Natural language syntax complies with the free-energy principle

Elliot Murphy[1,2], Emma Holmes[3], Karl Friston[3]

1. Vivian L. Smith Department of Neurosurgery, McGovern Medical School, University of Texas Health Science Center, Houston, Texas, 77030, USA
2. Texas Institute for Restorative Neurotechnologies, University of Texas Health Science Center, Houston, Texas, 77030, USA
3. The Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, London, WC1N 3AR, UK

Author note: Correspondence concerning this article should be addressed to Elliot Murphy (elliot.murphy@uth.tmc.edu).

Word count: 11,608

Abstract:

Natural language syntax yields an unbounded array of hierarchically structured expressions. We claim that these are used in the service of active inference in accord with the free-energy principle (FEP). While conceptual advances alongside modelling and simulation work have attempted to connect speech segmentation and linguistic communication with the FEP, we extend this program to the underlying computations responsible for generating elementary syntactic objects. We argue that recently proposed principles of economy in language design—such as “minimal search” and “least effort” criteria from theoretical syntax—adhere to the FEP. This permits a greater degree of explanatory power to the FEP—with respect to higher language functions—and presents linguists with a grounding in first principles of notions pertaining to computability. More generally, we explore the possibility of migrating certain topics in linguistics over to the domain of fields that investigate the FEP, such as complex polysemy. We aim to align concerns of linguists with the normative model for organic self-organisation associated with the FEP, marshalling evidence from theoretical linguistics and psycholinguistics to ground core principles of efficient syntactic computation within active inference.

Keywords: Free-energy principle; active inference; language; syntax; minimal search; complexity; minimum message length

1. Introduction

Implementational models of language must be plausible from the perspective of neuroanatomy (Embick & Poeppel 2015), but they must also be plausible from the perspective of how physical, organic systems must behave. We will argue that the structuring influence of the free-energy principle (FEP) from theoretical neuroscience and biology can be detected in language, not only—as has been argued recently—at the *complex* level such as narrative (Bouizegarene et al. 2020), interpersonal dialogue (Friston et al. 2020), cooperative/intentional communication (Vasil et al. 2020) and speech segmentation (Friston et al. 2021), but also at the *elementary* level of what linguists consider to be basic phrase-level computations (Berwick & Stabler 2019, Chomsky 1949, 1951, 1956, 1959, 2021a, 2021b).

Natural language syntax yields an unbounded set of hierarchically structured expressions, and adheres to language-general design principles. Our goal is to unite long-standing theories of linguistic computation with the FEP. We argue that many historical insights into syntax are consistent with the FEP—providing a novel perspective under which the principles governing syntax are not limited to language, but rather reflect domain-general processes that underpin a variety of cognitive computations. This is consistent with a strain within theoretical linguistics that explores how syntactic computation may adhere to “general principles that may well fall within extra-biological natural law, particularly considerations of minimal computation” (Chomsky 2011: 263), such that certain linguistic theories might be engaging with general properties of organic systems that impact language design (Chomsky 2004, 2014). Here, we consider the idea that many aspects of natural language syntax may be special cases of a variational principle of least free-energy.

At this stage, it is useful to consider that while the FEP has a substantial explanatory scope across a large range of behaviours of living organisms and cognitive systems, it should be seen as a methodological heuristic for multi-disciplinary research (Ramstead et al. 2019), in much the same way that the notion of economy is typically entertained in linguistics as a *programmatic* notion (Chomsky 1995). The FEP itself has been argued to be more of a conceptual-mathematical model for self-organizing systems (for some, it is a “generic” model; Barandiaran & Chemero 2009), or a guiding framework. Thus, when we argue that natural language syntax “complies” with the FEP, this is not to imply that the FEP necessarily bears any specific, direct predictions for linguistic behaviour. Rather, it motivates the construction of conceptual arguments for how some property of organic systems might be seen as realizing the FEP.

We begin by summarising the FEP, and describe how syntatic principles are consistent with it. We argue that the FEP allows us to understand how human language complies with the constraints imposed on worldly interactions, deriving certain features of language from first principles. We review key observations from linguistics that speak to the structuring influence of computational efficiency. We consider how the FEP is a variational principle of “least action”, such as those that describe systems with conserved quantities (Coopersmith 2017), and also consider how a relatively recent project in linguistics has suggested that natural language syntax adheres to principles of “least effort” and “minimal search” restrictions (Bošković & Lasnik 2007, Gallego & Martin 2018, Larson 2015). We believe that certain postulates from this field resonate with the FEP and its desiderata. We will also be concerned with the computational and implementational basis of language as viewed as a product of an individual’s mind/brain, following the standard ‘I-language’ (Chomsky 1986, 2000) perspective in generative linguistics (‘internal’, ‘individual’, ‘intensional’). In this connection, we later summarize how the FEP—and its associated process theory—can provide novel insights into syntax, for example by explaining human physiological responses during language processing, such as oscillatory coupling. We finish by highlighting directions for future research.

2. Active Inference and the Free-Energy Principle

Before we evaluate any work pertaining to linguistic behaviour, we need a clear grounding in some foundational concepts. This section introduces key elements of the FEP that motivate its appication to language.

2.1. The Free-Energy Principle

The FEP states that any adaptive change in the brain will minimise free-energy, either over evolutionary time or immediate, perceptual time. Free-energy is a function of sensory data and brain states: it is the upper bound on the ‘surprise’ of sensory data, given predictions that are based on an internal model of how those data were generated. Under simplifying assumptions, free-energy can be considered as the amount of prediction error. The brain can minimise free-energy either by optimising (i.e., inverting) its probabilistic generative model that specifies how hidden states cause sensory data (i.e., inferring the cause of sensory consequences), or by acting on the environment to sample data that are predicted by its model.

The FEP can be considered from the perspective of a Markov blanket, which specifies a statistical boundary between a system's internal states and its external environment. A Markov blanket partitions the world into 'things' which can then be seen as performing a form of inference about external states of affairs in the environment (Pearl 1988). Internal states are conditionally independent of external states, given the blanket (i.e., sensory and active) states. In other words, external states only affect internal states through sensory states, while internal states only change external states through active states. The implicit circular causality is formally identical to the perception-action cycle (Fuster 2004).

Under the FEP, 'surprise' is directly related to the log-evidence for sensory data under an internal model, so by minimising surprise the brain maximises evidence for (self-)existence (Bastos et al. 2012); or, we could turn to Vaas's assessment that the brain is "a self-referential, closed system, a functional reality emulator that constructs the world, rather than reconstruct it" (Vaas 2001: 88). The brain accrues evidence for itself by "explaining away sensory input" (Hohwy 2017: 1).

The difference between free-energy and 'surprise' (or equivalently, the negative log-evidence for sensory data) is the difference (specified by the Kullback-Leibler divergence) between probabilistic representations encoded by the brain and the true conditional distribution of the causes of sensory input. By minimising free-energy, the brain is, technically speaking, performing approximate Bayesian inference.

Variational free-energy (F) is an evaluable quantity, and we can formulate it in several ways, which are mathematically equivalent:

$$\begin{aligned} F &= E_Q[\ln Q(\tilde{s}) - \ln P(\tilde{o} \mid \tilde{s}) - \ln P(\tilde{s})] \\ &= E_Q[\ln Q(\tilde{s}) - \ln P(\tilde{s} \mid \tilde{o}) - \ln P(\tilde{o})] \\ &= \underbrace{D_{KL}[Q(\tilde{s}) \,\|\, P(\tilde{s} \mid \tilde{o})]}_{\textit{relative entropy}} - \underbrace{\ln P(\tilde{o})}_{\textit{log evidence}} \\ &= \underbrace{D_{KL}[Q(\tilde{s}) \,\|\, P(\tilde{s})]}_{\textit{complexity}} - \underbrace{E_Q[\ln P(\tilde{o} \mid \tilde{s})]}_{\textit{accuracy}} \end{aligned} \qquad \text{(Eq. 1)}$$

where $\tilde{s}$ represents the states of the internal model, $\tilde{o} = (o_1, \ldots, o_\tau)$ denotes sensory observations until the current time (τ), P indicates the probability under the internal model, Q indicates posterior beliefs, E_Q indicates the expected value, and D_{KL} denotes a Kullback-Leibler divergence. These equations demonstrate that we can consider free-energy as a trade-off between accuracy and complexity, whereby the best internal model is the one that accurately describes the data in the simplest manner. The complexity part of variational free energy will become important later. In effect, it reflects the degree of belief updating afforded by some new sensory data; in other words, how much some new evidence causes one to “change one’s mind”. A good generative model—with the right kind of priors—will minimise the need for extensive belief updating and thereby minimise complexity. Formally, complexity also appears in treatments of universal computation (Hutter 2006) and in the guise of minimum message or description lengths (Wallace & Dowe 1999). Indeed, in machine learning, variational free energy minimisation has been cast as minimising complexity—or maximising efficiency in this setting (MacKay 1995). One sees that same theme emerge in predictive coding—and related—formulations of free energy minimisation, where the underlying theme is to compress representations; thereby, maximising their efficiency (Barlow 1961, Linsker 1990, Rao & Ballard 1999).

2.2. Active Inference

The enactive component of active inference rests on the assumption that action is biased to realize preferred outcomes. Beliefs about which actions are best to pursue rely on predictions about future outcomes, and the probability of pursuing any particular outcome is proportional to the *expected free energy* if that action is taken. Expected free energy (G) can be expressed as the combination of extrinsic and epistemic value (Friston, Parr et al. 2017):

$$
\begin{aligned}
G(\pi) &= \sum_{\tau} G(\pi, \tau) \\
G(\pi, \tau) &= E_Q[\ln Q(s_\tau \mid \pi) - \ln Q(s_\tau \mid o_\tau, \pi) - \ln P(o_\tau)] \\
&= \underbrace{E_Q[\ln Q(s_\tau \mid \pi) - \ln Q(s_\tau \mid o_\tau, \pi)]}_{\text{(negative) mutual information}} - \underbrace{E_Q[\ln P(o_\tau)]}_{\text{expected log evidence}} \\
&= \underbrace{E_Q[\ln Q(o_\tau \mid \pi) - \ln Q(o_\tau \mid s_\tau, \pi)]}_{\text{(negative) epistemic value}} - \underbrace{E_Q[\ln P(o_\tau)]}_{\text{extrinsic value}}
\end{aligned}
\qquad \text{(Eq. 2)}
$$

where π is the series of actions (i.e., the policy) being persued. Extrinsic value is the expected evidence for the internal model if a particular series of actions is pursued, whereas epistemic value is the expected information gain: in other words, to what extent will a series of actions reduce uncertainty in the internal model?

Notice that the expected versions of relative entropy and log evidence in the free energy (Eq. 1) will now become intrinsic and extrinsic value respectively (Eq. 2). As such, selecting an action to minimise expected free energy reduces expected surprise (i.e., uncertainty) in virtue of maximising the information gain while—at the same time—maximising the expected log evidence; namely, actively self-evidencing (Hohwy 2016).

2.3. Belief Updating and Neuronal Dynamics

Through specifying a process theory that explains neuronal responses during perception and action, neuronal dynamics have previously been cast (Friston, FitzGerald et al. 2017) as a gradient flow on free energy (known as variational free energy in physics: introduced in Feynman 1972; see Hinton & Zemel 1993). That is to say, any neural process can be formulated as a minimisation of the same quantity used in approximate Bayesian inference (Hohwy 2016). The brain seeks to minimise free energy, which is mathematically equivalent to maximising model evidence. In machine learning, this is equivalent to maximisation of an evidence lower bound (ELBO) (Winn & Bishop 2005). This view of neuronal responses can be conceived with respect to Hamilton's principle of least action, whereby action is the path integral of free energy. Although a neural process theory is not critical for linking the FEP to syntactic principles, we will revisit this topic below.

2.4. Previous Applications

Active inference has already been applied to a variety of domains—from simple reflexive actions (Adams et al. 2013) to more complex behaviours, such as the tendency to conform to social norms and abide by shared values in decision making (Constant et al. 2018, 2019). In this section, we summarize some applications that we consider to be relevant to discussions about syntax.

Consider first how emotion recognition can be explained through active inference and Markov blankets (Demekas et al. 2020). While the concept of a Markov blanket, being not an intrinsically *physical* construct but a statistical one, can be applied to anything in nature—rocks, pencils, ice clouds on Mars—it is more readily and usefully applied to systems that actively engage with their environment and that are involved in protecting their integrity and identity: organisms, sub-organismal systems and superorganismal systems (Andrews 2020). Active inference allows the brain to make inferences about emotional states, conceptualising one brain's external states as the internal states of another brain. In this way, both brains are observers and recipients—or recognisers—of emotion, which gives rise to something that can be considered a shared (emotional) narrative between two brains. For our purposes, natural language carries some parallels with emotion: it can simultaneously result from sensation, but also cause subsequent action. The neural basis of emotion is a richly networked phenomenon, spanning subcortex and cortex (Pessoa 2017), and language has been shown to be no different (Copland & Angwin 2019, Murphy et al. 2021, Saravani et al. 2019, Woolnough et al. 2020). But in its mature state, language is also inherently generative and predictive.

Applying active inference relies on finding the right sort of generative model, and many different structures and forms of generative model are possible. Most relevant to the current application, deep temporal models accommodate a nesting of states that unfold at different temporal scales. Since language output is inherently temporal, this leads to a problem of how to map hierarchical structures onto serial outputs (Epstein et al. 1998), and models that are deep in time allow us to deconstruct associated nested structures. The factorization of a generative model is also relevant, because it allows structures that are like-for-like to be substituted into bigger structures. Defining these like-for-like structures was a key foundational goal of modern syntactic theory.

Recently, a deep temporal model for communication was developed based on a simulated conversation between two synthetic subjects, showing that certain behavioural and neurophysiological correlates of communication arise under variational message passing (Friston et al. 2020). The model of syntax here assumes that syntaxes are sequences of states (i.e., words) with a terminal node at the end of every sentence, with each form of syntactic structure being limited to ‘questions’ and ‘answers’ in a game of ‘Twenty Questions’. Yet, the conclusions in this model are in keeping with core assumptions from linguistics concerning the inherently constructive nature of language. Elementary syntactic units, highly robust and conserved across speakers of the same language, provide specific *belief structures* that are used to reduce uncertainty about the world, through rapid and reflexive categorization of events, states, objects and their relations, in compliance with the FEP. Sentential representations can be thought of as structures designed (partially) to consolidate and appropriately frame experiences, and to contextualise and anticipate future experiences. The range of parseable syntactic structures available to comprehenders provides alternate hypotheses that afford parsimonious explanations for sensory data and, as such, preclude overfitting. If the complexities of linguistic stimuli can be efficiently mapped to a small series of regular (and regularised) syntactic formats, this contributes to the brain’s goal of restricting itself to a limited number of characteristic states—and reduces the complexity cost of belief updating. Intuitively, knowing that you can only have said this or that greatly simplifies the inference problem, rendering exchanges between like-minded artefacts more efficient and economic.

By mapping syntactic structures to conceptual systems in a manner adhering to principles of economy (Chomsky et al. 2019), language can be seen as engaging in a series of 'questions' and 'answers' with sensory data (but also non-linguistic mental states) and conspecifics. Only through natural language (Berwick & Stabler 2019) can we generate the full complexity of *wh*-questions, cross-serial dependencies (e.g., in Dutch and Swiss German; Kallmeyer 2010, Shieber 1985), embedded filler-gap dependencies and other such syntactic phenomena, which permit an expansion of what kinds of querying the brain can execute over sampled data. Thus, all of the ways that language-centred aspects of human cognition can be motivated through conformity to the FEP and active inference (e.g., communication and the interpretation of narratives; Bouizegarene et al. 2020) can further be derived from this more elementary focus on syntactic computational complexity and the interface of this component with conceptual and articulatory systems. It is our intention to explore the compatibility of the FEP with an understanding of natural language as a computational system adhering to principles of economy, providing a novel format for—literally—deep thoughts and mutual understanding through shared narratives.

3. Computational Principles and Syntactic Hierarchies

3.1. A System of Discrete Infinity

How can the FEP contribute to our understanding of syntactic computation? Most immediately, it provides additional constraints on what a computational system can be physically realized as. Consider first the three principles in classical recursive function theory which allow functions to compose (Kleene 1952): substitution; primitive recursion; and minimisation. These are all designed in a way that one might think of as computationally efficient: they reuse the output of earlier computations. Substitution replaces the argument of a function with another function (possibly the original one); primitive recursion defines a new function on the basis of a recursive call to itself, bottoming out in a previously defined function (Lobina 2017); minimisation (also termed ‘bounded search’) produces the output of a function with the smallest number of steps (for Piantadosi 2020, human thought is essentially like Church encoding). One might position these restrictions as adhering to general principles of economy.

More broadly, the notion of minimising surprise can be used to ground observations pertaining to language’s propensity to: reduce the search space during syntactic derivations; permit “no tampering” (Chomsky 2008) of objects during syntactic derivations; limit the range of representational resources able to be called upon during any given stage of comprehension (aided by rapid prediction mechanisms; Forseth et al. 2020, Gastaldon et al. 2020). Examining some core principles of recursion, natural language clearly exhibits minimisation, while binary branching of structures (Radford 2016) limits redundant computation, reducing the range of possible computations.

Natural language syntax exhibits *discrete units* which lead to a *discreteness-continuity duality* (the boundary between syntactic categories can be non-distinct)[1]. Syntax is driven by *closeness of computation* (syntactic objects X and Y form a distinct syntactic object, {X, Y}). Its objects are *bounded* (a fixed list, e.g., N, V, Adj, Adv, P, C, T, *n*, *v*, Asp, Cl, Neg, Q, Det) and their hierarchical ordering is based on a specific functional sequence such as C-T-*v*-V (e.g., C is always higher than V; Starke 2004) which imposes direct restrictions on combinatorics (Adger & Svenonius 2011). These objects can be combined in *cycles* (Frampton & Gutmann 1999), which can be extended to form *non-local dependencies*. As we will discuss, these properties are in turn guided by principles of minimal search (an optimal tree search procedure, informed by notions from computer science; Aycock 2020, Ke 2019, Roberts 2019) and least effort (Larson 2015), fulfilling the imperatives of active inference to construct meaningful representations as efficiently as possible (Bouizegarene et al. 2020), contributing to surprise minimisation.

3.2. Compositionality

[1] The use of discrete—as opposed to continuous—states in generative models is an enormously potent way of minimising complexity. For example, if it is sufficient to carve the world (i.e., causes of my sensations) into a small number of hidden states, one can minimise the complexity of belief updating by not redundantly representing all the fine-grained structure within any one state. Similarly, factorisation plays a key role in structuring our hypotheses or expectations that provide the best explanation for sensations. Perhaps the clearest example here is the carving of a sensory world into *what* and *where*. This means one does not have to represent the location of every possible object in one's visual cortical hierarchy, just *where* an object is and *what* an object is—and then integrate these representations to explain visual input (Ungerleider & Haxby 1994).

Recently, certain efficiency principles at the conceptual interface (where syntax interfaces with general conceptualization) have been proposed (Pietroski 2018), such that the 'instructions' that language provides for the brain to build specific meanings are interpreted with notable simplicity and minimal computational effort. Leaving much of the technical details aside, this is ultimately achieved (in Pietroski's model) through M-join (e.g., F(_) + G(_) → F^G(_), which combines *monadic* concepts, like *red* + *boat* or *Saul* + *spoke*) and D-join (e.g., D(_,_) + M(_) → ∃[D(_,_)^M(_)], which merges *dyadic* concepts with monadic concepts, deriving the meaning of *X verb(ed) Y*, or lexical items such as *give*, which selects for agents, themes and recipients). These crucially restrict the number of resources (lexico-semantic features) able to be combined, and do not output any additional representations above and beyond what is inputted. As with models of syntax invoking a simple process of binary set-formation to derive recursion, by restricting the number of computational procedures able to generate semantic structures this model (and others like it; Heim & Kratzer 1998) restricts in highly predictable ways the possible range of outputs (as opposed to models of syntax/semantics invoking a larger range of compositional procedures and combinatorial apparatus).

This model also assumes that there are 'fetchable' concepts that language can use for compositionality, and non-fetchable concepts. For instance, many of the world's languages make use of quantification/numerosity (via a fronto-parietal network, closely linked to major language sites; Nieder 2016), but they make little use of colour (despite colour featuring just as prominently in ordinary experience) possibly due to the remoteness of occipital visual regions. For instance, one might imagine some functional morpheme coding brightness of coloration. Likewise, language seems to make considerable use of certain contentful concepts (e.g., evidentiality) but not others (e.g., worry/sollicitativity) (Adger 2019, Peterson 2016), seemingly making (neurally) efficient use of specific representational resources to map complex (monadic and dyadic) meanings onto natural language expressions.

Consider also how, in neo-Davidsonian event semantics, conjunction is limited to predicates of certain semantic types (Schein 1993, Pietroski 2005). Certain semantic rules of composition (in (1b)) have been claimed to arise directly from more elementary syntactic computations (Pietroski 2018) which adhere to principles of efficient computation.

(1) a. Dutch shot Micah quickly.

b. $\exists$e[Agent(e, Dutch) & Theme(e, Micah) & shot(e) & quickly(e)]

In this connection, it has further been observed that language acts as an artificial context which helps "constrain what representations are recruited and what impact they have on reasoning and inference" (Lupyan & Clark 2015: 283). Words themselves are "highly flexible (and metabolically cheap) sources of priors throughout the neural hierarchy" (Ibid). Meanwhile, González Escribano (2005) discusses the role of economy throughout language, most notably semantics. In their discussion of language evolution, Hauser et al. (2002: 1574) speculate that "at least" syntax involves computational efficiency (for discussion of simplicity in semantic computation, see Al-Mutairi 2014, Collins 2020, Gallego & Martin 2018).

Consider also the phenomenon of *multidimensionality*: "The [young], [happy], [eager], [either going to [Oxford or Cambridge]], [pleasant] man". This involves *unbounded, unstructured coordination* involving disjunction. Such rapid inferences about properties and states in the external world can be generated relatively effortlessly by language.

As some linguists have observed (Boeckx 2014, Hinzen & Sheehan 2013), there is a strong connection between specific levels of syntactic representation and forms of semantic representation. For example, two commonly discussed types of syntactic categories—CPs (Complementizer Phrases) and *v*Ps (“little” Verb Phrases)—provide distinct properties for conceptual interpretation (Chomsky 2004); CPs can be truth-evaluable and provide Tense and event structure along with Force features (i.e., whether the sentence is declarative, interrogative, and so forth), while *v*Ps are also propositional and provide crucial thematic information. The notion of *truth* is deeply wedded to complex syntax (Hinzen 2016, Hinzen & Sheehan 2013), and is a notion that spectacularly increases an agent’s ability to generate environmental inferences.

3.3. Considerations of Economy

The notion of simplicity has been a methological value which has guided linguistic inquiry for decades (Terzian and Corbalan 2020). As such, a number of economy principles have been proposed in theoretical linguistics: the No Tampering Condition (Chomsky 2008), Minimal Link Condition (Chomsky 1995), Last Resort (Chomsky 1995), Relativized Minimality (Rizzi 1990, 2001), Inclusiveness Condition (Chomsky 1995), Precedence Resolution Principle (Epstein et al. 1998), Phase Impenetrability Condition (Chomsky 2004), Full Interpretation (Freidin & Lasnik 2011, Lohndal & Uriagereka 2016), Input Generalisation (Holmberg & Roberts 2014), Maximise

Minimal Means (Biberauer 2019), Resource Restriction (Chomsky et al. 2019), and Equal Embedding (Murphy & Shim 2020).

These have been framed within a linguistic context, often invoking domain-specific notions, despite a core part of the intended project of modern theoretical linguistics being to embed linguistic theory within principles general to cognition. For example, the Inclusiveness Condition maintains that no new elements can be introduced in the course of a particular syntactic derivation, and only existing elements can be rearranged, restricting available computational resources. Motivating these language-specific computational generalizations by direct reference to the FEP may broaden the explanatory scope for the existence and prevalence of particular syntactic phenomena. Since linguists lack a general theory of computational efficiency for language (e.g., Gallego & Chomsky 2020: "To be sure, we do not have a general theory of computational efficiency, but we do have some observations that are pretty obvious and should be part of that theory"), additional support with respect to grounding these concerns within a well-motivated framework for general organic behaviour (i.e., the FEP) will likely prove productive. We can at least suppose that whatever definition will be forthcoming will be related to more generic notions of economy, such as Hamiltonian notions (distance × work), minimising energy expenditure during language processing (Rahman & Kaykobad 2005), shortening description length (Schmidhuber 2015), reducing Kolmogrov complexity (Ming & Vitányi 2008, Wallace & Dowe 1999) and the degree of necessitated belief updating (i.e., maintaining aspects of one's beliefs/generative model is less costly than forcing a revision).

Consider how the No Tampering Condition suggests that the merging of two syntactic objects, X and Y, leaves X and Y unchanged. The set {X, Y} created by MERGE (Chomsky et al. 2019) cannot be broken and no new features can be added[2]. The original structure in (2a) can be modified by the merging of a new element, λ, to form (2b), adhering to the No Tampering Condition, while (2c) violates this condition since the original structure (2a) is modified (Lohndal & Uriagereka 2016). The derivation of syntactic hierarchies appears to generate structures adhering to this principle, i.e., elements merge 'upstairs', as in (2b), while adjuncts that merge 'downstairs' do not alter the structure of the object they adjoin to, i.e., if "John saw Mary [in the park]", then the fact that John saw Mary does not change (subscripts denote syntactic heads/labels, standard script denotes lexical items, where (2a) could represent a sentence like 'The red boat', with α denoting a Determiner Phrase label and γ a Noun Phrase label).

(2) a. $[_{\alpha}\ [\beta\ [_{\gamma}\ [\delta\ \varepsilon]]]]$

b. $[_{\alpha}\ \lambda\ [_{\alpha}\ [\beta\ [_{\gamma}\ [\delta\ \varepsilon]]]]]$

c. $[_{\alpha}\ [\beta\ [_{\gamma}\ [\delta\ [_{\varepsilon}\ \varepsilon\ \lambda]]]]]$

Further, it is more *economical* to expand a structure than to backtrack and modify a structure that has already been built (see Lasnik & Lohndal 2013 for empirical motivations).

[2] MERGE has been defined as an operation on a workspace and its objects, formalized as follows (where WS = workspace; P/Q = workspace objects such as linguistic features; X = additional elements): MERGE(P,Q,WS) = WS' = $\{\{P,Q\},X_1,\ldots,X_n\}$.

Relatedly, a further observation in the literature concerns Relativized Minimality (Rizzi 1990), or the principle stating (in its recent formulation) that given a configuration, [X … Z … Y], “a local relation cannot connect X and Y if Z intervenes and Z fully matches the specification of X and Y in terms of the relevant features” (Starke 2001). In other words, if X and Y attempt to establish a syntactic relation, but some element, Z, can provide a superset of X’s particular features (i.e., X’s features plus additional features), this blocks such a relation[3]. Framing this in terms of structural relations between elements has led to disagreements about how best to frame nodal ‘distance’, but the generalization has proven to carry considerable explanatory power in its various guises across analyses of numerous languages (Rizzi 2001). Relativized Minimality has also been argued to emerge directly from minimal search (Aycock 2020), allowing this higher-level representational principle to emerge directly from properties of efficient computation.

In a similar connection, how a phrase is categorised (as a Noun Phrase, or a Verb Phrase, etc.) has been argued to be determined via a process of minimal search, which searches a structure to find the first (least embedded) categorial feature (e.g., N, V) suitable to a specific search algorithm for identifying the object (e.g., searching

[3] In (1), *which game* provides a superset of the features hosted by *how*, resulting in unacceptability. The equivalent does not obtain in (2), and so a relationship between both copies of *which game* can obtain, licensing interpretation (strikethroughs denote originally merged positions, now deleted after movement; asterisk denotes unacceptability).

(1) *[[How$_{[+Q]}$] [C$_{[+Q]}$ [do you wonder [[which game$_{[+Q, +N]}$] [C$_{[+Q]}$ [PRO to play ~~how~~$_{[+Q]}$]]]]]]

(2) [[Which game$_{[+Q,+N]}$] [C$_{[+Q]}$ [do you wonder [[how$_{+Q}$] [C$_{+Q}$ [PRO to play ~~which game~~$_{[+Q,+N]}$]]]]]]

With respect to minimal search perspectives, consider how when searching for matching features in (2) the search procedure would ‘skip’ *how* but find the original copy of *which game*.

for a type of linguistic feature to serve as a phrase category). This process determines, amongst other things, the distributional properties of a phrase. For instance, in (2c) the phrase category—the “label” or “head” (Frampton & Gutmann 1999)—would be determined by some property of β (being the least embedded, more computationally accessible object), namely its categorial feature (in this case, α), which could be a Determiner Phrase (‘The red boat’).

The principle of Resource Restriction (or ‘Restrict Computational Resources’, RCR; Chomsky 2019, Chomsky et al. 2019) states that when the combinatorial operation MERGE maps workspace *n* to workspace *n*+1, the number of computationally accessible elements (syntactic objects) can only increase by one. That is to say, in the course of constructing a legitimate interpretation, any given syntactic manipulation can only increase the number of ‘options’ for computation by one, and not by, say, three. Crucially, Resource Restriction interacts with other principles, such as minimal search, whereby only those elements that can be searched efficiently ‘count’ towards the number of objects in a given workspace. This principle has been argued to shed light on long-standing puzzles in theoretical syntax (Huybregts 2019, Komachi et al. 2019), and accords with what the FEP would demand of a linguistic computational system.

Another example of economy can be found in the principle of Full Interpretation. This bans superfluous symbols at the two linguistic interfaces: the conceptual and sensorimotor systems. This ensures that the system need not compute symbols that are ultimately superfluous to the goals of either interpretation or externalization. For example, (3) contains an argument that does not have a semantic role (‘theta role’, in semantic theory; Brody 1993) assigned to it, and so no compositional interpretation arises (asterisk denotes unacceptability).

(3) *Walt gave Jesse a gun to Saul.

The underlying operation of MERGE which constructs binary-branching structures, as in (2), can itself derive some set-theoretic properties of linguistic relations, such as *membership*, *dominate* and *term-of*, as well as the derived relation of *c-command* (=sister of) which is relevant for interpreting hierarchical relations between linguistic elements (Haegeman 1994). Exploring this issue further will take us too far afield, so we simply note that much of the complexities of syntactic relations (such as displacement, argument structure, discourse and scopal properties) can be derived from successive instances of this simple MERGE operation and how language-external systems access the structures generated by it. Related phenomena, like adjunction, do not involve modification of the semantic content of the structure the adjunct is concatenated with, as mentioned earlier (Chomsky 2004: 118): "[F]or β to lose some property when α adjoins to it would be a complication, an "imperfection"". These and many other cases suggest that there are properties of language design that exhibit computational efficiency, which often overrule the demands of other sub-systems of language (e.g., Dobashi 2010), a topic we will return to[4].

[4] Note that there are also numerous cases in which principles previously seen as purely syntactic have been re-framed as, for example, phonological in nature (e.g., string adjacency; Bobaljik 2002; see also Richards 2016, Samuels 2011).

Yet, regardless of whether one assumes constructs such as the No Tampering Condition, our point here has been to consider how the FEP can in principle provide a novel explanation for the prevalence of efficiency-encoded structures. We believe our thesis is in line with Fukui's (1996) observation that economy principles in linguistics resemble the principle of least action in physics, pointing to a deeper grounding for the types of linguistic rules we have been discussing. To further stress this point, consider Dasgupta and Gershman's (2021) assessment that mental arithmetic, mental imagery, planning, and probabilistic inference all share a common resource: memory that enables efficient computation. Other domains exhibiting computational efficiency include concept learning (Feldman 2003), causal reasoning (Lombrozo 2016) and sensorimotor learning (Genewein & Braun 2014). This suggests that the ideas we have been evaluating in connection to natural language syntax spill over into other domains of cognition, helping to motivate the more encompassing role of the FEP. Redundant computation can be avoided through storing partial solutions in memory, and structure-dependent phrase-chunking procedures are a clear example of such a solution. This perspective leads to a separation of optimality from language's proposed 'function' of mapping structures to the interfaces, since similar optimising principles are found elsewhere.

On the whole, the FEP has been equated with the principle of least effort, and it is effectively *a computational principle*; the probabilistic beliefs it is concerned with are directed *at* something, namely external states of a self-organising system. Linguists have developed theories of syntactic least effort (efficient search structured via nested hierarchical relations) and its process theory is less clear (on the notion of parsing, see Berwick & Stabler 2019), but we have argued that it may become clearer if it can be accommodated within active inference, given the rich grounding of this framework within biology and neuroscience. The next section will continue to explore these proposals in greater detail.

3.4. Minimising Free-Energy, Maximising Interpretability

The notions of linguistic *computation* and *inference* can be explored in parallel. For instance, the syntactic categories of words are not tagged acoustically, and yet sentential meaning is inferred from syntactic categorization (Adger 2019, Pietroski 2018). Interpreting thematic relations (*who did what to whom*) demands that relations between words are established, however spoken sentences are often ambiguous between distinct syntactic structures. For instance, consider the sentence “We watched a movie with Jim Carrey”. This can be hierarchically structured as (4a) or (4b), depending on whether we interpret Jim Carrey as starring in the movie, or sitting next to us (subscripts denote conventional phrase structures such as Tense Phrase, Noun Phrase, and so forth; Radford 2016).

(4) a. [$_{\text{TP}}$ [$_{\text{NP}}$We] [$_{\text{VP}}$watched [$_{\text{NP}}$a [$_{\text{N}}$movie [$_{\text{PP}}$with [$_{\text{NP}}$Jim Carrey]]]]]]

b. [$_{\text{TP}}$ [$_{\text{NP}}$We] [$_{\text{VP}}$[$_{\text{VP}}$watched [$_{\text{NP}}$a movie]] [$_{\text{PP}}$with [$_{\text{NP}}$Jim Carrey]]]]

Next, consider (5).

(5) Routinely, poems that rhyme evaporate.

In this instance, ‘routinely’ exclusively modifies ‘evaporate’. It cannot modify ‘rhyme’, despite this word being closer (in terms of linear distance) to ‘routinely’, and despite the fact that while poems routinely rhyme, they do not evaporate. The matrix predicate ‘evaporate’ is closer in terms of *structural distance* to ‘routinely’, since the relative clause embeds ‘rhyme’ more deeply (minimal search is partly “defined by least embedding”; Chomsky 2004: 109)[5]. Language computes over structural distance, not linear distance (Berwick et al. 2011, 2013, Friederici et al. 2017, Martin et al. 2020). The use of hierarchical structure to *limit* interpretation adheres to a core tenet of the FEP. While sensorimotor systems naturally impose linear order, linguistic expressions are complex *n*-dimensional objects with hierarchichal relations. Language prioritizes the demands of the syntax-semantics interface over other systems, such as morpho-phonological rules governing linearization (Gärtner & Sauerland 2007, Grohmann 2007, Kosta et al. 2014, Murphy 2016; although see Richards 2016). While structures may exhibit distinct linear orders, they may exhibit identical underlying hierarchical orders, as in (6) (English) and (7) (Basque).

(6)	John	has	read	**the book**
	John	auxiliary	read	the book
(7)	Jonek	**liburua**	irakurri	du
	Jon	book	read	auxiliary

[5] We refer the reader to Ke (2019: 44) and Aycock (2020: 3-6) for a detailed discussion of minimal search, which can be formally defined below, from Aycock (2020), adopting an Iterative Deepening Depth-First Search approach (Korf 1985); where MS = minimal search, SA = search algorithm, SD = search domain (where SA operates), ST = search target:

(1) MS = ⟨SA, SD, ST⟩

(2) SA:

a. Given ST and SD, match against every head member of SD to find ST [initial depth-limit of SD = 1; search depth-first].

b. If ST is found, return the head(s) bearing ST and go to d. Otherwise, go to c.

c. Increase the depth-limit of SD by 1 level; return to a.

d. Terminate Search.

In these examples, the Verb-Direct Object dependencies are the opposite but the interpretation is strictly conserved. This suggests that the syntax encodes the Verb and Direct Object as an abstract phrase which omits the subject; roughly [Subj [V DO]]. Different languages externalize these abstract relations in different ways (which could be considered as modifications to the structure of the generative model that contains the same core components), but the syntax itself does not consider linear order.

Throughout the various stages of language development, infants and children do not typically produce expressions that deviate from general grammatical principles pertaining to the structure-dependence of grammatical rules, even when they produce “mistakes”, suggesting that sensitivity to structure-dependence (and concomitant notions of computational efficiency) forms a core part of language design (Crain et al. 2017) and will be a crucial prior in any generative model. Relatedly, as Piantadosi (2020) reviews, human learners prefer to induce hypotheses that have a shorter description length in logic (Goodman et al. 2008), with simplicity preferences possibly being “a governing principle of cognitive systems” (Piantadosi 2020: 15; see Chater & Vitányi 2003). Simplicity-based preferences anchor a range of formal language models (Erdogon et al. 2015, Tomalin 2006), relating to the notion of minimum description lengths (Grünwald 2007, Kolmogorov 1965); relatedly, we might consult the principles of minimum redundancy and maximum efficiency in perception (Wipf & Rao 2007). Structure-dependence also poses serious difficulties for statistical methods of modelling language acquisition using linear order properties (bi/tri-gram sequences in a corpus), Moreover, the process of developing this linguistic knowledge through infancy and childhood is constrained by efficiency principles, such as the Subset Principle and the Tolerance Principle (Yang et al. 2017), which adhere to economy and use hierarchical knowledge to guide the learning of a range of linguistic sub-components, including morphology and phonology.

Another argument that has been made in recent literature is that linguistic computation is optimised for the generation of interpretable syntactic structures, rather than for the generation of maximally communicative messages to conspecifics. In other words, whenever there is a conflict between principles of computational efficiency and principles of communicative clarity, the former is prioritized (Asoulin 2016, Murphy 2020a). The normal functioning of syntax leads to instances which reduce communicative efficiency. For instance, consider (8).

(8) You persuaded Saul to sell his car.

The individual ('Saul') and the object ('car') can be questioned, but questioning the more deeply embedded object forces the speaker to produce a more complex circumlocution ('[ ]' denotes the originally merged position of the *wh*-expression).

(9) a. *[What] did you persuade who to sell [ ]?

b. [Who] did you persuade [ ] to sell what?

The structures in (9) involve the same words and interpretations, yet the more computationally costly process of searching for and then moving the more deeply embedded element cannot be licensed[6].

[6] Relatedly, it has been suggested that the behaviour of constructions involving optional movement of contrastive foci can be explained through recourse to principles of economy during interpretation, rather than the positing of a particular syntactic feature (i.e., a part of the lexicon) that triggers such behaviour (Titov 2020).

Other examples show that the acceptability of sentences can be impacted based on the extent to which the construction makes a novel, non-redundant contribution to one's mental models/beliefs, rather than those of conspecifics, again reinforcing the role of syntactic processing in inference generation. The degraded acceptability in (10b), relative to (10a), seems to stem from the fact that speakers are unlikely to be ignorant of the relevant content (Paillé & Schwarz 2019) ('?' denotes questionably acceptable).

(10) a. Kim knows whether Saul is in bed.
b. ?Kim knows whether I am in bed.

Even anaphoric relations (e.g., how pronouns are distributed) call upon complex syntactic and semantic processes (Sundaresan 2020; for computational efficiency in anaphora, see Reuland 2011). Other cases such as (11) reveal how superficially simple processes such as contraction are also sensitive to certain hierarchical phrase boundaries (Riemsdijk & Williams 1986).

(11) Saul's taller than [Kim is] / *[Kim's].

This view of language as an "instrument of thought" (Asoulin 2016, Hinzen 2006) is ultimately compatible with the FEP and should be seen as a unique case of surprise minimisation and active inference. In the realm of computational modelling, minimalist grammars typically exhibit a "shortest move/attract" condition for generating sentence meanings, with the usefulness of hierarchical structure for modelling being clear (Hunter et al. 2019, Stabler 1997).

Crucially, this is not to say that when language is used for communication that this process is not also structured via criteria of efficiency; for instance, Gibson et al.

(2019) provide a comprehensive examination of how tools from probability and information theory can explain certain properties of the lexicon, and even syntax (although see Galantucci et al. 2020 for evidence that people often fail to—and, indeed, do not care to—communicate faithfully). Rather, it is to stress that the fundamental unit of natural language—elementary phrase structure building via recursive hierarchical compositionality—is just as susceptible to the organizational influence of the FEP. In addition, one can still explore the role of linguisitic communication in, for instance, the exchange between the sensorium and human actions geared towards changing environmental states, where it seems plain that linguistic communication forms a core part of environment-modifying actions.

Lastly, some of the ideas we have made reference to—such as the I-language perspective through which linguistic communication can be achieved only because two individual's internal grammar sufficiently overlap (i.e., there is no external construct, "English", to which they both independently approximate)—find close analogues in the active inference literature. Consider a model of dialogue through which internal generative models used to infer one's own behaviour are deployed to infer the communicative intentions of another, given both parties have similar generative models (Friston & Frith 2015a). As such, the core notion of surprise might be a crucial adjudicator in framing the distinction between I-language and more social conceptions of language production: Reducing surprise/effort internally serves I-language functions of computational efficiency, but it can also serve to encourage mutual predictability between speakers, while, on the other hand, intentionally inducing surprise in others can often directly serve certain communicative goals (Giorgi & Dal Farra 2019).

In summary, core computational properties of natural language syntax are readily entailed by the FEP. There is increasing evidence supporting the claim that "syntax is organized in such a way as to simplify the computation" of linguistic structures (Frampton & Gutmann 1999: 26). The remaining sections will discuss some specific implications, and possible further applications, of this claim.

4. Further Applications of the FEP to Syntax

4.1. Hidden States

A recent application of active inference concerns the reading of a narrative. Considering hierarchical models with deep temporal structure, active inference has been applied to state transitions, or sequences over time, applying free-energy to sequential scene construction of the kind found in reading a narrative (Friston, Rosch et al. 2017). Equipping an agent with deep temporal models allows them to accumulate evidence over different temporal scales, seeking the best explanation for their sensations. For instance, perisaccadic updating (Marino & Mazer 2016) is associated here with neurobiological responses such as delay period activity and perisaccadic local field potentials, accounting for how epistemically rich information is judiciously sampled by the eyes. Under this framework, sequences of hierarchically nested hidden states (e.g., the location and category of an object, or some other unknown variable about the world) are generated by the brain using probability transitions specified by a particular policy (action or plan), such that hidden states can influence expected free energy and thus influence policies determining transitions among subordinate states. Agents infer the hidden states under each policy that they entertain before evaluating the evidence for each policy based on observed outcomes and beliefs. Posterior beliefs about policies are used to inform a Bayesian model average of the next outcome, realized via action.

Relating this back to earlier ideas, Markov blankets may be applicable to sub-components of the language system and their interfaces/boundaries. For instance, interfaces between syntax and sensorimotor systems can readily be seen as blanket states. The blanket between systems is fundamentally an achievement, not a given, being maintained via active inference (Ramstead et al. 2019). The possiblity of nested blankets has also been entertained, whereby, for example, an integrated blanket could comprise the brain's sensory epithelia and motor/autonomic efferents, while the internally nested Markov blankets would be a feature of cortical hierarchies (Shipp 2016). Cortical hierarchies would presumably play a role in the blanketing of cognitive sub-systems. The FEP licenses an interpretation of internal states (e.g., neurons inside a brain) as instantiating probabilistic beliefs about external states (e.g., properties of the external world). As such, internal states parameterize a probability density over external states and can consequently be interpreted in terms of a basic form of inference. Natural language syntax remains 'shielded' by a number of Markov blankets, only being able to uniquely interface directly with conceptual and articulatory formulation systems[7]. By opening up possibilities with respect to planning and inference, natural language syntax and its capacity for structural complexity allows

[7] These ideas may relate to how the hominization of the brain appears to have increased cortical network depth, thereby enhancing the abstraction capacities of humans (Changeux et al. 2020).

individuals to expand the scope of their predictions concerning future positions in state space: the topic we now turn to.

4.2. Reference to Abstractions

Syntactic structures can be framed as contributing to policies used to perform particular free-energy minimising actions. The rapid and reflexive identification of objects, states and events in the external world through simple linguistic means (even at early stages of childhood; Lightfoot 2020) can yield complex, flexible interpretations for some of the most common words (Chomsky 2000, Collins 2017, Gotham 2016, Murphy 2021, Pustejovksy & Batiukova 2019, Vicente 2015), aiding in the successful generation of internal models of the environment using a limited number of resources. Complex forms of polysemy generated via multi-word constructions, as in (12), allow for a more precise localization in conceptual space than discrete symbols/signs and gestures, with syntax allowing the generation of unique unveilings of hidden states in the world (subscripts denote polysemous sense types denoted by predicates: α = INFORMATION, β = PHYSICAL, γ = ORGANIZATION).

(12) The poorly [written]$_\alpha$ newspaper that I [held]$_\beta$ this morning has been [sued]$_\gamma$ by the government.

Since there cannot possibly be any object in the external world that a complex polysemous word like *newspaper* can index a one-to-one mapping with, under the framework we are developing here lexical items could be seen as hypotheses about the structure of likely co-occurring sensory input, and about ontological and mereological relations between objects and states (Arapinis 2013, Arapinis & Vieu

2015). From the perspective of active inference, things only exist as a label for a hypothesis or inference about hidden states. Our contention is that the forms of complex meaning derived from natural language semantics form a core component of this labelling. Simple lexical items like *book*, *table* and *walk* are effectively hypotheses composed from distinct core knowledge systems (e.g., geometry, place, social relations; Dillon et al. 2013, Spelke 2016, Spelke & Kinzler 2007, Strictland 2017) which can elucidate environmental regularities essential to active inference.

By permitting a refined positioning in conceptual space, natural language syntax aids agents in the formation of novel policies to navigate, and make inferences about, the environment (Yon et al. 2019). Cognition is an ongoing process of dynamic interaction between an organism and its environmental niche (Hutto & Myin 2013), yet notions like *event* are also not pre-defined, external entities but are actively generated (and parcellated by the language system). In active inference terms, the examples we have provided show the language system carving a continuous world into discrete state spaces in a way that minimises complexity.

There is precedence in recent literature for these suggestions. The FEP may be compatible with an instrumentalist theory of mental representations, via which representations are useful fictions for explanatory goals (Ramstead et al. 2020, 2021). This is compatible with certain models in philosophy of language (Collins 2017, Pietroski 2018), which assume that lexical items have no one-to-one direct referent in the external world, but are composites of distinct representational domains that are used for successful, efficient interpretation. It is also compatible with internalist models of Markov blankets, which Hohwy (2017) argues are forms of neo-Kantian and Helmholtzian accounts of cognition which emphasize the interactive, constructive nature of higher cognition in generating interpretable, actionable concepts of external hidden states (Anderson 2017).

As such, the long-term storage of conceptually rich lexical items, alongside the combinatorial rules underlying their creative deployment in language production and comprehension, allows speakers to categorize novel sensory data into a discrete set of objecthood and eventhood representations, contributing to the avoidance of surprising states.

4.3. Ideal and the Real: Aligning Grammatical Knowledge with Performance

A related issue concerns how readily language comprehension and production map onto active inference assumptions about perceptual sampling, policy formation and action. We would here like to suggest that what has come to be known as the *One-System Hypothesis* in psycholinguistics accords well with the FEP, since the basic claim is that the computational system that parses linguistic stimuli into hierarchically structured phrases is identical to the system that conceptualizes, formulates and articulates linear externalizations (via speech, sign, etc.) out of internal hierarchical representations (Bianchi & Chesi 2014, Momma & Phillips 2018). The mapping between priors and posteriors would presumably be optimized, and the recruitment of a single system serves to minimise surprise. Lewis and Phillips (2015) argue that grammatical theories and language processing models describe the same cognitive system, as evidenced by the fact that syntax-parser misalignments only appear to occur as a consequence of domain-general system limitations (e.g., memory access, cognitive control) (Sprouse & Almeida 2013). This provides the simplest possible format for language-related belief updating, recruiting a single system for action and perception.

From an implementational perspective, the categorization of phrase structures can be seen as the process of accessing phrasal nodes in memory by using syntactic category as a retrieval cue, with syntactic category independently impacting the speed of parsing (Trueswell & Kim 1998) and production (Melinger & Dobel 2005). More suggestively, there is evidence that syntactic aspects of a sentence are planned prior to lexical access and the 'grain size' of units in language production planning is smaller than a clause (Momma & Phillips 2018). Syntax thus renders meaning-making and higher-order inference a computationally efficient process.

Since the range of sentence structures is unbounded, it is likely that the most parsimonious strategy would be for the language system to avoid predictions relating to optional syntactic elements (such as adjuncts) and only generate predictions related to core structural elements (such as verbs and their arguments). Martorell (2018) reviews empirical evidence for this, such as the finding that arguments lead to faster reading times than adjuncts (Tutunjian & Boland 2008). These findings dovetail into prioritising the unveiling of hidden states and the reduction of variational free energy, which we can also ground within the terms of predictive coding (Martin 2020, Morillon et al. 2009), at least insofar as predictive coding can yield amortized elements of the language system's full computational complexity. Active inference assumes an imperative to find the most accurate explanation for sensory observations that is minimally complex—recruited in Barlow's (2001) exploration of minimum redundancy—and which seems in accord with how the language system provides the most computationally efficient format for solving the problem of mapping linear sensory linguistic input to hierarchical syntactic representations (Hauser & Watumull 2017).

How else might this issue of linguistic prediction be related to active inference and the FEP? Much work in psycholinguistics (e.g., exploring the processing of filler-gap dependencies leading to 'hyperactive gap-filling', island constructions, verb predictions; Berwick & Stabler 2019, Chow et al. 2016, Omaki et al. 2015) shows that phrase structures are generated predictively, in anticipation of upcoming stimuli (see also Gwilliams & Davis 2020, Reuters et al. 2020). Even something as simple as adjective-noun syntax (e.g., "red boat") is constructed predictively (Berwick & Stabler 2019). Finding evidence that some feature of cognition is concerned with surprise avoidance at once suggests that such a feature will comply with active inference (Da Costa et al. 2020; see also Armeni et al. 2018, Di Liberto et al. 2018, Keshev & Meltzer-Asscher 2021, Wang et al. 2018 and much other work for evidence for predictive processing in language, supporting the idea that prediction is a "canonical computation" implemented in domain-specific circuits; Keller & Mrsic-Flogel 2018).

Lastly, when exploring the issue of subgoaling—found, for example, when monkeys solve planning problems—Maisto et al. (2015) conclude that favoured subgoals are ones that permit planning solutions and controlling behaviour using less information resources (yielding parsimony in inference and control), and conclude that of those string that represent procedures returning the same output, strings with lower descriptive complexity are more probable, and should more likely be selected by probabilistic inference. They argue that this framing is compatible with the FEP, and we have further argued here that these assumptions of computational efficiency in subgoals can be found in syntactic combinatorics in natural language (e.g., Momma &

Phillips 2018 discuss linguistic parsing and production in terms of discrete subgoals). We find seeds for these ideas in the foundational principles of universal computation, where free energy is often discussed in terms of minimum description or message lengths (MacKay 2003, Schmidhuber 2010).

One of the conclusions we wish to draw from this literature concerns the apparent existence of principles of computational efficiency in the generation of syntactic structures and the mapping of these structures to the conceptual interfaces. From the perspective of active inference, individuals need to minimise the effort involved in meaning-making. We have proposed that there is increasing evidence from theoretical linguistics, but also psycholinguistics, that natural language syntax exhibits design principles in keeping with least effort criteria, both during comprehension and production.

4.4. Epistemic Foraging

In the active inference literature, epistemic foraging concerns the use of beliefs about hidden states to prescribe active sampling of new information to resolve uncertainty quickly. We would like to extend these considerations and suggest that the existence of syntactic structures in natural language can contribute to a unique form of epistemic foraging, through maximising model evidence and minimising surprise and variational free energy. For instance, that the highly restricted cartography of syntactic projections achieves this goal (Abels 2015, Cinque 2005, Martin et al. 2020). We assume that, during comprehension, language users select the phrase structure that is least surprising (and least costly to construct) from the perspective of their hierarchical generative model, an internal 'action' that selects among competing hypotheses (syntactic nodal categories and nestings) for the most likely causes of sensory stimuli. In support of this view, predictive processes have been found not just in sensorimotor systems (Bourguignon et al. 2020, Hovsepyan et al. 2020, Poeppel & Assaneo 2020), but also in the conceptual-intentional system (García & Ibáñez 2016). Crucially, these predictive processes are apparent once syntax interfaces with either system (Zaccarella et al. 2021), with the generative faculty of syntax being purely a reflex of laws of elementary computation.

Phrase structure processing requires a syntactic memory buffer (Berwick & Stabler 2019) alongside more local lexical memory (Davis et al. 2002). Insofar as the brain is assumed to work towards inferring the hidden causes of sensory states, the structuring apparatus for novel sensory stimuli afforded by phrase structures (permitting "complex meaning" going beyond monadic lexico-semantic concepts denoted by single words; Hagoort 2020) allows the generation of predictions concerning relations between entities, states and events. Even seemingly simple sentences like 'The second blue ball on the left', referring to a bundled collection of objects, can only be generated through hierarchical, recursive phrase structures (Adger 2019). Indeed, generating complex, structured recursive thoughts *expands* the range of possible hidden states, such that there are an increased number of states that the world could look like to the organism. Possibilities for uncertainty resolution rapidly expand with complex syntax.

Lastly, and pertaining more directly to existing work on active inference and linguistic communication (e.g., Vasil et al. 2020), it is also notable that the recursive combinatorial apparatus of natural language syntax has been argued to facilitate recursive theory of mind (‘I know that you know that Mary thinks that …’) (Arslan et al. 2017, Jamali et al. 2021), and as such could in turn be seen as deriving some active inference-based properties of higher-order cognition, unveiling the latent or hidden states that are other people’s mental states. Epistemic foraging aims to reduce uncertainty about the environment, and this is precisely what syntactic phrases do (Hippisley 2015).

4.5. Endogenous Synchronicity

In this section, we will briefly outline some intermediate levels of description which might negotiate the lower-level implementation of FEP-pointing computations.

Just as how any neural model of active inference is subject to the dictates of physiological and anatomical constraints, so too have recent models of language attempted to ground syntactic computation in endogenous oscillatory synchronicity (Ding & Simon 2014, Ding et al. 2016, Kaufeld et al. 2020). Parallel to these developments, in neighbouring fields, the FEP has shed light on a broad range of empirical findings, such as the hierarchical arrangement of cortical areas (Williams 2018), functional asymmetries between forward-backward connections in the brain, and long-latency (endogenous) components of evoked cortical responses, and it is our contention that one can derive some elementary properties of linguistic computation through a direct line of communication from the FEP, endogenous oscillatory synchronicity and linguistic behaviour. Under the FEP, endogenous oscillations are the type of dynamics that neurons would expect to encounter, since they have genetically encoded ‘beliefs’ that the cause of excitatory post-synaptic potentials follows a certain pattern. Active inference can synthetize various in-silico neurophysiological responses via a gradient descent on free energy (Friston, FitzGerald et al. 2017), such as the mismatch negativity, phase-precession, theta sequences, place-cell activity and theta-gamma coupling.

Moving forward with these concerns, neuronal dynamics and plasticity appear to minimise variational free energy under a simple generative model, which entails prior beliefs that presynaptic inputs are generated by an external state with a quasi-periodic orbit (Palacios et al. 2019). The implication is that ensembles of neurons make inferences about each other while individual neurons minimise their own free energy. Generalized synchrony is an emergent property of free energy minimisation; desynchronization is induced by exogenous input (explaining event-related desynchronization); and structure learning emerges in response to causal structure in exogenous input, explaining the functional segregation of neuronal clusters (Palacios et al. 2019). An external (i.e., neuron-external) state with a quasi-periodic orbit is assumed to generate the presynaptic inputs of a given neuron. Low-frequency phase synchronization emerges directly from this assumption, and also the coupled assumption that neuronal dynamics maximize variational free energy under a simple generative model.

One implication of this research avenue is that models of syntactic computation grounded in these dynamics (e.g., Giraud 2020, Kaufeld et al. 2020, Tavano et al. 2021) can be said to comply with foundational principles of the FEP (see also Guevara Erra et al. 2017). For instance, endogenous low-frequency (delta) tracking of syntactic nodes (Kaufeld et al. 2020) could be seen as emerging as a direct function of generative belief updating in accord with active inference, supplementing the association of delta oscillations with the cortical computations responsible for creating hierarchical linguistic structures. The active inference framework provides clear predictions about the neural dynamics of language, and can help bring together research programs that are presently pursued independently.

In keeping with this framework, oscillations are thought to provide the most energy-efficient physical mechanism for synchrony and temporal coordination (Mirollo & Strogatz 1990; see also Bergmann & Born 2017). By grounding principles of minimal computation from language into particular oscillatory mechanisms, one might position natural language syntax as being able to be fully accommodated by the FEP. Yet unlike properties of speech, natural language syntax has to be wholly inferred, and it has been suggested that intrinsic oscillatory synchronicity indexes such inferential processing (Meyer et al. 2020). We suggest that active inference can help elucidate the general schema through which the brain generates these inferences. Consider how the question of whether visually-evoked neuronal responses are caused by external events or by saccadic eye-movements has been discussed in the context of Markov blankets (Mirza et al. 2016)—do internal states predict external causes of sensory states, or actively cause them? Some recent debates in the language sciences concern the nature of neuronal entrainment versus endogenous synchronicity (Giraud 2020, Meyer et al. 2020) in response to linguistic stimuli. We see this issue as being readily able to be captured via the active inference framework, whereby endogenous oscillatory synchrony uses sensory states to infer linguistic representations.

Exploring the possible neurobiological basis of a core feature of language, Hovsepyan et al. (2020) argue that theta-gamma phase-amplitude coupling in language (coding syllable recognition) and predictive coding can be brought together (Martin 2020). This theta-gamma coupling has been applied to syllable parsing, modelling theta-gamma coupling for dialogue, which appears to form part of belief updating for active inference (whereby beliefs are simultaneously updated at high, ‘fast’ levels and also lower, ‘slow’ levels, depending on the representational level of interest) (Friston et al. 2020). In this model, phrase-level inferences generate words contained within the phrase before lower levels reset for the next phrase, manifested as theta phase-alignment (i.e., each transition at the higher level is accompanied by a resetting of lower-level states). This is in line with the suggestion that low frequency phase can coordinate the bundling of lexical features (indexed by fast gamma cycles) within a given structure, ensuring serial read-out of features, alongside the transfer of syntactic categorial identities to language-external systems (Murphy 2020b).

Theta-gamma coupling has also formed part of recent models of scheduling/updating the list of syntactic-semantic features being associated with a given ‘chunk’ of linguistic stimuli, with gamma cycles indexing distinct data structures being coordinated by theta phase (Benítez-Burraco & Murphy 2016, 2019, Murphy 2018, 2020b, Murphy & Benítez-Burraco 2016, 2017, Wilkinson & Murphy 2016; for linguistic details, see Adger & Svenonius 2011, Embick 2021). We see these proposals as potentially analogous from a neurocomputational perspective. That is to say, theta-coordinated serial representation association may form the basic mechanistic unit for interpretation out of which belief updating can emerge. Alternatively, these codes may index separable lower-level processes that merely converge at the level of cross-frequency coupling, since theta-gamma coupling has been attributed a number of distinct roles.

We stress that these issues, though promising, are not directly related to either the success of theoretical syntax or independent advances in active inference research. Our central thesis has no direct stake in whether neural oscillations index certain processes pertaining to computational efficiency outlined here, and it could well be that investigating other avenues will yield more promising outcomes, and we suspect that other research groups might prioritize different neurobiological mechanisms from those we have highlighted here.

4.6. Creativity

Lastly, active inference has been used to account for *creative functions* that have to do with exploration and novelty (Parr & Friston 2019). The hallmark of natural language is its creative aspect, or the ability to freely construct an unbounded array of hierarchically organised phrase structures, with novel interpretations. Linguistic creativity can be framed in terms of an adherence to physical, thermodynamic conservativity if it serves to minimise uncertainty and unveil hidden states within an individual's model of external states.

5. Future Work

The linguistic phenomena we have briefly surveyed here constitutes a fragment of current theorizing about computational simplicity in syntax, and indeed we are particularly interested in the extent to which other phenomena can be grounded within the FEP and the ways that other researchers would consider their areas of linguistic research compatible (or not) with the active inference framework. Given that language is a highly complex system, far from monolithic, with different sub-components arising at distinct evolutionary stages through distinct pressures and extra-linguistic interfaces (Fitch 2010, Murphy 2019), it may transpire that certain sub-components cannot easily be couched within active inference.

We have arrived at a number of suggestive (and preliminary) explanations for the way language implements the construction of hierarchical syntactic objects: to minimise uncertainty about the causes of sensory data; to unveil a species-unique format of external hidden states; to adhere to a "least effort" natural law when composing sets of linguistic features for interpretation, planning, consolidation and, in some cases, externalization. Exploring our proposal empirically may demand a more mature development of the science of computational complexity in the brain (Singer et al. 2019); for a similar roadblock in physics, see Conlon (2016: 107).

Plainly, there are many issues with the framework we have outlined here that need to be further unpacked and clarified before the explanatory power of the FEP with respect to natural language syntax can be evaluated further. As mentioned, simulation work on the issue of endogenous neuronal responses to hierarchically organized linguistic phrases could contribute to this goal. Currently, our proposals are part of what van Rooj and Baggio (2021) term the *theoretical cycle*, with the ultimate goal being to progress into the *empirical cycle*: Theories "can be assessed for their plausibility on formal, computational grounds", before qualifying for empirical treatment; we hope that our proposals will be critically engaged in service of this goal.

Overall, we hope to have provided a new perspective on how to think about principles of economy in models of natural language syntax, with broader implications scaling up to the level of higher-order conceptual interpretations. That is to say, if the

FEP can derive core features of syntactic combinatorics, then whatever structures are ultimately generated should also be driven by goals pertaining to active inference. A challenge for future research is to appropriately frame the interfacing of syntax with other systems in terms that accord with the minimisation of surprise and variational free energy. Since the FEP has attendant process theories (e.g., active inference), one of the latent pay offs of our suggestions here is the development of generative models of active inference that fully ground specific factors in syntactic theory (such as phrase boundary sensitivity, structure-dependent rules) and, through simulation work, may align with recent advances in the electrophysiology and neural dynamics and harmonics of syntax (Brennan & Martin 2020, Kaufeld et al. 2020, Keitel et al. 2017, Tavano et al. 2021). For example, one crucial factor in any syntax model seems to be phrasal category information since it appears to drive cortical tracking of hierarchical structures (Burroughs et al. 2021), in line with assumptions and predictions in work highlighting the unique computational contribution of phrase structure labeling (Adger 2013, Hornstein 2009).

6. Conclusions

We have reviewed how the FEP, that underwrites active inference, is an expression of the principle of least action, which is additionally a principle implemented in models of theoretical syntax. Ultimately, both the FEP and syntactic theory are empirically and conceptually well-supported constructs, and they share a number of intriguing commonalities. The FEP provides suggestions for *why* the brain computes the way it does, while theoretical linguistics provides a means of establishing *what* is computed. An intuition from 70 years ago—that the extent to which a model of grammar is simple seems to determine its explanatory power—echoes in the modern framing of syntactic computation as a system of economy: “[T]he motives behind the demand for economy are in many ways the same as those behind the demand that there be a system at all” (Chomsky 1951; see also Goodman 1951). This appeal to simplicity may even emerge from an evolved bias shared across our species to favour simple solutions, (see the ‘cognitive simplicity hypothesis’; Terzian and Corbalán 2020). The FEP has produced formal, simulation-supported models of complex cognitive mechanisms such as action, perception, learning, attention and communication, while theories of syntax embracing computational efficiency have led to empirically successful outcomes, explaining grammaticality intuitions (Adger 2003, Martin et al. 2020, Sprouse 2011, Sprouse & Almeida 2017), certain poverty of stimulus issues (Berwick et al. 2011, Crain et al. 2017, Wexler 2003, Yang et al. 2017) and the pervasive organizational role that hierarchy has in language (Friederici et al. 2017, Grimaldi 2012).

Active inference and the implicit Bayesian brain hypothesis that the FEP subsumes (that the brain is an inference generator or Helmholtz machine; Knill & Pouget 2004) concern the divergence between the inferences drawn by the brain and the hidden surroundings in the external world. The connection between the FEP and language has, we hope, become clearer in the light of our claim that the unique format for thought provided by language permits individuals to unveil their own constructed hidden states: Complex conceptual relations pertaining to who-did-what-to-whom are maximally and most readily available with syntax, which imposes upon sensory data a particular interpretation in order to then act “as if” such relations exist to better maximise the evidence for internal generative models. The notion of uniqueness is important here. As Zaccarella et al. (2021: 9) stress, while human motor planning and execution exhibits certain hierarchies and looped subroutines, the computation of action does not adhere to the principles governing linguistic recursive structures. Natural language syntax appears to contribute a computationally inimitable tool for guiding transactions between cognitive modules.

A core objective of current theories of syntax is to explain *why* language design is the way it is (Adger 2019, Narita 2014), and we have suggested that the FEP can contribute to this goal. The more efficiently a language user can internally construct meaningful, hierarchically organized syntactic structures, the more readily they can use these structures to contribute to the planning and organization of action, reflection and consolidation of experience, exogenous and endogenous monitoring, imagination of possible states, and adaptive environmental sampling.

More broadly, what the FEP can offer theoretical linguistics is proof of principle: a foundational grounding and means of additional motivation for investigating language in terms of efficient computation. The FEP is fundamentally a normative model (Allen 2018) which can, but does not have to, aid the generation of implementational/process models (such as predictive coding or active inference) and can place constraints on feasibility. Further simulation and modeling work is required to push these ideas further for natural language and its various sub-systems, and we envisage that this type of future work will provide fruitful insights into natural language syntax.

References

Abels, K. (2015). Syntax. Braber, N., Cummings, L., & Morrish, L. (Eds). *Exploring Language and Linguistics*. Cambridge: Cambridge University Press. 137–167.

Adams, R.A., Shipp, S., & Friston, K.J. (2013). Predictions not commands: active inference in the motor system. *Brain Structure & Function* 218(3): 611–643.

Adger, D. (2003). *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press.

Adger, D. (2013). *A Syntax of Substance*. Cambridge, MA: MIT Press.

Adger, D. (2019). Language Unlimited: The Science Behind Our Most Creative Power. Oxford: Oxford University Press.

Adger, D., & Svenonius, P. (2011). Features in minimalist syntax. Boeckx, C. (Ed). *The Oxford Handbook of Linguistic Minimalism*. Oxford: Oxford University Press. 27–51.

Allen, M. (2018). The foundation: mechanism, prediction, and falsication in Bayesian enactivism. Comment on Answering Schrödinger's question: a free-energy formulation, by Maxwell James Desormeau Ramstead et al. *Physics of Life Reviews* 24: 17–20.

Anderson, M. (2017). Of Bayes and bullets: an embodied, situated, targeting-based account of predictive processing. Metzinger, T., & Wiese, W. (Eds). *Philosophy and Predictive Processing* 2. Frankfurt am Main: MIND Group. 1–14.

Al-Mutairi, F.R. (2014). The Minimalist Program: The Nature and Plausibility of Chomsky's Biolinguistics. Cambridge: Cambridge University Press.

Andrews, M. (2020). The math is not the territory: navigating the free energy principle. http://philsci-archive.pitt.edu/18315.

Arapinis A. (2013). Referring to institutional entities: semantic and ontological perspectives. *Applied Ontology* 8: 31–57.

Arapinis, A., & Vieu, L. (2015). A plea for complex categories in ontologies. *Applied Ontology* 10: 285–296.

Armeni, K., Willems, R.M., van den Bosch, A., & Schoffelen, J.-M. (2018). Frequency-specific brain dynamics related to prediction during language comprehension. *NeuroImage* 198: 283–295.

Arslan, B., Hohenberger, A., & Verbrugge, R. (2017). Syntactic recursion facilitates and working memory predicts recursive theory of mind. *PLoS ONE* 12(1): e0169510.

Asoulin, E. (2016). Language as an instrument of thought. *Glossa: A Journal of General Linguistics* 1(1): 46. 1–23.

Aycock, S. (2020). A third-factor account of locality: explaining impenetrability and intervention effects with minimal search. *Cambridge Occasional Papers in Linguistics* 12(1): 1–30.

Barandiaran, X.E., & Chemero, A. (2009). Animats in the modeling ecosystem. *Adaptive Behavior* 17(4): 287–292.

Barlow, H. (1961). Possible principles underlying the transformations of sensory messages. In W. Rosenblith (Ed.), *Sensory Communication*. Cambridge, MA: MIT Press. 217-234.

Barlow, H. (2001). Redundancy reduction revisited. *Computation and Neural Systems* 12: 241–253.

Bastos, A. M., Martin Usrey, W., Adams, R. A., Mangun, G. R., Fries, P., Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76: 695–711.

Benítez-Burraco, A., & Murphy, E. (2016). The oscillopathic nature of language deficits in autism: from genes to language evolution. Frontiers in Human Neuroscience 10: 120.

Benítez-Burraco, A., & Murphy, E. (2019). Why brain oscillations are improving our understanding of language. *Frontiers in Behavioral Neuroscience* 13: 190.

Bergmann, T.O., & Born, J. (2017). Phase-amplitude coupling: a general mechanism for memory processing and synaptic plasticity? *Neuron* 97: 10–13.

Berwick, R.C., Friederici, A.D., Chomsky, N., & Bolhius, J.J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Sciences* 17(2): 89–98.

Berwick, R.C., Pietroski, P., Yankama, B., & Chomsky, N. (2011). Poverty of the stimulus revisited. *Cognitive Science* 35(7): 1207–1242.

Berwick, R.C., & Stabler, E.P. (Eds). (2019). *Minimalist Parsing*. Oxford: Oxford University Press.

Bianchi, V., & Chesi, C. (2014). Subject islands, reconstruction, and the flow of the computation. *Linguistic Inquiry* 45: 525–569.

Biberauer, T. (2019). Factors 2 and 3: towards a principled approach. *Catalan Journal of Linguistics Special Issue*. 45–88.

Bobaljik, J. (2002). A-chains at the PF-interface: copies and covert movement. *Natural Language & Linguistic Theory* 20(2): 197–267.

Boeckx, C. (2014). Elementary Syntactic Structures: Prospects of a Feature-Free Syntax. Cambridge: Cambridge Univeristy Press.

Bošković, Ž., & Lasnik, H. (Eds). (2007). *Minimalist Syntax: The Essential Readings*. Oxford: Blackwell.

Bouizegarene, N., Ramstead, M., Constant, A., Friston, K., & Kirmayer, L. (2020). Narrative as active inference. *PsyArXiv.* doi:10.31234/osf.io/47ub6.

Bourguignon, M., Molinaro, N., Lizarazu, M., Taulu, S., Jousmäki, V., Lallier, M., Carreiras, M., & Tiège, X.D. (2020). Neocortical activity tracks the hierarchical linguistic structures of self-produced speech during reading aloud. *NeuroImage* 216: 116788.

Brennan, J.R., & Martin, A.E. (2020). Phase synchronization varies systematically with linguistic structure composition. *Philosophical Transactions of the Royal Society B* 375(1791): 20190305.

Brody, M. (1993). Theta-theory and arguments. *Linguistic Inquiry* 24: 1–23.

Burroughs, A., Kazanina, N., & Houghton, C. (2021). Grammatical category and the neural processing of phrases. *Scientific Reports* 11: 2446.

Changeux, J-P., Goulas, A., & Hilgetag, C.C. (2020). A connectomic hypothesis for the hominization of the brain. *Cerebral Cortex* bhaa365.

Chater, N., & Vitányi, P. (2003). Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences* 7(1): 19–22.

Chomsky N. (1949). Morphophonemics of Modern Hebrew. Undergraduate Honors Thesis, University of Pennsylvania, Philadelphia.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory* 2: 113–124.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control* 2: 137-167.

Chomsky N. (1951/1979). *Morphophonemics of Modern Hebrew*. New York: Garland.

Chomsky, N. (1986). Knowledge of Language: Its Nature, Origin, and Use. Westport, CT: Praeger.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, N. (2004). Beyond explanatory adequacy. Belletti, A. (Ed). *Structures and Beyond*. Oxford: Oxford University Press. 104–131.

Chomsky, N. (2008). On phases. Freidin, R., Otero, C.P., & Zubizaretta, M.L. (Eds). *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud.* Cambridge, MA: MIT Press. 133–166.

Chomsky, N. (2011). Language and other cognitive systems. What is special about language? *Language, Learning and Development?* 7(4): 263–278.

Chomsky, N. (2014). Minimal recursion: exploring the prospects. Roeper, T., & Speas, M. (Eds). Studies in Theoretical Psycholinguistics 43. *Recursion: Complexity in Cognition*. London: Springer. 1–15.

Chomsky, N. (2019). Some puzzling foundational issues: the Reading Program. *Catalan Journal of Linguistics Special Issue* 263-285.

Chomsky, N. (2021a). Linguistics then and now: some personal reflections. *Annual Review of Linguistics* 7: 16. 1–11.

Chomsky N. (2021b). Simplicity and the form of grammars. *Journal of Language Modelling* 9(1).

Chomsky, N., Gallego, Á.J., & Ott, D. (2019). Generative grammar and the faculty of language: insights, questions, and challenges. *Catalan Journal of Linguistics Special Issue* 226–261.

Chow, W.Y., Smith, C., Lau, E., & Phillips, C. (2013). A “bag-of-arguments” mechanism for initial verb predictions. *Language, Cognition and Neuroscience* 31(5): 577–596.

Cinque, G. (2005). Deriving Greenberg’s Universal 20 and its exceptions. *Linguistic Inquiry* 36(3): 315–332.

Collins, J. (2017). The copredication argument. *Inquiry* 60(7): 675–702.

Collins, J. (2020). Conjoining meanings without losing our heads. *Mind & Language* 35: 224–236.

Conlon, J. (2016). *Why String Theory?* Boca Raton, FL: CRC Press.

Constant, A., Ramstead, M.J.D., Veissiere, S.P.L., Campbell, J.O., & Friston, K. J. (2018). A variational approach to niche construction. *Journal of the Royal Society Interface* 15: 20170685.

Constant, A., Ramstead, M.J., Veissière, S.P., & Friston, K. (2019). Regimes of expectations: an active inference model of social conformity and decision making. *Frontiers in Psychology* 10: 679.

Coopersmith, J. (2017). The Lazy Universe: An Introduction to the Principle of Least Action. Oxford: Oxford University Press.

Copland, D.A., & Angwin, A.J. (2019). Subcortical contributions to language. Zubicaray, G.I., & Schiller, N.O. (Eds). *The Oxford Handbook of Neurolinguistics*. Oxford: Oxford University Press. 850–876.

Crain, S., Koring, L., & Thornton, R. (2017). Language acquisition from a biolinguistic perspective. *Neuroscience and Biobehavioral Reviews* 81(B): 120–149.

Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., & Friston, K. (2020). Active inference on discrete state-spaces: a synthesis. *Journal of Mathematical Psychology* 99: 102447. 1–24.

Dasgupta, I., & Gershman, S.J. (2021). Memory as a computational resource. *Trends in Cognitive Sciences* 25(3): 240–251.

Davis, M.H., Marslen-Wilson, W.D., & Gaskell, M.G. (2002). Leading up the lexical garden path: segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 28 (1): 218–244.

Demekas, D., Parr, T., & Friston, K.J. (2020). An investigation of the free energy principle for emotion recognition. *Frontiers in Computational Neuroscience* 14: 30.

Di Liberto, G.M., Lalor, E.C., & Millman, R.E. (2018). Causal cortical dynamics of a predictive enhancement of speech intelligibility. *NeuroImage* 166: 247–258.

Dillon, M.R., Huang, Y., & Spelke, E.S. (2013). Core foundations of abstract geometry. *Proceedings of the National Academy of Sciences* 110: 14191–14195.

Ding, N., & Simon, J.Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Frontiers in Human Neuroscience* 8: 311.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience* 19: 158-164.

Dobashi, N. (2010). Computational efficiency in the syntax–phonology interface. *The Linguistic Review* 27(3): 241–260.

Embick, D. (2021). The motivation for roots in Distributed Morphology. *Annual Reviews of Linguistics* 7: 8. 1–20.

Epstein, S.D., Groat, E., Kawashima, R., & Kitahara, H. (1998). *A Derivational Approach to Syntactic Relations*. Oxford: Oxford University Press.

Erdogan, G., Yildirim, I., & Jacobs, R.A. (2015). From sensory signals to modality-independent conceptual representations: a probabilistic language of thought approach. *PLoS Computer Biology* 11(11): e1004610.

Evans, D.J., & Searles, D.J. (1994). Equilibrium microstates which generate second law violating steady states. *Physical Review E* 50: 1645–1648.

Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science* 12(6): 227–232.

Feynman, R.P. (1972). *Statistical Mechanics*. Reading, MA: Benjamin.

Fitch, W.T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.

Forseth, K.J., Hickok, G., Rollo, P.S., & Tandon, N. (2020). Language prediction mechanisms in human auditory cortex. *Nature Communications* 11: 5240.

Frampton, J., & Gutmann, S. (1999). Cyclic computation, a computationally efficient minimalist syntax. *Syntax* 2(1): 1–27.

Friederici, A.D., Chomsky, N., Berwick, R.C., Moro, A., & Bolhuis, J.J. (2017). Language, mind and brain. *Nature Human Behaviour* 1: 713–722.

Freidin, R., & Lasnik, H. (2011). Some roots of minimalism. Boeckx, C. (Ed). *The Oxford Handbook of Linguistic Minimalism*. Oxford: Oxford University Press. 1–26.

Friston, K. (2013). Life as we know it. *Journal of the Royal Society Interface* 10: 20130475.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: a process theory. *Neural Computation* 29(1): 1–49.

Friston, K.J., & Frith, C.D. (2015a). A duet for one. *Consciousness and Cognition* 36: 390–405.

Friston, K.J., & Frith, C.D. (2015b). Active inference, communication and hermeneutics. *Cortex* 68: 129–143.

Friston, K.J., Parr, T., & de Vries, B. (2017). The graphical brain: belief propagation and active inference. *Network Neuroscience* 1(4): 381–414.

Friston, K.J., Parr, T., Yufik, Y., Sajid, N., Price, C.J., & Holmes, E. (2020). Generative models, linguistic communication and active inference. *Neuroscience and Biobehavioral Reviews* 118: 42–64.

Friston, K.J., Rosch, R., Parr, T., Price, C., & Bowman, H. (2017). Deep temporal models and active inference. *Neuroscience and Biobehavioral Reviews* 77: 388–402.

Friston, K.J., Sajid, N., Quiroga-Martinez, D.R., Parr, T., Price, C.J., & Holmes, E. (2021). Active listening. *Hearing Research* 399: 107998.

Fukui, N. (1996). On the nature of economy in language. *Cognitive Studies* 3: 51–71.

Fuster, J.M. (2004). Upper processing stages of the perception-action cycle. *Trends in Cognitive Sciences* 8(4): 143-145.

Galantucci, B., Langsteim, B., Spivack, E., & Paley, N. (2020). Repair avoidance: when faithful informational exchanges don't matter that much. *Cognitive Science* 44(10): e12882.

Gallego, Á.J., & Chomsky, N. (2020). The faculty of language: a biological object, a window into the mind, and a bridge across disciplines. *Revista de la Sociedad Española de Lingüística* 50(1): 7–34.

Gallego, Á.J., & Martin, R. (Eds). (2018). *Language, Syntax and the Natural Sciences*. Cambridge: Cambridge University Press.

García, A.M., & Ibáñez, A. (2016). A touch with words: dynamic synergies between manual actions and language. *Neuroscience & Biobehavioral Reviews* 68: 59–95.

Gärtner, H.-M., & Sauerland, U. (Eds). (2007). Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics. Berlin: De Gruyter Mouton.

Gastaldon, S., Arcara, G., Navarrete, E., & Peressotti, F. (2020). Commonalities in alpha and beta neural desynchronizations during prediction in language comprehension and production. *Cortex* 133: 328–345.

Genewein, T., & Braun, D.A. (2014). Occam's razor in sensorimotor learning. *Proceedings of the Royal Society B* 281(1783): 2013–2952.

Gibson, E., Futrell, R., Piantadosi, S.P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences* 23(5): 389–407.

Giraud, A-L. (2020). Oscillations for all? A commentary on Meyer, Sun & Martin (2020). *Language, Cognition and Neuroscience* 35(9): 1106–1113.

Giraud, A-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience* 15(4): 511–517.

Giorgi, A., & Dal Farra, C. (2019). On the syntax/pragmatics interface: expressing surprise and disapproval. *Intercultural Pragmatics* 16(3): 335–361.

González Escribano, J.L. (2005). Semantocentric minimalist grammar. *Atlantis* 27(2): 57–74.

Goodman N. (1951). *The Structure of Appearance*. Cambridge, MA: Harvard Univ. Press.

Goodman, N., Tenenbaum, J., Feldman, J., & Griffiths, T. (2008). A rational analysis of rule-based concept learning. *Cognitive Science* 32(1): 108–154.

Gotham, M. (2016). Composing criteria of individuation in copredication. *Journal of Semantics* 34(2): 333–371.

Grimaldi, M. (2012). Toward a neural theory of language: old issues and new perspectives. *Journal of Neurolinguistics* 25(5): 304–327.

Grohmann, K. (Ed). (2007). InterPhases: Phase-Theoretic Investigations of Linguistic Interfaces. Oxford: Oxford University Press.

Grünwald, P.D. (2007). *The Minimum Description Length Principle*. New York: MIT Press.

Guevara Erra, R., Perez Velazquez, J.L., & Rosenblum, M. (2017). Neural synchronization from the perspective of non-linear dynamics. *Frontiers in Computational Neuroscience* 11: 98.

Gwilliams, L., & Davis, M. (2020). Extracting language content from speech sounds: an information theoretic approach. *The Auditory Cognitive Neuroscience of Speech Perception*. hal-03013496.

Haegeman, L. (1994). *Introduction to Government and Binding Theory*. 2nd edition. Oxford: Blackwell.

Hagoort, P. (2020). The meaning-making mechanism(s) behind the eyes and between the ears. *Philosophical Transactions of the Royal Society B* 375: 20190301.

Hauser, M., 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.

Hauser, M., & Watumull, J. (2017). The universal generative faculty: the source of our expressive power in language, mathematics, morality, and music. *Journal of Neurolinguistics* 43(B): 78–94.

Heim, I., & Kratzer, A., (1998). *Semantics in Generative Grammar*. Oxford: Blackwell.

Hinton, G.E., & Zemel, R.S. (1993). Autoencoders, minimum description length and Helmholtz free energy. *Proceedings of the 6th International Conference on Neural Information Processing Systems.* Denver, Colorado: Morgan Kaufmann Publishers Inc.

Hinzen, W. (2006). *Mind Design and Minimal Syntax*. Oxford: Oxford University Press.

Hinzen, W. (2016). On the grammar of referential dependence. *Studies in Logic, Grammar and Rhetoric* 46(1): 11–33.

Hinzen, W., & Sheehan, M. (2013). *The Philosophy of Universal Grammar*. Oxford: Oxford University Press.

Hippisley, A. (2015). The word as a universal category. Taylor, J.R. (Eds). *The Oxford Handbook of the Word*. Oxford: Oxford University Press. 246–269.

Hohwy, J. (2016). The self-evidencing brain. *Noûs* 50: 259–285.

Hohwy, J. (2017). How to entrain your evil demon. Metzinger, T., & Wiese, W. (Eds). *Philosophy and Predictive Processing* 2. Frankfurt am Main: MIND Group. 1–15.

Hohwy, J. (2020). New directions in predictive processing. *Mind & Language* 35(2): 209–223.

Holmberg, A., & Roberts, I. (2014). Parameters and the three factors of language design. Picallo, C. (Ed). *Linguistic Variation in the Minimalist Framework*. Oxford: Oxford University Press. 61–81.

Hornstein, N. (2009). *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge: Cambridge University Press.

Hovsepyan, S., Olasagasti, I., & Giraud, A-L. (2020). Combining predictive coding and neural oscillations enables online syllable recognition in natural speech. *Nature Communications* 11: 3117.

Hunter, T., Stanojević, M., & Stabler, E.P. (2019). The active-filler strategy in a move-eager left-corner minimalist grammar parser. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics (CMCL)*. Association for Computational Linguistics, Minneapolis, Minnesota. 1–10.

Hutter, M. (2006). *Universal Artificial Intellegence: Sequential Decisions Based on Algorithmic Probability*. Dordrecht: Springer-Verlag Berlin and Heidelberg & Co. KG.

Hutto, D.D., & Myin, E. (2013). *Radicalizing Enactivism: Basic Minds without Content*. Cambridge, MA: MIT Press.

Huybregts, M.A.C. (2019). Infinite generation of language unreachable from a stepwise approach. *Frontiers in Psychology* 10: 425.

Jamali, M., Grannan, B.L., Fedorenko, E., Saxe, R., Báez-Mendoza, R., & Williams, Z.M. (2021). Single-neuronal predictions of others' beliefs in humans. *Nature* DOI:10.1038/s41586-021-03184-0.

Kallmeyer, L. (2010). *Parsing Beyond Context-Free Grammars*. London: Springer-Verlag Berlin and Heidelberg.

Kaufeld, G., Bosker, H.R., Oever, S., Alday, P.M., Meyer, A.S., & Martin, A.E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *Journal of Neuroscience* 40(49): 9467–9475.

Ke, H. (2019). *The Syntax, Semantics and Processing of Agreement and Binding Grammatical Illusions*. PhD dissertation. University of Michigan.

Keitel, A., Ince, R.A.A., Gross, J., & Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *NeuroImage* 147: 32–42.

Keller, G.B., & Mrsic-Flogel, T.D. (2018). Predictive processing: a canonical cortical computation. *Neuron* 100(2): 424–435.

Keshev, M., & Meltzer-Asscher, A. (2021). Noisy is better than rare: comprehenders compromise subject-verb agreement to form more probable linguistic structures. *Cognitive Psychology* 124: 101359.

Kleene, S.C. (1952). *Introduction to Metamathematics*. New York: Van Nostrand.

Kolmogorov, A.N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission* 1: 1–7.

Komachi, M., Kitahara, H., Uchibori, A., & Takita, K. (2019). Generative procedure revisited. *Reports of the Keio Institute of Cultural and Linguistic Studies* 50: 269–283.

Korf, R. (1985). Depth-first iterative-deepening: an optimal admissable tree search. *Artificial Intelligence* 27(1): 97–109.

Kosta, P., Schürcks, L., Franks, S., & Radev-Bork, T. (Eds). (2014). *Minimalism and Beyond: Radicalizing the Interfaces*. Amsterdam: John Benjamins.

Knill, D.C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences* 27(12): 712–719.

Larson, B. (2015). Minimal search as a restriction on merge. *Lingua* 156: 57–69.

Lasnik, H, & Lohndal, T. (2013). Brief overview of the history of generative syntax. den Dikken, M. (Ed). *The Cambridge Handbook of Generative Syntax*. Cambridge: Cambridge University Press. 26–60.

Lewis, S., & Phillips, C. (2015). Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research* 44: 27–46.

Lightfoot, D. (2020). Born to Parse: How Children Select Their Languages. Cambridge, MA: MIT Press.

Linsker, R. (1990). Perceptual neural organization: some approaches based on network models and information theory. *Annual Review of Neuroscience* 13: 257–281.

Lobina, D.J. (2017). Recursion: A Computational Investigation into the Representation and Processing of Language. Oxford: Oxford University Press.

Lohndal, T., & Uriagereka, J. (2016). Third-factor explanations and universal grammar. Roberts, I. (Ed). *The Oxford Handbook of Universal Grammar*. Oxford: Oxford University Press. 114–128.

Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences* 20(10): 748–759.

Lupyan, G., & Clark, A. (2015). Words and the world: predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science* 24(4): 279–284.

MacKay, D.J. (1995). Free-energy minimisation algorithm for decoding and cryptoanalysis. *Electronics Letters* 31: 445–447.

MacKay, D.J. (2003). *Information Theory, Inference, and Learning Algorithms.* Cambridge: Cambridge University Press.

Maisto, D., Donnarumma, F., & Pezzulo, G. (2015). Divide et impera: subgoaling reduces the complexity of probabilistic inference and problem solving. *Journal of the Royal Society Interface* 12(104): 20141335.

Marino, A.C., & Mazer, J.A. (2016). Perisaccadic updating of visual representations and attentional states: linking behavior and neurophysiology. *Frontiers in Systems Neuroscience* 10: 3.

Martin, A., Holtz, A., Abels, K., Adger, D., & Culbertson, J. (2020). Experimental evidence for the influence of structure and meaning on linear order in the noun phrase. *Glossa: A Journal of General Linguistics* 5(1): 97. 1–21.

Martin, A.E. (2020). A compositional neural architecture for language. Journal of Cognitive Neuroscience 38: 1–20.

Martorell, J. (2018). Merging generative linguistics and psycholinguistics. *Frontiers in Psychology* 9: 2283.

Melinger, A., & Dobel, C. (2005). Lexically-driven syntactic priming. *Cognition* 98: B11–20.

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *European Journal of Neuroscience* 48(7): 2609–2621.

Meyer, L., Sun, Y., & Martin, A.E. (2020). Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience* 35(9): 1089–1099.

Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *Trends in Cognitive Sciences* 11(4): 143–152.

Ming, L., & Vitányi, P. (2008). *An Introduction to Kolmogorov Complexity and Its Applications*. 3rd ed. New York: Springer.

Mirollo, R.E., Strogatz, S.H. (1990). Synchronization of pulse-coupled biological oscillators. *SIAM Journal on Applied Mathematics* 50(6): 1645–1662.

Mirza, M.B., Adams, R.A., Mathys, C.D., & Friston, K.J. (2016). Scene construction, visual foraging, and active inference. *Frontiers in Computational Neuroscience* 10: 56.

Momma, S., & Phillips, C. (2018). The relationship between parsing and generation. *Annual Review of Linguistics* 4: 233–254.

Morillon, B., Kell, C.A., & Giraud, A.L. (2009). Three stages and four neural systems in time estimation. *Journal of Neuroscience* 29: 14803–14811.

Murphy, E. (2016). Phasal eliminativism, anti-lexicalism, and the status of the unarticulated. *Biolinguistics* 10: 21–50.

Murphy, E. (2018). Interfaces (travelling oscillations) + recursion (delta-theta code) = language. Luef, E. & Manuela, M. (Eds). *The Talking Species: Perspectives on the Evolutionary, Neuronal and Cultural Foundations of Language*. Graz: Unipress Graz Verlag. 251–269.

Murphy, E. (2019). No country for Oldowan men: emerging factors in language evolution. *Frontiers in Psychology* 10: 1448.

Murphy, E. (2020a). Language design and communicative competence: the minimalist perspective. *Glossa: A Journal of General Linguistics* 5(1): 2.

Murphy, E. (2020b). *The Oscillatory Nature of Language*. Cambridge: Cambridge University Press.

Murphy, E. (2021). *Linguistic Representation and Processing of Copredication*. PhD dissertation. University College London.

Murphy, E., & Benítez-Burraco, A. (2016). Bridging the gap between genes and language deficits in schizophrenia: an oscillopathic approach. *Frontiers in Human Neuroscience* 10: 422.

Murphy, E., & Benítez-Burraco, A. (2017). Language deficits in schizophrenia and autism as related oscillatory connectomopathies: an evolutionary account. *Neuroscience and Biobehavioral Reviews* 83: 742–764.

Murphy, E., Hoshi, K., & Benítez-Burraco, A. (2021). Subcortical syntax: reconsidering the neural dynamics of language. *PsyArXiv* doi:10.31234/osf.io/29cjw.

Murphy, E., & Shim, J-Y. (2020). Copy invisibility and (non-)categorial labeling. *Linguistic Research* 37(2): 187–215.

Narita, H. (2014). Endocentric Structuring of Projection-free Syntax. Amsterdam: John Benjamins.

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience* 17: 366–382.

Omaki, A., Lau, E.F., Davidson White, I., Dakan, M.L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology* 6: 384.

Paillé, M., & Schwarz, B. (2019). Knowing *whether* and ignorance inferences. *Proceedings of the 36th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project. 306–315.

Palacios, E.R., Isomura, T., Parr, T., & Friston, K. (2019). The emergence of synchrony in networks of mutually inferring neurons. *Scientific Reports* 9: 6412.

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Francisco, CA: Morgan Kaufmann.

Pessoa, L. (2017). A network model of the emotional brain. *Trends in Cognitive Sciences* 21(5): 357–371.

Peterson, T. (2016). Mirativity as surprise: evidentiality, information, and deixis. *Journal of Psycholinguistic Research* 45(6): 1327–1357.

Piantadosi, S.T. (2020). The computational origin of representation. *Minds and Machines* DOI: 10.1007/s11023-020-09540-9.

Pietroski, P. (2005). *Events and Semantic Architecture*. Oxford: Oxford University Press.

Pietroski, P. (2018). *Conjoining Meanings: Semantics Without Truth Values.* Oxford: Oxford University Press.

Poeppel, D., & Assaneo, M.F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience* 21(6): 322–334.

Pulvermüller, F. (2014). The syntax of action. *Trends in Cognitive Sciences* 18(5): 219–220.

Pustejovsky, J., & Batiukova, O. (2019). *The Lexicon*. Cambridge: Cambridge University Press.

Radford, A. (2016). *Analysing English Sentences*. Cambridge: Cambridge University Press.

Rahman, M.S., & Kaykobad, M. (2005). On Hamiltonian cycles and Hamiltonian paths. *Information Processing Letters* 94: 37–41.

Ramstead, M.J., Kirchhoff, M.D., Constant, A., & Friston, K.J. (2019). Multiscale integration: beyond internalism and externalism. *Synthese* DOI: 10.1007/s11229-019-02115-x.

Ramstead, M.J.D., Friston, K.J., & Hipólito, I. (2020). Is the free-energy principle a formal theory of semantics? From variational density dynamics to neural and phenotypic representations. *Entropy* 22: 889. 1–30.

Ramstead, M.J.D., Hesp, C., Tschantz, A., Smith, R., Constant, A., & Friston, K. (2021). Neural and phenotypic representation under the free-energy principle. *Neuroscience and Biobehavioral Reviews* 120: 109–122.

Rao, R.P., & Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience* 2(1): 79–87.

Reuland, E. (2011). *Anaphora and Language Design*. Cambridge, MA: MIT Press.

Reuter, T., Dalawella, K., & Lew-Williams, C. (2020). Adults and children predict in complex and variable referential contexts. *Language, Cognition and Neuroscience* DOI: 10.1080/23273798.2020.1839665.

Richards, N. (2016). *Contiguity Theory*. Cambridge, MA: MIT Press.

Riemsdijk, H. van, & Williams, E. (1986). *Introduction to the Theory of Grammar*. Cambridge, MA: MIT Press.

Rizzi, L. (1990). *Relativized Minimality*. Cambridge, MA: MIT Press.

Rizzi, L. (2001). Relativized minimality effects. Baltin, M., & Collins, C. (Eds). *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell. 89–110.

Roberts, I. (2019). *Parameter Hierarchies and Universal Grammar*. Oxford: Oxford University Press.

Samuels, B.D. (2011). *Phonological Architecture: A Biolinguistic Perspective*. Oxford: Oxford University Press.

Saravani, G.A., Forseth, K.J., Tandon, N., & Pitkow, X. (2019). Dynamic brain interactions during picture naming. *eNeuro* 6(4).

Schein, B. (1993). *Plurals and Events*. Cambridge, MA: MIT Press.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990-2010). *IEEE Transactions on Autonomous Mental Development* 2(3): 230–247.

Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Networks* 61: 85–117.

Shieber, S. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy* 8(3): 333–343.

Shipp, S. (2016). Neural elements for predictive coding. *Frontiers in Psychology* 7: 1792.

Singer, W., Sejnowski, T.J., & Rakic, P. (2019). *The Neocortex*. Cambridge, MA: MIT Press.

Spelke, E.S. (2016). Core knowledge and conceptual change: a perspective on social cognition. Barner, D., & Baron, A.S. (Eds). *Core Knowledge and Conceptual Change*. Oxford: Oxford University Press. 279–300.

Spelke, E.S., & Kinzler, K.D. (2007). Core knowledge. *Developmental Science* 10: 89–96.

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43: 155–167.

Sprouse, J., & Almeida, D. (2013). The empirical status of data in syntax: a reply to Gibson and Fedorenko. *Language and Cognitive Processes* 28(3): 222–228.

Sprouse, J., & Almeida, D. (2017). Setting the empirical record straight: acceptability judgments appear to be reliable, robust, and replicable. *Behavioral and Brain Sciences* 40: e311.

Stabler, E.P. (1997). Derivational minimalism. Retoré, C. (Ed). *Logical Aspects of Computational Linguistics*. LNCS 1328. New York: Springer-Verlag. 68–95.

Starke, M. (2001). *Move Dissolves into Merge*. PhD thesis. University of Geneva.

Starke, M. (2004). On the inexistence of specifiers and the nature of heads. Belletti, A. (Ed). *The Cartography of Syntactic Structures, Vol. 3: Structures and Beyond.* Oxford: Oxford University Press. 252–268.

Strictland, B. (2017). Language reflects “core” cognition: a new theory about the origin of cross-linguistic regularities. *Cognitive Science* 41(1): 70–101.

Sundaresan, S. (2020). Distinct featural classes of anaphor in an enriched person system. Hartmann, K., Mursell, J., & Smith, P.W. (Eds). *Agree to Agree: Agreement in the Minimalist Programme*. Open Generative Syntax series. Berlin: Language Science Press. 425–461.

Tavano, A., Blohm, S., Knoop, C.A., Muralikrishnan, R., Scharinger, M., Wagner, V., Thiele, D., Ghitza, O., Ding, N., Menninghaus, W., & Poeppel, D. (2021). Neural harmonics of syntactic structure. *bioRxiv* DOI:10.1101/2020.04.08.031575.

Terzian, G., & Corbalán, M.I. (2020). Simplicity of what? A case study from generative linguistics. *Synthese* DOI: 10.1007/s11229-020-02649-5.

Titov, E. (2020). Optionality of movement. *Syntax* 23(4): 347–374.

Tomalin, M. (2006). Linguistics and the Formal Sciences: The Origins of Generative Grammar. Cambridge: Cambridge University Press.

Trueswell, J.C., & Kim, A.E. (1998). How to prune a garden path by nipping it in the bud: fast priming of verb argument structure. Journal of Memory and Language 39(1): 102–123.

Tutunjian, D., & Boland, J.E. (2008). Do we need a distinction between arguments and adjuncts? Evidence from psycholinguistic studies of comprehension. *Language and Linguistics Compass* 2: 631–646.

Ungerleider, L.G., & Haxby, J.V. (1994). ‘What’ and ‘where’ in the human brain. *Current Opinion in Neurobiology* 4(2): 157–165.

Vaas, R. (2001). It binds, therefore I am! Review of Rodolfo Llinás’s *I of the Vortex*. *Journal of Consciousness Studies* 8(4): 85–88.

van Rooj, I., & Baggio, G. (2021). Theory before the test: How to build high-versimilitude explanatory theories in psychological science. *Perspectives on Psychological Science* DOI:10.1177/1745691620970604.

Vasil, J., Badcock, P.B., Constant, A., Friston, K., & Ramstead, M.J.D. (2020). A world unto itself: human communication as active inference. *Frontiers in Psychology* 11: 417.

Vicente, A. (2015). The green leaves and the expert: polysemy and truth-conditional variability. *Lingua* 157: 54–65.

Wallace, C.S., & Dowe, D.L. (1999). Minimum message length and Kolmogorov complexity. *Computer Journal* 42(4): 270–283.

Wang, L., Hagoort, P., & Jensen, O. (2018). Language prediction is reflected by coupling between frontal gamma and posterior alpha oscillations. *Journal of Cognitive Neuroscience* 30(3): 432–447.

Wexler, K. (2003). Lenneberg's dream: learning normal language development and specific language impairment. Schaffer, J., & Levy, Y. (Eds.). *Language Competence across Populations: Towards a Definition of Specific Language Impairment*. London: Lawrence Erlbaum. 11–61.

Wilkinson, C., & Murphy, E. (2016). Joint interventions in autism spectrum disorder: relating oscillopathies and syntactic deficits. *UCL Working Papers in Linguistics* 28: 1–7.

Williams, D. (2018). Predictive processing and the representation wars. *Minds and Machines* 28(1): 141–172.

Winn, J., & Bishop, C.M. (2005). Variational messge passing. *Journal of Machine Learning Research* 6: 661–694.

Wipf, D.P., & Rao, B.D. (2007). An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Transactions on Signal Processing* 55(7): 3704–3716.

Woolnough, O.W., Donos, C., Rollo, P.S., Forseth, K.J., Lakretz, Y., Crone, N.E., Fischer-Baum, S., Dehaene, S., & Tandon, N. (2020). Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *Nature Human Behaviour* https://doi.org/10.1038/s41562-020-00982-w.

Yang, C., Crain, S., Berwick, R.C., Chomsky, N., & Bolhuis, J.J. (2017). The growth of language: universal grammar, experience, and principles of computation. *Neuroscience and Biobehavioral Reviews* 81(B): 103–119.

Yon, D., de Lange, F.P., & Press, C. (2019). The predictive brain as a stubborn scientist. *Trends in Cognitive Sciences* 23(1): 6–8.

Zaccarella, E., Papitto, G., & Friederici, A.D. (2021). Language and action in Broca's area: computational differentiation and cortical segregation. *Brain and Cognition* 147: 105651. 1–13.