# On linguistic asymmetry: an investigation of the reality and significance of skewed patterns

Jesper Olof Eric Carlson
Department of Linguistics
Sidney Sussex College
University of Cambridge
2009 – 2010

*This thesis is submitted for the degree of Master of Philosophy*

**Abstract**

Asymmetrical, disharmonic and skewed distributions are widespread in several domains of the human language system as a whole. These patterns, which, if conceived as a logical space generated by a 2-by-2 factorial design, often appear to fall into a skewed, asymmetrical configuration; frequently, a clear-cut 3-out-of-4 pattern can be established but, if not, a general imbalance is usually still predominant. *Inter alia*, examples of such configurations are Holmberg's Generalisation, FOFC, the Square of Opposition (*id est* the nonlexicalisation of the O corner), Romance Interrogatives, Greenberg's Universal 25, Monotonicity and constraints on negation, Subject-Pro-Drop, N-drop, clause syntax and illocutionary force mismatching etc.

The central question to pose is whether such 3/4-patterns underlyingly are significant for the Faculty of Language or whether they simply are epiphenomena without further meaning or implications. In this thesis I argue that linguistic structures are sets of sets of purely privative atomic features which display surface phenomena such as asymmetries or skewings as a consequence of underlying feature interaction; I call this the Atomic Feature Conjecture. I suggest that features are combined by means of Merge and I demonstrate that there are three types of feature relationships: independence, dependence and antagonism. As far as asymmetry is concerned, I define two types of asymmetrical patterns: Type B, simultaneous incompatibility; and Type M, feature dependence. I furthermore outline consequences of these findings for Linguistic Theory, emphasising, on the one hand, the importance of privative features, and, on the other hand, the notion of the empty set as a fundamental component of the computational system. Ultimately, I also address remaining theoretical dilemmas stemming from the proposal and discuss future directions that cross-disciplinary research within the field could take.

Word count: 19951[1]

[1] Excluding: abstract, acknowledgements, table of contents, matrices and illustrations in data set, repeated examples, page headings, page numbers, and bibliography.
Including: section headings and language examples.

# Acknowledgements

First, I would like to express my utmost gratitude to my supervisors: Professor Ian Roberts and Dr Theresa Biberauer for their theoretical insight, invaluable comments, helpful suggestions and cordial encouragements.

Second, I wish to praise Sidney Sussex College of the University of Cambridge, my home for the last four years. As a fantastic and inspiring institution, it has enabled me to reach very far and for this I am for ever thankful.

Third, I express my indebtedness to the following people with whom I have discussed ideas or from whom I have received precious advice during the course of this project: Prof Kasia Jaszczolt, for two years of generous supervision; Dr Bert Vaux for being cool, generous and encouraging; Dr Neil Wright for teaching me Latin over the last three years; Dr Regina Karousou-Fokkas, for her efforts with Modern Greek Grammar instruction; Dr John Williams, for allowing me to attend his class on psychological processes; and, Neil Myler, for his will to read a draft and for his provision of ample suggestions.

Forth, I thank a selection of friends whose support has been crucial in moments of doubt and weakness:
Varunjay Ahluwalia and Sophie Harris, for our dinners and our fun.
Rowena Eason, for being wonderful.
Parmesh Gajjar for inspiration to use Latex, for his stimulating enthusiasm, and for his cooking.
Prakash Kapadia, who, despite being very far away this year, never stopped supporting me.
Olesya Kryshevich, for linguistic team work as well as for all the incessant laughter and banter.
Hattie Pickup, for her wit, concern and many visits during the year.
Jessica Vardy and Ellie Kim for 'co-inhabitational' moments of happiness.
The 'kids' of Portugal Street, whom I have come to cherish, mainly because of their provision of chaos...
My Modern Greek Class – Alexandra, Calum and Grace; thanks for all the fun and for putting up with my linguistic predilection.

Fifth, I am grateful to my parents, Susan and Anders Carlson, thanks to whom I am multilingual, and to my sisters, Jennifer and Johanna Carlson, for being extraordinary. Their trust and reassurance have been dearly appreciated.

Sixth, I praise the secretaries of the department of Linguistics: Rachel Deadman and Louise Radok, for their help with all things administrative.

Last, I wish to mention The Cambridge European Trust, for their generous financial support and belief in me, without which this project would have been impossible.

*Die Sprache ist eine Labyrinth von Wegen. Du kommst von einer Seite und kennst dich aus; du kommst von einer andern zur selben Stelle, und kennst dich nicht mehr aus.*

– Ludwig Wittgenstein,
Philosophische Untersuchungen, §203

# Contents

# 1 Introduction

In this thesis, against the working hypothesis that feature synergy is the driving force of the human language system, I attempt to identify and analyse the constraints and mechanisms of feature interaction. In order to do so I survey the extent to which phenomena that are conspicuous by their absence rather than by their presence can inform us about the Faculty of Language (FL). Following from the Borer-Chomsky Conjecture (Baker, 2008), namely that the locus of language variation resides in lexical items, my core assumption is that linguistic features and, in particular, their interaction constitute a window into the nature and the architecture of the FL. The main idea is that each unattested logically possible configuration within any language system provides insight into and information about the make-up of the elements involved in the configuration. Instrumental to this endeavour is therefore the concept of linguistic features; it will be posited that ultimately the core linguistic elements that hold the key to understanding human language are features of some sort. An increased understanding of the nature of features may in turn lead to further uncovering of universal aspects of all languages. In other words, it is the distribution, the accurate identification and, perhaps most crucially, the establishment of the interaction properties of features, that should propel linguistic research. Evidently it is easy to propose such an agenda but implementing it may prove to be difficult. Bringing about such an enterprise, however, should be a long-term goal for general theoretical linguistics. The immediate goal of the present thesis is then not to undertake a full-scale analysis of current feature theory, but it is rather to present a systematic classification of the dynamics of feature interaction. To achieve this I undertake a study of an empirically broad-ranging selection of asymmetrical or skewed patterns, which will exemplify as well as demonstrate the mechanics of feature interaction.

The structure of the thesis is as follows: first I discuss current theories of the design of the FL and summarise the current research into universals (section 2); second, I present a series of implicational feature-interaction patterns yielded by the interaction of two features (section 3); then I discuss and analyse these patterns attempting to extract sensible generalisations and constraints on feature interaction (section 4); thereafter I envisage integrating the findings into a language theory based on features,

discussing in particular the role of Merge as well as remaining dilemmas (section 5); I summarise the dissertation and its major ideas in the conclusion (section 6).

# 2 Theoretical Background and Assumptions

This section outlines the theoretical platform upon which the subsequent discussion is founded: it recapitulates current assumptions in linguistic theory as regards the architecture of the FL, (section 2.1); it presents current stances to universals and typology alongside this conception of the FL (section 2.2); and, it presents a short history of feature theory together with a summary of the current mainstream approaches to features (section 2.3).

## 2.1 Theoretical Assumptions and the Faculty of Language

It is necessary to be specific about the object of study in linguistics because 'language' is too fuzzy a concept. The stance I take in this thesis is a biolinguistic (Jenkins, 2002; Hauser and Bever, 2008) and a generative one. It must be pointed out, however, that the boundaries between notions such as biology, psychology, formal and applied/functional linguistics etc. with regard to the FL are increasingly starting to loosen. Albeit a slow process, this tendency towards collaboration is viewed as a positive development within the language sciences by Mairal and Gil (2006). For most generative linguists, the central desiderata for language theory were set out by Chomsky in seminal works such as Chomsky (1968, 1986), *inter alia*. The assumption of a separate mind-unit uniquely dedicated to language, the FL, was made and the important distinction was drawn between E-language, effectively external linguistic output in its most superficial sense, and I-language, the internal processing system. Since the early days of generativism, Chomskyans have always assumed a FL distinct from other cognitive systems, with unique properties, and have only been concerned with the study of I-language, the underlying 'pan-human language system'. As this proposal is contentious, subscription to this view varies from rejection to embrace across the other branches of the cognitive sciences.

More recently, however, it has become apparent that even further fine-tuning is required as regards the object of inquiry of linguistics. Hauser et al. (2002) present the view that abstracting away from 'language' is not enough; they establish a dichotomy

between the FLN and the FLB, the FL in the narrow and the broad sense. For them, essentially, FLN is the only uniquely human component of FL and in its most basic interpretation it only comprises recursion. The FLB comprises the sensory-motor-system, the conceptual-intentional system and the mechanisms for implementing recursion. Furthermore, these authors make the claim that 'I-language is the primary object of study for the evolution of the FL' and they call for the use of the comparative method to establish exactly how the FL came about. For clarity, this 'comparative method' does not refer to classical philology but rather to the phylogenic comparative method (PCM) used in evolutionary biology. The latter incorporates the study of various biological traits and, through computational analyses, evolutionary relationships between organisms are thus determined. The results from PCM analyses can be represented in a phylogenic tree which shows both hierarchical relationships between biological traits but also the time-span to the most recent common ancestor for each pair of traits that are compared.

The idea of a separation of FL into an FLN and an FLB was a controversial move which has enabled a more fine-tuned framework of study. In spite of this, the proposal of these ideas did not come without opposition. Pinker and Jackendoff (2004) agree with the distinction between FLB and FLN but criticise Hauser et al. (2002) on the grounds that their attack on the argument of design is wrong. Briefly put, the argument of design proposes that many aspects of language have recently evolved by natural selection for enhanced communication. This question – whether 'language' evolved for communication or for abstract thought, consciousness, or something else – is a fundamental dilemma for all of the cognitive sciences. The evolutionary argumentation was pursued in Fitch et al. (2005) and Jackendoff and Pinker (2005) but, though interesting and very important, it will not be taken into account further here. For my present purposes I concord with Hauser et al. (2002) and Fitch et al. (2005) in that defining 'language' requires a separation between the computational system, which generates linguistic expressions, and the broader communication system.

I also embrace the ideas about three factors in language design first defined in Chomsky (2005). Chomsky suggests that language acquisition and indeed adult linguistic competence are conditioned by three factors:

- Factor I – The Genetic Endowment, UG: uniform for the species.
- Factor II – Language Experience, PLD. Diverse across the species.
- Factor III – Mind-General Cognitive Principles not specific to the FL.

Linguists do have direct access to Factor II (F2) but not to Factor I (F1) and not 'directly' to Factor III (F3). Earlier, the main goal of generative grammar was to define the contents of F1, namely UG. Within the biolinguistic perspective and the Minimalist Programme (Chomsky, 1995, 2005), there has been a gradual move towards studying the relationship between F1 and F3, namely to what extent F1 can be conceived as a product of F3. UG is more and more thought of as being content-poor, underspecified and thus rather simple. The main reason behind most claims of this type is again an evolutionary one based on the recent emergence of our species and its prowess for language. Spelling out the whole rationale behind 'Evo-Devo' science and evolutionary linguistics, lies beyond the scope of this thesis. The gist of it is, nevertheless, that a cognitive apparatus of 'thought manipulation' together with the physiology for externalisation (vocal tract, hands etc.) were already in place prior to 'a great leap forward', which, essentially, provided the missing link in the machinery, namely FLN, the optimal solution to interface (CI and AP) legibility conditions, (cf. further details in Deutscher 2005; Fitch 2010, *inter alia*). In evolutionary terms this was one of the most recent developments in the species and must thus biologically constitute a very small modification to the genome. Analogously it follows that FLN is most likely content-poor rather than content-rich. Whether UG's sole component is a process of recursion, as proposed by Hauser et al. (2002), under one reading, is highly debatable and most likely a too simple proposal.

In general, as far as language models are concerned, I subscribe to the more reductionist and more 'UG-external' philosophy of the Minimalist Programme (MP) and its 'inverted Y-model' (cf. Chomsky 1995, chap 1, Chomsky et al. 2002; Chomsky 2006; Adger 2003; Boeckx 2006). By UG-external philosophy I refer to the recent paradigm-shift in theoretical linguistics that led from asking questions pertaining the nature of UG to asking questions about the world and constraints on the 'mind'. More epistemic inquiries of the following kind, 'what is the content of UG such that it could

yield the world's languages', have gradually been substituted by more world-oriented questions, like 'what are the [third] factors of general cognition that have produced the UG of humans'.

To be specific, I endorse the following view of generativism, reformulated with current terminology in broad terms: a Lexicon, the locus of lexical items and their (idiosyncratic) make-up, a generative computational system (FLN) whose mechanics essentially are structure-building, and two interfaces (pertaining to the FLB), namely PF (Articulatory-Perceptual) and LF (Conceptual-Intentional). Given these assumptions, the crucial question is that of the locus of language variation. Provided that UG is underspecified there are essentially two alternatives given that LF most likely is independent of structural variation: the Lexicon or PF. In the standard model, PF is an interface of externalisation of I-language-generated structures, but I also interpret it as strictly 'content-free'. Externalisation could in principle take any form but the most common output-types are speech or hand-made signs. It is important to realise that Externalisation is the process whereby any syntactic structure, which fundamentally has the property of hierarchy but not order, is inexorably subject to linearisation (although linearisation takes different formats depending on modality; e.g. sign languages allow a certain degree of simultaneity which is impossible in spoken languages). Given certain assumptions – *inter alia* briefly explained as LF not depending on linearisation – string surface order and linearisation are immaterial to an I-theory of the FL but a crucial window to how the whole system ultimately works. If Externalisation is a mechanism that executes linearisation of morpho-syntactic structures, then the structures must contain information relevant to such linearisation somewhere. Granted that structure building only stems from recursive merging processes executed by FLN, FLN does not hold the relevant information. Arguments have been put forward (cf. Berwick and Chomsky 2008; Richards 2004, 2009) that PF specifies linearisation information. That is just shifting from a previously content-heavy UG to a content-heavy PF, a move which does not fit the implementation of the minimalism I have described. By this I mean that FLN, PF and LF are only computational systems in the sense that they do not contain information themselves, they can only process. To reiterate, if FLN and PF are not the loci of Externalisation then it therefore follows that Externalisation-information resides within the Lexicon and, more precisely, most

likely on each lexical item.

Essentially this is spelling-out the Borer-Chomsky Conjecture (BCC), namely that the locus of linguistic variation is the lexical items (LI) of the Lexicon. If LIs are taken to be feature bundles then linguistic variation must stem from the featural make-up of LIs. With this theoretical stance borne in mind, my aim is to analyse how features behave vis-à-vis other features so as to establish and define the constraints of feature interaction. These constraints would then be instrumental in furthering the understanding of asymmetry within the language system.

## 2.2 An Appraisal of Linguistic Universals

Articulating the fundamental properties of the human language system, language universals, has been a driving force for grammarians and philosophers for millennia. Since Chomsky (1957), a shift was made from studying languages and their grammars to studying the abstract properties of grammars. Chomsky writes,

> [...] generally, linguists must be concerned with the problem of determining the fundamental underlying properties of successful grammars. The ultimate outcome of these investigations should be a theory of linguistic structure in which the descriptive devices utilized in particular grammars are presented and studied abstractly, *with no specific reference to particular languages.* (*Syntactic Structures*, introduction, Chomsky 1957; my emphasis)

Chomsky thus propounds the idea of an abstract level of linguistic analysis which would operate within any language system. This modern development of the philosopher Roger Bacon's older concept of a 'common grammar' was the seed of generativism and from it sprang the notion UG (cf. Chomsky 1965). With UG, a quest for 'universal grammar rules' was initiated and, since then, linguistic universals have been a key concept within the field. As stated by Barwise and Cooper (1981),

> Linguistic theory is concerned, in part, with natural language universals, facts which hold for all naturally occurring languages and which *distinguish them from other logically possible languages.* (My emphasis).

The notion of 'logical possibilities' is crucial in that although many logically conceivable instances of 'language' in a loose sense are possible, there are, seemingly, universal constraints which prohibit them from occurring. Beyond the universals themselves, seeking to axiomatise language, however, would thus be one of the ultimate goals for linguistic theory. This stronger aim entails, in the very mathematical sense of its terms, to find and define the fundamental atomic building blocks of language, to determine how these interact and how these yield universals – and perhaps also skewings. Furthermore, this would contribute to establishing the mechanisms of the generative system that produce instances of language, which, consequently, is a product of the axiomatic building blocks and their combinatorics. To what extent is this possible, however?

In virtue of being properties shared by all languages, universals are windows into the architecture of I-language. The Konstanz Universals Archive (KOA)[2] is an 'omnium gatherum' of a majority of seriously proposed universals that have been in and out of vogue in the literature. The list contains ca 2000 entries of universals which have since been modified or confronted with straightforward counter-examples. Currently, the debate on universals in language theory is still an area of major controversy, (cf. *inter alia* Evans and Levinson 2009; Longobardi and Roberts 2010; van der Hulst 2008). Without entering the debate itself, I shall here outline the major views on universals within the field as their various tenets are highly relevant to the present work. It is particularly important to understand the nature of universals since the primary empirical data of this thesis focuses on instances of what can be thought of as implicational universals.

Since the genesis of linguistic typology with Greenberg (1963), a lot has been written about universals in linguistics (Bach 1968; Comrie 1989; Croft 2002; Mairal and Gil 2006; Scalise et al. 2009 to name a few). The debate is older, however, but it also involves a complex network of several dichotomies of stance that mainly stem from the philosophical or the anthropological traditions. It is difficult to do them all justice but a very good overview is provided by the special issue of *The Linguistic Review*, volume 25 (2008). The issue provides a résumé of the state of universals in

[2] http://typo.uni-konstanz.de/archive/intro/

general (van der Hulst, 2008), universals in syntax (Newmeyer, 2008), in semantics (von Fintel and Matthewson, 2008), in phonology (Hyman, 2008) and in morphology (Bobaljik, 2008). I would like to highlight a few important points from these sources.

As far as universals are concerned in general, there are two schools: linguistic relativism and linguistic universalism. Taken in their extreme forms, they maintain that:

> [Linguistic Relativism] ... there are no linguistic universals, each language is a specific time- and space-bound solution to the communicative needs of people in some culture. In this view languages differ from each other in unlimited ways.
>
> [Linguistic Universalism] ... all of language is universal, ultimately a property of the human genome.
>
> (paraphrasing van der Hulst 2008).

The debate between these two stances pertains to the larger philosophical dichotomy between Nominalism and Rationalism. The former can be characterised as the notion that resemblances need or have no explanation whereas the latter maintains that properties of observable entities reflect entities of a more abstract nature, namely universals. This old conundrum is still a highly pertinent question to ask and it is tied to another very important question which is insisted upon by van der Hulst (2008), namely the notion of human languages as a natural class. The problems can be characterised by the following quotation:

> It is intrinsic to any scientific enterprise to pursue the formulation of general laws about some domain of inquiry. The phenomena that constitute such a domain cannot, in advance, or perhaps ever with certainty, be designated as forming a truly unified domain, i.e., a natural class. However, apparently, some classes of phenomena strike people, in a pre-theoretical, intuitive sense as forming such a unified domain and the goal of the scientist is to try and 'reconstruct' (and justify) this intuition. [...] Sometimes, however, it might appear that certain commonalities among languages only appear to be language-specific universals only because linguists have failed to see that the same characteristics appear in a broader class of phenomena. (van der Hulst, *ibidem*).

Without going into the debates about the modularity of mind (Fodor 1983) or the relationship between thought and language (Pinker 1994, 1998, and Fodor 2005), the crucial point is that it is difficult to delimit Language as a 'cognitive natural class'. It therefore follows that – and this is the important issue – it is difficult to judge the extent to which a given universal is domain-specific (i.e. language-unique), 'multi-modular' (affecting some but not all modules of the mind, e.g. a universal could affect both the language and music modules of the mind) or domain-general (effectively 'mind-general'). This is important because it has repercussions on the object of study, the FL. Considering the aforementioned theoretical scope of the MP and Hauser et al. (2002), van der Hulst writes that 'by reducing the apparatus [...] according to Boeckx (2006: 149) it became possible to ask which properties of grammars are unique to grammars and which are shared with other cognitive systems'. He then reduces the main point in Hauser et al., to the notion that perhaps there is nothing unique to language and that 'in the end it is all a cognitive apparatus'. According to this strict view there would be no language-unique universals. Following that path, I can see two types of universals: universals of human cultural behaviour (as summarised in the slightly controversial Brown 1991) and cognitive universals, in minimalist terms also known as third factors. It is evident, as argued by Mairal and Gil (2006); van der Hulst (2008) and Newmeyer (2008) among others, that many of the previously rather antagonistic fields of anthropology, psychology, neurology, evolutionary biology, theoretical or applied linguistics, and philosophy, are starting to unite with respect to language theory. Unpacking further intricacies of the important debate between domain-specific (language universals) versus domain-general universals (cognitive universals), however, lies beyond the scope of this thesis. To sum up, the nature of universals, whether more cognitive or linguistic, is a very important but also difficult question. Its answers depend on fundamental assumptions about the language system and about cognition, both of which are far from being thoroughly understood.

I now turn to the ontological debate between relativists and universalists (or nominalists/rationalists), which focuses on the rationale for whether universals do exist at all. Evans and Levinson's recent article (2009) attempts to make a considerable impact on the 'cognitive sciences' in that its aim was to destroy the mythical ideal of

language universals. Their stance is very much a relativist one, and their main point seems to be the lack of empirical data in formal linguistics, pointing out that '[...] there are vanishingly few universals of language in the direct sense that all languages exhibit them. Instead, diversity can be found at almost every level of linguistic organisation.' Thus they adduce significant diversity as a counterexample to abstract unity, seemingly misconceiving the notion of what UG is as a construct. Arbitrarily, they discard most substantive universals as un-enlightening and claim that the fundamental universals of constituency and recursion in syntax are 'non-universals'. As a selection among many replies from the universalist camp, Longobardi and Roberts (2010) grant it true that a lot of formal work suffers from a very limited cross-linguistic data span but they nevertheless maintain that there are universals that hold across all languages and they argue for the use of Modularized Global Parameterisation (cf. Longobardi 2003) to better control for the huge number of variables intervening in the study of any cross-linguistic generalisation. It is also worth pointing out that Evans and Levinson, despite their desire to debunk the myth of universals, cite a large number of seemingly valid universal properties of language in their paper (cf. the conclusion in particular). I concur with the universalists/realists in that I maintain that Evans and Levinson are too harsh in their dismantling of the universal edifice, on which more than 50 years of formal linguistics has focused. There is also, however, a modicum of truth in van der Hulst's claim (*ibidem* pp. 2) that a lot of absolute universals can be seen as flawed in that they depend on the 'theory of the day'.

My view is that universal research is possible and that it should indeed be as data-driven as possible. In this instance, data-driven particularly refers to the use of larger databases with information pertaining to a more broad-ranging selection of languages. Simultaneously, it should be assured that the cross-linguistic analyses used are theoretically up to date, sufficiently deep/fine-grained, but also compatible with one another (e.g. earlier comparisons of phenomena between, for instance, languages deemed to be non-configurational with configurational languages are flawed under the current emerging assumptions that all languages are in fact configurational). Such an approach would maximise the adequacy of 'surface' generalisations and minimise 'theory-of-the-day' flaws while still enabling formalist, more abstract, aims since these are capable of capturing, the 'deeper' universals which are difficult to induce directly

from the 'surface level'. Eventually, such an approach would break down the 'conceptual and empirical barriers between surface and deep universals' (Newmeyer, *ibidem* pp. 73).

The third point to elucidate about universal research is the very kinds of universals which can be proposed. By this I mean the different types of universals. All of the authors of the aforementioned *Linguistic Review*, issue 25 (2008), define different kinds of universals either based on their content or on their logical conception. Divisions like substantive/formal, deep/surface, absolute/relative etc. are abundant. In any case, provided that epistemological methods are borne in mind, namely that a universal must be falsifiable – that is, the situation *has to be able to be otherwise* – I believe that it is possible to define universals in a straightforward sense and I concur with Baker and McCloskey (2007) as well as with Longobardi and Roberts (2010) and many others. Baker and McCloskey, in particular, recapitulate the commonly made observation that typology has to do with discovering universals sampling widely among the world's attested natural languages, so as to obtain valid generalisations. There are three commonly mentioned types, which stem from Greenberg (1963):

- Absolute universals: no language/every language has X.
- Statistical regularities/relative universals: Languages with X are very (un)common.
- Implicational universals: If a language has X then it will also have Y. There are two classes: absolute and statistical.

Baker and McCloskey then discuss these different types and their implications for linguistic theory. The first type of universal is the most relevant. Universals of this type are, because of their very nature, rare and seemingly rather un-insightful, however – that is, on the contrary, a false conception probably partly due to their striking simplicity in comparison to otherwise more convoluted linguistic findings. As an example the authors cite 'all languages have vowels'; yet this is of course not a felicitous absolute universal since there are sign-languages. 'All languages have syllables' might be a better candidate; something that, despite its mundane character for the non-linguist, is anything but banal, for, as recently pointed out by Kayne

(2010) the syllable structure may in fact constitute evidence for Merge operating within phonology.[3] I return to aspects of this point subsequently (cf. section 5.1). In sum, those universals are often very simple in their formulation but very deep in content.

The second type concerns statistics, relative universals. Baker and McCloskey also maintain that 'many strands of work converge on the view that distributional regularities often reflect the action of functional pressures of one kind or another. These pressures can be perceived in a historical or an evolutionary sense; in functionalism, these pressures are seen in synchronic terms, with the fact that they are always in play in the same way giving rise to observed diachronic patterns.' Ultimately Baker and McCloskey suggest the controversial marriage of John Hawkins's theory of performance (Hawkins, 1983, 1994) with Minimalism. In his thesis, Mobbs (2008), has the same idea, recasting Hawkins's functionalist principles in terms of more minimalist, Third Factor considerations – a controversial proposal. It is worth pointing out here, that such collaboration across theories or frameworks is rare and not representative of the general situation within the field of linguistics; it seems fair to claim, however, especially in order to respect minimalist tenets, that such collaboration ought to take place.

The third type of universal may be the most interesting and significant for linguistic theory. Thus, Baker and McCloskey write, 'if there is a true partnership between grammatical theory and typology, this is where we might expect it to be centred. If sufficiently fine-tuned (the protasis of the if-clause can be expanded [*sic*]), statistical implications can turn absolute.'

The notion of implicational universals is highly relevant to my investigation into features. I espouse Baker and McCloskey's characterisation of universals and, largely based on Chomsky's Uniformity Principle,[4] I also hold that statistical universals – of the I-language variety, as opposed to other more contingent regularities such as Se-

[3]Note that, Samuels (2009) argues against the idea that syllable structure is analogous to phrase structure and also against the idea that syllables are part of UG.

[4]In the absence of compelling evidence to the contrary, assume languages to be uniform, with variety restricted to easily detectable properties of utterances, (Chomsky, 1999).

mantic Primes (cf. Wierzbicka 1972, 1996) etc. – do not exist; in fact the very notion of a 'relative universal' is oxymoronic. I agree with the strong claim of Baker and McCloskey (2007), namley that statistical relations are either immaterial epiphenomena without further significance for linguistic theory or disguised absolute universals (either absolute or implicational). In the latter case, the disguise is caused by noise in the data stemming from inadequate analyses of the language phenomena in question.

If a majority of linguistic variation can be reduced to some form of feature distribution or feature interaction, then the notion of universals and, in particular, implicational universals ought to be intrinsically linked with features. From this I infer, that clearly articulating the ways in which features interact and the extent to which features influence other features, will shed light on the language system and its internal workings.

## 2.3 Lexical Items and Linguistic Features

What is the nature of linguistic features – and what part do they play within linguistic theory? This is to ask what linguistic phenomena a feature-based theory can explain and that is the central question of this thesis. Based on the previous discussion, in this section I recapitulate the architecture of the FL showing where features fit in (section 2.3.1); I trace the origins of feature theory (2.3.2); and I outline recent developments and a blueprint of the nature of features (section 2.3.3) while underscoring some methodological caveats (section 2.3.4).

### 2.3.1 Features within the architecture of the language faculty

In both a primordial and an abstract fashion, I suggest that a human language system could look like the following: in the mind-external realm there is the PLD, also known as E-language (Factor 2). Mind-internally, the PLD is a manifestation of macroscopic feature-bundle combinations (molecular LIs, 'words') that are engaged in a hierarchical relationship (syntactic structure) and externalised into an ordered string (a 'sentence'). From a general perspective, all idiosyncratic aspects of the out-

put, loosely characterised as its morphology and phonology (whether sound or gesture etc.), the order of components, the prosody or intonation, etc. and ultimately the meaning structure,[5] are encoded as, and result from, the workings of atomic features in the Lexicon. If this highly contentious proposition is true, then most linguistic universals can be defined in terms of features or feature interaction. The initial state of a new-born baby would thus be (the architectural skeleton of) an empty Lexicon, a Computational System of Narrow Syntax (FLN), and two interface systems. The Lexicon would be akin to a storage of feature-combinations. The FLN has readily been described elsewhere (section 2.1) but it is the locus of recursive Merge and structure building. The interfaces could be conceived as two feature-processing/reading, only differing in their interpretative/processing objective. This feature-reading-device metaphor of the interfaces is intended to convey the idea that the interfaces would automatically execute whatever is dictated by the feature series with which they are presented.

### 2.3.2 The history of feature theory

The term 'feature' is somewhat versatile and in order to develop a theory based around the mechanics of features, these need to be defined, at least a priori.[6] Insofar as I am aware, features were first introduced into linguistics as a concept in morphology and phonology. The initial ideas have several sources. Early classifications of properties of morphemes, which today can be thought of as features, were developed within structuralist morphology (Matthews, 1991), but the major breakthrough came with the advent of speech analysis theories in the 50s, Jakobson et al. (1952), which were refined and deployed in Chomsky and Halle (1968), where the notion distinctive feature was developed within phonology. The features represent aspects of articulation and perception, and are binary in nature so that they take the values [+] or [-]. The features come in classes and examples of these are: the major classes of sounds, e.g. [±Continuant]; laryngeal state, e.g. [±voice]; manner of articulation, e.g. [±nasal]

[5] For proponents of the 'argument of design' (cf. Pinker and Jackendoff 2004; Jackendoff and Pinker 2005) this would be the major driving force of linguistic creation.

[6] Cf. overview and further reading in Odden (2005)

etc.

In parallel to these developments within phonology, seminal anthropological work on kinship terms is worth mentioning. Starting in the 1950s through the 1960s, work was carried out to describe socio-cultural kinship systems within different cultures/languages. Lounsbury (1956) first put forward the notion of a componential definition applied to kinship, whereby kinship terms were broken down into smaller units, kinship primes. These were then used to create genealogies, to express different kinship relationships, and to compare different kinship systems vis-à-vis a framework. A good overview is provided in Wallace and Atkins (1960), where it is spelt out that more '[...] elaborate kinship categories are conceived as relative products of these [primitive] categories'. Subsequently these ideas were incorporated into linguistics and lexical semantics. A good example is Wierzbicka (1972), and the notion of semantic primes – an anti-ethnocentric framework of universal primes used to express meaning with fundamental primitive building blocks common to all cultures and languages. She deploys this framework when she discusses principles of definition, translatability and semantic analysis with regard to kinship terms in Wierzbicka (1987). Another precursor to feature theory with roots in componential analysis is the school of cognitive semantics. It essentially argues against a strict Chomskyan view of the language system – that only syntax is properly generative – by proposing that semantics can be generative too. Jackendoff (1983) propounds one form of such anti-syntactocentric views based around the concept of a semantic marker. For Jackendoff meaning is a separate combinatorial unit, which does not entirely depend on syntax. A updated formulation of his theory can be found in Jackendoff (2002). Furthermore, it is also worth mentioning the framework called Head-driven Phrase Structure Grammar (HPSG), which proposes that the Lexicon is not just a list of entries but in fact internally highly structured, (Pollard and Sag, 1994).

By putting together these various notions and by expanding as well as adapting these different frameworks, rather intuitively, linguists formulated the idea of LIs as being feature-bundles. These more general features would fall into classes like the following: Semantic, Syntactic, Externalisational, etc. Generative grammar is fundamentally a feature-driven theory but too little work has been carried out in

order to define and understand what features are. Generic classifications of the type semantic, syntactic etc. might also be misleading since this presupposes that the language system separates features according to such properties when this may not be the case. Although there is empirical neurocognitive evidence (e.g. Ullman 2001) for this type of distinction, I still doubt that such differences have an actual impact on the functional mechanisms of the language system, at least for the purpose of pure computation *per se*. The most appropriate assumption to start from is that features are plainly 'bearers of (atomic) information' which participates in the generation of linguistic structures.[7]

### 2.3.3 More recent proposals

I shall now turn towards more recent developments. Again starting with phonology, ideas were formulated to the effect that features could be organised into a complex network of dependencies: a feature-geometry. A brief recapitulation of early attempts is provided in Harley and Ritter (2002a). They refer to pioneering initiators as represented by Clements (1985) and Sagey (1986). By the time of Noyer (1992), a Universal Feature Hierarchy was proposed. The problem with this treatment, however, as remarked in Harley and Ritter (2002a,b), is that 'under Noyer's treatment the feature groups themselves are still represented as unordered bundles'. What Harley and Ritter are proposing is to 'structure the bundle' (cf. also Müller 2008, 2010). It is my goal to provide further insight into how that can be achieved and how such a theory can be stretched from the fields of morphology and phonology to covering all the core levels of linguistic theory.

Hereafter I outline the gist of what features are taken to be. I base this condensed account on Harley (1994); Harley and Ritter (2002a) and Adger (2008). Although the account is rather intricate in both Harley (1994) and Adger (2008), I shall limit my description to the most necessary for the time being. Adger argues that an LI is a finite set of features. Both Harley and Adger, in his so called privative proposal, assume that each feature can be either present or absent but have no other properties. Concretely

[7] Henceforth I shall use the notion 'linguistic structure', rather than sentence, as a general term for a language structure of unspecified size and complexity.

this means that features appear only if they have a positive value. *Videlicet*, feature [F] is part of the feature set if it is active and positive, noted as [+F]. A feature [-F] does not pertain to the feature set. Moreover, the same feature cannot occur twice. While Harley more casually assumes monovalent features, Adger proceeds to argue for the possibility of features having several properties such as interpretability or uninterpretability; such a system is better than the present/absent alternative in some ways but not sufficient for dependencies between features such as morphological or semantic commonalities. Claiming that several feature theories are possible he ultimately concludes that the simplest way to deal with such links or dependencies is to allow a property called Value. The Value property is binary, it can take the option Positive (+) or Negative (-). In essence a feature is thus an ordered pair [Attribute; Value]; this can be written as [A:±] or [±A]. I shall opt for the latter. Adger points out that the system has the advantage that [+A] is the complement of [-A], a simple binary system. He subsequently discusses the possibility of a multivalent feature system but he ultimately dismisses that idea and propounds monovalency as well. In any case, note that Harley's theory is a theory for morphology and phonology while Adger's theory is a theory for syntactic features. I shall apply the same reasoning *mutatis mutandi* to any linguistic feature more generally, with their alleged subfield origin neglected.

Two very important points related to this framework are worth unpacking further. Ontologically, I do not endorse the existence – in the sense of substance – of negative features, e.g. [-F]. Within the system only positive features, namely information-carrying elements, can exist. In other words, this means that the property Value is superfluous and that I advocate an underspecified privative system.[8] Moreover I do not assume defaults, namely preselected preferences within the system that become operational in the absence of other overruling information. Instead I take the system to be entirely underspecified according to the tenets of Underspecification Theory in the sense of (Zwicky, 1977; Archangeli, 1988; Steriade, 1995) and, with evidence

[8] A felicitous analogy to this theory is the biological mechanisms of DNA encoding. All present DNA nucleotide sequences in a given gene code for a specific protein synthesis; there are no 'negative genes' coding for 'negative proteines'. A gene present in the genome is inexorably expressed; an absent gene has no substance in the genome and does not express anything.

from psycholinguistics, Penke et al. (2004). It should then naturally follow that the notion [-F] equates with the empty set. In a strict sense, this entails that negative features ought to be thought of exclusively as the absence of the feature, and nothing else. Consequently, the notion of a 'positive feature' also becomes unnecessary. For representational purposes, especially in small feature typologies or matrices, the use of the representation [-F] is convenient, however, but it must always be borne in mind that such a representation does not imply the existence of *something* within the system. This is a general preliminary as further specifications are provided in sections 4 and 5.

### 2.3.4 Methodological caveats

An issue with defining a new feature is the danger of designating as a feature something which, in fact is a complex phenomenon. Because language properties can cluster together and, as it were, mask each other, there is a risk of conceiving such properties as being dependent on the presence of one single macro-feature. A linguistic structure is ultimately a large-scale feature structure which, when externalised, does not necessarily make all of that structure apparent. The macroscopic language level does not therefore provide a one-to-one match with the microscopic level. This analytic ambiguity stems from the idea that all features do not seem to have to be explicitly externalised. It is hence the case that individual features might participate in the generation of the linguistic structure, as part of feature-complexes, while at the same time remaining opaque in the output. It may therefore be very difficult to be aware of the atomic features that operate at the lower levels of the string computation. It is plausible, therefore, that 'true' atomic features can be 'conceptually very tiny', perhaps much smaller than what contemporary feature theories suggest in the current literature. It is unproblematic within current theory to think of this type of higher- and lower-level properties, each of which may be associated with binary choices. My point is a methodological and epistemological one, aimed at showing that it is good scientific practice to start with an *a priori* more coarse-grained binary opposition as long as one is aware that there may be a more fine-grained reality beyond the initial assumptions.

## 2.4 Summary of Basic Assumptions

In this brief section I summarise the conclusions of my theoretical deliberations carried out so far.

I have theorised that universals are central to linguistic theory and that they may be formulated if a systematic and rigorous methodology within a significantly large cross-linguistic span is adhered to. Furthermore I hold that there are no statistical universals as tendencies are immaterial to the underlying properties of the system. Universals are taken to be higher-level properties of the language system that derive from underlying 'axiomatic building-blocks'. These are, ultimately, atomic language features.

Regarding features I assume or hypothesise that:

1. (atomic) features are information-carrying elements with a single Attribute
2. the information carried by a feature can in principle vary infinitely; since features are atomic, there cannot be *intra*-feature constraints on their nature, only constraints on *inter*-feature interaction
3. only 'positive features' occur in string derivations
4. the absence of a feature (noted as [-F]), does not entail a default output; the system itself is thus underspecified
5. features combine to form feature-complexes, or feature bundles, which make up linguistic structures
6. all superficial language phenomena can be linked to feature interaction and, moreover, the number of ways in which features can interact, at the atomic level, is finite.

I have thus provided a preliminary discussion of what features are and I shall now procede to present data of tetrachoric feature matrixes (section 3), after which I discuss potential generalisations regarding the nature of feature interaction (section 4).

# 3 Skewed Patterns and Asymmetrical Observations

This section is concerned with presenting observable data regarding skewed patterns and asymmetrical configurations. It starts with a few remarks on how the patterns can be conceived as well as represented (section 3.1), whereafter a series of configurations will be presented so as to demonstrate the extent to which asymmetry is present in the human language system and to what extent it is important (section 3.2).

## 3.1 Initial Remarks

My goal is to study feature interaction and how this can be linked to asymmetry. In order to do so I have proceeded by a contrastive analysis on the most minimal level in a systematic fashion. By this I mean that I have systematically compared at most two features at a time so as to single out how their value-setting (+/–)interacts with the other feature. Thus, with any two features F1 and F2, each having two value settings, I obtain a logical space in the form of a 2-by-2 factorial design, a matrix with four possibilities. The observation is that the matrix frequently generates a skewed pattern, an asymmetry of 3 against 4.

Important note: For feature-complexes the notation is arbitrary. [+F] is the same as [-$\bar{F}$], where $\bar{F}$ would be the complement formulation of the feature F.
Remember, however, that in theory, at the level of atomic features it is only the positive setting that corresponds to something being present within the set of features of the LI; in other words, the negative value of a feature always corresponds to the empty set '{∅}'.

## 3.2 Feature Interaction Patterns

For each pattern I shall define the two features by their full attribute name, and, if unclear what they refer to, examples will be provided. Necessary explanations and sources will be provided when needed. Note that the order of the features in each output pair is immaterial. Permutations that are ungrammatical/unattested are

marked with an asterisk and a generalisation as well as further comments are provided in each case. A theoretical matrix example is given below for clarity:

(P0) **Matrix Example**

[± Feature 1; ± Feature 2]
[+F1; +F2]
[+F1; –F2]
[–F1; +F2]
*[–F1; –F2]

Hereafter patterns observed in natural languages are presented. They have been organised approximately according to the linguistic properties of the relevant features. These are as follows: Syntactic relationships; Semantic relationships; Phonetic relationships and relationships pertaining to DPs, Pronouns and Morphology.

### 3.2.1 Syntactic relationships

- Head-Complement Relationships

(P1) **Greenberg's Universal 25**, (Greenberg, 1963).
*If the pronominal object follows the verb, so does the nominal object.*
With the present convention that can be rendered as:

[±Pro-V; ±VO]
[+Pro-V; +VO]. English: John saw the girl. John saw her.
[+Pro-V; –VO]. Latin: Puer puellam vidit. Puer eam vidit.
[–Pro-V; +VO]. Romance: Jean voit la fille. Jean la voit.
*[–Pro-V; –VO]. Unattested.

Generalisation: 'Reverse Romance' is impossible.
A possible account of this situation is provided by (Roberts, 2010, pp. 99-101). Roberts essentially derives the constellation from the distribution of an EPP feature on v* vis-à-vis whether the language is OV or VO. The first two cases are straightforward; regarding Romance, object pronouns are

defective goals for v and incorporate to v*, thus preceding the verb. In the unattested case, for the DP-object to precede V, v* requires an EPP feature but this would trigger the pronoun to move before the verb too.

(P2) **The Final-Over-Final Constraint**, (Biberauer et al., 2010).
*If $\alpha$ is a head-initial phrase and $\beta$ is a categorially non-distinct phrase immediately dominating $\alpha$, then $\beta$ must be head-initial. If $\alpha$ is a head-final phrase, and $\beta$ is a phrase immediately dominating $\alpha$, then $\beta$ can be head-initial or head-final.*
[±Head-Finality Lower Node; ±Head-Finality Higher Node]

| | |
|---|---|
| [+HFL; +HFH] | e.g. O V Aux |
| [+HFL; –HFH] | e.g. AUX O V |
| *[–HFL; +HFH] | e.g. V O AUX |
| [–HFL; –HFH] | e.g. AUX V O |

Generalisation: The higher node cannot be disharmonic if the lower node is head-initial; final-over-initial is unattested. For clarity, I provide a schematic representation of this situation below.

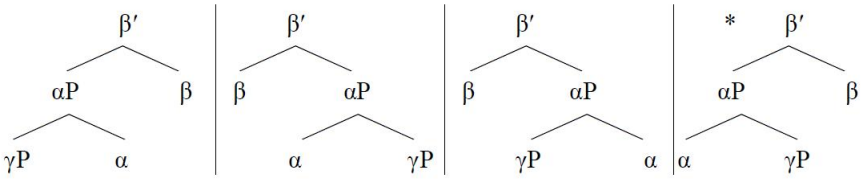


Figure 1: FOFC schemata

(P3) **Two types of VO languages but only one OV type**, (Pearson, 2000). In his paper Pearson proposes a typology of languages based on the order of objects and adverbs within the predicate phrase. He then contrasts this dichotomy of direct versus inverse languages with the fundamental VO/OV word order types. I illustrate these definitions below (cf. Pearson 2000 pp. 332)

| | *Direct* | *Direct* | *Inverse* |
|---|---|---|---|
| | **OV** | **VO** | **VO** |
| Double Obj. Constr. | IO DO V | V IO DO | V DO IO |
| Adv Order | Adv1 Adv2 V | V Adv2 Adv1 | V Adv1 Adv2 |
| Object Shift | $DP1_i$ Adv $t_i$ V | V $DP_i$ Adv $t_i$ | V $t_i$ Adv $DP_i$ |

Direct languages: English, French, Indonesian, Dutch, Turkish...
Inverse languages: Malagasy, Tzotzil, Q. Zapotec, Palauan...

The following matrix thus becomes apparent:
[±Direct; ±VO], where [-Direct] = [+Inverse] etc.
[+Dir; +VO] 'Direct-VO'
[+Dir; –VO] 'Direct-OV'
[–Dir; +VO] 'Inverse-VO'
*[–Dir; –VO] 'Inverse-OV'

Generalisation: In this typology the fourth possibility of an inverse-OV language is unattested.

(P4) **Conjunction type and Head-Finality**, (Zwart, 2009).
Zwart's article deals with minimalism and typology. Among other points he presents a typology of conjunction-species across a set of languages. Zwart defines two types of conjunctions: Final and Initial based on the relative position of the 'conjunction phrase head' to the second conjunct. The conjunction patterns are as follows:

i) A&B : Initial
ii) AB& : Final
iii) &AB : Non-existent

He then compares the distribution of Conjunction-finality with general Head-finality (cf. also Hawkins 1983; Dryer 1992, 2009). If mixed Head-finality (e.g. a system like German with different head-finality settings for different functional heads; e.g. V, Aux are final but C,D,N and most P are initial) is

ignored I obtain the following feature matrix:
[±Conjunction-Finality; ±Head-Finality], with [+CI] and [+HI] refers to the Initial value and the negative to the Final value.

[+CI; +HI]
[+CI; –HI]
*[–CI; +HI]
[–CI; –HI]

Generalisation: in sum, there are no Head-initial languages that have final conjunctions.
*Nota bene*: If conjunction is analysed as a form of syntactic 'andP', with 'and' being the head then this situation can be explained as part of the FOFC-generalisation (cf. pattern P2).

(P5) **Directional and Locative Preposition Order vs. NP**, (Cinque, 2008).
Cinque compares the distinction prepositional/postpositional and the order of directional versus locative prepositions in relation to the NP, which gives:

[±Prepositional; ±Directional first].
[+Prep; +DirF]
*[+Prep; –DirF]
[–Prep; +DirF]
[–Prep; –DirF]

Generalisation: No language is prepositional with locative prepositions preceding directional prepositions.

- Clause-type and the CP-domain

(P6) **Free-standing clause types**.
By contrasting the following features within the C domain I obtain:

[±Interrogative Q-feature; ±Wh-movement]
[+Q; +Wh] Interrogative clause.
[+Q; –Wh] Echo question (with declarative syntax) or yes/no-question.
*[–Q; +Wh] A sub-clause structure; impossible as bare, without a matrix clause.
[–Q; –Wh] Declarative clause.

Generalisation: The clause [-Q; +Wh] cannot occur on its own under normal circumstances of intonation.
With appropriately altered intonation this type of clause could be an exclamative as in E1. Note, however, that this does not work with any Wh-element, as shown by E2.

(E1) (What a nice man) you are!

(E2) *Which nice man you are!

(E3) How friendly of you!

(E4) *[To what extent] friendly of you!

Interestingly, these examples also demonstrate that the 'what' in E1 or the 'how' in E3 function like morphologically overt markers of exclamative illocutionary force and not as question particles, despite their syntax. Cf. further explanations below in pattern P7.

(P7) **Clause-type feature interaction**.
By contrasting the following features within the C domain I obtain:
[±Imperative feature; ± Interrogative Q-feature]

| | |
|---|---|
| *[+Imp; +Q] | Imperative and interrogative clause simultaneously: impossible. |
| [+Imp; –Q] | Imperative clause. |
| [–Imp; +Q] | Interrogative clause. |
| [–Imp;–Q] | Declarative clause. |

Generalisation: as noted by Sadock (1974) so called 'Wh-imperatives' (e.g. *Why don't you close the window?*) do seemingly occur. Structurally, the ambiguity between imperative and request does not pose a problem but when externalised a choice has to be made as it is seemingly impossible to apply the intonation and prosody of imperatives and questions simultaneously. In other words, an utterance cannot be a question and a request simultaneously under normal circumstances of intonation.[9]

*Nota bene*: Two more identical patterns could be derived if [Exclamative] was paired up with either of [Imp] or [Q]. The same explanations as above would

[9]Furthermore these 'wh-imperatives' also have implications for pragmatics and politeness theory.

also apply. Note that *[+Excl; + Imp] is definitely more improbable than *[+Excl; +Q], at least in English; this might have a phonetic explanation. Nevertheless, the two previous patterns demonstrate that there is a mismatch between illocutionary force and clause type, in the sense that these two do not necessarily map onto each other.

- Relationships of Syntactic Copying

(P8) **Holmberg's Generalisation**, (Holmberg, 1986, 1999).
I here present Holmberg's generalisation as a 3/4 pattern:
[±Verb Raising; ±Object Shift], with examples in Swedish from Holmberg (1999).

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| a) | [+VR; +OS] | Jag | kysste | henne | inte | [VP $t_v$ $t_o$] | |
| | | I | kissed | her | not | | |
| b) | [+VR; –OS] | Jag | talade | inte | [PP med | henne] | |
| | | I | spoke | not | to | her | |
| c) | *[–VR; +OS] | Jag | har | henne | inte | [VP kysst] | |
| | | I | have | her | not | kissed | |
| d) | [–VR; –OS] | [CP | att | jag | inte | kysste | henne] |
| | | | that | I | not | kissed | her |

Generalisation: Object shift only occurs if the verb raises.
Holmberg further explains that OS can be blocked by 'any phonologically visible category inside VP preceding the object position'. This is illustrated by the constellation b), where the preposition 'med' blocks the Object Shift. In MSc, in general, there is no V-raising in subclauses and therefore there is no OS in constellation d). Bearing these clarifications in mind, one might refine the generalisation as: Verb-raising is a necessary but not a sufficient condition for Object Shift.

(P9) **Romance Interrogative Clauses**.
The interrogative syntax of Romance languages manifests several variations. When contrasting subject-verb inversion with Wh-movment the following matrix is produced. It is taken for granted that each output is externalised

with the appropriate interrogative intonation (e.g. rising intonation, which, if present, modifies the illocutionary force of an utterance with declarative syntax to a question). Each case is here exemplified with French.

[±Inversion-trigger; ±Wh-movement]

| | |
|---|---|
| [+Inv; +Wh] | *Où es tu?* |
| *[+Inv; –Wh] | **Es tu où?* |
| [–Inv; +Wh] | *Où tu es?* |
| [–Inv; –Wh] | *Tu es où?* |

Generalisation: Subject-verb inversion is subject to Wh-movement to the left periphery.

A significant counterexample is associated with this distribution. In Munaro (1997) it is explained that the dialect 'Bellunese' (from the area Val Belluna in Italy) manifests interrogative clauses with optional wh-*in situ* despite subject-verb inversion as in the example below (Munaro 1997, from pp. 30):

(E5)

| | | |
|---|---|---|
| *con qual* | à-tu | parlà? |
| with whom | have$_{2sg}$ you | spoken? |

(E6)

| | | |
|---|---|---|
| à-tu | parlà | *con qual*? |
| have$_{2sg}$ you | spoken | with whom? |

I interpret this situation as representative of different feature-interaction conditions than those that operate in other Romance languages or dialects. The explanation given in Munaro (1997) is essentially that the Bellunese dialect does in fact not have wh-*in situ*, after all. It is argued that in interrogatives the VP undergoes further leftward movement after movement of the wh-phrase has occured (Bellunese thus falls under [+Inv; +wh]). The surface order of the interrogative clause therefore gives the impression of a wh-*in situ* situation, although this is not the case underlyingly. I return to this pattern subsequently to raise another independent point (section 4.1.5).

(P10) **Syntax copying and Phonological Spell-out**, (Barbiers et al., 2010).

In their paper, Barbiers et al. discuss aspects of syntactic copying and phonological spell-out; they argue that copying processes can be full or par-

tial and that phonology can spell out one or more chain links. They thus obtain the following situation (pp. 7-8):

i) Full copying and both chain members are spelt out: identical doubling

ii) Full copying, but (for some reason) only the higher chain member is spelt out: non-doubling

iii) Partial copying and both chain members are spelt out: non-identical doubling

iv) Partial copying but (for some reason) only the higher chain member is spelt out.

They then discard option iv) on the grounds that it creates a recoverability problem. In featural terms I summarise it as follows:
[±Full copying; ±Highest chain-link spell-out only] → *[–FC; +HCL].

Generalisation: Higher chain member spell-out is conditional upon the copying process being full. This clearly follows from the principle of Recoverability, namely the constraint permitting the application of deletion rules only in such a way that any item deleted can be (meaningfully) recovered.

### 3.2.2 Semantic relationships

(P11) **Aristotelian Square of Opposition**.
In traditional logic the square of opposition summarises the logical relationships between the four forms of subject-predicate proposition known as A, E, I, O: All X are Y; no X are Y; some X are Y; some X are not Y. This is shown by the diagram below:

Thence I derive the following matrix, contrasting universal/existential quantification with the presence/absence of logical negation within morphological structure of the monolectic quantifier:

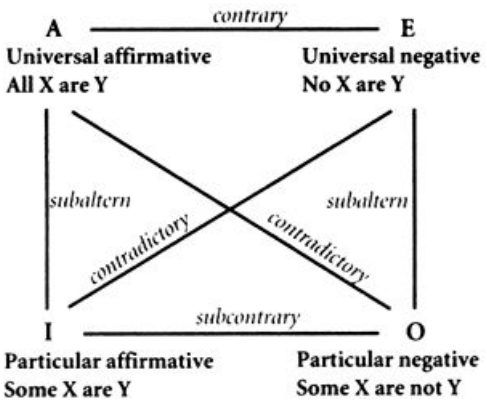


Figure 2: Aristorelian Square of Opposition

[±Universal; ±Neg]

| | |
|---|---|
| *[+U; +Neg] | **Nall*, periphrastic 'not all' is required |
| [+U; –Neg] | *All* |
| [–U; +Neg] | *None* |
| [–U; –Neg] | *Some* |

Generalisation: The concept *Not All* is unattested in the sense that it is not a single morphologically monolectic LI, e.g. *'NALL'*. This O-corner pattern of the Aristotelian Square can be replicated for many different semantic features as outlined below:

(a) *NEVERY

(b) *NAND

(c) *NALWAYS

(d) etc...

It is worth noting that this generalisation follows from Barwise and Cooper (1981), (also cf. pattern 13), in that only mon↑ quantifiers can be negated with an extra-lexical negation item. Therefore: *not all, not few, not every, not many etc.* are valid, while *not each, not some, not John, not no, ?not few, not the NP etc.* cannot be negated in this way, unless they occur in focus

or denial contexts as observed by Horn (1989). Also, see further theorising in under pattern 13.

(P12) **Semantic Primes of Colour Terms**, (Berlin and Kay, 1969).
Provided the hypothesis that colour terms are typologically distributed in languages according to a fixed hierarchy such as:
Colour C1 < Colour C2... < Colour $C_n$ then one can define the following:

[± $C_n$; ± $C_{n+1}$]
[+$C_n$; +$C_{n+1}$]
[+$C_n$; –$C_{n+1}$]
*[–$C_n$; +$C_{n+1}$]
[–$C_n$; –$C_{n+1}$]

Generalisation: If a language has a colour $C_n$ then it has all the colours between $C_1$ and $C_n$ in the hierarchy. As an instance of a semantic pattern, this observation might be less convincing, but, in principle, from the viewpoint that semantic features may be ordered in an hierarchy it is representative of some underling cognitive principle.

(P13) **Monotonicity Constraints and Negation**, (Barwise and Cooper, 1981).
In their now seminal paper, Barwise and Cooper develop a theory of generalised quantifiers in natural language. Therein they define Monotonicity, a property of NPs. In section 4.11 of their article they contrast monotonicity with lexical negation and derive their Universal 9 (op.cit, pp. 198). I extract the following matrix from their reasoning:
[±Monotonicity; ±Lexical Negation] where [+Mon]/[-Mon] stands for Monotone Increasing/Decreasing.

[+Mon; +Not] 'Not All'
[+Mon; –Not] 'All'
*[–Mon; +Not] 'Not few'
[–Mon; –Not] 'Few'

Generalisation: As Barwise and Cooper concludes, only mon↑ can be negated in this way.
*Nota bene*: These observations are connected to the observations in (P11).

**Possible feature-based explanation** (note that the notions atomic and feature-complex (cf. section 4) are relevant for this proposal): Consider the feature [Mon↓]. This feature could be conceived as a feature-complex composed of the following atomic features:

$$[\text{Mon}\uparrow]\cup[\text{Negation}].$$

Under the hypothesis that the same atomic feature cannot be incorporated twice into the same structure (I discuss this further in section 4.1, cf. Conjecture 1) within certain boundaries,[10] as also briefly posited by Adger (2008, pp. 4), it follows naturally that

$$[\text{Negation}]\cup[\text{Mon}\downarrow] \equiv [\text{Negation}]\cup[\text{Negation}]\cup[\text{Mon}\uparrow]$$

which is invalid because of the double appearence of the feature [Negation] within the boundaries of the same quantifier. Moreover, as noted by Horn (1989), double negation is a very unnatural structure in natural languages, except for special focus or denial contexts. This corroborates the idea that double negation cannot be lexicalised into a monolectic form expressing precisely such a meaning structure.[11]

(P14) **Internal Nominal Structure**, (Chierchia, 1998, p. 353).
Chierchia postulates that a nominal can bear the features Argumental or Predicative, thus:
NP: [±Arg; ±Pred]

| | | |
|---|---|---|
| [+Arg; +Pred] | Bare arguments allowed. | → e.g. Slavic (no articles) |
| | | → e.g. Germanic (articles) |
| [+Arg; –Pred] | Every noun is mass. | → e.g. Chinese |
| [–Arg; +Pred] | Bare arguments disallowed | → e.g. Romance |
| *[–Arg; –Pred] | Impossible | |

Generalisation: Essentially, nominals are required to be either argumental or predicative, or both. Nominals cannot be neither.

[10] Yet to be defined.
[11] This is a simplified account; cf. further details in section 4.1.5.

As far as providing definitions of the core lexical categories is concerned, this is an important generalisation adding to the understanding of what nouns fundamentally are. A further interesting observation from this matrix is the distinction mass/count languages versus mass languages, which can be captured by the presence of the feature [+Predicative]. Seemingly, languages lacking this feature cannot make the distinction count/mass, thus lacking the possibility of combining a numeral quantifier to an NP (eg. Chinese). Similarly, the feature [+Argumental] captures the possibility of having bare arguments.

### 3.2.3 Phonetic relationships

(P15) **Vowel Quality**, (Chomsky and Halle, 1968).
As famously proposed, the features [+high] or [+low] qualify one vowel dimension. Their interaction gives the following:
[±Low; ±High].

| | |
|---|---|
| *[+L; +H] | Impossible |
| [+L; –H] | Low/Open [a] |
| [–L; +H] | High/Close [i] |
| [–L; –H] | Mid [ə] |

Generalisation: A vowel cannot be both high and low simultaneously.
This follows from the definitions of the distinct features but even if more intricate systems are considered, such as a system with three or four levels of vowel height, the same generalisation capturing the impossibility of producing two vowel heights simultaneously would hold. Even if the property vowel height was conceived as a continuum the vocal tract could yet only produce one sound at a time. In other words, externalisation forces discrete renditions of vowel height values. The same property would apply to other phonological phenomena, notably sign language. In this case, even if the signing space might be divided into a continuous infinity of different positions, the hands have to produce a discrete output. E.g. it is impossible with one hand to make a sign appear both high and low in the signing space

simultaneously.

### 3.2.4 DPs, Pronouns and Morphology

- The DP-domain

(P16) **Article typology**, (Chierchia, 1998, p.362).
Chierchia suggests the following typology of having or not having Definite versus Indefinite Articles:

[±Def; ±Indef]

[+Def; +Indef]

[+Def; –Indef]

*[–Def; +Indef]

[–Def; –Indef]

Generalisation: There is no language with indefinite articles that does not also have definite articles.
Chierchia's rationale for proposing this generalisation had to do with the difference in scopal properties of indefinite singulars in languages with articles (e.g. English) versus the indefinite reading of bare nominals in languages without articles (e.g. Russian). He also claims that it would be important to consider languages with only one type of article and conjectures – hypothetically – the above generalisation. According to WALS, (Haspelmath et al., 2008), however, there are 44 such languages attested in the database. This pattern is therefore not valid and ought to be refined, perhaps taking mass/count and singular/plural distinctions into account.

To this effect, using WALS Online, I combined map 34 (Nominal Plurality), with the maps 37 (definite articles) and 38 (indefinite articles) and found the following: comparing languages with 'No nominal plural' vis-à-vis their article distribution I found that there are no languages with indefinite articles only that do not have a nominal plural. In other words, I suggest the following: all languages with indefinite articles can pluralise their nouns.

- Pronouns and Dropping

(P17) **Italian Null-Subject vs. Null-Object**,[12] (Rizzi, 1986).

[±Null-Subject; ±Null-Object]

[+NS; +NO]

[+NS; –NO]

*[–NS; +NO]

[–NS; –NO]

Generalisation: Null-objects only occur if Null-Subjects occur in the language.[13]

Cf. further comments below in patters P18 and P19.

(P18) **Agglutinative Morphology versus Radical Pro**, (Neeleman and Szendröi, 2007).

Neeleman and Szendröi analyse the prevalence of radical pro drop as a function of agglutinating pronoun morphology. From their reasoning I derive the following:

[±Agglutinating-Morphology; ±Radical-ProDrop]

[+Agg; +RPD] → e.g. Chinese (agglutinating pronouns with RPD)

[+Agg; –RPD] → e.g. Finnish (agglutinating pronouns, yet no RPD)

*[–Agg; +RPD] → unattested

[–Agg; –RPD] → e.g. Germanic (fusional pronouns and no RPD)

Generalisation: Radical Pro-Drop can only occur if the pronoun morphology is agglutinating.

Neeleman and Szendröi's proposal is based on three assumptions: a) Spell-out rules for pronouns may target nonterminal categories; b) Pro drop is zero spell-out of regular pronouns; c) Competition between spell-out rules is gov-

[12] My gratitude to Ian Roberts for pointing this out

[13] Singaporean English does manifest structures which could be taken as counterexamples against this generalisation, according to Sato (2010). These are more akin to null topics – plausibly due to contact-phenomena with Cantonese – rather than null subjects, however. The comparison is therefore probably inadequate.

erned by the Elsewhere Principle (Kiparsky, 1973). In particular, RPD occurs when pronouns are agglutinating for case and number, or perhaps some other nominal feature, regardless of whether determiners occur or whether there is rich agreement.

(P19) **Verb morphology agreement and subject pro-drop**, (Roberts, 1985). Here I suggest a comparison of the availability of subject pro-drop vis-à-vis verb morphology; I obtain the following considering Romance and Germanic languages, which do not have agglutinating pronoun morphology:

[±Verb-morphology; ±Subject Pro-Drop]

| | |
|---|---|
| [+VM; +SPD] | Spanish, Italian... |
| [+VM; –SPD] | German, French... |
| *[–VM; +SPD] | Unattested |
| [–VM; –SPD] | English, MSc... |

Generalisation: As argued by Roberts (and initially Rizzi 1982), subject pro-drop (SPD) only occurs if the verb has some morphological agreement. One may object against this generalisation, claiming that languages with radical pro-drop (RPD) constitute counterexamples. Indeed, they do demonstrate the occurrence of nominal dropping without verbal agreement; RPD languages are, however, very different from the Indo-European languages where SPD occur. For instance, in RPD languages all nominals can be elided whereas the present typology concerns subjects only. This is to show that very different factors come into play in the two systems and therefore the initial observations can be taken to hold within the more confined scale of Indo-European SPD languages. This is a good example of where superficial similarity (some form of DP-ellipsis) should not licence comparison since the featural make-up of the languages in question are most likely totally different.

(P20) **DP-morphology and noun-drop**, (Kester, 1996).

Here I suggest a comparison of the availability of noun-drop vis-à-vis DP-internal morphology, namely adjectival agreement, as in the examples below:

a) Der Gute *ec* [German] *the good (one/person/thing)*
b) Den gamle *ec* [Swedish] *the old (one/person/thing)*
c) Le grand *ec* [French] *the tall (one/person/thing)*
d) El raro *ec* [Spanish] *the unusual (one/person/thing)*

I obtain the following matrix:

[±DP-morphology; ±Noun-Drop]

| | |
|---|---|
| [+DPM; +ND] | German, MSc, Spanish, Italian... |
| [+DPM; –ND] | *ut supra* |
| *[–DPM; +ND] | Arguably unattested |
| [–DPM; –ND] | English... |

Generalisation: Noun-drop is only attested when there is rich morphological agreement within the DP.

A caveat has to be put forward with regard to this proposal: many East Asian languages which do not manifest adjectival agreement at all, do manifest various degrees of noun-dropping (most often NP-ellipsis, though). This is a potential counterexample to the proposal above. Analogously with the difference between SPD and RPD (cf. P20), however, it is arguably not an appropriate counterexample on the basis that the two phenomena, although seemingly similar on the surface, in fact, are underlyingly disparate – N-drop, as in the above pattern, is very different from NP-ellipsis.

(P21) **Reflexivity**, (Reinhart and Reuland, 1993).

Reinhart and Reuland propose a typology of referential DP-elements based around the two following features:

[±Reflexivizer; ±R(eferential) Independence]

| | |
|---|---|
| *[+Reflex; +R] | NOT ATTESTED |
| [+Reflex; –R] | SELF |
| [–Reflex; +R] | PRONOUNS |
| [–Reflex; –R] | SE |

Generalisation: The DP-category bearing both the feature Reflexivizer and Referential Independence is not attested.

Reinhart and Reuland do not explicitly explain why [+Reflex; +R] is impossible. Speculatively, I propose the following to explain the unattested combination: the feature [+Reflex], held by SELF anaphors which are NP-elements, act locally to reflexivise the predicate (a verbal element) but they do not cause the SELF-NP to function as arguments alone – a D-element is required to fulfil this role. SELF is probably therefore a feature pertaining to the functions of predicates. This explains why in some languages morphologically it can appear either within DP-arguments that take part in a predicate structure (e.g. reflexive pronouns in English) or within the VP itself (e.g. such as in Modern Greek, where the passive form of verbs can carry the reflexivising feature). As a feature, it is therefore dependent on another category which is referentially (anaphorically) connected to a higher DP. The feature [+R], conversely, is not dependent on a determiner and can thus act as a nominal argument on its own. In sum, though only speculative, I conclude that [+Reflex] is somehow verbal in nature and that [+R] is comparatively nominal in nature and that, therefore, the two cannot occur together. Note also that this conclusion may well be related to the features [+Arg] and [+Pred], as discussed in pattern P14.

- Morphological relationships

(P22) **Root-Compounding versus Resultative constructions**, (Snyder, 2001). Snyder presents data arguing for effects between argument structure and complex word formation, especially in child language. I infer the following typology from his article:

[±Root-compounding; ±Resultative clauses]
[+RC; +Res]
[+RC; –Res]
*[–RC; +Res]
[–RC; –Res]

Generalisation: as Snyder concludes, 'root-compounding is necessary but not a sufficient condition for the availability of resultatives.

(P23) **Baker's Morphological Visibility Condition**, (Baker, 1996, chap. 1, p. 17-19).

In his book *The Polysynthesis Parameter*, Baker defines the Morphological Visibility Condition, and suggests that it can be broken down into two independent parameters that make phrases visible for $\theta$-role assignment, namely agreement or incorporation. For my purposes such parameters could be thought of as features. From his arguments, I derive the following matrix, which in essence divides languages into $\theta$-role assignment types:

[±Agreement; ±Noun-Incorporation]

| | |
|---|---|
| [+Agr; +NI] | Polysynthetic proper. |
| [+Agr; –NI] | Nonconfigurational head-marking languages |
| *[–Agr; +NI] | Impossible |
| [–Agr; –NI] | A wide range of configurational languages... |

Generalisation: Only languages with agreement have noun incorporation; the configuration *[–Agr; +NI] is unattested.

(P24) **Grammatical Number**, (Adger, 2008).

With regard to grammatical number, Adger points out that for a Singular/Plural system one binary-valent feature is sufficient. Phenomenologically, however, there are three numbers attested in the world, namely Singular, Dual and Plural. To capture this, at least another feature is required. The features used are [Atomic], a set of atomic individuals, and [Augmented], a set with non-atomic proper subsets. Further details are to be found in (Noyer, 1992; Harbour, 2007). The feature patterns is as follows:

[±Atomic; ±Augmented].

| | |
|---|---|
| *[+At; +Aug] | Contradictory, as atomic individuals cannot have subsets. |
| [+At; –Aug] | Singular |
| [–At; +Aug] | Plural |
| [–At; –Aug] | Dual |

Generalisation: the features [Atomic] and [Augmented] are incompatible with each other. Adger further explains that with this system 'there is a way to refer to the class of singulars and duals together, namely [-Augmented],

and of duals and plurals together, namely [-Atomic], but no way to refer to the class of singulars and plurals to the exclusion of duals. This practice is indeed unheard of. Furthermore, concerning paucal or trial number, a feature geometry, which is more elaborate than Adger's system (i.e. involving more features) is discussed in Harley and Ritter (2002a).

(P25) **Grammatical Person,**[14] (Sigurdsson, 2003, 2004).

Sigurdsson devises a typology of person which he derives from contrasting the following:

[±Logophoric Agent; ±Logophoric Patient].

| | |
|---|---|
| *[+LA; +LP] | Seemingly impossible. |
| [+LA; –LP] | First Person |
| [–LA; +LP] | Second Person |
| [–LA; –LP] | Third Person |

Generalisation: the three commonly attested persons are derived, leaving the fourth possibility as impossible.

*Nota bene*: In Bobaljik (2008), among other things, a thorough review is provided of the grammatical category 'Person'. Therein is discussed the so called 'First Person Inclusive', which takes the features [+speaker; +hearer]. In the above analysis, however, I rule out *[+LA; +LP] on the strict reading that one single person cannot be the logophoric agent and patient simultaneously. I also rule out self-addressed speech from this since the pragmatics of self-speech is rather complex.[15]

(P26) **Noun type vs. number**, (Chierchia, 1998).

Under the view that nouns can come under the variety Mass or Count, one obtains the following when considering interaction with a plural/singular distinction.

[±MassN; ±Plural]. with [-MassNP] and [-Plural] representing Count Noun and Singular.

[14] My gratitude to Jeffrey Watumull for finding this pattern.

[15] Cf. Holmberg (2010b).

| | |
|---|---|
| *[+MassNP; +P] | → *The milks |
| [+MassNP; –P] | → The milk |
| [–MassNP; +P] | → The biscuits |
| [–MassNP; –P] | → The biscuit |

Generalisation: As is well known, Mass Nouns cannot be pluralised. Traditionally, it is thought that 'next to a count domain there is a second domain homomorphic to the former, which is not required to be atomic. Mass nouns take their denotation from the latter domain' (Chierchia, 1998, 346). Chierchia, however, proposes that mass nouns are inherently plural and that mass nouns are true in an undifferentiated manner of singular instances of masses (a piece of furniture) as well as of pluralities thereof (the furniture). Crucially, he claims that mass nouns are 'quite literally the neutralisation of the singular/plural distinction'. Technically, mass nouns are provided by the Lexicon under the group forming operation '∪' (cf. pp. 345, 347); hence pluralising them makes no sense because they already have a plural reading. In terms of atomic features, it could be possible to think of mass nouns as already having a feature [+Plural],[16] but, features cannot occur twice within the same bundle. Therefore, analogously with the analysis provided for P13 adding another feature [+Plural] within the same set would be inadequate. This is again an instance of the ungrammaticality of double feature occurrence. I return to this idea in section 4.

Summary: In section 3, a selection of a seemingly larger quantity of asymmetrical distributions has been presented. As shown by the widely ranging spread of linguistic domains, the notion of a 3/4-configuration is central to feature interaction; its relevance for linguistic theory is apparent.

[16] Evidence for this notion of 'inherent plural' could come from agreement contexts like:
a) I have a large family. It/They is/are very vociferous.
b) My team is great. I have so much fun with it/them.
In both cases the pronoun or verb could agree with the semantic quality of an 'inherent plural' pertaining to the given noun.

# 4 Discussion

As Greenberg (1963, pp. 58) points out about Implicational Universals: 'where the two sets of characteristics are binary, the typical distribution in a tetrachoric table is a zero as one of the four entries', in other words a 3/4 pattern. The 3/4-patterns as such are thus a well known distribution but the central question to pose is whether these characteristic 3/4 patterns underlyingly are significant *per se* – that is, *qua* abstractions of implications – for the comprehension of the FL. My hypothesis is that these configurations ultimately can be explained in terms of how features interact and affect each other within a feature structure (whether a bundle, phrase, clause or sentence). In this section I attempt to classify the patterns presented previously and to formulate reasonable abstractions (section 4.1). I then elaborate the idea of feature interaction (section 4.2), prior to discussing a more general theory of features and their nature (section 5).

## 4.1 Possible Abstractions

In order to examine the patterns from Section 3, I shall classify them according to their internal combinatorial structure, namely according to the way in which the missing output (the option with an asterisk) is yielded from the feature matrix. As apparent in (P0) the 2-by-2 factorial design generates 4 output possibilities, two harmonic and two disharmonic configurations. I posit, however, that any matrix output obeys commutativity – because the order of the two features is arbitrary – and therefore that each output is an unordered set. Thence it follows that the patterns can fall into three logically different classes as below:

a) Type A, (both Absent) - 'Harmonic Negative'

b) Type B, (Both present) - 'Harmonic Positive'

c) Type M, (Mixed) - Disharmonic

Before further analysis is possible a distinction regarding the 'level of complexity' of the features themselves has to be made. I posit the following definitions:

**Definition 1 (Atomic Feature)** *Any feature which cannot be decomposed and which, representationally, when its attribute is valued negatively, corresponds to the empty set. The two different values of such a feature thus correspond to the presence or absence of a given primordial linguistic characteristic.*[17]

Plausible examples of atomic features would be:

1. clause type features such as [+Question], [+Imperative], [+Exclamative];
2. logical negation [+NOT] (and perhaps also other logical operators, [+AND], [+OR], [+IF] etc.);
3. pronominal reference features such as [+Reflexivizer], [+Referential];
4. logophoric properties such as [+Logophoric Agent], [+Logophoric Patient];
5. number features [+Atomic], [+Augmented];
6. etc.

**Definition 2 (Feature-Complex)** *Any linguistic characteristic whose two attributes correspond to two different linguistic aspects which are complementary in some way and neither of which corresponds to the empty set. The two different varieties of such a feature-complex thus correspond to the presence of one of two distinct, somehow complementary, linguistic characteristics.*

Plausible examples of feature-complexes would be:

1. Head-Complement relationships, [±HI]

[17] This is Adger's Privative Feature System, Adger (2008), pp. 4.

2. Adpositionality, [±Preposition]

3. The FOFC characterising features of [±Head-Finality on Lower Node]

4. Morphological features such as [±Agglutinating-Morphology], [±DP-internal agreement], [±Verbal-agreement]

5. Dropping or Ellipsis features such as [±SPD], [±RPD]

6. etc.

Unlike atomic features, note that feature-complexes are more akin to the traditional conception of a linguistic switch-board parameter because they do not have a 'void value'. In a sense parameters encapsulate the idea of a binary setting between two complementary non-empty phenomena.[18] The idea with feature-complexes is that beyond the surface level of description they are, as it were, aggregates which, at some (opaque) level, differ in their internal make-up of atomic features. I return to this notion subsequently.

Having preliminarily classified the patterns, I shall now discuss each type in turn, demonstrating the different dynamics of features that operate for each type. This means investigating type A, B and M vis-à-vis the level of feature complexity - namely whether the feature is atomic or a complex.

### 4.1.1 Type A

1. Type A occurs when neither of the two features in the invalid output (the 'asterisked' one) are valued positively. The first observation to make is that it follows from the definition of atomic features that a negatively valued feature is equal to the empty set and thus the combination of two such features is also the empty set; such a combination yields the absence of both linguistic characteristics. This does not necessarily imply a 'void lexical term' or a null-utterance etc. since other features may be present in the feature-bundle too. As far as the two features are concerned the output would be 'uninteresting' but not necessarily

[18] Provided a default/marked definition of parameters is not endorsed.

ungrammatical. To illustrate this concretely, consider the patterns P24 and P25, grammatical person and number.

The following are possible subsets of feature bundles of pronouns:

[+LA; –LP; +At; –Aug; +ACC], ACC=Accusative Case

1) = [+LA; ∅; +At; ∅; +ACC]

= [+LA; +At; +ACC] → 'me'

[–LA; –LP; +At; –Aug; +ACC; +Fem], Fem=Feminine Gender

2) = [∅; ∅; +At; ∅; +ACC; +Fem]

= [+At; +ACC; +Fem] → 'her'

[–LA; –LP; +At; –Aug; +ACC; +Fem; +Masc], Masc=Masculine Gender

3) = [∅; ∅; +At; ∅; +ACC; +Fem]

= [+At; +ACC; +Fem; +Masc] → '??'; a problematic complex.

In any case, it is valid to conclude that a Type A output with two atomic features is trivial and also, importantly, that an empty output is not the same as a crashed derivation in the broadest sense. The empty output would be yielded by the absence of features while a crashed derivation is problematic somehow because of the presence of one or more features that are inadequate in the context of the given string, as illustrated above. In other words, more ontologically speaking, the empty-set output of a matrix – simply due to the sheer absence of features – is not necesarily of the same nature as an ungrammatical or nonsensical output due to the *presence* of one or more features in the relevant structure. (Cf. further comments in section 5).

2. Type A with one complex feature is plausible but that is akin to only having the complex feature (a 1-by-2 matrix), in which case the output is unlikely to be ungrammatical because, per definition, both values of the complex feature are possible on their own. The only example of this variety seems to be P13, *[–Mon; +Neg], where [Monotonicity] is a complex feature, as demonstrated, and [lexical negation – NOT], for the time being a more atomic feature. In the case of negation, again, it is a matter of arbitrary labelling but as per the definition it is the presence of the attribute that is characterised as bearing the positive

symbol and therefore the harmonic negative output corresponds to the presence of the quantifier mon↓ and the absence of the negation; as postulated, this is reducible to the presence of the [–Monotonicity] feature only.[19]

3. Type A with two complex features is logically conceivable but, to reiterate, a contradiction in terms. It is not a substantial or meaningful classification because, as I have pointed out, complex features are designated their value in an arbitrary way. To exemplify this point consider the patterns P1 and P3, the only two Type A patterns of this kind in the data set. In each case, none of the features involved are atomic. This means that P1, which is given as *[–Pro-V; –VO], could be re-thought as [+V-Pro; +OV]. In other words, the presence of two negative features is a pure accident stemming from the arbitrariness of the feature definition.

   In sum, the Type A configuration has no real *raison d'être*.

### 4.1.2 Type B

1. Type B occurs when both of the features are valued positively, namely when the two features are present simultaneously within the feature set. From the data set the relevant patterns are:

| | | |
|---|---|---|
| P7 | Clause Type | *[+Imp; +Q] |
| P11 | Square of Opposition | *[+U; +Neg] |
| P15 | Vowel quality | *[+L; +H] |
| P21 | Reflexivity | *[+Reflex; +R] |
| P24 | Number | *[+At; +Aug] |
| P25 | Person | *[+LA; +LP] |
| P26 | Noun type/Plurality | *[+MassNP; +P] |

   All of the examples from the sample (with a reservation for P11 and P26 perhaps) involve atomic features, suggesting that the presence of both features is inadequate because the two features are incompatible when they occur simul-

[19] Remember that in my analysis of this phenomenon, that would be broken down to [+Neg; +Mon].

taneously in the same set. *Videlicet*, one feature at a time on its own yields an output (cf. 2 and 3 below) and so does their simultaneous absence (cf. 4 below) – seemingly provided that another feature, which can assure a non-empty grammatical output, is present.[20] I illustrate this with pattern P21:

| | | | | |
|---|---|---|---|---|
| 1) | *[+Reflex; +R] | → | [+Reflex; +R] | → PROBLEMATIC COMPLEX |
| 2) | [+Reflex; –R] | → | [+Reflex] | → SELF |
| 3) | [–Reflex; +R] | → | [+R] | → PRONOUNS |
| 4) | [–Reflex; –R] | → | [∅] | → SE |

2. Type B with one feature of each 'level of complexity' is computable. This is seen in pattern P11, the square of opposition, where the feature [+universal quantifier] most likely is complex, while negation is atomic. A mixed-complexity Type B output would mean that the presence of the atomic feature and the given variety of the complex feature is an incompatible configuration.

3. Type B with two complex features is also plausible. To iterate, this would imply that the given varieties of the two features are incompatible with each other. Pattern P26 might be such a case, where the features [MassNP] and [Plural] could be thought of as being composed of a set of more atomic features.

4. A special case of the Type B configuration is the double occurrence of the same feature. This particular form of antagonism was noted in the patterns P13 (Double occurrence of [Negation]) and P26 ([Plural]). Here follows further examples of such antagonism:

   a) *Nall, (P13), [Negation]

   b) *The milks, *the dusts, P26, [Intrinsic Plural with Extrinsic Plural morpheme]

   c) *The formulae-s, the men-s [Extrinsic Plural morpheme]

   d) *Ran-ed, *went-ed, *sang-ed, [Past-tense morpheme]

[20] Indeed in example 4, it is apparent that at least one other feature with reference properties is present.

e) *Je vais chanterai (French, *I am going to will sing*), [Future-tense morpheme]

f) *I did not see nobody (ungrammatical in SBE), [Lexical negation with Neg-prefix]

g) *The man that which I saw [Conjunction Phrase]

h) *The scholarship-ness, *Your Highness-hood [Nominalising morpheme]

i) *Linguistic-al, *dramatic-al, *thematic-al [Adjectivising morpheme]

j) *Actor-ess, *Doctor-er, *Mathematician-er [Agent morpheme]

k) *The best-est, *the worst-est [Comparative morpheme]

l) *The boy and as well as the girl [Coordinating conjunction]

m) *One a boy [Numeral and indefinite article morphemes in D]

n) *The such that woman [Multiple Ds]

o) *The linguist to whom the prize was given to. [?Preposition]

p) *The s-song, *the w-woman, *Bach-ch, *pr-rocrastination [Phoneme repetition]

Based on the intuition that this represents an underlying more universal property of feature interaction I posit the following:

**Conjecture 1 (Feature Doubling Constraint Conjecture)** *Within the boundaries of a feature aggregate up to a certain level of complexity yet to be defined, the double occurrence of a feature is strictly prohibited.*

This conjecture is an excellent candidate for a genuine Third Factor, as it defines a fundamental cognitive constraint on language processing, which seemingly applies to any type of feature, regardless of its nature.[21]

In sum, in terms of interaction, the Type B patterns are instances of an incompatibility between features within a feature set.

[21] This can potentially be linked to Chomsky's notion of Full Interpretation, (Chomsky, 1995).

### 4.1.3 Type M

Type M is the most complex pattern type; it occurs when the invalid output is obtained from a disharmonic configuration of the two features, one feature being positive and the other negative. Most patterns in the data sample are of Type M.

There are few clear instances of Type M with only atomic features:

| | | | |
|---|---|---|---|
| P6 | Clause Types | *[–Q; +Wh] | → [+Wh] |
| P9 | Romance Interrogative | *[+Inv; –Wh] | → [+Inv] |

As indicated above, these configurations are in fact manifestations of the solitary presence of one feature without the other. The majority of patterns in the data set, however, are of Type M with mixed-complexity:

| | | |
|---|---|---|
| P2 | FOFC | *[–HFL; +HFH] |
| P4 | Conjunction Type | *[–CI; +HI] |
| P5 | Preposition Order | *[+Prep; –DirF] |
| P8 | Holmberg's Generalisation | *[–VR; +OS] |
| P10 | Syntax Copying | *[–FC; +HCL] |
| P12 | Focal Colours | *[–$C_n$; +$C_{n+1}$] |
| P13 | Monotonicity | *[–Mon; +Neg] |
| P17 | Italian Null Elements | *[–NS; +NO] |
| P18 | Radical Pro Drop | *[–Agg; +RPD] |
| P19 | Subject Pro Drop | *[–VM; +SPD] |
| P20 | DP-morphology | *[–DPM; +ND] |
| P22 | Compounding/Resultatives | *[–RC; +Res] |
| P23 | Baker's MVC | *[–Agr; +NI] |

The Type M pattern is very interesting as it demonstrates some form of 'dependence' of one (atomic) feature on another (atomic feature). Take *[–F1; +F2] to be the generic representation of the classical Type M scenario, then this means that while [+F1] can obtain on its own when [F2] is absent (or of the wrong specification, if it is

a complex feature), the reverse does not hold: [+F2] is contingent on [+F1].
Evidently this 'implicational scenario' will hold for any combination of the feature-complexity involved, as should be apparent.

### 4.1.4 Refining the explanation

As repeatedly alluded to, features come in different flavours. It is crucial to appreciate that most linguistic structures must inherently be highly complex, while their surface level can only provide a very poor mirror of the underlying feature make-up, which could be highly elaborate. One can think of atomic features as ideally atomic and feature-complexes as molecular, the latter disguising an underlying aggregate of several atomic features. I now proceed to postulate the following:

**Conjecture 2 (Atomic Feature Conjecture)** *Any linguistic phenomenon can be ascribed to an atomic feature or an interaction of atomic features present somewhere in the string computation. All LIs, and therefore all linguistic structures, in the broadest sense, are non-empty sets of atomic features, or feature-complexes of such atomic features.*

First, this conjecture follows entirely from the properties of Merge as I shall discuss in section 5.1. Second, it follows from this conjecture (AFC) that any consistent 'sub-bundle' of such an LI is therefore a subset of atomic features. These subset clusters, in turn, may be what has until now been thought of as molecular feature-complexes – these can be, but do not necessarily have to be, LIs. To be precise, the conjecture suggests that any complex feature, in fact, can be broken down into at least two or more atomic features. I shall illustrate this by considering the familiar head-parameter, which I fundamentally assume to be extremely complex but for the sake of the argument it will be as good as any other linguistic phenomenon.

The head-parameter, as coined by (Travis, 1984.) and (Koopman, 1984), can be defined as the parameter establishing complement-head order. The parameter has two settings: Head-Initial and Head-Final. Consider now, for the sake of the argument

that these two options in fact are atomic features: [±HI] and [±HF]. As should be familiar now this yields a matrix:

a) [+HI; +HF] → ?*[+HI; +HF]; Incompatible?

b) [+HI; –HF] → [+HI]; Head-Initial

c) [–HI; +HF] → [+HF]; Head-Final

d) [–HI; –HF] → [∅]; (potentially free word order?)

Generalisation: a Head-initial phrase bears the feature [+Head-Initial] and a Head-final phrase bears the feature [+Head-Final], and, crucially, not the 'feature' [–Head-Initial], which per definition is the empty set; and vice versa.

In the above case, it is not necessary but I shall assume that either a) or d) is invalid. a) could mean that both [+HF] and [+HI] are equally possible for a given structure (and that PF just 'makes do') or it could be seen as invalid because there would be a clash between the two values, an overflow of too much information, (Richards, 2004, 2009). d) could mean that no linearisation information is provided, namely that it is underspecified, and that the output is thus randomised – a nice way of accounting for a built-in optionality, *videlicet* that both outputs can be produced – or it could equally well be impossible just because no linearisation information is provided. This discussion is interesting from a more philosophical stance but from the point of view of linguistic theory the question is which theory is most useful. As indicated, my intuition is that alternative a) is the invalid one based on the analysis of Type A patterns with atomic features only, namely that alternative d) is *valid* because it is, in fact, the empty set. Moreover, a theory based on impossible feature combinations is more desirable than one which is based solely on the absence of features. The quintessential point with the example is that any complementary, feature-complex in my parlance, can be broken down in this way.

In this section, I provided an initial classification of the implicational 3/4-patterns into three types. After having formulated the AFC I now propose that Type A is in

fact not an appropriate classification. That is to say that the allegedly invalid output of the Type A variety is not problematic because if only atomic features are conceived as the fundamental building blocks of language then the simultaneous combination of two negatively valued features just yields the empty set. Following Watumull (2010), I take the empty set to be a fundamentally necessary unit of Language structure and not in any respect an invalid output or, as it were, a derivationally crashing output. The externalisation of the empty set is 'nothing' and in essence my point is that 'nothing' (at least on its own) is not ungrammatical. It immediately follows from this that there are only two types of patterns: Type B and type M. This implies that the patterns P1, P3 and P14 must then be re-thought.

### 4.1.5 Towards an insight into feature-complexes

Here I shall demonstrate an approach which might be useful in elucidating the nature of feature-complexes and in fine-tuning feature-interaction constraints. For this purpose I shall reanalyse a few patterns (P1, P3, P14 and P13) in more depth. I reproduce the relevant data for convenience.

(P1) **Romance Interrogatives:**

[±Pro-V; ±VO]

| | |
|---|---|
| [+Pro-V; +VO]. | English: John saw the girl. John saw her. |
| [+Pro-V; –VO]. | Latin: Puer puellam vidit. Puer eam vidit. |
| [–Pro-V; +VO]. | Romance: Jean voit la fille. Jean la voit. |
| *[–Pro-V; –VO]. | Unattested. |

The most straightforward reanalysis is to break down [±Pro-V; ±VO] into four individual more atomic features, as follows: [+Pro-V]; [+V-Pro]; [+VO]; [+OV]. This yields a different matrix (the irrelevant permutations are left out):

| | |
|---|---|
| [+Pro-V; +VO]. | English: John saw the girl. John saw her. |
| [+Pro-V; +OV]. | Latin: Puer puellam vidit. Puer eam vidit. |
| [+V-Pro; +VO]. | Romance: Jean voit la fille. Jean la voit. |
| *[+V-Pro; +OV]. | Unattested. |

It is immediately clear that this emerges as a Type B configuration, with [+V-Pro] and [+OV] being incompatible.

(P3) **Types of VO/OV languages**

[±Direct; ±VO], where [-Direct] = [+Inverse] etc.

| | |
|---|---|
| [+Dir; +VO] | 'Direct-VO' |
| [+Dir; –VO] | 'Direct-OV' |
| [–Dir; +VO] | 'Inverse-VO' |
| *[–Dir; –VO] | 'Inverse-OV' |

Similarly, this configuration can be thought of as:

| | |
|---|---|
| [+Dir; +VO] | 'Direct-VO' |
| [+Dir; +OV] | 'Direct-OV' |
| [+Inv; +VO] | 'Inverse-VO' |
| *[+Inv; +OV] | 'Inverse-OV' |

By postulating the more atomic features [+Dir] and [+Inv] instead of [±Dir], this pattern also emerges as a Type B configuration.

(P14) **Internal Nominal Structure**

| | |
|---|---|
| [+Arg; +Pred] | Bare arguments allowed. |
| [+Arg; –Pred] | Every noun is mass. |
| [–Arg; +Pred] | Bare arguments disallowed. |
| *[–Arg; –Pred] | Impossible |

Effectively, this is:

| | | |
|---|---|---|
| [+Arg; +Pred] | →[+Arg; +Pred] | → [+Arg; +Pred] |
| [+Arg; –Pred] | →[+Arg; ∅] | → [+Arg] |
| [–Arg; +Pred] | →[∅; +Pred] | → [+Pred] |
| *[–Arg; –Pred] | →[∅; ∅] | → [∅] |

I now integrate the feature [+Nom], for nominal.

| | |
|---|---|
| [+Nom; +Arg; +Pred] | →[+Nom; +Arg; +Pred] |
| [+Nom; +Arg; –Pred] | →[+Nom; +Arg] |
| [+Nom; –Arg; +Pred] | →[+Nom; +Pred] |
| *[Nom; –Arg; –Pred] | →[+Nom; ∅] |

It should thus be clear that this is a Type M configuration, wherein [+Nom] depends on [+Arg] or [+Pred]. [+Nom] cannot exist on its own and since the features [+Arg] and [+Pred] are not incompatible with each other, the feature set [+Nom; +Arg; +Pred] is grammatical. This makes sense as nouns cannot be semantically vacous.

(P13) **Monotonicity, revisited.**

This method does not come without caveats, however. Along the lines of reasoning above, the complex feature [±Monotonicity], whose values would be [+Mon] = Mon↑ and [-Mon] = Mon↓, could be reanalysed as two more atomic features: [+Mon↑] and [+Mon↓]. I have shown, however, that a more felicitous analysis of monotonicity takes the following from:

| | |
|---|---|
| [+Mon↑] | → Mon↑ |
| [+Mon↑; +NegOp] | → Mon↓ |

*Nota bene*: 'NegOp' is an inherent (logical) negation operator; this is not lexical negation, but rather something more akin to a logical non-veridical operator, in the sense of Giannakidou (1998).

The reason why this analysis is more desirable is simply because of economy. The former formulation entails more features in the system while the latter is more minimal, drawing on an independently needed feature.

In sum, feature-complexes give an illusion of consistency at the superficial level. They can, nevertheless, be broken down into sets of more atomic features, from which more adequate feature-constraint analyses can be drawn.

## 4.2 Conclusions regarding Feature Interactions

Here I recapitulate the findings from the discussion undertaken in section 4. If not specified, only atomic features are considered.

### 4.2.1 Interaction Types

The proposed Atomic Feature Conjecture maintains that all linguistic structures are hierarchical aggregates of LIs, which in turn are sets of linguistic features. It follows, therefore, that all linguistic structures are large-scale sets of sets of features. Within a given structure, features can either interact or not. If they do interact then the interaction can take two forms: Type B or Type M. I shall discuss the formal characteristics of each of these in turn together with a preliminary on the empty set:

Type $\{\emptyset\}$ Feature Absence

(a) **The empty set case:** When features that could logically be present within a structure are absent, then the cardinality of the given (sub)set of features is zero and its externalisation is 'nothing'.

Type B Simultaneous Presence

(a) **Feature compatibility:** If the two features are compatible with each other, both of them can be present simultaneously and the configuration will be valid.

(b) **Feature incompatibility:** If the two features are incompatible with each other, the configuration is impossible and will not obtain.

Type M Feature Dependence

(a) **Disharmonic Valid:** In this case, one feature is present and the other absent; of the two features only one is present within the feature set. The configuration obtains nevertheless. The feature which is present does not depend upon the absent feature.

(b) **Disharmonic Invalid:** In this case, one feature is present and the other absent; of the two features only one is present within the feature set. The configuration, however, does not obtain and is unattested. The feature which is present therefore depends upon the presence of the other feature.

In sum, as far as interaction is concerned, features do seem to exercise an influence on each other. The most important conclusions from the typology above is that:

- if features are compatible then they can occur simultaneously, otherwise only one or none of them can occur at a time.
- within certain boundaries, a feature cannot occur twice as this violates the Feature Doubling Constraint.
- if one feature can occur on its own but not the other, these features stand in a dependency relationship.

As has been postulated before, especially within feature-geometric theories (cf. Harley 1994; Harley and Ritter 2002a,b, etc.), the considerations adduced so far also suggest that features may be engaged in hierarchical relationships. I have concluded that features can be either antagonistic (they cannot occur together), independent (they can occur together), or dependent (one cannot occur without the other). To reiterate, in terms of relationships, features can logically be engaged in four types of constraint configurations: antagonism, independence, subordination and superordination. Presumably a given feature [F] can be engaged in all of these hierarchical relationships at once with various other features. Given that all universal implications give rise to a 3/4 pattern, I posit that they are excellent windows into establishing feature relationships of the kind just alluded to. This would be the object of further, more systematic and extensive empirical research.

### 4.2.2 Summary

In section 4, I classified the 3/4-patterns for the sake of deriving properties of features. First I showed that there were three ways of viewing the patterns but then I demonstrated that in fact there are only two fundamental types: a Type B pattern has an invalid possibility because the presence of both features is impossible – the Feature Doubling Constraint disallows a specific instance of this pattern, namely when the same feature occurs twice; a type M pattern has an invalid possibility because one feature is contingent upon the other. In essence the main idea put forward is that the output of the implicational matrix depends on the nature and the relationship between the features involved in it.

Moreover I first defined two levels of complexity of features, atomic and complex, which subsequently lead to the postulation of the Atomic Feature Conjecture; *videlicet*, all linguistic structures are sets of complex features which in turn are sets of atomic features. Finally therefore, I also stressed and exemplified that any feature-complex is in fact only an 'illusion of consistency' in that it can be broken down into (more) atomic features.

# 5 Remarks towards a Theory of Features

Against the background of what has been concluded so far, I shall here sketch a few conjectures and proposals regarding a theory of language based on the notion of features. First, I discuss the role of Merge within this theory (section 5.1); then, I revisit the architecture of the FL, focusing on the ontology of features and the fundamental principles of the language system (section 5.2); and third, I also consider remaining dilemmas (section 5.3).

## 5.1 The Role of Merge and Copying

An important question that has not yet been addressed is how the atomic features are put together into structures. In order to obtain a set of features, these have to combine somehow. Merge is argued to be a fundamental property of the language system; to be exact, a property of FLN, whose role is to fuse syntactic elements. I argue that Merge operates on a 'broader' level than that, and that it is the mechanism whereby set-formation of features is facilitated.

I shall assume the standard view of Merge in Minimalism from Chomsky (1995) onwards. I cannot indulge in a proper exposition of all the developments and fine-tunings of the assumed properties of Merge since its first proposal but I refer the reader to Watumull (2010) and his elaboration of Merge, properly defined as a mathematical function. I endorse that view of Merge and suggest that Merge should be applied to features as follows.

**Definition 3 (Fundamental Feature Merge)** *Merge is the operation which merges atomic features into sets of features. The output of this operation is an unordered set of features.*

Recently, Zwart (2009) postulated that Merge yields an *ordered pair* instead of an unordered set. Following the design of Watumull (2010), *inter alia*, I see this stipulation as erroneous. The notion of an unordered set also justifies the aforementioned postulation that feature sets obey commutativity.

When possible to externalise, a set of features can be thought of as an LI. In fact, this is very similar to the view propounded by the Distributive Morphology framework (Halle and Marantz, 1993). Merge can merge atomic features or sets of atomic features either as LIs or within a syntactic structure. It follows from this postulate that any language structure (category, phrase, sentence) is a set of sets of features; in my parlance, the Atomic Feature Conjecture follows directly from the properties of Merge. Note that it does not follow from this that the features or the sets necessarily pertain to the same cognitive dimension. By this I refer to the question as to whether there is a syntactic projection for each feature in a language structure (phrase, sentence etc.). This is a difficult question. An affirmative answer would fall under some version of the approach commonly know as Cartography (cf. overview in Cinque and Rizzi 2008), whereas a negative answer would require another perspective. To my knowledge, there are few prominent such suggestions regarding the latter, although the Nanosyntax framework or the approach taken by Giorgi and Pianesi (1997) could be good alternatives. In fact, as argued by Roberts and Roussou (2003), given the nature of the system as it is defined, for syntax a cartographic perspective is virtually necessary

In any case, the issue revolves around having either light objects in the syntax at the expense of a heavy structure or heavy objects in the syntax at the price of a more compact but less insightful structure. Solving the dilemma most likely requires a third-factor solution based around optimal processing. It is worth noting that sets of features that constitute LIs could be 'orthogonal' to the syntax structure. This is to suggest that each LI can have a feature structure (perhaps a form of a feature geometry), which is *sui generis* to each LI, and that projects in a 'different space' from the more conventional 'phrase structure' (2D-)syntax.[22]

Disregarding that particular issue, Merge must also be capable of performing copying of features or sets thereof. Conventionally this is known as Internal Merge and the only modification I wish to elucidate is that if the language structure in question is indeed a set of sets of features, then Merge should be able to copy any number of features from one set/LI to another at any time of the derivation where this is possible. This postulation provides an elegant solution to the problem of partial copying, in

[22] In this view, Agree could be a relation on the 'orthogonal' level, etc.

which it is thought that copying and spell-out processes are 'incomplete' with regard to each other (cf. Barbiers et al. 2010 for further details on the phenomenon; note that they assume a certain level of structure).

In sum, structure building of linguistic structures is a function of Merge at various levels. In each case, features, the building blocks of language constituents, are combined – Fundamental Merge (cf. pp. 63) and External Merge – or copied – Internal Merge. From this, interestingly, one can derive the DM Principle known as 'Uniformity of Syntactic Hierarchical Structure', which states that relationships among elements within words are structurally identical to those relationships that hold among words (assuming word-internal structure). This must be the case given that from the view of feature theory entertained here, there is no boundary between the intra-word and inter-word levels; a linguistic structure is merely a very complex set of sets of features.

## 5.2 Architecture of the Faculty of Language – Revisited

### 5.2.1 The ontology of features

Considering the proposals made so far, what reconsiderations are necessary for the architecture of the FL which was presented in section 2? One of the central questions is the degree of innateness. That is: where do features come from? In other words, are features innate or not? Initially, as was the case for Phonology, it was thought that (distinctive) features were innate and part of UG. As pointed out, however, there has been a gradual move towards a content-poor UG. Moreover, for instance as pointed out by van der Hulst (2008), the distinctive features do not apply to sign languages so they cannot be universal and thus not innate.[23] Given the multitude of languages in the world the number of linguistic features is most likely a legion. Therefore, proposing a 'universal pool' of features, which would be part of UG (and hence the human genome) appears highly unlikely, because of the unreasonable information load this would pose for the genome.

My view is hence that features cannot be innate and part of UG; they must all

[23] Sceptics might claim that this depends on the level of abstraction.

be acquired as part of language acquisition. Without going too far afield into metaphysics, this essentially posits that the ontological status of features is akin to the notion of mathematical objects – these are taken to exist independently of the mind albeit lacking physical reality. Features would thus be 'grasped' in the same fashion as mathematical objects are 'grasped' (cf. Burgess and Rosen 1997 for details on the philosophy of mathematics). In any case, this stance fits well with the recent trend which propounds a minimisation of the computational system and a maximisation of learning. Originally a concept within economical game theory, this theory applied to language has been summarised thus by Janet Fodor: [the FL] is a Minimax solution – 'minimize genetic information' and 'maximize/optimize the amount of learning'. (paraphrased from Fodor 2009). What I propose fits well into this conception; Language is fundamentally a very simple system, meaning that its computational mechanisms are as few as possible and as minimal as possible. By minimal I mean that each process is the simplest possible function that can fulfil the required need. The notion of 'need' is difficult to define, and it is most appropriately left for the attention of cross-disciplinary research. *A priori*, however, it entails aspects of the language system such as structure building, LF-coherence, etc. and perhaps also 'successful communication', even though that is a controversial claim. In any case, Merge is a good example of such a minimal function. The building-blocks of language, the features, are also minimal in that they are atomic, in the sense of 'impossible to split'.

Evidently the above entails considerable poverty-of-the-stimulus problems. Ergo, there must be *something* in UG facilitating feature acquisition, a feature skeleton or a feature template; within the Chomskyan framework this is inevitable. Furthermore, there must also be inbuilt in the system a propensity for perceiving and acquiring features from the PLD; a type of 'language(-feature) instinct'. Regarding templates, in section 2, the abstraction '[Feature Attribute; Value]' was suggested. As should be apparent from the previous exposition, I have argued for a very radical underspecification of feature values, according to which only positive feature specifications, in the sense of Archangeli (1988); Steriade (1995) and Penke et al. (2004), are part of the cognitive representation/manipulation within the language system. This entails that features present in the system are 'positive' and substantial only – no negative features can be acquired or be part of the system (Eisenbeiss, 2002; Samuels, 2009).

Samuels' thesis, in particular, contains much supporting evidence for the idea that at least phonological features are acquired rather than universal. This reduces the feature template, which is part of UG, to '[Attribute]', and nothing else. From a theoretical perspective this is also the most parsimonious proposal, in the minimalist sense of minimising the elements/aspects of the computation. Further confirmations of this radical underspecification proposal are left to cross-disciplinary efforts as it goes beyond the scope of this thesis.

### 5.2.2 The boundary problem

A last remark on features is motivated by the conclusions drawn so far in this thesis. Namely: what are the boundaries for feature interaction? When and at what levels within the linguistic structure do features interact and when do they not? To illustrate this point I consider different levels of feature aggregation (also cf. 52), attempting to contrast narrower, more 'word-internal' configurations with broader, more 'word-external' ones:

(E7)

| | | | |
|---|---|---|---|
| (1) | *Nall, nevery... | vs. | Not all, not every |
| (2) | *The formulae-s, the men-s | vs. | [Q The various$_{plural}$] balls$_{plural}$ |
| (3) | *He went-ed, she ran-ed... | vs. | *He did ran, *she did went |
| (4) | — | vs. | *Je vais$_{Future}$ chanter-ai$_{Future}$ |
| | | | I am going to will sing |
| (5) | — | vs. | *I did not$_{Neg}$ see no$_{Neg}$body |
| (6) | *Puer 'et-que' Puella | vs. | Puer puellaque, but *Puer et puellaque |
| | Boy and-and girl | | Boy and girl but boy and girl-and |
| (7) | — | vs. | *The linguist to whom the prize was given to. |

The point to highlight is that there are evidently boundaries for how far feature interaction stretches, namely exactly when/where the the notions of feature antagonism or feature dependence operate. Seemingly, judging from the examples in E7, the boundaries are not consistent and interaction appears to operate over varying distances ranging from very short to very long. Possible candidates for such boundaries are LIs, phrases or perhaps phases. This is potentially a very important area for

further investigation.

### 5.2.3 Recapitulation

An important notion for the theory which I have proposed is the empty set. For the sake of formality I adopt the following definition derived from algebraic set theory:

> The empty set is counted as a finite set, with order $|\emptyset|=0$. The empty set $\emptyset$ is regarded as a subset of any set A, an assumption which does not contradict the notion of subset. For the statement '$\emptyset \subseteq$A' should mean 'if $x \in \emptyset$, then $x \in A$'. Since there is no x such that $x \in \emptyset$, this last statement is never contradicted'. (From Green 1995, pp. 6).

By taking the principles of Merge and Bare Phrase Structure into account, as described above, it immediately follows from this definition that the empty set is a subset of any language structure. I hence conclude the following axiom:

**Conjecture 3 (The Empty set Axiom)** *The empty set, with the externalisation 'void' is a grammatical entity, and a fundamental building-block which is part of the language system. It is a subset of any and all possible linguistic structures.*

Essentially a linguistic structure is thus a macro-structure of feature sets, all of which provide various pieces of information/instructions for the interfaces. This is compatible with the SMT, namely that language is the optimal solution to legibility conditions. Thus, departing from the empty set, sentence structures are incrementally assembled by combining atomic features, like lego. The feature-assembly process is subject to two constraints: feature antagonism and feature dependence. Differences between languages stem from a variation in the feature make-up of the linguistic structures. Overall, a language will be very consistent in its 'feature choices' because consistent use of the same features can become automatised easily and thus efficiency of computation is achieved. There are, however, abundant exceptions to this stipulation, which by that very fact demonstrate the existence of a variation in feature-make up intra-linguistically. Consider the following examples from English:

(E9) Melvin is sufficiently able for the job.

(E10) Melvin is able enough for the job.

In E9 the adverb 'sufficiently' precedes the adjective 'able', which is the norm in English. In E10, however, the adverb 'enough', idiosyncratically but consistently follows the adjective. Under the assumption that the feature-driven theory is correct this illustrates how the adverb in E9 has a feature [X] akin to [+Initial], causing it to precede the adjective. Similarly the adverb in E10 has a feature [Y] akin to [+Final], causing it to be externalised as following the adjective. These examples demonstrate what I believe is the fundamental acting force responsible for all differences between idiolects, dialects and ultimately, on the largest scale, between different languages. In other words, the Borer-Chomsky conjecture would be correct in that variation can be localised to the LIs but the conjecture can now be 'pushed down' a level to making the more refined stipulation that the locus of language-variation factors is the features used in the language – that is to say the nature of the feature used and the ways in which complex features may or may not be assembled. Any form of asymmetry, or symmetry for that matter, is thus a product of feature distribution. In the same way as various asymmetrical or symmetrical relationships between quarks in nature make up nucleons and electrons, which in turn make up atoms, which in turn make up molecules, which in turn make up complex substances, which ultimately make up the world – analogously, linguistic features can, in various more or less symmetrical configurations, make up language structures.

I thus sketch a very sparse architecture of the FL obeying the SMT and the Minimax theorem. The idea is that Language, which primordially is a biological product of evolution, and which, as a system, thus should be subsumed under the law of Parsimony, is a compact, simple system. It comprises a finite set of operating modules and mechanisms but can operate on an infinity of contingent items thus capable of yielding a very diverse spread of outputs. The design of the system is therefore minimal while the potential output is maximal. Consider the old analogy with chess: Chess has a finite number of minimal rules (operating mechanisms), which can produce an infinity of different games of a very widespread complexity and variety.

In sum, Language is capable of maximal diversity while, as a processing system, it is also demonstrating the utmost unity across humans and across time.

## 5.3 Remaining Desiderata

In this last section I address various remaining desiderata which are crucial for the proposals put forward in this thesis but which cannot be done justice properly without going too far afield or which remain to be solved through further research. I shall speculate regarding each of the following: the problem of order (section 5.3.1); feature geometries (section 5.3.2); the freedom of merge (section 5.3.3); and, the possibility of universals (section 5.3.4).

### 5.3.1 The problem of order

The 'problem of order' refers to the question of whether or not feature sets, after having been aggregated by Fundamental Merge, are in fact strictly unordered, in the sense of the now looser definition of Merge in Chomsky (1995), or the more rigid mathematical approach of Watumull (2010). Evidently, the problem is also related to the boundary problem (cf. section 5.2.2). Intuition into this dilemma comes from configurations with scope, where hierarchy are crucial. Consider for instance pattern P13:

| | | |
|---|---|---|
| a) | [+Mon; +Not] | 'Not All' |
| b) | [+Mon; –Not] | 'All' |
| c) | *[–Mon; +Not] | 'Not few' |
| d) | [–Mon; –Not] | 'Few' |

If the sets are unordered then there could be two possibilities for both a) and c):

| | | | |
|---|---|---|---|
| a) | [+Mon; +Not] | 'Not All' | Not all people came to the party. |
| a') | [+Mon; +Not] | 'All Not' | All people did not come to the party |
| c) | *[–Mon; +Not] | 'Not few' | *Not few people came to the party |
| c') | *[–Mon; +Not] | 'Few Not' | ?Few people did not come to the party. |

It should be apparent from the above, that the order of elements and the constraints on their boundaries of interaction do matter. In c) 'Not' has scope over 'few' which causes ungrammaticality, but, provided one accepts it, in c') it is 'few' that has scope over not, which produces a grammatical derivation. The problem is thus two-fold: what does 'order' mean concretely in terms of computation? – where does this kind of feature-ordering take place – and, if Merge does not provide order, how can one account for order effects in feature interaction? The simplest way out of this dilemma is to assume that the hierarchy provided by Merge is sufficient to assign correct interpretation of scopal realtions (under C-command); thus order is an immaterial consideration for a set-theroretical approach to langauge structure.

### 5.3.2 Feature geometries

The idea that features can stand in dependence relationships is not new. Feature-geometric approaches (e.g. Harley and Ritter 2002a,b) are aimed at capturing such relationships. The question to pose then is what the locus of geometries is, namely what the cognitive nature of geometries is and how they participate in language computation. There are two main alternatives:

1. *Geometries are active objects within the syntax*: This option entails that there is a geometric template that participates in structure-building, licensing or not licensing subordinate features to be present according to the presence or absence of superordinate features. In other words, the geometric structures are necessary components of the system.

2. *Geometries are passive abstractions stored in memory*: This option entails that the geometric template is a derived abstraction stored in memory as part of acquisition. It represents feature relationships without participating in the aggregation of features. In this perspective, the features themselves are sufficient to instantiate relationships of interaction during computation without a geometric structure. In other words, the geometric structures are contingent components of the system.

This dilemma is very important and each alternative has very different consequences. On the one hand the former has the advantage of explaining how dependence relationships take effect within the syntax at the expense of presupposing more templates within the system. On the other hand, the latter is more minimal as it relies purely on features *per se* without presupposing more structure with UG, at the price of not providing a clear explanation of how the dependence relationships come about within the computation.

### 5.3.3 The Freedom of Merge

The proposal that Merge 'should be able to copy any number of features from one set to another at any point of the derivation where this is possible' (cf. pp. 64), predicts a rather chaotic situation, where Merge seemingly has total freedom. As argued by Barbiers et al. (2010), specific assumptions of structure are required to assure that Merge copy the adequate components of the structure and nothing else. In sum, Merge cannot be granted total operational freedom; it must be constrained somehow. The question is thus how the initial structure is yielded and how the system 'knows' what components are allowed to be copied. If one adopts the Atomic Feature Conjecture, then the information pertaining to what elements to copy is a feature itself, causing the copying. In other words, the system requires second-order features.[24] Enriching the system with such features might not be ideal from a perspective of minimalism, but they may be virtually necessary. In any case, such second-order features immediately entail the same problems as discussed above, namely how the 'copying-feature' is paired together with the relevant feature to copy. Given that current linguistic theory (syntax in particular) is a theory of features this type of question is of the essence for future research.

[24] In syntax, several such 'second order' features have been incorporated into the theory: feature strength, EPP etc.

### 5.3.4 The Possibility of Universals

The Atomic Feature Conjecture dictates that all language phenomena can be traced to features or interactions thereof. The way features have been thought of in this dissertation, however, is that they are cross-linguistically relative; there is no universal pool of features. (cf. pp. 65). That is to say that a feature in one language (e.g. [Feminine Gender] in German) responsible for a given phenomenon (e.g. gender agreement) does not necessarily have to be identical (and thus universal) to the feature causing a very similar phenomenon in another language (e.g. feminine gender agreement in Modern Greek). Although this severely limits the rather positive scope for universals outlined in section 2.2, such strict relativism can be moderated by the Wittgenstinian notion of *family resemblance* (cf. Wittgenstein 1953, §66-71), to allow generalisations to be formulated. Consequently the possibility of universals is not ruled out all together. It is ceratin, however, that succesful language universals must be *very* abstract and not rooted in any concrete language-specific phenomena. Language-processing properties like the interaction types Type B, in particular the Feature Doubling Constraint, and Type M or principles like the Empty Set Axiom, do probably come closer to the appropriate level of abstraction and could indeed turn out to be true absolute universals, in the sense of 'pertaining to all languages without exception'.
In sum, languages are relative at the level of features but universal at the level of processing.

### 5.3.5 Summary

This section has provided a discussion of a few caveats or remaining questions that deserve further consideration. The problems of boundary, feature-geometries, order and freedom of Merge are all related, and for a feature theory to be successful, they need to be solved. It cannot be sufficiently emphasised that these issues call for serious cross-disciplinary efforts.

# 6 Conclusion

This thesis sought to explain asymmetry in language departing from the notion of features. It was put forward that linguistic structures are merely a very complex set of sets of atomic features, which are aggregated by the fundamental process Merge. The features were taken to be radically underspecified and privative, thus only allowing substantial, positive features in the computation. Arguments supporting this view were based around minimalist optimality and the lack of psychological reality of negative features or solid evidence for the acquisition of such features. Within the assumption of the Atomic Feature Conjecture it follows that there is no boundary between the intra-word and inter-word levels. This allows for a wide conception of interaction possibilities between features within a structure, even though it remains to define those boundaries.

After examining data from various linguistic domains, the following abstractions were made: the 3/4-patterns hold across different linguistic domains which shows unity in the underlying system; and, there exist two types of constraints on feature interaction, Type B, simultaneous feature incompatibility, and Type M, feature dependence. Regarding the former type, a conjecture called the Feature Doubling Constraint was proposed, prohibiting the reiteration of the same feature within too narrow boundaries. Furthermore, bearing those relationships in mind, a proposition towards a methodology aimed at breaking down feature-complexes into more basic units was propounded, showing how seemingly complex phenomena are necessarily rooted in either of the two types of interaction patterns at a deeper level. In other words, all linguistic phenomena are yielded by interaction at the atomic feature level.

Although the theory presented in this dissertation provides deeper insight into Linguistic Theory, it does leave a few important questions unanswered. The problems of feature-interaction boundaries, and of the operational freedom of Merge, are required to be solved in order to obtain a more robust explanatory theory. Cross-disciplinary collaboration is called for to acquire answers to these important desiderata.

As far as linguistic asymmetry is concerned one can conclude that skewings, like the 3/4-distribution, are necessarily part of the system, as a consequence of its processing properties. Asymmetries are therefore excellent and useful windows into the atomic processing mechanisms of the language system. To finish, in terms of the relativist/universalist debate, it emerges from this thesis that languages are more relative on the atomic feature level but more universal on the processing level. This concretely means that there is an infinity of possible features while the human system of linguistic computation is finite, being limited in its architecture. True absolute language universals are the foundation axioms of this architecture.

## References


Adger, D., D.H. and Watkins, L. (2009). *Mirrors and Microparameters: Phrase Structure Beyond Free Word Order*. Cambridge: CUP.

Adger, D. (2003). *Core Syntax*. Oxford: OUP.

Adger, D. (2008). A minimalist theory of feature structure (to be published). *http://ling.auf.net/lingBuzz/000583* .

Aldridge, E. (2007). The Heterogeneity of Austronesian VOS order. Unpublished ms: University of Washington (Seattle). [available on-line under Variation in Austronesian VOS order at: http://faculty.washington.edu/eca1/manuscripts.shtml].

Archangeli, D. (1988). Aspects of Underspecification Theory. *Phonology* 5, 183–207.

Bach, E. (1968). *Universals in Linguistic Theory*. Holt, Rinehart & Winston.

Baker, M. (1996). *The Polysynthesis Parameter*. Oxford: OUP.

Baker, M. (2001). *'The natures of non-configurationality'*, 407–438. Oxford: Blackwell.

Baker, M. (2008). *The Syntax of Agreement and Concord*. Cambridge: CUP.

Baker, M. and McCloskey (2007). On the Relationship of typology to Theoretical Syntax. *Linguistic Typology* 11, 273–284.

Barbiers, S., Koeneman, O. and Lekakou, M. (2010). Syntactic doubling and the nature of wh-chains. *Journal of Linguistics* 46, 1–46.

Barwise, J. and Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy* 4, 159–219.

Berlin, B. and Kay, P. (1969). *Basic Color Terms, their universality and evolution*. University of California Press.

Berwick, R. and Chomsky, N. (2008). Biolinguistics: Its current state of development. *Unpublished ms: MIT.* .

Biberauer, T., Holmberg, A. and Roberts, I. (2010). A syntactic universal and its consequences. *(Unpublished ms: Universities of Cambridge and Newcastle)* .

Biberauer, T. (2008). 'Introduction' *in The limits of Syntactic Variation*. Amsterdam: Benjamins.

Bobaljik, J. (2002). A-chains at the interface: copies and covert movement. *Natural*

*Language and Linguistic Theory* 20(2), 197–267.

Bobaljik, J. and Thrainsson, H. (1998). Two heads arent always better than one. *Syntax* 1, 37–71.

Bobaljik, J.D. (2008). Missing persons: A case study in morphological universals. *The Linguistic Review* 25, 203–230.

Boeckx, C. (2008). 'Approaching parameters from below' in *The Biolinguistic Enterprise: new perspectives on the evolution and nature of the human language faculty,*A. Di Sciullo and C. Boeckx (eds.). Oxford: OUP.

Boeckx, C. (2006). *Linguistic Minimalism.* Oxford: OUP.

Boskovic, Z. (2009). Semantic correlates of the NP/DP parameter. *Proceedings of NELS* 39.

Brown, D. (1991). *Human Universals.* McGraw-Hill.

Burgess, J. and Rosen, G. (1997). *A Subject with No Object.* Oxford: OUP.

Carlson, J. (2010a). Towards an understanding of the bilingual experience: What are the effects of bilingualism on cognition? Cambridge University 2010.

Carlson, J. (2010b). Understanding the bilingual experience: a brief psychological appraisal. Cambridge University.

Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics* 6, 336–405.

Chomsky, N. (1957). *Syntactic Structures.* The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of Syntax.* Cambridge, Mass: MIT Press.

Chomsky, N. (1968). *Language and Mind.* Harcourt Brace Jovanovich.

Chomsky, N. (1986). *Knowledge of language: its nature, origins and use.* New York: Praeger.

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

Chomsky, N. (1999). Derivation by phase. *MIT occasional papers in Linguistics* 18.

Chomsky, N. (2005). Three factors in Language Design. *Linguistic Inquiry* 36, 1–22.

Chomsky, N. (2006). *Language and Mind.* Cambridge: CUP.

Chomsky, N., Belletti, A. and Rizzi, L. (2002). *On Nature and Language.* Cambridge: CUP.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English.* New York: Harper and Row.

Cinque, G. (2008). *Universals of Language Today.* Berlin: Springer.

Cinque, G. and Rizzi, L. (2008). The cartography of styntactic structures. *Studies in Linguistics CISCL Working Papers* 2.

Clements, G. (1985). The geometry of phonological featrues. *Phonology Yearbook* 2225–52.

Comrie, B. (1989). *Language and Linguistic Typology.* Chicago: University of Chicago Press.

Croft, W. (2002). *Typology and Universals.* Cambridge: CUP.

Dayal, V. (2009). *Bare NPs crosslinguistically' in The Handbook of Semantics (to appear).*

Dechaine, R.M. and Wiltschko, M. (2002). Decomposing pronouns. *Linguistic Inquiry* 33, 409–442.

Deutscher, G. (2005). *The Unfolding of Language.* New York: Henry Holt and Company.

Dryer, M.S. (2009). Problems testing typological correlations with the online WALS. *Linguistic Typology* 13, 121–135.

Dryer, M. (1992). The Greenbergian word order correlations. *Language* 68, 81–138.

Eisenbeiss, S. (2002). *Merkmalsgesteuerter Grammatikerwerb: Eine Untersuchung zum Erwerb der Struktur und Flexion von Nominalphrasen.* Ph.D. thesis, University of Duesseldorf.

Evans, N. and Levinson, S.C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 429–492.

von Fintel, K. and Matthewson, L. (2008). Universals in Semantics. *The Linguistic Review* 25, 139–201.

Fitch, T., Hauser, M. and Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition* 97, 179–210.

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

Fodor, J. (2009). Contribution to Round Table on Linguistic Universals in *Of Minds and Language: The Basque Country Encounter with Noam Chomsky,* M. Piatelli Palmarini and J. Uriagereka and P. Salaburu (eds.), 199–206. Oxford: OUP.

Fodor, J. (1983). *The modularity of Mind.* Cambridge, Mass: MIT Press.

Fodor, J. (2005). Reply to Steven Pinker 'So How Does The Mind Work?'. *Mind and Language* 20.

Giannakidou, A. (1998). *Polarity Sensitivity as (Non)Veridical Dependency.* Amsterdam: Benjamins.

Gianollo, C., C.G. and Longobardi, G. (2008). 'Three fundamental issues in parametric linguistics' in *The Limits of Syntactic Variation*, 109–142. Amsterdam: Benjamins.

Giorgi, A. and Pianesi, F. (1997). *Tense and Aspect: From Semantics to Morphosyntax.* New York and Oxford: Oxford University Press.

Green, J.A. (1995). *Sets and Groups.* Chapman & Hall.

Greenberg, J. (1963). 'Some universals with special reference to the order of meaningful elements' *in Universals of Language*, 73–113. Cambridge, Mass: MIT Press.

Halle, M. and Marantz, A. (1993). *The View from Building 20.* Cambridge, MA: MIT Press.

Harbour, D. (2007). *Morphosemantic Number.* Dordrecht: Kluwer.

Harley, H. (1994). Hug a tree: Deriving the morphosyntactic feature hierarchy. *Papers on phonology and morphology: MIT Working Papers in Linguistics* 21, 289–320.

Harley, H. and Ritter, E. (2002a). Person and number in pronouns: a feature-geometric analysis. *Language* 78, 409–442.

Harley, H. and Ritter, E. (2002b). *Pronouns: Features and representation.* Amsterdam: Benjamins.

Harvey, P.H. and Pagel, M.D. (1991). *The comparative method in evolutionary biology.* Oxford: OUP.

Haspelmath, M., Dryer, M.S., Gil, D. and Comrie, B. (2008). WALS Online. URL `http://wals.info/index`.

Hauser, M., Chomsky, N. and Fitch, T. (2002). The Faculty of Language: What is it, who has and how did it evolve? *Science* 298, 1566–1579.

Hauser, M. and Bever, T. (2008). A biolinguistic Agenda. *Scienec* 322, 1057–1059.

Hawkins, J. (1983). *Word Order Universals.* Academic Press.

Hawkins, J. (1994). *A performance theory of order and constituency.* Cambridge: CUP.

Hermon, G. (2009). Language typology and universal grammar: a commentary on the paper by Eric Potsdam. *Natural Language and Linguistic Theory* 27, 773–787.

Hockett, C. (1960). The Origin of Speech. *Scientific American* .

Holmberg, A. (1999). Remarks on Holmberg's generalization. *Studia Linguistica 53/1* 1–39.

Holmberg, A. (1986). *Word order and syntactic features in the Scandinavian languages and English.* Ph.D. thesis, Stockholms Universitet.

Holmberg, A. (2010a). 'Null subjec Parameters' *in Parametric Variation. Null Subjects in Minimalist Theory*, 88–124. Cambridge: CUP.

Holmberg, A. (2010b). Referring tou yourelf in Self-talk. (SyntaxLab, Cambridge University, 23 February 2010).

Horn, L.R. (1989). *A natural history of negation.* Chicago: University of Chicago Press.

van der Hulst, H. (2008). On the question of linguistic universals. *Linguistic Review* 25, 1–34.

Hyman, L. (2008). Universals in phonology. *The Linguistic Review* 25, 83–137.

Jackendoff, R. and Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser and Chomsky). *Cognition* 97, 211–225.

Jackendoff, R. (1983). *Semantics and Cognition.* Cambridge, Mass.: MIT Press.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford: Oxford University Press.

Jakobson, R., Fant, G. and Halle, M. (1952). *Preliminaries to speech analysis.* Cambridge, MA: MIT Press.

Jenkins, L. (2002). *Biolinguistics.* Cambridge: CUP.

Kayne, R. (2010). Why are there no directionality parameters? (Conference talk -

Linguistic Variation in the Minimalist Framework; 15 January 2010, Barcelona).

Kester, E.P. (1996). Adjectival inflection and the licensing of empty categories in DP. *Journal of Linguistics* 32, 57–78.

Kiparsky, P. (1973). "Elsewhere" in phonology in *A festschrift for Morris Halle*, 93–106. New York: Holt, Rinehart & Winston.

Kiss, K. (2008). Free Word Order, (Non-)Configurationality and Phases. *Linguistic Inquiry* 39, 441–476.

Koopman, H. (1984). *The syntax of verbs.* Dordrecht: Foris.

Lardiere, D. (2009a). Further thoughts on parameters and features in second language acquisition: a reply to peer comments on Lardieres Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research* 25, 409–422.

Lardiere, D. (2009b). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research* 25, 171–225.

Legate, J. (2001). The configurational structure of a non-configurational language. *Linguistic Variation Yearbook* 1, 61–104.

Levinson, S. and Wilkins, D. (2006). *Grammars of Space: Explorations in Cognitive Diversity.* Cambridge: CUP.

Longobardi, G. (2003). Methods in Parametric Linguistics and Cognitive History. *Linguistic Variation Yearbook* 3, 101–138.

Longobardi, G. and Roberts, I. (2010). Universals, diversity and change in the science of language. *Lingua* Unpublished.

Lounsbury, F.G. (1956). A Semantic Analysis of the Pawnee Kinship Usage. *Language* 32, 158–194.

Mairal, R. and Gil, J. (2006). *Linguistic Universals.* Cambridge: CUP.

Matthews, P. (1991). *Morphology.* Cambridge: CUP.

Mobbs, I. (2008). *'Functionalism', the design of the language faculty, and (disharmonic) typology.* Master's thesis, Cambridge University.

Montrul, S. and Yoon, J. (2009). Putting parameters in their proper place. *Second Language Research* 25, 291–311.

Müller, G. (2008). Syncretism without Underspecification in Optimality Theory: The Role of Leading Forms. *http://www.uni-leipzig.de/ muellerg/mu233.pdf. Universität Leipzig* .

Müller, G. (2010). On Deriving CED Effects from the PIC. *Linguistic Inquiry* 41, 35–82.

Munaro, N. (1997). *Sintagmi interrogativi nei dialetti italiani settentrionali.* Ph.D. thesis, Universita di Padova.

Neeleman, A. and Szendröi, K. (2007). Radical pro-drop and the morphology of pronouns. *Linguistic Inquiry* 38(4), 671–714.

Newmeyer, F. (2005). *Probable and Possible Langauges.* Oxford: OUP.

Newmeyer, F. (2008). Universals in Syntax. *The Linguistic Review* 25, 35–82.

Newmeyer, F., J. (2004). Against a parameter-setting approach to typological variation. *Linguistic Variation Yearbook* 4.

Noyer, R. (1992). *Features, positions, and affixes in autonomous morphological structure.* Ph.D. thesis, MIT.

Odden, D. (2005). *Introducing Phonology.* Cambridge: CUP.

Pearson, M. (2000). *The Derivation of VO and OV*, 327–364. Amsterdam: Benjamins.

Penke, M., Janssen, U. and Eisenbeiss, S. (2004). Psycholinguistic evidence for the underspecification of morphosyntatic features. *Brain and Language* 90, 423–433.

Pinker, S. and Jackendoff, R. (2004). The faculty of language: what's special about it? *Cognition* 95, 201–236.

Pinker, S. (1994). *The language Instinct.* W. Morrow and Co.

Pinker, S. (1998). *How the Mind Works.* London: Allen Lane.

Pollard, C. and Sag, I.A. (1994). *Head-driven phrase structure grammar.* Chicago: University of Chicago Press.

Potsdam, E. (2009). Austronesian verb-initial languages and wh-question strategies. *Natural Language and Linguistic Theory* 27, 737–771.

Ramchand, G. and Svenonius, P. (2008). 'Mapping a parochial lexicon onto a universal semantics' in *The Limits of Syntactic Variation*, 219–246. Amsterdam: Benjamins.

Reinhart, T. and Reuland, E. (1993). Reflexivity. *Linguistic Inquiry* 24, 657–720.

Richards, M. (2004). *Object Shift and Scrambling in North and West Germanic. A Case Study in Symmetrical Syntax. PhD Dissertation.* University of Cambridge.

Richards, M. (2009). Two kinds of variation in a minimalist system. *Linguistische Arbeitsberichte* 87, 113–162.

Rizzi, L. (1982). *Issues in Italian Syntax.* Dordrecht: Foris.

Rizzi, L. (1986). Null objects in Italian and the theory of pro. *Linguistic Inquiry* 17, 501–557.

Roberts, I. (1985). Agreement Parameters and the Development of English Modal Auxiliaries. *Natural Language and Linguistic Theory* 3, 21 – 58.

Roberts, I. (2010). *Agreement and Head Movement: Clitics, Incorporation and Defective Goals.* Forthcoming.

Roberts, I. and Holmberg, A. (2010). *Parametric Variation. Null Subjects in Minimalist Theory.* Cambridge: CUP.

Roberts, I. and Roussou, A. (2003). *Syntactic Change: A Minimalist Approach to Grammaticalization.* Cambridge: CUP.

Sadock, J. (1974). *Toward a linguistic theory of speech acts.* New York: Academic Press.

Sagey, E. (1986). *The representation of featrues and relations in non-linear phonology.* Ph.D. thesis, MIT.

Saito, M. (2007). Notes on East Asian Argument Ellipsis. *Language Research* 43, 203–227.

Samuels, B.D. (2009). *The Structure of Phonological Theory.* Ph.D. thesis, Harvard University.

Sato, Y. (2010). Radical Pro Drop and Fusional Pronominal Morphology in Colloquial Singaporean English: Reply to Neeleman and Szendröi. URL `http://ling.auf.net/lingBuzz/000991`.

Scalise, S., Magni, E. and Bisetto, A. (eds.) (2009). *Universals of Language Today.* Springer.

Sigurdsson (2003). Meaningful silence, meaningless sound. *Ms* .

Sigurdsson (2004). The syntax of person, tense, and speech features. *Italian Journal of Linguistics* 16, 219–251.

Smith, N. and Law, A. (2009). On Parametric (and non-parametric) variation. *Biolinguistics* 3(4), 332–343.

Snyder, W. (2001). On the nature of syntactic variation: evidence from complex predicates and complex word-formation. *Language* 77, 324–342.

Snyder, W. (2007). *Child Language. The Parametric Approach.* Oxford: OUP.

Steriade, D. (1995). 'Underspecification and Markedness' in *Handbook of Phonological Features*, 114–74. Blackwell.

Travis, L. (1989). 'Parameters of Phrase Structure' in *Alternative Conceptions of Phrase Structure*, 263–279. Chicago: University of Chicago Press.

Travis, L. (2008). Bahasa Indonesia: a window on parameters. *Lingua* 118, 1583–1602.

Travis, L. (1984.). *Parameters and effects of word order variation.* Ph.D. thesis, MIT, Cambridge, Mass.

Ullman, M. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nature Reviews Neuroscience* 2, 717–726.

Wallace, A. and Atkins, J. (1960). The Meaning of Kinship Terms. *American Anthropologist* 62, 58–80.

Watumull, J. (2010). *Merge as a Minimax Solution to the Optimization Problem of Generativity.* Master's thesis, Cambridge University.

Wierzbicka, A. (1972). *Semantic Primitives.* Frankfurt: Athenum.

Wierzbicka, A. (1987). Kinship Semantics: Lexical Universals as a Key to Psychological Reality. *Anthropological Linguistics* 29, 131–156.

Wierzbicka, A. (1996). *Semantics: Primes and Universals.* Oxford: OUP.

Wittgenstein, L. (1953). *Philosophical Investigations.* Oxford: Blackwell.

Yang, C. (2005). On productivity. *Linguistic Variation Yearbook* 5, 265–302.

Zwart, J.W. (2009). Relevance of typology to minimalist inquiry. *Lingua* 119, 1589–1606.

Zwicky, A. (1977). Hierarchies of person. *Chicago Linguistics Society* 13.