

Exploring Sound Change Over Time: A Review of Computational and Human Perception

Siqi He

Shanghai Jiao Tong University
University of Heidelberg
hesiqiid@sjtu.edu.cn

Wei Zhao

University of Aberdeen
wei.zhao@abdn.ac.uk

Abstract

Computational and human perception are often considered separate approaches for studying sound changes over time; few works have touched on the intersection of both. To fill this research gap, we provide a pioneering review contrasting computational with human perception from the perspectives of methods and tasks. Overall, computational approaches rely on computer-driven models to perceive historical sound changes on etymological datasets, while human approaches use listener-driven models to perceive ongoing sound changes on recording corpora. Despite their differences, both approaches complement each other on phonetic and acoustic levels, showing the potential to achieve a more comprehensive perception of sound change. Moreover, we call for a comparative study on the datasets used by both approaches to investigate the influence of historical sound changes on ongoing changes. Lastly, we discuss the applications of sound change in computational linguistics, and point out that perceiving sound change alone is insufficient, as many processes of language change are complex, with entangled changes at syntactic, semantic, and phonetic levels.

1 Background

There has been ongoing scholarly interest in sound change over time for decades. A popular historical sound change is the Great Vowel Shift (Lass, 1992), where the long vowel [i:] in Middle English shifted to a diphthong /aɪ/ in Modern English for example. It took place over time from the 15th to 18th centuries, and greatly changed the English vowel system. Other examples include the loss of voiceless velars like [ç] in Modern English (Dobson, 1968), reduction of consonant clusters like [kn] → [n] (Turville-Petre and Burrow, 2020), and vowel reduction in unstressed syllables (Minkova, 2013). Many ongoing sound changes took place in the 20th century. For instance, in American

regional dialects, a notable shift such as [ʌ] → [ɛ] in the vowel system occurred in the Northern Cities around the mid-20th century (Wolfram and Schilling, 2016).

While many works proposed computational approaches to perceive historical sound changes (Mielke, 2008; Dekker, 2018; Boldsen and Paggio, 2022) and others suggested using the listener-driven model to perceive ongoing sound changes (Janson, 1983; Sanker, 2018a; Quam and Creel, 2021), few works explored the intersection of both computational and human perception. The benefits of doing so can be substantial. Firstly, computational approaches perceive historical sound change by analyzing IPA transcriptions in etymological datasets, but these datasets lack acoustic features that human listeners/speakers can produce and perceive. Secondly, human perception observes ongoing changes through participation surveys over recording corpora, which lacks the considerations of acoustic and phonetic alignments between speakers that computational approaches can produce and perceive. Lastly, connections between etymological datasets and recording corpora are little explored; by combining both, one could conduct a comparative analysis of historical and ongoing sound changes, e.g., examining how historical changes impact ongoing changes. Thus, there is a need for a comparative review of computational and human perception.

In this work, we aim to fill the gap between two distant perception of sound change over time, with computational models on one hand and human observation on the other. To achieve this, we first review the tasks and methods from each perspective, and then present a unified view that combines both perception to explore sound change. Moreover, we discuss the connections of sound change to semantic and syntactic change, as well as the applications of sound change in computational linguistics.

2 Computational Perception

Sound change detection. Boldsen and Paggio (2022) connected semantic change detection with sound change detection and argued that diachronic distributional embeddings used for semantic change detection can track historical sound change. For lexical semantic change, diachronic word embeddings are guided by the distributional hypothesis suggesting that words that occur in similar contexts appear to have similar meanings. Interestingly, this idea also applies to phonology. Previous works showed that phonemes occurring in similar phonetic contexts likely belong to the same phonological class, demonstrating the applicability of the distributional hypothesis to phoneme embeddings (Mielke, 2008; Silfverberg et al., 2018).

Boldsen and Paggio (2022) proposed using phoneme embeddings trained on a historical Danish corpus to track sound changes over time. Their approach compared the embeddings of a phoneme pair across different periods to observe sound changes. For instance, [p] → [b] is observed when the distance between the phoneme embeddings of [p] and [b] becomes smaller over time. The results showed a decrease in distance between three phoneme embedding pairs: [p] and [b], [t] and [d], [k] and [g] over time, meaning that their approach recognized the phonetic changes from voiceless plosives to their voiced counterparts in Danish.

Phonetic alignment between cognate words. Phonetized cognate words consist of paired IPA transcriptions in two languages: either a proto-language and its descendant language or two descendant languages. Each transcription represents a sequence of phonemes. Translating a sequence of phonemes from one language to another can be framed as a machine translation task, as both execute a cross-lingual sequence-to-sequence task (Dekker, 2018; Fourrier and Sagot, 2020a).

For instance, Fourrier and Sagot (2020a) proposed using both statistical and neural machine translation models to perform phoneme-level translations between cognate words. The models investigated include Moses (Koehn et al., 2007) and MEDeA (Luong et al., 2015). The languages considered include Latin, Italian, and Spanish. For evaluation, the generated translations were compared to the ground-truths through BLEU (Papineni et al., 2002)—which calculates the overlap of n -grams phonemes between translations and the ground-truths. The results showed that the pho-

netic translations between cognate words from the proto-language to a descendant language (or from a descendant language to another) are much better than those from a descendant language to the proto-language. Moreover, the results demonstrated the superiority of the statistical model over the neural MT model on small datasets, whereas the neural model showed a greater ability to handle many-to-one mappings from various proto-forms to the same descendant form. It is important to note that the ground-truth translations were collected by automatically phonetizing cognate word pairs via Espeak (Duddington, 2007), and the automatic phonetization process is prone to errors, meaning that comparing generated translations with the ground truths may lead to inaccurate model assessment.

Markedness of phonemes. Markedness is a linguistic label separating common from less common phonemes in a phonological system. In English, the voiceless consonants [p], [t], and [k] are unmarked as they are more common compared to their voiced, marked counterparts [b], [d], and [g]. For vowels, peripheral high vowels such as [i] and [u] are marked while mid-central vowels like [ə] are unmarked (Jakobson, 1968; Haspelmath, 2006).

Ceolin and Sayeed (2019) proposed a probabilistic approach to model sound change by estimating the frequency of phonemes over time. Interestingly, they found that their approach could also recognize the markedness of a phoneme. Their approach was to estimate the frequency of each phoneme at a later time based on the frequencies of other phonemes observed at an earlier time through the split-merger process. The results showed that the unmarked phonemes at a later time appear to have higher frequencies compared to marked counterparts as postulated, meaning that their approach could separate unmarked from marked phonemes. We note that their approach considered neither phonetic nor acoustic features and was only evaluated on three phonemes in an artificial setup.

Sound convergence. Unlike historical sound change, which takes place gradually over centuries, sound convergence is a process of ongoing sound change through which speakers adjust their speech to align acoustically and phonetically with other speakers (Natale, 1975). Research showed that native speakers often phonetically converge to non-native speakers in interactive environments (Giles, 1973; Pardo, 2006; Babel, 2010; Yu et al., 2013). Recently, works by Lewandowski and Nygaard

(2018); Wagner et al. (2021) showed that sound convergence can also occur in non-interactive settings. For instance, Wagner et al. (2021) recruited 76 native Dutch speakers and a non-native speaker to read aloud a careful selection of words, and their speech was recorded. Importantly, although the native and non-native speakers have no interaction, the speech of non-native speakers was made available to the native speakers. This allows the native speakers to potentially adjust their speech. All the speech was transformed into acoustic features by using Praat (Boersma, 2011). Acoustic features (e.g., vowel and fricative duration) of the native speakers are compared to the non-native speaker by calculating the difference-in-distance score over these acoustic features. The positive score indicates convergence, otherwise divergence. The results showed that (a) the Dutch native speakers show sound convergence to the non-native speaker in the scenario where interaction is minimal and (b) the degree of sound convergence is affected by how much native speakers think the speech by the non-native speaker is native-like.

Despite being useful, relying on acoustic features to observe sound convergence has at least two limitations: Firstly, the convergence on the phonetic level is overlooked; for example, speakers who adjust their speech in terms of place and manner of articulation are not recognized as sound adaptation. Secondly, the outcome of sound convergence is affected by the quality of acoustic features—which relies on the recording quality and the efficacy of computer tools to extract these features from speech.

3 Human Perception

Perceptual similarity. The work by Goldinger (1998) introduced the perceptual similarity task that is concerned with how phonetic changes are received, processed and interpreted by listeners (Martin and Bunnell, 1981; Sanker, 2018b). This approach is known as the listener-driven model of sound change, where a group of listeners heard recordings of a speaker’s initial speech and the later speech (after the speaker listened to a target speech by another speaker). The listeners were then asked to determine which of the two recordings sounded closer to the target speech the speaker was exposed to.

Sound convergence. Wagner et al. (2021) employed the perceptual similarity task to study sound

convergence from native speech to non-native speech in Dutch. They recruited 16 listeners native in Dutch and asked them to perform the perceptual similarity task where the listeners heard participants’ initial and later speech and had to choose which production sounded more similar to that of the model speaker non-native in Dutch. Moreover, the listeners were asked to rate the model speaker’s speech in terms of how accented, comprehensible, and familiar it sounded. These ratings were included in the analyses to determine how they affected the degree of observed convergence. The results showed that the overall sound convergence score was slightly above the random chance, indicating a weakly perceived convergence in participants’ speech after the target speech was exposed to them. Secondly, they found that several speech samples showed more sound convergence than others. Moreover, they noted that perceived convergence was affected by how strongly the model speaker’s foreign accent was perceived.

Although human perception can observe ongoing sound changes, listeners may misperceive the acoustic and phonetic features of a speaker, resulting in incorrect judgments of sound changes (Babel and Johnson, 2010; Ohala, 2017; Sanker, 2018b).

4 A Unified Perspective

Computer-aided human perception. Using computational methods can partly automate and possibly refine the human perception process of sound changes. This is because doing so allows for observing subtle changes on phonetic and acoustic levels, such as vowel duration shift and nasal place assimilation that are sometimes not obvious to perceive by listeners. Additionally, combining computational and listener-driven methods would create a feedback loop where computational results could refine the human perception process and insights from listeners could be used to improve computational models.

Cross-studying etymological datasets and recording corpora. Etymological datasets contain phonetic transcriptions that reflect historical sound changes, which are commonly used for computational models to observe changes over centuries. In contrast, recording corpora are a database for listeners to perceive ongoing sound changes. Despite their different aims, it is intriguing to know the influence of historical sound changes on ongoing changes. A potential idea is to

start with collecting shared words in etymological datasets and recording corpora, and then inspect their phonetic similarity and difference. Note that unlike recording corpora, phonetic transcriptions do not include acoustic features; therefore, a comparative study of historical and ongoing sound changes at the acoustic level is not possible.

5 Discussions

5.1 Connections to Other Changes

While there have been many works on computational modeling of semantic and syntactic change (Hamilton et al., 2016; Schlechtweg et al., 2020; Ma et al., 2024a,b; Merrill et al., 2019; Krielke et al., 2022; Chen et al., 2024), they often lack connections to sound change. Such connections are crucial because many changes simultaneously affect multiple linguistic levels. A notable case is homograph where two words share the same spelling but have different meanings and pronunciations. Examples include “present” ([ˈprezənt] vs. [priˈzɛnt]) and “bow” ([bau] vs. [boʊ]). Another case is grammaticalization—a process that incurs semantic, syntactic and phonetic changes. For example, “going to” grammaticalizes into “gonna”, shifting from a verb to a future marker. This process changes the original meaning, impacts syntactic structure, and incurs phonetic reduction. To identify homographs and grammaticalization, it might be necessary to develop computational and human approaches to model/observe changes across multiple linguistic levels at once.

5.2 Applications in Computational Linguistics

Phylogenetic Inference. This task aims to reconstruct the evolutionary relationships among languages based on their shared linguistic features. For example, Proto-Indo-European, as the ancestral language, gives rise to many descendant languages within the Indo-European language groups such as Indo-Iranian, Germanic, and Celtic. Linguists construct a phylogenetic language tree by taking the ancestor language as the root and connecting it to descendant languages, based on the laws of sound changes over time (Hoenigswald, 1965). For instance, there exists a phoneme correspondence between High German [ts], Dutch [t], English [t], Swedish [t], and Icelandic [t], all of which are inherited from the proto-phoneme [*t] in their ancestry Proto-Germanic language group (where [*t] → [t] in High German). This phoneme correspondence

is one of many reasons that these languages are the descendants of Proto-Germanic.

However, computer-based language phylogenies for major language groups like Dravidian (Kolipakam et al., 2018), Sino-Tibetan (Sagart et al., 2019), and Indo-European (Heggarty et al., 2023) often rely on cognate sets from semantically aligned word lists across languages. Campbell and Poser (2008) questioned the use of cognate sets for phylogenetic inference, as meanings in cognate words might undergo changes over time, resulting in the instability of a phylogenetic tree. Other works proposed reconstructing phylogenetic language trees using sound correspondences between cognate words instead of lexical cognates (Chacon and List, 2016; Cathcart, 2019; Chang et al., 2023; Häuser et al., 2024). For instance, Häuser et al. (2024) presented a framework that first identifies phonetic alignment between cognate words using LingPy and then uses BMrBayes (Ronquist and Huelsenbeck, 2003) and RAxML-NG (Kozlov et al., 2019) to reconstruct phylogenetic trees. For evaluation, the generated phylogenetic trees are compared to the ground-truth Glottolog tree (Hammarström et al., 2019) by computing their topological distance via generalized quartet distance (Pompei et al., 2011). The results showed that sound-based phylogenetic trees underperform cognate-based counterparts, i.e., that cognate-based trees are topologically closer to the gold Glottolog tree. This might be attributed to the lack of consideration for borrowing. For instance, two languages might not be related, although the phoneme sequences of their cognate words could be similar. Loanword is the example, where phonemes are borrowed from a third, unrelated language, rather than inherited from the proto-phoneme.

Quality assessment of etymological datasets.

Etymological datasets are a crucial resource for phylogenetic inference, low-resource machine translation, and historical linguistic tasks. Many such datasets have been made available and are automatically generated from various data sources. For instance, EtymWordNet (De Melo, 2014) and CogNet (Batsuren et al., 2019) are derived from WordNet across hundreds of languages, while EtymDB 1.0 (Sagot, 2017) and 2.0 (Fourrier and Sagot, 2020b) are sourced from Wiktionary across over two thousand languages. However, the quality of these datasets remains unclear. Firstly, many datasets use a loose definition of cognacy to enlarge

data coverage. Secondly, the automatic processes used to generate these datasets are prone to errors. Therefore, there is a need to estimate the quality of these etymological datasets.

Wettig et al. (2012) proposed using the degree of phonetic alignment between cognate words as a measure of the internal consistency of an etymological dataset. They postulated that the more phonetically similar cognates words are, the better quality a dataset would be. For instance, the English word ‘house’ and the German word ‘Haus’ are phonetically equivalent [haʊz], implying that this cognate word pair is likely correct. To achieve this idea, they use the Minimum Description Length, a dynamic programming algorithm, to calculate the cost of an optimal phoneme-level alignment between cognate word pairs for the Uralic language group. The alignment operates on phonetic features such as plosive/fricative and labial/dental. The challenge arises from the fact that phonemes inherited from the proto-phoneme may undergo sound changes over time, resulting in phonemes in one language potentially different from another. For evaluation, the generated alignments were not compared to the ground-truths due to the lack of gold phoneme-level alignments. Instead, their approach was evaluated in three scenarios: compression rates, rules of correspondence and imputation.

Note that phoneme-level alignments were not compared against the ground-truths. Thus, the efficacy of the measure based on these alignments in estimating the quality of etymological datasets remains unclear. Moreover, their approach only considers one-to-one phoneme-level alignment and ignores one-to-many. In doing so, their approach could wrongly penalize correct cognate word pairs with one-to-many alignments, such as [kæt] in ‘cat’ and [katsə] in the German word ‘Katze’.

6 Conclusions

As two rarely connected disciplines, computational and human perception have their own interests, tasks and methods. However, we showed that these two perception benefit each other from the perspective of methods and datasets. Additionally, we showed that the applications of sound change are manifold in computational linguistics, including phylogenetic inference and quality assessment of datasets. Despite these positive aspects, we argue that a unified perception of multi-faceted change is crucial, as many changes are entangled across

phonetics, syntax and semantics.

Acknowledgements

We thank the anonymous reviewers for their thoughtful feedback that greatly improved the texts.

References

- Molly Babel. 2010. Dialect divergence and convergence in new zealand english. *Language in Society*, 39(4):437–456.
- Molly Babel and Keith Johnson. 2010. Accessing psycho-acoustic perception and language-specific perception with speech sounds. *Laboratory phonology*, 1(1):179–205.
- Khuyagbaatar Batsuren, Gábor Bella, Fausto Giunchiglia, et al. 2019. Cognet: A large-scale cognate database. In *ACL 2019 The 57th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference*, pages 3136–3145. Association for Computational Linguistics.
- Paul Boersma. 2011. Praat: doing phonetics by computer [computer program]. <http://www.praat.org/>.
- Sidsel Boldsen and Patrizia Paggio. 2022. Letters from the past: Modeling historical sound change through diachronic character embeddings. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6713–6722, Dublin, Ireland. Association for Computational Linguistics.
- Lyle Campbell and William J Poser. 2008. Language classification. (*No Title*).
- Chundra Cathcart. 2019. Gaussian process models of sound change in Indo-Aryan dialectology. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 254–264, Florence, Italy. Association for Computational Linguistics.
- Andrea Ceolin and Ollie Sayeed. 2019. Modeling markedness with a split-and-merger model of sound change. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 67–70.
- Thiago Costa Chacon and Johann-Mattis List. 2016. Improved computational models of sound change shed light on the history of the tukanoan languages. *Journal of Language Relationship*, 13(3-4):177–204.
- Kalvin Chang, Nathaniel Robinson, Anna Cai, Ting Chen, Annie Zhang, and David Mortensen. 2023. Automating sound change prediction for phylogenetic inference: A tukanoan case study. In *Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change*, pages 129–142, Singapore. Association for Computational Linguistics.

- Yanran Chen, Wei Zhao, Anne Breitbarth, Manuel Stoeckel, Alexander Mehler, and Steffen Eger. 2024. Syntactic language change in english and german: Metrics, parsers, and convergences. *arXiv preprint arXiv:2402.11549*.
- Gerard De Melo. 2014. Etymological wordnet: Tracing the history of words. In *LREC*, pages 1148–1154.
- Peter Dekker. 2018. Reconstructing language ancestry by performing word prediction with neural networks. *Master. Amsterdam: University of Amsterdam*.
- Eric John Dobson. 1968. English pronunciation, 1500–1700. (*No Title*).
- Jonathan Duddington. 2007. 2015. espeak text to speech.
- Clémentine Fourier and Benoît Sagot. 2020a. Comparing statistical and neural models for learning sound correspondences. In *LT4HALA 2020-First Workshop on Language Technologies for Historical and Ancient Languages*.
- Clémentine Fourier and Benoît Sagot. 2020b. Methodological aspects of developing and managing an etymological lexical resource: Introducing etymdb 2.0. In *LREC 2020-12th Language Resources and Evaluation Conference*.
- Howard Giles. 1973. Accent mobility: A model and some data. *Anthropological linguistics*, pages 87–105.
- Stephen D Goldinger. 1998. Echoes of echoes? an episodic theory of lexical access. *Psychological review*, 105(2):251.
- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.
- Harald Hammarström, Martin Haspelmath, Robert Forkel, and Sebastiaan Bank. 2019. Glottolog. version 4.0. *Max Planck Institute for the Science of Human History, Jena*.
- Martin Haspelmath. 2006. Against markedness (and what to replace it with). *Journal of linguistics*, 42(1):25–70.
- Luise Häuser, Gerhard Jäger, Johann-Mattis List, Taraka Rama, and Alexandros Stamatakis. 2024. Are sounds sound for phylogenetic reconstruction? In *Proceedings of the 6th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, pages 78–87, St. Julian’s, Malta. Association for Computational Linguistics.
- Paul Heggarty, Cormac Anderson, Matthew Scarborough, Benedict King, Remco Bouckaert, Lechosław Jocz, Martin Joachim Kümmel, Thomas Jügel, Britta Irlinger, Roland Pooth, et al. 2023. Language trees with sampled ancestors support a hybrid model for the origin of indo-european languages. *Science*, 381(6656):eabg0818.
- Henry M Hoenigswald. 1965. Language change and linguistic reconstruction.
- Roman Jakobson. 1968. *Child language: aphasia and phonological universals*. 72. Walter de Gruyter.
- Tore Janson. 1983. Sound change in perception and production. *Language*, 59:18.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions*, pages 177–180. Association for Computational Linguistics.
- Vishnupriya Kolipakam, Fiona M Jordan, Michael Dunn, Simon J Greenhill, Remco Bouckaert, Russell D Gray, and Annemarie Verkerk. 2018. A bayesian phylogenetic study of the dravidian language family. *Royal Society open science*, 5(3):171504.
- Alexey M Kozlov, Diego Darriba, Tomáš Flouri, Benoit Morel, and Alexandros Stamatakis. 2019. Raxml-ng: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21):4453–4455.
- Marie-Pauline Krielke, Luigi Talamo, Mahmoud Fawzi, and Jörg Knappen. 2022. Tracing syntactic change in the scientific genre: Two Universal Dependency-parsed diachronic corpora of scientific English and German. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4808–4816, Marseille, France. European Language Resources Association.
- Roger Lass. 1992. Phonology and morphology. *The Cambridge history of the English language*, 2:1066–1476.
- Eva M Lewandowski and Lynne C Nygaard. 2018. Vocal alignment to native and non-native speakers of english. *The Journal of the Acoustical Society of America*, 144(2):620–633.
- Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421, Lisbon, Portugal. Association for Computational Linguistics.

- Xianghe Ma, Dominik Schlechtweg, and Wei Zhao. 2024a. Presence or absence: Are unknown word usages in dictionaries? In *Proceedings of the 5th Workshop on Computational Approaches to Historical Language Change*, Bangkok. Association for Computational Linguistics.
- Xianghe Ma, Michael Strube, and Wei Zhao. 2024b. [Graph-based clustering for detecting semantic change across time and languages](#). In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1542–1561, St. Julian’s, Malta. Association for Computational Linguistics.
- James G Martin and H Timothy Bunnell. 1981. Perception of anticipatory coarticulation effects. *The Journal of the Acoustical Society of America*, 69(2):559–567.
- William Merrill, Gigi Stark, and Robert Frank. 2019. [Detecting syntactic change using a neural part-of-speech tagger](#). In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 167–174, Florence, Italy. Association for Computational Linguistics.
- Jeff Mielke. 2008. *The emergence of distinctive features*. Oxford University Press.
- Donka Minkova. 2013. *Historical phonology of English*. Edinburgh University Press.
- Michael Natale. 1975. Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills*, 40(3):827–830.
- John J Ohala. 2017. Phonetics and historical phonology. *The handbook of historical linguistics*, pages 667–686.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. [Bleu: a method for automatic evaluation of machine translation](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.
- Jennifer S Pardo. 2006. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4):2382–2393.
- Simone Pompei, Vittorio Loreto, and Francesca Tria. 2011. On the accuracy of language trees. *PloS one*, 6(6):e20109.
- Carolyn Quam and Sarah C. Creel. 2021. [Impacts of acoustic-phonetic variability on perceptual development for spoken language: A review](#). *Wiley interdisciplinary reviews. Cognitive science*, page e1558.
- Fredrik Ronquist and John P Huelsenbeck. 2003. Mr-bayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12):1572–1574.
- Laurent Sagart, Guillaume Jacques, Yunfan Lai, Robin J Ryder, Valentin Thouzeau, Simon J Greenhill, and Johann-Mattis List. 2019. Dated language phylogenies shed light on the ancestry of sino-tibetan. *Proceedings of the National Academy of Sciences*, 116(21):10317–10322.
- Benoît Sagot. 2017. Extracting an etymological database from wiktionary. In *Electronic Lexicography in the 21st century (eLex 2017)*, pages 716–728.
- Chelsea Sanker. 2018a. [A survey of experimental evidence for diachronic change](#). *Linguistics Vanguard*, 4.
- Chelsea Sanker. 2018b. A survey of experimental evidence for diachronic change. *Linguistics Vanguard*, 4(1):20170039.
- Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. 2020. [SemEval-2020 task 1: Unsupervised lexical semantic change detection](#). In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1–23, Barcelona (online). International Committee for Computational Linguistics.
- Miikka P Silfverberg, Lingshuang Mao, and Mans Hulden. 2018. Sound analogies with phoneme embeddings. *Society for Computation in Linguistics*, 1(1).
- Thorlac Turville-Petre and John Anthony Burrow. 2020. *A book of Middle English*. John Wiley & Sons.
- Mónica A Wagner, Mirjam Broersma, James M McQueen, Sara Dhaene, and Kristin Lemhöfer. 2021. Phonetic convergence to non-native speech: Acoustic and perceptual evidence. *Journal of Phonetics*, 88:101076.
- Hannes Wettig, Kirill Reshetnikov, and Roman Yangarber. 2012. Using context and phonetic features in models of etymological sound change. In *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*, pages 108–116.
- Walt Wolfram and Natalie Schilling. 2016. *American English: Dialects and variation*, 3 edition. Number 25 in Language in Society. Wiley Blackwell, Chichester, UK.
- Alan CL Yu, Carissa Abrego-Collier, and Morgan Sonderegger. 2013. Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and “autistic” traits. *PloS one*, 8(9):e74746.