

Is it the end of Generative linguistics as we know it?

Gillian Catriona Ramchand

Faculty of Linguistics, Phonetics and Philology, University of Oxford, UK

< gillian.ramchand@ling-phil.ox.ac.uk >

Department of Language and Culture, UiT the Arctic University of Norway, Norway

< gillian.ramchand@uit.no >

Are there certain core beliefs of generative grammar that are fatally undermined by the recent successes of Large Language Models and the unsupervised learning that trains them? Do Large Language Models then constitute a rival (and superior) ‘theory’ that can and should take over now from (all) previous theories in pushing the science forward? In this short article, I argue that the answer to both these questions is ‘no’. On the positive side, I make an urgent case for maintaining theory at the centre of the new era of linguistic science, and for generative grammar to expand its energies into theorizing the link between competence and various aspects of performance in order to shore up its claims to explanatory adequacy.

KEYWORDS: Large Language Models, linguistics, poverty of the stimulus.

1. Introduction

The recent rise of Large Language Models (LLMs) as foundation models for a wide variety of artificial intelligence applications has led many to predict the end of the world ‘as we know it’. Many of our old certainties are set to change in the wake of these new advances: the value and marketability of different kinds of labour and expertise (with its knock-on effects for education); improved pattern discovery tools for scientific and medical applications; threats to the reliability of our information ecosystems, with the global instability that can give rise to; the exacerbation of global inequalities in an era where a small number of stakeholders control access to the new technology and profit from it, while a large number will suffer from its energy requirements. To this dystopian list, we apparently need to add the demise of theoretical linguistics (or ‘generative’ linguistics as Chesi puts it) ‘as we know it’. As a generative linguist myself, I must confess I am much less worried for my field than I am about the planet more generally, but I welcome the opportunity to revisit the deep and fascinating questions that inform the scientific study of language systems and their natural embedding in human cognition in the light of these recent developments in natural

language processing models. The title of Chesi's (*this issue*) contribution that I am responding to alludes to the 'the end of generative grammar as we know it', and so in my own response I feel obliged to address the specific elements of that research program that have been claimed to have been refuted. For this reason, I will also respond to the particular claims in Piantadosi (2024) as part of my response to Chesi's own assessment. In the second half of this article, I will return to address Chesi's own main focus as I see it, which is the diagnosis of what generative grammar needs to do to forestall its alleged 'irrelevance' in the modern era.

So, what characterizes generative grammar as a field of inquiry, and what would it mean for it to be over? Chesi (*this issue*) does not define what he means by 'generative grammar' directly, calling it only 'a prototypical theoretical perspective on language'. The general thrust of his critique is that generative grammar needs to get its act together in terms of developing more precise and rigorous formulations of its model, which in turn would allow it to participate in the joint enterprise of evaluating its descriptive adequacy against a standardized dataset, comparing it with the LLMs that Piantadosi (2024) claims also count as 'genuine theories of language'. Piantadosi's own position seems to be more foundationally critical, asserting directly that LLMs 'refute' certain core principles of generative grammar (or rather, Chomsky's general approach to grammar). While Chesi is more circumspect here, the general thrust of his critique is that generative grammar needs to clean up its act as a theory, or risk becoming obsolete.

There are two issues that I want to address separately in this response article: (i) Are there certain core beliefs of generative grammar that are fatally undermined by the recent successes of LLMs and the unsupervised learning that trains them? (ii) Do LLMs then constitute a rival (and superior) 'theory' that can and should take over now from (all) previous theories in pushing the science forward?

2. What is 'Chomsky's approach', and how will I know when it has been refuted?

I first remind the reader that 'Chomsky's approach' to grammar and 'Generative Linguistics' should not be considered the same thing. Chesi criticizes minimalism in particular for not being sufficiently precisely and consistently formalized to generate the kinds of predictions required for a theory (with the exception of Stabler 1997, Collins & Stabler 2016). But other branches of generative grammar such as LFG and HPSG

have a long history of being tightly controlled and formalized and have always gone hand in hand with work in computational linguistics. So criticizing minimalism should not be considered a criticism of generative linguistics as a whole. Maybe Chesi, like Piantadosi (2024), actually just means ‘Chomskian approaches’ here. In the next section, I take a look at a number of the ideas that might be considered to be criterial of ‘Chomskian’ approaches in this sense, to assess whether or not they have been refuted by the success of LLMs.

2.1. Innateness

What then defines the generative linguistics school of scientific endeavour? <thoughtco.com> says it is the idea that all humans are born with an innate capacity for language. Nativism therefore seems to be a good candidate for a principle that characterizes generative grammar in the minds of those who consider it to have been debunked. In Piantadosi (2024), the list of ‘refutations’ contributed by the success of LLMs includes the idea ‘Hierarchical structure need not be innate’.

In fact, it is widely acknowledged that the concept of innateness is far more complex and subtle than this simple characterization suggests, and in ways that undercut the usefulness of trying to use it as a deciding ideological commitment to separate scientific schools of thought. Firstly, acknowledging the reality of an innate contribution is compatible with many different frameworks and positions. In *Rethinking Innateness*, Bates *et al.* (1996) point out that no scientist working on questions involving nature/nurture balance thinks that there is a simple relationship between genes and phenotype, and so it is increasingly difficult to draw the line between the kind of information contributed by the genome and that contributed by the developmental environment. Secondly, the kind of information that could be considered part of genetic endowment can also be quite abstract – in addition to being genetically hardwired to know certain things or perform certain behaviours, we could also possess genetic information that gives rise to the timing of certain developmental stages affecting rate and strategies for learning itself, the gross anatomy of certain brain regions and how they interact developmentally etc. Given what we currently know about brain plasticity and learning in humans, it is highly implausible that the kind of ‘representational innateness’ for language-specific principles that people may first think of when hearing the terms ‘Universal Grammar’ (UG) or ‘the Language Acquisition Device’ could be the case. Moreover, Chomsky himself, possibly in response to science’s improving understanding of brain, genes and development, moved away from a crude kind of representational

innateness in the old fashioned sense, to one that is much more minimal (Hauser *et al.* 2002). The more modern claim is that human minds have an innate capacity for recursion and that this allows hierarchical syntactic structures to emerge as a solution to the problem of language acquisition.

This is a far cry from the idea of UG as a templatic blueprint of abstract structures and rules given innately, in advance of any linguistic experience. I do not personally know of (many) working generative syntacticians who believe in that early simplistic version of UG. In short, the idea that hierarchical representations do not have to be innately given as ‘representations’ contradicts nothing in minimalist theorizing, and nor does it contradict the idea that there is some aspect of our genetic endowment that makes language possible in humans, an idea conceded already by the connectionist tradition in Bates *et al.* (1996).

2.2. A note on poverty of the stimulus

To what extent does language data in principle furnish enough information for the learner to infer the system that generates it? The lesson from LLMs and the transformer architecture seems to be that if a large enough data set is provided to the learner, then it appears that the distributional properties of language tokens provide enough implicit information to allow the model to infer the correct hidden layers of structure to perform correctly on a text generating task. Of course, a huge amount of data is required to achieve this feat, so the second relevant question here is whether this fact about LLMs at all defuses the poverty of the stimulus argument as it applies to actual human children. Children learn grammar from far less input than this, and also in a rather different kind of context, and motivated by a different kind of holistic ‘task’. We know that humans do engage in predictive processing, but the driving force behind language learning for the child is not plausibly a game to guess what word is coming next, or indeed to figure out which sentences are grammatical or not. The child’s ‘task’ is more likely to be the drive to understand the content and emotional value of what is being said to them, to predict the macro behaviour of other humans, and to acquire the toolbox to communicate in return. And for that children have a additional information coming from the context and shared experience of the world with their interlocutor. They also bring some things to the task including independent domain general facts about their own shared cognition with the speakers of the language they are learning, as well as possibly certain analytic and learning biases built in to their developing cognitive systems. So, is there enough information

coming from the incoming signal to acquire the system the child is faced with, or do we need to take account of cognitive and learning biases? To my mind it is a fascinating and still open question understanding how a child achieves this feat. But it is equally clear to me that the current evidence from the behaviour of LLMs does not advance our understanding of the question either way. Concerning the training algorithms themselves, it is widely acknowledged that the back propagation training algorithms that LLMs use are very different from human learning in rather deep ways (Hinton 2022, Evanson *et al.* 2023), and so cannot be considered good models for human learning, even independently of the huge disparities in data size required to achieve the same results.

2.3. The relation between syntax and semantics

Another place where Piantadosi (2024) claims that LLMs refute a basic principle of Chomsky-inspired grammar lies in the relationship between syntax and semantics. The claim is that in the training data (and in the model that is built in response to it), syntactic and semantic information are ‘integrated’ and cannot meaningfully be separated. This supposedly contradicts Chomsky’s view on the autonomy of syntax. I must confess that I am not sure I understand the point that is being made here. Nobody would deny that the language that the child is faced with is a combination of syntactic and semantic properties; the data that an LLM is trained on is the same, because it is just language, the same as for the child. Nevertheless, the model learns syntactic generalizations from this input, where nobody has given it the meta-information concerning how to separate the syntax from the semantics in principle. But this once again is the same for the child. The ‘integration’ of syntax and semantics which forms the ‘refutation’ here must lie in the implicit analysis or model that the LLM ends up embodying. While it is well known that inspection of the detailed representations of these models is quite difficult in principle (because of the ‘black box’ nature of the system), Piantadosi admits (and even elsewhere makes a virtue of the fact that) these models do seem to end up inferring syntactic generalizations, representing sentences hierarchically, implicitly characterizing word class membership, and tracking long distance dependencies. In other words, LLMs do end up representing syntactic information as a result of their training (Manning *et al.* 2020, Futrell *et al.* 2019, Linzen & Baroni 2021), even though they were not told to look for syntactic generalizations in advance, and even though the information is hopelessly entangled with semantics in the form of the distributional properties of linguistic tokens. In particular, they also seem to organize information in a way

that is similar to tree structures (Manning *et al.* 2020), and the extent to which this is true even predicts the model's performance on generalization (Murty *et al.* 2022). The models do well, in addition, on function words (Kim *et al.* 2019, and filler-gap dependencies (Wilcox *et al.* 2018).

I do in fact think there is something interesting and remarkable about what these models achieve. Recall that the task that the LLMs are set in training is to 'predict' the next token, and they are constantly given positive and negative feedback on that task and undergo mind bogglingly gargantuan amounts of training on it. They are allowed to use any information they can to help them succeed at the task, and they come to the conclusion that they can predict better if they start to classify words into classes, and build a syntax around them, in addition to the more fine grained lexical distributional statistics. So LLMs figure out for themselves in some sense that Chomsky is right – general statistics between words is not enough, they must also infer and build in hidden syntactic structure to do a good job on this prediction task!

Piantadosi thinks that it is obvious that syntax and semantics are not separated in the model's analysis, but no-one has convincingly shown that they can simply and reliably inspect what is in these models' analysis (which is why e.g. BERTology is its own distinct industry and object of inquiry – Rogers *et al.* 2021). In seeming contradiction with the above point, when discussing the problem of the (inhumanly) huge amounts of training data required to achieve appropriate linguistic behaviours, Piantadosi informs us that the syntactic part of LLM competence is reached with much less data, with semantics and real world knowledge being the data-guzzling culprits, suggesting that "syntactic knowledge requires a small number of bits of information, especially when compared to semantics (Mollica & Piantadosi 2019)". It sounds to me as though the different components of the model's knowledge are being (at least implicitly) separated here by Piantadosi at the acquisitional and implementational level.

In this discussion, Piantadosi also makes some non-standard assumptions about what semantics is. For him, 'semantics' is being proxied by the distributional properties of individual words and tokens of the language. But if there is one thing we know about LLMs it is that they have no mapping between the tokens of language and anything at all that exists 'outside language'. They, in other words, have no concept of denotation, reference or truth in the mapping to a world outside of the language-internal system of dependencies and relations. This mapping to an external reality is what most semanticists would define as semantics, and this is in fact the conception of semantics that Chomsky was most keen to excise from syntax in his autonomy of syntax thesis (Chomsky

1995b). Now, for LLMs, it turns out that not having semantics in the referential sense (or embodied cognition, or shared attention, or communicative urges) is absolutely no impediment at all to acquiring the ability to produce grammatically acceptable and appropriate sentences! So once again, it seems to me that the lesson of learning in LLMs shows us that Semantics (with a capital S, in the semanticists' sense) is not necessary in order to acquire syntax, thus confirming Chomsky's point rather than refuting it.

3. Do LLMs constitute a theory of grammar?

Piantadosi claims that LLMs are a theory of grammar, and Chesi implicitly agrees with this position at least to the degree that he thinks that minimalist proposals should be assessed side by side with them on commonly agreed benchmarks. Ambridge & Blything (2024) claim LLMs do better than theoreticians on all the jobs that grammatical theory was supposed to do. Müller (2024) on the other hand disagrees, arguing that LLMs are not theories in the same sense at all. That paper argues that an LLM is a successful piece of engineering that matches the patterns in the corpus of textual data it is fed, but is not a theory. Specifically, it argues that LLMs are not just a different theory, or a wrong theory, they are not theories at all. Why the disagreement?

Here it is instructive to look at Piantadosi's own justification of the status of LLMs as theory in the form of his own illustrative analogy. How does the model create a theory by setting parameters? Piantadosi asks us to imagine a situation where physicists might for example be uncertain about whether gravitational force falls off as an inverse function of distance r , or of the square of r . We could imagine them constructing a super equation for gravitational force F which has a constant α whose value between zero and 1 represents the effect of the two different characterizations of the situation as in (1). If α is zero then the equation reduces to a function where $1/r^2$ is the correct determinant, whereas when α is 1, the equation reduces to one where only $1/r$ is the determinant.

$$(1) \quad F(r, \alpha) = \alpha \cdot 1/r + (1 - \alpha) \cdot 1/r^2$$

So now the job of the model is to inspect the data and select the value of the parameter α that maximizes the likelihood of getting the correct (i.e. matching with reality) answer. So in this case, inferring a parameter in this equation is tantamount to evaluating distinct theories

against the data and coming up with the preferred one. So is this what LLMs are doing as well when they are setting their billions of parameters in response to the data? Almost. Piantadosi admits that in this case, we do not give the model a specific super equation in advance like (1), which embodies distinct theoretical proposals. Instead, there are some ‘natural bases’ or starting points, for which you can set parameters that will allow you to approximate essentially ANY COMPUTATIONAL THEORY. As Piantadosi puts it “Parameter fitting in these models is effectively searching over a huge space of possible theories to see which one works best, in a well-defined, quantitative sense.”

And here is the crux of the matter. Because of these universal natural bases, you do not need to have a theory, or even a hunch, or a specific question to ask the oracle when you set one of these neural nets loose on the data. It is a well known proof about these kinds of neural nets that they are capable of approximating to any degree of precision, any function that is in principle computable no matter what it is (Cybenko 1989). So if the training algorithm is sound and the data contains the right information, then the neural net will end up mimicking the input data to an arbitrary degree of precision. And here is the other thing. You cannot at that point go back into the model and reconstitute which particular equation or hypothesis was being piecemeal approximated by its elaborate parameter settings. It would be as if your machine was capable of correctly predicting the value for gravitational force when fed with distance information, but you would have no way of figuring out post hoc what the equation was! I quote from Piantadosi again lest I be accused of unwarranted negativity “In fact, we don’t deeply understand how the representations these models create work (see Rogers *et al.* 2021). It is a nontrivial scientific program to discover how their internal states relate to each other and to successful prediction.”

Collins (2024) is the only response on the topic that I have seen that makes this point most clearly and trenchantly. He argues that LLMs are not theories because they can represent any complex relationship, and they represent them all in essentially the same kind of way. So it is no good inspecting the representations arrived at by these LLMs for theoretical insight into the workings of language (even if that were easy to do), because the LLM representations work by piecemeal approximation and flatten out any computational specificities inherent in the thing being approximated. If we want the physicists’ equivalent of the equation for gravitational force, we are going to have to come up with it ourselves via human scientific theorizing and explicit hypothesis testing. It is possible that what Piantadosi and others are claiming here is that LLMs prove that humans themselves could also be just ‘universal approx-

imation machines', in which case the human child is just a supreme pattern matcher who will learn whatever language they are exposed to. This seems unlikely for a number of reasons. Firstly, it underplays the fact that children (and humans in general) are quite a bit worse at the sorts of computations that LLMs seem to excel at especially when it comes to seeing patterns in extremely large amounts of complex data. Secondly, in concentrating on the text prediction task, it misses the fact that human minds 'created' language systems (in all locations where humans can be found) in the first place, with a world-language relationship in mind. We would get no explanation of this phenomenon simply by asserting that the human brain is a massive approximating machine capable of imitating patterns it is exposed to in the form of disembodied language. In Müller (2024), another of the points raised is that studying LLMs trained on particular languages is unlikely to give us any purchase on crosslinguistic similarities and variation that exist. This is because comparative information of this kind is simply not extractable in a way that we as scientists can make sense of or interpret at a higher level. It is unclear therefore, what we gain from our marvellous engineering successes, other than a monetizable object for capitalism to chomp on. In terms of the kind of tangible, symbolically expressed theory that humans need in order to extend and generalize understanding into other domains, we do not seem to be able to extract something of that level from the LLMs we have created. The existence of SyntaxGym¹ provides a useful set of cross-model benchmarks. But it does not mean that the things being compared are all 'theories' in the same sense.

There are of course differences among different LLMs and how well they perform, but in general it appears that the differences between different neural networks come from the interplay between the nature of the training algorithms and the data they are being fed (Collins 2024). We could make a study of those systems and algorithms, but it since we already know there are deep differences between us and the neural nets in both the learning strategies and data exposure, it does not buy us anything to study the LLM instead of the human directly.

4. Where next for linguistic theory?

In this response so far, I have mostly concentrated on the claim that LLMs and their successes have seriously undermined specifically Chomskian approaches to grammar, that they are a refutation of and alternative to those theories. I have argued firstly that nothing criterial

to the enterprise has been refuted, and secondly that what has been produced is not qualitatively the same thing as a ‘theory of grammar’.

To my mind, what characterizes generative grammar, and the Chomskian approach in particular, was the radical reconception of the object of inquiry when it came to language science. Chomsky reframed the scientific question away from the cataloguing and analysis of linguistic behaviour, towards the psychological questions concerning the nature of linguistic knowledge in human minds that allows them to produce these behaviours. We want to understand the system that generates linguistic behaviour, not simply analyze the produced patterns that are measurable and recordable as outputs of that system. In that sense, the school of thought that considers the LLM model itself to be the ‘theory’, the desired endpoint of scientific endeavour, is basically a return to the crudest kind of behaviourism where the model is evaluated by how well it succeeds in mimicking the externally observed data, and not in how it helps us to understand human minds.

So indeed, I do think we are seeing a shift that threatens to strike at the heart of the Chomskian enterprise, but not in the way Piantadosi imagines. It is a shift that reifies the patterns of external data as an object of inquiry in its own right, where the goal is to produce generative systems that will demonstrably reproduce the fine detailed patterning of that data. Engineering success is defined by matching output behaviour, not by achieving an understanding of how this happens within the engineered device (let alone how it happens within human minds which are quite different). Dataism is the real existential threat, and one which should be resisted, if we have the goal of understanding the role of language within human cognition. Dataism left unchecked can give rise to a kind of theoretical nihilism, which will lead to dead ends as soon as solutions need to be extended, or generalized over, or metatheorized.

If we are threatened with the demise of generative grammar, it is at this level, as part of a general distrust of symbolic theories and an enthusiasm for bottom up, theory-free, engineering solutions. I think that this is not progress. I think that humans and human scientists have made great strides in understanding the world by using the cognitive ‘gadgets’ of language and symbolic theorizing. My hunch is that representing information symbolically is the gadget that allows us to generalize explicitly and metacognize in increasingly sophisticated ways. In this new era of big data and artificial intelligence aids to pattern-discovery, we need to maintain a pivotal role for human scientific expertise and theorizing. Generative grammar (or theoretical linguistics more generally) is a natural constituency for where that expertise can continue to

be nurtured and from where it can contribute to multidisciplinary questions.

Returning to Chesi (*this issue*), on the question of whether theoreticians, or generative grammar (or specifically the Chomsky inspired linguistic tradition) need to change in response to recent advances in this technology, my answer is yes, but with a perspective slightly different from that articulated in Chesi (*this issue*). I agree with Chesi that there are problems both in formalization, and in the nature of the data that can be used as falsifying evidence. While generative grammar can claim to be a theory (and we need theories!), it is fair to note that it has not been doing a particularly good job of showing that it is in fact a 'good theory'. This is because, in my opinion, it has not really made good on its own goals of 'explanatory adequacy', and ironically for a theory that started off by placing the scientific object of study within the realm of individual cognition, it has not really seriously engaged with the results or observations from cognitive science. It is true that generative grammar has always situated itself squarely at Marr's computational level (Marr 1982), and has used this to justify the lack of theorizing to the next step algorithmic and implementational levels. But even Marr, in his work on vision, saw filling in those other levels as part of the scientific enterprise he was engaged in, and he theorized about those too. The problem is that if we as scientists do not form linking theories between the computational level and how these tasks are achieved in real brains, then we are in possession of theories that make no predictions whatsoever about data gathered by psycholinguists or neurolinguists. This in turn means that we cannot claim that such theories have higher levels of explanatory adequacy than others with the same descriptive coverage. It seems to me that many of the early claims to 'explanatory adequacy' in the rhetoric of the Government and Binding era rested on the potential to account for the acquisition of language. In fact, it turned out to be much more difficult than anticipated to formalize a concrete implementation of a learning algorithm under a principles and parameters conception that achieves the right results deterministically given the input data and the kinds of cues available to the child (Gibson & Wexler 1994, but see also Fodor 1998). Moreover, acquisition is not the only explanatory adequacy benchmark. Generative grammar (and most especially the minimalist program), needs to begin to seriously theorize about the relationship between its computational theories and how they are embedded within more domain general theories of mind/brain. In other words, it needs to engage with a variety of different performance tasks directly and produce theories of them. The aim should be to get to the stage where our best theories do make predictions about the cognitive pro-

cessing behaviours we can measure when we are deploying our ‘knowledge of language’ whether in production or comprehension or acquisition. Only in this way can we bridge the commensurability gap between linguistic theory and the cognitive sciences, and only in this way can we assess the explanatory adequacy of these theories. So far, generative grammar has failed to do this. Not because it has tried and failed, but because it seems to have exerted a lot of rhetorical effort in arguing that it should not be required to try.

It is here that I wholeheartedly agree with Chesi that generative grammar needs to up its game if it is to take its place at the table where scientists take the next step towards understanding how language functions within human cognition. Chesi argues that currently the state of formalization within minimalist theorizing lags behind what is necessary if minimalist grammar is to join the conversation at the highest level. This is because formalization is not consistent and is often piecemeal. I agree with this, but to it I would add that a more precise and complete formalization of the computational theories we have is not enough, because such theories do not make predictions beyond that of grammaticality at the level of the sentence. Grammatical theory needs to take the step toward a more algorithmic understanding, and thus open itself up to potential falsification from a wider range of data types. In doing so, it needs to avail itself of all the new mathematical tools and methodologies on offer. I would argue that one of the things we are seeing in this new era of big data and LLMs is that the computational tools and methodologies at our disposal are greater and more sophisticated than ever before, and the data we are able to measure from human brains is getting more and more precise and detailed. It is no longer a pipe dream to think of making good on the promises of real explanatory adequacy. For the same reason, it is no longer defensible to stick cosily to our computational level and virtuously deny any ambitions to make predictions about ‘processing’. In making the transition, we will have to engage in new mathematical techniques and methodologies, including computational modelling. And here we come back again full circle to artificial intelligence, neural networks and computational modelling, this time not as theories, but as tools for cognitive science. This is the position laid out by van Rooij *et al.* (2024) who point out that this was originally part of the idea of computational modelling and later connectionist networks (see also Bates *et al.* 1996), not as rivals to a particular theory, or rivals to our human intelligence (AGI anyone?), but as a tool for testing the different theoretical ideas concerning computation and cognition. Language scientists need to reclaim the space for real theory, and not be afraid to use computational modelling as part of their toolbox. Collins

(2024) argues that an expansion of the methodological toolbox has already begun within formal grammar, for reasons that are independent of the sudden success of LLMs. For example, he notes a number of recent approaches that fuse formal linguistic models with information theoretic methodologies (Levy 2008, Goldsmith & Riggle 2012, Rasin *et al.* 2021). He also points out work like Allott *et al.* (2021) which discusses quantitative approaches to sentence processing and how they depend on the Chomskyan program. Collins himself advocates for generativists adopting methods from statistical physics, which can be used to tackle issues of complexity and emergence from a robustly theoretical and rationalistic point of view. When it comes to using the most advanced tools available, modelling via neural nets themselves are valuable tools in the pursuance of the Chomskian questions relating to human minds. I am in complete agreement with van Rooij *et al.* (2024) that we need to reclaim AI from the artificial general intelligence builders and put them to work in the service of answering the scientific questions we care about. Much of what is going on in the AI community is highly relevant and shows great potential for modelling and testing theoretical hypotheses. I sincerely hope that theoreticians and generative grammarians embrace these new possibilities and engage with the advances in the field of computational modelling instead of interpreting LLMs as a rival theory to be attacked. It certainly does not help if LLMs are being hyped as actual theories that are ‘refutations’ of a whole linguistic tradition and its research questions.

The bottom line is that we have a unique opportunity in this new era to create a genuinely multidisciplinary science of language which involves a whole host of methodologies, with distinct though related research questions. There will of course be the inevitable challenges that arise for generativists in moving into this more multidisciplinary space, and not all generativists need choose to do so. It is necessary to emphasize that there is still also important work to be done in pure language description and comparison that theoretical generativist linguists of all stripes continue to do, using shared or at least intertranslatable analytic vocabularies. I would not be unhappy however to be witnessing the demise of lazy repetition of ideological tropes and thought experiments associated with certain tribal memberships, which stand in the way of genuine re-examination of the issues from first principles as we learn more and more from different fields. If we are lucky, and if we have time to do it before capitalism consumes the planet, then we are on the brink of making genuine breakthroughs in cognitive neuroscience.

Understanding how language fits into cognition is one of the important aspects of cognitive neuroscience more generally, and I think that

Gillian Catriona Ramchand

theoretical linguists can and should be a central part of that scientific conversation going forward. Some of those will be generativists.

Note

¹ Chesi (*this issue*) explains: “One platform designed for performing such linguistic benchmarks is SyntaxGym (Hu *et al.* 2020): an on-line, open-source repository that includes a significant set of linguistic contrasts – 39 test suites that include a total of about 4k sentences. For each relevant contrast included, human generalizations have been gathered in various studies. Direct comparisons of these data with the predictions provided by the models under evaluation is then possible.”

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/jamaintern-med.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; Lukošiuūtė, 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; Zymła, 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.cog-sys.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 Merriënboer, 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. <csli-publications.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/iia217>.
- 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 Bilingual 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.; Guerrero, 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, Kleantes 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; Cordonì, 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 linguisticity 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](https://arxiv.org/abs/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, 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. 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, Tamar (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. *Métabolites: 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; Franzluebbbers, 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 & Žygis, 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/SilverDialoguesMarantz.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.pat-ter.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=SUTHgAACAAJ>.
- 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>.
- Sproue, 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> .
- Strivastava, 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/oxford-hb/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](https://arxiv.org/abs/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.
- Voltaire, 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 Entwicklung 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; Mohanoney, 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 extracausal 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>.