Large Language Models and Minimalism: Theories, grammars, and computational modeling

Jason Ginsburg

Graduate School of Human and Environmental Studies, Kyoto University, Japan
< ginsburg.jasonrobert.2h@kyoto-u.ac.jp >

Chesi (*this issue*), responding to Piantadosi (2024), compares Large Language Models with work in Generative Linguistics. Chesi points out that syntactic tests exist for Large Language Models, but not for theories of Generative Linguistics, as well as that theories in Generative Linguistics lack proper formalization, which has led to Generative Linguistics becoming marginalized. In this paper, I take the position that Generative Linguistics develops theories of language, but Large Language Models are not theories. Also, while syntactic tests that examine the validity and scope of theories in Generative Linguistics could be useful, a number of large hurdles exist for their development.

KEYWORDS: large language models, theories, grammar, Minimalism, computer modeling.

1. Introduction

Large Language Models (LLMs) have been the focus of a great deal of attention in the media and in academia. This is most likely because they appear to be a huge leap forward in Artificial Intelligence. They can be used to create seemingly authentic language, and they may be able to pass some versions of the Turing Test (Turing 1950).[1] These models potentially have a wide variety of applications. Some applications may be helpful; e.g. responding to patients for health care (Ayers *et al.* 2023), helping with language learning (Klimova *et al.* 2024), increasing productivity on workplace writing tasks (Noy & Zhang 2023), etc. Some applications/ aspects of these models may be harmful; e.g. replacement of human jobs (Demirci *et al.* 2024), increase of plagiarism in academia and elsewhere (Chen *et al.* 2024; Kwon 2024), spread of authentic-looking misinformation (Raman *et al.* 2024; Spitale *et al.* 2023), use to commit fraud and deceive people (Park *et al.* 2024), misdiagnose illnesses (Barile *et al.* 2024), cause damage to the environment (Crawford 2024), etc.

Chesi (*this issue*), responding to a paper by Piantadosi (2024) that skewers Generative Linguistics, argues that Piantadosi is partially correct with respect to criticism of Generative Linguistics. In this paper, I focus on the following topics from Chesi's paper.[2] (Below, I use

Minimalist theories to refer to work in Minimalism generally following Chomsky (1995), which is the latest incarnation of work within the Generative Grammar framework, following a long tradition led by Noam Chomsky himself.)

Several claims from Chesi's paper:
A) LLMs are more observationally adequate than Minimalist theories, but they "lack explanatory adequacy" (p. 32). Compared to LLMs, Minimalist theories, on the other hand, may be superior with respect to descriptive adequacy.
B) Minimalist theories are unable to perform adequately in "comprehensive syntactic tests" (p. 5) and it is necessary to "adopt a modern approach to theory evaluation that relies on shared datasets and metrics" (p. 6). Furthermore, there is no "shared test/reference set" (p. 18) for Minimalist theories.
C) A problem for developing linguistic benchmarks for Minimalist theories and for testing how well Minimalist theories do on benchmarks is that Minimalism has not been properly formalized.
D) The lack of comprehensive linguistic benchmarks and the lack of a fully explicit theory has led to Minimalism becoming marginalized.

I attempt to address these issues in the following sections. In section 2, I address the issue of whether or not LLMs can be called theories and/or grammars, and how they compare with Minimalist theories. In section 3, I examine the issue of whether or not it is possible to create benchmarks for Minimalist theories.

2. LLMs as theories and grammars

Some researchers refer to LLMs as theories of language. Baroni (2023: 1) writes that "deep networks should be treated as theories making explicit predictions about the acceptability of linguistic utterances." Baroni further writes

> we can think of a deep net architecture, before any language-specific training, as a general theory defining a space of possible grammars, and of the same network trained on data from a specific language as a *grammar*, that is, a computational system that, given an input utterance in a language can predict whether the sequence is acceptable to an idealized speaker of the language. (Baroni 2023: 7)

Piantadosi (2024: 360) writes that LLMs "develop representations of key structures and dependencies", and that "[a]s argued by Baroni (2022), this means that language models should be treated as bona fide linguistic *theories*." He argues that "a space of possible theories is parameterized by the models and compared to data to find which theory is best in a formal sense (Baroni 2022: 360)."

A theory provides an explanation (typically a hypothesis) about some fact(s) of the natural world. Linguistic theories attempt to provide explanations for how language works. An LLM is a computer model that produces language output that is similar (although not identical) to the language input (its training data). Specifically, an LLM is a type of Generative AI which produces output that is similar to the input, and Generative AI can be used for tasks other than language. While there may be terminological issues regarding definitions of theories, I think that there are flaws to viewing an LLM as a theory. An LLM, as well as other neural net models, are not theories in the sense that they do not provide explanations for facts of the world. A computer model, such as an LLM, can be used to implement, test, and/or develop a theory, but an LLM is not a theory.

LLMs are engineering tools that produce language and there are theories behind the designs of LLMs, but these are not theories of language. The engineering methods used to construct LLMs do not take into account the structures of language. Rather, LLMs are designed to model the training data in optimal ways.

Assume, as Baroni and Piantadosi suggest, that an LLM produces language in a way that is similar to a human. Then the question arises of how it produces and 'understands' language. While it is possible to explain how an LLM is built, it is not clear exactly how the various parameters within an LLM are set to produce a given output. The inner workings of an LLM are often referred to as a black box, in that we do not know exactly what is going on within them (Dobson 2023, Fazi 2020, etc.). They can have billions of parameters that are connected in ways that we do not fully understand. Fazi (2021: 59) writes

> once a deep neural network is trained (or self-trained…), it can be extremely difficult to explain why it gives a particular response to some data inputs and how a result has been calculated. The strength of a deep neural network lies in its capacity to find non-linear patterns in large datasets and improve this extraction through iterative interactions.

An LLM finds patterns, but we do not necessarily know how it comes up with these patterns, and even if we were to develop theories

to explain these patterns, these patterns do not necessarily correspond to how humans use language.

Fox & Katzier (2024: 72) write "We might be impressed by an LLM generating a Shakespearean sonnet or by LLM activity correlating with data from brain imaging, but unless these observations bear on theory selection, they are not going to tell us much about underlying machinery." Linguists are interested in the underlying machinery of language. Looking at the output of an LLM is like looking at a corpus of language produced by humans (although there are likely some differences). Just because an LLM produces language that is generally identical to human language does not necessarily tell us anything unique about how language works.

Theories can be developed based on LLMs (and other neural networks). For example, Manning *et al.* (2020) examine how self-supervised neural nets represent linguistic structure. Lakretz *et al.* (2021) investigate how language models represent agreement. These works develop theories that attempt to account for how neural-net language models represent aspects of language. These theories may be useful for understanding aspects of the inner-workings of LLMs (and other types of similar models), and they might also be of use in building better LLMs in the future. If LLMs work in the same ways that humans brains do, then these theories could be useful for explaining how language works. However, it is not at all clear if LLMs work in the same way that human language does. Katzir (2023: 2) writes "Since LLMs were designed to be useful engineering tools, discovering that they teach us about how humans work would be startling indeed, akin to discovering that a newly designed drone accidentally solves an open problem in avian flight."

LLM training is quite different from the human situation. LLMs are exposed to much more data than a human child is exposed to. As Piandadosi (2024: 354) notes, they are "trained on huge datasets of internet-based text to predict upcoming linguistic material." According to Piantadosi (2024: 354), "a typical language model might be trained on hundreds of billions of tokens, estimated to cost millions of dollars in energy alone (Piantadosi 2024: 354)." Estimates are that children hear between 2 million to 11 million words per year (Warstadt & Bowman 2022, Hart & Risley 1992, Gilkerson *et al.* 2017). GPT-3 was reportedly trained on over 200 billion words (Brown *et al.* 2020, Warstadt & Bowman 2022). Piandadosi (2024: 357) writes that LLMs "are imperfect, to be sure, but my qualitative experience interacting with them is like talking to a child, who happened to have memorized much of the internet." I note that I have never spoken with a child who has memorized much of the Internet. Also, as Piandadosi (2024: 358) notes, these models "are trained only on text prediction." Humans are not trained on text prediction.

Presumably, due to their ability to generally produce correct language, Chesi points out that "the computational perspective appears to be leading in terms of observational adequacy, and possibly in terms of descriptive adequacy as well" (p. 19). Furthermore, compared with Minimalist theories, Chesi writes that LLMs "are observationally more adequate but lack explanatory adequacy" (p. 32). LLMs clearly lack explanatory adequacy, but I also think they lack observational adequacy and descriptive adequacy.

Chomsky (1964: 63) writes that "[a] grammar that aims for observational adequacy is concerned merely to give an account of the primary data (e.g. the corpus) that is the input to the learning device." If an LLM could be considered a grammar, then it might be accurate to consider it to be more observationally adequate than Minimalist theories. However, I do not think that an LLM can be considered a grammar that corresponds to grammar used by humans for language. Chomsky (1965: 6) writes that "by a generative grammar I mean simply a system of rules that in some explicit and well-defined way assigns structural descriptions to sentences." Carnie (2021: 27) defines a grammar as "the set of mental rules or procedures that generate a sentence." Radford (1997: 4) writes that "a grammar of a language is a model of the grammatical competence of the fluent native speaker of the language." Given these definitions, ChatGPT (or other LLMS) is not a grammar, or at least not a human language grammar. There is no clear evidence that it assigns structural descriptions to sentences, nor that it has components that directly correspond to the mental rules or procedures that humans use to generate (and parse) sentences. There is no evidence that it is an accurate model of the grammatical competence of a human, in the sense that, as far as we know, it does not produce sentences in the same way that a human does.

Regarding descriptive adequacy, Chomsky (1964: 63) writes that "a grammar that aims for descriptive adequacy is concerned to give a correct account of the linguistic intuition of the native speaker." If you look at the output of an LLM, it produces grammatically correct sentences. On the other hand, it is not clear if it has any intuitions. You can ask an LLM questions about grammar, and it might be able to answer correctly. But these answers result from patterns found in the training data. For example, I asked ChatGPT to discuss the grammaticality of the sentences in Table 1, which require that the main verb agree with the head noun of the subject, even though the head noun of the subject is not the closest noun to the main verb. The answer in Table 1 is correct. However, this correctness does not show that ChatGPT actually has a notion of subject-verb agreement. Instead, ChatGPT was able to come up with an

appropriate answer due to what it 'learned' from patterns its training data, not from some internal grammar.

Discuss the grammaticality of these sentences.
a. The girl who likes her school friends studies hard.
b. The girl who likes her school friends study hard.

Summary
· Sentence (a) is grammatical because it correctly uses the singular verb *studies* with the singular subject *the girl*.
· Sentence (b) is ungrammatical because it incorrectly uses the base form *study* with the singular subject *the girl*.

Table 1. Sample discussion (abbreviated) with ChatGPT.

3. Comprehensive syntactic tests (benchmarks) for Minimalist theories

Chesi argues that a problem for Minimalist theories is an inability "to perform adequately in similarly complete and extensive benchmarks as the ones presented in SyntaxGym" (p. 6). SyntaxGym (<syntaxgym. org>, Hu *et al.* 2020) is an online platform that enables the performance of different computational language models to be compared. Having syntactic benchmarks that could be used to evaluate Minimalist theories could potentially be beneficial, but complex issues arise.

Note that the basic types of phenomena that Hu *et al.* (2020) compare with SyntaxGym come from research in generative linguistics. They write that they chose 16 "[o]f the 47 empirical phenomena reviewed in the summary sections at the end of each chapter (Hu *et al.*: 1727)" of the introductory syntax textbook Carnie (2013).[3] Some of these types of constructions are given in Table 2.

	CONSTRUCTION TYPES	EXAMPLES
(a)	Agreement	The author that the senators hurt is/*are good.
(b)	Center Embedding	The painting that the artist painted deteriorated. *The painting that the artist deteriorated painted.
(c)	Garden path	The woman brought the sandwich from the kitchen fell in the dining room.
(d)	Subordination	*As the doctor studied the book. As the doctor studied the book, the nurse walked into the room.

| (e) | Negative polarity licensing | No author that the senators liked has had any success. |
| (f) | Long distance dependencies (pseudo relative clause) | What the young man planted was the crops. |

Table 2. Types of examples found in SyntaxGym.

Although Chesi correctly points out that the performance of Minimalist theories are not tested with benchmarks such as those given in Table 2, it is notable that all of these types of constructions can be accounted for with Minimalist theories. These types of constructions are taken from a syntax textbook, and thus it is not surprising that accounts of these types of constructions can be found in textbooks such as Radford (2016) and Carnie (2021). There is also a wide body of literature in the Generative linguistics literature (not necessarily just Minimalism) regarding these types of constructions.

Although Minimalist theories can likely account for the various constructions in SyntaxGym, there is no one agreed-upon Minimalist theory that accounts for all of these types of constructions, which is a valid point that I think Chesi is making. The question then arises of whether or not it is possible to test Minimalist theories on these types of phenomena so that they can be compared with computer models of language such as LLMs. I think that it may be possible, but there are several complicating issues.

It is possible to create a computer model based on Minimalist theories and to test the model's ability to parse/generate target syntactic constructions. However, I do not think that it is currently possible to create a model of a single over-arching Minimalist theory that all (or most) researchers working in the Minimalist framework would agree on.

There are at least two approaches that I know of to computational modeling of Minimalist theories. One is the Minimalist Grammar approach of Stabler (1997, 2011) and related work.[4] The other approach, which I have been directly involved with, is attempts to model the latest Minimalist theories with computer programs.

Beginning at least with Fong's (1991) Government and Binding Theory-based parser, there has been research attempting to computationally model theories of Generative Linguistics. Some representative works are the following. Fong & Ginsburg (2012) models constructions with pronouns and antecedents. Fong & Ginsburg (2014) models *tough*-constructions. Ginsburg (2016) models basic statements and *wh*-questions from the perspective of Labeling theory (Chomsky 2013). Fong & Ginsburg

(2019) discusses the architecture of a computational model based on Phase Theory (Chomsky 2001), and Ginsburg & Fong (2019) discusses how this single model accounts for a variety of basic syntactic phenomena and constructions including multiple agreement, constructions with expletives, thematization/extraction, the *that*-trace effect, subject *vs* object *wh*-movement, and relative clauses. Fong & Ginsburg (2023) presents a model that accounts for a wide-variety of English relative clause constructions. Ginsburg (2024) presents a model that permits Merge operations to apply with a limited amount of freedom to account for basic statements, control constructions, *wh*-questions, and yes/no-questions. These types of Minimalist theory-based computational models are possible, and their scope can be extended, but they face several hurdles.

One important issue for these models is that linguists working on Minimalist theories do not agree on one particular Minimalist theory. Among linguists, there are often varying views about particular phenomena, as well as about the core nature of the Minimalist program. This means that no matter which theory is implemented, there will be researchers who may not be pleased. For example, Ginsburg (2024) discusses a computational implementation of some of the latest work in Minimalism, in particular following recent proposals in Chomsky (2001, 2013, 2015, 2021b, 2024). This work is recent, and the majority of linguistics papers in linguistics journals do not make use of the latest notions from Minimalist theories.[5] In order to construct the computer model described in Ginsburg (2024), I had to make assumptions about a variety of controversial topics (see Ginsburg 2024 for references). Without making specific assumptions, the model would not have been able to sufficiently implement a Minimalist theory. For example, a large body of recent work in Minimalism assumes that features on T are inherited from C. But due to the complexity of feature inheritance, and what I see as a lack of strong evidence for a complex feature inheritance operation, I did not implement feature inheritance. Head movement has been the subject of numerous analyses. Some work argues that head movement is problematic and should not exist in the syntax, whereas a large body of work assumes that it occurs in the syntax. Of the work that makes use of syntactic head movement, there are a variety of differing proposals ranging from adjunction of one head with a higher head, internal pair-Merge of a head with a higher head, and movement of a head to a higher specifier position. There are also accounts that posit that some head movement occurs in the syntax and some head movement occurs post-syntactically, as well as accounts for purely post-syntactic head-movement. Following the latest work in Chomsky (2021b, 2024), I assumed that head-movement is a post-syntactic operation.

Furthermore, Chomsky (2021b) proposes a novel FormCopy relation that accounts for how two NPs can be given the same reference. In more recent work, Chomsky (2024) develops a new account of *wh*-movement and focus-movement effects. I incorporated these latest theories into my model, but these theories are not necessarily accepted (or even well-known) by the mainstream community working on Minimalist theories.

Chesi argues that Minimalist theories have not been properly formalized and that a fully explicit theory is lacking, writing that "I think the original sin of most generative linguists is that they have gotten used to incomplete pseudo-formalizations and data fragment explanations" (p. 40). I think that this is one reason why creating computer models for Minimalist theories is useful, as well as a problem. In an ideal world, there would be one Minimalist theory generally agreed upon by linguistics researchers, and this theory could be implemented. But that currently is not the situation. The purpose of linguistic research is to understand how language works. If linguists currently have not come to a consensus about how to account for language, then that reflects the state of our understanding. The development of theories that truly explain how language works could help lead to a more unified theory, and computational models, which can demonstrate how well a single theory accounts for a variety of diverse syntactic phenomena, can be useful for this purpose.

Another problem for development of computational models of Minimalist theories is the technical skills that are required. Most computer scientists are not well-versed in linguistics and most linguists are not well-versed in computer programming. Theoretical syntacticians typically develop theories about how language works. They do not implement their theories on a computer. The development of an easy-to-use software application that enables theoretical linguists to test their theories could help to deal with this problem. How exactly this software should be designed and what exactly it should do are open questions though. But something of this sort could potentially be extremely beneficial.

Chesi argues that Minimalism has become marginalized because of its lack of a fully explicit theory. This may be the case. Development of a fully explicit theory should be a goal of linguistic research. I do not know of a clear way to achieve this, but developing a better understanding of language could help. Linguists (not just those working in Generative Grammar) should be striving to find out how language actually works. The goal of linguistics research is not to come up with clever theories of new (and occasionally old) data. The goal is to actually understand the human faculty of language. Linguists should view theories that are overly complex with suspicion – too much complexity likely indicates a lack of understanding of a phenomenon. Linguists in

different disciplines should talk to each other more and see if there is common ground. Common ground may lead to progress.

4. Conclusion

I have focused on a few of the issues that Chesi raises in his paper. Notably, I think that referring to LLMs as theories and/or as grammars is potentially problematic. A set of benchmarks for testing Minimalist theories could potentially be beneficial, but it faces obstacles, in that there is very little work done modeling Minimalist theories, and in that Minimalist theories at this point are not at all unified. Generative AI, and the LLMs that it produces, is a truly impressive technology (not necessarily in a good way) that is having a large impact (not necessarily good) on human society. However, notions that an LLM can account for how language works may be misguided. LLMs can be useful (as well as harmful), and they are easily accessible. Thus, they will overshadow theoretical work in linguistics. But theoretical work in linguistics gives us insight into how language works. In my opinion, LLMs do not.

Notes

[1]   It is quite difficult (and often not possible) to reliably distinguish language produced by an LLM from language produced by a human.
[2]   Due to lack of space I do not discuss some of the other important issues that Chesi raises.
[3]   Hu *et al.* cite Carnie (2012), but I assume that they are referring to the 2013 edition.
[4]   While this approach is influenced by Minimalism, it does not appear to have closely followed recent work in Minimalism.
[5]   I did an informal survey of syntax-related articles to see how often the latest theories of Chomsky (Chomsky 2013 and later) are discussed. In *Glossa: A Journal of Linguistics*, I counted 19 articles from 2024 that discuss syntax. Of these, only 3 referenced some of the latest work by Chomsky. In the journal *Syntax*, for 2023, I counted 14 articles, of which only 2 cited some of the latest Chomsky work. It looks like the majority of syntax-related publications in my field are not focused on the latest theories of Chomsky (which is primarily what I focus on in my research). I think that these facts show that the field of Generative Linguistics, and linguistics in general, is not exactly unified.

Bibliographical References

See the unified list at the end of this issue.

Unified Bibliographical References

Abels, Klaus & Neeleman, Ad 2012. Linear Asymmetries and the LCA: Linear Asymmetries and the LCA. *Syntax* 15,1. 25-74. <doi.org/10.1111/j.1467-9612.2011.00163.x>.

Abney, Steven 1996. Statistical methods. In Klavans, Judith L. & Resnik, Philip (eds.), *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*. Cambridge, MA: MIT Press. 1-26.

Acemoglu, Daron 2024. *The Simple Macroeconomics of AI*. Working paper 32487. Cambridge, MA: National Bureau of Economic Research. <DOI: 10.3386/w32487>.

Achinstein, Peter 1985. *The Nature of Explanation*. Oxford: Oxford University Press.

Aksënova, Alëna & Deshmukh, Sanket 2018. Formal restrictions on multiple tiers. In *Proceedings of the society for computation in linguistics (SCiL) 2018*. 64-73.

Aksënova, Alëna; Graf, Thomas & Moradi, Sedigheh 2016. Morphotactics as tier-based strictly local dependencies. In *Proceedings of the 14th SIGMORPHON workshop on computational research in phonetics, phonology, and morphology*. 121-130.

Aksënova, Alëna; Rawski, Jonathan; Graf, Thomas & Heinz, Jeffrey 2024. The computational nature of hamony patterns. In Ritter, Nancy & van der Hulst, Harry (eds.), *Handbook of vowel harmony*. Oxford, UK: Oxford University Press. 437-451.

Allott, Nicholas; Kush, Dave & Dillon, Brian 2021. Sentence processing and syntactic theory. In Lohndal, T. & Rey, G. (eds.), *A Companion to Chomsky*. Wiley Publishing. 305-324.

Ambridge, Ben & Blything, Liam 2024. Large language models are better than theoretical linguists at theoretical linguistics. *Theoretical Linguistics* 50,1-2. 33-48.

Anderson, Chris 2008. The end of theory: The data deluge makes the scientific method obsolete. *Wired* 23 June.

Askell, Amanda; Bai, Yuntao; Chen, Anna; Drain, Dawn; Ganguli, Deep; Henighan, Tom; Jones, Andy; Joseph, Nicholas; Mann, Ben; DasSarma, Nova *et al.* 2021. A general language assistant as a laboratory for alignment. <arXiv:2112.00861>.

Ayers, John W. *et al.* 2023. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Internal Medicine*. 589-596. <DOI: 10.1001/jamainternmed.2023.1838>.

Baker, Mark 2001. *The atoms of language* (1st ed.). New York: Basic Books.

Baker, Mark 2009. Formal generative typology. In Heine, Bernd & Narrog, Heiko (eds.), *The Oxford Handbook of Linguistic Analysis*. 1st edition. Oxford: Oxford University Press. 285-312.

Baker, Mark 2013. On agreement and its relationship to case: Some generative ideas and results. *Lingua* 130. 14-32.

Baker, Mark 2021. On Chomsky's legacy in the study of linguistic diversity. In Allott, Nicholas; Lohndal, Terje & Rey, George (eds.), *A companion to Chomsky*. Hoboken, NJ: Wiley Blackwell. 158-171. <doi:10.1002/9781119598732.ch10>.

Baker, Mark & McCloskey, Jim 2007. On the relationship of typology to theoretical syntax. *Linguistic Typology* 11. 285-296.

Bai, Yuntao; Kadavath, Saurav; Kundu, Sandipan; Askell, Amanda; Kernion, Jackson; Jones, Andy; Chen, Anna; Goldie, Anna; Mirhoseini, Azalia; McKinnon, Cameron *et al.* 2022. Constitutional AI: Harmlessness from AI feedback. <arXiv:2212.08073>.

Baltin, Mark 2017. Extraposition. In Everaert, Martin & van Riemsdijk, Henk C. (eds.), *The Wiley Blackwell Companion to Syntax, Second Edition*. Hoboken, NJ: John Wiley & Sons, Inc. 1-33. <doi.org/10.1002/9781118358733.wbsyncom111>.

Barile, Joseph *et al.* 2024. Diagnostic accuracy of a Large Language Model in pediatric case studies. *JAMA Pediatrics*. 313-315. <DOI: 10.1001/jamapediatrics.2023.5750>.

Baroni, Marco 2022. On the proper role of linguistically oriented deep net analysis in linguistic theorizing. In Lappin, Shalom & Bernardy, Jean-Philippe (eds.), *Algebraic structures in natural language*. Boca Raton: CRC Press, Taylor & Francis. 1-16. *ICoRR* <arxiv.org/abs/2106.08694> (2021).

Barton, G. Edward; Berwick, Robert C. & Ristad, Eric Sven 1987. *Computational complexity and natural language*. Cambridge, MA: MIT Press.

Bates, Elizabeth; Elman, Jeffrey L.; Johnson, Mark H.; Karmiloff-Smith, Annette; Parisi, Domenico & Plunkett, Kim 1996. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press. <doi.org/10.7551/mitpress/5929.001.0001>.

Beghelli, Filippo & Stowell, Tim 1997. Distributivity and Negation: The Syntax of Each and Every. In Szabolcsi, Anna (ed.), *Ways of Scope Taking* (Vol. 65). Dordrecht: Springer Netherlands. 71-107. <doi.org/10.1007/978-94-011-5814-5_3>.

Beier, Eleonora J. & Ferreira, Fernanda 2022. Replication of Cutler, Anne & Fodor, Jerry A. 1979, Semantic focus and sentence comprehension. *Journal of Memory and Language* 126. <doi.org/10.1016/j.jml.2022.104339>.

beim Graben, Peter & Potthast, Roland 2014. Universal neural field computation. In Coombes, Stephen; beim Graben, Peter; Potthast, Roland & Wright, James (eds.), *Neural Fields*. Berlin: Springer. <doi.org/10.1007/978-3-642-54593-1_11>.

Belkin, Mikhail; Hsu, Daniel; Ma, Siyuan & Mandal, Soumik 2019. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences* 116. 15849-15854. <doi.org/10.1073/pnas.1903070116>.

Belletti, Adriana 2004. *Structures and Beyond: The Cartography of Syntactic Structures, Volume 3*. Oxford, UK: Oxford University Press.

Bender, Emily M.; Gebru, Timnit; McMillan-Major, Angelina & Shmitchell, Shmargaret 2021. On the dangers of stochastic parrots: Can language models be too big? New York, NY: Association for Computing Machinery. 610-623. <DOI: 10.1145/3442188.3445922>.

Bender, Emily M. & Hanna, Alex 2025. *The AI Con: How to Fight Big Tech's Hype and Create the Future We Want*. Harper Collins.

Bender, Emily & Koller, Alexander 2020. Climbing toward NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th annual meeting of the Association for Computational Linguistics*. 5185-5198. <www.aclweb.org/anthology/2020.acl-main.463>.

Benesty, Michaël 2023. *Unexpected description of GPT4 architecture*. <x.com/pommedeterre33/status/1671263789914677248>.

Bengio, Yoshua; Hinton, Geoffrey; Yao, Andrew; Song, Dawn; Abbeel, Pieter; Darrell, Trevor; Harari, Yuval Noah; Zhang, Ya-Qin; Xue, Lan; Shalev-Shwartz, Shai; Hadfield, Gillian; Clune, Jeff; Maharaj, Tegan; Hutter, Frank; Baydin, Atilim Gunes; McIlraith, Sheila; Gao, Qiqi; Acharya, Ashwin; Krueger, David; Dragan, Anca; Torr, Philip; Russell, Stuart; Kahneman, Daniel; Brauner, Jan & Mindermann, Soren 2024. Managing extreme AI risks amid rapid progress. *Science* 384. 842-845. <doi.org/10.1126/science.adn0117>.

Berwick, Robert C. & Chomsky, Noam 2016. *Why only us: Language and evolution*. Cambridge, MA: MIT Press.

Berwick, Robert C.; Pietroski, Paul; Yankama, Beracah & Chomsky, Noam 2011. Poverty of the stimulus revisited. *Cognitive Science* 35,7. 1207-1242. <DOI: 10.1111/j.1551-6709.2011.01189.x>.

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. *Cognition and the Development of Language*.

Bever, Thomas G. & Townsend, David J. 2001. Some Sentences on Our Consciousness of Sentences. In Dupoux, Emmanuel (ed.), *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler*. Cambridge, MA: MIT Press. 143-155.

Bianchi, Valentina & Chesi, Cristiano 2014. Subject islands, reconstruction, and the flow of the computation. *Linguistic Inquiry*. 525-569. <doi.org/10.1162/LING_a_00166>.

Bjorkman, Bronwyn M. 2017. Singular *they* and the syntactic representation of gender in English. *Glossa: A Journal of General Linguistics* 2,1. <DOI: 10.5334/gjgl.374>.

Blank, Idan 2016. *The Functional Architecture of Language Comprehension Mechanisms: Fundamental Principles Revealed with fMRI*. PhD dissertation. MIT. <doi.org/1721.1/7582>.

Bloom, Paul A. & Fischler, Ira 1980. Completion norms for 329 sentence contexts. *Memory & Cognition* 8,6. 631-642. <doi.org/10.3758/BF03213783>.

Bobaljik, Jonathan D. 2012. *Universals in comparative morphology: Suppletion, superlatives, and the structure of words*. Cambridge, MA: MIT Press.

Bobaljik, Jonathan D. & Wurmbrand, Susi 2008. Case in GB / Minimalism. In Malchukov, Andrej & Spencer, Andrew (eds.), *The Handbook of Case*. New York: Oxford University Press. 44-58.

Bobrow, Daniel G.; Cheslow, Bob; Condoravdi, Cleo; Karttunen, Lauri; Holloway King, Tracy; Nairn, Rowan; de Paiva, Valeria; Price, Charlotte & Zaenen, Annie 2007. PARC's bridge and question answering system. In *Proceedings of the Grammar Engineering Across Frameworks Workshop (GEFA 2007)*. CSLI Publications Online. 46-66.

Bock, J. Kathryn 1986. Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 12,4. 575-586. <doi.org/10.1037/0278-7393.12.4.575>.

Boeckx, Cedric & Leivada, Evelina 2013. Entangled parametric hierarchies: Problems for an overspecified Universal Grammar. *PLOS ONE* 8,9. <doi:10.1371/journal.pone.0072357>.

Bögel, Tina; Freiseis, Mila; Hill, Romi; Wambach, Daniel & Zhao, Tianyi 2024. Language redundancy and acoustic salience: An account in LFG. In Butt, Miriam; Findlay, Jamie A. & Toivonen, Ida (eds.), *The proceedings of the lfg'24 conference*. 90-115.

Bögel, Tina & Zhao, Tianyi 2025. From speech signal to syntactic structure: A computational implementation. *Journal of Language Modeling* 13,1. 1-42.

Borer, Hagit 2005. *Structuring sense: In name only*. Oxford: Oxford University Press.

Bošković, Željko 2005. On the locality of left branch extraction and the structure of NP. *Studia Linguistica* 59. 1-45.

Bošković, Željko 2016. Introduction. *The Linguistic Review* 33,1. 1-16. <doi.org/10.1515/tlr-2015-0012>.

Bowman, Samuel R.; Hyun, Jeeyoon; Perez, Ethan; Chen, Edwin; Pettit, Craig; Heiner, Scott; Lukošiūtė, Kamilė; Askell, Amanda; Jones, Andy; Chen, Anna *et al.* 2022. Measuring progress on scalable oversight for large language models. <arXiv:2211.03540>.

Brayton, Flint; Laubach, Thomas & Reifschneider, David 2014. *The FRB/US Model: A Tool for Macroeconomic Policy Analysis*. Washington, DC: Board of Governors of the Federal Reserve System. <DOI: 10.17016/2380-7172.0012>.

Brennan, Jonathan R.; Stabler, Edward P.; Van Wagenen, Sarah E.; Luh, Wen-Ming & Hale, John T. 2016. Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language* 157-158. 81-94. <doi.org/10.1016/j.bandl.2016.04.008>.

Bresnan, Joan 1982. Control and complementation. *Linguistic Inquiry* 13,3. 343-434.

Bresnan, Joan 2016. Linguistics: The Garden and the Bush. *Computational Linguistics* 42,4. 599-617. <doi.org/10.1162/COLI a 00260>.

Bresnan, Joan; Cueni, Anna; Nikitina, Tatiana & Baayen, R. Harald 2007. Predicting the dative alternation. In Bouma, Gerlof; Krämer, Irene & Zwarts, Joost (eds.), *Cognitive Foundations of Interpretation*. Amsterdam: Royal Netherlands Academy of Science. 69-94.

Bressan, Veronica; Piccini Bianchessi, Maria Letizia; Fusco, Achille; Rossi, Sarah; Neri, Sofia & Chesi, Cristiano 2025. BLiMP-IT. <doi.org/10.17605/OSF.IO/2JKFN>.

Brown, Tom B.; Mann, Benjamin; Ryder, Nick; Subbiah, Melanie; Kaplan, Jared; Dhariwal, Prafulla; Neelakantan, Arvind; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D. M.; Wu, J.; Winter, C.; … Amodei, D. 2020. Language Models are Few-Shot Learners. In Larochelle, Hugo *et al.* (eds.), *Advances in Neural Information Processing Systems 33 (NeurIPS 2020) Proceedings*. <arxiv.org/abs/2005.14165>.

Brunato, Dominique; Chesi, Cristiano; Dell'Orletta, Felice; Montemagni, Simonetta; Venturi, Giulia & Zamparelli, Roberto 2020. AcCompl-it@ EVALITA2020: Overview of the acceptability & complexity evaluation task for Italian. *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020), Online. CEUR. Org.*

Burness, Phillip; McMullin, Kevin & Chandlee, Jane 2021. Long-distance phonological processes as tier-based strictly local functions. *Glossa* 6. 1-37. <doi.org/10.16995/glossa.5780>.

Burness, Phillip; McMullin, Kevin & Nevins, Andrew 2024. Revisiting locality in vowel harmony. In Ritter, Nancy & van der Hulst, Harry (eds.), *Handbook of vowel harmony*. Oxford, UK: Oxford University Press. 269-290.

Butt, Miriam; Bögel, Tina; Zymla, Mark-Matthias & Mumtaz, Benazir 2024. Alternative questions in Urdu: from the speech signal to semantics. In Butt, Miriam; Findlay, Jamie & Toivonen, Ida (eds.), *Proceedings of the LFG'24 Conference*. Konstanz: PubliKon. 141-164. <lfg-proceedings.org/lfg/index.php/main/article/view/65/50>.

Butt, Miriam; Holloway King, Tracy; Niño, María-Eugenia & Segond, Frédérique 1999. *A Grammar Writer's Cookbook*. Stanford: CSLI Publications.

Butt, Miriam & Ramchand, Gillian 2005. Complex aspectual structure in Hindi/Urdu. In Ertishik-Shir, Nomi & Rappaport, Tova (eds.), *The Syntax of Aspect*. Oxford: Oxford University Press. 117-153.

Cahill, Aoife 2008. Treebank-based probabilistic phrase structure parsing. *Language and Linguistics Compass* 2,1. 36-58.

Cann, Ronnie; Kempson, Ruth & Marten, Lutz 2005. *The Dynamics of Language: An introduction*. Elsevier Academic Press.

Cao, Rosa & Yamins, Daniel 2024. Explanatory Models in Neuroscience, Part 2: Functional Intelligibility and the Contravariance Principle. *Cognitive Systems Research* 85. 101200. <doi.org/10.1016/j.cogsys.2023.101200>.

Carnie, Andrew 2013. *Syntax: A Generative Introduction, Third Edition*. Malden, MA: Wiley Blackwell.

Carnie, Andrew 2021. *Syntax: A Generative Introduction, Fourth Edition*. Malden, MA: Wiley Blackwell.

Cauchy, Augustin 1847. Méthode générale pour la résolution des systèmes d'équations simultanées. *Comptes rendus hebdomadaires des séances de l'Académie des sciences* 25. 536-538.

Cecchetti, Gabriele; Tomasini, Cedric A.; Herff, Steffen A. & Rohrmeier, Martin A. 2023. Interpreting rhythm as parsing. *Cognitive Science* 47. e13389. <doi.org/10.1111/cogs.13389>.

Chaitin, Gregory J. 1969. On the Simplicity and Speed of Programs for Computing Infinite Sets of Natural Numbers. *Journal of the ACM* 16,3. 407-422. <doi.org/10.1145/321526.321530>.

Chandlee, Jane 2014. Strictly local phonological processes. PhD dissertation. University of Delaware.

Chandlee, Jane 2017. Computational locality in morphological maps. *Morphology* 27. 599-641.

Chandlee, Jane 2022. Less is more: Reexamining assumptions through the narrow focus of subregularity. *Theoretical Linguistics* 48. 205-218.

Chandlee, Jane & Heinz, Jeffrey 2018. Strict locality and phonological maps. *Linguistic Inquiry* 49. 23-60.

Charchidi, Vincent J. 2024. Creative Minds Like Ours? Large Language Models and the Creative Aspect of Language Use. *Biolinguistics* 18. 1-31.

Charpentier, Lucas Georges Gabriel & Samuel, David 2023. Not all layers are equally as important: Every Layer Counts BERT. *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*. 210-224. <doi.org/10.18653/v1/2023.conll-babylm.20>.

Chen, Binglin; Lewis, Colleen M.; West, Matthew & Zilles, Craig 2024. Plagiarism in the age of Generative AI: Cheating method change and learning loss in an Intro to CS Course. In *L@S '24: Eleventh ACM Conference on Learning @ Scale, Atlanta GA USA*. New York, NY: ACM. 75-85. <DOI: 10.1145/3657604.3662046>.

Chen, Tianlong; Frankle, Jonathan; Chang, Shiyu; Liu, Sijia; Zhang, Yang; Wang, Zhangyang & Carbin, Michael 2020. The lottery ticket hypothesis for pre-trained BERT networks. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M. F. & Lin, H. (eds.), *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*. Online: Curran Associates, Inc. 15834-15846.

Chen, Zhong & Hale, John T. 2010. Deforesting logical form. *Procs. Mathematics of Language*. Berlin: Springer. LNCS 6149. <doi.org/10.1007/978-3-642-14322-9_2>.

Cheng, Lisa L.-S.; Heycock, Caroline & Zamparelli, Roberto 2017. Two levels for definiteness. In Erlewine, M. Y. (ed.), *Proceedings of GLOW in Asia XI – Vol. 1*. Volume 84 of *MIT Working Papers in Linguistics*. MIT.

Cheng, Lisa L.-S. & Sybesma, Rint 1999. Bare and not-so-bare nouns and the

structure of NP. *Linguistic Inquiry* 30,4. 509-542.

Chesi, Cristiano 2007. An introduction to phase-based minimalist grammars: why move is top-down from left-to-right. In Moscati, V. (ed.), *STIL – Studies in Linguistics*, Volume 1. CISCL Press. 38-75.

Chesi, Cristiano 2021. Expectation-based Minimalist Grammars. <arxiv.org/abs/2109.13871>.

Chesi, Cristiano 2023. Parameters of cross-linguistic variation in expectation-based Minimalist Grammars (e-MGs). *Italian Journal of Computational Linguistics* 9,1. 21.

Chesi, Cristiano *forthcoming*. Linearization (as Part of Core Syntax). In Grohmann, Kleanthes & Leivada, Evelina (eds.), *Cambridge Handbook of Minimalism*. Cambridge (UK): Cambridge University Press. <ling.auf.net/lingbuzz/006689>.

Chesi, Cristiano; Barbini, Matilde; Bressan, Veronica; Neri, Sofia; Piccini Bianchessi, Maria Letizia; Sarah, Rossi & Sgrizzi, Tommaso 2024. Different Ways to Forget: Linguistic Gates in Recurrent Neural Networks. In *Proceedings of the BabyLM Challenge at the 28th Conference on Computational Natural Language Learning*.

Chesi, Cristiano & Bianchi, Valentina 2014. Subject islands, reconstruction, and the flow of the computation. *Linguistic Inquiry* 45,4. 525-569.

Chesi, Cristiano & Moro, Andrea 2015. The subtle dependency between Competence and Performance. *MIT Working Papers In Linguistics* 77. 33-46.

Chesi, Cristiano; Vespignani, Francesco & Zamparelli, Roberto *to appear*. Large language models under evaluation: An acceptability, complexity and coherence assessment in Italian. *Italian Journal of Computational Linguistics*.

Chierchia, Gennaro 1998. Reference to kinds across languages. *Natural Language Semantics* 6. 339-405.

Cho, Kyunghyun; van Merrienboer, Bart; Gulcehre, Caglar; Bahdanau, Dzmitry; Bougares, Fethi; Schwenk, Holger & Bengio, Yoshua 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Moschitti, Alessandro; Pang, Bo & Daelemans, Walter (eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar: Association for Computational Linguistics. 1724-1734. <DOI: 10.3115/v1/D14-1179>.

Chomsky, Noam 1956. Three models for the description of language. *IEEE Transactions on Information Theory* 2,3. 113-124. <doi.org/10.1109/TIT.1956.1056813>.

Chomsky, Noam 1957. *Syntactic Structures*. Berlin: Mouton de Gruyter.

Chomsky, Noam 1959. A Review of B. F. Skinner's Verbal Behavior. *Language* 35,1. 26. <doi.org/10.2307/411334>.

Chomsky, Noam 1964. *Current Issues in Linguistic Theory*. Berlin: De Gruyter.

Chomsky, Noam 1965. *Aspects of the Theory of Syntax* (Vol. 11). Cambridge, MA: MIT Press.

Chomsky, Noam 1966. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. New York, NY: Harper & Row.

Chomsky, Noam 1968. *Language and Mind*. New York, NY: Harcourt, Brace & World.

Chomsky, Noam 1968b. Quine's Empirical Assumptions. *Synthese* 19,1-2. 53-68. <doi.org/10.1007/bf00568049>.

Chomsky, Noam 1969. Quine's empirical assumptions. In Davidson, Donald & Hintikka, Jaakko (eds.), *Words and Objections: Essays on the Work of W.V. Quine*. Dordrecht, Netherlands: Springer Dordrecht. 53-68. <DOI: 10.1007/978-94-010-1709-1_5>.

Chomsky, Noam 1975. *Questions on Form and Interpretation*. Lisse: Peter de Ridder. <doi.org/10.1007/978-3-642-14322-9_2>.

Chomsky, Noam 1981. *Lectures on government and binding: The Pisa lectures*. Walter de Gruyter.

Chomsky, Noam 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Chomsky, Noam 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam 1995b. Language and Nature. *Mind* 104 (413). 1-61.

Chomsky, Noam 2001. Derivation by phase. In Kenstowicz, Michael (ed.), *Ken Hale: A life in language*. Cambridge, MA: MIT Press. 1-52.

Chomsky, Noam A. 2004. *The generative enterprise revisited. Discussions with Riny Huybregts, Henk van Riemsdijk, Naoki Fukui and Mihoko Zushi*. De Gruyter Mouton.

Chomsky, Noam A. 2005. Three Factors in Language Design. *Linguistic Inquiry* 36,1. 1-22.

Chomsky, Noam 2008. On phases. In Freidin, Robert; Otero, Carlos P. & Zubizarreta, Maria Luisa (eds.), *Foundational issues in linguistic theory: Essays in Honor of Jean-Roger Vergnaud* (Vol. 45). Cambridge, MA: MIT Press. 133-166.

Chomsky, Noam 2012. Language and Limits of Understanding. <www.nets.iusspavia.it/dox/chomsky2012-LLU-IUSS_Pavia.pdf>.

Chomsky, Noam 2013. Problems of projection. *Lingua* 130. 33-49.

Chomsky, Noam 2015. Problems of projection: Extensions. In Di Domenico, Elisa; Hamann, Cornelia & Matteini, Simona (eds.), *Linguistik Aktuell/ Linguistics Today* (Vol. 223). Amsterdam: John Benjamins. 1-16. <doi.org/10.1075/la.223.01cho>.

Chomsky, Noam 2021a. Simplicity and the form of grammars. *Journal of Language Modelling* 9,1. <doi.org/10.15398/jlm.v9i1.257>.

Chomsky, Noam 2021b. Minimalism: where are we now, and where can we hope to go. *Gengo Kenkyu* 160. 1-42.

Chomsky, Noam 2024. The Miracle Creed and SMT. In Greco, M. & Mocci, D. (eds.), *A Cartesian dream: A geometrical account of syntax: In honor of Andrea Moro*. Rivista di Grammatica Generativa / Research in Generative Grammar 17-40.

Chomsky, Noam & Lasnik, Howard 1977. Filters and Control. *Linguistic*

Inquiry 8,3. 425-504.

Chomsky, Noam; Roberts, Ian & Watumull, Jeffrey 2023. Noam Chomsky: The False Promise of ChatGPT. *New York Times* 8 March.

Chomsky, Noam; Seely, T. Daniel; Berwick, Robert C.; Fong, Sandiway; Huybregts, M. A. C.; Kitahara, Hisatsugu; McInnerney, Andrew & Sugimoto, Yushi 2023. *Merge and the Strong Minimalist Thesis* (1st ed.). Cambridge: Cambridge University Press. <doi.org/10.1017/9781009343244>.

Chowdhury, Shammur Absar & Zamparelli, Roberto 2018. RNN Simulations of Grammaticality Judgments on Long-distance Dependencies. *Proceedings of the 27th International Conference on Computational Linguistics*. 133-144. <aclanthology.org/C18-1012>.

Cinque, Guglielmo 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford, UK: Oxford University Press.

Cinque, Guglielmo 2002. *Functional Structure in DP and IP: The Cartography of Syntactic Structures, Volume 1*. Oxford, UK: Oxford University Press.

Cinque, Guglielmo 2005. Deriving Greenberg's Universal 20 and Its Exceptions. *Linguistic Inquiry* 36,3. 315-332. <doi.org/10.1162/0024389054396917>.

Cinque, Guglielmo & Rizzi, Luigi 2010. The Cartography of Syntactic Structures. In Heine, B. & Narrog, H. (eds.), *The Oxford Handbook of Linguistic Analysis*. Oxford / New York: Oxford University Press. 51-65.

Clark, Alexander & Lappin, Shalom 2010. Computational learning theory and language acquisition. *Philosophy of Linguistics*. 445-475.

Clark, Alexander & Lappin, Shalom 2011. *Linguistic Nativism and the Poverty of the Stimulus*. Chichester: Wiley-Blackwell.

Clifton, Charles Jr; Ferreira, Fernanda; Henderson, John M.; Inhoff, Albrecht W.; Liversedge, Simon P.; Reichle, Erik D. & Schotte, Elizabeth R. 2015. Eye movements in reading and information processing. *Journal of Memory and Language* 86. 1-19.

Collins, Chris; Kayne, Richard & Koopman, Hilda 2009. *Syntactic structures of the world's languages (SSWL)*. <terraling.com/groups/7>.

Collins, Chris & Stabler, Edward P. 2016. A Formalization of Minimalist Syntax. *Syntax* 19,1. 43-78. <doi.org/10.1111/synt.12117>.

Collins, Joe 2024. The simple reason LLMs are not scientific models (and what the alternative is for linguistics). <lingbuzz.net/lingbuzz/008026>.

Conneau, Alexis; Kruszewski, German; Lample, Guillaume; Barrault, Loïc & Baroni, Marco 2018. What you can cram into a single $&!#* vector: Probing sentence embeddings for linguistic properties. In Gurevych, Iryna & Miyao, Yusuke (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics. 2126-2136. <DOI: 10.18653/v1/P18-1198>.

Corbett, Greville G. 2010. Implicational hierarchies. In Song, Jae

Jong (ed.), *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press. 190-205. <doi.org/10.1093/oxford-hb/9780199281251.013.0011>.

Cottier, Ben; Rahman, Robi; Fattorini, Loredana; Maslej, Nestor; Besiroglu, Tamay & Owen, David 2025. The rising costs of training frontier AI models. <arXiv:2405.21015>.

Crain, Stephen & Nakayama, Mineharu 1987. Structure Dependence in Grammar Formation. *Language* 63,3. 522. <doi.org/10.2307/415004>.

Crain, Stephen & Thornton, Rosalind 2021. Universal grammar and language acquisition. In Allot, Nicholas; Lohndahl, Terje & Rey, Georges (eds.), *A Companion to Chomsky*. Wiley. <doi.org/10.1002/9781119598732.ch21>.

Crawford, Kate 2024. Generative AI's environmental costs are soaring – and mostly secret. *Nature* 626. 693. <DOI: 10.1038/d41586-024-00478-x>.

Crystal, David 2011. *Internet Linguistics: A Student Guide*. London: Routledge.

Cutler, Anne & Fodor, Jerry A. 1979. Semantic focus and sentence comprehension. *Cognition* 7. 49-59. <doi.org/10.1016/0010-0277(79)90010-6>.

Cybenko, George 1989 Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems* 2,4. 303-314.

Dahl, Östen 2020. Morphological complexity and the minimum description length approach. In Arkadiev, Peter & Gardani, Francesco (eds.), *The complexities of morphology*. Oxford: Oxford University Press. 331-343.

D'Alessandro, Roberta 2019. The achievements of Generative Syntax: A time chart and some reflections. *Catalan Journal of Linguistics*. 7-26.

Dalrymple, Mary (ed.) 2023. *The Handbook of Lexical Functional Grammar*. Empirically Oriented Theoretical Morphology and Syntax. Berlin: Language Science Press. <10.5281/zenodo.10037797>.

Dalrymple, Mary; Gupta, Vineet; Lamping, John & Saraswat, Vijay 1999. Relating resource-based semantics to categorial semantics. In Dalrymple, Mary (ed.), *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*. Language, Speech, and Communication. Cambridge, MA: MIT Press. 261-280.

Dalrymple, Mary; Patejuk, Agnieszka & Zymla, Mark-Matthias 2020. XLE+Glue – A new tool for integrating semantic analysis in XLE. In Butt, Miriam & Toivonen, Ida (eds.), *Proceedings of the LFG'20 Conference*. Stanford, CA: CSLI Publications. 89-108. <cslipublications.stanford.edu/LFG/2020/lfg2020-dpz.pdf>.

De Santo, Aniello 2019. Testing a Minimalist Grammar Parser on Italian Relative Clause Asymmetries. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*. 93-104. <doi.org/10.18653/v1/W19-2911>.

De Santo, Aniello 2020. Structure and memory: A computational model of storage, gradience, and priming. PhD dissertation. Stony Brook University.

Deacon, Terence W. 1997. *The symbolic species: The co-evolution of language and the human brain*. Allen Lane: The Penguin Press.

Delétang, Grégoire; Ruoss, Anian; Grau-Moya, Jordi; Genewein, Tim; Wenliang, Li Kevin; Catt, Elliot; Cundy, Chris *et al.* 2022. Neural Networks and the Chomsky Hierarchy. <doi.org/10.48550/ARXIV.2207.02098>.

Demirci, Ozge; Hannane, Jonas & Zhu, Xinrong 2024. Who is AI replacing? The impact of Generative AI on online freelancing platforms. *SSRN Electronic Journal.* <DOI: 10.2139/ssrn.4991774>.

Demirdache, H.; Hornstein, N.; Lasnik, H.; May, R.; Rizzi, L. 2024. Structured Sentences and the Computational Theory of Mind: Roundtable. In *Festschrift for Howard Lasnik*. Cambridge: Cambridge University Press.

Dennett, Daniel C. 1978. Why you can't make a computer that feels pain. *Synthese* 38. 415-456.

Dentella, Vittoria; Günther, Fritz & Leivada, Evelina 2023. Systematic testing of three Language Models reveals low language accuracy, absence of response stability, and a yes-response bias. *Proceedings of the National Academy of Sciences* 120,51. e2309583120. <doi.org/10.1073/pnas.2309583120>.

Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton & Toutanova, Kristina 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Burstein, Jill; Doran, Christy & Solorio, Thamar (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Vol. 1. Minneapolis, MN: Association for Computational Linguistics. 4171-4186. <DOI: 10.18653/v1/N19-1423>.

Dijkstra, Edsger W. 1982. *Selected Writings on Computing*. Berlin: Springer.

Dobson, James E. 2023. On reading and interpreting black box deep neural networks. *International Journal of Digital Humanities* 5. 431-449. <DOI: 10.1007/s42803-023-00075-w>.

Dryer, Matthew S. 2006. Descriptive theories, explanatory theories, and basic linguistic theory. In Ameka, Felix K.; Dench, Alan & Evans, Nicholas (eds.), *Catching language: The standing challenge of grammar writing*. Berlin: Mouton de Gruyter. 207-234. <www.acsu.buffalo.edu/~dryer/desc.expl.theories.pdf>.

Dryer, Matthew & Haspelmath, Martin 2022. *The World Atlas of Language Structures Online* (v2020.3) [dataset]. Zenodo. <doi.org/10.5281/ZENODO.7385533>.

Edinger, Harald 2022. Offensive ideas: structural realism, classical realism and Putin's war on Ukraine. *International Affairs* 98,6. 1873-1893. <DOI: 10.1093/ia/iiac217>.

Elman, Jeffrey L. 1990. Finding Structure in Time. *Cognitive Science* 14,2. 179-211. <doi.org/10.1207/s15516709cog1402_1>.

Elman, Jeffrey L. 1991. Distributed representations, simple recurrent net-

works, and grammatical structure. *Machine Learning* 7,2. 195-225. <DOI: 10.1023/A:1022699029236>.

Elman, Jeffrey L. 1993. Learning and development in neural networks: The importance of starting small. *Cognition* 48,1. 71-99. <doi. org/10.1016/0010-0277(93)90058-4>.

Engelfriet, Joost; Lilin, Eric & Maletti, Andreas 2009. Extended multi bottom-up tree transducers: Composition and decomposition. *Acta Informatica* 46. 561-590. <doi.org/10.1007/s00236-009-0105-8>.

Epstein, Samuel David; Groat, Erich M.; Kawashima, Ruriko & Kitahara, Hisatsugu (eds.) 1998. *A derivational approach to syntactic relations*. Oxford, UK: Oxford University Press.

Ermolaeva, Marina 2023. Evaluating syntactic proposals using Minimalist grammars and minimum description length. *Journal of Language Modelling* 11. 67-119. <doi.org/10.15398/jlm.v11i1.334>.

Espinal, Maria Teresa & Cyrino, Sonia 2022. A syntactically-driven approach to indefiniteness, specificity and antispecificity in Romance. *Journal of Linguistics* 58. 535-570.

Ettinger, Allyson 2020. What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. *Transactions of the Association for Computational Linguistics* 8. 34-48. <doi.org/10.1162/tacl_a_00298>.

Evans, Lyndon 2007. The Large Hadron Collider. *New Journal of Physics* 9,9. 335-335. <doi.org/10.1088/1367-2630/9/9/335>.

Evans, Nicholas & Levinson, Stephen C. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32,5. 429-448. <DOI:10.1017/S0140525X0999094X>.

Evanson, Linnea; Lakretz, Yair & King, Jean-Rémi 2023. Language acquisition: do children and language models follow similar learning stages? <arXiv:2306.03586>.

Fazi, M. Beatrice 2021. Beyond human: Deep learning, explainability and representation. *Theory, Culture & Society* 38. 55-77.

Feyerabend, Paul K. 1962. Explanation, reduction, and empiricism. In Feigl, Herbert & Maxwell, Grover (eds.), *Scientific explanation, space, and time*. Vol. 3. Minneapolis, MN: University of Minnesota Press. 28-97.

Fisher, Cynthia 2002. The role of abstract syntactic knowledge in language acquisition: A reply to Tomasello (2000). *Cognition* 82. 259-278.

Fleck, Ludwik 1935. *Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv*. Basel, Switzerland: Benno Schwabe & Co.

Fodor, Janet Dean 1998. Unambiguous triggers. *Linguistic Inquiry* 29. 1-36.

Fodor, Jerry A. 1980. *The Language of Thought*. Harvard: Harvard University Press.

Fodor, Jerry A. 1983. *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Fodor, Jerry A. 2010. *LOT 2: The Language of Thought Revisited*. Oxford, UK: Oxford University Press.

Fodor, Jerry A. & Bever, Thomas G. 1965. The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior* 4. 414-420. <doi.org/10.1016/s0022-5371(65)80081-0>.

Fong, Sandiway 1991. *Computational properties of principle-based grammatical theories*. PhD dissertation. MIT, Cambridge (MA).

Fong, Sandiway & Ginsburg, Jason 2012. Computation with doubling constituents: Pronouns and antecedents in Phase Theory. In Di Sciullo, Anna Maria (ed.), *Towards a Biolinguistic Understanding of Grammar: Essays on interfaces*. Amsterdam: John Benjamins. 303-338.

Fong, Sandiway & Ginsburg, Jason 2014. A new approach to tough-constructions. In Santana-LaBarge, Robert E (ed.), *Proceedings of the 31st West Coast Conference on Formal Linguistics (WCCFL 31)*. Somerville, MA: Cascadilla Proceedings Project. 180-188.

Fong, Sandiway & Ginsburg, Jason 2019. Towards a Minimalist Machine. In Berwick, Robert C. & Stabler, Edward P. (eds.), *Minimalist Parsing*. Oxford: Oxford University Press. 16-38.

Fong, Sandiway & Ginsburg, Jason 2023. On the computational modeling of English relative clauses. *Open Linguistics* 9. 1-35. <DOI: 10.1515/opli-2022-0246>.

Forster, Kenneth I.; Guerrera, Christine & Elliot, Lisa 2009. The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods* 41,1. 163-171. <doi.org/10.3758/BRM.41.1.163>.

Fox, Danny & Karzir, Roni 2024. Large Language Models and Theoretical Linguistics. *Theoretical Linguistics* 50. 71-76. <DOI: 10.1515/tl-2024-2005>.

Fox, Danny & Nissenbaum, Jon 1999. Extraposition and scope: A case for overt QR. *Proceedings of the 18th West Coast Conference on Formal Linguistics* 18,2. 132-144.

Fox, Melvin J. & Skolnick, Betty P. 1975. *Language in Education: Problems and Prospects in Research and Teaching*. New York, NY: Ford Foundation.

Frampton, John & Gutmann, Sam 2002. Crash-Proof Syntax. In Epstein, Samuel David & Seely, T. Daniel (eds.), *Derivation and Explanation in the Minimalist Program* (1st ed.). Wiley. 90-105. <doi.org/10.1002/9780470755662.ch5>.

Frank, Anette; Holloway King, Tracy; Kuhn, Jonas & Maxwell, John T. III 2001. Optimality theory style constraint ranking in large-scale LFG grammars. In Sells, Peter (ed.), *Formal and Empirical Issues in Optimality Theory*. Stanford: CSLI Publications. 367-397.

Frank, Robert 1990. Licensing and tree adjoining grammar in government binding parsing. *28th Annual Meeting of the Association for Computational Linguistics*. 111-118.

Frank, Robert 2002. *Phrase structure composition and syntactic dependencies*. Cambridge, MA: MIT Press.

Frankle, Jonathan & Carbin, Michael 2019. The lottery ticket hypothesis: Finding sparse, trainable neural networks. In *ICLR 2019 Conference Track*. New Orleans, LA: OpenReview.

Friedmann, Naama; Belletti, Adriana & Rizzi, Luigi 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua* 119,1. 67-88.

Fusco, Achille; Barbini, Matilde; Piccini Bianchessi, Maria Letizia; Bressan, Veronica; Neri, Sofia; Rossi, Sarah; Sgrizzi, Tommaso & Chesi, Cristiano 2024. Recurrent Networks Are (Linguistically) Better? An Experiment on Small-LM Training on Child-Directed Speech in Italian. In *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-It 2024)*. Aachen: CEUR.

Futrell, Richard; Gibson, Edward & Levy, Roger P. 2020. Lossy-Context Surprisal: An Information-Theoretic Model of Memory Effects in Sentence Processing. *Cognitive Science* 44,3. <doi.org/10.1111/cogs.12814>.

Futrell, Richard & Levy, Roger 2017. Noisy-context surprisal as a human sentence processing cost model. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. 688-698.

Futrell, Richard; Wilcox, Ethan; Morita, Takashi; Qian, Peng; Ballesteros, Miguel & Levy, Roger 2019. Neural language models as psycholinguistic subjects: Representations of syntactic state. <arXiv:1903.03260>.

Gauthier, Jon; Hu, Jennifer; Wilcox, Ethan; Qian, Peng & Levy, Roger 2020. SyntaxGym: An online platform for targeted evaluation of language models. In Celikyilmaz, Asli & Wen, Tsung-Hsien (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Online: Association for Computational Linguistics. 70-76. <DOI: 10.18653/v1/2020.acl-demos.10>.

Gehrke, Berit & McNally, Louise 2019. Idioms and the syntax/semantics interface of descriptive content vs. reference. *Linguistics* 57,4. 769-814. <10.1515/ling-2019-0016>.

Gerth, Sabrina 2015. Memory limitations in sentence comprehension. A structure-based complexity metric of processing difficulty. PhD dissertation. University of Potsdam.

Gianollo, Chiara; Guardiano, Cristina & Longobardi, Giuseppe 2008. Three fundamental issues in parametric linguistics. In Biberauer, Theresa (ed.), *Linguistik Aktuell/Linguistics Today* (Vol. 132). Amsterdam: John Benjamins. 109-142. <doi.org/10.1075/la.132.05gia>.

Gibson, Edward; Futrell, Richard; Piantadosi, Steven T.; Dautriche, Isabelle; Mahowald, Kyle; Bergen, Leon & Levy, Roger 2019. How efficiency shapes human language. *Trends in Cognitive Sciences* 23,5. 389-407. <doi:10.1016/j.tics.2019.02.003>.

Gibson, Edward & Wexler, Ken 1994. Triggers. *Linguistic Inquiry* 25,3. 407-454.

Gilkerson, Jill *et al.* 2017. Mapping the early language environment using

all-day recordings and automated analysis. *American Journal of Speech-Language Pathology* 26. 248-265. <DOI: 10.1044/2016_AJSLP-15-016>.

Ginsburg, Jason 2016. Modeling of Problems of Projection: A non-countercyclic approach. *Glossa: A Journal of General Linguistics* 1,1:7. 1-46. <DOI: 10.5334/gjgl.22>.

Ginsburg, Jason 2024. Constraining free Merge. *Biolinguistics* 18, e14015. 1-60. <DOI: 10.5964/bioling.14015>.

Ginsburg, Jason & Fong, Sandiway 2019. Combining linguistic theories in a Minimalist Machine. In Stabler, Edward P. & Berwick, Robert C. (eds.), *Minimalist Parsing*. Oxford, UK: Oxford University Press. 39-68. <doi.org/10.1093/oso/9780198795087.003.0003>.

Giusti, Giuliana 2015. *Nominal Syntax at the Interfaces: A Comparative Analysis of Languages With Articles*. Cambridge: Cambridge Scholars Publishing.

Gold, E. Mark 1967. Language identification in the limit. *Information and Control* 10,5. 447-474. <doi.org/10.1016/S0019-9958(67)91165-5>.

Goldsmith, John & Riggle, Jason 2012. Information theoretic approaches to phonological structure: The case of Finnish vowel harmony. *Natural Language and Linguistic Theory* 30. 859-896.

Gorman, Kyle 2016. Pynini: A Python library for weighted finite-state grammar compilation. In *Procs. SIGFSM Workshop on Statistical NLP and Weighted Automata*. <doi.org/10.18653/v1/W16-2409>.

Goyal, Anirudh & Bengio, Yoshua 2022. Inductive Biases for Deep Learning of Higher-Level Cognition. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 478 (2266). <doi.org/10.1098/rspa.2021.0068>.

Graf, Thomas 2020. Curbing feature coding: Strictly local feature assignment. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2020*. 362-371.

Graf, Thomas 2022c. Typological implications of tier-based strictly local movement. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2022*. 184-193.

Graf, Thomas 2022b. Subregular linguistics: Bridging theoretical linguistics and formal grammar. *Theoretical Linguistics* 48. 145-184. <doi.org/10.1515/tl-2022-2037>.

Graf, Thomas 2022a. Diving deeper into subregular syntax. *Theoretical Linguistics* 48. 245-278. <doi.org/10.1515/tl-2022-2043>.

Graf, Thomas 2023. Subregular tree transductions, movement, copies, traces, and the ban on improper movement. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2023*. 289-299. <doi.org/10.7275/tk1n-q855>.

Graf, Thomas *to appear*. Minimalism and computational linguistics. In Grohman, Kleanthes K. & Leivada, Evelina (eds.), *Handbook of Minimalism*. Cambridge: Cambridge University Press.

Graf, Thomas & Abner, Natasha 2012. Is syntactic binding rational?

In *Proceedings of the 11th international workshop on Tree Adjoining Grammars and related formalisms (TAG+11)*. 189-197.

Graf, Thomas & Kostyszyn, Kalina 2021. Multiple wh-movement is not special: The subregular complexity of persistent features in Minimalist grammars. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2021*. 275-285.

Graf, Thomas & Mayer, Connor 2018. Sanskrit n-retroflexion is input-output tier-based strictly local. In *Proceedings of SIGMORPHON 2018*. 151-160.

Graf, Thomas; Monette, James & Zhang, Chong 2017. Relative clauses as a benchmark for Minimalist parsing. *Journal of Language Modelling* 5,1. 57-106. <doi.org/10.15398/jlm.v5i1.157>.

Graf, Thomas & Shafiei, Nazila 2019. C-command dependencies as TSL string constraints. In Jarosz, Gaja; Nelson, Max; O'Connor, Brendan & Pater, Joe (eds.), *Proceedings of the Society for Computation in Linguistics (SCiL) 2019*. 205-215.

Grice, Herbert Paul 1975. Logic and conversation. In Cole, Peter & Morgan, Jerry L. (eds.), *Syntax and Semantics*. New York, NY: Academic Press. 41-58.

Grillo, Nino 2008. *Generalized minimality: Syntactic underspecification in Broca's aphasia.* LOT.

Grünwald, Peter D. 2007. *The minimum description length principle.* Cambridge, MA: MIT Press.

Guardiano, Cristina & Longobardi, Giuseppe 2016. Parameter Theory and Parametric Comparison. In Roberts, Ian (ed.), *The Oxford Handbook of Universal Grammar*. Oxford, UK: Oxford University Press. 376-398. <doi.org/10.1093/oxfordhb/9780199573776.013.16>.

Guardiano, Cristina; Longobardi, Giuseppe; Cordoni, Guido & Crisma, Paola 2020. Formal Syntax as a Phylogenetic Method. In Janda, Richard D.; Joseph, Brian D. & Vance, Barbara S. (eds.), *The Handbook of Historical Linguistics* (1st ed.). Wiley. 145-182. <doi.org/10.1002/9781118732168.ch7>.

Guasti, Maria Teresa 2017. *Language acquisition: The growth of grammar.* Cambridge, MA: MIT Press.

Guéron, Jacqueline & May, Robert 1984. Extraposition and Logical Form. *Linguistic Inquiry* 15,1. 1-31.

Gulordava, Kristina; Bojanowski, Piotr; Grave, Edouard; Linzen, Tal & Baroni, Marco 2018. Colorless green recurrent networks dream hierarchically. In Walker, Marilyn; Ji, Heng & Stent, Amanda (eds.), *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. New Orleans, LA: Association for Computational Linguistics. 1195-1205. <DOI: 10.18653/v1/N18-1108>.

Haider, Hubert 2023. Is Chat-GPT a grammatically competent informant? <lingbuzz/007285>.

Hale, John 2001. A Probabilistic Earley Parser as a Psycholinguistic Model.

Second Meeting of the North American Chapter of the Association for Computational Linguistics. <aclanthology.org/N01-1021>.

Hale, John 2011. What a rational parser would do. *Cognitive Science* 35,3. 399-443.

Hale, John 2016. Information-theoretical Complexity Metrics. *Language and Linguistics Compass* 10,9. 397-412. <doi.org/10.1111/lnc3.12196>.

Hanson, Kenneth 2025. Tier-based strict locality and the typology of agreement. *Journal of Language Modelling* 13,1. 43-97. <doi.org/10.15398/jlm.v13i1.411>.

Hanson, Kenneth 2024. Tiers, paths, and syntactic locality: The view from learning. In *Proceedings of the society for computation in linguistics (SCiL) 2024*. 107-116. <doi.org/10.7275/scil.2135>.

Hao, Sophie 2022. *Theory and Applications of Attribution for Interpretable Language Technology*. PhD dissertation. Yale University, New Haven, CT.

Hao, Sophie; Angluin, Dana & Frank, Robert 2022. Formal language recognition by hard attention transformers: Perspectives from circuit complexity. *Transactions of the Association for Computational Linguistics* 10. 800-810. <DOI: 10.1162/tacl_a_00490>.

Hao, Sophie; Mendelsohn, Simon; Sterneck, Rachel; Martinez, Randi & Frank, Robert 2020. Probabilistic predictions of people perusing: Evaluating metrics of language model performance for psycholinguistic modeling. In Chersoni, Emmanuele; Jacobs, Cassandra; Oseki, Yohei; Prévot, Laurent & Santus, Enrico (eds.), *Workshop on Cognitive Modeling and Computational Linguistics*. Online: Association for Computational Linguistics.

Hao, Sophie & Andersson, Samuel 2019. Unbounded stress in subregular phonology. In *Proceedings of the 16th Sigmorphon workshop on computational research in phonetics, phonology and morphology*. 135-143. <doi.org/10.18653/v1/W19-4216>.

Hart, Betty & Risley, Todd R. 1992. American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology* 28,6. 1096-1105. <doi.org/10.1037/0012-1649.28.6.1096>.

Haspelmath, Martin 1993. *A grammar of Lezgian.* Mouton Grammar Library 9. Berlin: Mouton de Gruyter.

Haspelmath, Martin 2007. Pre-established categories don't exist – consequences for language description and typology. *Linguistic Typology* 11. 119-132.

Haspelmath, Martin 2008. Parametric versus functional explanations of syntactic universals. In Biberauer, Theresa (ed.), *The limits of syntactic variation.* Amsterdam: Benjamins. Accessed 27 May 2016.

Haspelmath, Martin 2010a. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86,3. 663-687. <doi:10.1353/lan.2010.0021>.

Haspelmath, Martin 2010b. Framework-free grammatical theory. In Heine,

Bernd & Narrog, Heiko (eds.), *The Oxford Handbook of Linguistic Analysis.* Oxford: Oxford University Press. 341-365.

Haspelmath, Martin 2018. How comparative concepts and descriptive linguistic categories are different. In Van Olmen, Daniël; Mortelmans, Tanja & Brisard, Frank (eds.), *Aspects of linguistic variation: Studies in honor of Johan van der Auwera.* Berlin: De Gruyter Mouton. 83-113. <zenodo.org/record/3519206>.

Haspelmath, Martin 2020. Human linguisticality and the building blocks of languages. *Frontiers in Psychology* 10,3056. 1-10. <doi:10.3389/fpsyg.2019.03056>.

Haspelmath, Martin 2021. General linguistics must be based on universals (or nonconventional aspects of language). *Theoretical Linguistics* 47,1-2. 1-31. <doi:10.1515/tl-2021-2002>.

Haspelmath, Martin *to appear*. Breadth versus depth: Theoretical reasons for system-independent comparison of languages. In Nefdt, Ryan (ed.), *Oxford Handbook of Philosophy of Linguistics.* Oxford: Oxford University Press. <ling.auf.net/lingbuzz/008437>.

Hauser, M. D.; Chomsky, N. & Fitch, W. T. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298 (5598). 1569-1579. <doi.org/10.1126/science.298.5598.1569>.

Hawkins, John A. 2014. *Cross-linguistic variation and efficiency*. New York: Oxford University Press.

Heim, Johannes & Wiltschko, Martina 2025. Rethinking structural growth: Insights from the acquisition of interactional language. *Glossa: A journal of general linguistics* 10,1. <doi.org/10.16995/glossa.16396>.

Heinz, Jeffrey 2010. Learning long-distance phonotactics. *Linguistic Inquiry* 41. 623-661. <doi.org/10.1162/LING_a_00015>.

Heinz, Jeffrey 2018. The computational nature of phonological generalizations. In Hyman, Larry & Plank, Frank (eds.), *Phonological typology*. Mouton De Gruyter. 126-195.

Hewitt, John & Manning, Christopher D. 2019. A structural probe for finding syntax in word representation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 4129-4138.

Hey, Tony; Tansley, Stewart; Tolle, Kristin & Gray, Jim (eds.) 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft Research.

Hinton, Geoffrey 2022. The forward-forward algorithm: Some preliminary investigations. <arXiv:2212.13345>.

Hochreiter, Sepp; Bengio, Yoshua; Frasconi, Paolo & Schmidhuber, Jürgen 2001. Gradient flow in recurrent nets: The difficulty of learning long-term dependencies. In Kremer, S. C. & Kolen, J. F. (eds.), *A Field Guide to Dynamical Recurrent Neural Networks*. IEEE Press.

Hochreiter, Sepp & Schmidhuber, Jürgen 1997. Long short-term memory. *Neural Computation* 9,8. 1735-1780.

Hockenmaier, Julia & Steedman, Mark 2007. CCGbank: A corpus of CCG derivations and dependency structures extracted from the Penn Treebank. *Computational Linguistics* 33,3. 355-396. <DOI: 10.1162/coli.2007.33.3.355>.

Holmes, V. M. & Forster, K. 1972. Click location and syntactic structure. *Perception and Psychophysics* 12. 9-15. <doi.org/10.3758/bf03212836>.

Hornik, Kurt; Stinchcombe, Maxwell & White, Halbert 1989. Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks* 2,5. 359-66. <doi.org/10.1016/0893-6080(89)90020-8>.

Hosseini, Eghbal A. *et al.* 2024. Artificial neural network language models align neurally and behaviorally with humans even after a developmentally realistic amount of training. *Neurobiology of Language.* Apr 1.5,1. 43-63.

Hsu, Anne S. & Chater, Nick 2010. The Logical Problem of Language Acquisition: A Probabilistic Perspective. *Cognitive Science* 34,6. 972-1016. <doi.org/10.1111/j.1551-6709.2010.01117.x>.

Hsu, Anne S.; Chater, Nick & Vitányi, Paul 2013. Language Learning From Positive Evidence, Reconsidered: A Simplicity-Based Approach. *Topics in Cognitive Science* 5,1. 35-55. <doi.org/10.1111/tops.12005>.

Hu, Jennifer; Gauthier, Jon; Qian, Peng; Wilcox, Ethan & Levy, Roger 2020. A Systematic Assessment of Syntactic Generalization in Neural Language Models. In Jurafsky, Dan; Chai, Joyce; Schluter, Natalie & Tetreault, Joel (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics. 1725-1744. <doi.org/10.18653/v1/2020.acl-main.158>.

Hu, Michael Y.; Mueller, Aaron; Ross, Candace; Williams, Adina; Linzen, Tal; Zhuang, Chengxu; Cotterell, Ryan; Choshen, Leshem; Warstadt, Alex & Wilcox, Ethan 2024. Findings of the Second BabyLM Challenge: Sample-Efficient Pretraining on Developmentally Plausible Corpora. <doi.org/10.48550/ARXIV.2412.05149>.

Huang, C.-T. James 1982. *Logical relations in Chinese and the theory of grammar.* Cambridge, MA: MIT Press.

Huang, Lei; Yu, Weijiang; Ma, Weitao; Zhong, Weihong; Feng, Zhangyin; Wang, Haotian; Chen, Qianglong; Peng, Weihua; Feng, Xiaocheng; Qin, Bing *et al. 2023. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions.* <arXiv:2311.05232>.

Huh, Minyoung; Cheung, Brian; Wang, Tongzhou & Isola, Phillip 2024. Position: The Platonic Representation Hypothesis. In Salakhutdinov, Ruslan; Kolter, Zico; Heller, Katherine; Weller, Adrian; Oliver, Nuria; Scarlett, Jonathan & Berkenkamp, Felix (eds.), *Proceedings of the 41st International Conference on Machine Learning.* 235. 20617-42. Proceedings of Machine Learning Research. PMLR. <proceedings.mlr.press/v235/huh24a.html>.

Hume, David 1739. *A Treatise of Human Nature: Being an Attempt to Introduce*

the Experimental Method of Reasoning Into Moral Subjects*. London: John Noon.

Hume, David 1748. *Philosophical Essays Concerning Human Understanding*. London: A. Millar.

Hunter, Tim; Stanojević, Miloš & Stabler, Edward P. 2019. The active-filler strategy in a move-eager left-corner Minimalist grammar parser. In *Proceedings of the workshop on cognitive modeling and computational linguistics*. 1-10.

Ibbotson, Paul & Tomasello, Michael 2016. Evidence rebuts Chomsky's theory of language learning. *Scientific American* 315, 5. 70.

İdrisoğlu, Işıl & Spaniel, William 2024. *Information problems and Russia's invasion of Ukraine. Conflict Management and Peace Science* 41,5. 514-533. <DOI: 10.1177/07388942241238583>.

Ionin, Tania & Matushansky, Ora 2006. The composition of complex cardinals. *Journal of Semantics* 16. 315-360.

Jackendoff, Ray 1988. Why are they saying these things about us? *Natural Language and Linguistic Theory* 6,3. 435-442.

Jardine, Adam 2016. Computationally, tone is different. *Phonology* 33. 247-283. <doi.org/10.1017/S0952675716000129>.

Ji, Ziwei; Lee, Nayeon; Frieske, Rita; Yu, Tiezheng; Su, Dan; Xu, Yan; Ishii, Etsuko; Bang, Ye Jin; Madotto, Andrea & Fung, Pascale 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys* 55,12. 248:1-248:38. <DOI: 10.1145/3571730>.

Jurafsky, Dan & Martin, James H. 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd edition. Russell, Stuart & Norvig, Peter (eds.). Upper Saddle River, NJ: Prentice Hall.

Kalouli, Aikaterini-Lida 2021. *Hy-NLI: A hybrid system for state-of-the-art natural language inference*. University of Konstanz dissertation.

Kalouli, Aikaterini-Lida; Crouch, Richard & de Paiva, Valeria 2020. Hy-NLI: A hybrid system for natural language inference. In *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona, Spain (Online): International Committee on Computational Linguistics. 5235-5249. <aclanthology.org/2020.coling-main.459>.

Kaplan, Jared; McCandlish, Sam; Henighan, Tom; Brown, Tom B.; Chess, Benjamin; Child, Rewon; Gray, Scott; Radford, Alec; Wu, Jeffrey & Amodei, Dario 2020. *Scaling Laws for Neural Language Models*. <doi.org/10.48550/ARXIV.2001.08361>.

Kaplan, Ronald M. 1987. Three seductions of computational linguistics. In Whitelock, P.; Wood, M. M.; Somers, H.; Johnson, R. & Bennett, P. (eds.), *Linguistic Theory and Computer Applications*. London: Academic Press. 149-188.

Kaplan, Ronald M. 2019. Computational psycholinguistics. *Computational Linguistics* 45,4. 607-626. <doi:10.1162/coli a 00359>. <aclanthology.org/J19-4001>.

Kaplan, Ronald M.; King, Tracey H. & Maxwell, John T. III 2002. Adapting

existing grammars: The XLE experience. In *COLING-02: Grammar Engineering and Evaluation*.

Katz, Phillip 1986. PKZIP. Commercial Compression System, Version 1.1. <www.pkware.com/pkzip>.

Katzir, Roni 2023. *Why large language models are poor theories of human linguistic cognition. A reply to Piantadosi (2023)* [Lingbuzz]. <lingbuzz/007190>. *Biolinguistics* 17. <doi.org/10.5964/bioling.13153>.

Kawahara, Shigeto; Noto, Atsushi & Kumagai, Gakuji 2018. Sound symbolic patterns in Pokémon names. *Phonetica* 75,3. 219-244. <DOI: 10.1159/000484938>.

Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Keine, Stefan 2020. *Probes and their horizons*. Cambridge, MA: MIT Press.

Kempson, Ruth; Meyer Viol, Wilfried & Gabbay, Dov M. 2001. *Dynamic Syntax: The Flow of Language Understanding*. Wiley.

Kennedy, Christopher 2015. A "de-Fregean" semantics (and neo-Gricean pragmatics) for modified and unmodified numerals. *Semantics & Pragmatics* 8. 1-44. <dx.doi.org/10.3765/sp.8.1>.

Kerr, Dara 2024. How Memphis became a battleground over Elon Musk's xAI supercomputer. *NPR* 11 September 2024. <www.npr.org/2024/09/11/nx-s1-5088134/elon-musk-ai-xai-supercomputer-memphis-pollution>.

Kharitonov, Eugene & Chaabouni, Rahma 2021. What they do when in doubt: A study of inductive biases in seq2seq learners. In *ICLR 2021 Conference Track*. Online: OpenReview.

Kim, Najoung; Patel, Roma; Poliak, Adam; Wang, Alex; Xia, Patrick; McCoy, R. Thomas; Tenney, Ian; Ross, Alexis; Linzen, Tal & van Durme, Benjamin 2019. Probing what different NLP tasks teach machines about function word comprehension. <arXiv:1904.11544>.

Kingma, Diederik P. & Ba, Jimmy Lei 2015. Adam: A method for stochastic optimization. In *ICLR 2015 Conference Track*. San Diego, CA: OpenReview.

Kirov, Christo & Cotterell, Ryan 2018. Recurrent Neural Networks in Linguistic Theory: Revisiting Pinker and Prince (1988) and the Past Tense Debate. *Transactions of the Association for Computational Linguistics* 6 (December). 651-665. <doi.org/10.1162/tacl_a_00247>.

Kitaev, Nikita; Cao, Steven & Klein, Daniel 2019. Multilingual constituency parsing with self-attention and pre-training. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (ACL 2019). 3499-3505.

Kitchin, Rob 2014. Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1,1. <DOI: 10.1177/2053951714528481>.

Klein, Daniel & Manning, Christopher D. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Meeting of the Association for Computational Linguistics*. 423-430.

Kleyko, Denis; Rachkovskij, Dmitri; Osipov, Evgeny & Rahimi, Abbas

2023. A survey on hyperdimensional computing aka vector symbolic architectures, parts 1 and 2. *ACM Computing Surveys* 55. 130. <doi.org/10.1145/3538531>.

Klimova, Blanka; Pikhart, Marcel & Al-Obaydi, Liqaa Habeb 2024. Exploring the potential of ChatGPT for foreign language education at the university level. *Frontiers in Psychology* 15. <DOI: 10.3389/fpsyg.2024.1269319>.

Knight, Chris 2016. *Decoding Chomsky: Science and Revolutionary Politics*. New Haven, CT: Yale University Press. <DOI: 10.12987/9780300222159>.

Kobele, Gregory M. 2023. Minimalist Grammars and Decomposition. In Kleanthes, Grohmann & Leivada, Evelina (eds.), *The Cambridge Handbook of Minimalism*. Cambridge University Press.

Kobele, Gregory M.; Gerth, Sabrina & Hale, John T. 2013. Memory resource allocation in top-down Minimalist parsing. In Morrill, Glyn & Nederhof, Mark-Jan (eds.), *Formal grammar: 17th and 18th international conferences, FG 2012, Opole, Poland, August 2012, Revised selected papers, FG 2013, Düsseldorf, Germany, August 2013*. 32-51. Berlin / Heidelberg: Springer. <doi.org/10.1007/978-3-642-39998-5_3>.

Kodner, Jordan; Payne, Sarah & Heinz, Jeffrey 2023. Why linguistics will thrive in the 21st century: A reply to Piantadosi (2023). <arxiv.org/abs/2308.03228>.

Koerner, Konrad 1983. The Chomskyan 'revolution' and its historiography: A few critical remarks. *Language & Communication* 3,2. 147-169. <DOI: 10.1016/0271-5309(83)90012-5>.

Kojima, Takeshi; Gu, Shixiang (Shane); Reid, Machel; Matsuo, Yutaka & Iwasawa, Yusuke 2022. Large language models are zero-shot reasoners. In Koyejo, S.; Mohamed, S.; Agarwal, Al; Belgrave, D.; Cho, K. & Oh, A. (eds.), *Advances in Neural Information Processing Systems 35 (NeurIPS 2022) Main Conference Track*. New Orleans, LA: Curran Associates, Inc. 22199-22213.

Kolmogorov, Andrey N. 1963. On Tables of Random Numbers. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* 25,4. 369-376.

Kuhn, Thomas 1962. *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.

Kwon, Diana 2024. AI is complicating plagiarism. How should scientists respond? *Nature*. <DOI: 10.1038/d41586-024-02371-z>.

Lake, Brenden M. & Baroni, Marco 2023. Human-like systematic generalization through a meta-learning neural network. *Nature* 623. 115-121. <doi.org/10.1038/s41586-023-06668-3>.

Lakretz, Yair; Hupkes, Dieuwke; Vergallito, Alessandra; Marelli, Marco; Baroni, Marco & Dehaene, Stanislas 2021. Mechanisms for handling nested dependencies in neural-network language models and humans. *Cognition* 213. 1-24. <DOI: 10.1016/j.cognition.2021.104699>. <www.sciencedirect.com/science/article/pii/S0010027721001189>.

Lakretz, Yair; Kruszewski, German; Desbordes, Theo; Hupkes, Dieuwke; Dehaene, Stanislas & Baroni, Marco 2019. The emergence of number and syntax units in LSTM language models. In Burstein, Jill; Doran, Christy & Solorio, Thamar (eds.)*, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Vol. 1*. Minneapolis, MN: Association for Computational Linguistics. 11-20. <DOI: 10.18653/v1/N19-1002>.

Lan, Nur; Chemla, Emmanuel & Katzir, Roni 2024. Large language models and the argument from the poverty of the stimulus. *Linguistic Inquiry*. 1-28. <doi.org/10.1162/ling_a_00533>.

Lan, Nur; Geyer, Michal; Chemla, Emmanuel & Katzir, Roni 2022. Minimum Description Length Recurrent Neural Networks. *Transactions of the Association for Computational Linguistics* 10 (July). 785-99. <doi.org/10.1162/tacl_a_00489>.

Landman, Fred 2003. Predicate-argument mismatches and the adjectival theory of indefinites. In Coene, M. & d'Hulst, Y. (eds.), *From NP to DP: The syntax and semantics of noun phrases*. Volume 1. 211-237. Amsterdam: John Benjamins.

Lasnik, Howard & Lidz, Jeffrey L. 2016. The argument from the poverty of the stimulus. In Roberts, Ian (ed.), *The Oxford Handbook of Universal Grammar*. Oxford: Oxford University Press. 221-248.

Latour, Bruno 1984. *Les Microbes: Guerre et paix, suivi de Irréductions*. Paris, France: A. M. Métailié.

Law, John & Lodge, Peter 1984. *Science for Social Scientists*. London: Palgrave Macmillan UK. <DOI: 10.1007/978-1-349-17536-9>.

Lawson, Alex 2024. Google to buy nuclear power for AI datacentres in 'world first' deal. *Guardian* 15 October 2024. <www.theguardian.com/technology/2024/oct/15/google-buy-nuclear-power-ai-datacentres-kairos-power>.

Lee, So Young & De Santo, Aniello. A computational view into the structure of attachment ambiguities in Chinese and Korean. In *Proceedings of the north east linguistics society*. 189-198.

Levesque, Hector J. 2014. On our best behaviour. *Artificial Intelligence* 212. 27-35. <doi.org/10.1016/j.artint.2014.03.007>.

Levshina, Natalia 2023. *Communicative efficiency: Language structure and use*. Cambridge: Cambridge University Press.

Levy, Roger 2008. Expectation-based syntactic comprehension. *Cognition* 106,3. 1126-1177.

Li, Jixing; Bhattasali, Shohini; Zhang, Shulin; Franzluebbers, Berta; Luh, Wen-Ming; Spreng, R. Nathan; Brennan, Jonathan R.; Yang, Yiming; Pallier, Christophe & Hale, John 2022. *Le Petit Prince* multilingual naturalistic fMRI corpus. *Scientific Data* 9. 530. <doi.org/10.1038/s41597-022-01625-7>.

Li, Jixing & Hale, John 2019. Grammatical predictors for fMRI time-courses.

In Berwick, Robert C. & Stabler, Edward P. (eds.), *Minimalist Parsing*. Oxford, UK: Oxford University Press. 159-173. <doi.org/10.1093/oso/9780198795087.003.0007>.

Li, Ming & Vitányi, Paul 2008. *An Introduction to Kolmogorov Complexity and Its Applications*. New York: Springer. <doi.org/10.1007/978-0-387-49820-1>.

Lidz, Jeffrey & Gleitman, Lila R. 2004. Argument structure and the child's contribution to language learning. *Trends in Cognitive Sciences* 8,4.

Lillicrap, Timothy P.; Santoro, Adam; Marris, Luke; Akerman, Colin J. & Hinton, Geoffrey 2020. Backpropagation and the Brain. *Nature Reviews Neuroscience* 21,6. 335-46. <doi.org/10.1038/s41583-020-0277-3>.

Lin, Stephanie; Hilton, Jacob & Evans, Owain 2022. TruthfulQA: Measuring how models mimic human falsehoods. In Muresan, Smaranda; Nakov, Preslav & Villavicencio, Aline (eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Vol. 1*. Dublin, Ireland: Association for Computational Linguistics. 3214-3252. <DOI: 10.18653/v1/2022.acl-long.229>.

Ling, Jacqueline 2001. Power of a human brain. In *Physics Factbook*. <hypertextbook.com/facts/2001/JacquelineLing.shtml>.

Link, Godehard 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In Bauerle, Rainer; Schwarze, Christoph & von Stechow, Arnim (eds.), *Meaning, Use, and the Interpretation of Language*. Berlin / New York: de Gruyter. 302-323.

Linzen, Tal & Baroni, Marco 2021. Syntactic structure from deep learning. *Annual Review of Linguistics* 7. 195-212. <DOI: 10.1146/annurev-linguistics-032020-051035>.

Linzen, Tal; Dupoux, Emmanuel & Goldberg, Yoav 2016. Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies. *Transactions of the Association for Computational Linguistics* 4. 521-535. <doi.org/10.1162/tacl_a_00115>.

Liu, Lei 2023. Processing advantages of end-weight. *Proceedings of the Society for Computation in Linguistics* 6. 250-258.

Lohninger, Magdalena & Wurmbrand, Susi 2025. Typology of Complement Clauses. In Benz, Anton; Frey, Werner; Gärtner, Hans-Martin; Krifka, Manfred; Schenner, Mathias & Żygis, Marzena (eds.), *Handbook of clausal embedding*. Berlin: Language Science Press.

Longobardi, Giuseppe 1994. Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic Inquiry* 25. 609-665.

Manning, Christopher D.; Clark, Kevin; Hewitt, John; Khandelwal, Uravashi & Levy, Omer 2020. Emergent linguistic structure in artificial neural networks trained by self-supervision. In Gavish, Matan (ed.), *Proceedings of the National Academy of Science of the United States of America* 117. 30046-30054. <DOI: 10.1073/pnas.1907367117>.

Manzini, Maria Rita 1983. Syntactic conditions on phonological rules. *MIT Working Papers in Linguistics* 5. 1-9.

Marantz, Alec 2019. What do linguists do? In *The Julius Silver, Roslyn S. Silver, and Enid Silver Winslow Dialogues in Arts and Science, New York University*. <as.nyu.edu/content/dam/nyu-as/as/documents/silverdialogues/SilverDialogues_Marantz.pdf>.

Marcus, Gary 2022. Noam Chomsky and GPT-3 [Blog post]. *Marcus on AI*. <garymarcus.substack.com/p/noam-chomsky-and-gpt-3>. Last accessed 24/02/2025.

Marcus, Mitchell *et al*. 1994. The Penn Treebank: Annotating predicate argument structure. In *Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey*.

Marr, David 1982. *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: Freeman.

Marr, David & Poggio, Tomaso 1976. *From Understanding Computation to Understanding Neural Circuitry*. Cambridge, MA: MIT Press.

Martinetz, Julius; Linse, Christoph & Martinetz, Thomas 2024. Rethinking generalization of classifiers in separable classes scenarios and over-parameterized regimes. *International Joint Conference on Neural Networks* 2024. 1-10. <doi.org/10.1109/IJCNN60899.2024.10650680>.

Marvin, Rebecca & Linzen, Tal 2018. Targeted Syntactic Evaluation of Language Models. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 1192-1202. <doi.org/10.18653/v1/D18-1151>.

May, Robert 1985. *Logical form: Its structure and derivation* (Vol. 12). Cambridge, MA: MIT Press.

Mayer, Connor & Major, Travis 2018. A challenge for tier-based strict locality from Uyghur backness harmony. In Foret, Annie; Kobele, Greg & Pogodalla, Sylvain (eds.), *Proceedings of formal grammar 2018*. Berlin: Springer. 62-83.

McCawley, James D. 1976. Introduction. In McCawley, James D. (ed.), *Notes From the Linguistic Underground*. New York, NY: Academic Press. 1-19.

McClelland, James L. & Rumelhart, David E. 1991. *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*. 2nd print. Computational Models of Cognition and Perception. Cambridge, MA: MIT Press.

McCoy, Richard; Frank, Robert & Linzen, Tal 2018. Revisiting the poverty of the stimulus: Hierarchical generalization without a hierarchical bias in recurrent neural networks. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. Madison, WI: Cognitive Science Society. 2096-2101.

McCoy, R. Thomas; Yao, Shunyu; Friedman, Dan; Hardy, Matthew & Griffiths, Thomas L. 2023. Embers of autoregression: Understanding large language models through the problem they are trained to solve. <arxiv.org/abs/2309.13638>.

McCullough, Gretchen 2019. *Because Internet: Understanding the New Rules of Language*. New York, NY: Riverhead Books.

McGee, Thomas & Blank, Idan 2024. Evidence against syntactic encapsulation in large language models. *Procs. Cognitive Science Society* 46.

McKenzie, Ian R.; Lyzhov, Alexander; Pieler, Michael Martin; Parrish, Alicia; Mueller, Aaron; Prabhu, Ameya; McLean, Euan; Shen, Xudong; Cavanagh, Joe, Gritsevskiy, Andrew George *et al.* 2023. Inverse scaling: When bigger isn't better. *Transactions on Machine Learning Research*.

McNally, Louise & Boleda, Gemma 2004. Relational adjectives as properties of kinds. *Empirical Issues in Syntax and Semantics* 5. 179-196. <doi.org/ISSN1769-7158>.

Merrill, William; Sabharwal, Ashish & Smith, Noah A. 2022. Saturated transformers are constant-depth threshold circuits. *Transactions of the Association for Computational Linguistics* 10. 843-856. <DOI: 10.1162/tacl_a_00493>.

Michaelis, Jens 2001. Derivational Minimalism Is Mildly Context-Sensitive. In Moortgat, Michael (ed.), *Logical Aspects of Computational Linguistics* (Vol. 2014). Berlin / Heidelberg: Springer. 179-198. <doi.org/10.1007/3-540-45738-0_11>.

Mikolov, Tomáš 2012. *Statistical Language Models Based on Neural Networks*. PhD dissertation. Brno University of Technology, Brno, Czech Republic.

Milewski, Bartosz 2020. *Category Theory for Programmers*. <bartoszmilewski.com>.

Miller, George A. & Chomsky, Noam 1963. Finitary Models of Language Users. In Luce, D. (ed.), *Handbook of Mathematical Psychology*. John Wiley & Sons. 2-419.

Milway, Daniel 2023. A response to Piantadosi (2023). <ling-buzz/007264>.

Mishra, Swaroop; Khashabi, Daniel; Baral, Chitta & Hajishirzi, Hannaneh 2022. Cross-task generalization via natural language crowdsourcing instructions. In Muresan, Smaranda; Nakov, Preslav & Villavicencio, Aline (eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Vol. 1*. Dublin, Ireland: Association for Computational Linguistics. 3470-3487. <DOI: 10.18653/v1/2022.acl-long.244>.

Mitchell, J.; Kazanina, Nina; Houghton, Conor J.; Bowers, Jeffrey S. 2019. Do LSTMs know about Principle C? In *2019 Conference on Cognitive Computational Neuroscience*.

Mollica, Frank & Piantadosi, Stephen 2019. Humans store about 1.5 megabytes of information during language acquisition. *Royal Society Open Science* 6,3.

Mollica, Frank & Piantadosi, Stephen 2022. Meaning without reference in large language models. <arXiv:2208.02957>.

Momma, Shota & Phillips, Colin 2018. The Relationship Between Parsing and Generation. *Annual Review of Linguistics* 4,1. 233-254. <doi.org/10.1146/annurev-linguistics-011817-045719>.

Moro, Andrea 2023. Embodied syntax: Impossible languages and the irreducible difference between humans and machines. *Sistemi intelligenti* 2.

321-328. <doi.org/10.1422/108132>.

Moro, Andrea; Greco, Matteo & Cappa, Stefano F. 2023. Large languages, impossible languages and human brains. *Cortex* 167. 82-85. <doi.org/10.1016/j.cortex.2023.07.003>.

Müller, Stefan 2024. Large language models: The best linguistic theory, a wrong linguistic theory, or no linguistic theory at all. *Zeitschrift für Sprachwissenschaft.*

Mullins, Nicholas C. 1975. A sociological theory of scientific revolution. In Knorr, Karin D.; Strasser, Hermann & Zilian, Hans Georg (eds.), *Determinants and Controls of Scientific Development.* Dordrecht, Netherlands: Springer Netherlands. 185-203.

Murray, Stephen O. 1994. *Theory Groups and the Study of Language in North America.* Amsterdam, Netherlands: John Benjamins.

Murty, Shikhar; Sharma, Pratyusha; Andreas, Jacob & Manning, Christopher D. 2022. Characterizing intrinsic compositionality in transformers with tree projections.

Naveed, Humza; Asad Ullah Khan; Shi Qiu; Saqib, Muhammad; Anwar, Saeed; Usman, Muhammad; Akhtar, Naveed; Barnes, Nick & Mian, Ajmal 2024. A comprehensive overview of large language models. <arxiv.org/abs/2307.06435>.

Newmeyer, Frederick J. 1980. *Linguistic theory in America: The first quarter century of Transformational Generative Grammar.* New York: Academic Press.

Newmeyer, Frederick J. 1986. Has there been a 'Chomskyan revolution' in linguistics? *Language* 62,1. 1-18. <DOI: 10.2307/415597>.

Newmeyer, Frederick J. 2004. Against a parameter-setting approach to typological variation. *Linguistic Variation Yearbook* 4,1. 181-234. <doi:10.1075/livy.4.06new>.

Newmeyer, Frederick J. 2021. Complexity and relative complexity in generative grammar. *Frontiers in Communication* 6. <doi:10.3389/fcomm.2021.614352>.

Newmeyer, Frederick J. & Emonds, Joseph 1971. The linguist in American society. In *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society.* Chicago, IL: Chicago Linguistic Society. 285-303.

Nivre, Joakim; Agić, Željko; Ahrenberg, Lars; Antonsen, Lene; Aranzabe, Maria Jesus; Asahara, Masayuki; Ateyah, Luma; Attia, M.; Atutxa, A.; Augustinus, L. *et al.* 2017. *Universal Dependencies 2.1.*

Norvig, Peter 2017. On Chomsky and the two cultures of statistical learning. In Pietsch, Wolfgang; Wernecke, Jörg & Ott, Maximilian (eds.), *Berechenbarkeit der Welt? Philosophie und Wissenschaft im Zeitalter von Big Data.* Wiesbaden, Germany: Springer Fachmedien. 61-83.

Nosengo, Nicola 2014. *I robot ci guardano: Aerei senza pilota, chirurghi a distanza e automi solidali.* Bologna: Zanichelli.

Noy, Shakked & Zhang, Whitney 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381, 6654. 187-192. <DOI: 10.1126/science.adh2586>.

Nvidia *n.d.* meta/llama-3.1-405b-instruct. *Nvidia API reference.* <docs.api.nvidia.com/nim/reference/meta-llama-3_1-405b>.

Nye, Maxwell; Andreassen, Anders Johan; Gur-Ari, Guy; Michalewski, Henryk; Austin, Jacob; Bieber, David; Dohan, David; Lewkowycz, Aitor; Bosma, Maarten; Luan, David; Sutton, Charles & Odena, Augustus (2022). Show your work: Scratchpads for intermediate computation with language models. In *ICLR 2022 Workshop DL4C*. Online: OpenReview.

Oepen, Stephan; Toutanova, Kristina; Shieber, Stuart; Manning, Christopher; Flickinger, Dan & Brants, Thorsten 2022. The LinGO Redwoods treebank: Motivation and preliminary applications. In *COLING 2002: The 17th International Conference on Computational Linguistics: Project Notes*. Taipei, Taiwan: Association for Computational Linguistics.

Oerter, Robert 2006. *The theory of almost everything: The Standard Model, the unsung triumph of modern physics*. New York: Pi Press.

Oh, Byung-Doh & Schuler, William 2023. Why does surprisal from larger transformer-based language models provide a poorer fit to human reading times? *Transactions of the Association for Computational Linguistics* 11. 336-350. <DOI: 10.1162/tacl_a_00548>.

OpenAI 2023. *GPT-4 Technical Report* <arxiv.org/abs/2303.08774>.

Ouyang, Long; Wu, Jeff; Jiang, Xu; Almeida, Diogo; Wainwright, Carroll L.; Mishkin, Pamela; Zhang, Chong; Agarwal, Sandhini; Slama, Katarina; Ray, Alex *et al.* 2022. Training language models to follow instructions with human feedback. <arXiv:2203.02155>.

Ozaki, Satoru; Santo, Aniello De; Linzen, Tal & Dillon, Brian 2024. CCG parsing effort and surprisal jointly predict RT but underpredict garden-path effects. *Society for Computation in Linguistics* 7. 362-364. <doi.org/10.7275/scil.2229>.

Papineni, Kishore; Roukos, Salim; Ward, Todd & Zhu, Wei-Jing 2001. BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, 311. Philadelphia, Pennsylvania: Association for Computational Linguistics. <doi.org/10.3115/1073083.1073135>.

Park, Peter S.; Goldstein, Simon; O'Gara, Aidan; Chen, Michael & Hendrycks, Dan 2024. AI deception: A survey of examples, risks, and potential solutions. *Patterns* 5,5. 100988. <DOI: 10.1016/j.patter.2024.100988>.

Pascanu, Razvan; Mikolov, Tomas & Bengio, Yoshua 2013. On the difficulty of training recurrent neural networks. In Dasgupta, Sanjoy & McAllester, David (eds.), *ICML'13: Proceedings of the 30th International Conference on International Conference on Machine Learning. Vol. 28*. Atlanta, GA: Proceedings of Machine Learning Research. 1310-1318.

Pasternak, Robert & Graf, Thomas 2021. Cyclic scope and processing difficulty in a Minimalist parser. *Glossa* 6. 1-34. <doi.org/10.5334/gjgl.1209>.

Pasteur, Louis 1876. Études sur la bière, ses maladies, causes qui les provoquent, procédé pour la rendre inaltérable, avec une théorie nouvelle de la fermentation. Paris: Gauthier-Villars.

Pasteur, Louis 1880. De l'extension de la théorie des germes à l'étiologie de quelques maladies communes. In *Comptes rendus hebdomadaires des séances de l'Académie des sciences*. *Vol. 90*. Paris: Gauthier-Villars. 1033-1034.

Pasteur, Louis; Joubert, Jules & Chamberland, Charles 1878. La théorie des germes et ses applications à la médecine et à la chirurgie. In *Comptes rendus hebdomadaires des séances de l'Académie des sciences*. *Vol. 86*. Paris: Gauthier-Villars. 1037-1043.

Pater, Joe 2019. Generative linguistics and neural networks at 60: Foundation, friction, and fusion. *Language* 95. 41-74. <doi.org/10.1353/lan.2019.0009>.

Pearl, Lisa 2022. Poverty of the stimulus without tears. *Language Learning and Development* 18,4. 415-454. <DOI: 10.1080/15475441.2021.1981908>.

Pennington, Jeffrey; Socher, Richard & Manning, Christopher D. 2014. Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532-1543.

Pereira, Fernando 2000. Formal grammar and information theory: Together again? *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 358,1769. 1239-1253.

Perez, Ethan; Huang, Saffron; Song, Francis; Cai, Trevor; Ring, Roman; Aslanides, John; Glaese, Amelia; McAleese, Nat & Irving, Geoffrey 2022. Red teaming language models with language models. In Goldberg, Yoav; Kozareva, Zornitsa & Zhang, Yue (eds.), *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. 3419-3448. <DOI: 10.18653/v1/2022.emnlp-main.225>.

Pesetsky, David 2024. Is there an LLM challenge for Linguistics? Or is there a Linguistics challenge for LLMs?. Paper presented at the Academia Română, Bucarest, 22 May 2024.

Petroni, Fabio; Rocktäschel, Tim; Riedel, Sebastian; Lewis, Patrick; Bakhtin, Anton; Wu, Yuxiang & Miller, Alexander 2019. Language models as knowledge bases? In Inui, Kentaro; Jiang, Jing; Ng, Vincent & Wan, Xiaojun (eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics. 2463-2473. <DOI: 10.18653/v1/D19-1250>.

Phillips, Colin 1996. *Order and structure*. PhD dissertation. Cambridge, MA: MIT Press.

Phillips, Colin 2003. Linear order and constituency. *Linguistic Inquiry* 34. 37-90.

Piantadosi, Steven T. 2023. Modern language models refute Chomsky's approach to language. <lingbuzz.net/lingbuzz/007180>.

Piantadosi, Steven T. 2024. Modern language models refute Chomsky's approach to language. In Gibson, Edward & Poliak, Moshe (eds.), *From*

fieldwork to linguistic theory: A tribute to Dan Everett. Berlin: Language Science Press. 353-414.

Pinker, Steven 1984. *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Plate, Tony A. 1994. *Holographic Reduced Representation*. Stanford: CSLI.

Plato 380 BCE. *Meno*.

Poggio, Thomas; Rifkin, Ryan; Niyogi, Partha & Mukherjee, Sayan 2004. General conditions for predictivity in learning theory. *Nature* 428. 419-422. <doi.org/10.1038/nature02341>.

Pollard, Carl & Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, IL: University of Chicago Press.

Pollock, Jean Yves 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic inquiry* 20,3. 365-424.

Popper, Karl 1934. *Logik der Forschung*. Berlin: Springer. <doi.org/10.1007/978-3-7091-4177-9>.

Prasanna, Sai; Rogers, Anna & Rumshisky, Anna 2020. When BERT plays the lottery, all tickets are winning. In Webber, Bonnie; Cohn, Trevor; He, Yulan & Liu, Yang (eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics. 3208-3229. <DOI: 10.18653/v1/2020.emnlp-main.259>.

Pullum, Geoffrey K. & Scholz, Barbara C. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 18,1-2. 9. <DOI: 10.1515/tlir.19.1-2.9>.

Purnell, Thomas; Idsardi, William & Baugh, John 1999. Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology* 18,1. 10-30. <DOI: 10.1177/0261927X99018001002>.

Quine, Willard Van Orman 1960. *Word and Object*. Cambridge, MA: MIT Press. <doi.org/10.7551/mitpress/9636.001.0001>.

Quinlan, Philip T. (ed.) 2004. *Connectionist Models of Development* (0 ed.). Psychology Press. <doi.org/10.4324/9780203494028>.

Radford, Alec; Narasimhan, Karthik; Salimans, Tim; Sutskever, Ilya *et al.* 2018. *Improving language understanding by generative pre-training*.

Radford, Alec; Wu, Jeffrey; Amodei, Dario; Clark, Jack; Brundage, Miles & Sutskever, Ilya 2019a. Better language models and their implications [Blog post]. *OpenAI Research*. <openai.com/index/better-language-models>. Last accessed 24/02/2025.

Radford, Alec; Wu, Jeffrey; Child, Rewon; Luan, David; Amodei, Dario & Sutskever, Ilya 2019b. *Language Models Are Unsupervised Multitask Learners*. Technical report. San Francisco, CA: OpenAI.

Radford, Andrew 1997. *Syntax: A Minimalist Introduction*. Cambridge: Cambridge University Press.

Radford, Andrew 2016. *Analysing English Sentences, Second Edition*. Cambridge: Cambridge University Press.

Rafailov, Rafael; Sharma, Archit; Mitchell, Eric; Ermon, Stefano; Manning, Christopher D. & Finn, Chelsea 2023. Direct preference optimization: Your language model is secretly a reward model. In *ICLR 2023 Conference Track*. Kigali, Rwanda: OpenReview.

Raman, Raghu *et al.* 2024. Fake news research trends, linkages to generative artificial intelligence and sustainable development goals. *Heliyon* e24727. <DOI: 10.1016/j.heliyon.2024.e24727>.

Rasin, Ezer; Berger, Iddo; Lan, Nur; Shefi, Itamar & Katzir, Roni 2021. Approaching explanatory adequacy in phonology using minimum description length. *Journal of Language Modelling* 9,1. 17-66. <doi.org/10.15398/jlm.v9i1.266>.

Rawski, Jonathan & Heinz, Jeffrey 2019. No free lunch in linguistics or machine learning: Response to Pater. *Language* 95. 125-135.

Raymond, Louise & O'Reilly, Tim 1999. *The Cathedral and the Bazaar* (1st ed.). USA: O'Reilly & Associates, Inc.

Reinhart, Tanya 1976. *The syntactic domain of anaphora*. Cambridge, MA: MIT Press.

Retoré, Christian (ed.), *Logical Aspects of Computational Linguistics: Lecture Notes in Computer Science*. Berlin: Springer. 68-95.

Rickford, John R. & King, Sharese 2016. Language and linguistics on trial: Hearing Rachel Jeantel (and other vernacular speakers) in the courtroom and beyond. *Language* 92,4. 948-988.

Riesenhuber, Maximilian & Poggio, Tomaso 1999. Hierarchical models of object recognition in cortex. *Nature Neuroscience* 2,11. 1019-1025. <doi.org/10.1038/14819>.

Riezler, Stefan; Holloway King, Tracy; Kaplan, Ronald M.; Crouch, Richard; Maxwell, John T. III & Johnson, Mark 2002. Parsing the Wall Street Journal using a Lexical-Functional Grammar and discriminative estimation techniques. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia: Association for Computational Linguistics. 271-278.

Rissanen, Jorma 1978. Modeling by shortest data description. *Automatica* 14,5. 465-471. <doi.org/10.1016/0005-1098(78)90005-5>.

Rissanen, Jorma 1987. Stochastic Complexity. *Journal of the Royal Statistical Society: Series B (Methodological)* 49,3. 223-239. <doi.org/10.1111/j.2517-6161.1987.tb01694.x>.

Ritter, Elizabeth & Wiltschko, Martina 2014. The composition of INFL. An exploration of tense, tenseless languages and tenseless constructions. *Natural Language and Linguistic Theory* 32. 1331-1386.

Ritter, Elizabeth 1991. Two functional categories in Noun Phrases: Evidence from Modern Hebrew. *Syntax and Semantics* 25.

Rizzi, Luigi 1990. *Relativized minimality*. Cambridge, MA: MIT Press.

Rizzi, Luigi 1997. The Fine Structure of the Left Periphery. In Haegeman, Liliane (ed.), *Elements of Grammar*. Dordrecht: Springer Netherlands. 281-337. <doi.org/10.1007/978-94-011-5420-8_7>.

Rizzi, Luigi (ed.) 2004. *The structure of CP and IP*. Oxford, UK: Oxford University Press.

Rizzi, Luigi 2013. Locality. *Lingua* 130. 169-186.

Rizzi, Luigi 2016. Labeling, maximality and the head-phrase distinction. *The Linguistic Review* 33,1. 103-127.

Rizzi, Luigi 2021. *Complexité des structures linguistiques, simplicité des mécanismes du langage*, Leçon inaugurale, 2021, Collège de France – Fayard, Paris. English translation: *Complexity of Linguistic Structures, Simplicity of Language Mechanisms* (2024). OpenEdition Books, Collège de France. <DOI: 10.4000/books.cdf.16006>.

Rizzi, Luigi & Cinque, Guglielmo 2016. Functional Categories and Syntactic Theory. *Annual Review of Linguistics* 2,1. 139-163. <doi.org/10.1146/annurev-linguistics-011415-040827>.

Rizzi, Luigi & Savoia, Leonardo 1993. Conditions on /u/ propagation in Southern Italian Dialects: A Locality Parameter for Phonosyntactic Processes. In Belletti, A. (ed.), *Syntactic Theory and the Dialects of Italy*. Turin: Rosenberg & Sellier.

Roberts, Ian 2017. The final-over-final condition in DP: Universal 20 and the nature of demonstratives. In Sheehan, Michelle; Biberauer, Theresa; Roberts, Ian & Holmberg, Anders (eds.), *The Final-over-Final Condition: A Syntactic Universal* (Vol. 76). Cambridge, MA: MIT Press. 151.

Roberts, Ian 2019. *Parameter Hierarchies and Universal Grammar* (1st ed.). Oxford, UK: Oxford University Press. <doi.org/10.1093/oso/9780198804635.001.0001>.

Rogers, Anna; Kovaleva, Olga & Rumshisky, Anna 2021. A primer in BERTology: What we know about how BERT works. *Transactions of the Association for Computational Linguistics* 8. 842-866.

Ross, John Robert 1967. *Constraints on variables in syntax*. Cambridge, MA: MIT Press.

Rumelhart, David E.; Hinton, Geoffrey E. & Williams, Ronald J. 1986. Learning representations by back-propagating errors. *Nature* 323,6088. 533-536. <DOI: 10.1038/323533a0>.

Rumelhart, David E. & McClelland, James L. 1986. On Learning the Past Tenses of English Verbs. In *Parallel Distributed Processing*. Cambridge, MA: MIT Press. <doi.org/10.7551/mitpress/5237.003.0008>.

Rumelhart, David E.; McClelland, James L. & PDP Research Group (eds.) 1999. *Parallel distributed processing. 1: Foundations*. 12th print. Cambridge, MA: MIT Press. <doi.org/10.7551/mitpress/5236.001.0001>.

Russell, Bertrand 1947. *Human Knowledge: Its Scope and Limits*. New York, NY: Simon and Schuster.

Sampson, Geoffrey 1997. *Educating Eve: The 'language instinct' debate*. London / Washington, DC: Cassell.

Sanh, Victor; Webson, Albert; Raffel, Colin; Bach, Stephen; Sutawika, Lintang; Alyafeai, Zaid; Chaffin, Antoine; Stiegler, Arnaud; Raja, Arun;

Dey, Manan *et al. 2022. Multitask prompted training enables zero-shot task generalization. In ICLR 2022 Conference Track*. Online: OpenReview.

Sarlin, Paul-Edouard; DeTone, Daniel; Malisiewicz, Tomasz & Rabinovich, Andrew 2020. Superglue: Learning feature matching with graph neural networks. <arxiv.org/abs/1911.11763>.

Sartran, Laurent; Barrett, Samuel; Kuncoro, Adhiguna; Stanojević, Miloš; Blunsom, Phil & Dyer, Chris 2022. Transformer Grammars: Augmenting Transformer Language Models with Syntactic Inductive Biases at Scale. *Transactions of the Association for Computational Linguistics* 10 (December). 1423-39. <doi.org/10.1162/tacl_a_00526>.

Sathish, Vishwas; Lin, Hannah; Kamath, Aditya K. & Nyayachavadi, Anish 2024. LLeMpower: Understanding disparities in the control and access of large language models. <arXiv:2404.09356>.

Savitch, Walter J. 1993. Why it might pay to assume that languages are infinite. *Annals of Mathematics and Artificial Intelligence* 8. 17-25.

Shannon, C. E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27,3. 379-423. <DOI: 10.1002/j.1538-7305.1948.tb01338.x>.

Shieber, Stuart M. 1985. Evidence against the Context-Freeness of Natural Language. *Linguistics and Philosophy* 8,3. 333-43. <doi.org/10.1007/BF00630917>.

Siegelman, Noam; Schroeder, Sascha; Acartürk, Cengiz; Ahn, Hee-Don; Alexeeva, Svetlana; Amenta, Simona; Bertram, Raymond; Bonandrini, R.; Brysbaert, M.; Chernova, D.; Da Fonseca, S. M.; Dirix, N.; Duyck, W.; Fella, A.; Frost, R.; Gattei, C. A.; Kalaitzi, A.; Kwon, N.; Lõo, K.; … Kuperman, V. 2022. Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior Research Methods* 54,6. 2843-2863. <doi.org/10.3758/s13428-021-01772-6>.

Smith, Nathaniel J. & Levy, Roger 2013. The effect of word predictability on reading time is logarithmic. *Cognition* 128,3. 302-319. <DOI: 10.1016/j.cognition.2013.02.013>.

Smolensky, Paul 1990. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence* 46. 159-216. <doi.org/10.1016/0004-3702(90)90007-m>.

Solomonoff, Ray J. 1960. *A Preliminary Report on a General Theory of Inductive Inference*. United States Air Force, Office of Scientific Research. <books.google.it/books?id=SuTHtgAACAAJ>.

Spitale, Giovanni; Biller-Andorno, Nikola & Germani, Federico 2023. AI model GPT-3 (dis)informs us better than humans. *Science Advances* 9, 26. <DOI: 10.1126/sciadv.adh1850>.

Sprouse, Jon & Almeida, Diogo 2017. Design sensitivity and statistical power in acceptability judgment experiments. *Glossa* 2,1. 1-32. <doi.org/10.5334/gjgl.236>.

Sprouse, Jon & Hornstein, Norbert (eds.) 2013. *Experimental Syntax and*

Island Effects (1st ed.). Cambridge University Press. <doi.org/10.1017/CBO9781139035309>.

Srivastava, Aarohi; Rastogi, Abhinav; Rao, Abhishek; Shoeb, Abu Awal Md; Abid, Abubakar; Fisch, Adam; Brown, Adam R.; Santoro, Adam; Gupta, Aditya; Garriga-Alonso, Adrià *et al.* 2023. *Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models. Transactions on Machine Learning Research.*

Stabler, Edward 1991. Avoid the pedestrian's paradox. In Berwick, Robert C.; Abney, Steven P. & Tenny, Carol (eds.), *Principle-based Parsing: Computation and Psycholinguistics*. Dordrecht: Kluwer. 199-238. <doi.org/10.1007/978-94-011-3474-3_8>.

Stabler, Edward 1997. Derivational minimalism. In Retoré, Christian (ed.), *Logical Aspects of Computational Linguistics*. Berlin / Heidelberg: Springer. 68-95.

Stabler, Edward 2011. Computational Perspectives on Minimalism. In Boeckx, Cedric (ed.), *The Oxford Handbook of Linguistic Minimalism*. Oxford, UK: Oxford University Press. <doi.org/10.1093/oxfordhb/9780199549368.013.0027>.

Stabler, Edward 2013. Two Models of Minimalist, Incremental Syntactic Analysis. *Topics in Cognitive Science* 5,3. 611-633. <doi.org/10.1111/tops.12031>.

Starke, Michal 2001. *Move Dissolves into Merge: A Theory of Locality*. PhD dissertation. Université de Genève.

Steedman, Mark & Baldridge, Jason 2006. Combinatory categorial grammar. In Brown, Keith (ed.), *Encyclopedia of Language & Linguistics*. 2nd edition. Oxford: Elsevier. 610-621.

Steuer, Julius; Mosbach, Marius & Klakow, Dietrich 2023. Large GPT-like Models are Bad Babies: A Closer Look at the Relationship between Linguistic Competence and Psycholinguistic Measures. *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*. Singapore: Association for Computational Linguistics. 114-129. <doi.org/10.18653/v1/2023.conll-babylm.12>.

Stowe, Laurie A.; Kaan, Edith; Sabourin, Laura & Taylor, Ryan C. 2018. The sentence wrap-up dogma. *Cognition* 176. 232-247. <doi.org/10.1016/j.cognition.2018.03.011>.

Strubell, Emma; Ganesh, Ananya & McCallum, Andrew 2019. Energy and policy considerations for deep learning in NLP. In Korhonen, Anna; Traum, David & Màrquez, Lluís (eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics. 3645-3650. <DOI: 10.18653/v1/P19-1355>.

Sulger, Sebastian; Butt, Miriam; Holloway King, Tracy; Meurer, Paul; Laczkó, Tibor; Rákosi, György; Bamba Dione, Cheikh M.; Dyvik, Helge; Rosén, Victoria; De Smedt, Koenraad; Patejuk, Agnieszka; Çetinoglu, Özlem; Arka, I Wayan & Mistica, Meladel 2013. ParGramBank: The

ParGram parallel treebank. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, vol. 1. Sofia: Association for Computational Linguistics. 550-560. <www.aclweb.org/anthology/P13-1054.pdf>.

Sutton, Rich 2019. The bitter lesson [Blog post]. *Incomplete Ideas*. <www.incompleteideas.net/IncIdeas/BitterLesson.html>. Last accessed 24/02/2025.

Svenonius, Peter 2016. Significant mid-level results of generative linguistics. <blogg.uit.no/psv000/2016/08/30/significant-mid-level-results-of-generative-linguistics>.

Swanson, Logan 2024. Syntactic learning over tree tiers. In *Proceedings of ESSLLI 2024*. 187-196.

Taylor, Wilson L. 1953. "Cloze Procedure": A New Tool for Measuring Readability. *Journalism Quarterly* 30,4. 415-433. <doi.org/10.1177/107769905303000401>.

Torr, John 2017. Autobank: A semi-automatic annotation tool for developing deep Minimalist grammar treebanks. In *Proceedings of the demonstrations at the 15th conference of the European chapter of the Association for Computational Linguistics*. 81-86.

Torr, John 2018. Constraining MGbank: Agreement, L-selection and supertagging in minimalist grammars. In Gurevych, Iryna & Miyao, Yusuke (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Vol. 1.* Melbourne, Australia: Association for Computational Linguistics. 590-600. <DOI: 10.18653/v1/P18-1055>.

Torres, Charles & Futrell, Richard 2023. Simpler neural networks prefer subregular languages. In *Findings of the association for computational linguistics: EMNLP 2023*. 1651-1661.

Torres, Charles; Hanson, Kenneth; Graf, Thomas & Mayer, Connor 2023. Modeling island effects with probabilistic tier-based strictly local grammars over trees. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2023*. 155-164. <doi.org/10.7275/nz4q-6b09>.

Tran, Tu-Anh & Miyao, Yusuke 2022. Development of a multilingual CCG treebank via Universal Dependencies conversion. In Calzolari, Nicoletta; Béchet, Frédéric; Blache, Philippe; Choukri, Khalid; Cieri, Christopher; Declerck, Thierry; Goggi, Sara; Isahara, Hitoshi; Maegaard, Bente; Mariani, Joseph *et al. (eds.), Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association. 5220-5233.

Trinh, Trieu H. & Le, Quoc V. 2019. A simple method for commonsense reasoning. <arXiv:1806.02847>.

Trotta, Daniela; Guarasci, Raffaele; Leonardelli, Elisa & Tonelli, Sara 2021. Monolingual and Cross-Lingual Acceptability Judgments with the Italian CoLA corpus. *Findings of the Association for Computational Linguistics: EMNLP 2021*. Punta Cana, Dominican Republic: Association for Computational Linguistics. 2929-2940. <doi.org/10.18653/

v1/2021.findings-emnlp.250>.

Turing, Alan M. 1937. Computability and λ-definability. *Journal of Symbolic Logic* 2. 153-163. <doi.org/10.2307/2268280>.

Turing, Alan 1950. Computing machinery and intelligence. *Mind* 59. 433-460. <DOI: 10.1093/mind/lix.236.433>.

van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Oxford University Press. 97-157.

van Riemsdijk, Henk & Williams, Edwin 1986. *Introduction to the Theory of Grammar*. Cambridge, MA: MIT Press.

van Rooij, Iris; Guest, Olivia; Adolfi, Federico; de Haan, Ronald; Kolokova, Antonina & Rich, Patricia 2024. Reclaiming AI as a theoretical tool for cognitive science. *Computational Brain and Behaviour*.

Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N.; Kaiser, Lukasz & Polosukhin, Illia 2017. Attention Is All You Need. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S. & Garnett, R. (eds.), *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. Long Beach, CA: Curran Associates, Inc. 5998-6008. <arxiv.org/abs/1706.03762>.

Vermeerbergen, Myriam; Leeson, Lorraine & Crasborn, Onno Alex (eds.) 2007. *Simultaneity in signed languages: Form and function*. Amsterdam: John Benjamins.

Voldoire, A.; Sanchez-Gomez, E.; Salas y Mélia, D.; Decharme, B.; Cassou, C.; Sénési, S.; Valcke, S.; Beau, I.; Alias, A.; Chevallier, M. *et al.* 2013. The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dynamics* 40,9. 2091-2121. <DOI: 10.1007/s00382-011-1259-y>.

von Humboldt, Wilhelm 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwickelung des Menschengeschlechts*. Berlin, Prussia: Druckerei der Königlichen Akademie der Wissenschaften.

Wadler, Philip 1990. Deforestation: Transforming programs to eliminate trees. *Theoretical Computer Science* 73. 231-248. <doi.org/10.1016/0304-3975(90)90147-A>.

Warstadt, Alex & Bowman, Samuel R. 2022. What artificial neural networks can tell us about human language acquisition. In Lappin, Shalom & Bernardy, Jean-Phillipe (eds.), *Algebraic Structures in Natural Language*. Boca Raton: CRC Press, Taylor & Francis. 17-60.

Warstadt, Alex; Mueller, Aaron; Choshen, Leshem; Wilcox, Ethan; Zhuang, Chengxu; Ciro, Juan; Mosquera, Rafael; Paranjabe, B.; Williams, A.; Linzen, T. & Cotterell, R. 2023. Findings of the BabyLM Challenge: Sample-Efficient Pretraining on Developmentally Plausible Corpora. *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*. Singapore: Association for Computational Linguistics. 1-6. <doi.org/10.18653/v1/2023.conll-babylm.1>.

Warstadt, Alex; Parrish, Alicia; Liu, Haokun; Mohananey, Anhad; Peng, Wei; Wang, Sheng-Fu & Bowman, Samuel R. 2020. BLiMP: The

Benchmark of Linguistic Minimal Pairs for English. *Transactions of the Association for Computational Linguistics* 8. 377-392. <doi.org/10.1162/tacl_a_00321>.

Warstadt, Alex; Singh, Amanpreet & Bowman, Samuel R. 2018. Neural Network Acceptability Judgments. <arXiv:1805.12471>.

Warstadt, Alex; Singh, Amanpreet & Bowman, Samuel R. 2019. Neural network acceptability judgments. *Transactions of the Association for Computational Linguistics* 7. 625-641. <aclanthology.org/Q19-1040>.

Warstadt, Alex; Zhang, Yian; Li, Xiaocheng; Liu, Haokun & Bowman, Samuel R. 2020. Learning Which Features Matter: RoBERTa Acquires a Preference for Linguistic Generalizations (Eventually). *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 217-235. <doi.org/10.18653/v1/2020.emnlp-main.16>. <aclanthology.org/2020.emnlp-main.16>.

Waskan, Jonathan; Harmon, Ian; Horne, Zachary; Spino, Joseph & Clevenger, John 2014. Explanatory anti-psychologism overturned by lay and scientific case classifications. *Synthese* 191,5. 1013-1035. <DOI: 10.1007/s11229-013-0304-2>.

Wei, Jason; Bosma, Maarten; Zhao, Vincent; Guu, Kelvin; Yu, Adams Wei; Lester, Brian; Du, Nan; Dai, Andrew M. & Le, Quoc V. 2022a. Finetuned language models are zero-shot learners. In *ICLR 2022 Conference Track*. Online: OpenReview

Wei, Jason; Wang, Xuezhi; Schuurmans, Dale; Bosma, Maarten; Ichter, Brian; Xia, Fei; Chi, Ed; Le, Quoc V. & Zhou, Denny 2022b. Chain-of-thought prompting elicits reasoning in large language models. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K. & Oh, A. (eds.), *Advances in Neural Information Processing Systems 35 (NeurIPS 2022) Main Conference Track*. New Orleans, LA: Curran Associates, Inc. 24824-24837.

Wexler, Kenneth & Culicover, Peter W. 1980. *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.

Wickelgren, Wayne A. 1969. Context-Sensitive Coding in Speech Recognition, Articulation and Developments. In *Information Processing in The Nervous System: Proceedings of a Symposium Held at the State University of New York at Buffalo 21st-24th October, 1968*. Springer. 85-96.

Wilcox, Ethan; Futrell, Richard & Levy, Roger 2024. Using Computational Models to Test Syntactic Learnability. *Linguistic Inquiry*. 55,4. 805-848. <doi.org/10.1162/ling_a_00491>.

Wilcox, Ethan; Gauthier, Jon; Hu, Jennifer; Qian, Peng & Levy, Roger 2020. On the predictive power of neural language models for human real-time comprehension behavior. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. Online: eScholarship.

Wilcox, Ethan; Levy, Roger; Morita, Takashi & Futrell, Richard 2018. What do RNN Language Models Learn about Filler-Gap Dependencies? In

Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP. Brussels: ACL. 211-221. <arxiv. org/abs/1809.00042>.

Wilkenfeld, Daniel A. 2014. Functional explaining: A new approach to the philosophy of explanation. *Synthese* 191,14. 3367-3391. <DOI: 10.1007/s11229-014-0452-z>.

Wilkenfeld, Daniel A. & Lombrozo, Tania 2020. Explanation classification depends on understanding: Extending the epistemic side-effect effect. *Synthese* 197,6. 2565-2592.

Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie; Blomberg, Niklas *et al.* 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 3,1. 160018. <doi.org/10.1038/ sdata.2016.18>.

Williams, Edwin S. 1977. Discourse and Logical Form. *Linguistic Inquiry* 8,1. 101-139.

Wiltschko, Martina 2008. The syntax of non-inflectional plural marking. *Natural Language and Linguistic Theory* 26,3. 639-694.

Wiltschko, Martina 2014. *The universal structure of categories. Towards a formal typology*. Cambridge: Cambridge University Press.

Wiltschko, Martina 2018. Discovering syntactic variation. In Hornstein, N.; Lasnik, H.; Patel-Grosz, P. & Yang, Ch. (eds.), *Syntactic Structures after 60 Years. The Impact of the Chomskyan Revolution in Linguistics*. Studies in Generative Grammar [SGG] 129. 427-460.

Wiltschko, Martina 2021a. *The grammar of interactional language*. Cambridge: Cambridge University Press.

Wiltschko, Martina 2021b. Universal underpinnings of language-specific categories. A useful heuristic for discovering and comparing categories of grammar and beyond. In Alfieri, Luca; Ramat, Paolo & Arcodia, Giorgio Francesco (eds.), *Linguistic Categories, Language Description and Linguistic Typology*. 59-99.

Wiltschko, Martina 2022. Language is for thought and communication. *Glossa: A Journal of General Linguistics* 7,1. <doi.org/10.16995/glossa.5786>.

Wiltschko, Martina & Heim, Johannes 2016. The syntax of confirmationals. A neo-performative analysis. In Kaltenböck, Gunther; Keizer, Evelien & Lohmann, Arne (eds.), *Outside the Clause. Form and function of extraclausal constituent*. John Benjamins. 303-340.

Wiltschko, Martina & Heim, Johannes 2020. Grounding Beliefs: Structured Variation in Canadian English Discourse Particles. In Achiri-Taboh, B. (ed.), *Exoticism in English tag questions: Strengthening arguments and caressing the social wheel*. Cambridge: Cambridge Scholars Publishing.

Yang, Andy; Chiang, David & Angluin, Dana 2024. Masked hard-attention transformers recognize exactly the star-free languages. In Globerson, A.; Mackey, L.; Belgrave, D.; Fan, A.; Paquet, U.; Tomczak, J. &

Zhang, C. (eds.), *Advances in Neural Information Processing Systems 37 (NeurIPS 2024) Main Conference Track*. Vancouver, BC, Canada: Curran Associates, Inc. 10202-10235.

Yang, Charles D. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: MIT Press.

Yang, Yuan & Piantadosi, Steven T. 2022. One model for the learning of language. *Proceedings of the National Academy of Sciences* 119,5. e2021865119. <doi.org/10.1073/pnas.2021865119>.

Yi, Sanghyun; Goel, Rahul; Khatri, Chandra; Cervone, Alessandra; Chung, Tagyoung; Hedayatnia, Behnam; Venkatesh, Anu; Gabriel, Raefer & Hakkani-Tur, Dilek 2019. Towards coherent and engaging spoken dialog response generation using automatic conversation evaluators. In van Deemter, Kees; Lin, Chenghua & Takamura, Hiroya (eds.), *Proceedings of the 12th International Conference on Natural Language Generation*. Tokyo, Japan: Association for Computational Linguistics. 65-75. <DOI: 10.18653/v1/W19-8608>.

Zhang, Chiyuan; Bengio, Samy; Hardt, Mortiz; Recht, Benjamin & Vinyals, Oriol 2021. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM* 64. 107-115. <doi.org/10.1145/3446776>.

Zhang, Yian; Warstadt, Alex; Li, Haau-Sing & Bowman, Samuel R. 2021. When Do You Need Billions of Words of Pretraining Data? In Zong, Chengqing; Xia, Fei; Li, Wenjie & Navigli, Roberto (eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Vol. 1*. Online: Association for Computational Linguistics. 1112-1125. <DOI: 10.18653/v1/2021.acl-long.90>. <arxiv.org/abs/2011.04946> (2020).

Zhao, M.; Golaz, J. C.; Held, I. M.; Guo, H.; Balaji, V.; Benson, R.; Chen, J. H.; Chen, X.; Donner, L. J.; Dunne, J. P. *et al.* 2018a. The GFDL global atmosphere and land model AM4.0/LM4.0: 1. Simulation characteristics with prescribed SSTs. *Journal of Advances in Modeling Earth Systems* 10,3. 691-734. <DOI: 10.1002/2017MS001208>.

Zhao, M.; Golaz, J. C.; Held, I. M.; Guo, H.; Balaji, V.; Benson, R.; Chen, J. H.; Chen, X.; Donner, L. J.; Dunne, J. P. *et al.* 2018b. The GFDL global atmosphere and land model AM4.0/LM4.0: 2. Model description, sensitivity studies, and tuning strategies. *Journal of Advances in Modeling Earth Systems 10,3. 735-769. <DOI: 10.1002/2017MS001209>*.

Zymla, Mark-Matthias 2024. Ambiguity management in computational Glue semantics. In Butt, Miriam; Findlay, Jamie & Toivonen, Ida (eds.), *Proceedings of the LFG'24 Conference*. Konstanz: PubliKon. 285-310. <lfg-proceedings.org/lfg/index.php/main/article/view/59>.