Modelling non-specific linguistic variation in cognitive disorders

WOLFRAM HINZEN
MARTINA WILTSCHKO
ICREA & UPF

Abstract. Linguistic theory has long sought to capture neurotypical linguistic variation through a model of biologically fixed ‘principles and parameters’ that limit diversity, and it has tended to limit investigation of clinical linguistic diversity to disorders specific to language. However, non-specific linguistic variation, where cognition changes beyond its linguistic dimensions, is abundant in clinical populations. We ask how to capture the latter type of variation through a theoretical linguistic model, which can predict possible and impossible changes in our linguistic phenotype and how language and cognition will covary in their patterns of development and decline. The model we articulate here depicts linguistic cognition as resting on two pre-linguistic pillars: i) perceptual object categorization and ii) social-communicative interaction. Grammar viewed as a bridge crossing between them mediates the lexicalization of categories and, based on this, the forms of social interaction conveying thought based on grammar. To illustrate this model, we make a case that it provides a long-needed fit with known linguistic variation along the autism spectrum.

Keywords: clinical linguistic diversity, linguistic variation, autism spectrum disorder
1. INTRODUCTION

Linguistic theory has long conceptualized linguistic variation in the frame of limits to variation set by the biologically fixed, universal ‘principles and parameters’ of our language faculty (Chomsky, 1981). However, variation is of the essence in any biological trait, and our capacity for language itself could be subject to systematic, biologically conditioned, within-species variation. Numerous imperfections affect our genetic code, giving rise to diversity in all domains of mental and physical functioning. This variation is uncontroversial for the case of our cognitive phenotype, as the prevalence of highly heritable common mental disorders in all human populations suggests. All of these involve fundamental changes in this phenotype in its human-specific aspects, whether in early development, young adulthood, or old age (Manolio et al., 2009; Brainstorm Consortium, 2018). It would thus be a surprising and unexpected finding if language, a central component of cognition, would not be subject to pathological change as well. Here we present and model evidence for this conclusion, that what is true of cognition generally is true of its linguistic part: both are subject to change, and they significantly travel together in how they deviate in neurodevelopmental and neurodegenerative disorders.

Countless neurodevelopmental disorders, such as autism spectrum conditions (ASC), Cri du Chat, Coffin Siris, Rett, Angelman, Fragile-X, Landau-Kleffner, Down’s, Williams, and Phelan-McDermid syndromes, not only imply fundamental changes in how cognition develops, but also in language. Any visit to a large, mixed special-education school reveals a bewildering variety of such linguistic phenotypes, within and across disorders such as the above. The language profiles in question may include non- or minimally verbal children with ASC, now estimated to make up 25-30% of school-age children on the autism spectrum (Norrelgen et al., 2015; Tager-Flusberg & Kasari, 2013).
Language is also widely affected both structurally and functionally in the verbal remainder of the ‘low-functioning’ part ASC (Boucher, 2012), showing such deviances as restrictions to single phrases, verbal stereotypes, or idiosyncratic word choice and meanings, and pervasive deficits of comprehension. All of the other neurogenetic syndromes mentioned above can involve absence, regressions and atypicalities of language development, beyond merely delays, often at a severe end of the spectrum of language dysfunction, and in most cases incomparable to that seen in children with Specific Language Impairment (SLI) / Developmental Language Disorder (DLD) attending normal schools.

The picture will not change with any visit to a large adult psychiatric or neurological ward, where language phenotypes will range from the fluent but formally disorganised and unintelligible speech of patients with the clinical symptom of formal thought disorder (McKenna & Oh, 2005), to the patterns of hallucinated speech in other patients on the schizophrenia spectrum (Tovar et al., 2019), the empty speech and the eventual loss of a language capacity in advanced stages of Alzheimer’s disease, or the syntactically disorganized speech of patients with Huntington’s disease (Hinzen et al., 2016). Faced with this massive onslaught of linguistic variation, and with language affected in so numerous ways and across many dimensions, we are facing an impasse: there are no models that can capture it, allowing us to map out the variation we see. Within linguistics, traditional ‘principles and parameters’ models will not serve: designed to model linguistic variation in neurotypical brains, they have largely abstracted from non-linguistic cognition and cognitive variation. In line with this, insofar as clinical linguistic variation has been mapped onto a theoretical model of language, the key disorders investigated have been SLI/DLD, which by its traditional conception leaves ‘cognition’
intact, and acquired post-stroke aphasia, which is equally ‘specific’ to language by definition. In line with this, current neurological models of language originated in aphasiology, although it was Wernicke himself who urged the expansion of aphasiological models to neuropsychiatric disorders.¹

The need for integrated models of language and cognition has come to the fore even in ‘specific’ language disorders. Thus, in acquired aphasia, general cognitive processing limitations may mask the integrity of core linguistic competence (Bates et al., 1997; Kolk, 1998; Grillo, 2009; Mirman & Britt, 2014). Such integrity can also be indicated by the possibility of enhancing linguistic performance through prosthetic devices (Linebarger et al., 2007). A case has also long been made for an impact of aphasia on cognition at large (Baldo et al., 2005; Baldo et al., 2010; Fonseca et al., 2016), casting doubt on the idea that language and cognition are separable in this condition. In turn, in SLI, there is evidence that the language impairments in question are not, in fact, specific to language, and could at least in part reflect general cognitive deficits in auditory processing, working memory, procedural learning, processing speed, or motor developmental delays (Bishop, 2010; Leonard et al., 2007; Miller et al., 2001; Kohnert & Windsor, 2004; Tsimpli et al., 2017; Schaeffer, 2018). Together, this spectrum of findings raises the question of the extent to which there is a theoretically or clinically useful sense in which there are ‘specific’ language disorders at all.

¹ ‘From time immemorial, people have hoped that aphasia might be a starting point which leads to an understanding of mental illnesses.’ (Wernicke, 1874, in Miller & Dennison (eds.), 2015, p.5).
In the more obviously ‘non-specific’ language disorders that we will focus on here, the need for integrated models arises more clearly and strongly. Such models should be general, since a piecemeal approach suited to modelling language dysfunction in, say, ASC, but not Williams syndrome, would be as unsatisfactory as a linguistic theory of only French and English. Claiming that the linguistic variation in question is not in fact ‘linguistic’ but ‘cognitive’, taking us beyond the scope of linguistic theory, is no escape from this task, since this would either amount to an arbitrary stipulation (linguistic variation involving cognitive change is not ‘linguistic’ by definition), or else, if intended as an empirical assertion, it would be extremely implausible: there is currently no empirical basis that the variation involved is fully accounted for by non-linguistic cognitive dysfunctions. Crediting it to variation in general IQ scores is particularly implausible, since fundamental differences in language capacity often appear between clinical groups matched on IQ or general cognitive functioning (for some examples see Bartak et al., 1977; Maljaars et al., 2011; Cokal et al., 2018; Frizelle et al., 2019). Moreover, some of the crucial linguistic phenomena – e.g. failure of language development in ASC or thought-disordered speech in SZ – can occur with normal non-verbal or even verbal IQ scores (see Hus Bal et al., 2017; McKenna & Oh, 2005, respectively). Methodologically, the claim of the independence of language from ‘cognition’ remains untestable if linguistic variation is not investigated in the context of cognitive variation.

In this predicament, neurologists and psychiatrists investigating anomalous speech patterns in clinical populations have often used broad quantitative measures such as number of utterances, Mean Length of Utterance in words or morphemes, the type-token ratio, or the noun ratio (e.g. Ahmed et al., 2013, in the context of Alzheimer’s
disease). Linguists in turn have sought to group anomalies under such labels as ‘syntax’, ‘semantics’, ‘phonology’ and ‘pragmatics’ (e.g. Covington et al., 2005 for schizophrenic speech). But variables of the first type are linguistically of little informational value, and even the conventional variables of the second type are problematic in themselves, as most linguistic phenomena of interest will intersect between these domains, making them barely separable. Moreover, connections between these linguistic variables and cognition are opaque and have not been centrally thematized in traditional linguistic theory. What, in particular, is the cognitive function of syntactic organization in language, if any? Answers to such questions could make syntactic dysfunction interpretable in the context of cognitive dysfunction. Without such answers, we lack a general notion of how or why, if our cognitive phenotype changes fundamentally, language should change too. The basic problem would remain: we do not know which cognitive function involves which linguistic function, or which aspect of language plays which cognitive role.

---

2 Thus, for example, any form of semantics outside of purely lexical semantics necessarily depends on syntax, hence would be misdescribed as ‘semantic’ in any sense that would exclude it’s being ‘syntactic’; in turn, pragmatics today defines a domain largely fused with semantics; and difficulties captured descriptively as ‘phonological’ may be due to underlying syntactic or pragmatic mechanisms. For example, sentence-final rise in English is an intonational tune that might be considered a phonological phenomenon. However, like all prosodic phonology, intonation has to map onto structure and hence syntax is implicated. Moreover, rising declaratives receive a question interpretation unlike declaratives with falling intonation, which are interpreted as assertions. Hence semantics is implicated.
One answer to both questions could be: ‘none’. Language, in other words, plays no role in relation to cognition, which would then be what it is, whether we are linguistic creatures or not. This, however, apart from being implausible in itself, would make puzzling why language is so pervasively affected in cognitive disorders, and in so many systematically different ways. To whatever extent the answer is not ‘none’, we expect that specific aspects of ‘language’ and of ‘cognition’ can be mapped onto each other and covary in pathological conditions, potentially illuminating their neuronal basis (Hinzen & Sheehan, 2015). The model here proposed provides such a mapping. It can be tested for its usefulness by making predictions for how our linguistic phenotype can or cannot change, and which cognitive changes would occur alongside. It can also be tested by casting light on already documented clinical linguistic variation, which so far has no models or only unsatisfactory ones. We will pursue both of these strategies here, and start now developing the question of what the cognitive function of language might be.

2. THE COGNITIVE FUNCTION OF LANGUAGE

In one extreme end on the spectrum of views on the language-cognition interface, the cognitive role of language is restricted to the ‘expression’ or ‘externalization’ of an otherwise already independently functional thought system existing ‘internally’. This expressivist view is associated both with traditional rationalist theories of language (Arnauld & Lancelot, 1676/1966), who saw language in its grammatical structure as a ‘mirror’ of the rational structure of thought as governed by logic and viewed as already given; and with their contemporary ‘Cartesian’ rationalist heirs (Chomsky, 1966), insofar as they attribute language as encoding independently constituted thoughts, perhaps generated in a ‘Language of Thought’ (Fodor, 2008; Pinker & Jackendoff, 2005); or
depict language development as ‘latching onto’ or communicating a world of ‘concepts’ and ‘thoughts’ that we already possess for whatever other reason (Pinker, 1994). But an expressivist view in this sense could equally be attributed to the contemporary cognitive-functionalist tradition (Tomasello, 2003), insofar as it regards language as a social construction for conventionally representing ‘communicative intentions’ and their contents again assumed to be independently given.

The expressivist view is challenged not only by the co-morbidity of cognition and language in neurodevelopmental disorders noted above, but also by pervasive evidence from neurotypical developmental reviewed below, which suggests that language is a critical factor in cognitive development from the very beginning of human life, mediating species-specific forms of categorization, learning, and social cognition (Vouloumanos & Waxman, 2014; Arunachalam & Waxman, 2010). We will here start out with this challenge by considering one possible way in which language might play a fundamental role in relation to our cognitive phenotype.

A familiar claim is that in language the relation between form (sound) and meaning is mediated by grammar. We could imagine several ways in which, in a given species, meaning was not mediated by grammatical structure in this sense. One would be that the meaning is exceedingly simple and associated with sound without the mediation of structure, as e.g. in monkey call systems (Seyfarth & Cheney, 2017). These may resemble innate human vocalizations like crying, grunting, sobbing or laughter more than speech (Deacon, 2006). In our own species, another such way would be that all meanings are abstract objects, denizens of a Platonic-Fregean third realm. This would make natural languages and the forms they exhibit no more relevant to the existence of these meanings than the language of arithmetic may be to the existence of the numbers referred to in that
language (Katz & Postal, 1991). On another possible view, all meaning is effectively causal, controlled by word-object causal links, with language merely being an arbitrary way of referring to the objects in question and conventionalizing the link between them and concepts of them (Fodor, 1998). In either of these cases, one would not have to look at the structure of the system in which such meaning is grammatically configured and conveyed, in order to understand the principles on which it was based. Even if it was true in some sense that meaning in language is mediated by grammatical form, it would still not follow that the meaning in question depended on that form for its existence. For example, ‘grammatical form’ could be theoretically unpacked as simply meaning a function generating binary sets embedding other binary sets (‘Merge’, Chomsky, 2008). In this case, all semantic content would either be lexical or be located on the other, non-linguistic side of the semantic interface – grammar would contribute no meaning, as Merge doesn’t have any to give. This function would merely manipulate units that have those meanings as semantic values attached to them already, for reasons external to the functioning of grammar. While content could come from a universal set of lexical features, there is no restrictive theory of features at present and all forms of meaning arising at a grammatical level would have to be pre-coded lexically.

That the form of language bears on the meaning conveyed and configured in it could mean something deeper, however: that without grammatical forms of organization of the kind seen in every human language, the meanings conveyed in those languages would not exist – the kinds of thoughts ceaselessly crossing through our mental spaces and manifest in language use would simply not be available. This would immediately make sense of the empirical fact that whatever cognitive process we dignify with the term ‘thinking’ in nonverbal species, cognitive phenotypes across types and hence styles of
thinking differ (Penn et al., 2008); and that we seem to never find the same kinds of meanings when language is absent, as in the case of (declarative) referential meaning, which is absent in apes (Tomasello & Call, 2018), absent in humans when language is absent (Maljaars et al., 2011; Slušná et al., 2018), and highly correlated with language in neurotypical development (Iverson & Goldin-Meadow, 2005; Colonnesi et al., 2010).

In line with this proposal, Hinzen & Sheehan (2015) propose an ‘un-Cartesian’ model of language functioning, which claims that the kind of meaning mediated by linguistic form in the strong sense defined above is referential meaning, as coming about through embedding lexical concepts in the context of an act of speech mediated by grammar. Though abstract poetry takes this idea to its limits, there is no other way to use words than the referential one – language could not be used like music, say, or as a calculation device. Moreover, and crucially, mere lexical concepts in the sense of isolated content words – HOUSE, BARK, DOG, etc. – are not as such referential, since they only capture general classes of objects or events but no referential distinctions (e.g. reference to a particular dog or house, as opposed to the dog I saw in that house, or the specific event of his barking that I witnessed yesterday). Reference is thus an aspect of grammatical organization inherently, giving grammar a defined cognitive role. But it is equally important, on this view, that unlike a monkey call system, reference is internally mediated by a store of lexicalized concepts (codified semantic memory). These can be activated at any moment of our mental lives for purposes of thought and reference, irrespective of perceptual stimuli in the here and now.

On this proposal, grammar is the cognitive function converting a lexicalized semantic memory store into referential expressions. By giving a specific content to the claim that form mediates meaning, the un-Cartesian model connects language to thought
directly: referentiality is as intrinsic to normal language use as it is to the thoughts expressed in it. Linguistic evidence for this model comes from the fact that (declarative) reference is not available non-grammatically, is in all of its forms mediated by specific grammatical configurations, and language is never used non-referentially. Starting from and refining Longobardi’s (2005) Topological Mapping Theory (TMT) for quantificational (Russellian definite descriptions) vs. rigid (proper name) forms of object reference, Sheehan & Hinzen (2011) specifically propose an Extended TMT, identifying three (rather than two) configurationally distinct types of object reference: (i) quantificational, (ii) definite-referential, and (iii) rigid. They further identify this same threefold distinction in the domain of clausal reference: (i) intensionally interpreted (non-referential) clauses, (ii) factive clauses that function as clausal equivalents of definities (referring to facts), and (iii) matrix clauses asserted as true, thereby incorporating truth into this scheme as well, by interpreting it as reference at the sentence-level (Arsenijević & Hinzen, 2012). Martin & Hinzen (2014) expand the topology of object reference further to deictic and personal forms of reference involving still more configurational complexity. By integrating reference in all of its forms into grammar and showing it to be configurational, the widely noted developmental link between reference and grammar

---

3 Again, abstract poetry and expressives provide limiting cases for this, of different kinds. Expressives are sometimes not conceptualized as referential, yet they (e.g. oh, wow, oops) cannot be uttered without the presence of a specific contextually given event and hence without a referent being implied in that broad sense. However, this referent appears to be always restricted to the here and now, there is no lexical description of the referent, and no anaphoric relations are possible.
noted above, as seen in pointing (Iverson & Goldin-Meadow, 2005; Tomasello & Call, 2018), is now predicted: pointing itself is a signal of linguistic cognition unfolding, a form of cognition integrated with language and involving reference inherently.

The un-Cartesian claim, however, does not address a typological challenge. How can we understand morpho-lexical and morpho-syntactic variation across neuro-typical languages? Suppose that such variation simply leads to different ways in which grammar plays its hypothesized functional role with language-specific lexical and morphological resources as the prime loci of variation. This would then still not address the typological challenge that languages also differ in the syntactic categories they involve and how this may impact the forms of reference in question. We now introduce a model that can handle such typological variation in categories, which we will later (Section 5) expand into a model that can handle neuro-pathological variation of language as well.

3. **THE UNIVERSAL SPINE HYPOTHESIS**

In recent years, formal approaches towards typology have already moved in the direction of basing grammatical categories on abstract cognitive functions that are linked to grammar, so as to address a long-standing challenge: The identification of syntactic categories in a given language has to employ language-specific diagnostics; but then, how are universal categories identified? How do we know whether there is indeed universal substance to structure, which would serve in the grammatical configuration of reference and propositional meaning, and if there is such substance, what it would be? In this context, Wiltschko (2014) introduces patterns of *multi-functionality* as a universal diagnostic for categorial status. Multi-functionality is the phenomenon whereby a given unit of language can be interpreted in multiple ways depending on its grammatical
distribution. For example, in its use as a main verb (*Virginia has a room of her own*), *have* denotes the relation of possession; in its use as an auxiliary (*Virginia has written her essay*) it has a bleached meaning which expresses grammatical content (Aspect) only. It follows that a given word (or unit of language) may have one meaning in one grammatical context, and a different (albeit often related) meaning in another context. Patterns of multi-functionality are ubiquitous in natural languages calling for a model which predicts them to exist. Assuming that structure comes with substance mediating different kinds of meaning does just that. For any given unit of language, the interpretation will differ depending on its structural position, because the substance associated with structure will affect its interpretation. In this way, syntax not only mediates the relation between sound and meaning for complex expressions but also for simplex ones. The substance of structure adds meaning to words. We refer to this as the substantivist view of grammar.4

Based on cross-linguistic variation, Wiltschko (2014) argues for a substantivist view, according to which grammatical categories are constructed on a language-specific basis. Specifically, units of language associate with a *universal spine* (Wiltschko’s term for a hierarchically layered set of structures that define the substance of grammar). Since units of language are necessarily language-specific, as they involve the conventional bundling of sound and meaning that make up the lexicon of a language, it follows that the categories constructed in this way are also language-specific. Each layer is defined by a function essential for the configuration of reference and propositional meaning. We here

4 This proposal differs significantly from another substantivist approach, namely the cartographic view (Cinque, 1999). For discussion see Wiltschko (2014).
adopt Wiltschko's (2014) four spinal functions depicted in Fig.1, which are hierarchically organized in the sense that all ‘higher’ ones presuppose the foregoing as necessary parts, reflecting an inherent increase in grammatical complexity and meaning: (i) *classification* serves to classify events and individuals into subcategories (e.g. telic vs. atelic events; mass vs. count nouns, etc.); (ii) the introduction of a *point of view* serves to map the classified event or individual to a particular perspective on it (e.g. viewing it as perfective or imperfective); (iii) *anchoring* serves to map the perspectivized event or individual to the utterance context, and (iv) *linking* serves to map the anchored event or individual to the ongoing discourse. Since configuring these layers (first an event, then perspectivizing it, then embedding it in context, then linking it) inherently corresponds to building the formal ontology of a normal thought itself, building this grammatical complexity is conceptually inseparable from building the ontology of neurotypical human thought.

Figure 1: The universal spine

![Diagram](attachment:image.png)

Crucially, these four functions may be instantiated by different grammatical categories. For example, in English clauses, classification yields AKTIONSART categories, point-of-view yields temporal ASPECT, anchoring yields TENSE, and linking yields CLAUSE-TYPING categories. But languages differ dramatically in the surface categories these functions yield. For example, (Ritter & Wiltschko, 2014) argue that in Halkomelem...
(Salish), anchoring is instantiated by LOCATION while in Blackfoot (Algonquian) it is instantiated by PERSON. While superficially TENSE, LOCATION and PERSON appear to differ in their content, their function in the configuration of propositional meaning is identical: they serve to anchor the event to the utterance by establishing ‘when’ or ‘where’ relative to the utterance the event took place or ‘who’ relative to the utterance participants participated in the event (cf. also Zubizarreta & Pancheva, 2017). Similar variation in the specific content of underlyingly identical grammatical categories is found for Case (Lochbihler, 2012; Wiltschko & Ritter, 2015), Viewpoint Aspect (Wiltschko, 2014; Bliss et al., 2011), Aktionsart (Jacobs, 2011), and Number (Wiltschko, 2008; Kim et al., 2017).

The initial goal of postulating the spine was purely linguistic: namely to provide a tertium comparationis (in the sense of Humboldt, 1829/1963) for the comparison of categories across languages. The universal spine hypothesis postulates a structure with substance, which restricts the types of categories languages construct and the hierarchical order they display. However, the functions of each hierarchical layer of the spine as depicted in Fig. 1 are equally ‘cognitive’, though clearly a distinctively linguistic cognitive phenotype is involved. In this way, the universal spine model arrives at a similar conclusion as the un-Cartesian proposal, though on a different route: namely, by showing how the uniform cognitive functions of grammar interact with language-specific resources to create the same effect across languages: a form in which meaning is organized in a particular way and along several layers, spanning the space of thought.

Here we take this idea a step further, by noting that the effect in question goes beyond merely configuring truth-evaluable information (which we could call ‘the grammar of truth’, see further Hinzen & Wiltschko, 2018), but also concerns aspects of linguistic communication and social interaction, which are not typically assumed to be
part of the grammatical system but could be called the ‘grammar of use’. These latter grammatically conditioned configurations constrain the way language is used in social interaction, beyond merely mediating referential meaning. This implies that the way in which grammar structures the space of thought is in fact deeper and indeed potentially exhausts it: it configures both truth- and use-conditional meaning. Evidence that the language function implicates this interactional dimension comes from the fact that speakers of neuro-typical human populations have clear judgments about the appropriate use of language dedicated to encoding it. Note that even core grammatical categories such as Tense and Person have meaning directly relating to the speech act situation: the relational meaning of Tense connects an event and/or reference time to the utterance time, Person connects event participants to utterance participants (Sigurðsson, 2004). But grammar extends to structure the mental attitudes that speakers convey as well as linguistic social interaction and can be modelled as being part of the substance that comes with. It includes the encoding of speech act types (Ross, 1970, Rizzi, 1997), epistemic attitudes (Speas & Tenny, 2003; Bhadra, 2017), tag questions (Wiltschko & Heim, 2016), discourse particles (Haegeman 2014, Thoma, 2016), politeness markers (Miyagawa, 2012), addressee agreement (Miyagawa, 2017), vocatives (Hill 2007) and response markers (Farkas & Bruce, 2009; Wiltschko, 2017). There is significant convergence in this body of work, which suggests that the syntactic spine includes a layer of structure dedicated to hosting the language of interaction. Specifically, the evidence from the interactional language indicates that there are several layers with distinct spinal functions which encode i) **subjective** meaning (the speaker’s attitude toward what is being said), ii) **intersubjective** meaning (the speaker’s evaluation of how the addressee relates to what is being said), iii) **meta-communicative** content (such as turn-taking management)
This matters in the present context, since linguistic diversity in cognitive disorders critically affects the social-communicative dimensions of language.

These dimensions are different from meaning configured in lower layers of the spine in a number of respects: i) They consist of the non-propositional dimensions of meaning which do not contribute to the computation of truth-conditions, as they are configured after truth value is assigned and do not affect this truth value either. They regulate what people do with propositions. ii) Interactive language allows for limited recursion only: there is recursion to the extent that several (potentially articulated layers) of non-propositional language are attested; but it is limited in that interactive language does not allow for self-embedding. iii) Interactive language is typically (though not always) found in sentence-peripheral positions. For this reason, interactive language is sometimes considered to be outside the clause (Kaltenboeck et al., 2016) and thus not part of grammar proper. However, interactive language displays all of the characteristics we expect if it were associated with the spine (Wiltschko, 2019): its meaning is mediated by grammatical form, it shows patterns of multi-functionality (many discourse markers,

---

5 Under current assumptions regarding the units of analysis of syntactic theory this is not an argument since clausehood is no longer a primitive. In Minimalist grammar (Chomsky, 2008) in particular, syntactic analysis concerns itself with sound-meaning correspondences or ‘phases’ of a derivation, which do not necessarily correspond to classic notions such as ‘sentence’ or ‘clause’. Crucially, interactive language is prosodically integrated into utterances and hence may be considered part of a phase in that sense.
e.g., *right*, serve double duty in that they can be used both within the grammar of truth and the grammar of use.), it is structure-dependent, it exhibits patterns of contrast and may be paradigmatic, and finally it displays ordering restrictions. Hence, the interactive dimension, too, has to be part of the universal spine, dominating propositional structure or the grammar of truth, as schematized in Fig. 2.

Figure 2: Extending the universal spine hypothesis

![Diagram](image)

In this way, both the un-Cartesian model and the universal spine hypothesis make grammar and the type of thought expressed in it inseparable and seemingly co-extensive. Yet the empirical question remains where human cognition as structured by the spine begins: there is mental and cognitive life both before and outside of language. Our proposal, which we now present, is based on the fact that there are two pre-linguistic pillars on which language rests, and which, we will argue, it connects: first, infants are able to group stimuli into perceptual classes, forming object categories long before they can produce or even understand words. Second, infants are agents in a space of social interaction. A model that seeks to integrate language and thought has to minimally
address the question as to how grammatical cognition is grafted onto these two critical capacities (categorization and interaction), resulting in a thought system that is as species-specific as it is linguistic.

4. **The Bridge Model**

As perceptual creatures, we form object categories capturing commonalities among perceptual objects. Categorization in this specific sense is neither dependent on language in humans nor human-specific (Hespos & Spelke, 2004; Mareschal & Quinn, 2001; Plunkett et al., 2008; Santos et al., 2002). Humans, on the other hand, are special in that perceptual categories become lexicalized as *words*, and these play a critical role in how, when, and which object categories are formed. From at least 3 months (Ferry et al., 2010), infants group perceptually different objects into categories when these are named but not when they are presented simultaneously with tones. This suggests an incipient understanding of the fact that words refer to things but tones do not. In line with this, 4-month olds appreciate that speech, apart from being a social signal, has *content*: it communicates, with words picking out objects in the world about which a thought is entertained (Marno et al., 2015; Vouloumanos & Waxman, 2014). By 6 months, this property of speech is not only abstractly appreciated, but words are analyzed in accordance with semantic features that make up their meanings: babies for example are sensitive to the fact that cars are more similar to strollers than they are to bananas (Bergelson & Aslin, 2017). In the course of the second year, this process can be seen to further depend on parts-of-speech distinctions such as noun, adjective and verbs, which infants use to map percepts onto categories of different formal-ontological types: objects, properties, and events, respectively (Arunachalam & Waxman, 2010). As infants grow
into this linguistic space of shared meanings, words also exert a top-down influence on visual object perception, as electrophysiological evidence from 1-year olds suggests (Gliga et al., 2010). In these ways, while object perception as such is clearly at least partially language-independent, it is also partially linguistic in humans, taking off on a different course early on, through the way it is connected to speech, and, via speech, to thought.

The perception of speech itself structures human mental and social life from birth. It is one of the earliest stimuli processed prenatally (as early as 24-28 weeks: Eggermont & Moore, 2012), the subject of preferential attention relative to non-speech sounds from birth (Vouloumanos & Curtin, 2014), and it activates perisylvian language regions similar to those seen activated in language tasks in adults (Dehaene-Lambertz & Spelke, 2015). As speech is a social stimulus and crucial social bond, a preferential bias for speech processing makes the space of social interaction inherently linguistic from the start. As an interactive creature, moreover, the infant is not merely a recipient of linguistic utterances directed at it, but an active partner from the very beginning. This can be seen in the shape of the early vocalizations described in the literature as coos and murmurs (Oller, 2000), which occur as part of rapid vocal exchanges with adult partners. These resemble conversations and feature-alternating vocalizations separated by clearly defined pauses (Bateson, 1975; Gratier et al., 2015). By around 2 months, maternal and infant vocalizations are separated by pauses ranging from 500ms to 1s (Jaffe et al., 2001). The coos and murmurs in question are described to be more ‘speech-like’ as compared with vocalizations outside of a turn-taking format (Bloom et al., 1987). They also elicit emotional and motivated responses from social partners, with vocalizations in general
eliciting responses from adult partners more frequently than gaze and smiling (Van Egeren et al., 2001), thereby further forging a speech-related social bond.

These findings demonstrate that there is a linguistic signature to the long-noted precocious social capacities of young infants. Newborns are more likely to vocalize while the mother is speaking (Rosenthal, 1982) and they ‘tune’ their behaviour and movement patterns to the caregiver’s actions more generally. They are particularly sensitive to rhythm, to which they synchronize their crying, movement, sucking, heart rate, and breathing (Provasi et al., 2014), with first evidence of sensitivity to rhythm in foetuses 35 weeks old (Minai et al., 2017). This is a prerequisite for establishing neurotypical social bonds and is restricted to species which display vocal learning (Schachner et al., 2009).

Rhythm in turn is a feature of speech and essential for the acquisition of words through basic rhythmic patterns (stress-, syllable-, or mora-timed) that are relevant for word segmentation; and it is critical for acquisition of word order, which can be bootstrapped via prosodic phonological phrasing (Langus et al., 2017; Nespor & Vogel, 1986). In line with these patterns, neonates at 2-4 days of age show evidence for turn-taking behaviour in the temporal organization of their vocalizations in face-to-face communication with their mothers. Dominguez et al. (2016) found that 68.9% of these vocalizations occurred within a time window of 1 second following the mother’s turn, with 26.9% ‘latched’ onto them, i.e. occurring within the first 50ms, suggesting a surprising predictive grasp of when her turn would end, whose basis is currently unclear. At the same time, 30% percent of vocalizations are characterized by a pattern of overlap, differing from fully developed turn-taking behaviour, which is characterized by a ‘no gap/no overlap’ requirement (Sacks et al., 1974), and is acquired only later.
The full development of these adult-like no gap/no overlap turn-taking patterns takes an interesting trajectory hinting at how language acquires its role in providing the turns in question with linguistic content. As early as 5 months (Hilbrink et al., 2015), and maybe already at 2 months (Gratier et al., 2015), the amount of overlap in the interaction between infants and mothers decreases so as to display more of a turn-taking like structure, i.e. moving towards a no gap/no overlap pattern. However, at some point (Gratier et al., 2015; Leonardi et al., 2017) infants slowdown in their responses, apparently violating the no gap requirement of adult-like turn-taking behaviour. This may be the result of the time it takes for infants to process previous turns and (later on) to plan the appropriate response (Clark & Lindsey, 2015; Hilbrink et al., 2015; Casillas et al., 2016) when such turns start to involve linguistic content that has to be processed. The slow-down at 9 months of age reported in Hilbrink et al. (2015) may specifically correlate with the fact that skills relevant for communication, such as joint attention and pointing, start emerging around this time. As argued in Butterworth (2003), pointing is ‘the royal road to language’. Early pointing richly correlates with linguistic measures, both lexical and grammatical (Iverson & Goldin-Meadow, 2005; Colonnesi et al., 2010). In short, once pointing emerges, grammar does, and the slowdown in turn-taking behaviour could emerge due to its taking on its first lexical and proto-grammatical forms.

Turn-taking, then, like perceptual categorization, is neither dependent on language in humans nor human-specific, as shown by rudimentary turn-taking abilities in other prosocial species such as marmosets (Chow et al., 2015) or songbirds (Henry et al., 2015). Yet again, as in categorization, when language is grafted upon these precursor structures, these are also transformed: categories become words; words become parts of thoughts as configured in clauses; clauses define the contents of conversational turns; and turns are
managed with linguistic means as well. As grammar falls into place, these contents come to exhibit the full referential format of thought as structured by the spine layer by layer, up to and including the discourse and response markers as well as intonational tunes that are the outposts of grammar at its interactional end.

Grammar, in short, acts cognitively as a bridge that connects perception-based categorization to social interaction in its full scope as involving thought, spanning the space in between, where such thought is hierarchically built. We refer to this as the ‘Bridge model’ (Fig. 3). As depicted there, the extended universal spine is at the core of our language faculty. Its substance determines the grammatical organization of human thoughts and the ways these thoughts are communicated. It is grafted upon evolutionarily older cognitive domains, namely perceptual categorization and social interaction, which we have suggested form direct interfaces with grammar-based thought. The former defines an entry point after which lexical items appear, the latter an exit point as thought-sized units are packed into turns of social speech. The spine is the bridge in between.

Figure 3. Language as a bridge between categorization and social interaction
Since the bridge model incorporates cognitive functioning in verbal minds, it can be used to generate predictions for cognitive dysfunction when there is language dysfunction of different kinds. As a crucial test case and support of how clinical linguistic diversity might be approached in this way, we now turn to the challenge of modelling a spectrum of language abilities across the autism spectrum. We will do so with a view to addressing the following questions, which reflect constraints on any useful model for non-specific clinical linguistic variation:

A) Does ASC entail variation in the linguistic phenotype?

B) Does the model entail predictions for how this phenotype can (or cannot) change?

C) Does it have a good degree of fit with the clinical data?

5. The Broken Spine: Linguistic Diversity in ASC
While language dysfunction is not criterial for ASC anymore in current autism diagnostic schemes (DSM-5), and has no status in current theoretical schemes of its underlying neurocognitive basis, its practical and clinical significance in this disorder is uncontested. Thus, language is one of the most important early reasons of referral, and some of the earliest behavioural warning signs for autism are language-related, such as deviance in babbling (Patten et al., 2014), or production and perception of speech sounds (Arunachalam & Luyster, 2016), including an absence of the preference of speech over non-speech, and of infant-directed speech relative to adult-directed speech (e.g., Vouloumanos & Curtin, 2014; Droucker, Curtin, & Vouloumanos, 2013), mentioned as a feature critical to early social development above. Early language levels are also a crucial predictor of outcomes (Howlin et al., 2014), a role that recent neuroimaging studies using a range of methods confirm at the brain level (Lombardo et al., 2015; Eyler et al., 2012; Kuhl et al., 2014). ASC also harbours the most significant non- or minimally verbal population in our species: language is estimated to be absent or minimal in a substantial 25-30% of individuals on the autism spectrum (Norrildgen et al., 2015; Tager-Flusberg & Kasari, 2013), with earlier estimates of up to 50% (Rutter, 1978; Bryson et al., 1988). Non- or minimally verbal autism (NMVA) in this sense is crucially different from mutism, which is an affective disorder and involves individuals who have shown evidence for a language capacity (even if they do not speak anymore or only to selected people), and comprehension levels in NMVA or equal to or lower than the (minimal or absent) production levels (Hinzen et al., 2019). Language impairment as seen in NMVA, moreover, forms no special or extraneous case in ASC, but is continuous with, though more severe, than language impairment seen across verbal autism with ID, which comprises nearly half of the spectrum (Centers for Disease Control and Prevention, 2014).
Even in the other half, where people with ASC reach IQ scores in the normal range, language impairment at structural levels is abundant, and universal in aspects of language identified as ‘pragmatic’ (Boucher, 2012; Noterdaeme et al., 2010). If, in that half, language is not compromised on standardized tests, linguistic differences often show up in non-standardized tasks (e.g. narrative or cohesive discourse: Norbury & Bishop, 2003; Rumpf et al., 2012; Fine et al., 1994; Bartlett et al., 2005; Eigsti et al., 2016; Suh et al., 2014), in how language is processed in the brain (Mizuno et al., 2011; Moseley et al., 2016; Radulescu et al., 2013; Stigler et al., 2011; Mills et al., 2015), or in speech processing anomalies (Alcántara et al., 2004; Klin et al., 1991; Kujala et al., 2013; Vouloumanos & Curtin, 2014; Foss-Feig et al., 2017).

The autism spectrum thus literally is a language spectrum, with language function ranging all the way from nearly non-existent to nearly (but never quite) normal, reflecting the multifarious ways in which affected children use their preserved cognitive resources to come to terms with an external language model that does not fit their own internal makeup. One specific idea has been that non-verbal IQ, which tends to be higher in ASC than verbal IQ (Rutter, 1978), may help in compensating for language difficulties (Boucher, 2012). That idea, however, clearly does not work in a substantial minority of children with NMVA who have non-verbal IQ scores in the normal range (Hus Bal et al., 2017), and it does not predict that many children with ASC and total IQ scores in the normal range show language impairment (Kjelgaard & Tager-Flusberg, 2001). ASC with ID is also unique among other ID groups in the severity of the language disorder involved in the former case, and the degree to which comprehension is affected along with production, with the former tending to be more affected than the latter (Maljaars et al., 2012; Maljaars et al., 2011; Garrido et al., 2015; Slušná et al., 2018). In particular, in
Down syndrome, the most common cause of ID, no substantial subpopulation with no or minimal language is reported (Martin et al., 2009). Overall, a model of language dysfunction across ASC based on general cognitive functioning as measured by IQ does not seem promising.

Since ASC is diagnosed through deficits identified under such labels as ‘communication’, ‘reciprocal social interaction’, ‘nonverbal gestures’, or ‘restrictive and repetitive behaviours’, as in the DSM-5 today, a model based on a more specific mechanism than general IQ would need to relate to cognitive mechanisms underlying these behaviours. But this raises the conceptual problem of how a dysfunction in any of these ASC-criterial domains could possibly not relate to language. Thus, the problems of communication and reciprocal social interaction in question are evidently not intended to be problems in forms of communication and social interaction also seen in animals, where such behaviours do not depend on language. On the contrary, symptoms of ASC are located precisely in human-specific forms of communication and social interaction, which involve language function. Crucially, individuals with ASC, even at its lowest-functioning end, do both socially interact and communicate (Cantiani et al., 2016; DiStefano et al., 2016; Maljaars et al., 2011; Preissler, 2008), just in their own ways rather than the species-typical ones linked to language in its normal use. Moreover, human communication and social cognition is linguistic from birth. As reviewed above, one of the most paradigmatic reciprocal human interactions, starting from birth, involves speech. Paradigmatically autism-related so-called ‘nonverbal’ gestures like pointing, shared attention, and even social smiles, closely relate to language in development (on social smiles, see Hsu et al., 2001). Diagnostically significant behaviours such as echolalia or stereotyped and idiosyncratic phrases, moreover, descriptively are nothing other than
anomalies of normal language function, which is creative and interactive in its normal use by nature. Specific cognitive mechanisms postulated to explain social interaction deficits in ASC, such as ToM, are closely related to language in development as well (Farrar et al., 2017).

This leaves the option of a language-based model of the problems in communication and social interaction criterial for ASC, thus addressing question A above. But if this was a model of language impairment in ASC as reflecting a co-morbidity of SLI, it would just record this co-morbidity as fact, and it would face two additional problems: linguistic profiles in SLI and ASC diverge (for discussion, see Boucher, 2012; Taylor et al., 2012); and it cannot account for language impairment of the nature and scale seen in NMVA. The basic flaw of such a model, however, is not to create a conceptual and empirical connection between linguistic and cognitive dysfunction, which on the SLI model precisely are separate. A bridging hypothesis is needed of why language impairment shows up so pervasively in ASC, and why it is of the specific type that it is. Linguistic and social-cognitive dysfunction, particularly reciprocal social interaction, should connect, the moment the social function of language is recognized and we don’t confine ourselves to a traditional modularist model of language based on its syntactic function only. Note that at this junction, characterizing the linguistic phenotype of ASC under such labels as ‘syntax’, ‘semantics’, etc., would only allow us to record the facts for each of these domains assuming particular definitions of them, and again leave the issue of the connection between language and (social) cognition unanswered. More importantly, it would make no predictions for what linguistic dysfunction we should find or not find in a disorder like ASC. These conventional linguistic variables make no
predictions in relation to each other, either: for example, it is not clear what a ‘semantic’
dysfunction would entail for a ‘syntactic’ one.

Here is where the bridge model does something different, addressing question B) above. It starts from the observation that, in ASC, the bridge is affected, not merely the
pillars (communication/interaction and perceptual categorization in a non-linguistic
sense). Thus, deficits show in social interactions as mediated by language, not merely
their pre-linguistic or non-linguistic forms, and grammar-based dimensions of language
have indeed often been found to be proportionally more affected than vocabulary
(Boucher, 2012; Arunachalam & Luyster, 2016). Even vocabulary and lexical meaning,
though, are not normal in those children with ASC who develop words (Tek et al., 2008;
Arunachalam & Luyster, 2016), showing that the language change is global, covering
language at both a grammatical and lexical level. Since the model has this bridge
(language) structured into hierarchically ordered layers, moreover, each of which are
associated with syntactic, semantic, phonological, and pragmatic properties, predictions
can now be stated: it is in the nature of a hierarchy that higher entails lower, and an
immediate prediction is that as the bridge disintegrates, we will find that any intact layer,
\( n \), should entail the intactness or any lower layer \(<n\), while layers \( >n \) could be deviant.
This makes predictions for possible and impossible clinical linguistic change – and
associated cognitive dysfunction, which each of the layers is connected to.

In particular, a child who can put single phrases together, such as verb-noun
combinations, might not show evidence of sensitivity to Aspect (or episodicity, as relating
to Tense or ‘Anchoring’ higher up the spine); but the other way around should not be
observable. It could also be that children break into language in the pillars of the bridge:
A child with NMVA will seek to identify its own strategies of socially interacting so as
to achieve its goals, and it will be able to learn some words based on pre-linguistic perceptual categories or concepts that it has. But the prediction will be that as long as this child does not develop grammar, the use made of the words in question will not be the typical referential one. This is borne out by current research, which shows that to the extent that words are acquired in NMVA, they are not processed normally (Preissler, 2008; Cantiani et al., 2016), and declarative reference is absent (Slušná et al., 2018). Analogously, a child with Williams syndrome might break into language via its interactive end, showing great skills as a conversationalist, yet showing semantic deficits at a grammatical and lexical level, which have to do with how the spine grows from one pillar to the other.

Where there are productive two-word combinations, in turn, we would expect a child to understand basic events, but with or without an ability to make distinctions depending on a point of view on them, such as whether they are ongoing or completed (Aspect); and with or without an ability to anchor them in the utterance context, or in time (Tense). It is noteworthy in this regard that problems with time-deixis have been highlighted in ASC from early on (Bartolucci et al., 1980; Fine et al., 1994). Where, in terms of the layers of the spine, classification, point of view, and anchoring are in place, additional questions would arise, such as whether a child can link events placed in context to other events referred to in discourse. If linking is affected, this would predict pervasive findings of problems in the more grammaticalized forms of reference (pronouns, definite NPs and deictic devices) even in ASC without ID, which directly relate to discourse-based linking (Banney et al., 2015; Bartolucci et al., 1980; Modyanova, 2009; Norbury & Bishop, 2003; Rumpf et al., 2012). If linking is possible, in turn, the intricacies of social-interactive language including turn-taking and the interpretation of utterances in
relation to epistemic states, may or may not be neurotypical. That they are not, may specifically result in misuse of personal pronouns documented in ASC (Shield et al., 2015; Bartolucci et al., 1980; Mizuno et al., 2011), misuse of discourse markers which serve to indicate the epistemic states of the interlocutors (e.g., oh, huh,...), terms of address specifically geared towards conversational management (e.g. vocatives, intonational tunes, proper names in their normal uses), or response markers and backchannels which serve to indicate mutual understanding and (dis)agreement (yes, no, mhm, right).

Since to be linguistic is to master this hierarchy along the bridge, and the autism spectrum is a language spectrum, our model thus invites mapping language function seen in ASC systematically onto the different hierarchical layers of the spine, creating implicational predictions at each layer as well as for associated cognitive dysfunction as currently measured by non-linguistic variables as ‘perspective-taking’, ‘reciprocity’, or theory of mind. Language can be seen as not merely another behavioural and neurocognitive domain where a given syndrome shows impairments, but as an integral mechanism in how and why a social-cognitive phenotype unfolds on a different path. Where this mechanism fails, alterations of this phenotype are predicted and rationalizable on a cognitive basis that integrates language, such that typologies of cognitive disorders can be developed on an at least partially linguistic basis. This is not to say that the bridge model is the only model that has promise of a good fit with the clinical data, addressing question C. What we have argued is that (i) non-linguistic models based on IQ, say, are widely recognized not to be promising, (ii) linguistic models based on an SLI-type comorbidity are limited in scope, failing to account for language spectrum seen and its
specificity, and (iii) linguistic models integrating linguistic and cognitive function have not been articulated so far, with the bridge model opening up this pathway.

6. Conclusions

Two types of linguistic diversity mark the human linguistic phenotype: one staying within the confines of a neurotypical thought process, another co-varying with a disintegration and fundamental change in the latter. Our perspective suggests systematically linking such change to a correlated change in our language capacity as modelled by current models of UG that we have used here. While UG is the current term used, though, if the breakdown of UG turned out to be the breakdown of species-typical cognition more broadly, it would be mislabelled: UG, when intact, is a thought and social communication system as much as it is a grammar system. Yet if it is, it can hold a new key to making sense of the cognitive diversity that forms one of the deepest challenges for both human scientific understanding and societies today.

References


Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn-taking, timing, and planning in early language acquisition. *Journal of Child Language, 43*(6), 1310–1337. [https://doi.org/10.1017/S0305000915000689](https://doi.org/10.1017/S0305000915000689)


http://dx.doi.org/10.1016/j.neuron.2015.09.026


https://doi.org/10.1002/icd.1976


https://doi.org/10.1016/j.nicl.2015.11.014


Audition-specific temporal processing deficits associated with language function in 
https://doi.org/10.1002/aur.1820

Frizelle P, Thompson PA, Duta M and Bishop D. (2019). The understanding of 
complex syntax in children with Down syndrome [version 2; referees: 3 approved] 
Wellcome Open Research 3:140 
(https://doi.org/10.12688/wellcomeopenres.14861.2)

comprensión del lenguaje en niños no verbales con trastornos del espectro autista y 

Processing in 1-Year-Old Children. *Journal of Cognitive Neuroscience, 22*(12), 

715.

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. 
(2015). Early development of turn-taking in vocal interaction between mothers and 
https://doi.org/10.3389/fpsyg.2015.01167


https://doi.org/10.1093/brain/awr151


https://doi.org/10.1080/136820310000108133

https://doi.org/10.1177/1362361314556782

Adolescent Psychiatry, 19(6), 475–481. https://doi.org/10.1007/s00787-009-0057-0


