Mathematical comparison of Linguistic Categories without Universal Spines

Part IIB Dissertation

Jago Rowe

9994 Words

Abstract

This dissertation looks at the mathematical nature of Neo-Emergentist Generative theory (Biberauer, 2019b) in syntax and phonology, as well as in cognition more widely. Neo-Generativism attempts to reconcile insights about the formal properties of language systems from Generativist Theory, with a radical impoverishment of Universal Grammar on grounds of empirical data and psychological plausibility. Following Bosch (2022), properties of Dynamical Systems are used to model this emergent behaviour. I look at Successive Division (Dresher, 2003, and following) as a way for acquirers to postulate increasingly fine-grained featural distinctions, attempting to constrain the typology of relations between categories to highlight order preservation, and avoid chaos and 'rampant and unconstrained variation' (Biberauer, 2019b, pg. 59). I relate this to Song (2019)'s implementation of category-theoretic Adjunctions to mediate comparisons between emergent Extended Projections in syntax. I argue that these techniques allow for a maximally impoverished Universal Grammar, and that emergent categories do not need constraining by a small innate ordered set of functional, non-linguistic categories, such as proposals of Universal Spines and similar mechanisms (Wiltschko, 2014; Ramchand and Svenonius, 2014). This contrasts with Song's conclusion that a Universal Spine or similar is required for meaningful comparison of syntactic categories. This dissertation contributes to arguments in favour of the plausibility of maximally strong Neo-Emergentist theory as able to capture the properties of linguistic systems without unconstrained overgeneration, as well as linking to and making novel predictions about various cognitive domains.

Plain Language Summary

This dissertation looks at categories in Neo-Emergentist Generative theory (Biberauer, 2019b), a modern framework which combines Generativist ideas about linguistic structure with a significant reduction in what is innately assumed as part of Universal Grammar. I follow Bosch (2022) by using Dynamical Systems Theory to explain how reducing the content of Universal Grammar means the system is more than the sum of its parts. I look at Dresher (2003, and following)'s proposal of Successive Division as a way that acquirers can generate new features. I present different relations between these features, which are empirically supported and maintain order preservation, without leading to 'rampant and unconstrained variation' (Biberauer, 2019b, pg. 59). I also look at Song (2019)'s use of category theory to compare different syntactic Extended Projections. I suggest a generalisation of his approach, which is stronger as it does not require Universal Grammar to contain a small innate set of ordered functional categories, such as a Universal Spine or similar proposals (Wiltschko, 2014; Ramchand and Svenonius, 2014). This dissertation presents a strong form of Neo-Emergentism, arguing that it can find a balance between capturing the intricate detail of linguistic systems, while not being too powerful. I also link strong Neo-Emergentist theory to proposals for hierarchical structure in non-linguistic cognitive domains.

Acknowledgments

I would like to extend my immense graditude to Theresa Biberauer, my Director of Studies and Dissertation Supervisor. Theresa's incredible knowledge and support is visible throughout this work, through inspiration for this research, recommendations for interesting readings in the literature, and detailed commentary on all parts of the presentation and argumentation. Discussions with Núria Bosch and Julio Song have helped guide this dissertation, particularly with reference to mathematical subfields, and the influence of their work linking Neo-Emergentist linguistics to Dynamical Systems Theory and Category Theory respectively is particularly clear. I would also like to thank the audience at SyntaxLab in Cambridge and online, organised by Theresa, for the questions and comments on this work and its presentation to audiences. Finally, I would like to thank many of my fellow final-year linguistic undergraduates for collaborative study, sharing of resources, answering questions, and proofreading. All errors are my own.

Contents

1 Introduction

It is often stated that language is a system which makes 'infinite use of finite means' (Chomsky, 1965, quoting Von Humboldt (1836)). This has been pursued at the level of Phrase Structure by Mainstream Generative Grammar (MGG)[1], but is also relevant to higher level cognition including discourse (Polanyi and Scha, 1983; Watumull, 2012)[2]. The research strands in the literature on phonological and syntactic features have largely proceeded independently, and the idea that a finite set is innately specified has persisted[3]. However, in recent years, more approaches are questioning the plausibility of rich, innately specified feature sets on grounds of empirical data and evolutionary implausibility (i.a. Hall, 2007; Dresher, 2003; Cowper and Hall, 2014; Wiltschko, 2014; Ramchand and Svenonius, 2014; Biberauer, 2019b). These approaches wish to maintain the findings of formal generative linguistics, while minimising the required contents of Universal Grammar (UG), and so are termed Neo-Emergentist (Biberauer, 2019b). Emergentism here relates to the term in Dynamical Systems Theory (DST)[4], which refers to non-linear interactions between subparts of a system, creating a whole which is greater than the sum of its parts. Neo- situates these approaches within a Generativist background, as opposed to traditional usage-based Emergentist approaches (eg. Tomasello, 1992; Ellis, 1998). For Neo-Emergentism, 'infinite use of finite means' must also apply to the domains of linguistic features.

In this dissertation, I adopt the economic goals of Neo-Emergentist theory in modelling language systems, while making minimal assumptions. Such modelling should explain not only what the properties of language are, but why they are so, which is beyond explanatory adequacy (Chomsky, 1965, 2004). I look at and develop the mathematical modelling of these systems, situating linguistic theory within philosophical approaches to language and cognition as mathematical objects. However, achieving discrete infinity is trivial for even the most basic algebraic structures[5]. This dissertation looks at how constraining the possibilities can be modelled using mathematical theory, supported by empirical and cognitive data across modalities and domains. In particular, I will look from the perspective of Complex Adaptive Systems, within DST. The above characterisation of linguistic systems shows Edge-of-Chaos behaviour: The systems can see non-deterministic innovations, but must be learnable and avoid the chaos and 'rampant and unconstrained variation' (Biberauer, 2019b, pg. 59) that discrete infinity should tend to.

Section 2 looks at approaches to features, arguing that Neo-Emergentist approaches which highlight properties of Complex Adaptive Systems are needed for empirical coverage. Section 3 looks at the comparison of such categories using different mathematical techniques. Section 4 looks at the explanatory power for linguistic theory that comparison of Neo-Emegentist categories can give insight into.

2 Neo-Emergent Features

2.1 Approaches to Feature Systems

In this section, I will look at the debate around feature systems across different linguistic domains. As stated by Panagiotidis (2014, pg. 2), 'our understanding of the human Language Faculty will advance further only if we pay as much attention to features as we (rightly and expectedly) do to structural relations.' Features have been extensively studied from a number of perspectives, including different perspectives from phonology and syntax. In phonology, the physiological basis for many of the parameters of variation has influenced Generativist work in proposing extensive feature sets early on (Jakobson and Halle, 1956). Generative phonology has largely used binary features, although Element Theory uses privative features, making an analogy to chemical elements

[1] I use this term loosely, generally corresponding to the views held by Chomsky at a given time.

[2] Notably it has been argued that recursion in not universally found within syntax, with Pirahã being the most persistent argument (Everett, 2010), though cf. suggestions that Everett's conceptualisation of recursion is misguided (Zwart, 2011a).

[3] For example, (Chomsky, 2001, pg. 10) states it is a 'conventional assumption' that a language makes a 'one-time selection' from a set of innately specified syntactic features.

[4] In a technical sense, DST studies mathematical equations relating to time-based systems.

[5] Van Steene (2021) shows an infinite free magma (see Appendix B) generated from minimally a single feature.

(Backley, 2011). Element Theory does so to focus on the shared knowledge of speakers and hearers, rather than privileging articulation. It proposes privative 'elements' which relate directly to perceptible units of the speech signal, for example an element [I] which is pronounced /i/, and can be combined with other elements, rather than abstract feature bundles. The combination of Neo-Emergentist thinking with the direct relation to the linguistic signal of Element Theory has been previously noted (Tiflit and Voeltzel, 2014).
In syntax, extensive feature sets did not receive significant attention until the era of linguistic Minimalism (Chomsky, 1995), although distinctions between parts-of-speech were relevant significantly earlier (see i.a. Chomsky, 1970). The highly abstract nature of syntactic features means their structure and properties are extensively debated[6], and this hinders comparisons between approaches. Extensive overviews of this topic include Adger and Svenonius (2011); Panagiotidis (2014); Song (2019). I will review recurring themes in the literature below. I will pay particular attention to the properties of non-linear Dynamical Systems, showing that feature systems necessarily display these. DST has been applied to many cognitive systems, following the Dynamical Hypothesis and Mathematical Universe Hypothesis, (1), which provide a philosophical background for the mathematical investigation of language undertaken here.

(1) a. **The Dynamical Hypothesis**:
Cognitive agents are Dynamical Systems.
(van Gelder, 1998, pg. 615)

b. **The Mathematical Universe Hypothesis**:
Our external physical reality is a mathematical structure.
(Tegmark, 2014, pg. 254)

In (2.2), I loosely follow the properties of Dynamical Systems as presented by Bosch (2022). Other linguistic applications of DST include Van Geert (1991); Larsen-Freeman (1997); Irvin et al. (2016)[7]. Mathematical introductions include Gottlieb (2001); Thelen and Smith (2007).

2.2 Features and Dynamical Systems Theory

2.2.1 Emergence

Emergence in DST refers to non-linear interactions between subparts, leading to the whole being greater than the sum of the parts (Abler, 1989; Tucker and Hirsh-Pasek, 1993). Emergence of phonological features has been investigated extensively by Mielke (2008). Innatist approaches struggle to explain why the same segments should pattern in contradictory ways in relation to natural classes and phonological rules across languages. As an example, /i/ is the prototypical example of a front segment (cf. Element Theory in (2.1)). Hungarian vowel harmony operates on a front-back basis, as seen by alternations between /a/ and /e/ depending on the [±front] value of the preceding vowel. Unexpectedly, /i/ patterns as a neutral vowel, excluded from the language-specific class of front vowels (Törkenczy, 2011)[8]. The contradictory status between Hungarian and other languages like Turkish, where /i/ patterns as a front-vowel as expected, suggest that the higher level interactions between segments are emergent, and greater than the sum of the parts.
Similar arguments for emergence can be found in syntax. Languages make overt use of different features in different ways[9]. Wiltschko (2008) shows systematic differences between the plural in English and Halkomelem Salish, given below (from Wiltschko's Table 6):

[6]It is not even universally agreed within Generative syntax that features are necessary (Boeckx, 2014).
[7]Research by authors of the latter studies the child-mother dyad as a Dynamical System, which emphasises linguistic systems outside of a single mind.
[8]This particular case can be explained diachronically, but there is a synchronic mismatch.
[9]Although see Miyagawa (2009) for the opposite hypothesis in syntax, that all languages overtly realise the same features.

(2)

	English	Halkomelem Salish
Plural marking is obligatory for plural referents	yes	no
The absence of plural marking implies singularity	yes	no
Plural marking is restricted to nouns	yes	no
Possibility of pluralia tantum	yes	no
Possibility of plurals within compounds	no	yes

Wiltschko concludes that the differences here are reducible to the inflectional (English) vs non-inflectional (Halkomelem) status of plural marking. These distributional differences in (2) suggest differences in the syntax of plurality between the two languages. It would be highly stipulative to postulate two (or more) distinct innate features for plurality, but rather the inflectionality distinction must arise emergently through interactions between the subparts of the language system of the acquirer and the world, leading to non-linearities in the system.

Emergence as described here should be contrasted with maturational approaches (eg. Root Infinitive hypothesis, Rizzi, 1993), which assume innately specified feature sets. In these, full availability of features is modelled in a time-based system, but the whole is just the sum of the innately specified parts (Bosch, 2023). Arguments for emergence and against innateness of features in phonology also come from sign languages (Cowper and Hall, 2014; Sandler, 2014), where the phonological features are entirely distinct from the dominant spoken modality.

2.2.2 Sensitivity to Initial Conditions

Non-linear systems are highly sensitive to their starting conditions, or the initial state. The initial state of language acquisition has been perhaps the biggest debate in modern linguistics, in the form of whether there should be a UG. The three factor model of Chomsky (2005) provides a highly reductive model.

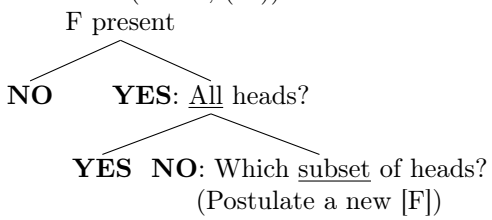
(3) Three factors of language design:

1. Universal Grammar.
2. Primary Linguistic Data (PLD).
3. Factors not domain-specific to language.

Of these, parts of Factors 1 and 3 describe the initial state of the system. Factor 2 is the input used in progressing to different states in the system. The parts of the input that children pay attention to differ across time in a 'Goldilock's' way - input which is not too easy or complex for the current state of the system (Biberauer, 2019a). This Goldilocks Effect arises as a consequence of third factors across cognition, as evident in vision and audition (Kidd et al., 2012, 2014).

Various frameworks implicitly weight the factors differently. MGG work of the Principles and Parameters era, and some modern work in traditions like Cartography (Cinque, 1999), assume a rich UG. This also occurs in some phonological approaches, where descriptive analyses of phonemic contrasts require an incredibly large number of proposed features to capture all differences (eg. Moran, 2012). However, the psychological burden of assuming a such a rich UG becomes increasingly more difficult as further study of languages sees an increasing number of attested categories (Ramchand and Svenonius, 2014), and the evolutionary requirements here diverge significantly from biolinguistic research (eg. Berwick and Chomsky, 2016; Huybregts, 2017). This would imply that at least some categories must be emergent, following a mechanism such as Successive Division (SD) (Dresher, 2003), or the None>Some>All learning path (Biberauer and Roberts, 2015), (4):

(4) Biberauer (2019b, (14))



The question then arises as to what the initial state of the system should be, or what is in UG. For MGG approaches, UG might consist of as little as (Simplest)[10] Merge. If Merge contributes to hierarchy in other domains like music (Katz and Pesetsky, 2011) or maths (Hiraiwa, 2017), then very little remains as domain-specific to language. Slightly larger proposals for UG come from Ramchand and Svenonius (2014) and Wiltschko (2014)'s Universal Spine (US)s. These approaches propose that there is an innate functional hierarchy, which influences the emergence of categories in language specific ways. Wiltschko's US is given below:

(5) Classification $\leq$ Point-of-view $\leq$ Anchoring $\leq$ Linking

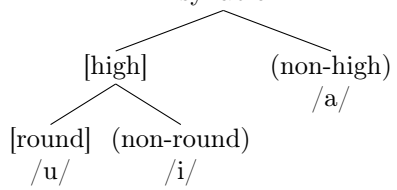
The question I will be concerned with here is whether the initial state of a linguistic system should be like a US, or not contain any domain-specific information beyond Merge, as in the strongest formulation of Maximise Minimal Means (MMM) (Biberauer, 2019b). The answer to this question cannot be answered on psychological grounds at present, given the more coarse granularities of objects investigated by psycho- and neurolinguistics when compared to formal syntax (Poeppel and Embick, 2005). However, the explicit existence of a US would be highly significant for cognition, and has not been proposed on non-linguistic grounds[11].

An empty categorial space as an initial state is a stronger hypothesis than something like (5) as innate. The former could theoretically lead to the latter, if the corresponding divisions are made. Therefore, assuming the weaker hypothesis (5) would be logical if the divisions in (5) are universal, and the stronger hypothesis is logical if not all possible linguistic systems make these distinctions.
Before answering which hypothesis is preferred on these grounds in syntax, I will switch attention to phonology, whose similarity to syntax for emergent theories has been noted (Bazalgette, 2015). Here, there is no equivalent to a totally ordered (Appendix A) US. One potential equivalent is the different division of labour between vowels and consonants (Nespor et al., 2003). Consonants are used in distinctions between lexical items, whereas vowels are more important for signalling grammatical properties. This division of labour is not found in other animals like rats (Bouchon and Toro, 2019). However, this division of labour is also found in sign languages, where movements pattern like vowels, and handshapes and locations like consonants (eg. Liddell, 1984; Brentari, 2019). This provides a strong argument against the vowel vs consonant distinction as being innate with any reference to acoustic/auditory form, and favours a model where this division of labour is emergent in the PLD (Biberauer, 2022).
This very early bootstrapping of the phonological system demonstrates the sensitivity to initial conditions well. Simple decisions, like the speech vs sign modality or the basic divisions into natural classes in the vowel inventory (as in (6)) have far reaching consequences for the subsequent paths the system may take. Equally, there are fewer possible points of variation at the very early stages of the system, as less is immediately accessible to the acquirer.

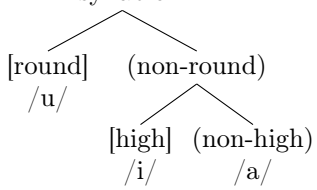
(6)

[10] Simplest Merge is solely binary set formation without other axioms, cf. other proposals for Merge types discussed in (3.1).
[11] Although Wiltschko does not require the US to have psychological reality.

a.



b.



Dresher (2014, pg. 168)

The different organisational possibilities for even basic tripartite vowel inventories as above is supported on acquisitional and typological grounds (Dresher, 2014; Bazalgette, 2015).

I return to syntax now, noting that the presence of a US would provide an asymmetry between syntax and phonology. This would be easiest to justify if US divisions characterise the earliest states of language acquisition. This is debatable. For Wiltschko, USs should aid in the anchoring of functional categories in both the nominal and verbal domains, and a symmetry is expected here (see (27) for arguments in favour of this, including specific symmetrical categories). However, major asymmetries have also been identified (Bruening, 2009). One of these arises from selectional differences. There are verbs which select for certain clausal elements, such as declarative or interrogative complementisers below, but there are no verbs which select for functional nominal elements, such as the presence or absence of a specific article. Such asymmetries mean that US-like divisions reflect clausal structure better than nominal structure for certain languages, which is not expected if USs characterise early syntactic divisions[12].

(7) a. Sue thinks that the world is flat.

b. *Sue thinks whether the world is flat.
Bruening (2009, (1))

A second possible location for US-type distribution of functional categories is word-internally (Marantz, 1997). This can be seen with Expressive Morphology, the use of morphological marking to signal speech act related information (Zwicky and Pullum, 1987).

(8) (Dutch)

a. baz-en
babble-INF
'to babble'

b. baz-**el**-en
babble-*el*-INF 'to waffle' (pejorative) (Cavirani-Pots et al., 2023, pg. 5)

The suffix *-el* can attach to nouns and verbs, has a prototypically aspectual (iterative) meaning, and also includes inherent pejorative information, presenting a challenge as elements of these types are expected to be in separate places in a hierarchy. Language users are able to make use of existing material in novel ways, such as when previous formal distinctions have been lost and excess material is repurposed for new distinctions such as this speech act meaning (Lass, 1990; Los, 2013). Again, US-like divisions between high and low categories like discourse and agreement appear to recur in several positions, which is expected in DST, but not in a way which lends support for a US as an innate linguistic object.

[12]The relation of clauses and nominals in embeddedness terms has been debated since Rosenbaum (1967), and so introduces debate here.

The final challenge for USs in syntax discussed here is differences between modality. Although the basic division between VP, TP and CP has been observed for sign languages (Pfau et al., 2018), the externalisation of these domains shows differences, primarily due to the greater simultaneity allowed by sign languages[13], where up to four propositions may be expressed at once in cases of storytelling (Napoli and Sutton-Spence, 2010). Information relating to agreement and speech acts is articulated using different articulators such as non-manual marking, where changes in facial expressions and similar are grammaticalised. This cuts across domains, as shown by minimal pairs in the lexicon disambiguated only by facial expressions. One example is PITY and FALL-IN-LOVE in Catalan Sign Language, where the former has a negative facial expression and the latter a positive one (Pfau and Quer, 2012).

The most compelling difference between USs as they are proposed for spoken languages, and sign language data, comes from the *anchoring* segment. Wiltschko (2008) devotes Chapter 4 and 5 to introducing the different possibilities for anchoring categories in the verbal domain, which include tense (as in English), person (Blackfoot), location (Halkomelem) and realis (Upper Austrian German). In sign languages, this variation is not attested. Anchoring only occurs in relation to space - it is not clear how this could otherwise occur in the visual modality. The explanatory power of a distinct anchoring segment as a maturational feature of language systems, whose formal realisation is then emergent, is therefore undermined by sign languages[14]. Across modalities a fully emergent approach seems more parsimonious.

To conclude, when considering variation across languages, there are no convincing reasons at this point to adopt the weaker hypothesis of USs, as opposed to the stronger hypothesis of a radically impoverished UG. Another potential reason for postulating a US-like structure will be introduced, and rejected, in (3.3.1).

2.2.3 Self-Similarity

In this section, I will look at the self-similar, or fractal, patterns of natural languages systems. These are a key property of Complex Adaptive Systems (Beggs, 2008). This has been widely investigated diachronically in grammaticalisation processes in Western languages. Self-similarity has also been investigated for linguistic elements which can occur synchronically in multiple positions (Robert, 2018). An extreme case of the latter is given for Baka (Niger-Congo) (Heine and Kilian-Hatz, 1994): the morpheme 'tɛ' can behave like a particle, a preposition, an auxiliary, or a co-ordinating or subordinating conjunction. The self-similarity idea arises as these elements are similar to one another, but non-identical as the size of the elements in relation to the linguistic system differs.

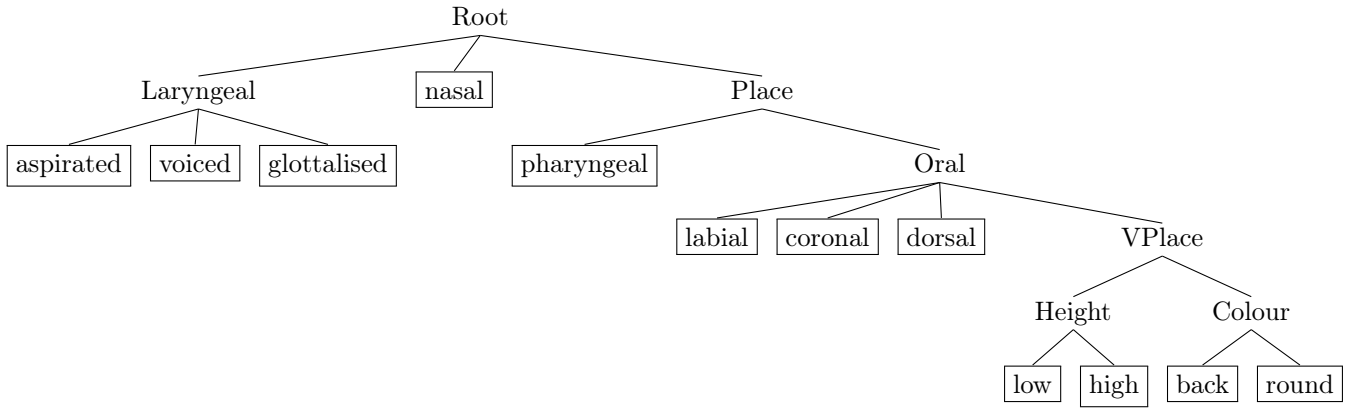
In this dissertation, I will investigate self-similarity from the perspective of granularity instead. I look at how large sets of linguistic categories behave in similar ways to individual categories. Large sets of categories in phonology are investigated in feature geometry approaches (McCarthy, 1998; Broe, 1992)[15]. These group the distinctive characteristics of different categories into a complex web of nodes, such as (Padgett, 2002, (3)):

[13]Morphological simultaneity is found in spoken languages in cases like ablaut or autosegments. This can be extensive, like non-concatenative exponence in Dinka nouns (Arkadiev and Klamer, 2019, pg. 447), but the spoken modality inherently restricts this more than sign languages, as the auditory signal has fewer independent articulators than the visual one.

[14]This may be seen as a 'no-choice parameter' (Biberauer et al., 2014b), although the question still arises of why a no-choice parameter in a specific modality should be innately determined in a significantly impoverished UG.

[15]Feature geometry approaches have been used, although more rarely, in syntax (Harley and Ritter, 2002).

(9)



Here a node like *Colour* consists of the set of features {back, round}[16]. Self-similarity arises by the recursive division into smaller natural classes - for example, analogous to *Colour* as a set is *VPlace* as a set consisting of *Colour* and *Height*.

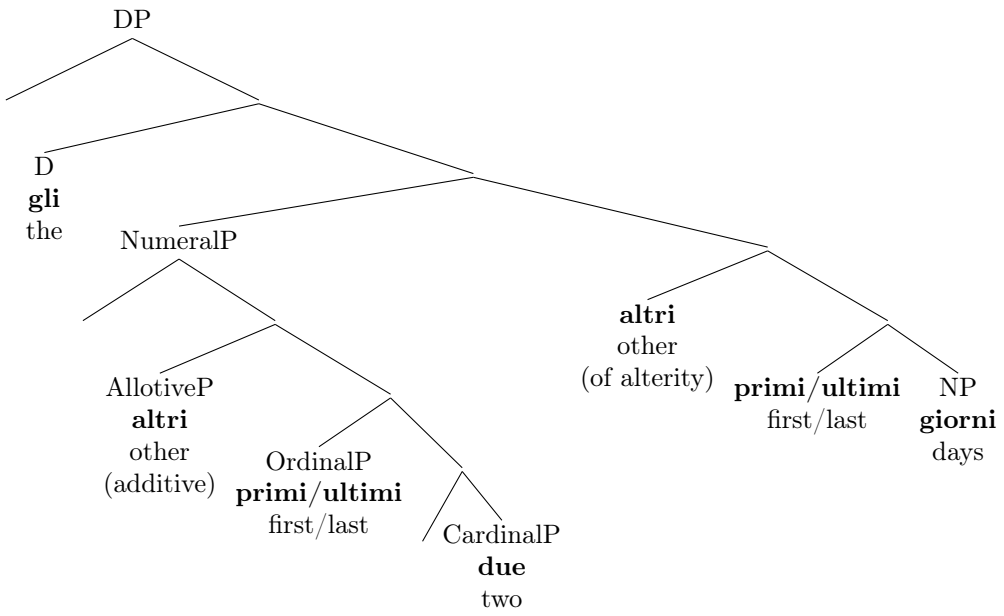
I compare these self-similar feature geometries with Extended Projection (EP)s, a representation of higher granularity categories in syntax (Grimshaw, 1991). A prototypical EP, as introduced by Grimshaw, should have a lexical head at the bottom, and related functional heads above. For example, the classical $V \leq v \leq T \leq C$ sequence, which can map neatly onto a US (5), can also be represented as an EP. The two classical EPs are therefore the Nominal Extended Projection (NEP) and the Verbal Extended Projection (VEP). However, the proposals in the literature related to EPs have significantly more variation.

On one hand, the number of proposed EPs is greater than two for some analyses. Song (2019) references the Pre-/Post-positional Extended Projection and the Adjectival Extended Projection. These two further categories have been the next most privileged in Generative Grammar after V and N, notably in the 'Amherst System' following Chomsky (1970), which decomposes adjectives into [+V, +N], and prepositions into [-V, -N], although their ontology as EPs is unclear if they are taken as emergent categories. A smaller-scale proposed Number EP by (Cinque, 2022, (29)), (10) here, explains how the same adjective can appear in two different positions with different semantics. This example makes the self-similar nature of EPs particularly obvious, as they can fit inside other EPs, with smaller semantic coverage.

(10) a. Gli altri primi/ultimi due altri primi/ultimi giorni.
'The other first/last two other first/last days.'

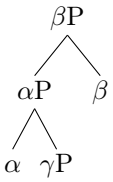
[16]For Padgett, Colour exists solely as this set, whereas for other feature geometric approaches non-terminal nodes have an independent status, cf. EPs below.

b.



On the other hand, the noun/verb distinction, which is taken as early acquired or even basic by several authors working on linguistic categories (eg. Panagiotidis, 2014), has been argued not to hold in all languages (Rijkhoff and van Lier, 2013). If this is the case, then there may be only one, or zero, EPs. Further support against the innateness of two EPs comes from investigation into fillers at the pre-morphemic stage of language acquisition, where categorial divisions into separate EPs of other categories do not appear to have occurred yet (Kilani-Schoch and Dressler, 2000; Veneziano and Sinclair, 2000; Veneziano, 2003). Fillers are single segmental elements which mark basic word class rather than lexical items, which appear in multiple languages before adult morphology for word classes has developed (Peters and Menn, 1997, for a review). This supports an emergent approach to EPs, whose number can vary. Another approach which reduces the number of postulated EPs is Mixed Extended Projection (MEP)s (Borsley and Kornfilt, 2000), which suggests a subset of a VEP is dominated by the NEP. Evidence for this in some languages comes from the Final-over-Final Condition (FOFC), a proposed universal on structure building. FOFC states that the structure below, where α, β and γ are syntactic heads, is not attested for certain combinations of heads (Sheehan et al., 2017).

(11) *



The locus of FOFC has been debated (see Biberauer and Rowe, 2025, for an overview). It has been claimed to hold within EPs (Biberauer et al., 2014a), and evidence against a stronger formulation than this comes from languages including German, where a head-initial NEP (with highest category DP) is dominated by a head-final verbal projection, but no FOFC-violation occurs.

(12) a. [$_{\text{CP}}$... dass$_{\text{C}}$ ich [$_{\text{TP/VP}}$ [**$_{\text{DP}}$ die$_{\text{D}}$** [**$_{\text{NumP}}$ drei$_{\text{Num}}$** [**$_{\text{NP}}$ Männer$_{\text{N}}$**]]] sehe.$_{\text{T/V}}$]]
[$_{\text{CP}}$... that$_{\text{C}}$ I [$_{\text{TP/VP}}$ [**$_{\text{DP}}$ the$_{\text{D}}$** [**$_{\text{NumP}}$ three$_{\text{Num}}$** [**$_{\text{NP}}$ men$_{\text{N}}$**]]] see.$_{\text{T/V}}$]]
'... that I see the three men.'

This contrasts with several languages in West Africa, including Bùlì, where FOFC seemingly holds across the two EPs, as seen in examples of nominalisation. In the Bùlì example, normal VO word order mandatorily changes to OV before a head-final nominaliser. This has been argued as support for the existence of MEPs in these languages, and that FOFC operates over these larger domains (Hein and Murphy, 2022). The relation between core EPs is emergent here, as it differs from the more symmetrical relation seen in other languages.

(13) Bùlì (Niger-Congo) (Cited in Hein and Murphy, 2022, pg. 341 (11))

a. *dɛ̀ mángò-kú-lá*
ate mango-DEM-DEF

'ate that mango'
(Hiraiwa (2005a)).

b. *mángò-kú dɛ̄-kā̄*
mango-DEM eat-NMLZ

'eating the mango'
(Hiraiwa (2005b)).

In summary, I have given examples of categories of different granularities across syntax (10b) and phonology (9). A self-similar aspect of these is how they all make divisions into subparts of larger categories, which I represent with the partial-order relation $\leq$ (Appendix A) here. Examples that have been shown of dividing a larger categorial space into a smaller one are given below:

(14)

	Syntax	Phonology
Individual Category	Anchoring Tense $\leq$ Verbal	Dresher 2014 (6a) [high] $\leq$ [syllabic]
Sets of Categories	Mixed Extended Projection VEP $\leq$ MEP	Feature Geometry Colour $\leq$ VPlace

Following the earlier discussion, I model these categories as emergent, arising from SD of a larger categorial space. This mechanism also allows for smaller granularities to arise. Empirical support for this comes from acquisition. There is evidence that a single C category for complementisers arises relatively early, but fine-grained division into various information structural heads, allowing for differentation between topicalisation, focalisation etc., occurs significantly later (Bosch and Biberauer, 2024). In later sections, I will expand on the use of Category Theory for modelling emergence of categories of different granularities (Appendix B has a brief mathematical introduction to Category Theory). Category Theory studies the relations between objects abstracted from different mathematical subfields. For the purposes here, Category Theory helps formalise how a larger category like an EP can be viewed on two levels - both as an indivisible object, and as a category with a collection of internal objects and relations.

2.2.4 Edge of Chaos

The above discussion has defended a view of language as a system characterised by emergence of structure rather than innate knowledge. Part of the challenge for such a system is Plato's Problem, or how the PLD seems deficient compared to the intricate system which is acquired. Research into Third Factors attempts to explain this. Another problem is the opposite: how the mechanisms which allow for linguistic variation should be constrained such that they do not tend towards unlearnable chaos. The Edge-of-chaos concept from DST is useful here. Non-linear systems arise at the edge of order and chaos, benefiting from both non-linear innovation and regularities (Waldrop, 1993). The rest of this dissertation looks at mathematical modelling which may limit variation to the edge-of-chaos. In this section I elaborate on what the edge-of-chaos should mean for linguistic

systems.

First, I will look at feature systems. It has been mentioned above that there are relatively few possibilities for the first acquisitional divisions. In phonology, the division between vowels and consonants (defined for their information-carrying properties rather than acoustic/auditory substance) holds across modalities, and in syntax, the division between predicates and arguments (corresponding to verbs and nouns) seems to hold for the vast majority of languages. It follows that there are increasingly greater possibilities for later divisions, which arise from less immediately salient information in the world, according to Goldilocks effects (2.2.2). The field of linguistic typology looks at this variation, and the findings from this field present incredible diversity, which continues to grow as more languages are studied. In phonology, certain phonemes appear to only have a single independent innovation, before being spread diachronically and through contact. These include linguolabials in Vanuatu (Lynch, 2019), and click consonants in Southern Africa, which developed early into the history of language and characterise the consonantal inventories of these languages to a significant degree, but do not occur outside the continent (Huybregts, 2017). However, grammaticalisation of features is not 'anything goes' as might be predicted from the above, as plausible features which are cognitively salient such as [±bright] or [±poisonous] are never grammaticalised (Leivada and Barceló-Coblijn, 2021). The possible reasons given for this include issues of learnability, adaptation constraints, and semantic restrictions. These may contribute to avoiding chaos.

The role of syntactic edges as a point of variation in contact situations has been investigated from a number of grounds. In Afrikaans, South African English and Singapore English, innovations in negation and modal particles are frequently incorporated into the linguistic system at salient edges, like clause final position (Biberauer, 2023; Erlewine, 2023). Similarly, conditional inversion, the ability to for subject-verb inversion to signify an irrealis interpretation, has been in decline since the Early Modern English period (Biberauer and Roberts, 2017). For modern conservative speakers, it is available with just three verbs (*were, should, had*, displayed below), although total loss is also common. This variation is chaotic in that it targets individual lexical items in an idiosyncratic manner, but is restricted to the upper clausal edge.

(15) a. Should I hear from him, I'll inform you immediately.

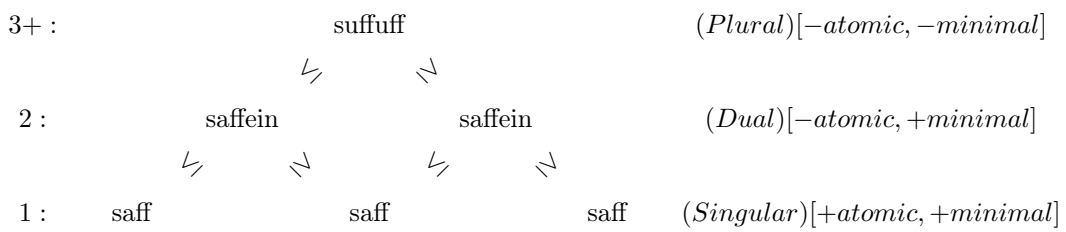
b. Had I heard from him, I'd have informed you immediately.

c. Were I able to hear from him, I'd inform you immediately.

d. *Could I hear from him, I'd inform you immediately.

On the other hand, loss of non-monotonic functional sequences in heritage languages has also been observed (D'Alessandro and Terenghi, 2024). This includes loss of the dual in heritage varieties of Arabic. The authors adopt Harbour (2014)'s analysis of number systems as deriving from two features, [±atomic] and [±minimal]. The semantics of these features are understood in a lattice framework (Harbour, 2016). [+atomic] selects for the atomic elements of such a lattice, whereas [+minimal] selects the smallest elements in a given region of the lattice. Informally, the potential combinations for counting up to three 'classes' (singular *saff* in Palestinian/Egyptian Arabic) are given below:

(16)



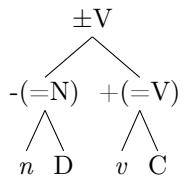
The singular form comprises the atomic elements of the set of classes, and vacuously the minimal units of this too. Plurals do not refer to atomic units, nor the minimal units of the non-atomic subset of the lattice. By contrast, duals display a non-monotonic pattern by having a [-atomic] value, but switching polarity in [+minimal], as they refer to the minimal elements of the non-atomic lattice. This non-monotonicity is hypothesised to favour the attested loss of dual in heritage Arabic varieties (Albirini and Benmamoun, 2012), and these changing values are also at the upper edge of the number domain, which fits with the edge-of-chaos phenomenon.
In the above discussion, the definition of syntactic edges seems to vary. Edges relate to locality, or information-transfer. The current primary MGG approaches to locality include Relativized Minimality (Rizzi, 1990) and Phase Theory (Chomsky, 2001). The details of these are unimportant here, but they are intended as salient edges in syntactic derivation for hierarchical phenomena to operate over. For Chomsky, phases are explicitly non-dynamic, and operate above the level of words[17]. Neo-Emergentist thinking of the whole being greater than the sum of the parts supports a dynamical approach to relativized minimality/phases, the latter of which has been independently suggested (Bošković, 2014), and better captures the derivational edges described in this section.

3 Comparison of Linguistic Categories

3.1 Structure of Features

So far, I have not discussed the assumed structure of various categories. As mentioned in (2.1), there is extensive debate about this. Song (2019)'s Chapter 2 discusses several parameters of variation for features, including specification, valuation, and typing, which will not be followed up on here. I follow the Neo-Emergentist literature in adopting a mixture of binary (boolean) and privative features. Dresher (2003, and following) favours boolean features, whereas (Cowper and Hall, 2014) favour privative features, although both note the general ease of switching between the two types. This can be seen in (17) (adapted from Biberauer (2019b)'s (24)), which has the first division represented as a boolean feature, and second divisions into privative features.

(17)



However, I do not believe that the model presented here rules out features with other characteristics, such as multivariate features, or secondary features (Adger and Svenonius, 2011), as are argued for in different theoretical frameworks. Following from the earlier discussion of granularity, and how an element can be represented as both an object and a category, there is value to investigating the internal structure of features. A logical comparison is with phrase structure, specifically Merge. (Simplest) Merge also obeys binarity, and can be used to describe the vast majority of cases of syntactic structure (Chomsky et al., 2023)[18]. However, some phenomena have motivated alternative formations, including Parallel Merge (Citko, 2005) and Late Merge (Lebeaux, 1988). I make no comment on the merits of these proposals, but focus on the deviation from binarity[19]. This can be resolved if binary branching is an attractor in the space of mathematical structures (Roberts et al., 2023), as it is the most efficient way to generate maximal structure from minimal inputs, as the output is one greater than the input. This is motivated by its similarity to the successor function, which returns an input plus one, and which is an attractor in the computational space as evidenced by simulating Turing machines (Minsky, 1967,

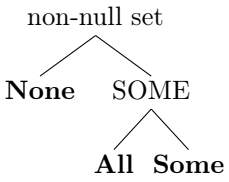
[17]Specifically CP and vP in the earliest work.
[18]The importance of Merge for some authors is described with the slogan 'all you need is Merge' (Berwick, 2011).
[19]A particularly interesting case of deviation from symmetric binarity is Zwart (2011b)'s idea of Merge as an ordered pair. This highlights asymmetry, in a similar way to how I discuss categories in (3.3.1).

1985). Under this view, the bias towards binarity is a reflex of a third-factor principle (Piattelli-Palmarini and Vitiello, 2017), and the default in the majority of cases, but atypical situations at edges of syntactic derivation, perhaps at various interfaces, might see more complex formulations, both in syntactic and feature structure. A consequence of this is that UG does not need to contain an innate feature template, as suggested as a possibility in (Biberauer, 2019b), but this can be emergent.

3.2 Successive Division

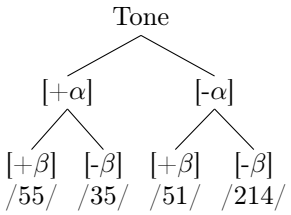
I will look at the mathematics of SD in more detail in this section. SD, or the None>Some>All learning path, has been independently noted in a number of linguistic and non-linguistic cognitive domains (Biberauer and Bosch, 2021). Applications outside of feature systems include in vision (Leeuwenberg and Van der Helm, 2013) and concept formation (Seuren and Jaspers, 2014). Figure (4a) from the latter is given as (18). This explains how *some* can felicitously mean *some-or-all*, but the non-adjacent pairs *some-or-none* and *none-or-all* are never lexicalised in language.

(18)



Dresher (2003, and following) has looked at deriving attested vowel inventories using SD. I will do this for Turkish and Hungarian, paying attention to the interactions between features, and how variation can be restrained. For n independent binary features, there are 2^n distinct possible values in a lattice framework. This minimises the number of features in the input, but massively overgenerates the possible feature combinations compared to attested contrasts for the majority of language systems.
For two features which are independent of each other, I use slightly different notation to Dresher. An example from Krekoski (2013), following Dresher, is given below:

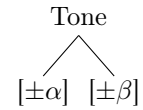
(19)



(19) gives feature specifications for tones in Beijing Mandarin, represented with pitch contours on a rising scale 1-5. [-α] represents a falling tone, and [-β] represents a rising tone[20]. Combinations of these features straightforwardly account for the flat, and falling-then-rising tones. (19) has two instances of the [$\pm\beta$] feature. This could be seen as suggesting that the division into [$\pm\beta$] is made independently as different points. If divisions into the same features could proceed randomly and many times independently, the predictive power of division as limited by third factors is lost, and there is nothing preventing a tendency towards chaos. Instead, while [$\pm\beta$] is distributed once over both possibilities for [$\pm\alpha$], I represent this in a Category Theoretic view with a single instance of both features(20).

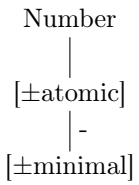
[20]Which value of the binary feature is taken as positive is unimportant in this model.

(20)



Some pairs of features are in a dependency relation, as captured by feature geometries. (16) gives an example which can be represented like this. [±minimal] only becomes active for negative values of [±atomic], which I represent to the right of the division line in (21).

(21)



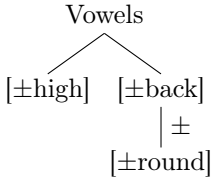
However, it is possible for two features to be related by SD, and also allow all four permutations. This is seen in the Turkish vowel system.

(22)

	Front		Back	
	Unround	Round	Unround	Round
High	i	ü	ı	u
low	e	ö	a	o

There is ample phonetic and phonological justification to suggest that backness and roundedness are closely connected in vowels, as represented with the node *Colour* in (9). Both effect the second formant and enhance one another perceptually (Stevens et al., 1986), so this relation is plausibly emergent, but consistent across languages. Vowel height is uncontroversially less closely related to these two articulatory parameters, although not universally unrelated: Map 11A in WALS (Maddieson, 2013) surveys a sample of 562 languages for the phonemic presence of front rounded vowels. Only 37 have such vowels, and 23 of these have them for both high and mid vowels. Given these relations between features, I propose the successive division of Turkish vowels as below:

(23)



Some justification that this analysis is on the right track comes from vowel Harmony. There are two vowel phonemes underspecified for certain features, notated /I/ and /E/ in the Turkish literature. /E/ alternates between /a/ and /e/ depending on the ±back feature of the preceeding vowel. /I/ varies between /i/, /ü/, /ı/ and /u/, depending on the ±back and ±round features of the previous vowel. Roundness harmony very commonly occurs with harmony of other parameters crosslinguistically (Rose and Walker, 2011), and it applies in a subset of the cases that backness harmony does in Turkish, so this position of roundness as derived from backness is logical.

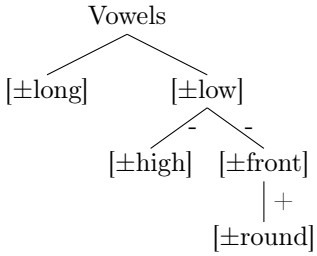
The Hungarian vowel inventory is given here too[21], followed by the SD analysis I propose for it:

[21]The right symbol of each pair represents a long vowel.

(24)

	Front		Back
	Unround	Round	
High	i í	ü ű	u ú
Mid	e é	ö ő	o ó
Low			a á

(25)



There are two vowel harmony alternations. One is between /a/ and /e/, the other is between /e/, /o/ and /ö/, both including long variations. Length acts independently of vowel quality. The first quality distinction is low /a/ separate from the other vowels. Next are independent distinctions of height and frontness, and finally front vowels have a further distinction of roundness. Although the back vowels are phonetically rounded, the features here have an arbitrary relation to acoustic qualities, and so roundness is only a specified contrastive feature for front vowels. Justification for this geometry of emergent features comes from acquisition data. Zajdó and Stoel-Gammon (2003) measured the proportions of correct vowel phoneme production for children acquiring Hungarian, stratified by age and gender. I give the short vowel data for girls age 2;0 (n=8), as gender and vowel length were not statistically significant:

(26)

Vowel	a	e	i	o	ö	u	ü
%	93	90	96	91	66	81	69

The authors note that physiological explanations for the late acquisition of lip roundedness, like two year olds not consistently having the muscle control to produce this (Ruark and Moore, 1997), do not predict such a difference between front and back rounded vowels. By contrast, (25) predicts roundness as a contrastive feature of front vowels to be later acquired than the front/back distinction, in which back vowels are vacuously round[22].

This section has looked at a few case studies to investigate the parameters available for Successive Division in phonological feature systems, although the mathematical relations here should apply to syntactic SD too. Binary features are optimal for maximising minimal means, but massively overgenerate if viewed completely independently of one another. Instead, there is evidence from phonological classes, rules, and acquisition data that features can differ in their relations to one another. The Turkish and Hungarian data may be particularly neat, as I do not believe that there are no cases where features are used independently in multiple positions, or that every division in a system must lead to a separate contrast, but the tendency to avoid overgeneration leads to concise and more plausibly learnable systems.

3.3 Order Preserving Relations

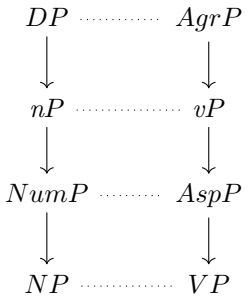
3.3.1 Comparison of Core Extended Projections

Having looked at how categories can arise within a system, and some of the ways this might be constrained, I will now look at mathematical comparison between categories which are not directly related by SD, when and why

[22]The authors note that not all children show the same temporal development, which is expected for Emergentist theory.

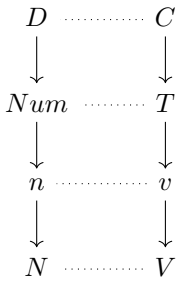
this should arise, and what implications it has. The primary case study I look at in this section is comparison between the NEP and VEP, which has been the focus of previous Neo-Emergent work (Song, 2019), although I reference other phenomena and intend this as widespread.
Innatist approaches to categories have proposed direct one-to-one comparisons between the structure of the two core EPs. I show two proposals for this in the literature below - notably the two are incompatible, as they distinct categories across the EPs.

(27) a.



(Megerdoomian, 2008, (42))

b.



(Roberts, 2024, (21))

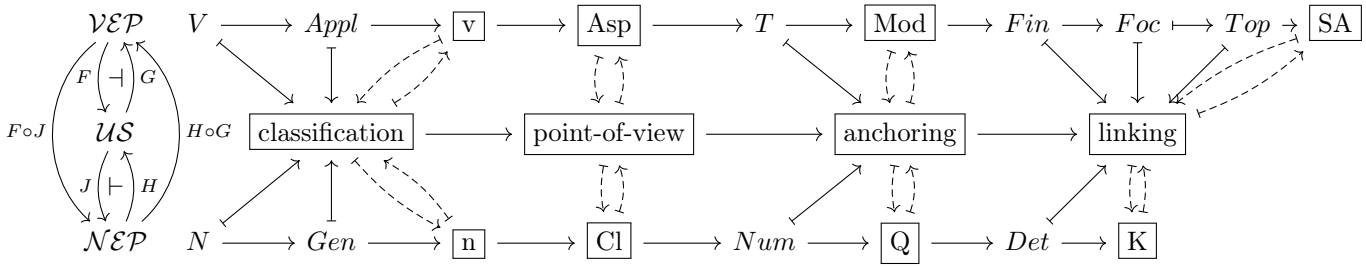
It is necessary to compare EPs in at least some cases, notably when features interact emergently across EPs within a language. This can be seen with number in Frankenduals[23], where a dual is formed from a plural noun and a singular inflected verb (Harbour, 2020).

(28) Innovation in Belfast English for some speakers, (Henry, 2005, 1610-1611, cited in Harbour 2020):
   a. 'The (two) men is talking.'

Frankenduals have been attested in several typologically distinct languages, but the reverse pattern of a singular noun and plural inflected verb is never attested, suggesting there is a meaningful asymmetric comparison here. Harbour argues [±atomic] is in the nominal domain and [±minimal] is in the verbal, but their combination leads to emergent duals.
However, if a comparison should exist between two EPs, it is not easy to state which categories in each should be comparable, except perhaps lexical categories. As mentioned earlier, the equivalence of functional N and V categories is debated due to asymmetries in selectional behaviour (Bruening, 2009). The DP hypothesis, which takes functional categories in N to be heads in the NEP, is widely accepted, at least for English (Abney, 1983), partially 'based on presumed parallelisms between the noun phrase and the clause' (Salzmann, 2020, pg. 1). For languages like Russian, there are more serious doubts about the validity of the DP hypothesis (Bošković, 2005). At a surface level, an obvious difference is the lack of articles in Russian. Another difference is the widespread possibility of Left-Branch Extraction in article-less languages[24]. For languages with articles, there are differences in their formal similarities. German *das* (neuter nominative) is homophonous with a complementiser, whereas Romance articles have developed from Latin demonstratives, which pattern like other adjectives.
The described facts suggest that comparisons should arise emergently in some, but not all circumstances. Importantly, the two categories to be compared do not need to have the same cardinality of elements. This rules out the strongest forms of comparison available in Category Theory - equivalence and isomorphism - as these generalise one-to-one functions. The next strongest form of comparison is an Adjunction (Appendix B), which Song (2019) has used to model comparison of EPs. An example comparison between the two core EPs with arbitrary categories[25] is given below (pg. 219):

[23]There is also verbal number, although this is not directly comparable to nominal number (Cabredo Hofherr and Doetjes, 2021).
[24]Although Korean is an example of a language with no overt D's and also no possibility of LBE (Kim, 2011).
[25]Taken from Song's work, acronyms explained in the relevant section.

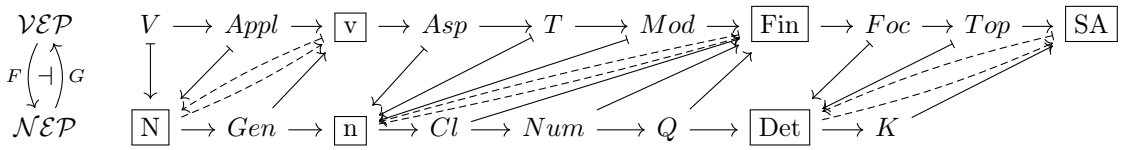
(29)



Here, an EP is represented as a totally-ordered set. The most notable point here is that Song requires a US to mediate the comparison between EPs. Every EP is left-Adjoint to a US, and more than two EPs can be Adjoined to the same one. This creates symmetry, in that comparison of categories returns the greatest element in a given segment of the EP.

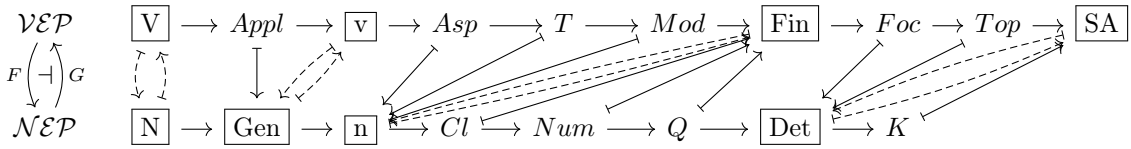
I propose that there is value in investigating direct Adjunctions between EPs, which are inherently asymmetric. This idea is presented by Song (pg. 214), but rejected due to this asymmetry, and the desire to avoid stipulation:

(30)



However, as discussed earlier, there are numerous asymmetries between the NEP and VEP. I introduce new ones here, with a particular focus on how an acquirer could postulate an asymmetric Adjunction. One strong argument for this is data from acquisition. Nouns and verbs are both core lexical categories acquired early in acquisition, but nouns are generally acquired in far greater numbers than verbs at the earliest stages (Gentner, 1982)[26]. For other EPs, timing differences in acquisition might be greater. These differences could be relevant for the smaller, fractal, EPs, such as (10b). Another asymmetry is semantic, as the combinatory and event related properties of nouns and verbs are distinct. In Combinatory Category Grammars (Steedman, 1993), nouns are type *NP* and verbs are type *(S\NP)/NP*. A final asymmetry is the salience of the EPs, particularly with regards to their edges. While the salient unit of a VEP is the clause, there are many suggestions for the top of the NEP across languages, including Determiner (Longobardi, 2001), Case (K) (Bittner and Hale, 1996), Quantifier (Shlonsky, 1991), Classifier (Cheng and Sybesma, 1998), or various other heads (see Corver, 2013, for a more extensive overview). At the lower edge of the EPs, V and N might be parallel, but this can be straightforwardly represented mathematically in the Adjunction, as Adjunctions are weaker than one-to-one comparisons:

(31)



These reasons suggest that asymmetric comparison of EPs is a productive path for future research. One major question is why some language systems should have clear comparisons between the core EPs, and others do not. In polysynthetic languages, verbs inflect for a large number of categories, whereas nouns do not (Baker, 1996). An Adjoint comparison here would not provide much explanatory power. While Neo-Emergent theory

[26]There are findings in samples of children acquiring Mandarin and Korean that verbs are acquired before nouns (Tardif, 1996; Gopnik and Choi, 1995). In either case, this is not symmetrical.

rejects the notion of an innate 'Polysynthesis Parameter' as Baker had proposed, the system defining properties of polysynthesis are expected to be acquired early. If the early states of the system are less chaotic, as argued earlier, then there is less motivation for comparison and constraining of the development of independent categories. Therefore, Adjoint comparisons are expected to be more meaningful when the differences between the behaviour of nouns and verbs are later acquired, and following similar trajectories helps to avoid chaos. English might be an example of this, as it is relatively poor in inflectional morphology and syntactic headedness is the same in both core EPs.

3.3.2 Comparison of Other Categories

I will now look at how this can be extended to different categories. The benefits of Adjoint comparison are in order-preserving relations among categories of different cardinalities. Biberauer (2017, 2018) looks at instances of monotonic behaviour across languages, which can be straightforwardly modelled with an Adjunction. One of these examples is from Pesetsky (2013), in Russian. A subset of Russian animate nouns with masculine grammatical gender may show feminine agreement when used with a feminine referent. One such noun is *vrač* (doctor). The lowest adjectives, displaying idiomatic composition with the noun, cannot show feminine agreement, whereas higher adjectives and verbal participles can. The restriction is that masculine agreement starts at the bottom, can switch to feminine at a given point in the structure, but no switching back to masculine is allowed.

(32) a. *Nov-yj* *vrač-∅* *prišla-∅.*
new-**M**.NOM.SG doctor-NOM.SG arrived-**F**.SG

'The new doctor arrived.' (26b)

b. **Nov-aja* *vrač-∅* *prišël-∅.*
new-**F**.NOM.SG doctor-NOM.SG arrived-**M**.SG

'The new doctor arrived.' (26c)

c. *Glavn-yj/*Glavn-aja* *vrač-∅* *poliklinik-i* *skazal-a,* *čtoby...*
head-**M**/***F**.NOM.SG doctor-NOM.SG clinic-Gen.SG say-PST.**F**.SG that.Subj...

'The (female) head doctor of the clinic ordered that...' (27a)

d. *U nas byl-a* *očen' xoroš-aja* *zubn-oj* *vrač-∅.*
by us COP-PST.**F**.SG very good-**F**.NOM.SG dental-**M**.NOM.SG doctor-NOM.SG

'We had a very good (female) dentist.' (28b) (Skoblikova, 1971, pg. 183)

The data can be seen as the order-preserving relation between the two Partially Ordered Set (poset) (Appendix A) categories below:

(33) a. ADJ(ectives): Low Adjectives ≤ High Adjectives ≤ Participles

b. GENDER: M ≤ F

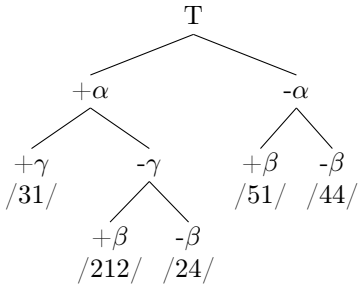
c. $\mathcal{ADJ} \;\underset{G}{\overset{F}{\perp}}\; \mathcal{GENDER}$

An Adjunction between the two can model how order is preserved, as there are different possibilities for how large a range feminine agreement applies to. This type of modelling can go beyond grammatical features. An obvious example of this is FOFC, which inspired Biberauer's discussion of monotonic phenomena in language.

Poset categories, which may differ in cardinality, as seen in the comparison of German (12) and Bùlì (13), are in an order-preserving relation with a poset *Head-Complement* $\leq$ *Complement-Head* relating to linearisation[27]. This left-right linearisation asymmetry is natural with the temporal flow, as well as relating to physiological factors such as decreasing subglottal pressure available across the production of an utterance, which has been formalised as 'biological codes' (Gussenhoven, 2004). Interestingly, unlike the previous examples where Adjunctions are used in relatively late acquired phenomena, basic headedness is acquired very early (Wexler, 1998; Tsimpli, 2014), making use of prosodic cues (eg. Christophe et al., 2001; Soderstrom et al., 2003; Guasti, 2016). An Adjoint comparison here would help an acquirer get into the system, rather than reducing the possibilities as later comparisons seem to. This multifunctionality gives the comparisons plausibility in Maximising Minimal Means.

In addition, order preservation might be used to model SD of categories. Krekoski (2017) looks at tonal systems in several varieties of Chinese, most of which are more complicated than the Beijing Mandarin system of (19). Modern Nantong Chinese, in (34), has five contrastive tones, so requires minimally three features to model this. Krekoski links SD of features here to tone sandhi between relations, which is a similar perspecitve to the vowel harmony discussion above (22):

(34)



(Modern Nantong Chinese Krekoski, 2017)

The phonetic correlates of the features here are not fully specified, but in rough terms $+\beta$ represents a fall, $-\gamma$ represents a rise, and $-\alpha$ represents peripheral position within the tonal space. It can be seen that the temporal sequence of acquisition is $\pm\alpha \leq \pm\gamma \leq \pm\beta$. However, a division into $\pm\beta$ happens twice. As stated earlier, I do not expect it to be impossible for features to be postulated in multiple positions, but this is dispreferred, as without limits this weakens the principles of monotonicity argued to help guide feature postulation. However, if there is an order relation between positive and negative values of a given feature, then the $\pm\beta$ division operates over a contiguous segment of the lattice defined by the previously postulated features. Arbitrarily taking $+ \leq -$ to be the relation, this can be seen below:

(35) $[+\alpha, +\gamma] \leq \underline{[+\alpha, -\gamma] \leq [-\alpha]}$

The division operates over the underlined segment of the poset. If this analysis is along the right tracks, it suggests a limited ability to look into categorial splits of different granularities when postulating new ones.

3.4 Conclusion

This section has looked at the literature presenting implementations for Neo-Emergentist category formation, and contributed to these suggestions, with a focus on restricting the powerful nature of the proposed methods.

[27]Sign languages differ with regards to linear order in their capability for rightward movement and right-linearised specifiers (Cecchetto et al., 2009). However, in that work, FOFC has been tentatively attested in Lingua dei segni italiana. An interesting proposal by Bross (2020) relates syntactic hierarchy to height of articulator in sign languages, which might pattern like linearity in Adjoint modelling.

The properties of DST are notable again here, including the emergence of order across categories, sensitivity to time of acquisition, and self-similar nature across granularity. It is suggested that SD has a preference for binary division, and there are a number of different empirically attested interactions between two features of this type. In particular, some features are independent of one another whereas some are in a dependency relation, which lowers the overall combinations in the system. It is also suggested that Adjoint comparisons, as applied to linguistic theory by Song (2019), have use in modelling and constraining the numerous many-to-one order preserving relations of natural languages. I have not commented on why specific comparisons should arise, and this is necessary in further research to understand how chaos is avoided here. However, the restrictions on order-preservation presented here restrict the occurrences of such comparisons to a subset of the logically possible ones. The suggestions here link SD to typological and acquisitional concerns, including phonological phenomena like vowel harmony. This successive development of features at the same time as the wider grammar is discussed by Nazarov (2014).

4 Explanatory Power of Category Comparison

Having looked at potential mathematical strategies for MMM within feature systems, I will now look at how this may relate to other linguistic and non-linguistic cognitive domains. First, I will continue the discussion of binarity (3.1), looking at how lattice structures could unify feature systems and Merge. Feature systems starting from an undivided categorial space can form a meet semilattice (Appendix A) if SD is $\leq$ the original category space. A syntactic derivation with a root can be seen as a join semilattice, as every two nodes in a derivation have a join, a maximal projection in the tree. This contrasts with defining Merge as operating as a free magma over the set of syntactic features (Watumull, 2015; Van Steene, 2021). The semilattice view adopted here is less powerful than the free magma view, but suggests some symmetry between feature systems and structure derivations. The associative axiom of lattices is required somewhere in syntax, as Merge is sensitive to hierarchical order distinctions, for example the difference between a complement and a specifier. However, associativity may be applied externally to basic structure building in a framework assuming free operation of Merge (Watumull and Roberts, 2023).

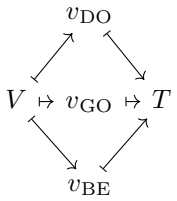
This links to the structure of c-command. C-command is somewhat unusual as the closure of a symmetrical relation (sisterhood) and an asymmetrical one (daughterhood). The same relation has been observed in multiple domains here, including (18), where lexicalisation cannot occur between elements in a c-command relation, and (34), where feature postulation may make reference to the c-command domain of features derived by SD. Song (2019) has suggested that c-command can be non-stipulatively derived from the Yoneda Lemma. In Neo-Emergentist theory, this can capture the emergent properties relating to scope that arise in a full linguistic system, and explains why this relation should recur in cognition.

Song (2019, 2020) has looked at the applications of Adjoint comparison to Phase Theory (see 2.2.4). For Song's comparison, phase boundaries can be defined as different contiguous segments of the EPs. As EPs are left-Adjoint to the US in Song's model, Song proposes that all phase heads must be indivisible. This theorem is intended as an argument against flavoured phase heads. Flavouring of heads refers to heads of the same general category, but with formal differences, relating to lexical semantics. Cuervo (2003)'s (16) ((36) below) shows some syntactic differences between v categories in Spanish. The divisions are not directly hierarchical in nature.

(36)

	vDO	vGO	vBE
character of event	dynamic, agentive	dynamic	stative
can take subject DP?	via voice	no	yes (+root)
can take object DP?	yes (via root)	yes	yes
can embed a vP?	yes, all types	yes, vPBE	no
can embed a small clause?	yes	yes	yes

Flavoured heads have also been proposed for other categories like number (Mathieu, 2012). In Neo-Emergentist theory, if they do exist, they should not be derived from each other by SD. Song's theorem bans flavoured v's, as in (37), if v's are to be taken as non-dynamic phase heads (Chomsky, 2001), as their colimit/join is T, which would be in a different phase.

(37)


As the model I present no longer requires EPs to be left-Adjoint to a US, there is no restriction against any categories being flavoured[28]. This makes this model less useful for making predictions about phase theory. However, future research could look into the link to dynamical phases (Bošković, 2014), which fits in naturally with Neo-Emergentist thinking, or the relation between granularity and the information-transfer role of phases.

In addition to arguments about core linguistic structure, Neo-Emergentist theory could model connections to other cognitive domains. It is hard to meaninfully delimit syntax from other domains, which has received increasing attention in recent years. One neighbouring domain is that of gesture accompanying spoken language: in a sample of spoken language by Nobe (2000), over 90% of spoken propositions involve gestures, and in many cases gestures are obligatory for an utterance to be felicitous. This has been incorporated into syntactic work (Jouitteau, 2004; Colasanti, 2023). The written modality also poses challenges, especially in the modern era with advancements such as emojis. Song (2019, 2024) has looked at the wider semiotics of emojis in syntactic structure using Category Theory, specifically phrase-final emojis and their relation to Sentence-Final Particles (SFP)s. SFPs have attracted interest due to their phonological and linearisation properties. In particular, they are notorious surface violators of the FOFC (Biberauer, 2017)[29]. Biberauer's approach is that these elements do not form part of an EP, due to their ability to attach to different categories and freedom of positioning and semantic meaning. If FOFC is defined over EPs, then this approach does not lead to a violation. Song (2019) further suggests that the availability of certain SFPs relates to the phonology of the language. This can be seen clearly in Mandarin Chinese, where SFPs are toneless vowels. This leads to inherent restrictions on the availability of such SFPs in the spoken modality, and predicts the ability of emojis to take on this role in the written modality, due to their orthographic idiosyncracies.

More widely, Emergentist linguistic theory has seen radical expansions to nonlinguistic domains. Song (2024)'s Category-theoretic analysis extends to semiotic information like memes and background music in social media posts. This relates to Super-linguistics work that investigates human cognition using linguistic techniques, in areas as varied as interpretive-dance, yoga, and resistance-training (i.a. Schlenker, 2018; Patel-Grosz and Mascarenhas, 2023; Hess and Napoli, 2008; Esipova, 2022). I make no comment on these individual domains, but the model presented here highlights the domain-general cognitive techniques used in allowing emergence and restricting chaos in hierarchical systems, and so application to other cultural endeavours which display structure without tending to chaos is an intriguing direction for future research.

[28] I make no comment on the usefullness of flavoured heads.

[29] There are several approaches to account for this other than Biberauer's. Song (2019) proposes that SFPs and their associated clause are in a (linguistic) adjunction relationship. Erlewine (2017) proposes that SFPs obligatorily occur at the bottom of phases, and FOFC is defined over phases.

As a final point on the freedom this model predicts, I look at potential explanatory power for studying first language acquisition in people with autism, which has not been a focus of MGG research[30]. While non-Generativist research into the language usage of autistic children has primarily focused on pragmatics, syntactic development has also been reviewed (Eigsti et al., 2011; Park et al., 2012). Multiple studies have found deficits in certain autistic children in the production of aspect (Chen et al., 2023) and tense (Barlotucci and Albers, 1974; Faroqi-Shah and Thomspon, 2007). The finding that morphemes can be learned in unexpected orders is particularly interesting. In particular, past tense markers in English can be acquired later than expected compared to other morphemes (Barlotucci et al., 1980). The Temporal Binding Deficit hypothesis of autism (Brock et al., 2002) has been invoked as a possible explanation. This suggests that people with autism have weak central coherence, in that they have some failure to integrate information from different specialised networks within the brain. This deficit would inhibit temporal binding, which allows contextual information to be integrated by the simultaneous firing of neurons to express composite information. This predicts people with autism should see larger objects as more of a collection of parts than a whole. For example, processing of individual words is unimpaired in people with autism, but context is not used successfully in resolving homography or ambiguity (i.a. Frith and Snowling, 1983; Eskes et al., 1990; Jolliffe and Baron-Cohen, 1999). The model presented here proposes that third-factor techniques are used to make emergent connections between objects in feature systems and structure derivation. It is entirely logical that weak central coherence could lead to less successful integration in autistic people, leading to the observed outputs in different contexts. This is harder to explain in approaches presupposing components of UG, such as a US, where connections should be innately wired. I leave it as a question for further research whether the techniques of mathematical comparison suggested here can make specific predictions about neurodivergence and language acquisition.

5 Conclusion

This dissertation contributes to the literature on Neo-Emergentism, across linguistic subfields and in human cognition more widely. It follows other Neo-Emergentist works in maintaining a commitment to linguistic minimalism (Chomsky, 1995), noting that minimalism needs to be defined over the entire linguistic system and the child's task of acquiring an adult grammar which does not overgenerate, rather than just over the factors of cognition. This dissertation particularly highlights third factor proposals, proposing that these alone can generate the adult grammars acquired. It argues that assuming even minimal content for UG, in the form of a Universal Spine or similar, is superfluous, and even Merge plausibly arises from non-linguistic third factors. The mathematical possibilities for Maximising Minimal Means without causing rampant variation are investigated from a DST perspective. The particular proposals are that Successive Division as proposed by Dresher (2003) is sensitive to several parameters of variation to be maximally efficient, and that Song (2019)'s approach of comparing larger linguistic categories with Adjunctions can give insight into Neo-Emergent theory. This latter point is particularly relevant for acquisition and typology of phonological and syntactic phenomena, and investigating the possible limits of these. Finally, I looked at the potential for future research in this domain, arguing that Neo-Emergent theory can model human cognition and the uses of language with a wider scope than other approaches in MGG. Further research could look at applying these proposals, as well as others in the Neo-Emergentist literature, to wider typological surveys of languages, or more difficult case studies, as current work focuses on simpler cases.

[30] I use autism as a case study here, leaving more general neurodiversity for further research.

Acronyms

ADJ Adjective. 19

Agr Agreement. 17

Appl Applicative. 18

Asp Aspect. 17, 18

Cl Classifier. 18

COP Copula. 19

DEF Definite. 11

DEM Demonstrative. 11

Det Determiner. 18

DST Dynamical Systems Theory. 3, 4, 7, 11, 21, 23

EP Extended Projection. 9–11, 17–19, 21, 22, 35

F Feminine. 19

Fin Finite. 18

Foc Focus. 18

FOFC Final-over-Final Condition. 10, 11, 19, 20, 22

Gen Genitive. 18, 19

INF Infinitive. 7

K Case. 18

M Masculine. 19

MEP Mixed Extended Projection. 10, 11

MGG Mainstream Generative Grammar. 3, 5, 6, 13, 23

MMM Maximise Minimal Means. 6, 21

Mod Modality. 18

Nat Natural Transformation. 37, 38

NEP Nominal Extended Projection. 9, 10, 17, 18

NMLZ Nominaliser. 11

NOM Nominative. 19

Num Number. 17, 18

PLD Primary Linguistic Data. 5, 6, 11

poset Partially Ordered Set. 19, 35, 36

PST Past. 19

Q Quantifier. 18

SA Speech Act. 18

SD Successive Division. 5, 11, 14–16, 20–23

SFP Sentence-Final Particles. 22

Subj Subjunctive. 19

Top Topic. 18

UG Universal Grammar. 3, 5, 6, 8, 14, 23

US Universal Spine. 6–9, 18, 21–23

VEP Verbal Extended Projection. 9–11, 17, 18

VPlace Vowel Place. 9, 11

References

Abler, W. (1989). On the Particulate Principle of Self-Diversifying Systems. *Journal of Social and Biological Structures*, 12:1–13.

Abney, S. P. (1983). *The English Noun Phrase in its Sentential Aspect*. PhD thesis, Massachusetts Institute of Technology.

Adger, D. and Svenonius, P. (2011). Feature in Minimalist Syntax. In Boeckx, C., editor, *The Oxford handbook of linguistic minimalism*, pages 27–51. Oxford University Press.

Albirini, A. and Benmamoun, E. (2012). Aspects of Second-Language Transfer in the Oral Production of Egyptian and Palestinian Heritage Speakers. *International Journal of Bilingualism*, 18(3):244–273.

Arkadiev, P. and Klamer, M. (2019). Morphological Theory and Typology. In Masini, F. and Audring, J., editors, *The Oxford Handbook of Morphological Theory*, pages 435–54. Oxford University Press.

Backley, P. (2011). *An Introduction to Element Theory*. Edinburgh University Press.

Baker, M. (1996). *The Polysynthesis Parameter*. Oxford University Press.

Barlotucci, G. and Albers, R. J. (1974). Deictic Categories in the Language of Autistic Children. *Journal of Autism and Childhood Schizophrenia*, 4:131–141.

Barlotucci, G., Pierce, S. J., and Streiner, D. (1980). Cross-Sectional Studies of Grammatical Morphemes in Autistic and Mentally Retarded Children. *Journal of Autism and Developmental Disorders*, 10:39–50.

Bazalgette, T. (2015). *Algorithmic Acquisition of Focus Parameters*. PhD thesis, University of Cambridge.

Beggs, J. M. (2008). The Criticality Hypothesis: How Local Cortical Networks Might Optimize Information Processing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, pages 329–343.

Berwick, R. (2011). All you Need is Merge: Biology, Computation, and Language from the Bottom up. In Di Sciullo, A. M. and Boeckx, C., editors, *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*, pages 461–491. Oxford University Press.

Berwick, R. and Chomsky, N. (2016). *Why Only Us: Language and Evolution*. MIT Press.

Biberauer, T. (2017). The Final-over-Final Condition and Particles. In Sheehan, M., Biberauer, T., Holmberg, A., and Roberts, I., editors, *The Final-Over-Final Condition: A Syntactic Universal*, chapter 9, pages 187–296. MIT Press.

Biberauer, T. (2018). Beyond Word Order - and Syntax: what we can learn from FOFC. Presentation at University of Göttingen.

Biberauer, T. (2019a). Children always go beyond the Input: The Maximise Minimal Means Perspective. *Theoretical Linguistics*, 45:211–224.

Biberauer, T. (2019b). Factors 2 and 3: Towards a Principled Approach. *Catalan Journal of Linguistics*, Special Issue: Generative Syntax. Questions, Crossroads, and Challenges:45–88.

Biberauer, T. (2022). The Recycling Imperative: an MMM Perspective on the Nature and Origin of Grammatical Features. Talk at Leiden University.

Biberauer, T. (2023). The Edge of Chaos in Acquisition, Variation and Change: A Minimalist Perspective. Talk at Ulster University.

Biberauer, T. and Bosch, N. (2021). The case for Maximise Minimal Means: Linguistic and General - Cognitive Perspectives. Talk at SyntaxLab, University of Cambridge. 16th November.

Biberauer, T., Holmberg, A., and Roberts, I. (2014a). A Syntactic Universal and Its Consequences. *Linguistic Inquiry*, 45(2):169–225.

Biberauer, T. and Roberts, I. (2015). Rethinking Formal Hierarchies: a Proposed Unification. *Cambridge Occasional Papers in Linguistics*, 7:1–31.

Biberauer, T. and Roberts, I. (2017). Conditional Inversion and Types of Parametric Change. In de Han, P. and Los, B., editors, *Word Order Change in Acquisition and Language Contact*, chapter 4, page 57–77. John Benjamins Publishing Company.

Biberauer, T., Roberts, I., and Sheehan, M. (2014b). No-Choice Parameters and the Limits of Syntactic Variation. In LaBarge, R. E. S., editor, *Proceedings of the 31st West Coast Conference on Formal Linguistics*, pages 46–55. Cascadilla Proceedings Project.

Biberauer, T. and Rowe, J. (2025). FOFC in 2025. Unpublished Manuscript. University of Cambridge.

Bittner, M. and Hale, K. (1996). The Structural Determination of Case and Agreement. *Linguistic inquiry*, 27:1–68.

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

Boisseau, G. and Gibbons, J. (2018). What you needa know about Yoneda: Profunctor Optics and the Yoneda Lemma (Functional Pearl). *Proceedings of the ACM on Programming Languages*, 2:1–27.

Borsley, R. and Kornfilt, J. (2000). Mixed extended projections. In Borsley, R., editor, *The nature and function of syntactic categories*, pages 101–131. Brill.

Bosch, N. (2022). Emergence, Complexity and Developing Grammars: a reinterpretation from a Dynamical Systems perspective. *Cambridge Occasional Papers in Linguistics*, 14(1):1–33.

Bosch, N. (2023). Emergent Syntax and Maturation: a neo-emergentist approach to syntactic development. Master's thesis, University of Cambridge.

Bosch, N. and Biberauer, T. (2024). Emergent Syntactic Categories and Increasing Granularity: Evidence from a Multilingual Corpus Study. In AlThagafi, H. A. A. and Ray, J., editors, *Proceedings of the 48th Boston University Conference on Language Development (BUCLD)*, pages 101–116.

Bouchon, C. and Toro, J. (2019). Is the Consonant Bias specifically Human? Long–Evans Rats encode Vowels better than Consonants in Words. *Animal Cogition*, 22:839–850.

Bošković, Z. (2005). On the Locality of Left Branch Extraction and the Structure of NP. *Studia Linguistica*, 59:1–45.

Bošković, Z. (2014). Now I'm a Phase, now I'm not a phase: On the Variability of Phases with Extraction and Ellipsis. *Linguistic inquiry*, 45:27–89.

Brentari, D. (2019). *Sign Language Phonology*. Cambridge University Press.

Brock, J.and Brown, C., Boucher, J., and Rippon, G. (2002). The Temporal Binding Deficit Hypothesis of Autism. *Development and psychopathology*, 14:209–224.

Broe, M. (1992). An Introduction to Feature Geometry. In Docherty, G. and Ladd, R., editors, *Papers in Laboratory Phonology II: Gesture, segment, prosody*, pages 149–165. Cambridge University Press.

Bross, F. (2020). *The clausal syntax of German Sign Language: A cartographic approach (Volume 5)*. Language Science Press.

Bruening, B. (2009). Selectional Asymmetries between CP and DP Suggest that the DP Hypothesis is Wrong. *University of Pennsylvania Working Papers in Linguistics*, 15:25–35.

Cabredo Hofherr, P. and Doetjes, J. (2021). Introduction. In Cabredo Hofherr, P. and Doetjes, J., editors, *The Oxford Handbook of Grammatical Number*, chapter 1, pages 1–12. Oxford University Press.

Cavirani-Pots, C., de Villiers, E., and Jaspers, D. (2023). More Species in the Affixal Forest. *Tijdschrift voor nederlandse taal*, 139:1–28.

Cecchetto, C., Geraci, C., and Zucchi, S. (2009). Another way to mark Syntactic Dependencies: The Case for Right-Peripheral Specifiers in Sign Languages. *Language*, 85:278–320.

Chen, L., He, X., and Durrleman, S. (2023). Acquisition of Grammatical Aspect by Mandarin-Speaking Preschool Children with Autism Spectrum Disorder. *International Journal of Language I& Communication Disorders*, 58:1697–1716.

Cheng, L. L.-S. and Sybesma, R. (1998). Yi-wan tang, yi-ge tang: Classifiers and Massifiers. *Tsing Hua journal of Chinese studies*, 28:385–412.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

Chomsky, N. (1970). Remarks on Nominalization. In Jacobs, R. and Rosenbaum, P., editors, *Readings in English Transformational Grammar*, pages 11–61. Ginn and Co.

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

Chomsky, N. (2001). Derivation by phase. In Hale, K., editor, *Ken Hale: A life in language*, pages 1–52. MIT Press.

Chomsky, N. (2004). Beyond explanatory adequacy. In Belletti, A., editor, *Structures and Beyond. The Cartography of Syntactic Structures*, volume 3, page 104–131. Oxford University Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36:1–22.

Chomsky, N., Seely, T. D., Berwick, R. C., Fong, S., Huybregts, M. A. C., Kitahara, H., McInnerney, A., and Sugimoto, Y. (2023). *Merge and the strong minimalist thesis*. Cambridge University Press.

Christophe, A., Mehler, J., and Sebastián-Gallés, N. (2001). Perception of Prosodic Boundary Correlates by Newborn Infants. *Infancy*, 2:385–394.

Cinque, G. (1999). *Adverbs and Functional Heads. A Cross-Linguistic Perspective*. Oxford University Press.

Cinque, G. (2022). On Two Sub-Projections of the Nominal Extended Projection: some Romance Evidence. *Probus*, 34:55–75.

Citko, B. (2005). On the Nature of Merge: External Merge, Internal Merge, and Parallel Merge. *Linguistic Inquiry*, 36(4):475–496.

Colasanti, V. (2023). Functional gestures as morphemes: Some evidence from the languages of southern italy. *Glossa*, 8.

Corver, N. (2013). Lexical Categories and (Extended) Projection. In den Dikken, M., editor, *The Cambridge Handbook of Generative Syntax*, pages 353–424. Cambridge University Press.

Cowper, E. and Hall, D. (2014). Reductio ad Discrimen: Where Features come from. *Nordlyd*, 41:145–164. Special Issue on Features, number 2.

Cuervo, M. C. (2003). *Datives at large.* PhD thesis, MIT.

D'Alessandro, R. and Terenghi, S. (2024). Non-Monotonic Functional Sequences: A New Metric for Complexity in Heritage Languages. In Polinsky, M. and Putnam, M., editors, *Formal approaches to complexity in heritage language grammars*, chapter 7, pages 154–180. Language Sciences Press.

Davey, B. A. and Priestley, H. A. (2002). *Introduction to lattices and order.* Cambridge University Press.

Dresher, B. (2003). Contrast and Asymmetries in Inventories. In di Sciullo, A. M., editor, *Asymmetry in Grammar: Volume 2: Morphology, phonology, acquisition*, pages 239–257. John Benjamins Publishing Company.

Dresher, B. E. (2014). The arch not the stones: Universal feature theory without universal features. *Nordlyd*, 41:165–181. Special Issue on Features, number 2.

Eigsti, I.-M., de Marchena, A. B., Schuh, J. M., and Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, 5:681–691.

Ellis, N. (1998). Emergentism, connectionism and language learning. *Language learning*, 48:631–664.

Erlewine, M. Y. (2017). Low sentence-final particles in mandarin chinese and the final-over-final constraint. *Journal of East Asian Linguistics*, 26:37–75.

Erlewine, M. Y. (2023). A Syntactic Universal in a Contact Language; the Story of Singlish already. In Nomoto, H. and McCready, E., editors, *Discourse Particles in Asian Languages Volume II*, pages 91–120. Routledge.

Esipova, M. (2022). Reps and Representations: A warm-up to a Grammar of Lifting. *Linguistics and Philosophy*, 46(4):871–904.

Eskes, G., Bryson, S. E., and McCormick, T. A. (1990). Comprehension of Concrete and Abstract Words in Autistic Children. *Journal of Autism and Developmental Disorders*, 20:61–73.

Everett, D. (2010). You Drink. You Drive. You Go to Jail. Where's Recursion? Retrieved from https://ling.auf.net/lingbuzz/001141.

Faroqi-Shah, Y. and Thomspon, C. (2007). Verb Inflections in Agrammatic Aphasia: Encoding of Tense Features. *Journal of Memory and Language*, 56:129–151.

Fong, B. and Spivak, D. (2019). *An Invitation to Applied Category Theory: Seven Sketches in Compositionality.* Cambridge University Press.

Frith, U. and Snowling, M. (1983). Reading for Meaning and Reading for Sound in Autistic and Dyslexic Children. *British Journal of Developmental Psychology*, 1:329–342.

Gentner, D. (1982). Why Nouns are Learned before Verbs: Linguistic Relativity versus Natural Partitioning. In Kuszaj, S., editor, *Language development, vol. 2: language, thought, and culture*, pages 301–334. Erlbaum.

Gopnik, A. and Choi, S. (1995). Names, Relational Words, and Cognitive Development in English and Korean Speakers: Nouns are not Always Learned before Verbs. In Tomasello, M. and Merriman, W. E., editors, *Beyond names for things*, pages 63–80. Psychology Press.

Gottlieb, G. (2001). *Individual Development and Evolution: The Genesis of Novel Behavior*. Psychology Press.

Grimshaw, J. (1991). Extended Projection. Master's thesis, Brandeis University.

Grätzer, G. (2008). *Universal algebra*. Springer Science & Business Media.

Guasti, M. T. (2016). *Language Acquisition: The Growth of Grammar*, chapter 3. MIT Press.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press.

Hall, D. C. (2007). *The Role and Representation of Contrast in Phonological Theory*. PhD thesis, University of Toronto.

Harbour, D. (2014). Paucity, Abundance, and the Theory of Number. *Language*, 90(1):185–229.

Harbour, D. (2016). *Impossible Persons*. The MIT Press.

Harbour, D. (2020). Frankenduals: Their Typology, Structure, and Significance. *Language*, 96:60–93.

Harley, H. and Ritter, E. (2002). Person and Number in Pronouns: A Feature-Geometric Analysis. *Language*, 78(3):482–526.

Hein, J. and Murphy, A. (2022). VP-Nominalization and the Final-over-Final Condition. *Linguistic inquiry*, 53:337–370.

Heine, B. and Kilian-Hatz, C. (1994). Polysemy in African Languages: An Example from Baka. In Geider, T. and Kastenholz, R., editors, *Sprachen un Sprachzeugnisse in Afrika, Eine Sammlung philologischer Beiträge Wilhelm JG Möhlig zum 60*, pages 177–187. Köppe Verlag.

Henry, A. (2005). Non-Standard Dialects and Linguistic Data. *Lingua*, 115:1599–1617.

Hess, S. and Napoli, D. J. (2008). Energy and Symmetry in Language and Yoga. *Leondardo*, 41:332–338.

Hiraiwa, K. (2005a). *Dimensions of Symmetry in Syntax: Agreement and Clausal Architecture*. PhD thesis, MIT.

Hiraiwa, K. (2005b). Predicate Clefts in Bùlì: Categories and Phases. *Linguistic Analysis*, 32:544–583.

Hiraiwa, K. (2017). The Faculty of Language Integrates the two Core Systems of Number. *Frontiers in Psychology*, 8.

Huybregts, M. (2017). Phonemic Clicks and the Mapping Asymmetry: how Language Emerged and Speech Developed. *Neuroscience and Biobehavioral Reviews*, 81:279–294.

Irvin, J., Spokoyny, D., and Moscoso del Prado Martín, F. (2016). Dynamical Systems Modeling of the Child–Mother Dyad: Causality between Child-Directed Language Complexity and Language Development. In *Proceedings of the Annual Meeting of the Cognitive Science Society. (Vol 38)*, volume 38. Retrieved from https://escholarship.org/uc/item/9gd973dk.

Jakobson, R. and Halle, M. (1956). *Fundamentals of Language*. de Gruyter.

Jolliffe, T. and Baron-Cohen, S. (1999). A Test of Central Coherence Theory: Linguistic Processing in High-Functioning Adults with Autism or Asperger Syndrome: Is Local Coherence Impaired? *Cognition*, 71(2):149–185.

Jouitteau, M. (2004). Gestures as Expletives, Multichannel Syntax. In Chand, V., Kelleher, A., Rodríguez, A. J., and Schmeiser, B., editors, *3rd West Coast Conference on Formal Linguistics (WCCFL 23), Apr 2004, Davis, CA, United States*, pages 422–435.

Katz, J. and Pesetsky, D. (2011). The Identity Thesis for Language and Music. retreived from https://ling.auf.net/lingbuzz/000959.

Kayne, R. (1994). *The Antisymmetry of Syntax*. MIT Press.

Kidd, C., Piantadosi, S. T., and Aslin, R. N. (2012). The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *PLoS ONE*, 7(5):e36399.

Kidd, C., Piantadosi, S. T., and Aslin, R. N. (2014). The Goldilocks Effect in Infant Auditory Attention. *Child Development*, 85(5):1795–1804.

Kilani-Schoch, M. and Dressler, W. (2000). Are Fillers as Precursors of Morphemes Relevant for Morphological Theory? In Dressler, W., Pfeiffer, O., Pöchtrager, M., and Rennison, J., editors, *Morphological Analysis in Comparison*, pages 89–112. John Benjamin's Publishing.

Kim, S.-W. (2011). A Note on NP/DP Parameter: Left Branch Extraction in korean. *Linguistic Research*, 28:257–269.

Krekoski, R. (2013). On tone and the nature of features. Master's thesis, University of Toronto.

Krekoski, R. (2017). *Contrast and complexity in Chinese tonal systems.* PhD thesis, University of Toronto.

Larsen-Freeman, D. (1997). Chaos/Complexity Science and Second Language Acquisition. *Applied Linguistics*, 18:141–165.

Lass, R. (1990). How to do Things with Junk: Exaptation in Language Evolution. *Journal of linguistics*, 26:79–102.

Lebeaux, D. (1988). *Language Acquisition and the Form of the Grammar*. PhD thesis, University of Massachusetts.

Leeuwenberg, E. and Van der Helm, P. (2013). *Structural Information Theory: The Simplicity of Visual Form*. Cambrudge University Press.

Leinster, T. (2014). *Basic Category Theory*. Cambridge University Press.

Leivada, E. and Barceló-Coblijn, L. (2021). Why Don't Languages Grammaticalize [±poisonous]? *Biolinguistics*, 14 (Special Issue):51–58.

Liddell, S. (1984). THINK and BELIEVE: Sequentiality in American Sign Language. *Language*, pages 372–399.

Longobardi, G. (2001). The Structure of DPs: some Principles, Parameters and Problems. In Baltin, M. and Collins, C., editors, *The Handbook of Contemporary Syntactic Theory*, pages 562–603. Blackwell.

Los, B. (2013). Recycling "Junk": A Case for Exaptation as a Response to Breakdown. In Kikusawa, R. and Reid, L. A., editors, *Historical Linguistics 2011: Selected papers from the 20th International Conference on Historical Linguistics, Osaka, 25-30 July 2011. Current Issues in Linguistic Theory, vol. 326*, pages 267–288.

Lynch, J. (2019). The Bilabial-to-Linguolabial Shift in Southern Oceanic: A Subgrouping Diagnostic? *Oceanic Linguistics*, 58:292–323.

Maddieson, I. (2013). Front Rounded Vowels (v2020.4). In Dryer, M. S. and Haspelmath, M., editors, *The World Atlas of Language Structures Online*. Zenodo.

Malčić, J. (2019). The Asymmetry and Antisymmetry of Syntax. Master's thesis, University of Cambridge.

Marantz, A. (1997). No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon. *U. Penn Working Papers in Linguistics*, 4:201–225.

Mathieu, É. (2012). Flavors of Division. *Linguistic Inquiry*, 43:650–679.

McCarthy, J. (1998). Feature Geometry and Dependency: A Review. *Phonetica*, 43:84–108.

Megerdoomian, K. (2008). Parallel nominal and verbal projections. In Keyser, S. J., editor, *Foundational Issues in Linguistic Theory*. MIT Press.

Mielke, J. (2008). *The Emergence of Distinctive Features*. Oxford University Press.

Minsky, M. (1967). *Computation: Finite and Infinite Machines*. Englewood Cliffs.

Minsky, M. (1985). Why Intelligent Aliens will be Intelligible. In *Extraterrestrials. Science and Alien Intelligence*, pages 117–128. Cambridge University Press.

Miyagawa, S. (2009). *Why Agree? Why Move?: Unifying Agreement-Based and Discourse-Configurational Languages*. MIT Press.

Moran, S. (2012). *Phonetics Information Base and Lexicon*. PhD thesis, University of Washington.

Napoli, D. J. and Sutton-Spence, R. (2010). Limitations on Simultaneity in Sign Language. *Language*, 86:647–662.

Nazarov, A. (2014). A Radically Emergentist Approach to Phonological Features: Implications for Grammars. *Nordlyd*, 41:21–58. Special Issue on Features.

Nespor, M., Peña, M., and Mehler, J. (2003). On the Different Roles of Vowels and Consonants in Speech Processing and Language Acquisition. *Lingue e linguaggio*, 2:203–230.

Nobe, S. (2000). Where do Most Spontaneous Representational Gestures actually Occur with Respect to Speech. In McNeill, D., editor, *Language and gesture*, chapter 9, pages 186–198. Cambridge University Press.

Padgett, J. (2002). Feature Classes in Phonology. *Language*, 78:81–110.

Panagiotidis, P. (2014). *Categorial Features*. Cambridge University Press.

Park, C., Yelland, G., Taffe, J., and Gray, K. (2012). Morphological and Syntactic Skills in Language Samples of Pre School aged Children with Autism: Atypical Development? *International Journal of Speech-Language Pathology*, 14:95–108.

Patel-Grosz, P. and Mascarenhas, S Chemla, E. (2023). Super Linguistics: an Introduction. *Linguistics and Philosophy*, 46:627–692.

Pesetsky, D. (2013). *Russian Case Morphology and the Syntactic Categories*. MIT Press.

Peters, A. M. and Menn, L. (1997). False starts and filler syllables: Ways to learn grammatical morphemes. *Language*, pages 742–777.

Pfau, R. and Quer, J. (2012). Nonmanuals: their Grammatical and Prosodic Roles. In Brentari, D., editor, *Sign Languages*, chapter 17, pages 381–402. Cambridge University Press.

Pfau, R., Salzmann, M., and Steinbach, M. (2018). The Syntax of Sign Language Agreement: Common Ingredients, but Unusual Recipe. *Glossa*, 3:107.

Piattelli-Palmarini, M. and Vitiello, G. (2017). Third Factors in Language Design: Some Suggestions from Quantum Field Theory. In McGilvray, J., editor, *The Cambridge Companion to Chomsky*, chapter 6, page 134–152. Cambridge University Press, 2 edition.

Poeppel, D. and Embick, D. (2005). Defining the Relation between Linguistics and Neuroscience. In Cutler, A., editor, *Twenty-first century psycholinguistics: Four Cornerstones*, pages 103–118. Routledge.

Polanyi, L. and Scha, R. (1983). On the Recursive Structure of Discourse. In Ehlich, K. and van Riemsdijk, H., editors, *Connectedness in Sentence, Discourse and Text*, pages 141–178. Tilburg University.

Ramchand, G. and Svenonius, P. (2014). Deriving the Functional Hierarchy. *Language Sciences*, 46:152–174.

Riehl, E. (2014). *Category Theory in Context*. Dover Publications.

Rijkhoff, J. and van Lier, E. (2013). *Flexible Word Classes: Typological Studies of Underspecified Parts of Speech.* Oxford Univesrity Press.

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

Rizzi, L. (1993). Some notes on Linguistic Theory and Language Development: The Case of Root Infinitives. *Language Acquisition*, 3:371–393.

Robert, S. (2018). The Challenge of Polygrammaticalization for Linguistic Theory: Fractal Grammar and Transcategorial Functioning. *Cognitive Linguistic Studies*, 5(1):106–132.

Roberts, I. (2024). Hierarchy and Categories II: Nominals, Features, Case and Agreement. Li9 Lecture 5. Lecture at University of Cambridge, 13th November 2024.

Roberts, I. G., Watumull, J., and Chomsky, N. (2023). Universal grammar. In Vakoch, D. and Punske, J., editors, *Xenolinguistics: Towards a Science of Extraterrestrial Language*, chapter 15, page 165–181. Routledge.

Rose, S. and Walker, R. (2011). Harmony Systems. In Goldsmith, J., Riggle, J., and Yu, A. C. L., editors, *The handbook of phonological theory*, chapter 8, pages 240–290. Blackwell.

Rosenbaum, P. S. (1967). Phrase Structure Principles of English Complex Sentence Formation. *Journal of Linguistics*, 3:103–118.

Ruark, J. L. and Moore, C. A. (1997). Coordination of Lip Muscle Activity by 2-Year-Old Children During Speech and Nonspeech Tasks. *Journal of Speech, Language, and Hearing Research*, 40(6):1373–1385.

Salzmann, M. (2020). The NP vs. DP Debate. Why Previous Arguments are Inconclusive and what a Good Argument Could Look like. Evidence from Agreement with Hybrid Nouns. *Glossa: a journal of general linguistics*, 5(1):83.

Sandler, W. (2014). The Emergence of the Phonetic and Phonological Features in Sign Language. *Nordlyd*, 41:183–212. Special Issue on Features, number 2.

Schlenker, P. (2018). What is Super Semantics? *Philosophical Perspectives*, 32:365–453.

Seuren, P. A. M. and Jaspers, D. (2014). Logico-Cognitive Structure in the Lexicon. *Language*, 90(3):607–643.

Sheehan, M., Biberauer, T., Roberts, I., and Holmberg, A. (2017). *The Final-Over-Final Condition: A Syntactic Universal*. MIT Press.

Shlonsky, U. (1991). Quantifiers as Functional Heads: A Study of Quantifier Float in Hebrew. *Lingua*, 84:159–180.

Skoblikova, E. S. (1971). *Soglasovanie i Upravlenie v Russkom*. Prosveščenie.

Soderstrom, M., Seidl, A., Nelson, D. G. K., and Jusczyk, P. W. (2003). The Prosodic Bootstrapping of Phrases: Evidence from Prelinguistic Infants. *Journal of Memory and Language*, 49(2):249–267.

Song, C. (2019). *On the Formal Flexibility of Syntactic Categories*. PhD thesis, University of Cambridge.

Song, C. (2020). Grammatical Types that Cannot be: A Category-Theoretic Theorem for Chomsky-School Generative Syntax. Talk at 3rd Annual International Applied Category Theory Conference, ACT 2020, Cambridge, USA, 6-10th July 2020.

Song, C. (2024). Sentence-Final Particle vs. Sentence-Final Emoji: The Syntax-Pragmatics Interface in the Era of Computer-Mediated communication. In Haralambous, Y., editor, *Grapholinguistics and Its Applications 9: Proceedings of Grapholinguistics in the 21st Century Conference 2022*, pages 157–192. Fluxus Editions.

Spivak, D. (2014). *Category Theory for the Sciences*. MIT Press.

Steedman, M. (1993). Categorial Grammar. *Lingua*, 90.

Stevens, K., Keyser, S., and Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. In Perkell, J. and Klatt, D., editors, *Invariance and variability in speech processes*, pages 426–463. Lawrence Erlbaum.

Tardif, T. (1996). Nouns Are Not Always Learned Before Verbs: Evidence from Mandarin Speakers' Early Vocabularies. *Developmental Psychology*, 32:492–504.

Tegmark, M. (2014). *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality*. Penguin Books.

Thelen, E. and Smith, L. B. (2007). Dynamic Systems Theories. In *Handbook of Child Psychology, vol. I. Theoretical Models of Human Development*, chapter 6, pages 358–312. Wiley.

Tiflit, A. and Voeltzel, L. (2014). From Binary Features to Elements: The Case of Scandinavian. *Nordlyd*, 41:59–83. Special Issue on Features.

Tomasello, M. (1992). The Social Bases of Language Acquisition. *Social development*, 1:67–87.

Törkenczy, M. (2011). Hungarian vowel harmony. In van Oostendorp, M., Ewen, C. J., Hume, E., and Rice, K., editors, *The Blackwell companion to phonology*, volume 5, pages 2963–2990. John Wiley and Sons.

Tsimpli, I. M. (2014). Early, Late or Very Late?: Timing Acquisition and Bilingualism. *Linguistic Approaches to Bilingualism*, 4(3):283–313.

Tucker, M. and Hirsh-Pasek, K. (1993). Systems and Language: Implications for Acquisition. In Thelen, E. and Smith, L. B., editors, *A Dynamic Systems Approach to Development*, page 359–384. The MIT Press.

Van Geert, P. (1991). A Dynamic Systems Model of Cognitive and Language Growth. *Psychological Review*, 98:3–53.

van Gelder, T. (1998). The Dynamical Hypothesis in Cognitive Science. *Behavioral and Brain Sciences*, 21(5):615–628.

Van Steene, L. (2021). Minimax Feature Merge: The Featural Linguistic Turing Machine. In Wiernik, L., editor, *Proceedings of ULAB 2021. University of Aberdeen 16-18th April 2021*, pages 128–149.

Veneziano, E. (2003). The Emergence of Noun and Verb Categories in the Acquisition of French. *Psychology of Language and Communication*, 7:23–36.

Veneziano, E. and Sinclair, H. (2000). The Changing Status of "Filler Syllables" on the Way to Grammatical Morphemes. *Journal of Child Language*, 27:1–40.

Von Humboldt, W. (1836). *Über die Verschiedenheit des Menschlichen Sprachbaues*. Königlichen Akademie der Wissensehaften.

Waldrop, M. (1993). *Complexity: The Emerging Science at the Edge of Order and Chaos*. Simon and Schuster.

Watumull, J. (2012). A Turing Program for Linguistic Theory. *Biolinguistics*, 6:222–245.

Watumull, J. (2015). *The Linguistic Turing Machine*. PhD thesis, University of Cambridge.

Watumull, J. and Roberts, I. (2023). Rebuttal to "Merge Is Not 'Lerge'". *Biolinguistics*, 17:1–5.

Wexler, K. (1998). Very Early Parameter Setting and the Unique Checking Constraint: A New Explanation of the Optional Infinitive Stage. *Lingua*, 106(1–4):23–79.

Wiltschko, M. (2008). The Syntax of Non-Inflectional Plural Marking. *Natural Language and Linguistic Theory*, 26:639–694.

Wiltschko, M. (2014). *The Universal Structure of Categories*. Cambridge University Press.

Zajdó, K. and Stoel-Gammon, C. (2003). The Acquisition of Vowels in Hungarian: Developmental Data. In Solé, M. J., Recasens, D., and Romero, J., editors, *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*, volume 3, pages 2229–2232.

Zwart, J.-W. (2011a). Recursion in Language: a Layered-Derivation Approach. *Biolinguistics*, 5:43–56.

Zwart, J.-W. (2011b). Structure and Order: Asymmetric Merge. In Boeckx, C., editor, *The Oxford handbook of linguistic minimalism*, pages 96–118. Oxford University Press.

Zwicky, A. and Pullum, G. (1987). Plain Morphology and Expressive Morphology. In Aske, J., Beery, N., Michaelis, L., and Filip, H., editors, *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, pages 330–340. Linguistic Society of America.

Appendices

A Order Theory

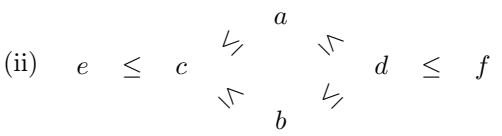
A *set* S is a collection of things/elements[31], which has no further structure. Sets can be equipped with *binary relations*, which relate elements of one set X to those of another Y. A *Function* $y = f(x)$ is a binary relation which assigns one element of Y to each element of X.
A Partially Ordered Set (poset) is a set equipped with the *binary relation* $\leq$. For elements a, $b \in X$, where X is a set, there can be $a \leq b$, $b \leq a$, or no relation. $\leq$ has the following properties for all elements a, b, $c \in X$:

(i) a. *Transitivity*: If $a \leq b$ and $b \leq c$ then $a \leq c$.

b. *Antisymmetry*: if $a \leq b$ and $b \leq a$ then $a = b$ - no two distinct elements may precede each other.

c. *Reflexivity*: $a \leq a$ - every element is related to itself.

Transitivity is necessary for the relation to generalise over more than two categories. Antisymmetry in syntax is discussed by Malčić (2019)[32], and prevents the same element from appearing in two non-contiguous places in linguistic hierarchies. Reflexivity allows for edge elements to be treated the same way as internal ones. For example, the join (see below) of an entire subset satisfies the $\leq$ relation with regards to itself, whereas it is not $<$ itself. This is useful for generalising to identity arrows in Category Theory (Appendix B).
An EP, as introduced in main text Section 2.2.3, can be represented as a poset. If every two elements of these a, b, must form an ordered pair, this is a totally ordered set, meaning there must be either $a \leq b$ or $b \leq a$. By contrast, main text Section 4 discusses flavoured categories, which are modelled with partial instead of total orders, as some pairs of elements are not comparable with $\leq$.
Two elements may, but must not necessarily, have a join (lowest upper bound) and meet (greatest lower bound):

(ii)


In (ii), the join of a and b is d, and the meet is c. Formally, the meet m of a and $b \in X$, if it exists, has the following properties:

(iii) a. $m \leq a$ and $m \leq b$

b. For all $y \in X$, if $\leq a$ and $m \leq b$, then $y \leq m$ - m is the greatest element satisfying (iii. a.).

The join can be similarly defined. For a poset P, if all pairs of elements have a meet and a join, this is a *lattice* over that set. If all pairs of elements have a meet but not necessarily a join, this is a *meet semilattice*, the reverse case is a *join semilattice*. The definition of lattice concepts here has proceeded from Order Theory (Davey and Priestley, 2002), although they can also be defined from Universal Algebra, a distinct subfield of mathematics which studies algebraic structures (Grätzer, 2008). Meets and joins (represented here with the notation $\wedge$ for meet) satisfy the following axioms, defined for elements *a, b, c* $\in$ in a set S:

(iv) a. *Associativity*: x $\wedge$ (y $\wedge$ z) = (x $\wedge$ y) $\wedge$ z - changing the order of operations does not change the result.

b. *Commutativity*: x $\wedge$ y = y $\wedge$ x - changing the order of operands does not change the result.

c. *Idempotency*: x $\wedge$ x = x - applying the relation to an element and itself does not change the element.

[31] Distinct from the usage of Element in Element Theory (Backley, 2011), which is instead inspired by chemical elements.
[32] Kayne (1994)'s antisymmetry refers to asymmetry, which does not have the reflexive property.

The only other algebraic structure used in this dissertation (in main text Section 3.1) is a free magma, whose relation is represented •. This is the most general structure that can be defined over a set. The only binary relation is that of closure:

(v) *Closure*: a • b ∈ S - the result of the relation on members of a set also a member of the set.

Returning to order theory, a function f between ordered sets is called *monotonic* if it preserves or reverses the $\leq$ relations. Formally, for every a, $b \in X$, where $a \leq b$, a monotonic function has all $f(a) \leq f(b)$ (increasing) or all $f(b) \leq f(a)$ (decreasing). For two posets, the posetal version of an Adjunction (Appendix B) is a *Galois connection*. For posets A and B, a Galois connection between them consists of two monotonic functions[33] f: $A \to B$ and g: $B \to A$ such that for all $a \in A$ and $b \in B$:

(vi) f(a) ≤ b iff a ≤ g(b).

f is the lower Adjoint, and g is the upper Adjoint. After a round trip between the two functors, a can only get bigger or stay the same, and b can only get smaller or stay the same:

(vii) a. a ≤ g(f(a))

b. f(g(b)) ≤ b

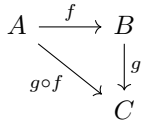
[33]Related to the above discussion of monotonicity, a monotone Galois connection uses order-preserving functions, and an antitone one uses order-reversing functions.

B Category Theory

Works used in writing this Appendix include Fong and Spivak (2019); Spivak (2014); Leinster (2014); Riehl (2014) and Song (2019, Section 6.3). In particular, Song's section introduces the notions used here in a linguistic context. This appendix only introduces concepts relevant for the dissertation.
Category Theory is a subfield of mathematics which generalises concepts found in other areas. A category $\mathcal{C}$ consists of:

- A collection of *objects*, represented *A, B, C...*
- A collection of *morphisms/arrows*, or maps between objects, represented *f, g, h...*
- Operations to assign a source **src** and a target **tar** to each morphism. If $A =$ **src** f and $B =$ **tar** f, this is represented $f\colon A \to B$.
- An operation for composition of morphisms. For $f\colon A \to B$ and $g\colon B \to C$, this is represented $g \circ f\colon A \to C$.
- An identity arrow for each object A, represented $1_A\colon A \to A$.

Category Theory uses commutativity diagrams for illustrations. In these, objects and morphisms (arrows) are displayed. A commutativity diagram for composition is given below:

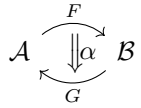
(viii)


I omit all identity morphisms and the labels of some other morphisms for readability in diagrams in the dissertation. All labels for morphisms in this dissertation are arbitrary and unimportant, and do not relate beyond individual diagrams. Commutativity is the equivalence of paths between a pair of objects, understood with composition of morphisms. For example, (viii) shows that the morphism $g \circ f$ is equivalent to the morphism f followed by the morphism g.
A few important categories are introduced here. The category **Set** has all sets as objects and functions between sets as arrows. The category **Pos** has all posets as objects and monotone functions between them as arrows. A single poset can be viewed as a category with elements as objects and the $\leq$ relation as a morphism between objects. Notation of posets in the dissertation varies between order-theoretic notation and the current conceptualisation of them as a category. As every two objects in a poset can only have one morphism between them, posetal categories are called thin.
A *functor* is a mapping between categories, which preserves objects and morphisms, including compositions and identity morphisms. It is represented $F\colon \mathcal{A} \to \mathcal{B}$. For posetal categories, a functor is simply a monotone function:

(ix) $\mathcal{A} \xrightarrow{F} \mathcal{B}$

The *hom-set* hom(A,B) is the collection of morphisms from objects A to B. For all categories considered here, hom-sets are true sets as opposed to proper classes[34]. Such categories are called *locally small*.
Given multiple functors between two categories $F\colon \mathcal{A} \to \mathcal{B}$ and $G\colon \mathcal{A} \to \mathcal{B}$, a natural transformation is a morphism between functors respecting the internal structure of functors $\alpha\colon F \Rightarrow G$. The natural transformations between functors F, G can also be notated Nat(F, G):

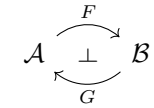
[34]Proper classes can very informally be defined as large set-like collections, which cannot be treated formally as sets to avoid paradoxes.

(x) 

For any morphism *f:* $A \to X$ between objects in $\mathcal{A}$, the following diagram commutes in $\mathcal{B}$:

(xi)
$$\begin{array}{ccc} F(A) & \xrightarrow{F(f)} & F(X) \\ {\scriptstyle \alpha_A}\downarrow & & {\scriptstyle \alpha_X}\downarrow \\ G(A) & \xrightarrow{G(f)} & G(X) \end{array}$$

An *Adjunction*[35] is a form of comparison between categories. It is a weaker form of comparison than equivalence. For categories and functors *F:* $\mathcal{A} \to \mathcal{B}$ and *G:* $\mathcal{B} \to \mathcal{A}$, F is left-Adjoint to G and G is right-Adjoint to F if there is a hom-set isomorphism $\mathcal{B}(F(A),B) \cong \mathcal{A}(A,G(B))$ which is natural for $A \in \mathcal{A}$ and $B \in \mathcal{B}$:

(xii) 

Explanations of this can be found in other sources, the importance is the generality of Adjunctions across different branches of mathematics. In the posetal categories considered in this dissertation, an Adjunction is a much more intuitive Galois connection, described in Appendix A. Left-Adjoints preserve colimits - the generalisation of joins, and right-Adjoints preserve limits - the generalisation of meets.

The *Yoneda lemma* is discussed in main text Section 4. For a functor F from a locally small category $\mathcal{C}$ to **Set**, each object $C \in \mathcal{C}$ has a *hom-functor* h_C, similar to hom-sets above:

(xiii) $h_C = \mathrm{Hom}(C, -)$

h_C sends an object $X \in \mathcal{C}$ to the set of morphisms *Hom(C,X)* and a morphism $X \to Y$, $Y \in \mathcal{C}$ to the morphism $f \circ -$, which itself sends a morphism g in *Hom(C,X)* to the morphism $f \circ g$ in *Hom(C,Y).*
Yoneda's lemma says that there is the following isomorphism:

(xiv) $\mathrm{Nat}(h_C,F) \cong F(C)$

Mathematical explanation of this should be found from other sources. In this dissertation, a philosophical interpretation of the Yoneda lemma is more important. The Yoneda lemma is 'arguably the most important result in category theory' (Riehl, 2014, pg. 57). This lemma is often described with idioms like 'tell me who your friends are, and I will tell you who you are' (Boisseau and Gibbons, 2018, pg. 1). This refers to the fact that the Yoneda lemma states how an object is uniquely determined up to isomorphism by its relations to other objects and morphisms in the locally small category. For posetal categories, the Yoneda Lemma splits each category into two segments, one of which is the principal lower set (the subposet of all elements less than or equal to it), which Song (2019) relates to scope as used in deriving c-command.

[35] I follow Song (2019) in capitalising this term to avoid confusion with the unrelated linguistic concept.