernard P. F. (1980): „Sur la richesse lexicale du Corpus d'Orléans. Contribution à l'analyse sociolinguistique d'un vocabulaire oral". In: Alkemade, Dick J. van (ed.): *Linguistic studies offered to Berthe Siertsema*. Unter Mitarbeit von Berthe Siertsema. Amsterdam: Rodopi (Costerus, New series, 25), 315–321.

Baroni, Marco / Bernardini, Silvia / Ferraresi, Adriano / Zanchetta, Eros (2009): „The [WaCky](#) Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora". In: *Language Resources and Evaluation* 43, 209–226. [Download](#)

Bacelar do Nascimento, M. F. (2000): "[Corpus de Référence du Portugais Contemporain](#)". In: BILGER, M. (ed.) *Corpus, Méthodologie et Applications Linguistiques*, Paris, H. Champion et Presses Universitaires de Perpignan (2000), 25-30. [Download](#)

Bellini, Daniele / Schneider, Stefan (2006): „Spoken Italian online: The [Banca dati dell'italiano parlato \(BADIP\)](#)". In: Bernhard Kettemann, Georg Marko (eds.). *Planing, Gluing and Painting Corpora. Inside the Applied Corpus Linguist's Workshop*. Frankfurt am Main: Lang, 13-26.

Cresti, Emanuela/Moneglia, Massimo (eds.) (2005): [C-ORAL-ROM](#). Integrated Reference Corpora for Spoken Romance Languages. Amsterdam / Philadelphia: Benjamins.

De Mauro, Tullio / Mancini, F. / Vedovelli, M. / Voghera, M. (1993): [Lessico di frequenza dell'italiano parlato](#). Mailand: ETAS.

Durand, Jacques / Laks, Bernard / Lyche, Chantal (2009).Le projet [PFC](#): une source de données primaires structurées. In: lid. (eds): *Phonologie, variation et accents du français*. Paris: Hermès, 19-61.

Eisele, Andreas / Chen, Yu (2010): "[MultiUN: A Multilingual Corpus from United Nation Documents](#)".In: Tapias, Daniel et al. (eds.): *Proceedings of the Seventh conference on International Language Resources and Evaluation*. La Valletta, Malta, European Language Resources Association (ELRA), 2868-2872. [Download](#)

Généreux, Michel, Iris Hendrickx, Amália Mendes (2012): "Introducing the [Reference Corpus of Contemporary Portuguese On-Line](#)". In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation - LREC 2012, Istanbul, 21-27 de Maio de 2012*, 2237-2244.

Kaminskaia, Svetlana / Poiré, François (2012): „Prosodie du français laurentien en milieu minoritaire : le corpus Windsor". In: Anne Catherine Simon (ed.): *La variation prosodique régionale en français*. Louvain-La-Neuve: De Boeck (Champs linguistiques: Recueils), 159–178.

Koehn, Philipp (2005): [Europarl](#): A Parallel Corpus for Statistical Machine Translation, MT Summit 2005. [Download](#)

MacWhinney, B. (1994): The [CHILDES](#) Project: Tools for analyzing Talk. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pusch, Claus D. (2000): „Das Corpus Occitano-Gascon als Beispiel multimedialer Sprachdatenaufbereitung". In: Rieger, A. (ed.): *Okzitanistik, Altokzitanistik und Provenzalistik. Geschichte und Auftrag einer europäischen Philologie*. Frankfurt a. M. et al.: Lang, 43-56.

Reese, Samuel et al. (2010): "[Wikicorpus](#): A Word-Sense Disambiguated Multilingual Wikipedia Corpus". In: *Proceedings of 7th Language Resources and Evaluation Conference (LREC'10)*. La Valleta, Malta. [Download](#)

Rossini Favretti, Rema (1998): "Using multilingual parallel corpora for the analysis of legal language: the [Bononia Legal Corpus](#)". In: *Translation Equivalence. Proceedings of the Third European Seminar*, (W .Teubert., E. Tognini Bonelli , N. Volz eds.), The TELRI Association - V., Institut für Deutsche Spräahe, The Tuscan Word Centre, 57-68. [Download](#)

Rossini Favretti, R. / Tamburini, F. / Martelli, E. (2007): "Words from [Bononia Legal Corpus](#)". In: *Text Corpora and Multilingual Lexicography* (W.Teubert ed.), John Benjamins, 11-30. [Download](#)

Schäfer, Roland (2015): "Processing and querying large web corpora with the [COW14](#) architecture". In: Proceedings of Challenges in the Management of Large Corpora (CMLC-3). [Download](#)

Wäschle, K., Simianer, P., Bertoldi, N., Riezler, S., and Federico, M. (2013). "Generative and Discriminative Methods for Online Adaptation in SMT". In: Proceedings of Machine Translation Summit XIV. Nice, France. [Download](#) / [PatTR-Korpus](#)

Zampieri, M. / Becker, M. (2013): [Colonia: Corpus of Historical Portuguese](#). In: ZSM Studien, Special Volume on Non-Standard Data Sources in Corpus-Based Research. Vol. 5. Shaker. [Download](#)

Korpora und gesprochene Sprache

Bellini, Daniele / Schneider, Stefan (2006): „Spoken Italian online: The Banca dati dell'italiano parlato (BADIP)". In: Bernhard Kettemann, Georg Marko (eds.). Planing, Gluing and Painting Corpora. Inside the Applied Corpus Linguist's Workshop. Frankfurt am Main: Lang, 13-26.

Baude, Olivier (2006): Corpus Oraux. Guide des bonnes pratiques. Orléans: Presses Universitaires d'Orléans / CNRS Éditions.

Boisvert, Lionel / Laurendeau, Paul (1988): „Répertoire des corpus québécois de langue orale". In: Revue québécoise de linguistique 17/2, 241–261. Online verfügbar unter: <http://www.erudit.org/revue/rql/1988/v17/n2/602634ar.pdf>.

Blanche-Benveniste, Claire (2002): „Réflexions sur les transcriptions de corpus de français parlé". Revue Parole 22/23/24, 91-118.

Carruthers, Janice (2008): „Annotating an oral corpus using the Text Encoding Initiative. Methodology, problems, solutions". In: Journal of French Language Studies 18/1, 103–119.

Durand, Jacques / Laks, Bernard / Lyche, Chantal (2002): La phonologie du français contemporain: usages, variétés et structure. In: Pusch, Claus / Raible, Wolfgang (eds.): Romanistische Korpuslinguistik- Korpora und gesprochene Sprache. Tübingen: Narr, 93-106.

MacWhinney, B. (1994): The CHILDES Project: Tools for analyzing Talk. Hillsdale, NJ: Lawrence Erlbaum Associates.

Murphy, Bróna (2010): Corpus and sociolinguistics. Investigating age and gender in female talk. Amsterdam: Benjamins (Studies in corpus linguistics, 38).

Pusch, Claus D. (2006): „Corpora of spoken discourse". In: Brown, K. (ed.): Encyclopedia of Language and Linguistics, 2. Auflage, Bd. 3. Amsterdam et al.: Elsevier, 226-230.

Pusch, Claus D. / Raible, Wolfgang (2002): Romanistische Korpuslinguistik: Korpora und gesprochene Sprache. Tübingen: Narr (ScriptOralia, 126).

Rühlemann, Christoph (2013): Narrative in English Conversation. A Corpus Analysis of Storytelling. Cambridge: Cambridge University Press.

Korpora und Varietätenlinguistik

Boisvert, Lionel / Laurendeau, Paul (1988): „Répertoire des corpus québécois de langue orale". In: Revue québécoise de linguistique 17/2, 241–261. Online verfügbar unter: <http://www.erudit.org/revue/rql/1988/v17/n2/602634ar.pdf>.

Diwersy, Sascha / Evert, Stefan / Neumann, Stella (2014): „A semi-supervised multivariate approach to the study of language variation". In: Benedikt Szmrecsanyi, Bernhard Wälchli (eds.): Aggregating dialectology, typology, and register analysis. Linguistic variation in text and speech. Berlin: De Gruyter (Linguae et litterae, 28).

Kaminskaia, Svetlana / Poiré, François (2012): „Prosodie du français laurentien en milieu minoritaire : le corpus Windsor". In: Anne Catherine Simon (ed.): La variation prosodique régionale en français. Louvain-La-Neuve: De Boeck (Champs linguistiques: Recueils), 159–178.

Pusch, Claus D. (2000): „Das Corpus Occitano-Gascon als Beispiel multimedialer Sprachdatenaufbereitung". In: Rieger, A. (ed.): Okzitanistik, Altokzitanistik und Provenzalistik. Geschichte und Auftrag einer europäischen Philologie. Frankfurt a. M. et al.: Lang, 43-56.

Korpora und Soziolinguistik

Baker, Paul (2010): Sociolinguistics and Corpus Linguistics. Edinburgh: Edinburgh University Press (Edinburgh Sociolinguistics).

Murphy, Bróna (2010): Corpus and sociolinguistics. Investigating age and gender in female talk. Amsterdam: Benjamins (Studies in corpus linguistics, 38).

Korpora und Morphologie

Baayen, Rolf H. (2009): „Corpus Linguistics in Morphology: Morphological Productivity“. In: Lüdeling, Anke / Kytö, Merja (eds.): Corpus Linguistics. An International Handbook. HSK 29.2. Berlin / New York: De Gruyter, 899–919.

Fradin, Bernard / Dal, Georgette / Grabar, Natalia / Lignon, Stéphanie / Namer, Fiametta / Tribout, Delphine / Zweigenbaum, Pierre (2008): „Remarques sur l'usage de corpus en morphologie“. In: Langages 171 (3), 34–59.

Hathout, Nabil / Montermini, Fabio / Tanguy, Ludovic (2008): „Extensive data for morphology: using the World Wide Web“. In: French Language Studies 18, 67–85.

von der Grün, Angela (1999): Wort-, Morphem- und Allomorphhäufigkeit in domänenspezifischen Korpora des Deutschen. Friedrich-Alexander-Universität Erlangen-Nürnberg. Erlangen (CLUE-Arbeitsberichte, 2/1).

Korpora und Syntax

Mensching, Guido (2005): „Variación sintáctica, lingüística de corpus y gramática generativa: Teorías, métodos y problemas“. In: Gabriele Knauer (ed.): Variación sintáctica en español. Tübingen: Niemeyer, S. 13–33.

Selig, Maria (2008): „Geschichte, Variation, Wandel. Sprachwandel und historische Korpora“. In: Elisabeth Stark, Roland Schmidt-Riese, Eva Stoll, Wulf Oesterreicher (eds.): Romanische Syntax im Wandel. Tübingen: Narr, 67–86.

Korpora und Sprachwandel

Selig, Maria (2008): „Geschichte, Variation, Wandel. Sprachwandel und historische Korpora“. In: Elisabeth Stark, Roland Schmidt-Riese, Eva Stoll, Wulf Oesterreicher (eds.): Romanische Syntax im Wandel. Tübingen: Narr, 67–86.

Taavitsainen, Irma / Jucker, Andreas H. / Tuominen, Jukka: Diachronic corpus pragmatics (Pragmatics & beyond new series, 243).

Korpora und Sprachenlernen

Aston, Guy (ed.) (2001): Learning with corpora. Bologna, Houston, TX: CLUEB (Biblioteca della Scuola superiore di lingue moderne per interpreti e traduttori, Forlì, 29).

Korpora und Internetkommunikation

Barbera, Manuel / Corino, Elisa / Onesti, Cristina (eds.) (2007): Korpora e linguistica in rete. Perugia: Guerra.

Baroni, Marco / Bernardini, Silvia / Ferraresi, Adriano / Zanchetta, Eros (2009): „The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora“. In: Language Resources and Evaluation 43, 209–226.

Hathout, Nabil / Montermini, Fabio / Tanguy, Ludovic (2008): „Extensive data for morphology: using the World Wide Web“. In: French Language Studies 18, 67–85.

Korpora und kognitive Linguistik

Stefanowitsch, Anatol (2006): *Corpus-based approaches to metaphor and metonymy*. Berlin: De Gruyter.

Stefanowitsch, Anatol (2008): *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*. Berlin/New York: De Gruyter.

Statistische Analysemethoden für die Korpuslinguistik

Baayen, Rolf H. (2008): *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: University Press.

Baayen, Rolf H. (2009): „Corpus Linguistics in Morphology: Morphological Productivity“. In: Lüdeling, Anke / Kytö, Merja (eds.): *Corpus Linguistics. An International Handbook*. HSK 29.2. Berlin / New York: De Gruyter, 899–919.

Bergenholtz, Henning / Mugdan, Joachim (1989): „Korpusproblematik in der Computerlinguistik: Konstruktionsprinzipien und Repräsentativität“. In: Bátori, István S. / Lenders, Winfried / Putschke, Wolfgang (eds.): *Computational Linguistics. An International Handbook on Computer Oriented Language Research and Applications*. HSK 4. Berlin / New York: De Gruyter, 141–149.

Gries, Stefan Thomas (2008): *Statistik für Sprachwissenschaftler*. Göttingen: Vandenhoeck & Ruprecht (Studienbücher zur Linguistik, 13).