DOI: 10.26346/1120-2726-241

The dose makes the poison: Chesi's vision and subregular syntax

Thomas Graf

Department of Linguistics, Stony Brook University, Stony Brook, NY, USA
mail@thomasgraf.net

In this reply to Chesi's *Is it the end of (generative) linguistics as we know it*, I argue that the specifics of his vision for generative syntax in the 21st century remain hazy. Depending on how one interprets Chesi's methodological desiderata, they may well have a chilling effect on novel approaches instead of fostering them. As a concrete example of this dynamic, I discuss the problems with Chesi's focus on benchmarks and Minimum Description Length and how it would undermine recent efforts in subregular syntax that are in fact closely aligned with Chesi's goals. I conclude that Chesi's vision has merit, but only in moderation.

KEYWORDS: Minimalist grammars, subregular syntax, syntactic benchmarks, minimum description length.

In *Is it the end of (generative) linguistics as we know it*, Chesi argues that Piantadosi's criticism of generative syntax contains more than just a grain of truth, and he proposes several changes to keep generative syntax competitive in this brave new world of LLMs. From the perspective of computational syntax, I agree with Chesi that there is a lot to criticize about how generative syntax, in particular Minimalism,[1] currently operates as a theory and as a field: published papers frequently require a fair amount of exegesis in order to arrive at a fully worked out analysis; there are systemic issues with what counts as data and how that data is collected, reported, and preserved; the analytical space is delineated by pre-theoretical claims about computation and learnability that have little grounding in the actual research on those topics. The reader can probably add a fair number of their own pet peeves to this list. But what isn't perfect isn't necessarily in dire straits. The *status quo* always falls short in comparison to a bold vision unencumbered by reality. This is why it is important for the critic to present their envisioned alternative with sufficient detail so that their audience can assess whether this alternative is SOUND and FEASIBLE.

Chesi's paper leaves key issues to the reader's imagination in this respect. One extreme interpretation, henceforth EI, is that Chesi wants the merit of generative proposals to be determined by EI1) their quantitative performance on a common test set such as SyntaxGym and EI2)

 (received November 2024)

their Minimum description length (MDL). This interpretation is neither sound nor feasible. A much more moderate interpretation, henceforth MI, is that generative syntax should MI1) integrate corpora and databases into its empirical base, and MI2) adopt a broader methodology of theory comparison that also includes quantifiable metrics. This much more modest (detractors may say milquetoast) interpretation is not only feasible, it is already reality, e.g. through recent work that grows out of *subregular syntax* (Graf 2022a; Graf 2022b; Hanson 2025). In the following, I will first argue against EI (Sec. 1) and then sketch how subregular syntax instantiates the spirit of MI but would be severely handicapped by the adoption of EI (Sec. 2).

1. Against benchmarks and MDL

Under EI, Minimalism would incorporate non-generative methodologies in response to the LLM paradigm Piantadosi champions. Syntacticians should measure theories' observational adequacy over curated datasets, and they should formalize descriptive adequacy via Minimum Description Length (MDL). I will briefly touch on the issues with this position, many of which have already been discussed in the literature much more in-depth than is possible here. This abstract methodological discussion will form the backdrop against which the very concrete, practical problems with EI will emerge in Sec. 2.

For starters, it is far from obvious that Piantadosi is indeed arguing from a position of strength, which diminishes the motivation for EI. Kodner et al. (2023) provide a detailed criticism of Piantadosi's claims, and Lan et al. (2024) argue that the findings in Wilcox et al. (2023), which Chesi perceives as a watershed moment, paint an incomplete picture: "We examine the evidence further, looking in particular at parasitic gaps and across-the-board movement, and argue that current networks do not succeed in acquiring or even adequately approximating wh-movement from training corpora roughly the size of the linguistic input that children receive" (Lan et al. 2024, 1). Hence one should not be too hasty in adapting the panacea offered by EI. And indeed the veneration of benchmarks (Sec. 1.1) and MDL (Sec. 1.2) fails to appreciate the unique challenges of studying syntax.

1.1 The problem with benchmarks

Consider first the case of "shared benchmarks such as SyntaxGym, CoLa, or BLiMP" (p. 39). Chesi calls LLMs "observationally more adequate" (p. 40) than Minimalism because LLMs seem to "perform prop-

erly on shared benchmarks" (p. 39). But shared benchmarks cannot do the heavy lifting of observational adequacy.

Even in NLP, which is utterly dominated by the use of benchmarks, there is increasing awareness that they provide a very limited way of measuring a model's performance. Numerical scores tell us how often a model gives the correct answer, but they do not shed light on whether the correct answer was given for the correct reasons. Nor do they capture how badly the model gets things wrong when it gets them wrong. And most importantly, model X can greatly outperform model Y in the current benchmark yet underperform relative to Y when tested against a new dataset that controls for some confounds of the old one. In their discussion of the model BERT, Bender and Koller (2020: 5186) note that "BERT's unreasonably good performance on the English Argument Reasoning Comprehension Task [...] falls back to chance if the dataset is modified by adding adversarial examples that just negate one piece of the original" and that "BERT's performance on the English Multi-genre Natural Language Inference dataset [...] is predicated on its ability to leverage syntactic heuristics [...]. In a dataset carefully designed to frustrate such heuristics, BERT's performance falls to significantly below chance." Bender and Koller (2020: 5186) warn that a model's performance on a given benchmark may just be "a mirage built on leveraging artifacts in the training data".

One might object that syntactic theories would be less likely to employ such heuristics, but in order to capture observed behavior, they effectively have to. This is because in the realm of corpora and experiments, there is no such thing as grammaticality judgments, only acceptability judgments. But acceptability invariably involves factors that go beyond syntax, such as lexical frequency, semantic plausibility, and processing difficulty. A syntactic theory that does not account for these factors cannot hope to replicate attested acceptability judgments, while a theory that does take them into account has access to non-syntactic heuristics that could allow it to perform well in benchmarks despite getting the syntax wrong.

But perhaps this merely shows that we need to keep building better and better benchmarks of increasing sophistication, controlling for more and more confounds? This retort presupposes that there is a finite number of confounds, that we know them all, and that there is a way of addressing each confound without creating new ones. Odds are that at least one of those three isn't true. And none of this even considers the time and resources generative syntax would have to pour into this ill-motivated enterprise of corpora creation. Nor does it price in the cost of maintaining benchmarks, the potential long-term effects of transcrip-

tion errors on theory building, and how the reliance on benchmarks marginalizes understudied, low-resource languages (all of which are well-known issues in NLP). A heavy focus on benchmarks would also reduce theoretical diversity because only a few approaches have enough practitioners to continuously tweak their theories for better benchmark performance. Every scientific field is subject to the Matthew effect, and in the short to medium term, a lousy theory with lots of manpower and resources is likely to outperform a good theory that only a few researchers are working on. Elevating benchmarks to the arbiters of observational adequacy would greatly exacerbate this. With no clear pay-off and many risks, the "benchmarkification" of observational adequacy looks like a methodological dead end.

1.2 The problem with MDL

The limitations of benchmarks and corpora also affect MDL, which Chesi presents as "a practical mathematical way to compare the 'descriptive adequacy' fit […] of a theory" (p. 9). This vastly oversimplifies the intricacies of MDL. As a concrete example, consider Ermolaeva (2023), who uses MDL to compare multiple Minimalist analyses. To this end, she formalizes them in terms of Minimalist grammars (Stabler 1997; Stabler 2011) and then measures two components: how many bits it takes to encode each grammar (*grammar cost*), and how many bits are needed to encode her test corpus based on the structural descriptions provided by each grammar (*corpus cost*). In an exemplary display of scholarly transparency she carefully lays out all the limitations of her approach:

> MDL can disagree with a linguistic intuition on what constitutes a simpler explanation of the data, if some aspect of the analysis is not taken into account by the encoding scheme, or cannot be expressed by the chosen formalism, or requires an overhead cost that does not pay off in the case of the chosen corpus. […] Some proposals in the linguistics literature are motivated by patterns that could only be translated into cost reduction under a sophisticated encoding scheme; […] extremely small datasets can favor overfitting grammars, if the reduction in corpus cost provided by introducing syntactic generalizations is insufficient to justify the initial investment in the grammar. This also applies to large but repetitive datasets. […] [W]ith a very large corpus of diverse sentences (which is a better representation of natural language as a whole) the MDL value is decided primarily by the corpus cost. (Ermolaeva 2023: 108-110)

MDL results aren't a universal arbiter of succinctness, they are dependent on the choice of corpus. This is problematic for all the rea-

sons mentioned in the previous section, but the problem goes even deeper. Given the limitations and confounds of small corpora, generative syntax would have to consider fairly extensive corpora, which makes grammar cost negligible. But grammar cost is what is actually of interest to syntacticians. A simplicity metric that pays little attention to the simplicity of the grammar simply misses the mark.

Of course one could carefully engineer all the MDL parameters to get a more appropriate result, but that is exactly the problem: MDL is not an easy, objective way of quantifying simplicity, it is an elaborate modeling task that is ripe with linguistically arbitrary decisions. It is replete with formal parameters that differ in subtle ways yet yield different results. Based on what criteria should linguists trust one MDL comparison but disregard another? And since there are few researchers who know both MDL and syntactic theory well enough to combine the two in an insightful manner, an excessive focus on MDL would supercharge the Matthew effect and stifle theoretical diversity in the field. The malleability of MDL disqualifies it as the be-all and end-all of linguistic simplicity and theory comparison.

At this point some readers may object that my arguments against EI have the flavor of conceptual nitpicking that erroneously allows *perfect* to be the enemy of *good*. But quite to the contrary, it is EI that allows *perfect* to be the enemy of *good* by pursuing an empiricist pipe dream of fully quantifiable theory comparison (and by extension, theory construction). In a perfect world, this approach would work and greatly accelerate linguistic progress by allowing syntacticians to skip the challenging and time-intensive task of theory comparison. But NLP is living proof that EI1 is far from perfect, and the work that has been done on combining MDL and syntactic theory casts major doubt on the feasibility of EI2. Insisting on EI ignores these warning signs to the detriment of all the good alternatives that are already around. In the next section, I present subregular syntax as one example of such an alternative, discuss how it meets many of Chesi's desiderata, and explain why EI would have a chilling effect on this enterprise.

2. Subregular syntax: An answer to Chesi's vision?

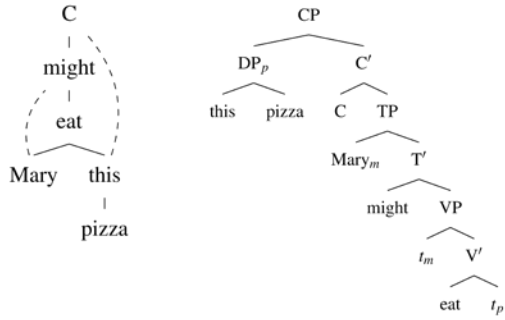
Subregular syntax is a program that combines generative syntax with notions from formal language theory (subregular complexity) that have also been fruitfully applied to phonology and morphology (Heinz 2010; Chandlee 2014; Aksënova et al. 2016; Jardine 2016; Chandlee 2017; Aksënova and Deshmukh 2018; Chandlee and Heinz 2018;

Graf and Mayer 2018; Heinz 2018; Mayer and Major 2018; Hao and Andersson 2019; Burness et al. 2021; Chandlee 2022; Aksënova et al. 2024; Burness et al. 2024). The central goal is to identify very restricted classes of dependencies and constraints that are powerful enough to capture a wide range of empirical phenomena, and to leverage these restrictions for learning, typology, and cognition. Since this isn't the place for a comprehensive discussion (see Graf 2022a,b), I will limit myself to a few brief examples of how this program opens up new analytical avenues for Minimalism (Sec. 2.1) and how it captures the spirit of Chesi's paper as expressed by MI (Sec. 2.2). After that, I will explain why this program could not thrive under EI, the extreme interpretation of Chesi's proposals (Sec. 2.3).

2.1 What is subregular syntax?

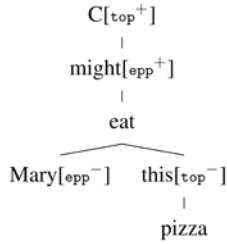
The central object of study in subregular syntax is the computations that underpin syntactic structure building, i.e. the syntactic derivation. A syntactic derivation is represented in a format similar to a dependency graph.

(1) Syntactic derivation and corresponding phrase structure tree for *this pizza, Mary might eat* (*v* omitted for brevity)



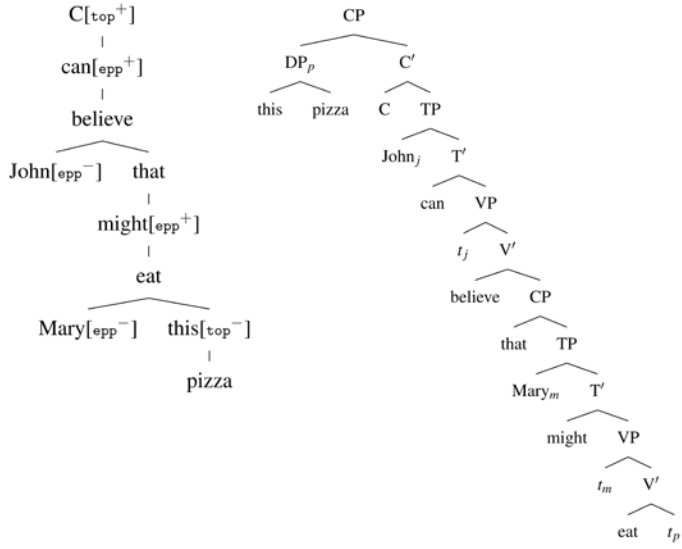
In Minimalist terms, solid branches in (1) indicate Merge steps (with the rightmost daughter of a node being merged as its complement and all other daughters being merged as specifiers). Dashed branches encode a long-distance dependency, usually movement. Since trees are mathematically easier to reason over than graphs, dashed branches are replaced with diacritics of opposite polarity such that minus (-) marks the dependent of a dashed branch and plus (+) its opposite end. The diacritics are usually chosen to reflect syntactic theory, e.g. `top` for topicalization and `epp` for subject movement.

(2) Syntactic derivation with dashed branches replaced by diacritics



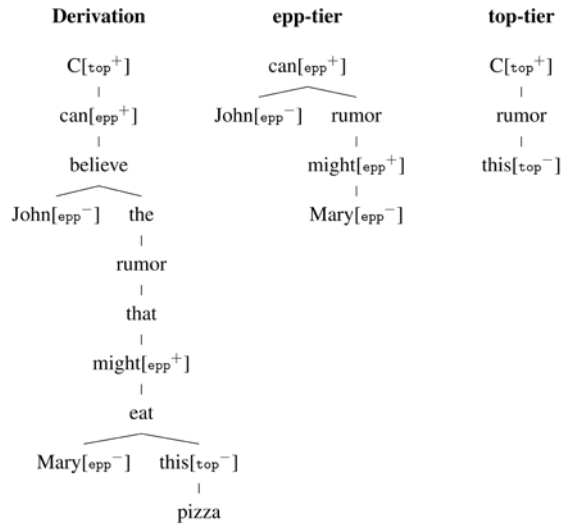
Note that Merge dependencies are STRICTLY LOCAL over these representations: for any given head, it suffices to inspect its string of daughters to determine that the head-argument configurations are well-formed (correct number of arguments, each argument with the requisite category, and so on). But long-distance dependencies can span arbitrarily large domains, as in (3).

(3) Syntactic derivation and phrase structure tree for *This pizza, John can believe that Mary might eat*



However, strict locality does obtain if specific nodes can be ignored for calculating locality. This can be visualized as SYNTACTIC TIERS. A tier contains a subset of a derivation's nodes, preserving their relative order.

(4) Syntactic derivation from (3) with epp-tier and top-tier



Over these tiers, it once again suffices to consider only mother-daughter configurations. In particular, if a node carries a plus-diacritic, say epp$^+$, then its string of epp-tier daughters has to contain exactly one instance of epp$^-$.[2] The difference between a Merge dependency and a long-distance dependency, then, is that the former puts restrictions on the string of DAUGHTERS whereas the latter puts restrictions on a string of TIER DAUGHTERS.

By expanding its focus beyond full structures to tiers, subregular syntax provides many new insights into syntax. Crucially, it does so in a manner that closely matches MI, the moderate interpretation of Chesi's paper.

2.2 How subregular syntax fits Chesi's vision

Subregular syntax avoids many of the criticisms Chesi levels against Minimalist syntax. It is rigorously formalized, grounded in computation and cognition, makes a connection to psycholinguistics (gradience, sentence processing, acquisition), allows for quantitative comparisons of analyses, and integrates corpora into its research methodology.

Subregular syntax grows out of Minimalist grammars and has a full formalization in terms of first-order logic (Graf 2023). This allows subregular syntax to draw from the rich body of computational work in that area (Stabler 2011; Graf to appear). Notably, subregular syntax is fully compatible with the Minimalist grammar work on sentence processing

that Chesi mentions (Lee and De Santo; Kobele et al. 2013; Gerth 2015; Graf et al. 2017; Hunter et al. 2019; De Santo 2020; Pasternak and Graf 2021; Liu 2023). But even without the links to Minimalist grammars, subregular syntax is a rigorous enterprise in the sense that formal rigor is indispensable for some of its key findings.

Consider, for example, the status of tiers in the grammar. While they may look like a new kind of syntactic representation, the mathematics reveals that tiers are a visual metaphor for a specific kind of cognitive architecture. This is easier to illustrate with strings. Suppose that some language has CV syllables and a vowel harmony system without neutral vowels where *a* and *o* cannot occur in the same word as *i*. Hence *baboba* and *bibibi* would be well-formed, but not *babobi*. From a cognitive perspective, this system only requires enough working memory to store the last two symbols and the ability to check whether the current symbol is a valid continuation of the previous two symbols.

(5) Memory configurations for vowel harmony when processing *babobi*

MEMORY CELL 2	MEMORY CELL 1	CURRENT SYMBOL	PERMITTED?
-	-	b	Yes
-	b	a	Yes
b	a	b	Yes
a	b	o	Yes
b	o	b	Yes
o	b	i	**No!**

Tiers modify this architecture with a cognitive least-effort principle: memory cells are updated only if relevant symbols are encountered. For our vowel harmony system, only vowels need to trigger a memory update, which is the same as saying that vowels PROJECT onto their own tier. As a welcome side-effect, this also reduces the necessary working memory to just one cell (and if the harmony featured neutral vowels, those simply would not trigger a memory update).

(6) Memory configurations for vowel harmony when processing *babobi* with a vowel tier

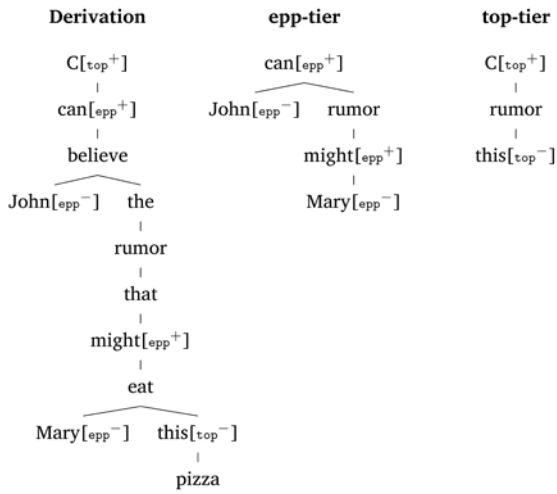
VOWEL MEMORY CELL	CURRENT SYMBOL	PERMITTED?
-	b	Yes
-	a	Yes
a	b	Yes

VOWEL MEMORY CELL	CURRENT SYMBOL	PERMITTED?
a	o	Yes
o	b	Yes
o	i	**No!**

Hence tiers represent a particular way of managing a finitely bounded amount of working memory. This insight can also be used to prove that tiers can only capture a very restricted subset of the regular (string/tree) languages. Yet this subset covers tremendous linguistic ground.

For example, island effects are unsurprising in the sense that they require no new cognitive resources beyond what is needed for long-distance dependencies. If *rumor* projects onto every movement tier as in (7), then $\texttt{top}^{+}$ on the matrix C-head won't have any instance of $\texttt{top}^{-}$ among its `top`-tier daughters and thus the derivation is ill-formed.

(7) Complex NP island *This pizza, John can believe the rumor that Mary might eat*



The same tier projection mechanism that allows for long-distance dependencies thus also allows for them to be disrupted by intervening elements, giving rise to island effects. In fact, this approach to islands can be pushed even further by using a probabilistic tier projection for *rumor* and similar heads, which makes it possible for the grammar to produce the kind of gradient acceptability judgments that have been observed in psycholinguistic experiments (Torres et al. 2023). And as explained in Graf

(2022a,c), the general idea of projecting non-movers onto movement tiers can also be leveraged for other phenomena such as Irish wh-agreement, extraction morphology, and German wh-copying. Tier projection thus ties together a diverse range of phenomena (movement, islands, extraction morphology, wh-copying, and more) in such a manner that the same structural analysis can produce categorical and gradient judgments.

Tiers also put restrictions on externalization that line up with the empirical facts. Graf (2023) shows that mapping a syntactic derivation to its bare phrase structure with copies requires no additional memory beyond what is provided by tiers, but identifying which of these copies should be pronounced is a harder problem. It can be solved with tiers only if syntax obeys a constraint similar to the Ban on Improper Movement, providing a computational motivation for the existence of such constraints. At the same time, tiers also allow us to distinguish between different implementations. A literal implementation of the Ban on Improper Movement requires n additional tiers, where n grows polynomially with the number of movement diacritics (`epp`, `top`, and so on). Probe horizons (Keine 2020), on the other hand, achieve the same effect without requiring any new tiers. Since each tier corresponds to a separate memory register, probe horizons are preferable to the Ban on Improper Movement because the latter imposes a polynomial memory load whereas the former has no additional cost at all. Tiers thus provide a theory-internal succinctness metric, and in contrast to MDL this metric is directly grounded in cognition.

There is also emerging work on how syntactic tiers can be learned from limited positive data (Swanson 2024), and how the Tolerance Principle (Yang 2016) could be applied to subregular learning (Hanson 2024). With recent findings that at least some neural architectures (LSTMs) have subregular biases (Torres and Futrell 2023), even a convergence of subregular syntax and neural networks is in the realm of possibilities.

On top of all that, subregular syntax also has principled uses for corpora. In Graf (2020), I conjectured that syntactic category systems are subregular in the sense that the syntactic category of a lexical item can be correctly inferred in a strictly local manner. Ongoing work by Kenneth Hanson and Logan Swanson suggests that this conjecture holds for *MGbank*, a fragment of the Penn treebank reanalyzed for use with Minimalist grammars (Torr 2017). Of course MGbank may be missing exactly those constructions that would disprove the conjecture, but this kind of corpus work nonetheless is an essential step in vetting subregular claims.

In sum, subregular syntax is interesting for the purposes of this reply because it instantiates a lot of what Chesi wants to see in a syntactic theory: rigorous formalization, computational grounding, connections to processing and learnability, modeling of experimental data, quantifiable notions of succinctness, and the use of corpora. Despite all that,

though, subregular syntax wouldn't flourish under EI, the extreme interpretation of Chesi's proposal.

2.3 Why Chesi's vision doesn't fit subregular syntax

The methodological issues from Sec. 1 that made EI unsound are also the reason why this extreme interpretation of Chesis' paper isn't feasible, either. The case of subregular syntax illustrates this quite clearly.

First, the use of shared benchmarks presupposes a shared vision of what the relevant phenomena are. While subregular syntax draws a lot from the Minimalist literature in its empirical analyses, it is not beholden to them. For example, subregular analyses of binding (Graf and Abner 2012; Graf and Shafiei 2019) have computational reasons to treat binding as a distributional constraint on pronominal forms rather than a constraint on co-indexation. They do not address whether *John told Peter that Bill likes him* is ill-formed when *him* refers to *Bill*, they merely observe that the string contains one or more viable antecedents for *him*. If linguistic theories had to prove their worth on standardized benchmarks, one of them being a test suite of co-indexed binding sentences, subregular syntax would be an immediate failure simply for factoring interpretation out of syntactic binding.

Even for the phenomena where subregular syntax marches in lockstep with generative orthodoxy, benchmarking it would be a laborious task. The first step of bechmarking subregular syntax is to annotate the datasets with subregular tree structures. As Chesi notes (p. 5), SyntaxGym currently contains about 4,000 sentences, and this number is bound to grow within the next few years. Annotating all these sentences would be a herculean task, even with sophisticated tooling for semi-automatic annotation. It would also be theoretically dubious because many constructions that show up in these datasets haven't been investigated from a subregular perspective yet. And barely any of that work would include phenomena that truly challenge subregular syntax, e.g. closest conjunct agreement. Why should subregular syntax spend its limited resources on making itself benchmark-able if those benchmarks are utterly misaligned with the priorities of the program?

Succinctness as formalized via MDL is also a problematic criterion for subregular syntax. Our discussion of vowel harmony in (5) and (6) already showed that the increased expressivity of tiers also comes with increased succinctness. This is a general fact of computation (Savitch 1993). If we wanted to optimize for succinctness, we should not stop at tiers, we should move all the way up to finite-state (string/tree) automata. But then we lose many of the upsides of subregular syntax: learnabil-

ity, restricted typology, and tiers as a cognitively grounded succinctness metric. MDL comparisons thus would penalize subregular syntax for the very thing it is built on: sacrificing some succinctness for the linguistic benefits of limited expressivity.

Both EI1 and EI2 thus would have the opposite effect of what Chesi seems to envision: rather than elevating innovative work like subregular syntax, EI would stop it dead in its tracks.

3. Conclusions

Chesi's paper presents a grand vision of generative syntax in the 21st century, but it remains unclear just what exactly this vision ought to look like in practice. The paper allows for many interpretations, from the moderate (MI) to the extreme (EI). While MI is fairly uncontroversial, EI is so strong that it would undermine even those enterprises that currently come closest to Chesi's vision. I discussed subregular syntax as a concrete example of this dynamic. Irrespective of how one feels about Chesi's vision, then, there is value to it, but only in moderation. The dose makes the poison.

Abbreviations

EI = extreme interpretation; LLM = Large Language Models; MDL = Minimum Description Length; MI = moderate interpretation.

Notes

[1] Piantadosi and Chesi each use the term *generative linguistics* in their papers even though they only consider the subfield of *generative syntax* and in particular Minimalist syntax. Notably, generative phonology had a very similar debate years ago prompted by the recent advances in NLP (Pater 2019, Rawski & Heinz 2019, a.o.), and many of the arguments made there carry over to the current conversation.

[2] Note that any system that can compute this "exactly one" requirement can also compute its weakened counterpart "at least one". Graf and Kostyszyn (2021) use this fact to explain the existence of multiple wh-movement.

Bibliographical References

See the unified list at the end of this issue.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Carnie, Andrew 2013. *Syntax: A Generative Introduction, Third Edition*. Malden, MA: Wiley Blackwell.
Carnie, Andrew 2021. *Syntax: A Generative Introduction, Fourth Edition*. Malden, MA: Wiley Blackwell.
Cauchy, Augustin 1847. Méthode générale pour la résolution des systèmes d'équations simultanées. *Comptes rendus hebdomadaires des séances de l'Académie des sciences* 25. 536-538.
Cecchetti, Gabriele; Tomasini, Cedric A.; Herff, Steffen A. & Rohrmeier, Martin A. 2023. Interpreting rhythm as parsing. *Cognitive Science* 47. e13389. <doi.org/10.1111/cogs.13389>.
Chaitin, Gregory J. 1969. On the Simplicity and Speed of Programs for Computing Infinite Sets of Natural Numbers. *Journal of the ACM* 16,3. 407-422. <doi.org/10.1145/321526.321530>.
Chandlee, Jane 2014. Strictly local phonological processes. PhD dissertation. University of Delaware.
Chandlee, Jane 2017. Computational locality in morphological maps. *Morphology* 27. 599-641.
Chandlee, Jane 2022. Less is more: Reexamining assumptions through the narrow focus of subregularity. *Theoretical Linguistics* 48. 205-218.
Chandlee, Jane & Heinz, Jeffrey 2018. Strict locality and phonological maps. *Linguistic Inquiry* 49. 23-60.
Charchidi, Vincent J. 2024. Creative Minds Like Ours? Large Language Models and the Creative Aspect of Language Use. *Biolinguistics* 18. 1-31.
Charpentier, Lucas Georges Gabriel & Samuel, David 2023. Not all layers are equally as important: Every Layer Counts BERT. *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*. 210-224. <doi.org/10.18653/v1/2023.conll-babylm.20>.
Chen, Binglin; Lewis, Colleen M.; West, Matthew & Zilles, Craig 2024. Plagiarism in the age of Generative AI: Cheating method change and learning loss in an Intro to CS Course. In *L@S '24: Eleventh ACM Conference on Learning @ Scale, Atlanta GA USA*. New York, NY: ACM. 75-85. <DOI: 10.1145/3657604.3662046>.
Chen, Tianlong; Frankle, Jonathan; Chang, Shiyu; Liu, Sijia; Zhang, Yang; Wang, Zhangyang & Carbin, Michael 2020. The lottery ticket hypothesis for pre-trained BERT networks. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M. F. & Lin, H. (eds.), *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*. Online: Curran Associates, Inc. 15834-15846.
Chen, Zhong & Hale, John T. 2010. Deforesting logical form. *Procs. Mathematics of Language*. Berlin: Springer. LNCS 6149. <doi.org/10.1007/978-3-642-14322-9_2>.
Cheng, Lisa L.-S.; Heycock, Caroline & Zamparelli, Roberto 2017. Two levels for definiteness. In Erlewine, M. Y. (ed.), *Proceedings of GLOW in Asia XI – Vol. 1*. Volume 84 of *MIT Working Papers in Linguistics*. MIT.
Cheng, Lisa L.-S. & Sybesma, Rint 1999. Bare and not-so-bare nouns and the

structure of NP. *Linguistic Inquiry* 30,4. 509-542.
Chesi, Cristiano 2007. An introduction to phase-based minimalist grammars: why move is top-down from left-to-right. In Moscati, V. (ed.), *STIL – Studies in Linguistics*, Volume 1. CISCL Press. 38-75.
Chesi, Cristiano 2021. Expectation-based Minimalist Grammars. <arxiv.org/abs/2109.13871>.
Chesi, Cristiano 2023. Parameters of cross-linguistic variation in expectation-based Minimalist Grammars (e-MGs). *Italian Journal of Computational Linguistics* 9,1. 21.
Chesi, Cristiano *forthcoming*. Linearization (as Part of Core Syntax). In Grohmann, Kleanthes & Leivada, Evelina (eds.), *Cambridge Handbook of Minimalism*. Cambridge (UK): Cambridge University Press. <ling.auf.net/lingbuzz/006689>.
Chesi, Cristiano; Barbini, Matilde; Bressan, Veronica; Neri, Sofia; Piccini Bianchessi, Maria Letizia; Sarah, Rossi & Sgrizzi, Tommaso 2024. Different Ways to Forget: Linguistic Gates in Recurrent Neural Networks. In *Proceedings of the BabyLM Challenge at the 28th Conference on Computational Natural Language Learning*.
Chesi, Cristiano & Bianchi, Valentina 2014. Subject islands, reconstruction, and the flow of the computation. *Linguistic Inquiry* 45,4. 525-569.
Chesi, Cristiano & Moro, Andrea 2015. The subtle dependency between Competence and Performance. *MIT Working Papers In Linguistics* 77. 33-46.
Chesi, Cristiano; Vespignani, Francesco & Zamparelli, Roberto *to appear*. Large language models under evaluation: An acceptability, complexity and coherence assessment in Italian. *Italian Journal of Computational Linguistics*.
Chierchia, Gennaro 1998. Reference to kinds across languages. *Natural Language Semantics* 6. 339-405.
Cho, Kyunghyun; van Merrienboer, Bart; Gulcehre, Caglar; Bahdanau, Dzmitry; Bougares, Fethi; Schwenk, Holger & Bengio, Yoshua 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Moschitti, Alessandro; Pang, Bo & Daelemans, Walter (eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar: Association for Computational Linguistics. 1724-1734. <DOI: 10.3115/v1/D14-1179>.
Chomsky, Noam 1956. Three models for the description of language. *IEEE Transactions on Information Theory* 2,3. 113-124. <doi.org/10.1109/TIT.1956.1056813>.
Chomsky, Noam 1957. *Syntactic Structures*. Berlin: Mouton de Gruyter.
Chomsky, Noam 1959. A Review of B. F. Skinner's Verbal Behavior. *Language* 35,1. 26. <doi.org/10.2307/411334>.
Chomsky, Noam 1964. *Current Issues in Linguistic Theory*. Berlin: De Gruyter.
Chomsky, Noam 1965. *Aspects of the Theory of Syntax* (Vol. 11). Cambridge, MA: MIT Press.

Chomsky, Noam 1966. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. New York, NY: Harper & Row.
Chomsky, Noam 1968. *Language and Mind*. New York, NY: Harcourt, Brace & World.
Chomsky, Noam 1968b. Quine's Empirical Assumptions. *Synthese* 19,1-2. 53-68. <doi.org/10.1007/bf00568049>.
Chomsky, Noam 1969. Quine's empirical assumptions. In Davidson, Donald & Hintikka, Jaakko (eds.), *Words and Objections: Essays on the Work of W.V. Quine*. Dordrecht, Netherlands: Springer Dordrecht. 53-68. <DOI: 10.1007/978-94-010-1709-1_5>.
Chomsky, Noam 1975. *Questions on Form and Interpretation*. Lisse: Peter de Ridder. <doi.org/10.1007/978-3-642-14322-9_2>.
Chomsky, Noam 1981. *Lectures on government and binding: The Pisa lectures*. Walter de Gruyter.
Chomsky, Noam 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
Chomsky, Noam 1995. *The minimalist program*. Cambridge, MA: MIT Press.
Chomsky, Noam 1995b. Language and Nature. *Mind* 104 (413). 1-61.
Chomsky, Noam 2001. Derivation by phase. In Kenstowicz, Michael (ed.), *Ken Hale: A life in language*. Cambridge, MA: MIT Press. 1-52.
Chomsky, Noam A. 2004. *The generative enterprise revisited. Discussions with Riny Huybregts, Henk van Riemsdijk, Naoki Fukui and Mihoko Zushi*. De Gruyter Mouton.
Chomsky, Noam A. 2005. Three Factors in Language Design. *Linguistic Inquiry* 36,1. 1-22.
Chomsky, Noam 2008. On phases. In Freidin, Robert; Otero, Carlos P. & Zubizarreta, Maria Luisa (eds.), *Foundational issues in linguistic theory: Essays in Honor of Jean-Roger Vergnaud* (Vol. 45). Cambridge, MA: MIT Press. 133-166.
Chomsky, Noam 2012. Language and Limits of Understanding. <www.nets.iusspavia.it/dox/chomsky2012-LLU-IUSS_Pavia.pdf>.
Chomsky, Noam 2013. Problems of projection. *Lingua* 130. 33-49.
Chomsky, Noam 2015. Problems of projection: Extensions. In Di Domenico, Elisa; Hamann, Cornelia & Matteini, Simona (eds.), *Linguistik Aktuell/ Linguistics Today* (Vol. 223). Amsterdam: John Benjamins. 1-16. <doi.org/10.1075/la.223.01cho>.
Chomsky, Noam 2021a. Simplicity and the form of grammars. *Journal of Language Modelling* 9,1. <doi.org/10.15398/jlm.v9i1.257>.
Chomsky, Noam 2021b. Minimalism: where are we now, and where can we hope to go. *Gengo Kenkyu* 160. 1-42.
Chomsky, Noam 2024. The Miracle Creed and SMT. In Greco, M. & Mocci, D. (eds.), *A Cartesian dream: A geometrical account of syntax: In honor of Andrea Moro*. Rivista di Grammatica Generativa / Research in Generative Grammar 17-40.
Chomsky, Noam & Lasnik, Howard 1977. Filters and Control. *Linguistic*

Inquiry 8,3. 425-504.
Chomsky, Noam; Roberts, Ian & Watumull, Jeffrey 2023. Noam Chomsky: The False Promise of ChatGPT. *New York Times* 8 March.
Chomsky, Noam; Seely, T. Daniel; Berwick, Robert C.; Fong, Sandiway; Huybregts, M. A. C.; Kitahara, Hisatsugu; McInnerney, Andrew & Sugimoto, Yushi 2023. *Merge and the Strong Minimalist Thesis* (1st ed.). Cambridge: Cambridge University Press. <doi.org/10.1017/9781009343244>.
Chowdhury, Shammur Absar & Zamparelli, Roberto 2018. RNN Simulations of Grammaticality Judgments on Long-distance Dependencies. *Proceedings of the 27th International Conference on Computational Linguistics*. 133-144. <aclanthology.org/C18-1012>.
Cinque, Guglielmo 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford, UK: Oxford University Press.
Cinque, Guglielmo 2002. *Functional Structure in DP and IP: The Cartography of Syntactic Structures, Volume 1*. Oxford, UK: Oxford University Press.
Cinque, Guglielmo 2005. Deriving Greenberg's Universal 20 and Its Exceptions. *Linguistic Inquiry* 36,3. 315-332. <doi.org/10.1162/0024389054396917>.
Cinque, Guglielmo & Rizzi, Luigi 2010. The Cartography of Syntactic Structures. In Heine, B. & Narrog, H. (eds.), *The Oxford Handbook of Linguistic Analysis*. Oxford / New York: Oxford University Press. 51-65.
Clark, Alexander & Lappin, Shalom 2010. Computational learning theory and language acquisition. *Philosophy of Linguistics*. 445-475.
Clark, Alexander & Lappin, Shalom 2011. *Linguistic Nativism and the Poverty of the Stimulus*. Chichester: Wiley-Blackwell.
Clifton, Charles Jr; Ferreira, Fernanda; Henderson, John M.; Inhoff, Albrecht W.; Liversedge, Simon P.; Reichle, Erik D. & Schotte, Elizabeth R. 2015. Eye movements in reading and information processing. *Journal of Memory and Language* 86. 1-19.
Collins, Chris; Kayne, Richard & Koopman, Hilda 2009. *Syntactic structures of the world's languages (SSWL)*. <terraling.com/groups/7>.
Collins, Chris & Stabler, Edward P. 2016. A Formalization of Minimalist Syntax. *Syntax* 19,1. 43-78. <doi.org/10.1111/synt.12117>.
Collins, Joe 2024. The simple reason LLMs are not scientific models (and what the alternative is for linguistics). <lingbuzz.net/lingbuzz/008026>.
Conneau, Alexis; Kruszewski, German; Lample, Guillaume; Barrault, Loïc & Baroni, Marco 2018. What you can cram into a single $&!#* vector: Probing sentence embeddings for linguistic properties. In Gurevych, Iryna & Miyao, Yusuke (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics. 2126-2136. <DOI: 10.18653/v1/P18-1198>.
Corbett, Greville G. 2010. Implicational hierarchies. In Song, Jae

Jong (ed.), *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press. 190-205. <doi.org/10.1093/oxfordhb/9780199281251.013.0011>.
Cottier, Ben; Rahman, Robi; Fattorini, Loredana; Maslej, Nestor; Besiroglu, Tamay & Owen, David 2025. The rising costs of training frontier AI models. <arXiv:2405.21015>.
Crain, Stephen & Nakayama, Mineharu 1987. Structure Dependence in Grammar Formation. *Language* 63,3. 522. <doi.org/10.2307/415004>.
Crain, Stephen & Thornton, Rosalind 2021. Universal grammar and language acquisition. In Allot, Nicholas; Lohndahl, Terje & Rey, Georges (eds.), *A Companion to Chomsky*. Wiley. <doi.org/10.1002/9781119598732.ch21>.
Crawford, Kate 2024. Generative AI's environmental costs are soaring – and mostly secret. *Nature* 626. 693. <DOI: 10.1038/d41586-024-00478-x>.
Crystal, David 2011. *Internet Linguistics: A Student Guide*. London: Routledge.
Cutler, Anne & Fodor, Jerry A. 1979. Semantic focus and sentence comprehension. *Cognition* 7. 49-59. <doi.org/10.1016/0010-0277(79)90010-6>.
Cybenko, George 1989 Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems* 2,4. 303-314.
Dahl, Östen 2020. Morphological complexity and the minimum description length approach. In Arkadiev, Peter & Gardani, Francesco (eds.), *The complexities of morphology*. Oxford: Oxford University Press. 331-343.
D'Alessandro, Roberta 2019. The achievements of Generative Syntax: A time chart and some reflections. *Catalan Journal of Linguistics*. 7-26.
Dalrymple, Mary (ed.) 2023. *The Handbook of Lexical Functional Grammar*. Empirically Oriented Theoretical Morphology and Syntax. Berlin: Language Science Press. <10.5281/zenodo.10037797>.
Dalrymple, Mary; Gupta, Vineet; Lamping, John & Saraswat, Vijay 1999. Relating resource-based semantics to categorial semantics. In Dalrymple, Mary (ed.), *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*. Language, Speech, and Communication. Cambridge, MA: MIT Press. 261-280.
Dalrymple, Mary; Patejuk, Agnieszka & Zymla, Mark-Matthias 2020. XLE + Glue – A new tool for integrating semantic analysis in XLE. In Butt, Miriam & Toivonen, Ida (eds.), *Proceedings of the LFG'20 Conference*. Stanford, CA: CSLI Publications. 89-108. <cslipublications.stanford.edu/LFG/2020/lfg2020-dpz.pdf>.
De Santo, Aniello 2019. Testing a Minimalist Grammar Parser on Italian Relative Clause Asymmetries. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*. 93-104. <doi.org/10.18653/v1/W19-2911>.
De Santo, Aniello 2020. Structure and memory: A computational model of storage, gradience, and priming. PhD dissertation. Stony Brook University.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

works, and grammatical structure. *Machine Learning* 7,2. 195-225. <DOI: 10.1023/A:1022699029236>.
Elman, Jeffrey L. 1993. Learning and development in neural networks: The importance of starting small. *Cognition* 48,1. 71-99. <doi.org/10.1016/0010-0277(93)90058-4>.
Engelfriet, Joost; Lilin, Eric & Maletti, Andreas 2009. Extended multi bottom-up tree transducers: Composition and decomposition. *Acta Informatica* 46. 561-590. <doi.org/10.1007/s00236-009-0105-8>.
Epstein, Samuel David; Groat, Erich M.; Kawashima, Ruriko & Kitahara, Hisatsugu (eds.) 1998. *A derivational approach to syntactic relations*. Oxford, UK: Oxford University Press.
Ermolaeva, Marina 2023. Evaluating syntactic proposals using Minimalist grammars and minimum description length. *Journal of Language Modelling* 11. 67-119. <doi.org/10.15398/jlm.v11i1.334>.
Espinal, Maria Teresa & Cyrino, Sonia 2022. A syntactically-driven approach to indefiniteness, specificity and antispecificity in Romance. *Journal of Linguistics* 58. 535-570.
Ettinger, Allyson 2020. What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. *Transactions of the Association for Computational Linguistics* 8. 34-48. <doi.org/10.1162/tacl a 00298>.
Evans, Lyndon 2007. The Large Hadron Collider. *New Journal of Physics* 9,9. 335-335. <doi.org/10.1088/1367-2630/9/9/335>.
Evans, Nicholas & Levinson, Stephen C. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32,5. 429-448. <DOI:10.1017/S0140525X0999094X>.
Evanson, Linnea; Lakretz, Yair & King, Jean-Rémi 2023. Language acquisition: do children and language models follow similar learning stages? <arXiv:2306.03586>.
Fazi, M. Beatrice 2021. Beyond human: Deep learning, explainability and representation. *Theory, Culture & Society* 38. 55-77.
Feyerabend, Paul K. 1962. Explanation, reduction, and empiricism. In Feigl, Herbert & Maxwell, Grover (eds.), *Scientific explanation, space, and time.* Vol. 3. Minneapolis, MN: University of Minnesota Press. 28-97.
Fisher, Cynthia 2002. The role of abstract syntactic knowledge in language acquisition: A reply to Tomasello (2000). *Cognition* 82. 259-278.
Fleck, Ludwik 1935. *Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv*. Basel, Switzerland: Benno Schwabe & Co.
Fodor, Janet Dean 1998. Unambiguous triggers. *Linguistic Inquiry* 29. 1-36.
Fodor, Jerry A. 1980. *The Language of Thought*. Harvard: Harvard University Press.
Fodor, Jerry A. 1983. *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

all-day recordings and automated analysis. *American Journal of Speech-Language Pathology* 26. 248-265. <DOI: 10.1044/2016_AJSLP-15-016>.
Ginsburg, Jason 2016. Modeling of Problems of Projection: A non-countercyclic approach. *Glossa: A Journal of General Linguistics* 1,1:7. 1-46. <DOI: 10.5334/gjgl.22>.
Ginsburg, Jason 2024. Constraining free Merge. *Biolinguistics* 18, e14015. 1-60. <DOI: 10.5964/bioling.14015>.
Ginsburg, Jason & Fong, Sandiway 2019. Combining linguistic theories in a Minimalist Machine. In Stabler, Edward P. & Berwick, Robert C. (eds.), *Minimalist Parsing*. Oxford, UK: Oxford University Press. 39-68. <doi.org/10.1093/oso/9780198795087.003.0003>.
Giusti, Giuliana 2015. *Nominal Syntax at the Interfaces: A Comparative Analysis of Languages With Articles*. Cambridge: Cambridge Scholars Publishing.
Gold, E. Mark 1967. Language identification in the limit. *Information and Control* 10,5. 447-474. <doi.org/10.1016/S0019-9958(67)91165-5>.
Goldsmith, John & Riggle, Jason 2012. Information theoretic approaches to phonological structure: The case of Finnish vowel harmony. *Natural Language and Linguistic Theory* 30. 859-896.
Gorman, Kyle 2016. Pynini: A Python library for weighted finite-state grammar compilation. In *Procs. SIGFSM Workshop on Statistical NLP and Weighted Automata*. <doi.org/10.18653/v1/W16-2409>.
Goyal, Anirudh & Bengio, Yoshua 2022. Inductive Biases for Deep Learning of Higher-Level Cognition. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 478 (2266). <doi.org/10.1098/rspa.2021.0068>.
Graf, Thomas 2020. Curbing feature coding: Strictly local feature assignment. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2020*. 362-371.
Graf, Thomas 2022c. Typological implications of tier-based strictly local movement. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2022*. 184-193.
Graf, Thomas 2022b. Subregular linguistics: Bridging theoretical linguistics and formal grammar. *Theoretical Linguistics* 48. 145-184. <doi.org/10.1515/tl-2022-2037>.
Graf, Thomas 2022a. Diving deeper into subregular syntax. *Theoretical Linguistics* 48. 245-278. <doi.org/10.1515/tl-2022-2043>.
Graf, Thomas 2023. Subregular tree transductions, movement, copies, traces, and the ban on improper movement. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2023*. 289-299. <doi.org/10.7275/tk1n-q855>.
Graf, Thomas *to appear*. Minimalism and computational linguistics. In Grohman, Kleanthes K. & Leivada, Evelina (eds.), *Handbook of Minimalism*. Cambridge: Cambridge University Press.
Graf, Thomas & Abner, Natasha 2012. Is syntactic binding rational?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Second Meeting of the North American Chapter of the Association for Computational Linguistics. <aclanthology.org/N01-1021>.
Hale, John 2011. What a rational parser would do. *Cognitive Science* 35,3. 399-443.
Hale, John 2016. Information-theoretical Complexity Metrics. *Language and Linguistics Compass* 10,9. 397-412. <doi.org/10.1111/lnc3.12196>.
Hanson, Kenneth 2025. Tier-based strict locality and the typology of agreement. *Journal of Language Modelling* 13,1. 43-97. <doi.org/10.15398/jlm.v13i1.411>.
Hanson, Kenneth 2024. Tiers, paths, and syntactic locality: The view from learning. In *Proceedings of the society for computation in linguistics (SCiL) 2024*. 107-116. <doi.org/10.7275/scil.2135>.
Hao, Sophie 2022. *Theory and Applications of Attribution for Interpretable Language Technology*. PhD dissertation. Yale University, New Haven, CT.
Hao, Sophie; Angluin, Dana & Frank, Robert 2022. Formal language recognition by hard attention transformers: Perspectives from circuit complexity. *Transactions of the Association for Computational Linguistics* 10. 800-810. <DOI: 10.1162/tacl_a_00490>.
Hao, Sophie; Mendelsohn, Simon; Sterneck, Rachel; Martinez, Randi & Frank, Robert 2020. Probabilistic predictions of people perusing: Evaluating metrics of language model performance for psycholinguistic modeling. In Chersoni, Emmanuele; Jacobs, Cassandra; Oseki, Yohei; Prévot, Laurent & Santus, Enrico (eds.), *Workshop on Cognitive Modeling and Computational Linguistics*. Online: Association for Computational Linguistics.
Hao, Sophie & Andersson, Samuel 2019. Unbounded stress in subregular phonology. In *Proceedings of the 16th Sigmorphon workshop on computational research in phonetics, phonology and morphology*. 135-143. <doi.org/10.18653/v1/W19-4216>.
Hart, Betty & Risley, Todd R. 1992. American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology* 28,6. 1096-1105. <doi.org/10.1037/0012-1649.28.6.1096>.
Haspelmath, Martin 1993. *A grammar of Lezgian*. Mouton Grammar Library 9. Berlin: Mouton de Gruyter.
Haspelmath, Martin 2007. Pre-established categories don't exist – consequences for language description and typology. *Linguistic Typology* 11. 119-132.
Haspelmath, Martin 2008. Parametric versus functional explanations of syntactic universals. In Biberauer, Theresa (ed.), *The limits of syntactic variation*. Amsterdam: Benjamins. Accessed 27 May 2016.
Haspelmath, Martin 2010a. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86,3. 663-687. <doi:10.1353/lan.2010.0021>.
Haspelmath, Martin 2010b. Framework-free grammatical theory. In Heine,

Bernd & Narrog, Heiko (eds.), *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press. 341-365.
Haspelmath, Martin 2018. How comparative concepts and descriptive linguistic categories are different. In Van Olmen, Daniël; Mortelmans, Tanja & Brisard, Frank (eds.), *Aspects of linguistic variation: Studies in honor of Johan van der Auwera*. Berlin: De Gruyter Mouton. 83-113. <zenodo.org/record/3519206>.
Haspelmath, Martin 2020. Human linguisticality and the building blocks of languages. *Frontiers in Psychology* 10,3056. 1-10. <doi:10.3389/fpsyg.2019.03056>.
Haspelmath, Martin 2021. General linguistics must be based on universals (or nonconventional aspects of language). *Theoretical Linguistics* 47,1-2. 1-31. <doi:10.1515/tl-2021-2002>.
Haspelmath, Martin *to appear*. Breadth versus depth: Theoretical reasons for system-independent comparison of languages. In Nefdt, Ryan (ed.), *Oxford Handbook of Philosophy of Linguistics*. Oxford: Oxford University Press. <ling.auf.net/lingbuzz/008437>.
Hauser, M. D.; Chomsky, N. & Fitch, W. T. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298 (5598). 1569-1579. <doi.org/10.1126/science.298.5598.1569>.
Hawkins, John A. 2014. *Cross-linguistic variation and efficiency*. New York: Oxford University Press.
Heim, Johannes & Wiltschko, Martina 2025. Rethinking structural growth: Insights from the acquisition of interactional language. *Glossa: A journal of general linguistics* 10,1. <doi.org/10.16995/glossa.16396>.
Heinz, Jeffrey 2010. Learning long-distance phonotactics. *Linguistic Inquiry* 41. 623-661. <doi.org/10.1162/LING_a_00015>.
Heinz, Jeffrey 2018. The computational nature of phonological generalizations. In Hyman, Larry & Plank, Frank (eds.), *Phonological typology*. Mouton De Gruyter. 126-195.
Hewitt, John & Manning, Christopher D. 2019. A structural probe for finding syntax in word representation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 4129-4138.
Hey, Tony; Tansley, Stewart; Tolle, Kristin & Gray, Jim (eds.) 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft Research.
Hinton, Geoffrey 2022. The forward-forward algorithm: Some preliminary investigations. <arXiv:2212.13345>.
Hochreiter, Sepp; Bengio, Yoshua; Frasconi, Paolo & Schmidhuber, Jürgen 2001. Gradient flow in recurrent nets: The difficulty of learning long-term dependencies. In Kremer, S. C. & Kolen, J. F. (eds.), *A Field Guide to Dynamical Recurrent Neural Networks*. IEEE Press.
Hochreiter, Sepp & Schmidhuber, Jürgen 1997. Long short-term memory. *Neural Computation* 9,8. 1735-1780.

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

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

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

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

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

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

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

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

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

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

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

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

the Experimental Method of Reasoning Into Moral Subjects. London: John Noon.
Hume, David 1748. *Philosophical Essays Concerning Human Understanding*. London: A. Millar.
Hunter, Tim; Stanojević, Miloš & Stabler, Edward P. 2019. The active-filler strategy in a move-eager left-corner Minimalist grammar parser. In *Proceedings of the workshop on cognitive modeling and computational linguistics*. 1-10.
Ibbotson, Paul & Tomasello, Michael 2016. Evidence rebuts Chomsky's theory of language learning. *Scientific American* 315, 5. 70.
İdrisoğlu, Işıl & Spaniel, William 2024. *Information problems and Russia's invasion of Ukraine. Conflict Management and Peace Science* 41,5. 514-533. <DOI: 10.1177/07388942241238583>.
Ionin, Tania & Matushansky, Ora 2006. The composition of complex cardinals. *Journal of Semantics* 16. 315-360.
Jackendoff, Ray 1988. Why are they saying these things about us? *Natural Language and Linguistic Theory* 6,3. 435-442.
Jardine, Adam 2016. Computationally, tone is different. *Phonology* 33. 247-283. <doi.org/10.1017/S0952675716000129>.
Ji, Ziwei; Lee, Nayeon; Frieske, Rita; Yu, Tiezheng; Su, Dan; Xu, Yan; Ishii, Etsuko; Bang, Ye Jin; Madotto, Andrea & Fung, Pascale 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys* 55,12. 248:1-248:38. <DOI: 10.1145/3571730>.
Jurafsky, Dan & Martin, James H. 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd edition. Russell, Stuart & Norvig, Peter (eds.). Upper Saddle River, NJ: Prentice Hall.
Kalouli, Aikaterini-Lida 2021. *Hy-NLI: A hybrid system for state-of-the-art natural language inference*. University of Konstanz dissertation.
Kalouli, Aikaterini-Lida; Crouch, Richard & de Paiva, Valeria 2020. Hy-NLI: A hybrid system for natural language inference. In *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona, Spain (Online): International Committee on Computational Linguistics. 5235-5249. <aclanthology.org/2020.coling-main.459>.
Kaplan, Jared; McCandlish, Sam; Henighan, Tom; Brown, Tom B.; Chess, Benjamin; Child, Rewon; Gray, Scott; Radford, Alec; Wu, Jeffrey & Amodei, Dario 2020. *Scaling Laws for Neural Language Models*. <doi.org/10.48550/ARXIV.2001.08361>.
Kaplan, Ronald M. 1987. Three seductions of computational linguistics. In Whitelock, P.; Wood, M. M.; Somers, H.; Johnson, R. & Bennett, P. (eds.), *Linguistic Theory and Computer Applications*. London: Academic Press. 149-188.
Kaplan, Ronald M. 2019. Computational psycholinguistics. *Computational Linguistics* 45,4. 607-626. <doi:10.1162/coli a 00359>. <aclanthology.org/J19-4001>.
Kaplan, Ronald M.; King, Tracey H. & Maxwell, John T. III 2002. Adapting

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Milway, Daniel 2023. A response to Piantadosi (2023). <lingbuzz/007264>.

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

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

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

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

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

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

321-328. <doi.org/10.1422/108132>.
Moro, Andrea; Greco, Matteo & Cappa, Stefano F. 2023. Large languages, impossible languages and human brains. *Cortex* 167. 82-85. <doi.org/10.1016/j.cortex.2023.07.003>.
Müller, Stefan 2024. Large language models: The best linguistic theory, a wrong linguistic theory, or no linguistic theory at all. *Zeitschrift für Sprachwissenschaft*.
Mullins, Nicholas C. 1975. A sociological theory of scientific revolution. In Knorr, Karin D.; Strasser, Hermann & Zilian, Hans Georg (eds.), *Determinants and Controls of Scientific Development*. Dordrecht, Netherlands: Springer Netherlands. 185-203.
Murray, Stephen O. 1994. *Theory Groups and the Study of Language in North America*. Amsterdam, Netherlands: John Benjamins.
Murty, Shikhar; Sharma, Pratyusha; Andreas, Jacob & Manning, Christopher D. 2022. Characterizing intrinsic compositionality in transformers with tree projections.
Naveed, Humza; Asad Ullah Khan; Shi Qiu; Saqib, Muhammad; Anwar, Saeed; Usman, Muhammad; Akhtar, Naveed; Barnes, Nick & Mian, Ajmal 2024. A comprehensive overview of large language models. <arxiv.org/abs/2307.06435>.
Newmeyer, Frederick J. 1980. *Linguistic theory in America: The first quarter century of Transformational Generative Grammar*. New York: Academic Press.
Newmeyer, Frederick J. 1986. Has there been a 'Chomskyan revolution' in linguistics? *Language* 62,1. 1-18. <DOI: 10.2307/415597>.
Newmeyer, Frederick J. 2004. Against a parameter-setting approach to typological variation. *Linguistic Variation Yearbook* 4,1. 181-234. <doi:10.1075/livy.4.06new>.
Newmeyer, Frederick J. 2021. Complexity and relative complexity in generative grammar. *Frontiers in Communication* 6. <doi:10.3389/fcomm.2021.614352>.
Newmeyer, Frederick J. & Emonds, Joseph 1971. The linguist in American society. In *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society*. Chicago, IL: Chicago Linguistic Society. 285-303.
Nivre, Joakim; Agić, Željko; Ahrenberg, Lars; Antonsen, Lene; Aranzabe, Maria Jesus; Asahara, Masayuki; Ateyah, Luma; Attia, M.; Atutxa, A.; Augustinus, L. *et al.* 2017. *Universal Dependencies 2.1*.
Norvig, Peter 2017. On Chomsky and the two cultures of statistical learning. In Pietsch, Wolfgang; Wernecke, Jörg & Ott, Maximilian (eds.), *Berechenbarkeit der Welt? Philosophie und Wissenschaft im Zeitalter von Big Data*. Wiesbaden, Germany: Springer Fachmedien. 61-83.
Nosengo, Nicola 2014. *I robot ci guardano: Aerei senza pilota, chirurghi a distanza e automi solidali*. Bologna: Zanichelli.
Noy, Shakked & Zhang, Whitney 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381, 6654. 187-192. <DOI: 10.1126/science.adh2586>.
Nvidia *n.d.* meta/llama-3.1-405b-instruct. *Nvidia API reference*. <docs.api.nvidia.com/nim/reference/meta-llama-3_1-405b>.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

fieldwork to linguistic theory: A tribute to Dan Everett. Berlin: Language Science Press. 353-414.
Pinker, Steven 1984. *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.
Plate, Tony A. 1994. *Holographic Reduced Representation*. Stanford: CSLI.
Plato 380 BCE. *Meno*.
Poggio, Thomas; Rifkin, Ryan; Niyogi, Partha & Mukherjee, Sayan 2004. General conditions for predictivity in learning theory. *Nature* 428. 419-422. <doi.org/10.1038/nature02341>.
Pollard, Carl & Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, IL: University of Chicago Press.
Pollock, Jean Yves 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic inquiry* 20,3. 365-424.
Popper, Karl 1934. *Logik der Forschung*. Berlin: Springer. <doi.org/10.1007/978-3-7091-4177-9>.
Prasanna, Sai; Rogers, Anna & Rumshisky, Anna 2020. When BERT plays the lottery, all tickets are winning. In Webber, Bonnie; Cohn, Trevor; He, Yulan & Liu, Yang (eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics. 3208-3229. <DOI: 10.18653/v1/2020.emnlp-main.259>.
Pullum, Geoffrey K. & Scholz, Barbara C. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 18,1-2. 9. <DOI: 10.1515/tlir.19.1-2.9>.
Purnell, Thomas; Idsardi, William & Baugh, John 1999. Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology* 18,1. 10-30. <DOI: 10.1177/0261927X99018001002>.
Quine, Willard Van Orman 1960. *Word and Object*. Cambridge, MA: MIT Press. <doi.org/10.7551/mitpress/9636.001.0001>.
Quinlan, Philip T. (ed.) 2004. *Connectionist Models of Development* (0 ed.). Psychology Press. <doi.org/10.4324/9780203494028>.
Radford, Alec; Narasimhan, Karthik; Salimans, Tim; Sutskever, Ilya *et al.* 2018. *Improving language understanding by generative pre-training*.
Radford, Alec; Wu, Jeffrey; Amodei, Dario; Clark, Jack; Brundage, Miles & Sutskever, Ilya 2019a. Better language models and their implications [Blog post]. *OpenAI Research*. <openai.com/index/better-language-models>. Last accessed 24/02/2025.
Radford, Alec; Wu, Jeffrey; Child, Rewon; Luan, David; Amodei, Dario & Sutskever, Ilya 2019b. *Language Models Are Unsupervised Multitask Learners*. Technical report. San Francisco, CA: OpenAI.
Radford, Andrew 1997. *Syntax: A Minimalist Introduction*. Cambridge: Cambridge University Press.
Radford, Andrew 2016. *Analysing English Sentences, Second Edition*. Cambridge: Cambridge University Press.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Rizzi, Luigi (ed.) 2004. *The structure of CP and IP*. Oxford, UK: Oxford University Press.
Rizzi, Luigi 2013. Locality. *Lingua* 130. 169-186.
Rizzi, Luigi 2016. Labeling, maximality and the head-phrase distinction. *The Linguistic Review* 33,1. 103-127.
Rizzi, Luigi 2021. *Complexité des structures linguistiques, simplicité des mécanismes du langage*, Leçon inaugurale, 2021, Collège de France – Fayard, Paris. English translation: *Complexity of Linguistic Structures, Simplicity of Language Mechanisms* (2024). OpenEdition Books, Collège de France. <DOI: 10.4000/books.cdf.16006>.
Rizzi, Luigi & Cinque, Guglielmo 2016. Functional Categories and Syntactic Theory. *Annual Review of Linguistics* 2,1. 139-163. <doi.org/10.1146/annurev-linguistics-011415-040827>.
Rizzi, Luigi & Savoia, Leonardo 1993. Conditions on /u/ propagation in Southern Italian Dialects: A Locality Parameter for Phonosyntactic Processes. In Belletti, A. (ed.), *Syntactic Theory and the Dialects of Italy*. Turin: Rosenberg & Sellier.
Roberts, Ian 2017. The final-over-final condition in DP: Universal 20 and the nature of demonstratives. In Sheehan, Michelle; Biberauer, Theresa; Roberts, Ian & Holmberg, Anders (eds.), *The Final-over-Final Condition: A Syntactic Universal* (Vol. 76). Cambridge, MA: MIT Press. 151.
Roberts, Ian 2019. *Parameter Hierarchies and Universal Grammar* (1st ed.). Oxford, UK: Oxford University Press. <doi.org/10.1093/oso/9780198804635.001.0001>.
Rogers, Anna; Kovaleva, Olga & Rumshisky, Anna 2021. A primer in BERTology: What we know about how BERT works. *Transactions of the Association for Computational Linguistics* 8. 842-866.
Ross, John Robert 1967. *Constraints on variables in syntax*. Cambridge, MA: MIT Press.
Rumelhart, David E.; Hinton, Geoffrey E. & Williams, Ronald J. 1986. Learning representations by back-propagating errors. *Nature* 323,6088. 533-536. <DOI: 10.1038/323533a0>.
Rumelhart, David E. & McClelland, James L. 1986. On Learning the Past Tenses of English Verbs. In *Parallel Distributed Processing*. Cambridge, MA: MIT Press. <doi.org/10.7551/mitpress/5237.003.0008>.
Rumelhart, David E.; McClelland, James L. & PDP Research Group (eds.) 1999. *Parallel distributed processing. 1: Foundations*. 12th print. Cambridge, MA: MIT Press. <doi.org/10.7551/mitpress/5236.001.0001>.
Russell, Bertrand 1947. *Human Knowledge: Its Scope and Limits*. New York, NY: Simon and Schuster.
Sampson, Geoffrey 1997. *Educating Eve: The 'language instinct' debate*. London / Washington, DC: Cassell.
Sanh, Victor; Webson, Albert; Raffel, Colin; Bach, Stephen; Sutawika, Lintang; Alyafeai, Zaid; Chaffin, Antoine; Stiegler, Arnaud; Raja, Arun;

Dey, Manan *et al. 2022. Multitask prompted training enables zero-shot task generalization. In ICLR 2022 Conference Track*. Online: OpenReview.
Sarlin, Paul-Edouard; DeTone, Daniel; Malisiewicz, Tomasz & Rabinovich, Andrew 2020. Superglue: Learning feature matching with graph neural networks. <arxiv.org/abs/1911.11763>.
Sartran, Laurent; Barrett, Samuel; Kuncoro, Adhiguna; Stanojević, Miloš; Blunsom, Phil & Dyer, Chris 2022. Transformer Grammars: Augmenting Transformer Language Models with Syntactic Inductive Biases at Scale. *Transactions of the Association for Computational Linguistics* 10 (December). 1423-39. <doi.org/10.1162/tacl_a_00526>.
Sathish, Vishwas; Lin, Hannah; Kamath, Aditya K. & Nyayachavadi, Anish 2024. LLeMpower: Understanding disparities in the control and access of large language models. <arXiv:2404.09356>.
Savitch, Walter J. 1993. Why it might pay to assume that languages are infinite. *Annals of Mathematics and Artificial Intelligence* 8. 17-25.
Shannon, C. E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27,3. 379-423. <DOI: 10.1002/j.1538-7305.1948.tb01338.x>.
Shieber, Stuart M. 1985. Evidence against the Context-Freeness of Natural Language. *Linguistics and Philosophy* 8,3. 333-43. <doi.org/10.1007/BF00630917>.
Siegelman, Noam; Schroeder, Sascha; Acartürk, Cengiz; Ahn, Hee-Don; Alexeeva, Svetlana; Amenta, Simona; Bertram, Raymond; Bonandrini, R.; Brysbaert, M.; Chernova, D.; Da Fonseca, S. M.; Dirix, N.; Duyck, W.; Fella, A.; Frost, R.; Gattei, C. A.; Kalaitzi, A.; Kwon, N.; Lõo, K.; … Kuperman, V. 2022. Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior Research Methods* 54,6. 2843-2863. <doi.org/10.3758/s13428-021-01772-6>.
Smith, Nathaniel J. & Levy, Roger 2013. The effect of word predictability on reading time is logarithmic. *Cognition* 128,3. 302-319. <DOI: 10.1016/j.cognition.2013.02.013>.
Smolensky, Paul 1990. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence* 46. 159-216. <doi.org/10.1016/0004-3702(90)90007-m>.
Solomonoff, Ray J. 1960. *A Preliminary Report on a General Theory of Inductive Inference*. United States Air Force, Office of Scientific Research. <books.google.it/books?id=SuTHtgAACAAJ>.
Spitale, Giovanni; Biller-Andorno, Nikola & Germani, Federico 2023. AI model GPT-3 (dis)informs us better than humans. *Science Advances* 9, 26. <DOI: 10.1126/sciadv.adh1850>.
Sprouse, Jon & Almeida, Diogo 2017. Design sensitivity and statistical power in acceptability judgment experiments. *Glossa* 2,1. 1-32. <doi.org/10.5334/gjgl.236>.
Sprouse, Jon & Hornstein, Norbert (eds.) 2013. *Experimental Syntax and*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

v1/2021.findings-emnlp.250>.
Turing, Alan M. 1937. Computability and λ-definability. *Journal of Symbolic Logic* 2. 153-163. <doi.org/10.2307/2268280>.
Turing, Alan 1950. Computing machinery and intelligence. *Mind* 59. 433-460. <DOI: 10.1093/mind/lix.236.433>.
van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Oxford University Press. 97-157.
van Riemsdijk, Henk & Williams, Edwin 1986. *Introduction to the Theory of Grammar*. Cambridge, MA: MIT Press.
van Rooij, Iris; Guest, Olivia; Adolfi, Federico; de Haan, Ronald; Kolokova, Antonina & Rich, Patricia 2024. Reclaiming AI as a theoretical tool for cognitive science. *Computational Brain and Behaviour*.
Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N.; Kaiser, Lukasz & Polosukhin, Illia 2017. Attention Is All You Need. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S. & Garnett, R. (eds.), *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. Long Beach, CA: Curran Associates, Inc. 5998-6008. <arxiv.org/abs/1706.03762>.
Vermeerbergen, Myriam; Leeson, Lorraine & Crasborn, Onno Alex (eds.) 2007. *Simultaneity in signed languages: Form and function*. Amsterdam: John Benjamins.
Voldoire, A.; Sanchez-Gomez, E.; Salas y Mélia, D.; Decharme, B.; Cassou, C.; Sénési, S.; Valcke, S.; Beau, I.; Alias, A.; Chevallier, M. *et al.* 2013. The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dynamics* 40,9. 2091-2121. <DOI: 10.1007/s00382-011-1259-y>.
von Humboldt, Wilhelm 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwickelung des Menschengeschlechts*. Berlin, Prussia: Druckerei der Königlichen Akademie der Wissenschaften.
Wadler, Philip 1990. Deforestation: Transforming programs to eliminate trees. *Theoretical Computer Science* 73. 231-248. <doi.org/10.1016/0304-3975(90)90147-A>.
Warstadt, Alex & Bowman, Samuel R. 2022. What artificial neural networks can tell us about human language acquisition. In Lappin, Shalom & Bernardy, Jean-Phillipe (eds.), *Algebraic Structures in Natural Language*. Boca Raton: CRC Press, Taylor & Francis. 17-60.
Warstadt, Alex; Mueller, Aaron; Choshen, Leshem; Wilcox, Ethan; Zhuang, Chengxu; Ciro, Juan; Mosquera, Rafael; Paranjabe, B.; Williams, A.; Linzen, T. & Cotterell, R. 2023. Findings of the BabyLM Challenge: Sample-Efficient Pretraining on Developmentally Plausible Corpora. *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*. Singapore: Association for Computational Linguistics. 1-6. <doi.org/10.18653/v1/2023.conll-babylm.1>.
Warstadt, Alex; Parrish, Alicia; Liu, Haokun; Mohananey, Anhad; Peng, Wei; Wang, Sheng-Fu & Bowman, Samuel R. 2020. BLiMP: The

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Wiltschko, Martina & Heim, Johannes 2016. The syntax of confirmationals. A neo-performative analysis. In Kaltenböck, Gunther; Keizer, Evelien & Lohmann, Arne (eds.), *Outside the Clause. Form and function of extra-clausal 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>.

Printed in June 2025
by Industrie Grafiche Pacini Editore Srl
Via A. Gherardesca • 56121 Ospedaletto • Pisa • Italy
Tel. +39 050 313011 • Fax +39 050 3130300
www.pacinieditore.it