# Structures and strings

by

Virginia Savova

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

October, 2006

# Abstract

A central question for linguistics is how strings and structures relate to one another. The majority of frameworks that exist today assume that the hidden structures contain complete information about word order. Hence, the theory of grammar is only a theory of the structure-generating component. However, it is possible to design a theory of grammar which views the structure-generating component as separate from the linearization component. As a result, structural descriptions can directly incorporate multi-dominance, and consequently eliminate the need for transformational devices like movement. Another benefit is that linearization can take into account any combination of structural, morpho-phonological and discourse features, paving the way for a word order typology in the style of Optimality Theory. This dissertation presents an analysis of basic word order typology in this formalism, and suggests that cross-linguistic frequency of word orders can be explained if constraints are viewed as priors for Bayesian iterative learning. Finally, it discusses the role of information structure and morpho-phonology for language-internal word order.

Author: Virginia Savova

Adviser: Dr. Robert Frank, Professor of Cognitive Science, Johns Hopkins University

# Acknowledgements

**The following acknowledgments are due:**

To my grandmother, for teaching me the English language, and playing silly card games to motivate me.

To my mother, for teaching me the intricacies of linguistic expression.

To my father, for teaching me about the human psyche.

To my husband, for introducing me to AI and machine learning.

To my Greek and Latin teachers in highschool, for letting me know how syntax really works.

To my professor at Harvard, who insisted that I should study linguistics.

To my professor at Harvard, who taught me that “what doesn’t happen is interesting”.

To my professor at MIT, who started it all.

To my professor at Johns Hopkins, who made it all fit together.

To my adviser, for letting me find my own way, while gently pushing in the right direction.

You taught me well: Vesela Kotchemidova, Christina Kotchemidova, Zdravko Savov, Leon Peshkin, Ogi Radev, Lili Iordanova, Calvert Watkins, Sam Epstein, Noam Chomsky, Paul Smolensky, Robert Frank.

Thank you!

# Contents

# List of Tables

# List of Figures

# Part I

# Introduction

## 0.1 Overview

Linguistic expressions are like icebergs: the string of words we hear is the small observable part above water, while the crucial syntactic and semantic information hides beneath. The objective of syntactic theory is to capture the shape of the hidden object and the nature of its correspondence with the observable string.

Thus, a central question for linguistics is how strings and structures relate to one another. The majority of frameworks that exist today assume that word order must be accounted for exclusively in structural terms. In phrase-structure grammar and its successors, structures are trees on which dominance and precedence relations are exhaustively specified. The words of the linguistic expression constitute the leaf nodes of the tree, and the precedence relations between them determine the string. Hence, the theory of grammar is only a theory of the structure-generating component. A theory of structure-string mapping is not required, beyond defining the latter as lower-dimensional projections of the former. However, it is possible to design a theory of grammar which views the structure-generating component as separate from the linearization component, which could take into account any combination of structural, phonological and discourse facts. Under this view, structural relationships influence the string form, but may be overwritten by more powerful considerations.

This dissertation will show that separating the theory of structures from the theory of strings is particularly fruitful in conjunction with an Optimality Theoretic view of word order variation. I will explore a theory of grammar in which structural descriptions capture the abstract syntactic relationships among words but make no direct reference to word order. Instead, the word order corresponding to the structure emerges through the interaction of mutually contradicting ranked constraints.

While such a system can be built on the basis of any number of existing formalisms, this work is situated within the dependency grammar framework. The choice is motivated by a desire to avoid unnecessary theory-driven structural complexity at this early stage, particularly because a significant portion of the massive theoretical machinery accumulated by constituent-based formalisms is due to the fusion of hiearchical structure with precedence. In defining a system that treats these components separately, it is preferable to start with the most transparent type of structures.

What is stucture? No pre-theoretical reason forces us to confine this notion to any particular set of abstract objects. There is general agreement that conditional and semantic dependencies among words, such as thematic roles and subcategorization frames, should be reflected at the structural level, be it configurationally or substantively. While some approaches treat agreement as a conditional dependency, other choose to attribute it to psycholinguistic factors external to the grammar. The situation is similar with respect to discourse features, such as prominence or novelty. In some frameworks these are presented on a par with selectional properties of lexical items, while others leave them to external components.

Some of these questions are partly contingent on the extent to which word order should treated as part of the strunctural description. If word order is taken to be a purely structural consequence, any difference in word order implies a structural difference. As a result, all factors which influence word-order have to be reified in the structure, albeit in the most economical way. Consider optional displacement phenomena such as scrambling. It has been shown that scrambling has no effect on either truth-conditional meaning or grammaticality. An order-free theory of structure can assign scrambled versions of the same sentence to a single structure. In contrast, a fixed-order theory has to assign them to different structures, even if they contain the same set of words and correspond to the same truth conditions. Theories of grammar that do not rely on a rich linearization component often have no other choice, but to treat word order as a de facto structural feature. A radical version of this approch is spelled out in Kayne's antisymmetry proposal [Kayne, 1994]. If a constituent is displaced from the expected postion mandated by its hierarchical place, it must be assigned to a different hierarchical position. Including word order as a structural feature is often circumvened by assuming that structure reflects not only truth-conditional semantics but also discourse or pragmatics. This is achieved through incorporating discourse features such as Topic and Focus into syntax. However, such an approach betrays the original idea that syntax is supposed to explain the difference between grammatical and ungrammatical strings. Instead, theory turns toward explaining felicitous versus unfelicitous strings within all possible contexts- a somewhat poorely defined enterprise.

If a difference in word order is assumed to imply structural difference, there is a massive increase in the number and type of structures permited within and across languages. It also guarantees that structures themselves cannot be thought of as invariant across languages. Given the cross-linguistic variety of word orders, the invariants are limited to abstract aspects of structures. In contrast, if word order was divorced from the notion of structure, sentences across languages that differ only in word order can be assigned the same structure, and the theory of structure itself can become the theory of invariance.

An important benefit from separating hierarchical descriptions from precedence is that structural descriptions can freely move to graphs instead of trees. While phrase-structure (PS) grammar was at first an attempt to formalize the traditional analysis of sentences in keeping with the tree-like scheme, it soon moved on to using more complex objects as structural descriptions within the transformational framework. Trees were replaced by tree sequences, on which various restrictions were imposed by subsequent formalisms. Abandonning simple tree structures was necessitated by the abundant evidence of multi-headedness (or multi-dominance) in natural language.

Multi-headedness occurs when a word appears to depend on more than one parent in the dependency structure of the sentence. It has long been noted in most frameworks that the wh-word in questions like (1) is both dependent on the questioning auxiliary, and on the main verb.

(1) What did the king buy?

Thus, the data-driven representation of dependencies should look like Figure (0.1).

$$
\begin{array}{lll}
did & & \\
\downarrow & \searrow & \\
what & \longleftarrow & buy \\
& \swarrow & \\
king & & \\
\downarrow & & \\
the & &
\end{array}
$$

Figure 0.1: Wh-questions involve multidominance

Similarly, according to at least one view, the structure of (2) should be Figure (0.2).

(2) The king liked owning camels.

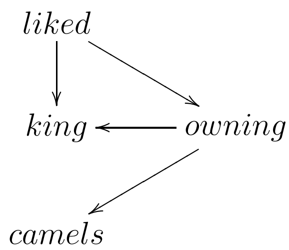



Figure 0.2: Control as multidominance

These type of phenomena are responsible for the introduction of theoretical devices such as movement and control. Movement is an operation applied to PS trees, first conceived with the advent of Transformational Grammar (TG). While it was thought of as a monolitic operation, it has been recently redefined in terms of two ordered primitive operations Copy and Delete ([Chomsky, 1995], pp.251-252). The first operations copies an object from one position in the PS tree to another. The second removes the phonological features of one the copy. For example, the question in (1) has the following simplified derivation history (Figure 0.3):

The history essentially enforces the relationship between the wh-elements and each of its heads in stages. Stage one (the D-structure) captures only the role of "what" within the subcategorization frame of "buy". Stage two creates the structural dependency of "what" and "did" by inserting the former into the specifier position of the latter. Stage three is a resolution of the resulting repetition. The paralellism between multi-headed representation and movement is self-evident. A more technical discussion of the mapping is provided in Chapter 3.

Multi-headedness involved in cases like (2) are handled through control theory, largely because movement between the sites involved is theoretically undesirable. Unlike the wh-word above, a movement account of (2) would involve positions semantically tied to two different lexical verbs. This entails that one verb will fail to assign its thematic role. Control theory simply states that the understood subject of an infinitival clause must be anaphorically dependent on a specific argument of the matrix clause.

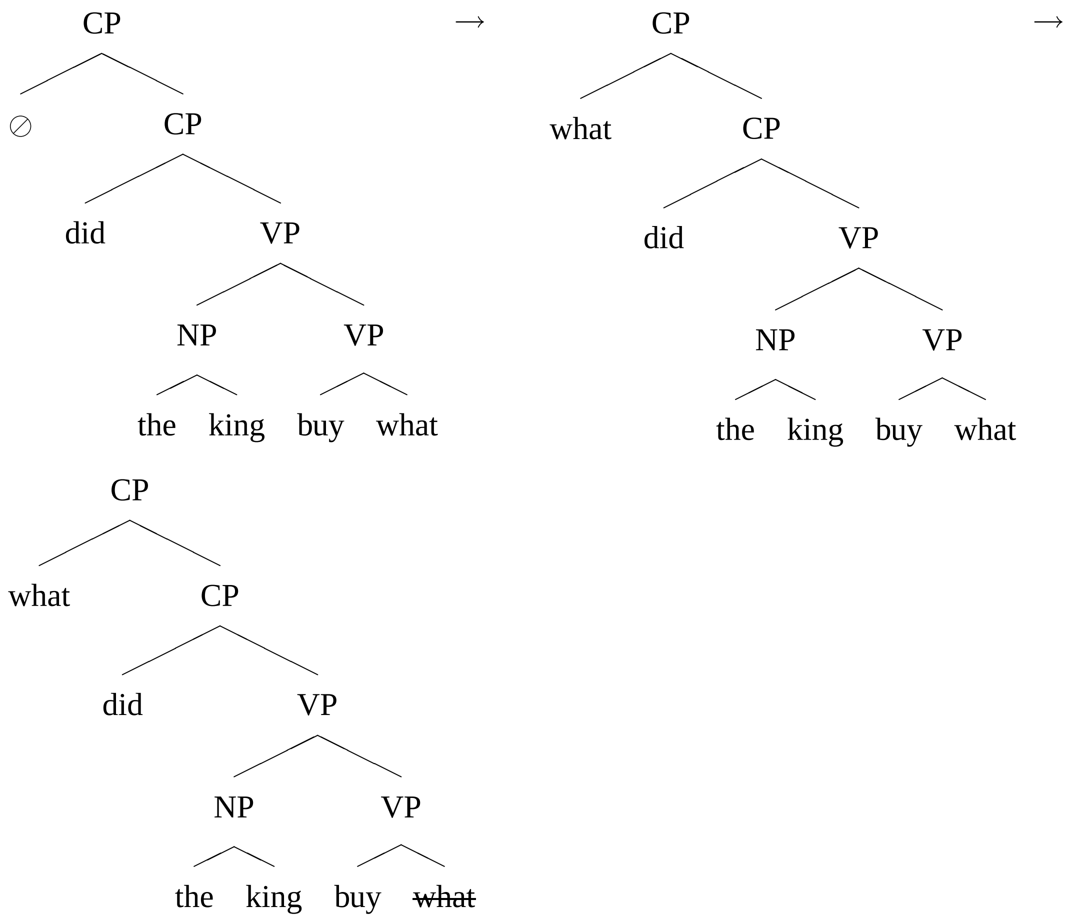



Figure 0.3: Minimalist structure of example (1)

Movement and control make it possible for syntactic theories to retain their tree representations. This is important because strings can be trivially read out from trees according to precedence. It is much more complicated to define precedence on an arbitrary graphical object. However, if precedence is to be dissociated from structure, it is much more transparent to represent multi-headedness with generalized acyclic graphs.

## 0.2 Organization of the dissertation

Chapter 1 focuses on the ways existing syntactic theories approach constituent discontinuity and long-distance dependencies. Theories are divided into two categories according to their attitude to structure-string correspondence. The strict correspondence class of theories abides by the principles of Exclusivity and Nontangling. Hence word order variation is necessarily reflected at the level of structure. I discuss the role of theoretical devices such as transformations, traces, movement and control in maintaining the strict correspondence assumption. The other class consists of approaches that explicitely separate dominance and precedence relationships. They do not abide by Exclusivity and Nontangling, which allows them to reflect word order variation independently of structure. I review various proposals for restricting the structural component, and solutions to the structure-string mapping problem, with particular emphasis on theories espousing multidominance.

Chapter 2 lays out problems in representing discontinuity in both classes of approaches. The strict correspondence approach runs into two major empirical issues: first, discontinuity and long-distance dependecies do not always show the hierarchical effects expected under the strict correspondence assumption; and second, hierarchical effects not always imply surface discontinuity- a problem that covert movement was designed to deal with. In addition, the approach raises some unsettled theoretical issues. Since word order variation within a language is largely discourse-driven, strict correspondence is forced to maintain that structure reflects features of discourse. This leads to a watered-down notion of structural grammaticality by putting discourse factors on a par with much more rigid selectional and agreement requirements. Furthermore, the strict-correspondence approach guarantees that structures will vary vastly from language to language, since the only way to account for cross-linguistic word-order variation is to make it structural. Finally, the requirement that each word order should correspond to a different structure leads to lack of parsimony in the number of possible structures required for language description. Representing discontinuity raises a different set of questions for the ID/LP approaches. Some fail to provide an adequate, or even explicit theory of word order given the dominance representations. But even those that venture into defining the linearization component fail to propose a cross-linguistic theory of linearization.

Chapter 3 introduces my solution to the issues raised above, in the form of multi-headed hierarchical dependency representation complemented by an OT-style linearization. I motivate and formally define the structural component of the grammar, including multidominance and locality restrictions on it. Then, I turn to the linearization component discussing the candidate set, the mode of evaluation and the types of constraint that may be involved

Chapters 4 and 5 demonstrate how the approach handles a cross-linguistic analysis of word order typologies in basic declarative sentences.

Chapter 5: Essentially, most types of observed word orders can be explained with five linearization constraints acting on a multi-headed dependency structures. Even more importantly, the same set of constraints excludes the vast majority of unobserved, but logically possible, word order permutations. This result is important because variation in basic word order is not accompanied by hierarchical effects, since in so far as we can tell, basic word order sentences mean the same thing cross-linguistically. Thus, basic word order variation presents a challenge to the strict correspondence assumption, because of the lack of observable differences of hierachical structure under different orders.

Chapter 5 presents an evolutionary view of cross-linguistic frequency of word orders. I propose that the constraints identified in Chapter 5 can be viewed as learning biases, or priors in Bayesian iterative learning. Thus, instead of a categorical destinction between possible and impossible word orders, quantitative predictions can be obtained.

Chapter 6 examines the role of discourse and phonology in linearization. Discourse-features are viewd as a stamp on dependency structures, rather than a property of the structure itself. This allows discourse-driven contrasts of word order to be attributed to linearization constraints targeting discourse features. The absence of such features inactivates the constraints, leading to neutral word order. As a result, the notion of structural grammaticality can be separated from the notion of discourse-appropriateness. A tentative analysis of clitic phenomena is also presented.

# Part II

# Word Order in Syntactic Theory

# Chapter 1

# Conflicts of structure and order

## 1.1 Discontinuity

The assumtpion that constituents must be linearly adjacent can be traced at least as far back as early American structuralism. It dominates theoretical syntax to this day, despite the fact that it has roused numerous empirical objections. Researchers who assume a tree-like organization of dominance relations can be roughly divided into two camps. Those that favor a relaxed version of structure-string correspondence are primarily driven by empirical considerations, while those that do not, tend to advance formal arguments for strict constituent continuity. Phenomena listed as problematic by the former include but are not limited to wh-questions, wh-relative clauses [Pike, 1943], definite noun phrases [Wells, 1947], particle verbs [Chomsky, 1955], parentheticals [McCawley, 1982], cross-serial constructions of the type found in Dutch and Swiss-German [Ojeda, 1988], as well as the existence of free word order [Pullum, 1982], VSO [Anderson and Chung, 1977, Blevins, 1990] and V2 [Kathol, 1995], [Kathol, 2000] languages. The response of the continuity camp since 1965 [Chomsky, 1965] has been generally to argue that the job of linear order rules can be substituted by operations on trees. The debate also relates to the stance of a third group of researchers, who question the assumption that tree structures provide the best descriptions of dominance relations. The formalisms which they advocate inevitably turn linearization into a non-trivial procedure requiring serious theoretical exploration.

Not all discontinuity effects have the same properties. Some, usually referred to as *long-distance dependencies* (LDD), exhibit linear displacement associated with hierarchical effects. The displaced element not only appears in a non-canonical position, but also posesses unexpected semantic scope and syntactic binding properties. For example, the original position of the wh-phrase before displacement authorizes coindexation between the wh-phrase and the pronoun in (3), compared to (4):

(3) Which picture of $himself_i$ does $Bill_i$ like best?

(4) *This picture of $himself_i$ is important to $Bill_i$?

Other disconinutities seem to have no known hierarchical consequences. For convenience, I will refer to them as string discontinuities (SD). The displaced element appears in a non-canonical position, but continues to be interpreted in the usual way (6).

(5) Bill woke the cat up.

(6) Bill woke up the cat.

While such word order variation often seems to convey nuances of discourse or aleviate comprehension, it has no bearing on the semantic and syntactic status of the sentence.

Generally speaking, those who support rigid linearization treat the lack of interpretive effects of SDs as a grammatical accident, and advance structural explanation for both LDDs and SDs. Similarly, it is often the case that those who argue for relaxing the linearization principles of phrase-structure grammars treat both LDDs and SDs as the product of linearization constraints, attributing differences of binding and interpretation to non-structural, often extra-grammatical factors. If they do acknowledge the structural aspect of LDDs, it is captured through devices separate from the linearization constraints, such that their string does not follow from its hierarchical position, but is a mere coincidence. Proposals that choose the middle ground are few and far between.

In 1933, Bloomfield [Bloomfield, 1933] proposed that the scientific study of language can be carried out via "immediate constituent analysis" -a procedure by which a sentence is recursively subdivided into contiguous subcomponents. Bloomfield's follower Kenneth Pike [Pike, 1943] pointed out difficulties in applying IC-analysis to sequences involving WH questions and relative clauses. He writes:

> There are many forms which at first sight one might regard as containing non-contiguous members of a constituent... Thus, *When he comes, I will go home* might be analysed as composed of the constituent *I* and the constituent *When he comes, will go home*...; similarly *Who did he send?* might be subdivided into *he* + *who did send* (partially analogous to *John/ran away*).

He goes on to suggest that "positional pressure" must weigh in against such analyses, favoing instead the contiguous solutions *Who/did he send* and *When he comes/ I will go home*. However, he is ambivalent as to whether "positional pressure" will be enough to force contiguous analyses like *Has/he gone*, because they would entail

> a rather complicated description of contituents such as *he gone*, and hence on the basis of convenience one would probably prefer to treat *has...gone* as non-contiguous members of a single constituent.

Similar considerations lead Zelig Harris [Harris, 1945] to seek a generalized notion of morpheme, which would allow it to apply to broken sequences of morphemes. Another Bloomfield follower, Rulon Wells [Wells, 1947] noted that the IC-analysis of a phrase like *the English king* would yield "constituents of a much greater independence and mobility" if discontinuity is allowed than if it were not (resulting in *English* and *the king*, or *the* and *English king*, respectively). He suggests that, in the discontinuous case,

> the pattern of *poor/John* and *English/literature* will be better immitated.

Wells is worried that discontinuity will make IC-analysis a "tremendously intricate affair" because "the possibilities requiring investigation would be enormously multiplied". The assumption of continuity would result in a "more orderly and manageable procedure". Thus, he seems to arrive at a preliminary formulation of later formal objections referring to the combinatorial explosion of candidate analyses. He ends up cautiously welcoming discontinuity, and provides a restrictive principle about where it should be sought, namely

A discontinuous sequence is a constituent if in some environment the corresponding continuous sequence occurs as a constituent in a construction semantically harmonious with the construction in which the discontinuous sequence occurs.

In his early adoption of string transformations, Chomsky [Chomsky, 1955] initially sided with those advocating a relaxed view of structure-string correspondence. String transformations can permute the leaves of the syntactic tree in arbitrary ways, which can be naturally represented as a tree with crossing branches ([Ojeda, 2005]). Thus, to address the problem of floating particle verbs, one could use a transformation like $[VParticleNP] \rightarrow [VNPParticle]$ and obtain *woke the cat up* from *woke up the cat* (1.1).

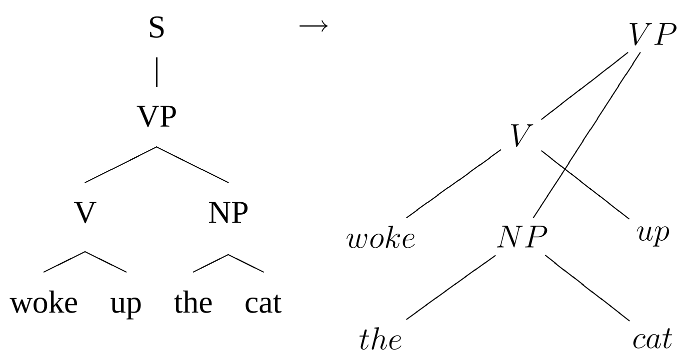



Figure 1.1: Chomsky's early view of split constituency

Other approaches place discontinuity at the level of the generative component. Yngve's computational model of sentence production also allows discontinuous constituents via special expansion rules of the form $X \rightarrow Y...Z$ [Yngve, 1960]. These wrapping rules require category $X$ to be expanded into $YWZ$, where W is the right sister of $X$. They were designed to deal with expletive subject chains in sentences like *It is true that he went*. Wraping rules were also proposed in categorial grammar by Bach [Bach, 1965] [Bach, 1984] and Bunt [Bunt, 1996] to account for particle verbs, quantifier raising and gapping.

Several formalizations and discussions of the dominance versus precedence distinction in mainstream generative grammar also deserve mentioning in this context. Curry

[Curry, 1961] proposes unordered constituent expansions of the type $S \rightarrow \{NP, VP\}$. He identifies two levels of grammar - tectogrammatics and phenogrammatics, built by recursive string functions (functions from strings to strings, functions from functions from strings to strings to strings). The tectogrammatics of a syntactic object is the history of function applications, the phenogrammatics is the result of evaluating these functions. Sixteen year later, Lasnik and Kupin [Lasnik and Kupin, 1977] point out that tree structures are accidental representations of generatve rules, and that it is possible (and empirically necessary) to define reduced phrase markers which cannot always be represented in the form of trees. Specifically, they propose relaxing the standard assumption is that every two objects in a syntactic structure are ordered with respect to precedence and dominance. This idea was taken up by Goodall [Goodall, 1984] in his treatment of coordination and gapping phenomena, in which he describes projection algorithms for linearizing the hieararchical structure. Goodall's formalization is implemented in a parser of conjunction and related constructions [Fong and Berwick, 1985]. A similar notion was used by Hale [Hale, 1983] in his description of Walpiri mobiles, where precedence realtions inside the verb phrase are completely relaxed. Finally, Bromberger and Halle [Bromberger and Halle, 1989] emphasize that the role of precedence in phonology is as crucial as the role of hierachy in syntax, thereby implying that syntax, unlike phonology, is entirely hierarchical in nature.

The advent of GPSG and subsequent representational formalisms prompted a renewed interest in separating the structural and the linearization component of the grammar. The starting point is the observation that a phrase-structure rule like $S \rightarrow NP\,VP$ is equivalent to two statements independent of each other [Pullum, 1982] [Gazdar et al., 1985]. One statement declares the immediate dominance relation between S and its constituents, while the other statement refers to the precedence relations among them. Thus, S is a possible mother of NP and VP, and if NP is a sister of VP, it precedes it. The original proposal limited precedence statements to the sisterhood relation due only to considerations of generative complexity. A metagrammar with this limitation can be easily shown to generate context-free grammars, which is desirable for computational reasons. A small relaxation of the limitation which still preserves generative capacity is offered by Pullum [Pullum, 1982], who proposes a metagrammar allowing subconstituents to scramble one level up from their phrases. For example, given the metagrammar in (7) a CFG like (8) can be generated.

(7) XP $\rightarrow$ YP[F] X YP $\rightarrow$ ZP[F] Y

(8) XP $\rightarrow$ ZP[F], Y, X

For every ID statement that allows XP to dominate a YP with feature set F and head X, there is another ID stament which allows XP to dominate X, Y and the complement of Y directly. LP rules can be defined on the second rule, resulting (seemingly, but not underlyingly) in one-level-up scrambling.

Ojeda ([Ojeda, 1988],[Ojeda, 2005]) has since extended the proposal to license any type discontinuity. Rather than specifying the order of sisters, he proposes partial ordering constraints refering to heads of phrases such as the one in (9) or (10)

(9) If NP and VP are sisters, the head of V must precede the NP.

(10) If NP and VP are sisters, the head of V must precede the NP.

The first rule allows him to account for VSO word order by generating the tree in (1.2). The second applies to the cross-serial dependencies in Dutch.

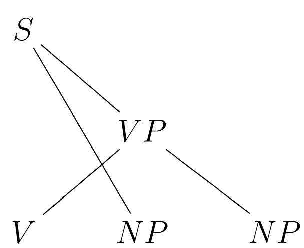



Figure 1.2: VSO word order in Ojeda's account

Reape tackled the same type of cross-serial dependencies from an HPSG perspective. In HPSG, each level of syntactic structure is a merger of the feature structures of its constituents, and defines a word order domain. According to the Constituent Ordering Principle, the word-order domain of the mother is defined in terms of the word-order domains of its daughters. Reape notes that the principle allows the mother domain to be composed either by concatenation or by interleaving. While mainstream research upholds constituent contiguity by choosing concatentation, he proposes substituting concatenation with the more general DOMAIN UNION operation. Given two sequences $\langle A, B\rangle$ and

$\langle C, D \rangle$, concatenation results in $\langle A, B, C, D \rangle$, while domain union yields several outputs: $\langle A, B, C, D \rangle$, $\langle A, C, B, D \rangle$, $\langle A, C, D, B \rangle$, $\langle C, A, D, B \rangle$, $\langle C, D, A, B \rangle$ and $\langle C, A, B, D \rangle$. Unlike concatenation, domain union is not a function, but a ternary relation over sequences. The sequence union of the sequences $\sigma_1$, $\sigma_2$ and $\sigma_3$ is true, iff each of the elements in $\sigma_1$, $\sigma_2$ occur in $\sigma_3$ in their original order.

(11) daßes ihm jemand zu lesen versprochen hat.
that it(Acc) him(Dat) someone(Nom) to read promised has

This allows Reape to present an HPSG grammar of Dutch, German and Swiss German cross-serial dependencies. The feature structure of each entry can specify linear precedence constraints for each of its possible sisters. When the entry is merged with another entry, the word-order domain is non-deterministically determined by domain union, subject to the LP constraints. For example, the sentence in (11) is derived by the sequence of mergers in Figure (1.1). The analysis assumes the LP constraint $NP > V$, which guarantees, for example, that the NP *es* is positioned before the verb *zu lesen* in the first merger. However, the constraint does not preclude another NP *ihm*, from being interlieved inside the word order domain *es zu lesen*. Hence, the problematic cross-serial dependencies example in (10 is non-deterministically derived (i.e. it is one of all possible derivations achievable through domain union).

Reape also derives German subordinate clauses involving extraposition. To do so, he suggests that extraposed constituents (marked [+Extra]) cannot be unioned. Furthermore, he posits the LP constraint [-Extra] precedes [+Extra]. This allows him to analyze sentences like (12) as the result of the derivation in Figure (1.1)

(12) daßder Mann versucht hat, das Buch zu lesen.
that the man tried had the book to read .
that the man had tried to read the book.

The normal word order of German subordinate clauses is SOV. In example (12), the VP *das Buch zu lesen* is extraposed. Hence, it is concatenated with *versucht* subject to the LP constraint. At the next step, *der Mann* and *hat* are unioned with the resulting larger VP constituent. Presumably, some linear order constraint bars *hat* from preceding *versucht*.

Reape's approach is best conceptualized as a theory of unordered D-structure, partially linearized via the projection of constituent domains. LP constraints determine the permissible order of elements within and across merged domains. Relative order is inherited monotonically bottom-up throughout the derivation. It is desined to

> incrementally get words *in order* by imposing partial constraints, rather than get them *out-of-order* via movements.

A very similar proposal is advanced independently by Dowty [Dowty, 1995], who defines a "minimalist" theory of syntax with the following components:

1. a Categorial Grammar with compositional semantics;
2. a default operation for combining two expressions into an unordered multiset;
3. a set of Linear Precedence Principles (LPP) that partially or fully determine the allowed word orders;
4. a set of Bounding Categories that delimit the domain of LPPs
5. an ordering operation
6. an attachment operation which "glues" two expressions into an unseparable whole

$$
\begin{tikzcd}
 & {[_S[_{NP_3}es][_{NP_2}ihm][_{NP_1}jemand][_{V_3}zulesen][_{V_2}versprochen][_{V_1}hat]]} & \\
{[_{NP_1}jemand]} \arrow[ur] & & {[_{V_1}hat]} \arrow[ul] \\
 & {[_S[_{NP_3}es][_{NP_2}ihm][_{V_3}zulesen][_{V_2}versprochen]]} \arrow[uu] & \\
{[_{V_2}versprochen]} \arrow[ur] & {[_S[_{NP_3}es][_{V_3}zulesen]]} \arrow[u] & {[_{NP_2}ihm]} \arrow[ul] \\
{[_S[_{NP_3}es]} \arrow[ur] & & {[_{V_3}zulesen]} \arrow[ul]
\end{tikzcd}
$$

7. a “Categorial Grammar” style morphological agreement system (e.g. unification-based)

This system allows one to analyse free word order languages like Finnish and rigid word order languages like English with equal ease. For example, Dowty claims (after Kartunnen) that, in Finnish, all permutations of S, O, V are acceptable as basic word order. This is easy to fomalize by postulating that no LPPs govern the relative position of NPs and V. Furthermore, Dowty’s theory accounts the ability of Finnish constituents to scramble out of an infinitival clause (data by Kartunnen (13)) with the simple asumption that VP is not a bounding node in this language.

(13) En minä näissä ole tennistä aikonut ruveta pelaamaan.
En minä tennistä näissä ole aikonut ruveta pelaamaan.
En minä tennistä ole aikonut näissä ruveta pelaamaan.
not I tennis have intend these-in start play.
I do not intend to start playing tennis in these (clothes).

Unlike Finnish, English treats infinitival VPs as bounding categories, prohibiting sentenses like (14)

(14) *I do not intend tennis in these (clothes) to start playing.

Another point of rigidity of English word order is its reluctance to separate verb and object with adverbials or parentheticals (15).

(15) I easily could have been passing the note to her.
I could easily have been passing the note to her.
I could have easily been passing the note to her.
I could have been easily passing the note to her.
$*$ I could have been passing easily the note to her.

Similarly, prepositions cannot be separated from their NP complements (16)

(16) *I left the note on, probably, the table.

This can be accounted for by the principle in (17, slightly restated for clarity)

(17) Whenever a functor combines with a NP complement in English, they combine via the operation of syntactic attachment.

Dowty extends the analysis to account for some peculiar properties of the English object pronouns, by imposing the additional requirement that the functor and pronominal complement must form a phonological phrase. Thus, while both (18) and (19) are permissible by (17), only (20) is allowed when the NP is pronominal.

(18) look the answer up

(19) look up the answer

(20) look it up

(21) *look up it

The ungrammaticality of object pronouns in double object constructions (22) follows from the phonological phrase requirement and Principle (17).

(22) *I gave Mary it.

The pronoun cannot be separated from the verb, but the direct object cannot be either. Hence, the accusative object can only be expressed as a full argument.

I will conclude the part of the literature review devoted to relaxation approaches by mentioning two recent dissertations entirely devoted to evidence for discontinuity and its implications. Andreas Kathol [Kathol, 1995] proposes and linear account of German clausal syntax, based on the notion of topological fields. Topological fields are equivalence classes of a partition of the clausal word order domain and correspond to clusters of material whose position is claimed to be essentially orthogonal to issues of constituency. Earlier, James Blevins [Blevins, 1990]

> explores the consequences of relaxing the undermotivated constraints that prohibit discontinuous and converging configurations.

In addition to discussing the already familiar Germanic cross-serial dependencies and English extraction, he provides an account of VSO languages (Welsh, Breton, Irish and Niuean) with an arboreal model of phrase-structure which admits discontinuity and multidomination (multidominance). His work naturally brings us to the next section, in which we review arguments for and against multidominance.

## 1.2 Multidominance

Proposals for allowing multidominance are motivated by the type of discontinuities classified as LDDs. Recall that unlike SDs, LDDs exhibit hierarchical effects presumably associated with their linear position. In Chomskian phrase structure grammar, they are handled through more or less limited tree transformations. Some researchers have maintained that tree transformations are merely an inconvenient way to preserve the undermotivated assumption about the tree character of structures. Transformations essentially generate a sequence of trees as the syntactic representation of a sentence. In his attack on transformational grammar, Karlgren ([Karlgren, 1976] argues that the formal properties of tree sequences are less understood than properties of graphs, and hence the latter should be the preferred formal device. Another formal argument against tree sequences is their representational redundancy. All relations unaltered by the transformation are repeated in every member of the sequence. While Karlgren proposes several graph grammars of differing genarative complexity, he presents no discussion of their application to linguistic structural description.

Other multidominance proposals have been advanced on the basis of empirical data. Unfortunately, they are primarily motivated through the problematic phenomena of pronominalization and deletion, instead of more clear-cut movement cases. [Sampson, 1975]'s argument begins with the observation that many transformations require identity, and no transformations require non-identity. He points out that the reflexive pronoun transformation requires identity and coreference, even as identity is redundant if coreference is present.

(23) $George_i$ likes $George_i$
$\# \, George_i$ likes $George_j$
$\#$ *$George_i$ likes $Dick_i$

$\# \, George_i$ likes $Dick_j$

In example (23, *mine*), $*$ denotes an impossible D-structure $\#$ denote the cases in which the refelxive transformation fails to apply. Sampson notes that nothing hinges on the coreference distinction when identity is lacking, since a D-structure with one but not the other is impossible. Similarly, cases of the control transformation require identity of the main and subordinate subject, while no analogous transformation exists for non-identical subjects. Presumably, we could imagine a language where the subordinate subject gets substituted by a non-identity pronoun as in (24, *mine*).

(24) George wants (*some other*) George to ride a camel. $\rightarrow$ George wants *not-he* to ride a camel.

However, such language does not exist. While tree structures require us to postulate this fact separately (in the recoverability principle[1] -n.a.), graphs naturally allow restricting transformations only to identical entities, which are subsumed under a single merged node.

Sampson also shows that multidominance can resolve the Bach-Peters paradox of pronominalization in infinite structures. The paradox claims that, if pronominalization is substitution, then the deep structure of sentences like (25) must be infinite. Suppose we undo the sustitution of *it* and replace with its original referent. The result is shown in (26). Now, lets undo the substitution of *him*. We are left with something like (27), which has the same number of *him* pronouns, prompting an infinite regression. The mechanisms for applying any sort of transformation to infinite sentences is unclear.

(25) The pilot who shot at it hit the MIG that chased him.

(26) The pilot who shot at the MIG that chased him hit the MIG that chased him.

(27) The pilot who shot at the MIG that chased the pilot who shot at the MIG that chased him hit the MIG that chased the pilot who shot at the MIG that chased the pilot who shot at the MIG that chased him.

[1] an element may be deleted only if it is fully determined by a structurally related phrase, or if it is a 'designated element'. (Lexicon of Linguistics [Ackema et al., ])

Suppose, however, that the structure of sentences like (25) is a graph with shared nodes for the pronouns and their referents (Figure 1.3). Then pronouns can be conceptualized as place holders for the double dependency arrows.

Anderson [Anderson, 1979] refutes Sampson's suggestion to explain pronominalization through graph-based representations. He notes that the recoverability principle already accounts for the lack of non-identity transformations. Furthermore, he points out that identity is not necessary for coreference, as evidenced by sentences like (28).

(28) That he had first seduced the elder $sister_i$ and then the younger was denied by the $first_i$ of his alleged victims.

As for the lack of transformations that apply to coreference but not identity, he stipulates that only identical items are marked for coreference.

Another argument by Anderson against Sampson's proposal is based on the fact that sometimes identical and coreferential material appears in multiple positions. Compare (29) with (30). Sampson suggests that (30)is the natural realization of a structure containing a single multi-headed object *go to Poland*.

(29) John went to Poland because he wanted to.

(30) John went to Poland because he wanted to go to Poland

What, then, is the structure of (29)? Assuming the two examples have different structure requires an explanation how both of them differ in a semantic sense. If both are realizations of the same structure, what sort of transformation will come up with (29)?

While these objections raise some relatively minor issues, the most substantial critique by Anderson lies in the seeming lack of an adequate linearization procedure. Take the structure in Figure (1.4).

What set of linearizing tranformations would account for the allowed and disallowed linarizations in (34)?

(31) John allowed those girls who wanted to to go to London

(32) John allowed to go to London those girls who wanted to

(33) ?John allowed those girls who wanted to go to London to

(34) *John allowed to those girls who wanted to go to London

Anderson formulates the following condition:

> The sequence of the doubly governed V is determined relative to the higher governor.

This excludes the unacceptable possibilities. However, it conflict with the data in (36)

(35) Because John desparately wanted to go to London, Fred allowed him to.

(36) Because John desparately wanted to, Fred allowed him to go to London.

Assuming that because is subordinate to the main clause, the first sentence should be unacceptable. Assuming the opposite, the second sentence should be unacceptable. Thus, no adequate linearization procedure can be formulated. The same objections apply to pronominalization transformations.

(37) That he is unpopular pleases John.

(38) That John is unpopular pleases him.

It appears that, according to Sampson, the examples in (38) should derive from the same deep structure. To what, then, should we attribute the difference in pronouns?

Despite the difficulties in linearization, Anderson is not entirely ready to give up the idea of multidominance. He proposes that such representations might be required to explain some properties of object raising. The dual status of the subject-object noun phrase is evident in the fact that it obeys the ban on part subject extraction, despite its surface object position (39).

(39) *Who do you expect stories about to terrify John?

This leads Anderson to conclude that the raised object is at the same time a subject of the lower clause in a dependency structure which represents both as a single node, dominated both by the lower and the upper verb. The question then is what prevents multiple dependencies from appearing elsewhere. The answer, according to Anderson, is that they are only permitted if they do not lead to violations of projectivity [2]. In such cases, linearization conflicts do not arise.

Hudson's treatment of LDDs in word grammar almost a decade later [Hudson, 1984] is similarly a balancing act between the desire to avoid projectivity violations and the need to allow multidominance. It starts with the observation that many displaced elements seem to be dependent on multiple heads. Such representations are the basis for formulated exceptions in the linearization specifications of the string. However, this means the linearization of such structure will contain projectivity violations. This is solved by representing the syntactic structure as a D-structure and S-structure pair. The D-structure contains multi-headed elements whose dependencies are pruned down to derive the S-structure via some (unspecified) type of "competition". At S-structure, unilke D-structure, projectivity holds.

The minimalist conceptualization of movement as re-merge has also given rised to the recent interest in multidominance. As Fitzpatrick and Groat [Fitzpatrick and Groat, 2005] write:

> Multidominance structures, which arise in a theory of syntax where the input to merge [...]is unrestricted and movement is in fact (re)merger ([Epstein et al., 1998], [Gärtner, 1999], [Chomsky, 2001] who dubs this internal merge, [Kracht, 2001], [Starke, 2001], [Wilder, 1999], inter alia), solve several empirical, computational, and conceptual problems in current syntactic theory. Remerger allows the elimination of xerox-copying, which is crucial in more traditional copy theory [Nunes, 2004], and thus removes the need to mark distinct elements as copy-identical through indices or chains (both violations of the hypothesis of inclusiveness, [Chomsky, 2001]). It also explains why lower copies seem to be affected by feature checking/valuation of higher copies. Furthermore, remerger, which is the simple addition of a relation/locational pointer to a structure, is inherently simpler than the computationally unwieldy copy operation, which we argue is nearly as computationally complex as re-

[2] A dependency is projective if its dependent is not separated, in the linear sequence, from the governor by anything apart from descendents of the governor.

building the relevant structure entirely. Finally, remerger explains the puzzling existence of displacement in human language: As [Chomsky, 2001] notes, with unrestricted merge, the surprising thing would be lack of displacement.

A detailed formalization of multidominance in minimalism is presented in Gartner's dissertation [Gärtner, 1997].

## 1.3 Transformations and the Linear Correspondence Axiom

Despite the proposals to relax structure-string correspondence and the occasional flirting with multidominance, the overwhelming majority of syntactic theories throughout the twentieth century have sought other ways of dealing with displacement phenomena. The representational approaches have portrayed LDDs as boundedly local constraints on structure while their derivational counterparts have gone the way of a maximally restricted transformational component. As far as SDs are concerned, strict tree approaches have had no choice but to reduce them to just another type of LDDs. Given the strict tree assumption, any change in an element's position in the string must be due to a change in its position in the dominance hierarchy.

In the transformational tradition after [Chomsky, 1965], a displaced element is typically seen as raised to a higher structural position, such as SpecCP, FocusP, and various adjunction sites associated with maximal projections. In the early days of Transformational grammar, a D-structure tree (such as in Figure 1.6) was transformed into an S-structure tree via some sequence of truly non-local rules, defined over indefinite strings of elements.

The D-structure was taken to represent the underlying logico-syntactic function of the element, while the S-structure position was merely responsible for its surface realization. Thus, sentences with identical truth conditions were generated with the same D-structure, but given the ephemeral role of S-structure, transformations were merely a way to license displacement.

A different view of S-structure gradually emerged with the necessity to impose locality conditions on movement transformations. Ross [Ross, 1967] noted that certain constituents represent barriers for the application of transformation operations. His island conditions included, for example, the Complex NP Constraint [3] The restrictions were soon afterward subsumed under the Subjacency condition [Chomsky, 1977b], which forbids movement across more than one bounding node. What nodes constitued bounding nodes was taken to be a language-particular parameter.

This approach to within-language word order variation was parallel to the one taken towards across-language variation. Faced with the differences in basic word order in particular, the strict correspondence camp adopted a number of strategies, only one of which involved minor relaxation of the correspondence assumption. The Principles and Parameters approach suggested that language-specific precedence relations among sister nodes are determined according to the value of a binary Head Parameter.

**Head Parameter:** The head precedes its complement.

The difference between SVO and SOV languages was due to the parameter being on in SVO languages, and off in SOV languages (1.7). This was a departure from strict correspondence, since the same hierarchical structure gives rise to different strings.

With the help of an additional, specifier placement parameter, two other word orders VOS and OVS can be generated. However, these to are far less common than the VSO order, which cannot be generated in this system. Under strict correspondence, VSO must be due to differences in the hierarchical organization of simple sentences, at least at the level of S-structure. According to one proposal, while the structure of Subject-Verb-Object(SVO) languages involved a hierarchical asymmetry between subject and object, the structure of VSO languages was assumed to be flat ([Schwartz, 1972],[Tallerman, 1990], [Stenson, 1981], [McCoskey, 1979], [McCoskey, 1980], [Awbery, 1976], Figure 1.8). Other proposals attributed VSO to verb movement to I or C (1.2). The flat structure analysis has been extensively argued against both by oponents and followers of the strict correspondence approach, mainly on the basis of evidence for subject/object asymmetry, and verb-object constituency.

The movement approach has the advantage of answering these objections. It is brought

[3] The Complex NP Contstraint (CNPC):No element contained in an S (CP) dominated by an NP with a lexical head noun may be moved out of that NP by a transformation.

to its logical extreme in Kayne's antisymmetry proposal [Kayne, 1994]. According to Kayne, all word orders are derived through movement from the basic SVO order, and without the help of a Head parameter. Precedence relations stand in one-to-one correspondence with asymmetric c-command relations, defined as follows:

**Asymmetric c-command:** X asymmetrically c-commands Y iff X c-commands Y and Y does not c-command X.[4]

Thus, X precedes Y if and only if X asymmetricaly c-commands Y. This translates into the following axioms:

**Axiom 1:** The specifier precedes the head.

**Axiom 2:** The head precedes the complement.

The second axiom holds because the complement is a phrasal constituent, and hence, the head asymmetrically c-commands all of its subconstituents.

Thus, the only available unrderlying word order is the Specifier-Head-Complement order. To explain the presence of head-final structures, Kayne's proposal relies extensively on movement operations. If a complement precedes its head, this constitutes evidence for raising from an underlying position c-commanded by the head, to a specifier position c-comanding the head (1.10). Kayne's fundamental premise is that linear order must be determined by hierarchical structure, and hence different linear orders must be associated with different hierarchical structures. His work paints what is probably the most articulate portrayal of strict correspondence to date. Since the goal of this dissertation is to provide a viable alternative to strict correspondence, the next chapter will lay out some philosophical and empirical arguments against Kayne's view. I will then explain why current alternatives of strict correspondence are not an adequate response to his approach.

[4] X c-commands Y iff X does not dominate Y and every node dominating X also dominates Y.

$$\begin{tikzcd}[column sep=small]
 & {[_S[_{NP} der Mann][_V versucht][_V hat][_{VP} das Buch zu lesen]]} & \\
{[_S[_{NP} der Mann]} \arrow[ur] & {[_{VP}[_V versucht][_{VP} das Buch zu lesen]]} \arrow[u] & {[_V hat]} \arrow[ul] \\
{[_{VP} das Buch zu lesen]} \arrow[ur] & & {[_V versucht]} \arrow[ul] \\
{[_{NP} das Buch]} \arrow[u] & {[_V zu lesen]} \arrow[ul] &
\end{tikzcd}$$

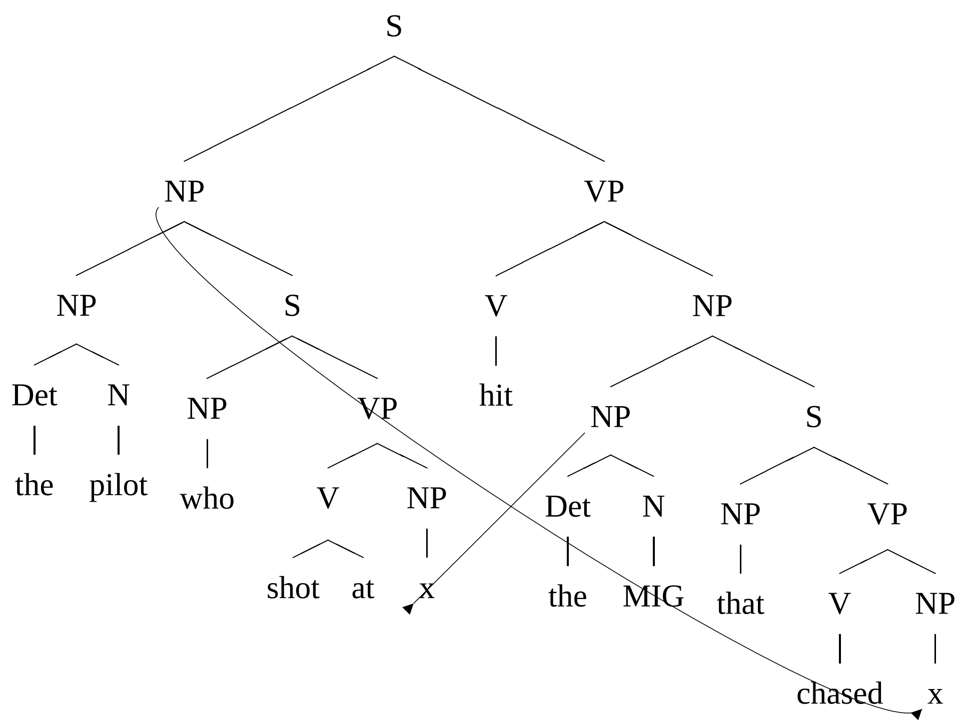



Figure 1.3: Sampson's solution to the Bach-Peters paradox

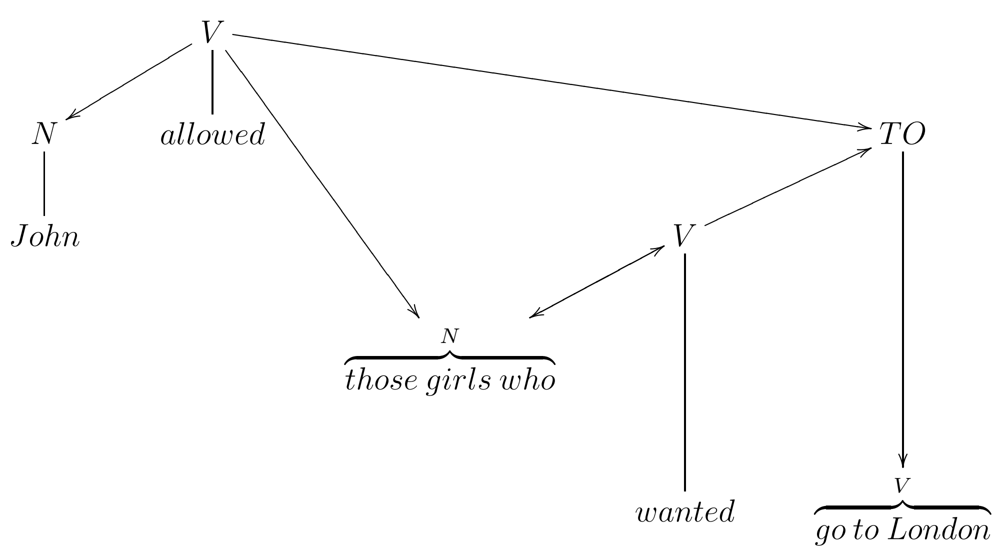



Figure 1.4: Multidominance presents problems for linearization

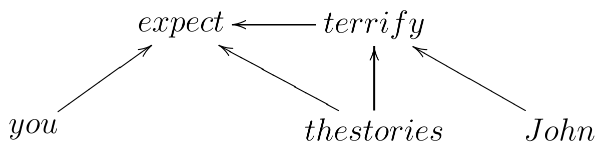



Figure 1.5: Object-raising according to Anderson

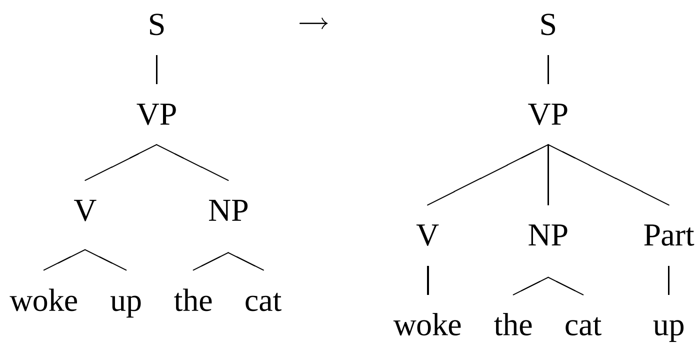



Figure 1.6: Chomsky's view of displacement


Figure 1.7: The Head Parameter


Figure 1.8: SVO versus VSO structure: the flat structure approach

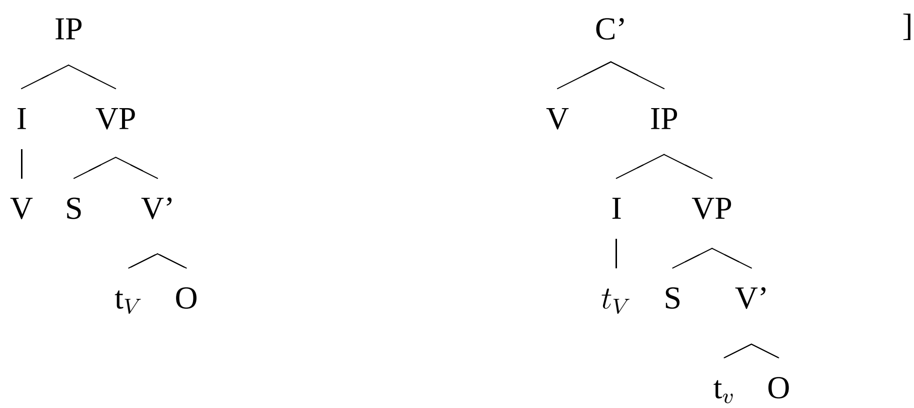



Figure 1.9: SVO versus VSO structure: the movement approach

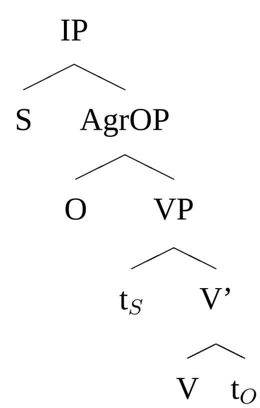



Figure 1.10: SOV: the movement approach

# Chapter 2

# Complexity and Universality

In Chapter 1, I discussed past and current views on the extent to which word order reflects hierarchical structure. The strict correspondence approach advocated by [Kayne, 1994], and often assumed in previous work, maintains that some aspect of syntactic dominance exhaustively determines precedence. It logically implies that any cross-linguistic or language-internal word order variation must result from corresponding differences in structure. Let us first examine how this implication affects linguistic analysis. One consequence is that languages turn out to differ along the dimension of descriptive complexity. Recall that Kayne's version of strict correspondence defines precedence as equivalent to asymmetric c-command. This constraint on structure-string mapping produces the shortest structural descriptions for SVO languages like English. For example, the Latin structure in Figure (2.1) is simply larger than its English counterpart. Thus, under any straightforward encoding scheme, the basic declarative Latin sentence is also more complex than its English counterpart. Since nothing in Kayne's theory suggests a counterbalancing advantage for Latin, we would expect a learning and/or processing pressure in favor of English [Culicover and Nowak, 2003]. One way to test this claim is to compare he number of SVO or English-type languages with the number of SOV or Latin-type languages. The simpler SVO language type is expected to be more frequent, but the opposite is true. The frequency of SOV languages is statistically equal and numerically greater than that of SVO languages. It is possible that descriptive complexity is in no way relevant for learning and processing, but the burden lies with the supporters of this proposition to provide an appropriate theo-

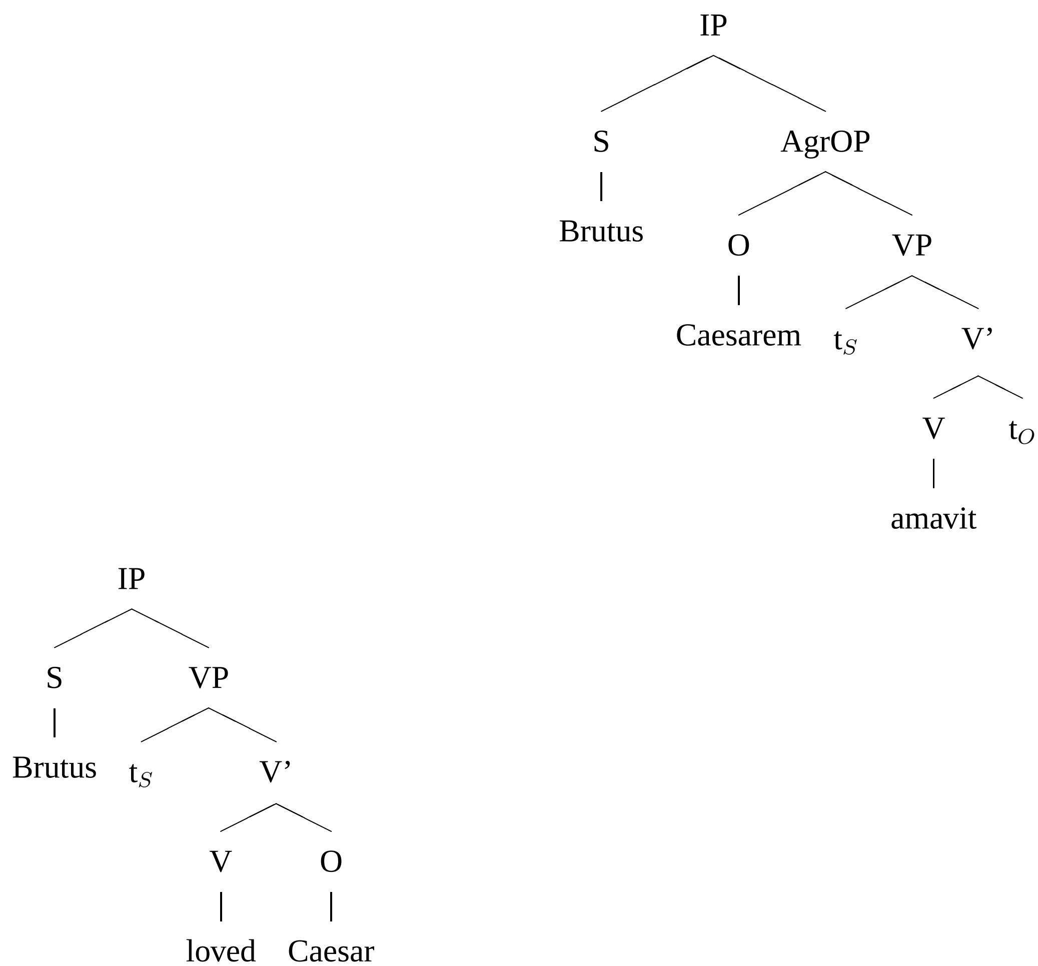



Figure 2.1: SOV versus SVO: the movement approach

retical model. Until then, one must take this failed prediction as a reason to doubt Kayne's hypothesis.

Another fundamental consequence of strict correspondence is that the mapping between structure and meaning cannot be regarded as cross-linguistically invariant. Instead, invariance is limited to some abstract aspects of this mapping. For example, the hierarchical position of the object with respect to the verb in Latin differs from English, despite the fact that they stand in a functor-argument relation in either language. Assuming that hierarchical position is what determines interpretation, strict correspondence at the PF interface implies a flexible view of correspondence at LF.

These undesirable consequences are valid only if we assume that the the structures in Figure 2.1 represent the whole syntactic structure of the sentences. Under the minimalist program, the syntactic structure is equated with the whole derivation, rather than with the tree structure at spell-out. Thus, the complexity of the tree structure at spell-out is irrelevant, since presumably the whole derivation has to be encoded regardless of where spell-out occurs (Figure 2.2). However, this leads to a new set of challenges. Consider

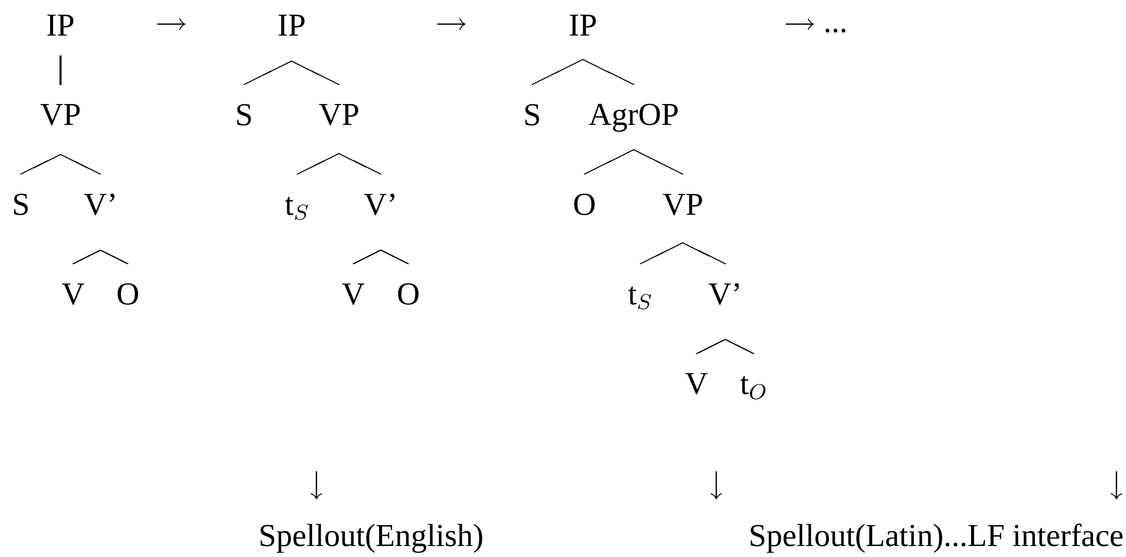



Figure 2.2: A single derivation

the computational problem of uncovering the meaning, given the sentence. First, we must parse the string into its spell-out tree form, then map the spell-out tree form to the full derivation, and only then map the appropriate subset of these trees to the interpretative component. This representation seems unnecessarily cumbersome, especially because it is also highly redundant. Since the derivational steps mapping each tree representation to the next are minimal, the bulk of structural information carries over unchanged. It is desirable to find a way to collapse this redundant multi-step representation of syntactic structure into a single representational component situated between the interpretative and linearization components.

Another questionable consequence of strict correspondence is that all factors which influence word order have to be reified in the structure, albeit in the most economical way.

Consider optional displacement phenomena such as scrambling. It has been shown that scrambling has no effect on either truth-conditional meaning or grammaticality. Thus, in order to ground word order in structure, we must include discourse and pragmatics as structural features. This is problematic if our goal is to distinguish between obligatory conditions for grammaticality (such phenomena as gender and number agreement), and optional variation (such as topicalization).

At the same time, strict correspondence has some obvious advantages for the theory of syntax. First, it allows us to take word order as direct evidence for constituency. To the extent that constituents tend to be continuous most of time, this is a good attitude to take. Second, it makes the strong claim in favor of uniform-branching structures, which seem to be preferred according to typological data. And third, it provides a clear theory of cross-linguistic word order variation by attributing it exclusively to parameter settings for the movement operation.

In contrast, proposals that allow ordering flexibility at the PF level *de facto* imply that word order cannot be considered evidence of constituency. In so far as there is no explicit theory of preferred and dispreferred linearization, tangled structures are as normal as non-tangled ones. Thus, the sentential string per se provides no basis for a prior over potential constituent structure. This leaves opportunities for serious criticism. First, without a theory of preferred and dispreferred mappings flexible approaches cannot explain the prevalence of uniform branching and constituent continuity across and within languages. Second, it raises serious questions regarding the learnability of such grammars. Both of these objections can be met by a theory of linearization which meaningfully constrains structure-string correspondence.

The goal of this dissertation is to address the concerns associated with flexible approaches by supplying a theory of preferred alignment in the spirit of optimality theory [Prince and Smolensky, 1993]. I will argue that cross-linguistic variation is the result of language-specific resolutions of conflicting alignment preferences, while language-internal variation reflects discourse-activated biases that override default preference resolutions. Once the hierarchical component has been fully dissociated from the precedence component, structural descriptions no longer have to be conceptualized as tree sequences. Instead, we can think of them as generalized graph structures. At a minimum, the structures must encode dependency relationships among lexical items. I will therefore assume that the hierarchical component generates multidominant dependency representations, which are linearized in an OT-style optimization. The next chapter defines and discusses in detail the formal system.

# Part III

# Generation and Linearization

# Chapter 3

# Separating structure from order

In this chapter, I will lay out a formalism suitable for providing a unified analysis of displacement phenomena. It is divided in two components: the hierarchical component, which generates the syntactic structure, the linearization component, which maps the structure to an output string (sentence).

## 3.1 The hierarchical component

In designing the hierarchical component, I was led exclusively by the question what is the essential minimum of information that the syntactic structure must encode. I have found that perhaps the single recurrent theme in syntactic analysis is the idea of binary and asymmetric dependency relations. It is through such relations that modification, agreement, and functor-argument pairs are realized.

It is entirely conceivable and even likely that this minimum is a necessary but not sufficient component of syntactic structure. However, there are many advantages to keeping the structural component of the system as simple as possible. For one, this will allow us to explore the extent to which phenomena usually handled at the level of structure can be outsourced to other components. By choosing not to represent constituency directly at the level of structure, I have left open the question whether or not distributional and interpretative facts attributed to constituency might fall out of individual head-dependent relationships, along with properties of discourse-situated interpretation. It appears to me that a decision to include constituency explicitly in the structural representation should be postponed until this possibility is ruled out.

Keeping the structural component simple is advantageous from the point of view of computational modeling. While the structural base advanced here may be incapable of providing all interpretative distinctions which are ultimately needed, they are sufficient for constructing meaning representations within the confines of propositional logic, which goes a long way toward interpretation. At the same time, the simplified format allows us to build simplified generative models as the basis for probabilistic and other processing methods [Eisner, 1996].

Finally, the simple representation allows for a cleaner formal treatment of the objects generated by the structural component. One of the main goals of this work is to include multidominance as a structural device. As a result, the said objects are no longer tree-like. Including multidominance in a structural component which explicitly represents constituency would raise a variety of formal questions which would ultimately have to resolved ad-hoc. For example, can a terminal be dominated by two non-terminals, or should only non-terminals have this privilege? How should the definition of constituency be revised to accommodate the formal modification? In contrast, multidominance in dependency grammars is implementable in a relatively straightforward way.

## 3.2 Overview of dependency grammars

A dependency structure was designed to represent the intuition that "particular occurrences of minimal units are directly related to one another" [Hays, 1964]. The notion *directly related* or *dependent* has a number of interpretations from a variety of theoretical angles. From a distributionalist perspective, *dependent* translates into *conditioned upon*. From a syntax-semantic perspective, it can be interpreted as *is selected by*, or *modifies* [Tesniere, 1959], The earliest formalisations of dependency grammar can be found in the work of [Hays, 1964] and [Gaifman, 1965]. These formalisations were driven by a desire to maintain weak equivalence to context-free grammars. To achieve this goal, they included a set of appropriate theoretical restrictions on dependency structures. The first restriction is that each word (except the ROOT) has exactly one head. This ensures there is only one path from each node to the ROOT and makes the structure a hierarchical tree. The second condition is that the structure must be projective. i.e. that the linear order of nodes in a dependency structure must be such that no dependency links intersect. This condition is also captured by Hudson's Adjacency Principle.

**Hudson's Adjacency Principle** : Every dependent D of a head H must be adjacent to H. D is adjacent to H provided that every word between D and H is a subordinate of H in the dependency tree.

However, it is important to note that these restrictions are not empirically motivated, and face the same issues as constituent-based approaches face with discontinuity. Not surprisingly, attempts to reconcile the theoretical requirements with empirical data lead to the same type of solutions adopted by constituent grammars. This is perhaps most evident in the work of Robinson [Robinson, 1970], who proposed a transformational grammar for dependency structures. Scaled-down versions of this idea give rise to the so-called multi-striatal theories, such as the one proposed by Richard Hudson [Hudson, 1991]. In his approach, a word may have more than one head at deep structure. Surface structure is derived from deep structure by severing one of the competing links in a language-specific fashion. Apparently, similar multistratal approaches have been advocated by the Prague school. For example, Petkevic [Petkevic, 1987] proposed five levels of representation, among which tectogrammatics (the equivalent of deep structure), and surface syntax, in addition to morphemics, phonemics and phonetics. Just like Hudson, he proposes rule-governed mapping from the tectogrammatical level to the surface syntactic level. Both proposals offer very little detail as to the nature of the required rule system. Another type of multistratal representation was developed the early 1970s by Igor Melčuk [Melčuk87], whose Meaning-Text Theory resembles LFG-style representations, consisting of a conceptual and syntactic stratum. Similarly, [Sgall, 1992] incorporated topic-focus annotations into the dependency framework at the tectogrammatical level.

Whether or not they rely on multiple levels of representation, variants of dependency theory diverge on other key parameters, such as whether dependencies are directed (asymmetric) or indirected (symmetric), the degree of lexicalization, and the types of dependency relations they allow. For example, Link Grammar [Sleator and Temperley, 1993] can be understood as a dependency grammar except that the links do not have a direction. While [Hudson, 1991], [Sleator and Temperley, 1993] employ lexical rules, other formalizations are strictly category-based [Hays, 1964]. Word grammar [Hudson, 1991] employs a large set of thematic primitives (agent, patient, benefactor etc.) to serve as a set of dependency relations. Other theorists explicitly seek to avoid this development [Kreps, 1996].

According to Kreps, the set of primitive binary relations contains a single, fundamental licensing relation, which grammatical roles like "subject", "object", and "modifier" are configurational derivatives of. In the traditional dependency structure shown in Figure 3.1, the arguments of the verb "bought" are structurally symmetrical in the dependency tree. Thus, the subject and the object of "buy" are only distinguishable in terms of linear precedence, but not in dominance. This is undesirable because the strength of selectional preference of the verb for its object has been shown to be stronger than that for its subject.

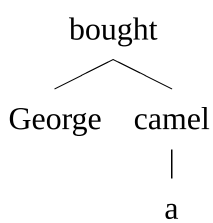



Figure 3.1: Simple dependency structure

A more elegant approach is to borrow the idea of functional head from the X-bar framework and postulate the subject as a dependent of an abstract FIN head, which carries the finite inflection of the verb. Under this approach, the new dependency structure is shown in Figure 3.2

This allows a configurational definition of the grammatical roles object and subject. Object is straightforwardly defined as the dependent of the verb, while subject is the closest qualifying dependent of a head that dominates the verb. In our example, the subject is the dependent of the immediate head of the verb.

The term "qualifying" aims to exclude as potential subjects any dependents that fall on the branch which dominates the verb (e.g. "bought" cannot be the subject of itself in

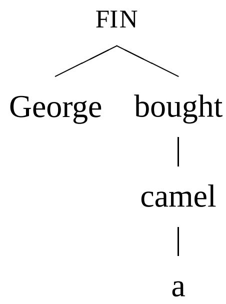



Figure 3.2: Dependency structure according to Kreps

Figure 3.2) [Kreps, 1996]. Another category of potential subjects that needs to be excluded are verbal adjuncts. For example, the structure of (40) should not allow the adverb to take on the role of subject.

(40) George bought a camel today.

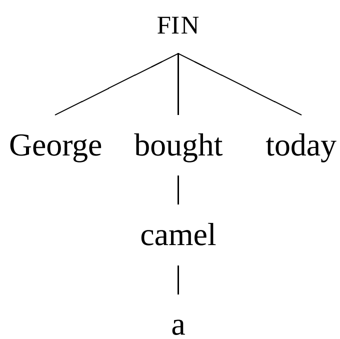



Figure 3.3: Dependency structure

To exclude this possibility, Kreps uses the fact that licensing can be subdivided into two types, depending on the initiator of the relation:

1. Mutual: The head requires the presence of the dependent and the dependent is sanctioned solely by the head.

2. Passive: The head allows the presence of the dependent and the dependent is sanctioned solely by the head.

Mutual contingency relations hold between a head and what X-bar theory terms its "specifier" and "complement", while the passive contingency relation ties the head and an optional adjunct. The relation between the verb and its object in the current example is mutual because the verb requires an object and the object is solely licensed by the verb. However, the relation between the verb and the adverb is not mutual, but passive. While the adverb needs to be licensed by the verb, the verb does not require the presence of the adverb. Thus, the subject is the closest qualifying mutual licensee of a head that dominates the verb.

To sum up the dependency theories so far, one can provide the following general definition of dependency grammar:

**Dependency Grammar** is a tuple $\{R,T,C, \Delta, \Sigma, \mu\}$, where:

R is a (set of) special root symbol(s)

T is a set of terminal symbols (words)

C is a set of category symbols (parts of speech)

$\Delta$ is a set of dependency types (e.g. "dependent", agent, modifier)

$\Sigma$ is a set of ordered valency (subcategorization) frames for C

$\mu$ is a set of ordered modifier motifs for C

A dependency grammar may be defined with one special root symbol, or with a set of symbols which can serve a root symbols. Assuming a special root symbol is convenient for certain applications, such as parsing, because its linking into the structure automatically signals the end of the parse procedure. Thus, computational applications of dependency grammars for natural language often rely on a silent root symbol added to every utterance. Alternatively, R may be a set of symbols which qualify (but not always are) root symbols (e.g. verbs). The terminal symbols may be words (morphemes) or part-of-speech tags, depending on the degree of lexicalization required. If the terminal symbols are words, then a special set of non-terminal category symbols may be used. However, note that these non-terminals are merely needed to specify part-of-speech tags, and never phrase labels. Ultimately, the choice of terminals and non-terminals depends on what the valency frames refer to. Fully lexicalized grammars do not require a set of category symbols.

The extent to which a grammar is lexicalized brings us to another question. How can partial lexicalization be implemented given the definition above. Under partial lexicalization, some valency frames are inherited from category labels, while others are word-specific. To accommodate such a system, the definition may be revised to include a structured set of category labels. This set will include sets of sets, instead of a list. For example, it may include the set of Nouns and Verbs, which in turn include the sets of Proper Noun and Common Noun, and Transitive and Intransitive Verbs etc.(Figure 3.4). Under this type

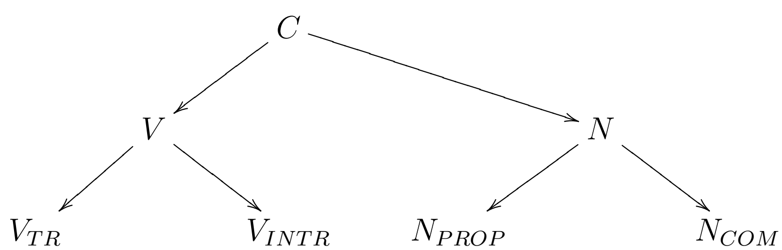



Figure 3.4: Structured set of category labels

of definition, the set of category labels may fully include the set of terminals.

Another question in relation to subcategorization frames is whether or not different types of dependencies need to be distinguished, and if so, what? Tesniere's proposal [Tesniere, 1959] to consider function words as overt markers of different relations has some obvious deficiencies. For one, it is not clear where the boundary between function and substantive words falls. It does not merely follow the closed/open class words distinction. Are pronouns to be regarded as dependency-type markers? In cases when the pronoun referent is part of the clause, this may be a tenable position, but not when the pronouns have external referents.

The difficulties with defining what words constitute overt markers and what are *bona fide nodes* in the dependency structure may be the reason why researchers overwhelmingly prefer to treat dependency types as covert. One approach is to assume a single dependency relation, and attribute the differences of interpretation to configurational properties. For example, *dependent of* is interpreted as modification in the context of adjective-noun, but as functor-argument in the context of verb-noun. The question is how this approach can handle cases like transitive verbs, which have more than one dependent of the same syntactic category. Given the structure in Figure (3.5), which noun is the agent and which one is the patient? One solution is to allow the interpretative component to consult the linear

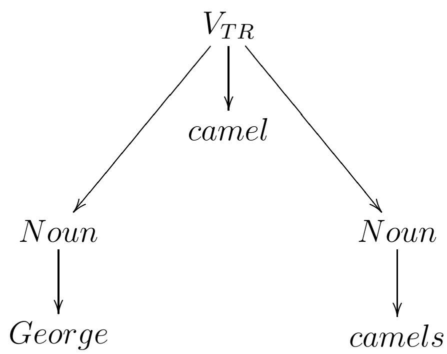



Figure 3.5: Interpreting dependency relations

order among the dependents and their head, morphology and pragmatics. For example, in English, the which precedes the verb and/or the second noun is the subject (41).

(41) George and Laura like camels.

The problem is that such a strategy fails in sentences with atypical word order, such as when topicalization has occurred (42).

(42) Camels, George and Laura like.

One might consider the possibility that the pragmatic component will save the day in some cases. However, in this example, both interpretation are possible (*camels* as subject, or *George and Laura* as subject), while only one is grammatical.

Similarly, morphology is not helpful in this case. Thus, the interpretative component has to entertain a rather complicated procedure for deciding which noun is the subject: If the order is Noun-Verb-Noun, the first noun is the subject, but if the order is Noun-Noun-Verb, the second noun is the subject. While this is not computationally unachievable, it is theoretically bizzare.

These considerations have forced many researchers to define dependency grammars with substantively motivated semantico-syntactic types or seek a stronger configurational formalization. Hudson's word grammar [Hudson, 1984] is dependency grammar with a rich dependency type component. A more complex versions of the configurational alternative has been explored by [Kreps, 1996]. His proposal, involving the introduction of phonologically null heads, is discussed at length below.

Each valency frame specifies the number and type of obligatory dependents of a symbol. Each modifier motif specifies the symbols that may modify each symbol. Unlike valency frames, modifier motifs do not specify number of modifiers. This is because recursion of optional modifiers is in principle unbounded.

Valency frames and modifier motifs may or may not specify the order of dependents with respect to each other and the symbol. If they do not, the order must be specified in a separate component of the grammar ($\Omega$), as in the following modified description.

**Dependency Grammar** is a tuple $\{$R,T,C, $\Delta$, $\Sigma$,$\mu$,$\Omega\}$, where:

R is a (set of) special root symbol(s)

T is a set of terminal symbols (words)

C is a set of category symbols (parts of speech)

$\Delta$ is a set of dependency types (e.g. "dependent", agent, modifier)

$\Sigma$ is a set of valency (subcategorization) frames for C

$\mu$ is a set of modifier motifs for C

$\Omega$ is a set of ordering constraints associated with valency frames and modifier frames

$\Omega$ specifies the ordering of every dependent with respect to the head and other dependents. If more than one ordering is possible, both are specified. There are of course, more economical descriptions of $\Omega$, but this simple version is sufficient for the present purpose.

A dependency grammar generates a (surface) structure according to the following set of requirements:

**Single-root** The special symbol R is independent and occurs exactly once in every structure.

**Single-head** Every symbol except R is dependent on exactly one other symbol.

**No cycles** No symbol is dependent on itself either directly or indirectly.

Given this definition, we can now specify a generative procedure for the grammar. Let us first specify a linearization procedure that obeys the criteria above:

**Linearize T, T a dependency tree**

**Step 1** Initialize S, S empty string.

**Step 2** Pick a set of non-contradicting ordering constraints from $\Omega$ for each node in T.

**Step 3** Insert ROOT in S. Let n=ROOT.

**Step 4** Insert all dependents of n in S and order them according to $\Omega$.

**Step 5** Insert n in S to the left of all dependents $d$ s.t $n > d_i \in \Omega$. Mark n.

**Step 6** Repeat Steps 4-6 for all unmarked nodes in T.

The linearization procedure is second part of the procedure for generating sentences. The first part is structure building Structure-building can be performed either top-down, starting from root, or bottom-up, starting from the dependent-free lexical items. In the top-down procedure, each time a word is added, it generates a set of dependents that comply with its subcategorization frame and the motif.

**Step 0** Initialize dependency structure T, T empty.

**Step 1** Insert ROOT.

**Step 2** Recursively For each unmarked node n $\in$ T:

Pick a word-tag pair p Pick a subcat frame F

Generate a set of dependents according to F and $\mu_p$ Mark n.

Alternatively, we can specify a bottom-up procedure which starts with a numeration, akin to the one proposed by Chomsky in the “Minimalist Program” [Chomsky, 1995].

**Step 0** Pick a set of word-tag pairs P.

**Step 1** For each p $\in$ P, pick a subcategorization frame $F_p$.

**Step 2** Generate all possible sets of links $L_1 ... L_k$ among members of P.

**Step 3** Pick an unmarked set $L_i$. If:

1. $L_i$ conforms to the dependency structure requirements in (3.3,

2. All frames F are satisfied

Return $L_i$; Else mark $L_i$ and go to Step 1.

Obviously, the second procedure will finish without returning a structure, if the numeration does not allow it to build one.

Both procedures are geared towards generating a tree structure, linearized in compliance with projectivity and the specified ordering. Such a structure is incapable of expressing a variety of empirically occurring dependency relations. For example, the fact that the subject agrees with the auxiliary in complex tense clauses suggests that it is dependent on it, just like the adjective is dependent on the noun it agrees with. At the same time, the subject must also agree with the verb regardless of tense, to satisfy its valency requirements. Thus, it seems that the common subject-auxiliary-verb construction already presents a challenge for dependency theory, because it requires the subject to form multiple dependencies.

This is where the multistratal approaches discussed above come into play. In Hudson's model, deep structure contains both subject dependencies. The surface structure is derived by severing one of the links through an unspecified *competition*. Kreps takes a different approach to this particular case by generalizing his solution of subject-object asymmetry. Recall that according to his definition, the <u>subject</u> is the closest qualifying dependent of a head that dominates the verb. Assuming that the subject is the dependent of the highest auxiliary, all lower components of tense are configurationally bound to it. This idea is easily extendable to a variety of raising constructions. Thus, the structure of the sentence in (43) is given in Figure 3.6.

(43) George seems to need a camel.

In this case, *George* is determined to be the subject of both seems and *need* solely through the configurational definition, without reference to any traces or empty categories. The strategy is especially truthful with respect to object raising constructions, as it straightforwardly predicts that the object of the controlling verb should be the subject of the infinitival.

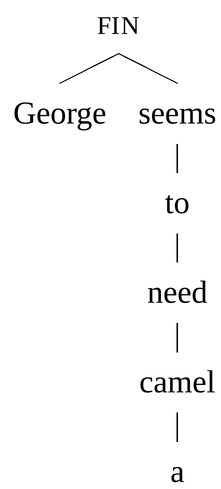



Figure 3.6: Dependency structure

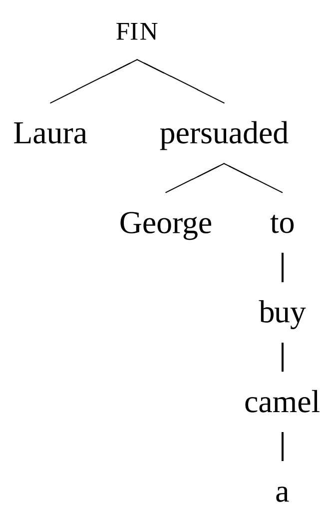



Figure 3.7: Dependency structure

(44) Laura persuaded George to buy a camel.

An apparent counterexample to the analysis [Kreps, 1996] is put forth by *promise*-constructions (45, Figure 3.8)). At first blush, *promise*-constructions seem exactly parallel to the *persuade* construction above, yet the subject of the infinitival is not the object of the control verb (George), but its own subject (Laura).

(45) Laura promised George to buy a camel.

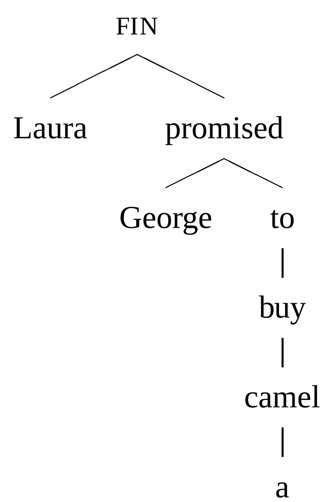



Figure 3.8: Dependency structure

One solution to this counterexample is to argue that the optionality of the argument of *promise* plays a crucial role. Compare (46) versus (47) and (48) versus (49). The ungrammaticality of (47, 49) points to a much closer relationship of the argument to *persuade* than to *promise*, which leads us to suggest that the argument of *promise* is a passive, rather than mutual licensee, and therefore does not qualify as a subject of the infinitival.

(46) Laura promised George to buy a camel.

(47) *Laura persuaded to buy a camel.

(48) Laura promised to George to buy a camel.

(49) *Laura persuaded to George to buy a camel.

Whether a relation is mutual or passive depends on the valency associated with the head. If the head actively seeks to license its dependent, the relation is mutual; otherwise it is passive. Note that the quality of the relation is not defined as part of the grammar, rather its an epiphenomenal property of the construction.

Both the multistratal and the configurational approach are partially successful at dealing with some aspects of multidominance. However, they also have significant shortcomings. The configurational approach leaves open the issue of why some dependencies are direct, while others are configurational. This is particularly troubling in cases when different types of dependencies have the same grammatical consequences. For example, why does the direct adjective-noun dependency result in morphological agreement, just as the configurational subject-verb dependency? The multistratal approach circumvenes this problem by guaranteeing that direct dependency holds between any morphologically connected pair at some level of the derivation. However, the different derivational levels are merely designed to reconcile multidominance and long-distance dependencies with the linearization requirements of traditional dependency theories. By separating the linearization component from the structural component of the formalism, I have cleared the road to merging the different levels into a single representation. The details of this representation are provided in the next section.

## 3.3 Dependency grammar with multidominance

Allowing multidominance in dependency theory has little effect in the definition of the grammar as a set of sets. Instead, the changes are confined to the generative and recognition procedures associated with the grammar. Thus, the foundation for the formalism developed here is a particular variant of the dependency grammar definition provided in the previous section. I begin by motivating the particular choices I have made in choosing the definition.

Given that the ultimate goal of this work is to create a formalism which handles word order entirely separately from structure, neither valency nor modifier frames specify order. In fact, ordering constraints do not constitute a part of my definition of grammar at all. The job of the hierarchical component is to generate an acyclic graph of directed dependencies, hence order is irrelevant.

To make subsequent formalization of parsing and generation easier, I will adhere to the computational tradition of reserving a special silent ROOT symbol. I will also view the set of category labels as structured in the manner described in the previous section. This allows valency and modifier frames to refer to symbols at any level of the category hierarchy. For example, the subcategorization frame of a locutive verb (e.g. *say*) includes simply some verb (e.g. transitive *like*, intransitive *sleep*, or locutive *note*) as its dependent. At the same time, different types of verbs specify their own subcategorization frames: two nouns for transitive, a single noun for intransitive.

What types of dependencies should be specified? The semantico-syntactic approach is not attractive from a theoretical point of view, because it fails to provide a mechanism for generalization across subcategorization frames. For example, if the frame of a transitive verb is specified as $V_{TR}[subject : N, object : N]$ and the frame of *OF*-adjectives $Adj_{OF}[complement : OF]$. However, formulating such similarities may be necessary, given the parallel interpretation of constructions involving these frames [Chomsky, 1995]. Furthermore, the semantico-syntactic approach is data-driven, without a priori restrictions on either the number or the type of relations. It is more parsimonious to postulate a single primitive relation which generates the syntactic structure, and treat all grammatical roles as epiphenomenal. The question is whether felicitous configurational definitions can be crafted for all syntactic asymmetries. In the previous section, I laid out Kreps's solution to the subject-object asymmetry. However, the subject-object asymmetry is just an instant of the specifier-complement asymmetry. The general question is whether to represent specifiers and complements configurationally, or simply to reify them. The latter strategy is a middle ground between Hudson and Kreps. While it does rely on more than one fundamental dependency type, it does so only to represent syntactic asymmetry, leaving semantic issues outside the scope of the grammar. A parallel question is how to distinguish between direct and indirect objects in the double object construction. One way is to introduce a null dependent of the double object verb which subcategorizes for the direct object complement. A different tack is to increase the number of fundamental types from specifier and complement to specifier, direct complement and indirect complement. While I accept that the ultimate goal of a minimalist theory of syntax is to reduce all dependency types to configurational properties, whether or not this is feasible remains to be seen empirically. Thus, I will initially assume all three types of dependency relations. Thus, subcategorization frames have the form Head [Spec:Dependent, D-Comp:Dependent, I-Comp:Dependent], with at most one dependent per type. Head and Dependent position can be filled up by category symbols at any level of the structured category set and the subcategorization frame can be satisfied by any subordinate of that symbol. Some examples are given below:

**Subcategorization frames** :

$V_{LOCUTIVE}[V]$

$V_{TRANSITIVE}[Spec : N, D - Comp : N]$

$V_{INTRANSITIVE}[Spec : N]$

$V_{DOUBLE-TRANSITIVE}[Spec : N, D - Comp : N, I - Comp : N]$

Unlike subcategorization frames, modifier motifs do not require a multitude of dependency types. They are of the form *Head*[*Modifier*$\bigvee$ *Modifier*$\bigvee$ *Modifier*...] where $\bigvee$ has the usual meaning of inclusive OR. In other words, a modifier motif lists all modifiers that can be used with a given head. Here are some examples:

**Subcategorization frames** :

$V[Adv \bigvee Prep]$

$N_{COMMON}[Adj \bigvee Prep]$

The dependency relation between a modifier and its head is always the modifier relation. Thus, the definition of the dependency grammar of the hierarchical component is:

**Dependency Grammar** is a tuple $\{$R,T,C, $\Delta$, $\Sigma$,$\mu\}$, where:

R is a special root symbol

T is a set of words

C is a structured set of parts of speech

$\Delta$ is a set of dependency types (Spec, D-Comp,I-Comp, Mod)

$\Sigma$ is a set of valency (subcategorization) frames for C

$\mu$ is a set of modifier motifs

Next, let us turn to the necessary modifications of the generation procedures for dependency grammars of this type. Recall the top-down procedure given in (3.2). To allow multiheaded structures, we need to introduce a look-back step.

**Step 0** Initialize dependency structure T, T empty.

**Step 1** Insert ROOT.

**Step 2** Recursively For each unmarked node n $\in$ T:

Pick a word-tag pair p

Pick a subcat frame F

If a node d $\in$ T is in a *qualifying* position and d is required by F

link d to n. Delete d from

F. Generate a set of dependents according to F and $\mu_p$ Mark n.

The bottom-up procedure requires a simpler modification. Instead of checking the current structure against the requirements in (3.2), it will check against a new set of requirements, which specifies the type of generalized acyclic graphs allowed, including whatever definition of *qualifying* is appropriate.

**Single-root** The special symbol R is independent and occurs exactly once in every structure.

**Obligatory head** Every symbol except R is dependent on at least one other symbol.

**Constraints on multidominance** If a symbol is dependent on more than one other symbol, the relationship among the head symbols must satisfy constraints on multidominance.

Recall that the original reference to "qualifying" belongs to Kreps's configurational analysis of subject/object roles and is influenced by two factors:the position of the candidate dependent and the type of dependency it already holds. In particular, in order to qualify as a remote subject dependent of a non-finite verb, a node must a)not dominate the verb and b)be a mutual dependent of a dominating finite verb. Obviously, this restriction was particular to the type of double-headed construction involving subjects of non-finite verbs. Different types of restrictions must be imposed on other types of movement constructions. We will examine the structural analysis of wh-movement in Chapter 6.

## 3.4 The linearization component

The linearization component must be designed to take into account different types of influences on word order- structural, discourse-based and phonological. The primary determinants of word order are the dependency relations among lexical items. Dependency relations influence word order in two ways. First, they overwhelmingly translate into adjacency relations. This observation has lead most grammarians to assume that phrases are continuous and natural languages are context-free. One formulation of this assumption is known as the Hudson Adjacency Principle.

**Hudson's Adjacency Principle** : Every dependent D of a head H must be adjacent to H. D is adjacent to H provided that every word between D and H is a subordinate of H in the dependency tree.

In addition, specific types of dependencies are linked to precedence relations. For example, English prefers to place subjects before verbs, and objects after them.

Information structure also plays a role in determining precedence relations. Topicalized constituents typically align with the left edge, and focused constituents with the right end of utterances. For example, Bulgarian direct objects normally precede indirect objects, but follow them when focused.

(50) Georgi kupi kamila na Lora.
George bought a camel for Laura.

(51) Georgi kupi na Lora kamila.
George bought Laura a camel.

Similarly, even though direct objects normally follow the verb in both Bulgarian and English, topicalization reverses the order (53):

(52) Kamilata, Georgi kupi na Lora.

(53) The camel, George bought for Laura.

The relative influence of structural, discourse and phonological factors differs from language to language. For example, while English prefers objects before verbs, Latin prefers the opposite order (cf. 54).

(54) Brutus Ceasarem interfecit.

Similarly, discourse factors influence word order only in so far as the particular language allows. While some languages, like Bulgarian, are considered *free word order* because they often defer order decisions to information-structure, others are *rigid word order*, allowing few if any deviations from the canonical order.

Consequently, a linearization component must be capable of handling cross-linguistic differences. Traditionally, this is done through language-specific rules. Thus, the set of universal linearization rules is considered a superset of the rules of any individual language. In other words, languages can pick and choose specific subsets of all available rules. The problem with this setup is that the universality claim becomes rather weak, since the universal character of any given rule cannot be disputed even if many languages do not comply with it. Optimality Theory offers a different way of looking at cross-linguistic variation. Instead of claiming that only a subset of universal rules is active in any given language, OT holds that all universal rules are active in all languages. However, the inherent conflicts among rules are resolved differently from language to language, resulting in variation.

### 3.4.1 Optimality Theory

Optimality Theory (OT) claims that linguistic expressions are restricted by a set of universal, mutually inconsistent and violable constraints [Prince and Smolensky, 1993]. Conflicts result in the satisfaction of higher ranked constraints at the expense of their lower ranked adversaries. The variations among languages are attributed to differences in the constraint rankings. In OT, a grammatical linguistic expression is a winner of an optimization. Given an underlying representation (UR), a generator function (Gen) produces a (potentially infinite) set of surface realizations (SRs), and a process of optimization picks the SRs that minimally violate the constraints according to a language-particular ranking.

OT is a general framework that can give rise to a variety of specific formal instantiations depending on the types of representations and constraints invoked.

Let us begin with a formal definition of an OT system, adapted from [Frank and Satta, 1998].

**An optimality system** is a 4-tuple $OS = \{\Sigma, \Gamma, Gen, C\}$ where $\Sigma$ and $\Gamma$ are the finite input and output alphabets, Gen is a relation over $\Sigma^* \times \Gamma^*$, and $C$ is a finite set of total functions from $\Sigma^* \times \Gamma^*$ to $N$.

As seen in this definition, Gen maps a UR to a set of SRs, while a constraint is a function from a candidate UR-SR pair to a natural number, which we take to represent the degree of violation incurred by that candidate on that constraint. An OS gives rise to a set of optimality grammars (OG), defined in (3.4.1):

**An optimality grammar** OG is an OS together with a total ordering $R$ on $C$, called a ranking.

For the purposes of linearization, the underlying representations are the structures generated from the hierarchical component, and the surface representation are strings. Since the constraint component must be able to act on discourse features, the latter must also be a part of the underlying representation. Without explicitly formalizing this idea, I will simply assume that the underlying representations are structures stamped with discourse features by the hierarchical component.

If Gen is as unrestricted as possible, its output would consist of all possible permutations of nodes in the underlying representation. There are many ways of restricting Gen. For example, it is possible to limit the candidate set to all top-down left-to-right traversals. However, such restrictions considerably limit the potential of the constraint component to introduce variation. That is why, in the absence of a solid pre-theoretical reason, the linearization component will benefit from a maximally unrestricted Gen.

## 3.5 The parsing component

Before concluding the technical description of the grammatical system proposed here, it is worth saying a few words about the counterpart of the linearization component, which

I assume to be responsible for mapping a string into a dependency structure. While this dissertation does not deal with this component in detail, no description of the system would be complete without a mention of its important role, and a rough sketch of its inner workings. I will call it the parsing component.

As we saw earlier, the function of the linearization optimization is to choose the best possible string for any structure generated by the Dependency Grammar. I assume that the parsing component also relies on an optimization to perform the opposite function, i.e. to choose the best possible structure given a string. It is particularly important that the parsing component be able to restore the dependency structure of any output of the linearization component. Thus, if the dependency structure in Figure 3.1 is linearized as the string in (55), the parsing component must be able to map the string in (55) into the structure of Figure 3.1.

(55) George bought a camel.

To accomplish this, the parsing component must rely on constraints referring to several types of information. First, it must "know" the ordering preferences of the linearization component. Second, it must "know" the dependency grammar, so as to be able to distinguished well-formed from ill-formed structures. This includes knowledge of the subcategorization frames of atomic elements, as well as knowledge of general constraints on structure. Finally, it must have access to intonational and discourse features.

While keeping in mind that the details of the parsing optimization are yet to be worked out, let us go through a an example of how it works. Suppose the input is the string in (55). The relevant parsing constraints include the ordering constraints Spc>Cmp, Hd>Cmp, HdCmp and HdEdge, as well as structural constraints on subcategorization for each lexical item: $Subcat(V_{tr})$, $Subcat(N_{PROP})$, $Subcat(N_{CNT})$ and $Subcat(D)$. The set of candidate structures includes the correct structure (CORR) in Figure 3.1, but also—under "richness of the base"— all other possible dependency graphs among the four lexical nodes. Some examples are given in Figures 3.9(CDS1) and 3.10(CDS2). In the parsing optimization, the wrong structures (CDS1 and CDS2) loose to the right structure (CORR). CDS1 involves a violation of two structural constraints, while CDS2 involves violations of two ordering constraints.

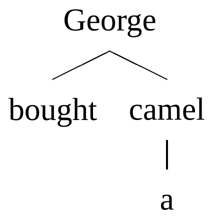



Figure 3.9: Competitor dependency structure (CDS1)

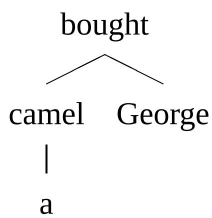



Figure 3.10: Competitor dependency structure (CDS2)

Note that the correct structure violates at least one constraint—$HdEdge$. However, since $HdEdge$ is low-ranked in English, the correct structure wins. In principle, we allow subcategorization constraints to be freely rerankable with respect to ordering constraints. It is possible that in another language, where $HdEdge$ is ranked higher than the subcategorization constraints, a nonce structure like CDS1 will win. Since CDS1 cannot be read by the interpretative component, such outcome will point to the ungrammaticality of the input string.

Along with the linearization optimization, the parsing optimization provides us with a way to explain ungrammaticality, especially when the latter is interpretative in nature. For example, it is likely that pronoun coreference is established by the parsing component, and the unavailability of certain types of coreference are attributable to the presence of more felicitous competing structures. This picture fits well with existing bidirectional optimization approaches to Optimality Theory, which claim that ungrammaticality arises whenever the reverse optimization fails to map the output of the forward optimization back into the same input. Further research on the precise mechanisms of the parsing optimization is undoubtedly necessary, even as it remains beyond the scope of this work.

| Input | Struct | $SbcV_{tr}$ | $SbcN_{PR}$ | $SbcN_{CNT}$ | Spc> Cmp | Hd> Cmp | Hd Edge |
|---|---|---|---|---|---|---|---|
| George bought a camel | CDS1 | * | * | | | | |
| | CDS2 | | | | * | * | * |
| | CORR | | | | | | * |

Table 3.1: Apparent interaction of word order and discourse-motivated linear precedence in German

# Part IV

# Cross-linguistic word order variation

# Chapter 4

# Basic Word Order

The relative order of the main sentential components-Subject, Object, and Tense [1] varies greatly from language to language. One source of variation is information structure. Some languages allow ordering to reflect fine grained topic-focus distinctions, while others impose rigid, discourse-independent rules. Another source of variation is the morphological marking on the verb, and the morpho-syntactic category of its arguments. While these facts are intriguing in their own right, the most profound mystery is why “stylistically neutral, independent, indicative clauses with full noun phrase participants” [Siewierska, 1988] succumb to different ordering requirements across languages.

Any universalist theory of language structure to explain why multiple surface forms are possible for the exact same (deep) structure. It is perhaps an even more pressing challenge to figure out why some orderings are unattested, and presumably therefore, impossible (or highly unlikely). I will demostrate that this can be accompished with a high degree of success if structure and linearization are relegated to separate components of the grammar, and word order is viewed as the product of universal but conflicting linearization constraints ranked in a language-specific fashion.

There are four types of declarative sentences that I will be concerned with here. Typological studies and theories of word order have largely focused on the basic subject, object,

[1] I am using the term Tense here to denote all elements of a verbal complex. In the case of simple tenses, Tense is equivalent to the main verb, while in composite tenses, Tense consists of the auxiliary as well as the main verb.

simple verb sentence (e.g. [Tomlin, 1986], [Costa, 1997]). While it is important to address this type of sentences, a theory of word order variation must be able to reach beyond the scope of this preliminary data to explain more complex sentence type, and the covariation among them. Therefore, I will examine the order of primary sentential constituents in the presence and absence of auxiliary, as well as in the context of main and subordinate clauses.

## 4.1 What is basic word order?

One issue is how to define the concept of basic word order in a way that would allow it to apply to most, if not all, languages of the world. According to Siewierska, basic word order is found in "stylistically neutral, independent, indicative clauses with full noun phrase participants, where the subject is definite, agentive and human, the object is a definite semantic patient, and the verb represents an action" [Siewierska, 1988]. Unfortunately, virtually all requirements run into problems on a language-particular basis, due to the availability of bound pronouns, clitics, noun incorporation and pro-drop in spontaneous speech. It is also clear from Siewierska's definition that word order can be affected by some rather arbitrary features of the lexical items involved, which were empirically identified after years of previous typological research (definiteness, animacy, agency). It is entirely conceivable that order in some languages may depend on other, hitherto unknown features of the verb or arguments. As a result, linguists with non-native proficiency and no *a priori* information about the variation, may come to radically different conclusions about the basic word order depending on what particular words they use in their elicitation examples.

The morphological markedness criterion is based on the observation that displaced constituents signal their position via some overt marker. While it may be considered an positive indicator of non-canonical order, it is not a universal one, since its importance depends on the morphological richness of the language involved. Furthermore, it is not entirely clear what elements should be considered displacement markers.

Statistical sriteria are formulated in the hope of avoiding confounding variables. Basic word order is often assumed to be simply the most frequent order, determined through statistical comparisons of corpus counts. However, word order distribution often varies widely from corpus to corpus, depending on the set of topics and the associated style. Creating a representative corpus of a language is a highly non-trivial task even in cases of rich written culture, let alone languages with little of no written text. The restricted context criterion runs into similar difficulties. It requires conditional probability estimates of word orders given the surrounding context. For example, while the English verb may always be followed by an adverb, it may be preceded by an adverb only in special contexts (56 versus 57)

(56) ?John is slowly walking.

(57) John is walking slowly.

Obviously, the restricted context criterion calls for a definition of context, which may be of arbitrary granularity. For example, the context in (56) may be defined as $Aux - V - .$ or as $is - walking-$. In the latter case, the generalization only pertains to a certain subset of adverbs that modify *walk*, while in the former, it hardly holds, due to counterexamples like (58, 59).

(58) ?John is going definitely.

(59) John is definitely going.

The simplicity criterion [Dryer, 2005a] may be considered the mirror image of the restricted context criterion. It assumes that non-canonical order is due to the complexity of constituents. Therefore, canonical word order is the order in which simple constituents appear. For example, ordinary English adjectives precede the noun, but complex adjectives follow it (60).

(60) the woman taller than John

This criterion may have some unintended implications which are at odds with Siewierska's definition. In some languages, word order varies depending on whether the object is a full NP or pronominal. For example, the Bulgarian sentence with full object noun phrase is SVO, while the corresponding sentence with an object clitic is SOV (61, 62).

(61) Georgi kupi kamilata.
George bought camel-*def*.
George bought the camel.

(62) Georgi ja kupi.
George her bought.
George bought it.

Since the pronominal clitic appears internally simpler than a full phrase, the simplicity criterion would entail that the word order of clauses with pronominal clitics is basic, while that with full arguments is derived. This is hardly a non-controversial conclusion.

For the purposes of this work, I consider basic word order to be the order of primary constituents in "stylistically neutral, independent, indicative clauses with full noun phrase participants", as Siewierska proposes. My goal is to characterize the order of the primary constituents: subject, object, and the components of tense (auxiliary and verb) in discourse-neutral declaratives. I have obtained most cross-linguistic data from the work of Matthew Dryer [Dryer, 1997], [Dryer, 2005a], [Dryer, 2005b], [Dryer, 2005a], who has dealth with difficult methodological issues in a disciplined manner. Whenever I have turned to other sources, I have done my best to ensure that the terminology and methodology are consistent with, or comparable to, that of Dryer. However, in the absence of standardization, the risk of experimental error is inherent in any typological study of word order.

## 4.2 Dependency structure of declaratives

The structure of the declarative sentence must reflect two kinds of dependencies among the main constituents. Subcategorization relations between a) the auxiliary and the main verb; and b) the verb and its arguments, constitute the first type of dependency. As discussed in Chapter 3, the asymmetric status of subject and object is expressed through imposing the specifier/complement distinction directly onto the dependency relations (Figure 4.1).

The other type of dependency is morphological. It links the subject to the auxiliary verb, and leads to a case of multi-headedness (Figure 4.2).

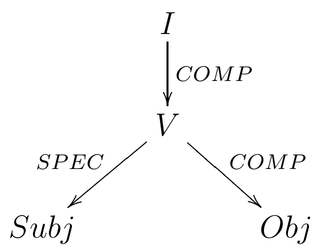



Figure 4.1: Subcategorization dependencies of the main declarative constituents

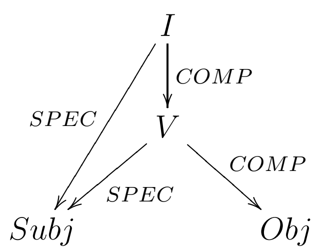



Figure 4.2: Subcategorization and morphological dependencies of the main declarative constituents

The resulting dependency structure can be considered fairly incontroversial for complex tense declaratives, especially in view of the fact that it represents a direct mapping from the well-known constituent structure (cf Figure 4.3). The subject is the specifier of

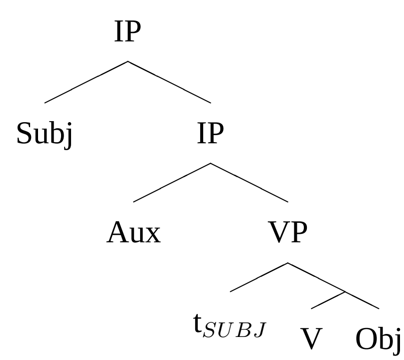



Figure 4.3: Constituent structure corresponding to Figure4.2

an inflectional phrase (IP), headed by the auxiliary with the verb phrase (VP) in complement position. Within the VP, the object is in the complement position, while the specifier position is occupied by the trace of the subject. The trace ensures that the subject remains in a subcategorization relation of the verb, even though it has also entered a secondary (morphosyntactic) relation with the auxiliary (i.e. with the inflectional head).

To obtain the structure of declaratives for simple tense forms, constituency analyses simply assume that the inflectional head is either phonologically null, or unpronounceable consisting solely of the inflectional ending (Figure 4.4). Head movement (or affix lowering)

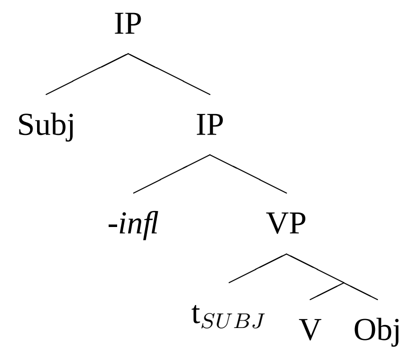



Figure 4.4: Constituent structure of simple tense clauses

unites the verb with the impoverished inflectional head to produce a well-formed structure (Figure 4.5). Essentially, what either of these operations does is unite the two heads into a

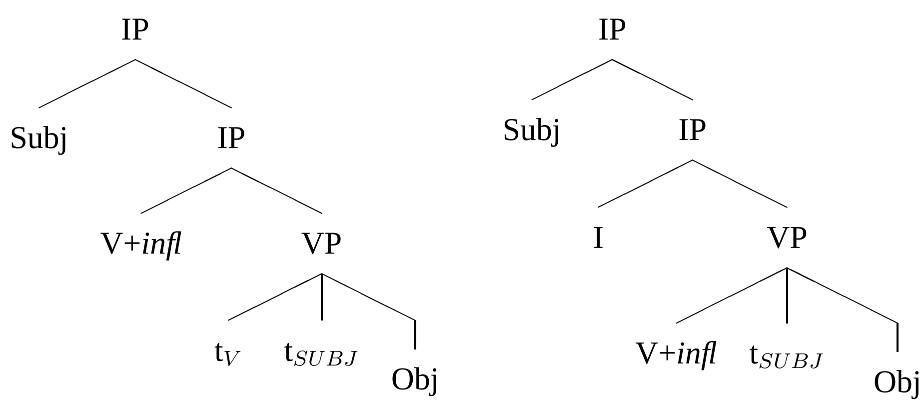



Figure 4.5: Constituent structure of simple tense clauses - head movement and affix lowering

single (compex) head.

Note that the nature of head movement is different from the phrasal movement operation. While phrasal movement simply alters, or adds to, the dominance and precedence relations of the object, head movement actually produces a new composite object. The surface similarity of the two in constituency analyses is somewhat deceptive. In fact, it disappears when conversion from constituency to dependency structure takes place. While phrasal movement corresponds to multi-headedness, there is no reason to impose the same on head movement. Rather, the most natural way of representing head movement is simply by directly merging the heads invloved into a complex head[2] (Figure 4.6). I suggest the

$$\begin{tikzcd} & V+i \arrow[dl, "SPEC"'] \arrow[dr, "COMP"] & \\ Subj & & Obj \end{tikzcd}$$

Figure 4.6: Simple-tense declarative

complex head is the result of a lexical operation that merges heads with their complements. Such operation has been independently proposed to account for semantic incorporation effects (see [Hale and Keyser, 1993]).

An important consequence of this purley representational shift is that the Minimal Lnk Condition [Chomsky, 1995][3] is no longer necessary to regulate head movement. Since I have reduced the latter to a local lexical operation, long head movement structures [4] as the one in Figure 4.7 cannot be generated.

$$\begin{tikzcd} & I \arrow[d, "COMP"] & \\ & V+c \arrow[dl, "SPEC"'] \arrow[dr, "COMP"] & \\ Subj & & Obj \end{tikzcd}$$

Figure 4.7: Structure violating LHM and implicitely disallowed by the current formalism

[2]Henceforth, small letters will be used to represent incorporated heads.

[3]MLC: A can raise to target K only if there is no operation (satisfying last resort) Move B targeting K where B is closer to K

[4]LHM is a peculiar type of head movement, allowing a lower head to move to a higher head by skipping over an intermediately positioned head

$$
\begin{tikzcd}
 & A : TypeT \arrow[dl, "SPEC"'] \arrow[dr, "COMP"] & \\
B : TypeQ \arrow[d, "COMP"] & & C : TypeQ \arrow[d, "COMP"] \\
D & & E
\end{tikzcd}
$$

Figure 4.8: Abstract dependency structure

## 4.3 Constraints on linearization of discourse-neutral declaratives

Dependency representations naturally lead to constraints that refer to precedence and adjacency relations defined on dependents and their heads. Both head and dependents may be addressed at a general or specific level, according to the particular category or type they belong to. For example, each of the constraints below can be used in the linearization of Figure 4.8:

1. Heads must precede their complements.

2. Heads of type T must precede their specifiers.

3. Heads must be adjacent to their complements.

The first constraint favors linearizations where $A > C$, $B > D$ and $C > E$, the second favors $(A > B)$. Some word orders compliant with these constraints are $ABCDE$, $ABCED$, and $ACEBD$. Obviously, the first constraint is more general than the second and therefore theoretically more desirable. Constraints of the second type should only be used when general constraints are unable to account for the typology.

The effect of the third constraint is slightly less strainghtforward. It is possible to interpret *adjacent* in two ways. The literal interpretation favors $ACEBD$, but not $AECBD$, where the daughter of the complement interferes on the path to its head. However, it is also possible to interpret *adjacent* as a local requirement in the depedency structure. Assuming that linearization is carried out top-down, the structure which constraints see at any given point excludes subordinates. Thus, the adjacency constraint would favor $ACEBD$ and $AECBD$ equally. Note however, that it would discriminate against $EACBD$ because $A$ - a superordinate, appears between $C$ and its complement.

Let us now turn our attention from the abstract structure in Figure 4.8 to the declarative structures discussed in the previous section. In the past, cross-linguistic evidence has contributed to the formulation of a number of language-specific and more general linearization principles from which soft linearization constraints can be derived.

One of the earliest observations in phrase-structure grammar was that certain elements (such as adverbs) can interfere between the subject and the verb but not between the subject and the object. This was taken as an indication that the verb and the object were a subconstituent of the sentence that excludes the subject. Similar observations prompted word order typologist W. P. Lehmann to formulate the Fundamental Principle of Placement, which states that the verb and the object noun phrase are primary concommitants of each other in the sentence and the components of the primary syntactic structure [Lehmann, 1973]. The idea of primary concommitants also underlies Tomlin's Verb-Object Bonding principle:

**Verb-Object Bonding Principle** : The object of a transitive verb is more tightly bonded to the verb than the subject.

As we discussed in Chapter 1, VSO languages are problematic for the assumption that verb and object form a subconstituent, unless a transformational view of structure is adopted. But what if this intuition is not taken as evidence of an inviolable princliple, but is instead treated as a soft adjacency preference between heads and complements? In that case, the existence of VSO is evidence simply that the soft preference can be overriden by a more important preference on a language-particular basis. Thus, I attribute the affinity between verb and its complement to an adjacency constraint that requires heads to be adjacent to their complements, interposition of subordinants notwithstanding (*HdCmp*)

(63) **HdCmp**: For all $x$ such that $Head > x > Compl$, $x$ is either null or a proper subordinate of Compl.
= Heads may not be separated from their complements, except by proper complement subordinates.

| Order | Number | Percent | Lower CI 95% | Higher CI 95% |
|---|---|---|---|---|
| SOV | 497 | 47.0 | 44.1 | 50.1 |
| SVO | 435 | 41.2 | 38.2 | 44.1 |
| VSO | 85 | 8.0 | 6.5 | 9.8 |
| VOS | 26 | 2.5 | 1.6 | 3.6 |
| OVS | 9 | 0.9 | 0.4 | 1.6 |
| OSV | 4 | 0.4 | 0.1 | 1.0 |

Table 4.1: Frequency of base word orders [Dryer, 2005a]

> X is a proper subordinate of Y iff X is a subordinate of Y and X does not depend on sisters or superordinates of Y.

If HdCmp is truly overruled by some other constraints in VSO languages, what might they be? There seem to be two factors involved - a requirement on the verb to be in initial position, and on the subject to precede the object. For now, let us leave aside the question of verb-initial position, and review the precedence relation between subject and oblect.

The evidence for a constraint on subject-object order is ample. Of the four word orders known to exist with reasonable frequency and certainty, only one allows objects to precede subjects. Combinatorially, there are six possible word orders of subject, object and main verb: SOV, SVO, VSO, VOS, OVS and OSV. It is often assumed that object-initial languages do not exist [Baker, 2001], despite the occasional counter-evidence presented in the literature [Derbyshire and Pullum, 1979], [Pullum, 1982]. What is indisputable is that object-before-subject languages are vanishingly rare. In a recent typological study of $1056$ languages with a clear dominant order, they constituted less that $4\%$ of the sample [Dryer, 2005a]. The results of the survey are presented in Table 4.3. It lists the number of languages found to have each basic order, what percent of the sample it constitutes, as well as the lower and higher boundary of the confindence interval, given the size of the sample. Previous studies reflect virtually the same distribution [Tomlin, 1986]. The difference between OVS and OSV languages is not significant. All other differences are significant.

It is evident that OVS and OSV languages are extremely rare. Further examination shows these languages are probematic for a number of reasons. Most are spoken by a very small group people; there is no unanimity on which order is basic; and a significant fraction are ergative or there is inconclusive evidence for ergativity. For example, the Carib language Hixkaryana is the first known and most cited OVS language (originally described by Desmond Derbyshire [Derbyshire, 1979]). It is spoken by just over 500 people on the Nhamund river, a tributary of the Amazon River in Brazil. The classification of Hixkaryana is not unproblematic, given that indirect objects follow the subject, and word order in non-finite embedded clauses is SOV. It is possible that the OVS order reflects a topicalization requirement on definiteness, rather than a genuine object-subject ordering preference. Three other OVS languages come from the same region: Bacairi [Wheatley, 1973], Apalai [Gordon, 2005], and Macushi ([Abbott, 1991]). Bacairi is spoken by 570 people, Apalai by 450 people. The estimates for Macushi $(11000 - 14000)$ vary widely due to rampant bilingualism [Gordon, 2005]. Furthermore, according to Dryer, Macushi as well as Apalai seems to position the subject freely either before or after the OV complex. It is quite possible that these languages are actually SOV. A fifth language - Selknam, was spoken in Terra del Fuego and apparently went extinct in $2003$. The remaining three of the nine OVS languages in Dryer's study are actually Absolutive-Verb-Ergative languages: Pri (Nilotic; Sudan; Andersen 1988), Mangarrayi (northern Australia; [Merlan, 1982]), and Ungarinjin (Wororan; northwestern Australia; [Rumsey, 1982]). The previously mentioned Bacairi language is also argued to be ergative [de Souza, 1994]. The situation is similar with the OSV languages. The most widely used of these languages is Warao. However, there is no general agreement on basic order in Warao. Some claim SOV word order [Osborn, 1966] others argue in favor of OSV [Romero-Figeroa, 1997]. Warao is perhaps best described as a "verb-final" language. Nedeb was spoken by three-hundred people, Tobati by three-hundred fifty [Gordon, 2005]. Wik Ngathana was used by a mere hundred and twenty-six people [WURM and HATTORI, 1981]. Given the rarity of these languages, these additional problems with the data must be taken very seriously. A language spoken by a small group of people is prone to grammatical instability. Furthermore, it presents the linguist with an inherent poverty of sources, which translates into noisy measurement, as evidenced by the lack of agreement among researchers.

There are also difficulties associated with mapping the ergative case system to conventional notions of subject and object. According to Dryer:

> “For most OVS languages, there isn’t a lot of evidence bearing on whether they are syntactically ergative, so we can’t really be too sure. But about half the OVS languages I’ve looked at have ergatively-based word order. I.e. they are SV for intransitive clauses so they are really Abs-V-Erg.” (Dryer, personal communication)

In Dryer’s work, object stands for patient and subject for agent, regardless of case. Thus, absolutive in a transitive clause will map to object, ergative to subject. In oither words, absolutive in a transitive clause will map to object, ergative to subject. In contrast, Edith Aldridge [Aldridge, 2004] describes the syntactically ergative language Seediq as follows:

> Seediq is a VOS language, the absolutive always appearing in clause-final position”

Obviously, she uses a structural case based definition of subject as the higher argument, which in this case happens to be the absolutive argument, or thematic patient. Both strategies can be defended and are equally valid. The problem is that the word order data looks different depending on which strategy is chosen. Furthermore, confusion may arise among researchers about the evidence, if data is cited without an explicit discussion of the chosen strategy. Therefore, it is best to leave these languages out of consideration, until a consensus on the mapping between ergative system and subject/object categories is found.

Thus, while the existence of SOV, SVO, VSO and, I believe, VOS, is established with reasonable certainty, the data on OVS and OSV languages is very problematic. If OVS and OSV are indeed possible, it is only in very rare circumstances. I would argue that consequently, typological theories of word order are better off if they do not predict their existence. The explanation of occasional data to the contrary probably lies in statistical or evolutionary perspectives of universality, which can be developed on top of the basic OT substrate.

Many researchers with otherwise different tacks on typology have agreed that some set of factors make subject-before-object order higly preferable. Kayne's antisymmetry approach advocates this view on purely formal grounds [Kayne, 1994]. He asserts that the order of subject and object merely reflects the hiearchical assymetry between specifier and complement. In the absence of structural transformations that introduce new hierarchical relations, the former is higher than the latter, and therefore precedes it. Earlier, Tomlin had approached the issue from a discourse-theoretic and semantic perspective. His ordering principles are based on the notion of thematicity and animacy. A given expression is thematic to the extent that the speaker assumes that the hearer already attnds to the information in the expression. Thus, pronominal elements are more thematic than definite NPs, which in turn are more thematic than their indefinite counterparts. Tomlin formulates the Theme-First Principle, which states that information which is more thematic precedes that which is less so.

**Theme-First Principle (TFP)**: Information inside the clause is ordered by thematicity, from high to low.

One implication of the TFP is that if subjects are more thematic than objects, the former must precede the latter.

Similarly, the Animate-First Principle states that expressions referring to more animated entities precedes those which refer to less animated entities. Animacy is a gradient notion that places humans above other animate entities, which in turn are higher than inanimates:

**Human** $>$ **Other animate** $>$ **Inanimate**

It is also linked to the hierarchy of semantic [5] roles.

**Agent** $>$ **Instrumental** $>$ **Benefactive** $>$ **Patient**

The roles intrinsically cary information about animacy by attributing agency or intent. When the two scales are in conflict, the semantic role hierarchy takes precedence over the pure animacy hierarchy. According to Tomlin, animacy and subjecthood are higly correlated in the basic transitive clause. Thus, subjects are forced to precede object by both the TFP and the AFP.

[5] Also known as thematic roles, not to be confused with the notion of thematicity above

The idea that animacy and thematicity are linked to subjecthood has been advanced by a number of other researchers. Aissen [Aissen, 1999] recounts Michael Silversteins [Silverstein, 1976] claim that elements on the upper end of a proposed person hierarchy are more likely to be agents in transitive propositions, while elements on the lower end are more likely to be patients.

**Local person** > **Pronoun 3rd** > **Proper Noun 3rd** > **Human 3rd** >

**Animate 3rd** > **Inanimate 3rd**

Aissen reformulats Silversteins generalization in the OT framework into a set of proposed universal prominence scales:

**Person scale: Local** > **3rd [Local = 1st, 2nd]**

**Role Scale: Agent** > **Patient**

**Relational Scale: Subject** > **Nonsubject**

By bringing scales into alignment [6], she established a set of constraints on the realization of particular categories as arguments. The set of constraints also links subject choice to discourse prominence - topicality, empathy, perspective, discourse coherence. While Aissen doesn't expressly address the ordering between subject and object, her approach largely backs Tomlin's claims about the influence of animacy and discourse status on subjecthood.

All of the approaches recounted here seek to explain the propensity of subjects to precede objects. However, it seems that two issues are often conflated. One issue is the selection of appropriate categories for the appropriate grammatical roles. Another issue is the position of the arguments once the categories have been selected. It is possble that the two issues influence one another in a global way during processing or on the scale of language evolution. For example, it may be that since subjects precede objects and thematic/animate referents precedes non-thematic/inanimate referents, a sentence in which subject, animacy and theme overlap is more harmonic than a sentence in which they do not. Alternatively,

[6] For a fomal definition of alignment, see [Prince and Smolensky, 1993]

it is possible that subjects precede objects because subject and object are a grammaticized form of the cognitive tendency to order thematic/animate information.

Whatever the interaction may be, it is beyond the scope of this work, which seeks to separate discourse factors from selectional or structural factors at the level of the abstract grammatical system. Consequently, it seems more appropriate to phrase ordering constraints acting in discourse-neutral environments in purely structural terms, *a la* Kayne, rather than in discourse terms. The constraint in (64) is structural in nature, and supported by the antisymmetry condition.

(64) **Spc**$>$ **Cmp**: Specifiers precede complements.

But while this may seem as a mere restatement of antisymmetry as a soft constraint, there is a substantive difference. Kayne's principle leads to the constraint in 64, in addition to two other constraints in (65), (66).

(65) **Hd**$>$ **Cmp**: Heads precede complements.

(66) **Spc** $>$ **Hd**: Specifiers precede heads.

As a consequence, the constraint in (64) is often assumed to be derived from the transitive closure of (65, 66), rather than being fundamental. I will argue that the constraint in (65) has no basis in cross-linguistic typology and that only the constraints in (64, 66) are truly active.

The verb and the specifier precede the complement in three out of the four attested word orders (SVO, VSO, VOS and SOV, SVO, VSO respectively). This strongly supports the reality of (64) and (66). Note however, that (65) is not similarly supported, since only half of the four orders comply with it (SVO, SOV). Furthemore, these two cases can easily be explained without appealing to (65). SVO is the result of interaction between HdCmp (63), Hd$>$Cmp and Spc $>$ Cmp. Assuming that these constraints are reprected by the language the verb must be next to the object and precede it - giving rise to VO, and the subject must be positioned before the object without separating it from the verb, resulting in SVO. SOV on the other hand, respects HdCmp and Spc $>$ Cmp, but not Hd$>$Cmp. Presumably, some other constraint must be overruling the latter. This cannot be our putative Spc $>$ Hd, since it would be obeyed by either SVO and SOV, and consequently leave SOV harmonically bounded. One approach is to create a constraint which is an exact opposite of Hd$>$ Cmp, namely Cmp $>$ Hd. For example, Grimshaw [Grimshaw, 1997] uses two constraints $HdLeft$ and $HdRight$ (definitions in (68))

(67) **HdLeft** :the head is leftmost in its projections.

(68) **HdRight**: the head is rightmost in its projections.

For Grimshaw, the projection of the verb is the VP containing the overt object and the subject (or possibly its trace). At first glance, it appears that the head directionality constraints differ from Cmp$>$Hd and Hd$>$Cmp because they regulate the position of the head with respect to the specifier as well as the complementizer. However, Grimshaw introduces another set of constraints- Spec Left/Right. When these are ranked higher than the head directionality constraints, the effects of the latter are equivalent to Cmp $>$ Hd (HdRight) and Hd$>$Cmp (HdLeft).

There is a general sense in which directly contradicting constraints are contrary to the spirit of OT because they strongly resemble the binary non-interacting parameters of the Principles&Parameters approach, and because they can hardly be independently motivated from a functionalist perspective ([Newmeyer, 2004a], [Newmeyer, 2004b]). One proposal has been to eliminate them by converting them into parametric variation at the base [Costa, 1997]. However, a better appoach would be to substitute them with constraints which have similar effects but are not entirely contradictory.

Note that when the specifier-positioning constraints proposed by Grimshaw are ranked lower than the head directionality constraints, the latter have the effect of pushing to verb to either the left or the right periphery of its phrase. A separate constraint with exaclty that effect has been proposed in [Zepter, 2003]- LexHdEdge (69).

(69) **LexHdEdge**: A lexical head must align with one edge of its phrase.

Whether the head is lexical or not is inconsequential for our purposes. I therefore reformulate LexHdEdge as HdEdge (70) in terms of the head-dependent relationship.

(70) **HdEdge**: The dependents must align on the same side of the head.

HdEdge favors SOV, VOS and VSO.

## 4.4 Order of simple-tense clauses (S, V, O)

With this our constraint set is complete. As Tableau 4.2 demonstrates, there is a ranking which favors each of the four existing orders. The remaining two are harmonically bounded and cannot win under any ranking.

| Candidate | Hd>Cmp | Spc>Cmp | HdEdge | HdCmp |
|---|---|---|---|---|
| SOV | * | | | |
| SVO | | | * | |
| VSO | | | | * |
| VOS | | * | | |
| OVS | * | * | * | |
| OSV | * | * | | * |

Table 4.2: Basic word order typology

## 4.5 Order of complex-tense clauses (S, O, Aux, V)

Despite the successful outcome, the analysis of word order in simple Subject-Verb-Object constructions is hardly satisfying. Since at least four of the candidate permutations are attested (SOV, SVO, VSO, VOS), we must include at least four binary constraints. This can be easily demonstrated by turning the problem into a binary encoding task. Each winning candidate must be "encoded" so that it's non-zero bits are not a subset of any other competitor's non-zero bits. This ensures the candidate is not harmonically bounded by any other candidate in the set. One way of achieving this is not to allow any overlap among the non-zero entries of each competitor. Thus, four winners require four constraints. Allowing overlap will not improve the situation in the case of four winners. [7]. There is hardly gain in an analysis which postulates as many parameters (constraints) as there are data points

[7] If we assume five winners (SOV, SVO, VSO, VOS, OVS), we can do slightly better by allowing some overlap, and reducing the constraint set from five members to four

(candidates). Each winning candidate violates exactly one constraint, which is not violated by other candidate.

In order to confirm the promising result above, I will apply the constraints to an expanded set of data which includes not only the basic subject-verb-object constructions, but also their complex tense counterparts involving auxiliaries. The precise position of the auxiliary can be very informative for theories of word order. Combinatorially, the four elements of complex-tense sentences present a greater challenge for cross-linguistic analysis than the three elements of simple-verb sentences. The possible permutations of subject, object, auxiliary, and verb are twenty-four. Distributional facts about these permutations are a serious test for theories of word order. For example, Kayne's view of word order runs into some difficulty when the auxiliary positon in SOV languages is considered. If SOV languages are derived from SVO, the derivation must involve V movement, and O-movement out of the VP to a position higher than the position of the auxiliary, but lower than that of the verb. Even casting aside the problems associated with postulating such a position, we are still left with the possibility that a language may involve one of the two types of movement alone, predicting the existence of SVAuxO and SOAuxV languages, both of which are unattested [8]. Complex-tense clauses involving auxiliaries are especially important for OT-style analyses of word order, since the abundance of unattested candidates allows for a good constraint-to-competitor ratio. Unlike the $4 : 2$ winner:loser ratio in simple tense clauses, the twenty-four permutations of complex tense clauses boast a challenging $18 : 6$ ratio of unattested to attested forms. This means that by utilizing harmonic binding, it is theoretically possible to arrive at an analysis which eliminates all eighteen competitors and ensures a ranking for the attested six word orders with five constraints. Whether such a minimal set of constraints makes linguistic sense is a different matter, but it is clear that the complex tense clauses have the potential for a much more insightful treatment than the simple tense data. If we take into account the fact that word order of simple and complex tense declaratives covaries across languages, the pool of competitors grows even larger. With six possible basic word order permutations and twenty-four possible complex-tense permutations, the number of possible language types is $24 \times 6$, or $144$ languages. Let us first

[8] Barring cases of auxiliary suffixation, which is irrelevant here

examine the issue using the four constraints we developed for dealing with the basic simple tense clauses. Tableau 4.3 lists each of the twenty-four permutations with their respective violations.

As before, HdCmp is violated whenever verb and object are not adjacent to each other. However, the presence of the auxiliary creates an additional source of violation, because the verb in turn is the complementand of the auxiliary (Figure 4.2). Consequently, a structure which separates the verb from the auxiliary is also in violation of HdCmp. Before proceeding further, let us remind ourselves what "separates" means in this context. In Section 4.3 we opted for a local definition of adjacency, which excluded from consideration subordinate elements (see the defintion of HdCmp in (63). Thus, if the verb and the auxiliary are separated only by the object, this does not constitute a violation of HdCmp. However, if the subject intervenes, HdCmp is violated, because the subject is not a proper subordinate of the verb, as a result of its secondary dependency on the auxiliary.

The change in input structure affects the other three constraints as well. Spc$>$Cmp is now violated not only whenever the object preceeds the subject, but also whenever the verb preceeds the subject, since the latter doubles as a specifier of the auxiliary head, where the verb is a complement. Violations of Hd$>$Cmp are incurred whenever the auxiliary does not precede the verb, and/or the verb does not precede the object. Finally, $HdEdge$ discourages linearizations that place the auxiliary in between the subject and the verb, in addition to those that place the verb in between the subject and the object.

The word orders predicted by the optimization in Tableau 4.3 are SOVAux, SAuxVO, AuxVSO, AuxSOV and AuxVOS. All of them are cross-linguistically attested. While systematic data on auxiliary placement has been collected only in relation to the position of the verb, individual studies of particular languages can be assembled to provide key data. According to Greenberg's original study of thirty languages, all VSO languages place the auxiliary before the verb, all SOV languages place it after the verb, and all but one SVO language behave like VSO. Dryer's larger study upholds this result. Unfortunately, neither study distinguishes between VSO-AuxSVO languages and VSO-AuxVSO languages. However, it is known that Niuean is a VSO-AuxVSO language, while Irish is a VSO-AuxSVO language. In addition to these well-known language types, Scottish Gaelic exibits VSO-AuxSOV alternation [Adger, 1996]. Similarly, two types of possible SVO languages

| Input | Candidate | HdCmp | Spc>Cmp | Hd>Cmp | HdEdge |
|---|---|---|---|---|---|
| Aux S V O | AuxVSO | 1 | 1 | 0 | 0 |
| | VAuxSO | 1 | 1 | 1 | 1 |
| | VSAuxO | 2 | 1 | 1 | 0 |
| | VSOAux | 2 | 1 | 1 | 0 |
| | AuxSVO | 1 | 0 | 0 | 1 |
| | SAuxVO | 0 | 0 | 0 | 2 |
| | SVAuxO | 1 | 0 | 1 | 2 |
| | SVOAux | 0 | 0 | 1 | 1 |
| | AuxSOV | 1 | 0 | 1 | 0 |
| | SAuxOV | 0 | 0 | 1 | 1 |
| | SOAuxV | 1 | 0 | 1 | 1 |
| | SOVAux | 0 | 0 | 2 | 0 |
| | AuxVOS | 0 | 2 | 0 | 0 |
| | VAuxOS | 1 | 2 | 1 | 1 |
| | VOAuxS | 0 | 2 | 1 | 1 |
| | VOSAux | 1 | 2 | 1 | 0 |
| | AuxVSO | 1 | 1 | 0 | 0 |
| | VAuxSO | 1 | 1 | 1 | 1 |
| | VSAuxO | 2 | 1 | 1 | 0 |
| | VSOAux | 2 | 1 | 1 | 0 |
| | AuxOSV | 2 | 1 | 1 | 0 |
| | OAuxSV | 2 | 1 | 1 | 0 |
| | OSAuxV | 1 | 1 | 1 | 1 |
| | OSVAux | 1 | 1 | 2 | 0 |
| | AuxOVS | 0 | 2 | 1 | 1 |
| | OAuxVS | 1 | 2 | 1 | 1 |
| | OVAuxS | 0 | 2 | 2 | 2 |
| | OVSAux | 1 | 2 | 2 | 1 |



Table 4.3: Order in complex tense clauses

center

| | Hd>Cmp | HdEdge |
|---|---|---|
| SAuxOV | 1 | 1 |
| SAuxVO | | 2 |
| SOVAux | 2 | |

Table 4.4: SAuxOV is harmonically bounded by other forms

are known- those that remain SVO in the presense of auxiliary (SAuxVO). Interestingly, virtually no examples of SVO/AuxSVO are attested. Whenever this alternation is found, it is linked to particular auxiliaries within a language that otherwise adheres to the more popular SVO/SAuxVO alternation (e.g. Breton [9]). As for SOV languages, the overwhelming majority of them is Aux-final (SOVAux). Similarly for VOS languages, it seems that prefixing the auxiliary is the only option (AuxVOS). Specific examples will be provided in Section 4.7, where I will discuss the link between simple and comlex-tense order.

However, this typology excludes at least one word order which is known to exist. Complex-tense clauses in German and Dinka are SAuxOV. This order is harmonically bounded by two other forms:SOVAux and SAuxVO (see Tableau 4.4). The three candidates are equally good on $HdCmp$ and Spc>Cmp. However, for each of the two remaining constraint there is a candidate which is better than SAuxOV. Therefore, it appears the analysis must be modified to allow for this order.

Note however, that both languages that exhibit the SAuxOV order do so only in main clauses. In subordinate clauses, German reverts to SOVAux, while Dinka migrates to AuxSOV. Thus, it is reasonable to assume that SAuxOV results from some structural and/or discourse property of main clauses, rather than from a fundamental ordering constraint. A subordinate clause is typically headed by a complementizer, which may attract or block the leftward movement (or "raising") of elements. The typological literature has often focused on subordinate clauses as indicators of the underlying word order. For example, German has been classified as SOV with an additional verb-second parameter specific to

[9] Although I consider Breton to be an $SVO/SAuxVO$ language, the issue is highly contentious (Legendre, pers. comm.)

main clauses ([Erdmann, 1990]). However, it is not clear why subordinate clauses are a better indicator of basic order than main clauses. Obviously, they are rarer than main clauses. Furthermore, they are aquired late and are not well represented in child-directed speech. A more theory-neutral approach would simply take into account that main/subordinate clause status affects word order in some languages.

## 4.6 Order and main/subordinate clause status

Subordinate clauses can be distinguished by the type of subordinating conjunction-adverbalizer, relativizer, or complementizer [Loos et al., 2004]. An adverbalizer indicates that the subordinate clause has an ”adverbial” or interpropositional relation to the main clause, indicating purpose, condition, time, and location. A relativizer introduces a clause which describes the referent of a head noun or pronoun. Finally, a complementizer heads a clause that acts as a complement to the main verb. Some languages employ different word orders for different types of subordinate clauses. Since the structural and discourse status of relative and adverbial clauses is generally more involved than that of complementizer clauses, I will confine myself to examining the latter. The term subordinate clause in this study should be understood to refer to complement clauses, unless otherwise noted.

It is reasonable to assume that word order discrepancies between main and subordinate clauses are due to some aspects of the structural difference. Researchers in other frameworks have made similar assumptions on a language particular basis. The widely-accepted transformational analyses of German word order appeals to the notion that the null complementizer head attracts the inflectional head via head-movement, and the subject (or any other noun phrase) to the specifier position. Thus, the V2 property of main clauses is a result of the coincidental head-movement and phrasal movement to the complementizer projection (73). The latter is usually referred to as some form of topicalization. In the absence of topicalized constituents, the subject serves as default topic. Head-movement, on the other hand, is attributed to a ban against empty C-head in German. The problem with this analysis is that it is not clear why German should ban empty C-heads in declaratives, while other languages, like English, behave just the opposite. Topicalization to SpecCP in En-

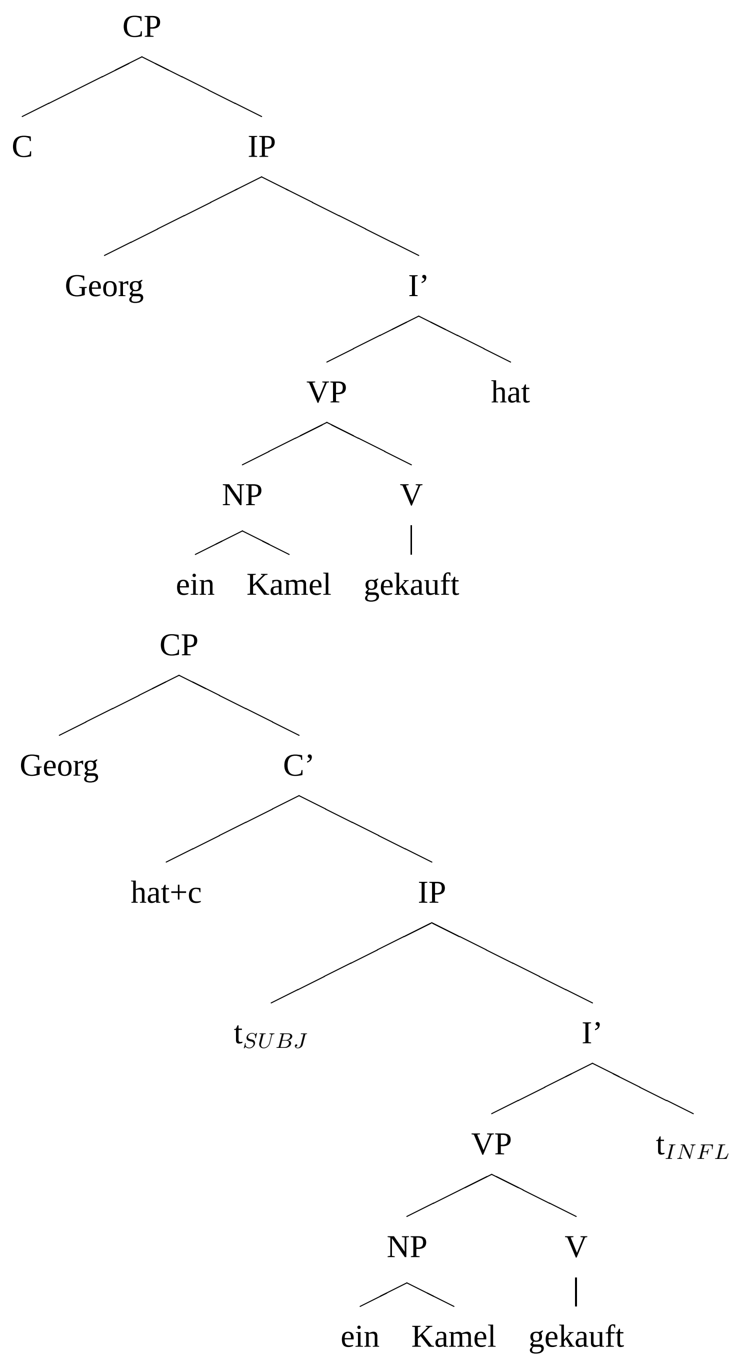



Figure 4.9: Movement to C in German

glish occurs strictly without head-movement to C, which is similarly absent in clauses that do not involve topicalization (i.e. the default topic case, Figure 4.10). There is something

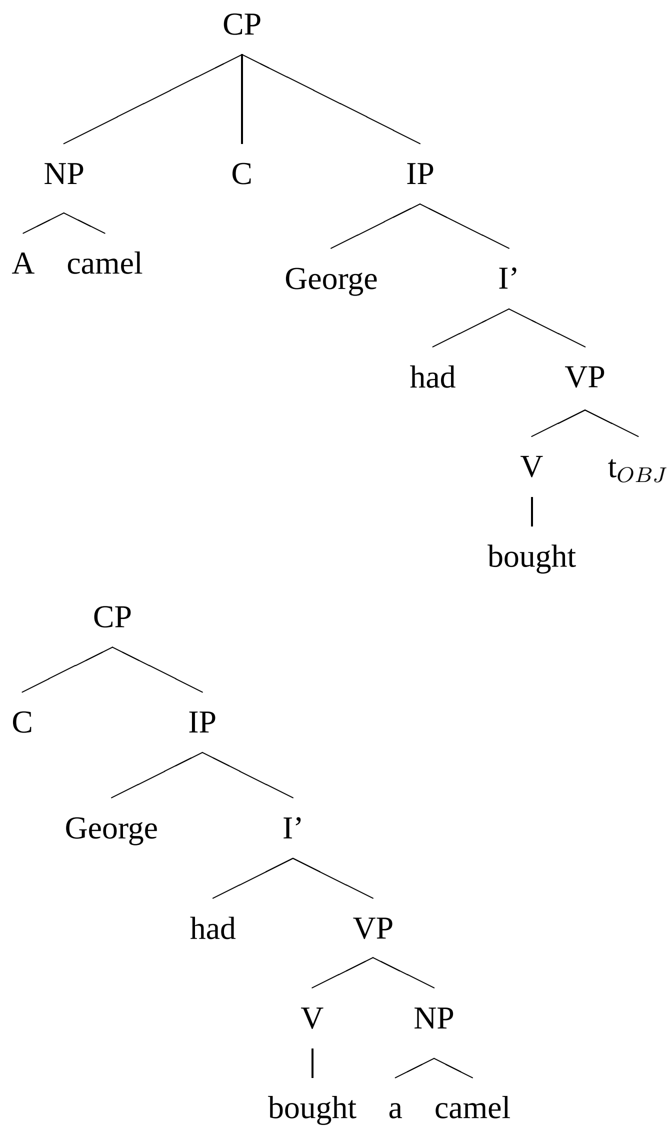



Figure 4.10: Lack of head-movement to C in English

troubling about a cross-linguistic theory permissive enough to allow languages to take either a positive or negative attitudes to empty complementizer heads. This is especially true because other phenomena demonstrate that empty heads are undesirable and languages take various measures to avoid them. Examples of such phenomena include verb movement to

Infl, and do-support, at least one of which is a property of English. Thus, languages should at best tolerate covert complementizers, not actively defend them from incorporation.

I propose that in essence, any null head, including complementizers, is banned from standing alone in any language. Rather, it must merge with its complement, giving rise to observable word order distortions, such as the main/subordinate order dichotomy and the simple/complex-tense covariance alluded to in Section 4.5. Earlier, I suggested that the structure of simple-tense clauses is derived by merging the covert inflectional head with the verb head (Figure 4.6). The structural differences between subordinate and main clauses can be handled analogously. Subordinates are headed by overt complementizers (e.g. English *that*, Bulgarian *che*), while main clauses are not. The dependency representation of a declarative with overt and covert complementizer is given in Figure 4.11 How does this

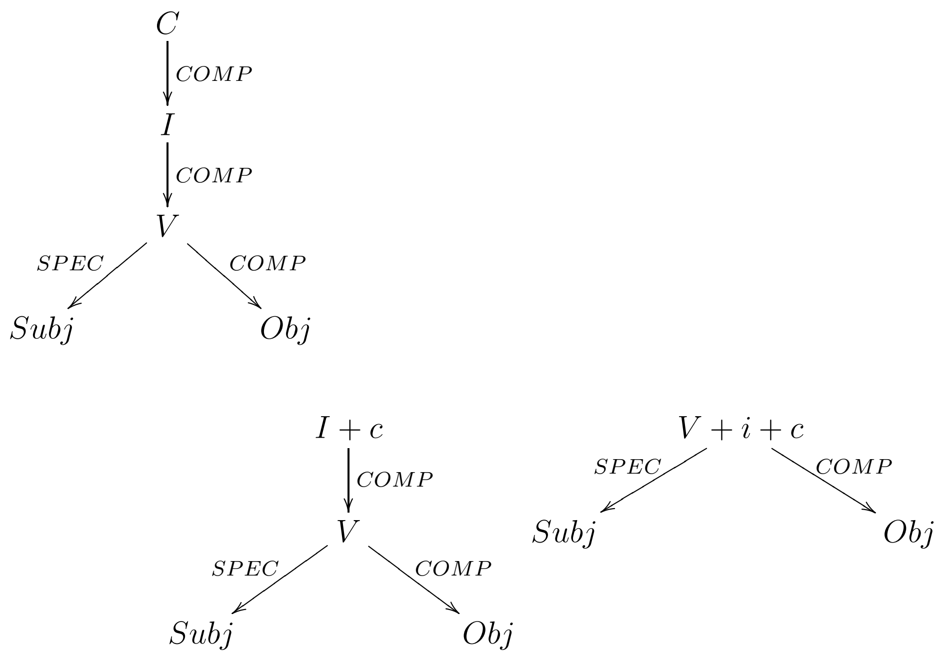



Figure 4.11: Structure of sentences with or without overt complementizers

difference in structure affect linearization? The absence of overt complementizer changes the number and type of competitors. When the complementizer is overt, all permutations of the five elements (C, Aux, V, S and O in the case of overt Infl) are part of the candidate set. The constraints I used so far determine little about the position of the complementizer. HdEdge is a local constraint which simply makes sure that all dependents are to one side of the head. The complementizer has only one dependent (Infl), which means HdEdge is vacuously satisfied. Since complementizers can precede or follow, but not interfere within the structure headed by Infl-Aux, it is necessary to introduce a constraint which would ensure that the complementizer is placed at the edge of the embedded phrase. One such constraint can be derived from the Hudson Adjacency Principle, widely embraced in the dependency grammar literature (see Chapter 3):

**Hudson's Adjacency Principle** : Every dependent D of a head H must be adjacent to H. D is adjacent to H provided that every word between D and H is a subordinate of H in the dependency tree.

Informally, it means that a head and its dependents may not be separated by superordinates of the head. It enforces a context-free structure, which allows nesting, but no crossing dependencies. The motivation for including this principle as a constraint is clear. Natural language is overwhelmingly context-free (cf. for example [Pullum, 1984]). However, it is important that the constraint, unlike a principle, is violable under certain conditions, since crossing dependencies as counterexamples exist in some languages [Shieber, 1985]. Introducing Hudson-Adjacency (HA) as a soft constraint allows us to both capture the overwhelming context-free tendency of natural language, and account for exceptions. It also ensures that the complementizer stays at the periphery of the embedded clause.

However, I will argue that we do not need HA as a separate constraint, since it overlaps in function with a constraint, which is already a necessary part of the anaylisis- HdEdge. If HdEdge were a global constraint, requiring that arguments align to one side along with their subordinates, its effect will mimic that of HdEdge-local and HA combined (71).

(71) **HdEdge** reformulated: The head must align with one edge of its phrase.

Having one constraint instead of two is more economical, and, as we will see below, fully adequate on empirical grounds. Global HdEdge will make sure that the complementizer remains at the edge of the subordinate clause. Hd$>$Cmp will keep it to the left, but pressure from outside the clause may further influence its position, if this constraint is ranked low. Thus, despite the increased number of initial competitors, the Tableau 4.3 remains valid, since candidates with internal complement positions are harmonically bounded.

Unlike the competitors in the linearization of subordinates, the candidates in the main clause linearization have at most four elements [10] (Aux+c, V, S and O, see Figure 4.11). In this situation, the status of S-Aux(+c)-OV must be promoted from harmonically bounded to a winner under some ranking. The intuition I will pursue is essentially the one captured in transformational analyses involving movement to C: Empty complementizers cannot be clause-intitial. However, I will claim that a) this is a soft prference; b) it pertains not to empty complementizers *per se*, but to the complex heads containing them (i.e. Aux+c and V+Aux+c, Figure 4.11) and c) it resticts not only clause-initial but also clause-final positions. Thus, I postulate a new constraint $*cEdge$

$*cEdge$ Complex heads containing complementizers cannot appear at the clausal edge.

Thus, SAux+cOV beats SOVAux+c and Aux+cSOV under some rankings (Tableau 4.5).

| Input | Candidate | HdEdge | Hd>Cmp | Spc>Cmp | *cEdge | HdCmp |
|---|---|---|---|---|---|---|
| Aux+cVSO | AuxVSO | 0 | 0 | 1 | 1 | 1 |
| | AuxSVO | 1 | 0 | 0 | 1 | 1 |
| | SAuxVO | 2 | 0 | 0 | 0 | 0 |
| | AuxSOV | 0 | 1 | 0 | 1 | 1 |
| | SAuxOV | 1 | 1 | 0 | 0 | 0 |
| | SOVAux | 0 | 2 | 0 | 1 | 0 |
| | AuxVOS | 0 | 0 | 2 | 1 | 0 |

Table 4.5: Complex-tense main clauses

By the same logic, the typology of main clauses without an auxiliary differs from the subordinate typology in Tableau 4.2. However, the difference is not in the predicted typology -it remains SOV, SVO, VSO, VOS. Instead, the difference is which word orders are licensed under which rankings. As we will see in the next section, this is crucial for deriving order correlations across constructions and arriving at a full linguistic typology of basic word order.

[10] Or three in the absence of an overt auxiliary

## 4.7 Order correlations across constructions

If all combinations of possible orders in main and subordinate clauses are considered, the number of possible languages rises to $144^2$, or a total of $20736$ languages! A solid theory of word order must be able to eliminate the vast majority of these possibilities. Transformational analyses attribute the correlation to an underlying word order surfacing in subordinate clauses, which gives rise to a derived order in the main clause. I will demonstrate that the analysis presented here captures the crucial fact that the word order of main and subordinate clauses with or without an auxiliary is not independent.

It is easy to see that the notable cross-linguistic tendency of basic and complex constuction is that of word order conservation. If the auxiliary is adjacent to the verb, the order of subject, object and verb remain the same in its absence. Languages complying with this tendency exhibit SOVAux/SOV (Latin), SAuxVO/SVO (English), AuxVSO/VSO (Niuean) and AuxVOS/VOS (Tsou) order in both main and subordinate clauses.

(72) m-i-ta e-mafe to-fuzu ’e-Pasuya.
AV-Rea-3S AV-like to eat Obl-wild boar Nom-Pasuya
’Pasuya liked to eat wild boar.’

The five constraints defined above can be ranked to account for each of these cases (see Tableau 4.6, Tableau 4.7, Tableau 4.8 and Tableau 4.9).

If the auxiliary is not adjacent to the verb, the verb replaces the auxiliary whenever the latter is not present. SVO languages may alternate with either VSO or SOV in the simple subordinate, while VSO and SOV themselves are not known to alternate. Whenever the presence of an auxiliary affects order, it generally pushes the verb further to the right, after the subject and sometimes after the object. The attested orders are AuxSVO/VSO, AuxSOV/VSO, AuxVOS/VOS, and SAuxOV/SVO. The first three alternations are known to occur in both main and subordinate clauses, either straight through or in combination with other orders. I have listed all examples I was able to find in Table 4.7. The source of the data is indicated in the last column, along with reference to examples for the less known languages whenever available.

(73) Georg hat ein Kamel gekauft.

| Input | Candidate | HdCmp | Spec>Com | HdEdge | Hd>Cmp | *cEdge |
|---|---|---|---|---|---|---|
| CAuxVSO | AuxVSO | 1 | 1 | 0 | 0 | 0 |
| | AuxSVO | 1 | 0 | 1 | 0 | 0 |
| | SAuxVO | 0 | 0 | 2 | 0 | 0 |
| | AuxSOV | 1 | 0 | 0 | 1 | 0 |
| | SAuxOV | 0 | 0 | 1 | 1 | 0 |
| ☞ | SOVAux | 0 | 0 | 0 | 2 | 0 |
| | AuxVOS | 0 | 2 | 0 | 0 | 0 |
| Aux+c VSO | cAuxVSO | 1 | 1 | 0 | 0 | 1 |
| | cAuxSVO | 1 | 0 | 1 | 0 | 1 |
| | ScAuxVO | 0 | 0 | 2 | 0 | 0 |
| | cAuxSOV | 1 | 0 | 0 | 1 | 1 |
| | ScAuxOV | 0 | 0 | 1 | 1 | 0 |
| ☞ | SOVcAux | 0 | 0 | 0 | 2 | 1 |
| | cAuxVOS | 0 | 2 | 0 | 0 | 1 |
| CAuxVSO | xVSO | 1 | 0 | 0 | 0 | 0 |
| | SxVO | 0 | 0 | 1 | 0 | 0 |
| ☞ | SOVx | 0 | 0 | 0 | 1 | 0 |
| | xVOS | 0 | 1 | 0 | 0 | 0 |
| V+i+c SO | cxVSO | 1 | 0 | 0 | 0 | 1 |
| | ScxVO | 0 | 0 | 1 | 0 | 0 |
| ☞ | SOVxc | 0 | 0 | 0 | 1 | 1 |
| | cxVOS | 0 | 1 | 0 | 0 | 1 |

Table 4.6: SOV/SOVAux ranking

| Input | Candidate | Hd>Cmp | Spc>Cmp | *cEdge | HdCmp | HdEdge |
|---|---|---|---|---|---|---|
| CAuxVSO | AuxVSO | 0 | 1 | 0 | 1 | 0 |
| | AuxSVO | 0 | 0 | 0 | 1 | 1 |
| ☞ | SAuxVO | 0 | 0 | 0 | 0 | 2 |
| | AuxSOV | 1 | 0 | 0 | 1 | 0 |
| | SAuxOV | 1 | 0 | 0 | 0 | 1 |
| | SOVAux | 2 | 0 | 0 | 0 | 0 |
| | AuxVOS | 0 | 2 | 0 | 0 | 0 |
| Aux+c VSO | cAuxVSO | 0 | 1 | 1 | 1 | 0 |
| | cAuxSVO | 0 | 0 | 1 | 1 | 1 |
| ☞ | ScAuxVO | 0 | 0 | 0 | 0 | 2 |
| | cAuxSOV | 1 | 0 | 1 | 1 | 0 |
| | ScAuxOV | 1 | 0 | 0 | 0 | 1 |
| | SOVcAux | 2 | 0 | 1 | 0 | 0 |
| | cAuxVOS | 0 | 2 | 1 | 0 | 0 |
| CAuxVSO | xVSO | 0 | 0 | 0 | 1 | 0 |
| ☞ | SxVO | 0 | 0 | 0 | 0 | 1 |
| | SOVx | 1 | 0 | 0 | 0 | 0 |
| | xVOS | 0 | 1 | 0 | 0 | 0 |
| V+i+c SO | cxVSO | 0 | 0 | 1 | 1 | 0 |
| ☞ | ScxVO | 0 | 0 | 0 | 0 | 1 |
| | SOVxc | 1 | 0 | 1 | 0 | 0 |
| | cxVOS | 0 | 1 | 1 | 0 | 0 |

Table 4.7: SVO/SAuxVO ranking

| Input | Candidate | HdEdge | Hd>Cmp | Spc>Cmp | *cEdge | HdCmp |
|---|---|---|---|---|---|---|
| CAuxVSO ☞ | AuxVSO | 0 | 0 | 1 | 0 | 1 |
| | AuxSVO | 1 | 0 | 0 | 0 | 1 |
| | SAuxVO | 2 | 0 | 0 | 0 | 0 |
| | AuxSOV | 0 | 1 | 0 | 0 | 1 |
| | SAuxOV | 1 | 1 | 0 | 0 | 0 |
| | SOVAux | 0 | 2 | 0 | 0 | 0 |
| | AuxVOS | 0 | 0 | 2 | 0 | 0 |
| Aux+c VSO ☞ | cAuxVSO | 0 | 0 | 1 | 1 | 1 |
| | cAuxSVO | 1 | 0 | 0 | 1 | 1 |
| | ScAuxVO | 2 | 0 | 0 | 0 | 0 |
| | cAuxSOV | 0 | 1 | 0 | 1 | 1 |
| | ScAuxOV | 1 | 1 | 0 | 0 | 0 |
| | SOVcAux | 0 | 2 | 0 | 1 | 0 |
| | cAuxVOS | 0 | 0 | 2 | 1 | 0 |
| CAuxVSO ☞ | xVSO | 0 | 0 | 0 | 0 | 1 |
| | SxVO | 1 | 0 | 0 | 0 | 0 |
| | SOVx | 0 | 1 | 0 | 0 | 0 |
| | xVOS | 0 | 0 | 1 | 0 | 0 |
| V+i+c SO ☞ | cxVSO | 0 | 0 | 0 | 1 | 1 |
| | ScxVO | 1 | 0 | 0 | 0 | 0 |
| | SOVxc | 0 | 1 | 0 | 1 | 0 |
| | cxVOS | 0 | 0 | 1 | 1 | 0 |

Table 4.8: VSO/AuxVSO ranking

| Input | Candidate | Hd>Cmp | HdCmp | HdEdge | *cEdge | Spc>Cmp |
|---|---|---|---|---|---|---|
| CAuxVSO | AuxVSO | 0 | 1 | 0 | 0 | 1 |
| | AuxSVO | 0 | 1 | 1 | 0 | 0 |
| | SAuxVO | 0 | 0 | 2 | 0 | 0 |
| | AuxSOV | 1 | 1 | 0 | 0 | 0 |
| | SAuxOV | 1 | 0 | 1 | 0 | 0 |
| | SOVAux | 2 | 0 | 0 | 0 | 0 |
| ☞ | AuxVOS | 0 | 0 | 0 | 0 | 2 |
| Aux+c VSO | cAuxVSO | 0 | 1 | 0 | 1 | 1 |
| | cAuxSVO | 0 | 1 | 1 | 1 | 0 |
| | ScAuxVO | 0 | 0 | 2 | 0 | 0 |
| | cAuxSOV | 1 | 1 | 0 | 1 | 0 |
| | ScAuxOV | 1 | 0 | 1 | 0 | 0 |
| | SOVcAux | 2 | 0 | 0 | 1 | 0 |
| ☞ | cAuxVOS | 0 | 0 | 0 | 1 | 2 |
| CAuxVSO | xVSO | 0 | 1 | 0 | 0 | 0 |
| | SxVO | 0 | 0 | 1 | 0 | 0 |
| | SOVx | 1 | 0 | 0 | 0 | 0 |
| ☞ | xVOS | 0 | 0 | 0 | 0 | 1 |
| V+i+c SO | cxVSO | 0 | 1 | 0 | 1 | 0 |
| | ScxVO | 0 | 0 | 1 | 0 | 0 |
| | SOVxc | 1 | 0 | 0 | 1 | 0 |
| ☞ | cxVOS | 0 | 0 | 0 | 1 | 1 |

Table 4.9: VOS/AuxVOS ranking

| Main | Main+Aux | Sub | Sub+Aux | Language | Source |
|---|---|---|---|---|---|
| SVO | SAOV | SOV | SOVA | German | Own data 73 |
| SVO | SAOV | VSO | ASOV | Dinka | [Dryer, 1997] |
| SVO | SAVO | VSO | ASVO | Breton | [Bury, 2000], 74 |
| VSO | ASVO | VSO | ASVO | Welsh | [Borsley, 2005],[Carnie, 1991], 75 |
| VSO | ASOV | VSO | ASOV | S. Gaelic | [Adger, 1996], 77 |
| SOV | SAOV | - | - | Wan[11] | Nikitina (pers.com.), 78 |

George has a camel bought.

George bought a camel.

Laura hat mir gesagt das Georg ein Kamel gekauft hat.

Laura has me-Dat. said George a camel bought has.

Laura has told me that George bought a camel.

(74) Me a re al levr dezhi breman. (VSO) [Bury, 2000]

I Prt. give the book to-her now.

I am giving the book to her now.

Mona a lavar e oar Yann ar respont

Mona Prt. say that knows Yann the answer

(75) Dywedodd Gwyn fod Emrys yn ddiog. (Welsh [Borsley, 2005])

said Gwyn be Emrys PRED lazy

Gwyn said Emrys was lazy.

T na Clingena ag scaoileadh na fasar (Welsh [Carnie, 1991])

Be-Pres the Klingons Prog fire the phasers-Gen[12]

The Klingons are firing the phasers

(76) Tha mi air an cat a bhualadh (Scottish Gaelic [Adger, 1996])

be-Pres. I *asp* the-COM cat-COM *prt* strike ‘I have struck the cat’

Aux Subj Obj Verb

[12] Accusative is also possible colloquially

(77) * Tha mi air a bhualadh an cat
be-PRES I ASP Prt close-VN the-COM cat-COM ‘I have struck the cat’
Aux Subj Verb Obj

Rather peculiarly however, the last order is only found in clauses without overt complementizer. In German, it appears as main-clause order and subordinate order in the absence of complementizer. Similar restrictions apply in Dutch [Santorini, 1992]. Dinka- a Nilotic language of Southern Sudan, is a complex blend of German word order in the main clause and that Scottish Gaelic in the subordinate. It employs SVO and SAuxOV in main clauses (like German), but switches to VSO and AuxSOV in subordinate clauses. The only case in which it appears throughout is the Mande language Wan, which completely lacks complementizers (78). According to Tatiana Nikitina, who did fieldwork on the SOV/SAuxOV alternation:

> In Wan...,there are no real syntactically subordinate clauses with auxiliaries, it only has embedded nonfinite clauses and nominalized clauses. In Wan, ”John said that Mary was eating fish” will be expressed as ”John said Mary is eating fish” (no subordinator; word order in the second clause is the same as in an independent sentence). It will look like that (tones and some vowels simplified):
>
> (78) e1 ge3 Marie a3 kpO3 lO1 le3
> he.nom said M. aux fish eat prgr
>
> When the subjects are coreferential (”John said he is eating fish”), a logophoric pronoun will be used, but the construction remains the same:
>
> (79) e1 ge3 bha2 a3 kpO3 lO1 le3
> he.nom said he.log aux fish eat prgr
>
> Synactically, such sentences look just like two independent clauses (same with other predicates like ’think’, ’decide’, etc.).

> With some predicates that take subordinate clauses in English, e.g., 'be sure (that)', the construction can look a little different, literally, "He will come, I am sure in THAT", where a pronoun refers back to a finite clause. In such cases, again, no evidence of syntactic subordination is present. Most importantly, the word order is always SAuxOV.

This unexpected observation does in fact follow from my analysis. Recall that SAuxOV would be harmonically bounded were it not for $*cEdge$, which is only active in clauses with null complementizers. Hence, it is never expected to appear in conjunction with a full complementizer [13]. The rankings deriving each of the mixed order languages is presented in Tableaux 4.10,4.12 ,4.11 ,4.13, 4.14 , and 4.15.

Now let us turn our attention to some other predicted languages. Tableau 4.16 shows the emergence of a language which is "underlyingly" VSO/AuxVSO in the subordinate clause, but employs subject raising in the main clause - SVO/SAuxVO. Another predicted language is presented by Tableau 4.17. It is also AuxVSO/VSO in the subordinate, but like German in the main clause. This language can be described as having separate underlying orders for the two types of clauses. If so, it involves V-to-I raising from an (Aux)SOV order in some types of clauses, and subject raising in main clauses.

A third predicted language (Tableau 4.18) is VOS/AuxVOS in the subordinate and SVO/SAuxOV in the main clause. Aside from the order of subject and object, this language is the same as the previous one. Finally, the last predicted language appears to actually exist, according to some data uncovered after the completion of this analysis. Just like the third predicted language, this one is VOS/AuxVOS in the subordinate. However, it is SVO/SAuxVO in the main clause (like English). The ranking deriving it is shown in Tableau tab:pr4 Miya is a West Chadic language, described for the first time by Schuh, who

[13] Even though Vikner 1991 calls some languages "generalized verb-second" languages said to exhibit word order in all types of clauses, the two he considers, Yiddish and Icelandic, are not SAuxOV (see example from Yiddish below)

(1) ... az di Kinder hobn zikh gelernt geschickte [Santorini, 1992]
that the children have Refl. learned history

For a further discussion of these types of verb second, see [Santorini, 1992] and [Iatridou and Kroch, 1992]

| Input | Candidate | HdCmp | Spc>Cmp | *cEdge | HdEdge | Hd>Cmp |
|---|---|---|---|---|---|---|
| CAuxVSO | AuxVSO | 1 | 1 | 0 | 0 | 0 |
| | AuxSVO | 1 | 0 | 0 | 1 | 0 |
| | SAuxVO | 0 | 0 | 0 | 2 | 0 |
| | AuxSOV | 1 | 0 | 0 | 0 | 1 |
| | SAuxOV | 0 | 0 | 0 | 1 | 1 |
| ☞ | SOVAux | 0 | 0 | 0 | 0 | 2 |
| | AuxVOS | 0 | 2 | 0 | 0 | 0 |
| Aux+c VSO | cAuxVSO | 1 | 1 | 1 | 0 | 0 |
| | cAuxSVO | 1 | 0 | 1 | 1 | 0 |
| | ScAuxVO | 0 | 0 | 0 | 2 | 0 |
| | cAuxSOV | 1 | 0 | 1 | 0 | 1 |
| ☞ | ScAuxOV | 0 | 0 | 0 | 1 | 1 |
| | SOVcAux | 0 | 0 | 1 | 0 | 2 |
| | cAuxVOS | 0 | 2 | 1 | 0 | 0 |
| CAuxVSO | xVSO | 1 | 0 | 0 | 0 | 0 |
| | SxVO | 0 | 0 | 0 | 1 | 0 |
| ☞ | SOVx | 0 | 0 | 0 | 0 | 1 |
| | xVOS | 0 | 1 | 0 | 0 | 0 |
| V+i+c SO | cxVSO | 1 | 0 | 1 | 0 | 0 |
| ☞ | ScxVO | 0 | 0 | 0 | 1 | 0 |
| | SOVxc | 0 | 0 | 1 | 0 | 1 |
| | cxVOS | 0 | 1 | 1 | 0 | 0 |

Table 4.10: German ranking

| Input | Candidate | HdCmp | Spc>Cmp | *cEdge | HdEdge | Hd>Cmp |
|---|---|---|---|---|---|---|
| CAuxVSO | AuxVSO | 0 | 1 | 0 | 0 | 1 |
| ☞ | AuxSVO | 0 | 0 | 0 | 1 | 1 |
| | SAuxVO | 0 | 0 | 0 | 2 | 0 |
| | AuxSOV | 1 | 0 | 0 | 0 | 1 |
| | SAuxOV | 1 | 0 | 0 | 1 | 0 |
| | SOVAux | 2 | 0 | 0 | 0 | 0 |
| | AuxVOS | 0 | 2 | 0 | 0 | 0 |
| Aux+c VSO | cAuxVSO | 0 | 1 | 1 | 0 | 1 |
| | cAuxSVO | 0 | 0 | 1 | 1 | 1 |
| ☞ | ScAuxVO | 0 | 0 | 0 | 2 | 0 |
| | cAuxSOV | 1 | 0 | 1 | 0 | 1 |
| | ScAuxOV | 1 | 0 | 0 | 1 | 0 |
| | SOVcAux | 2 | 0 | 1 | 0 | 0 |
| | cAuxVOS | 0 | 2 | 1 | 0 | 0 |
| CAuxVSO ☞ | xVSO | 0 | 0 | 0 | 0 | 1 |
| | SxVO | 0 | 0 | 0 | 1 | 0 |
| | SOVx | 1 | 0 | 0 | 0 | 0 |
| | xVOS | 0 | 1 | 0 | 0 | 0 |
| V+i+c SO | cxVSO | 0 | 0 | 1 | 0 | 1 |
| ☞ | ScxVO | 0 | 0 | 0 | 1 | 0 |
| | SOVxc | 1 | 0 | 1 | 0 | 0 |
| | cxVOS | 0 | 1 | 1 | 0 | 0 |

Table 4.11: Breton ranking

| Input | Candidate | Spc>Cmp | *cEdge | HdEdge | Hd>Cmp | HdCmp |
|---|---|---|---|---|---|---|
| CAuxVSO | AuxVSO | 1 | 0 | 0 | 0 | 1 |
| | AuxSVO | 0 | 0 | 1 | 0 | 1 |
| | SAuxVO | 0 | 0 | 2 | 0 | 0 |
| ☞ | AuxSOV | 0 | 0 | 0 | 1 | 1 |
| | SAuxOV | 0 | 0 | 1 | 1 | 0 |
| | SOVAux | 0 | 0 | 0 | 2 | 0 |
| | AuxVOS | 2 | 0 | 0 | 0 | 0 |
| Aux+c VSO | cAuxVSO | 1 | 1 | 0 | 0 | 1 |
| | cAuxSVO | 0 | 1 | 1 | 0 | 1 |
| | ScAuxVO | 0 | 0 | 2 | 0 | 0 |
| | cAuxSOV | 0 | 1 | 0 | 1 | 1 |
| ☞ | ScAuxOV | 0 | 0 | 1 | 1 | 0 |
| | SOVcAux | 0 | 1 | 0 | 2 | 0 |
| | cAuxVOS | 2 | 1 | 0 | 0 | 0 |
| CAuxVSO ☞ | xVSO | 0 | 0 | 0 | 0 | 1 |
| | SxVO | 0 | 0 | 1 | 0 | 0 |
| | SOVx | 0 | 0 | 0 | 1 | 0 |
| | xVOS | 1 | 0 | 0 | 0 | 0 |
| V+i+c SO | cxVSO | 0 | 1 | 0 | 0 | 1 |
| ☞ | ScxVO | 0 | 0 | 1 | 0 | 0 |
| | SOVxc | 0 | 1 | 0 | 1 | 0 |
| | cxVOS | 1 | 1 | 0 | 0 | 0 |

Table 4.12: Dinka ranking

| Input | Candidate | Hd>Cmp | Spc>Cmp | HdEdge | *cEdge | HdCmp |
|---|---|---|---|---|---|---|
| CAuxVSO | AuxVSO | 0 | 1 | 0 | 0 | 1 |
| ☞ | AuxSVO | 0 | 0 | 1 | 0 | 1 |
| | SAuxVO | 0 | 0 | 2 | 0 | 0 |
| | AuxSOV | 1 | 0 | 0 | 0 | 1 |
| | SAuxOV | 1 | 0 | 1 | 0 | 0 |
| | SOVAux | 2 | 0 | 0 | 0 | 0 |
| | AuxVOS | 0 | 2 | 0 | 0 | 0 |
| Aux+c VSO | cAuxVSO | 0 | 1 | 0 | 1 | 1 |
| ☞ | cAuxSVO | 0 | 0 | 1 | 1 | 1 |
| | ScAuxVO | 0 | 0 | 2 | 0 | 0 |
| | cAuxSOV | 1 | 0 | 0 | 1 | 1 |
| | ScAuxOV | 1 | 0 | 1 | 0 | 0 |
| | SOVcAux | 2 | 0 | 0 | 1 | 0 |
| | cAuxVOS | 0 | 2 | 0 | 1 | 0 |
| CAuxVSO ☞ | xVSO | 0 | 0 | 0 | 0 | 1 |
| | SxVO | 0 | 0 | 1 | 0 | 0 |
| | SOVx | 1 | 0 | 0 | 0 | 0 |
| | xVOS | 0 | 1 | 0 | 0 | 0 |
| V+i+c SO ☞ | cxVSO | 0 | 0 | 0 | 1 | 1 |
| | ScxVO | 0 | 0 | 1 | 0 | 0 |
| | SOVxc | 1 | 0 | 0 | 1 | 0 |
| | cxVOS | 0 | 1 | 0 | 1 | 0 |

Table 4.13: Welsh ranking

| Input | Candidate | HdEdge | Spc>Cmp | Hd>Cmp | *cEdge | HdCmp |
|---|---|---|---|---|---|---|
| CAuxVSO | AuxVSO | 0 | 1 | 0 | 0 | 1 |
| | AuxSVO | 1 | 0 | 0 | 0 | 1 |
| | SAuxVO | 2 | 0 | 0 | 0 | 0 |
| ☞ | AuxSOV | 0 | 0 | 1 | 0 | 1 |
| | SAuxOV | 1 | 0 | 1 | 0 | 0 |
| | SOVAux | 0 | 0 | 2 | 0 | 0 |
| | AuxVOS | 0 | 2 | 0 | 0 | 0 |
| Aux+c VSO | cAuxVSO | 0 | 1 | 0 | 1 | 1 |
| | cAuxSVO | 1 | 0 | 0 | 1 | 1 |
| | ScAuxVO | 2 | 0 | 0 | 0 | 0 |
| ☞ | cAuxSOV | 0 | 0 | 1 | 1 | 1 |
| | ScAuxOV | 1 | 0 | 1 | 0 | 0 |
| | SOVcAux | 0 | 0 | 2 | 1 | 0 |
| | cAuxVOS | 0 | 2 | 0 | 1 | 0 |
| CAuxVSO ☞ | xVSO | 0 | 0 | 0 | 0 | 1 |
| | SxVO | 1 | 0 | 0 | 0 | 0 |
| | SOVx | 0 | 0 | 1 | 0 | 0 |
| | xVOS | 0 | 1 | 0 | 0 | 0 |
| V+i+c SO ☞ | cxVSO | 0 | 0 | 0 | 1 | 1 |
| | ScxVO | 1 | 0 | 0 | 0 | 0 |
| | SOVxc | 0 | 0 | 1 | 1 | 0 |
| | cxVOS | 0 | 1 | 0 | 1 | 0 |

Table 4.14: Scottish Gaelic ranking

| Input | Candidate | HdCmp | Spc>Cmp | *cEdge | HdEdge | Hd>Cmp |
|---|---|---|---|---|---|---|
| Aux+c VSO | cAuxVSO | 1 | 1 | 1 | 0 | 0 |
| | cAuxSVO | 1 | 0 | 1 | 1 | 0 |
| | ScAuxVO | 0 | 0 | 0 | 2 | 0 |
| | cAuxSOV | 1 | 0 | 1 | 0 | 1 |
| ☞ | ScAuxOV | 0 | 0 | 0 | 1 | 1 |
| | SOVcAux | 0 | 0 | 1 | 0 | 2 |
| | cAuxVOS | 0 | 2 | 1 | 0 | 0 |
| V+i+c SO | cxVSO | 1 | 0 | 1 | 0 | 0 |
| ☞ | ScxVO | 0 | 0 | 0 | 1 | 0 |
| | SOVxc | 0 | 0 | 1 | 0 | 1 |
| | cxVOS | 0 | 1 | 1 | 0 | 0 |

Table 4.15: Wan ranking. *Note: Wan has no overt complementizer*

| | | Hd>Cmp | *c | HdEdge | Spc>Cmp | HdCmp |
|---|---|---|---|---|---|---|
| C Aux V S O ☞ | AuxVSO | 0 | 0 | 0 | 1 | 1 |
| | AuxSVO | 0 | 0 | 1 | 0 | 1 |
| | SAuxVO | 0 | 0 | 2 | 0 | 0 |
| | AuxSOV | 1 | 0 | 0 | 0 | 1 |
| | SAuxOV | 1 | 0 | 1 | 0 | 0 |
| | SOVAux | 2 | 0 | 0 | 0 | 0 |
| | AuxVOS | 0 | 0 | 0 | 2 | 0 |
| Aux+c V S O | cAuxVSO | 0 | 1 | 0 | 1 | 1 |
| | cAuxSVO | 0 | 1 | 1 | 0 | 1 |
| ☞ | ScAuxVO | 0 | 0 | 2 | 0 | 0 |
| | cAuxSOV | 1 | 1 | 0 | 0 | 1 |
| | ScAuxOV | 1 | 0 | 1 | 0 | 0 |
| | SOVcAux | 2 | 1 | 0 | 0 | 0 |
| | cAuxVOS | 0 | 1 | 0 | 2 | 0 |
| C V+i S O ☞ | xVSO | 0 | 0 | 0 | 0 | 1 |
| | SxVO | 0 | 0 | 1 | 0 | 0 |
| | SOVx | 1 | 0 | 0 | 0 | 0 |
| | xVOS | 0 | 0 | 0 | 1 | 0 |
| V+i+c S O | cxVSO | 0 | 1 | 0 | 0 | 1 |
| ☞ | ScxVO | 0 | 0 | 1 | 0 | 0 |
| | SOVxc | 1 | 1 | 0 | 0 | 0 |
| | cxVOS | 0 | 1 | 0 | 1 | 0 |

Table 4.16: Predicted language 1: SVO/SAuxVO and VSO/AuxVSO

| Input | Candidate | *cEdge | HdEdge | Hd>Cmp | Spc>Cmp | HdComp |
|---|---|---|---|---|---|---|
| CAuxVSO ☞ | AuxVSO | 0 | 0 | 0 | 1 | 1 |
| | AuxSVO | 0 | 1 | 0 | 0 | 1 |
| | SAuxVO | 0 | 2 | 0 | 0 | 0 |
| | AuxSOV | 0 | 0 | 1 | 0 | 1 |
| | SAuxOV | 0 | 1 | 1 | 0 | 0 |
| | SOVAux | 0 | 0 | 2 | 0 | 0 |
| | AuxVOS | 0 | 0 | 0 | 2 | 0 |
| Aux+c VSO | cAuxVSO | 1 | 0 | 0 | 1 | 1 |
| | cAuxSVO | 1 | 1 | 0 | 0 | 1 |
| | ScAuxVO | 0 | 2 | 0 | 0 | 0 |
| | cAuxSOV | 1 | 0 | 1 | 0 | 1 |
| ☞ | ScAuxOV | 0 | 1 | 1 | 0 | 0 |
| | SOVcAux | 1 | 0 | 2 | 0 | 0 |
| | cAuxVOS | 1 | 0 | 0 | 2 | 0 |
| CAuxVSO ☞ | xVSO | 0 | 0 | 0 | 0 | 1 |
| | SxVO | 0 | 1 | 0 | 0 | 0 |
| | SOVx | 0 | 0 | 1 | 0 | 0 |
| | xVOS | 0 | 0 | 0 | 1 | 0 |
| V+i+c SO | cxVSO | 1 | 0 | 0 | 0 | 1 |
| ☞ | ScxVO | 0 | 1 | 0 | 0 | 0 |
| | SOVxc | 1 | 0 | 1 | 0 | 0 |
| | cxVOS | 1 | 0 | 0 | 1 | 0 |

Table 4.17: Predicted language 2:SVO/SAuxOV in the main and VSO/AuxVSO in the subordinate clause

| Input | Candidate | *cEdge | HdCmp | HdEdge | Hd>Cmp | Spc>Cmp |
|---|---|---|---|---|---|---|
| CAuxVSO | AuxVSO | 0 | 1 | 0 | 0 | 1 |
| | AuxSVO | 0 | 1 | 1 | 0 | 0 |
| | SAuxVO | 0 | 0 | 2 | 0 | 0 |
| | AuxSOV | 0 | 1 | 0 | 1 | 0 |
| | SAuxOV | 0 | 0 | 1 | 1 | 0 |
| | SOVAux | 0 | 0 | 0 | 2 | 0 |
| ☞ | AuxVOS | 0 | 0 | 0 | 0 | 2 |
| Aux+c VSO | cAuxVSO | 1 | 1 | 0 | 0 | 1 |
| | cAuxSVO | 1 | 1 | 1 | 0 | 0 |
| | ScAuxVO | 0 | 0 | 2 | 0 | 0 |
| | cAuxSOV | 1 | 1 | 0 | 1 | 0 |
| ☞ | ScAuxOV | 0 | 0 | 1 | 1 | 0 |
| | SOVcAux | 1 | 0 | 0 | 2 | 0 |
| | cAuxVOS | 1 | 0 | 0 | 0 | 2 |
| CAuxVSO | xVSO | 0 | 1 | 0 | 0 | 0 |
| | SxVO | 0 | 0 | 1 | 0 | 0 |
| | SOVx | 0 | 0 | 0 | 1 | 0 |
| ☞ | xVOS | 0 | 0 | 0 | 0 | 1 |
| V+i+c SO | cxVSO | 1 | 1 | 0 | 0 | 0 |
| ☞ | ScxVO | 0 | 0 | 1 | 0 | 0 |
| | SOVxc | 1 | 0 | 0 | 1 | 0 |
| | cxVOS | 1 | 0 | 0 | 0 | 1 |

Table 4.18: Predicted language 3: VOS/AuxVOS in the subordinate, SVO/SAuxOV in the main clause

| Input | Candidate | Hd>Cmp | *cEdge | HdCmp | HdEdge | Spc>Cmp |
|---|---|---|---|---|---|---|
| CAuxVSO | AuxVSO | 0 | 0 | 1 | 0 | 1 |
| | AuxSVO | 0 | 0 | 1 | 1 | 0 |
| | SAuxVO | 0 | 0 | 0 | 2 | 0 |
| | AuxSOV | 1 | 0 | 1 | 0 | 0 |
| | SAuxOV | 1 | 0 | 0 | 1 | 0 |
| | SOVAux | 2 | 0 | 0 | 0 | 0 |
| ☞ | AuxVOS | 0 | 0 | 0 | 0 | 2 |
| Aux+c VSO | cAuxVSO | 0 | 1 | 1 | 0 | 1 |
| | cAuxSVO | 0 | 1 | 1 | 1 | 0 |
| ☞ | ScAuxVO | 0 | 0 | 0 | 2 | 0 |
| | cAuxSOV | 1 | 1 | 1 | 0 | 0 |
| | ScAuxOV | 1 | 0 | 0 | 1 | 0 |
| | SOVcAux | 2 | 1 | 0 | 0 | 0 |
| | cAuxVOS | 0 | 1 | 0 | 0 | 2 |
| CAuxVSO | xVSO | 0 | 0 | 1 | 0 | 0 |
| | SxVO | 0 | 0 | 0 | 1 | 0 |
| | SOVx | 1 | 0 | 0 | 0 | 0 |
| ☞ | xVOS | 0 | 0 | 0 | 0 | 1 |
| V+i+c SO | cxVSO | 0 | 1 | 1 | 0 | 0 |
| ☞ | ScxVO | 0 | 0 | 0 | 1 | 0 |
| | SOVxc | 1 | 1 | 0 | 0 | 0 |
| | cxVOS | 0 | 1 | 0 | 0 | 1 |

Table 4.19: Predicted language 4 (Miya: VOS/AuxVOS in the subordinate, SVO/SAuxVO in the main clause

classifies it as SVO and VOS in the main clause, but VOS in the subordinate [Schuh, 1998]. In personal communication, he states that AuxVOS order is required for nominal subjects in most types of subordinate clauses [14]. However, in elicited main clauses, speakers generally would give SAuxVO, which is also a common order in texts, although they also accept AuxVOS. While this data is not conclusive, it is definitely an indication that the language in Tableau 4.19 might actually be attested.

The results presented in this chapter indicate that a multi-headed version of dependency grammar coupled with an OT approach to linearization can provide an account of cross-linguistic word order in neutral declarative constructions. It is important to note that an OT linearization without the multi-headedness assumption would not be as successful. The current analysis crucially relies on the fact that Spc>Cmp affects not only S > O, but also S > V in complex-tense clauses. The typology without this assumption is presented in Tableau 4.20. Among other things, it fails to capture the Welsh variety of VSO languages

| sub. compl. | main compl. | sub. simp. | main simp. | |
|---|---|---|---|---|
| C Aux V S O | Aux+c V S O | C V S O | V+i+v S O | Language |
| AuxVSO | cAuxVSO | xVSO | cxVSO | Niuean |
| AuxVSO | ScAuxVO | xVSO | ScxVO | Pred. 1 |
| AuxVSO | ScAuxOV | xVSO | ScxVO | Pred. 2 |
| SAuxVO | ScAuxVO | SxVO | ScxVO | English |
| SOVAux | ScAuxOV | SOVx | ScxVO | German |
| SOVAux | SOVcAux | SOVx | SOVxc | Latin |
| AuxVOS | ScAuxVO | xVOS | ScxVO | Miya |
| AuxVOS | ScAuxOV | xVOS | ScxVO | Pred. 3 |
| AuxVOS | cAuxVOS | xVOS | cxVOS | Tsou |

Table 4.20: Typology predicted by the same analysis without subject multi-headedness

[14] although complement clauses can also use SAuxVO, I will assume this is not true optionality, but rather some weak topicality effect

(VSO/AuxSVO), which are well-attested and documented in Celtic linguistics and beyond. One question left open by my analysis is whether and how cross-linguistic frequency can be accounted for. The next chapter is devoted to some tentative proposals in this area.

# Chapter 5

# Cross-linguistic frequency of word order types

## 5.1 The role of frequency facts in linguistic theory

One of the most important theoretical question in the study of word order typology is how to interpret fact about cross-linguistic frequency in relation to grammatical theory. One view is that frequency is simply a historical accident, unrelated to universal grammar. According to this view, the existence of even a single language with some property $P$ provides as much evidence about that property as would a thousand languages.

There are two serious problems with this position:

1. From an empirical standpoint, it supports the highly unlikely assumption that there is no experimental error in word order "measurements." One language in a sample of one thousand might be a noisy aberration, especially if it is spoken by a small community of people and investigated by linguists of non-native proficiency. In other words, it is possible for a linguist every once in a while to register a wrong answer because of asking the wrong person, or the wrong question. Some indirect evidence of experimental error is already present in the literature, since different researchers often propose different base orders for the same language.

2. From a theoretical standpoint, it assumes that the language faculty licenses linguistic

properties in an all-or-none binary fashion. This is not necessarily the case. It is very possible that the language faculty favors or discourages some properties over others through a system of soft preferences. Under this interpretation, any language is in principle possible, but some are more unlikely than others.

In fact, some version of the latter idea plays a role in Chomsky's notion of an evaluation metric for grammars according to simplicity. Obviously, simple grammars require less evidence to be learned, while more complex grammars require more evidence. In the end, given the right amount of evidence any grammar could be learned, but the probability of receiving sufficient evidence is inversely related to the size of that evidence. Hence, grammars which require more evidence are less stable than those that do not. The same argument applies here: if a language violates many soft preferences, a lot of evidence is required to overcome the negative bias against it. On the other hand, if it complies with the soft preferences, very little evidence is needed.

Where would such preferences come from? I believe this issue is very important for cognitive science because its answer relates to the big questions of modularity, and evolutionary history in the human brain. For example, it is possible that a system of arbitrary soft preferences evolved under the pressure of communicative uniformity. In other words, children born without the soft preferences had difficulties learning to communicate with their peers, which made them less successful in their hunting practice, and prone to accidents which they would have avoided, if only they were able to understand the advice of their tribe. Another — in my opinion more likely, possibility is that these preferences are the result of design constraints on the physical and cognitive capabilities of humans.

Ultimately, each researcher must decide for him/herself the relevance of frequency data to the theory. Performing some type of statistical significance test may provide some clues about experimental errors in studies on word order variation. However, in the final analysis, the theoretical implications of relative frequency depend on the theoretical framework. For example, Tomlin links relative frequencies to functional principles [Tomlin, 1986]. He argues that the more principles are satisfied by a given order, the greater its frequency. Each of the possible subject, verb, and object permutations are scored according to three principles. It emerges that SOV and SVO languages satisfy all three, VSO satisfies two, VOS and OVS satisfy only one, while OSV satisfies none. This mirrors closely the relative frequency of orders in Tomlin's, and subsequently Dryer's data

Optimality theorists are split on the extent to which analyses are informed by cross-linguistic frequency. While cross-linguistic frequency has been considered one of the criteria for markedness [Croft, 1990], some researchers point out that "non-linguistic factors, including politics, famine and natural disasters, have no doubt skewed the sample of languages and, thus, the types of [constructions] found in those languages" [Hume, 2003]. At the same time arguments from frequency are often made in support of certain constraint. For example, [Zepter, 2003] motivates her *BranchingRight* constraint with the higher frequency of SVO and SOV over other existing orders.

## 5.2 The number-of-rankings hypothesis

While there are several ways of accounting for the ratio of alternate forms within a language, some can be extended to cross-linguistic analysis. One recent proposal is that relative frequency in cases of free variation reflects the number of rankings consistent with any given winner [Anttila, 1997]. If the same idea is applied to cross-linguistic variation, we would expect that frequent forms are winners under many rankings, while rare forms are winners under one or few rankings. For example, the frequent SOV word order should correspond to more rankings than VSO. This proposal has a theoretically stronger and weaker version. The strong version demands that the ratio of rankings for any two potential forms mimic the real-world distribution. If the real-world ratio of SOV to VSO is approximately $4:1$, the analysis should contain four times as many rankings for SVO than for VSO. The weak version demands simply that the number of SOV rankings exceed that of VSO rankings. Obviously, the strong version would also imply a higher minimal number of constraints than the weak version.

Could an extension of Antilla's proposal account for the cross-linguistic frequency of word order? The answer may depend on which version of the proposal we decide to adopt. At a minimum, our data set must satisfy the weak version, since any dataset that satisfies the strong version must also satisfy the weak version.

Thus, the first step is to check whether the number of rankings for each word order corresponds to its position in the frequency table. To do so, we will determine the number of total orders (linear extensions) that corresponds to the partial order of constraints required to generate each language type predicted in Chapter 4: HdCmp (= HComp), Hd>Cmp (= Head>Com) and Spc > Cmp (= Spec>Com), and HdEdge (= HEdge). The definition of each constraint is reviewed below:

1. **HdCmp**(= HComp): For all $x$ such that $Head > x > Compl$, $x$ is either null or a proper subordinate of Compl.

   = Heads may not be separated from their complements, except by proper complement subordinates.

   X is a proper subordinate of Y iff X is a subordinate of Y and X does not depend on sisters or superordinates of Y.

2. **Spc>Cmp**(= Spec>Com): Specifiers precede complements.

3. **Hd>Cmp**(= Head>Com): Heads precede complements.

4. **HdEdge**(= HEdge): The dependents must align on the same side of the head.

The unique partial order for each language derived in Chapter 4 is represented in a Hasse diagram (e.g. Figure 5.8; see all diagrams - Figures 5.8 to 5.20, at the end of this chapter). An arrow from X to Y means that X is ordered (i.e. ranked) higher than Y. Since transitivity of domination is assumed, arrows which follow from transitivity are omitted.

Computing the number of linear extensions of a poset is a #P-complete problem [G.Brightwell and P.Winkler, 1991]. For the purposes of this work, I used the Combinatorial Object Server [Ruskey, 2003]. The program computes the number of extensions in an exhaustive manner with a few shortcuts. If the width of the poset is small then a dynamic programming approach is used to compute the number of extensions more quickly.

The possible rankings for each language are given in Table 5.1 in order of frequency. Unfortunately, we have no direct frequency data for each language-type. Instead, what we have are basic word order frequencies from [Dryer, 2005a] (see Chapter 4); for convenience, the data are repeated here (Table 5.2).

| Main | Main+Aux | Sub | Sub+Aux | Language | Rankings |
|---|---|---|---|---|---|
| SxVO | SAuxVO | ScxVO | ScAuxVO | English | 30 |
| SOVx | SOVAux | SOVxc | SOVcAux | Latin | 18 |
| xVOS | AuxVOS | cxVOS | cAuxVOS | Tsou | 18 |
| SOVx | SOVAux | ScxVO | ScAuxOV | German | 12 |
| xVOS | AuxVOS | ScxVO | ScAuxVO | Miya | 8 |
| xVSO | AuxSOV | cxVSO | cAuxSOV | S. Gaelic | 7 |
| xVSO | AuxVSO | cxVSO | cAuxVSO | Niuean | 7 |
| SxVO | SAuxVO | cxVSO | cAuxSVO | Breton | 6 |
| xVSO | AuxSVO | cxVSO | cAuxSVO | Welsh | 4 |
| xVOS | AuxVOS | ScxVO | ScAuxOV | Pred. 3 | 4 |
| SxVO | SAuxOV | cxVSO | cAuxSOV | Dinka | 3 |
| xVSO | AuxVSO | ScxVO | ScAuxVO | Pred. 1 | 2 |
| xVSO | AuxVSO | ScxVO | ScAuxOV | Pred. 2 | 1 |

Table 5.1: Number of winner rankings for each word order predicted in Chapter 4

Thus, in order to evaluate the number-of-rankings hypothesis, we must collapse the word orders of Table 5.1 into basic word order types. Obviously, there is no single clear way of doing this.

One possibility is to consider only the order of simple main clauses as basic, giving rise to the grouping in Table 5.3. The predicted frequencies for basic word order obtained by summing over simple main clause orders do not correspond to the actual frequencies found in Dryer's study. SVO languages are predicted to be more frequent than all other types of languages.SOV, VSO and VOS languages come out as equally frequent. In reality, SOV is at least as frequent as SVO, and VSO is less frequent that SVO and more frequent than VOS.Thus, the number-of-ranking hypothesis has so far failed to account for cross-linguistic frequencies. This failure may be due either to the incorrectness of the hypothesis itself, or to the linking hypothesis which allows us to derive basic word order groupings.

| Order | Number | Percent | Lower CI 95% | Higher CI 95% |
|---|---|---|---|---|
| SOV | 497 | 47 | 44.1 | 50.1 |
| SVO | 435 | 41 | 38.2 | 44.1 |
| VSO | 85 | 8 | 6.5 | 9.8 |
| VOS | 26 | 2 | 1.6 | 3.6 |
| OVS | 9 | .9 | 0.4 | 1.6 |
| OSV | 4 | .4 | 0.1 | 1.0 |

Table 5.2: Frequency of base word orders [Dryer, 2005a]

A possible alternative grouping is by simple subordinate order. It results in Table 5.4. Unfortunately, this grouping does not correspond to the correct relative frequencies either. There are two remaining groupings to investigate - complex main and subordinate clauses. When grouping according to subordinate clauses, the auxiliary is ignored and the order of subject, object and verb is taken to be the basic word order. The grouping by complex main clauses is presented in Table 5.5. There are a number of problems with this grouping. Most prominently, the Celtic languages-like Welsh, are classified as SVO, while from the point of view of earlier word order research they are the prototypical VSO languages (e.g. [Tomlin, 1986],[Dryer, 2005a]). Furthermore, even if this unusual classification is granted, the predicted frequencies are incorrect. VOS languages appear to be more frequent than the VSO languages while the opposite is true. The last straightforward grouping - by complex subordinate clauses, is presented in Table 5.6. It is ridden by the same problematic premises and outcomes as the previous grouping. The Celtic languages are once again unusually classified and the VOS languages are more frequent than the VSO languages, contrary to fact. Thus, by any straightforward linking hypothesis, the number-of-rankings proposal fails to account for frequency data. It is possible that some type of non-trivial grouping may improve the chances of the proposal. However, due to the imprecise nature of word order typological research, it is impossible to know for sure which grouping would be correct. If we allow ourselves to explore virtually any conceivable grouping, we may or may not be able to arrive at a better result, but the result itself would be compromised by the number of attempts we have made at obtaining it. Thus, I will conclude that the proposal is simply inadequate.

There is another reason why cross-linguistic frequency is unlikely to correspond simply to the number of rankings. Recall the low-frequency orders which were excluded from our analysis. Now let us suppose that their documented existence is not an experimental error but a fact. Since no ranking derives them, the number-of-rankings hypothesis is at best moot on their existence. I will now turn to a different type of explanation of frequency data. Instead of deriving frequencies from properties of the OT analysis, I will view the suggested OT constraints as the components of a learning bias. The influence of this bias on language transfer from generation to generation results in frequency distributions not unlike the one we observe today.

## 5.3 The learning bias hypothesis

It is an idealizing assumption in theoretical linguistics that children learn language instantaneously and perfectly. However, at least one view of language change suggests that it is the result of cross-generational mis-learning. According to this view, language learning is an iterative process which effects changes the characteristics of the original input [Kirby, 2001]. At each iteration, the learner (or generation of learners) maps linguistic input (data) to a grammar (hypothesis). The input for the next iteration is generated from the grammar obtained in the current iteration (Figure 5.1). The languages which survive the

$$
\begin{tikzcd}
 & LEARNER \arrow[dr] & & & LEARNER & \\
DATA \arrow[ur] & & HYPOTHESIS \arrow[r] & DATA \arrow[ur] & & \cdots
\end{tikzcd}
$$

Figure 5.1: Kirby's iterative learning scheme

iterative process are the ones which are fit enough to be transmitted through the information bottleneck that accompanies this process. The question becomes what properties make a language capable of surviving over many generations.

One way of interpreting the cross-linguistic frequency of word order is as a measure of the relative fitness of each variant. If this interpretation is correct, explaining the observed distribution requires us to identify the source of linguistic fitness. Early work on iterative learning (e.g. [Kirby, 2001]) attributed fitness to the information bottleneck created by the mode of transmission. Thus, the emergence of linguistic universals was credited to the *process* of transmission, rather than to the learning algorithm used by the learners. In fact, it was shown that a particular aspect of human language *compositionality* was "selected for" under different circumstances, i.e. when learners used different learning algorithms [Brighton, 2002], [Smith, 2003]. Recently, Griffiths and Kalish [Griffiths and Kalish, 2005] examined iterative learning from a Bayesian perspective. They suggested a different interpretation of previous results, arguing instead that all above mentioned works implicitly define a prior distribution over hypotheses that favor compositional languages. Since a variety of learning algorithms can be formulated as Bayesian inference, iterative learning can be analyzed as the interaction of rational Bayesian agent equipped with a hypothesis space and a prior over it. The hypothesis space is the space of possible grammars. The prior is a learning bias which influences the mapping of data to hypothesis. It turns out that the biases fully determine the types of languages that would be stable over many generations of learners. In particular, the authors showed that the distribution of languages in the limit is guaranteed to converge to the prior. In their words:

> ... the asymptotic probability that a language is used does not depend at all upon the properties of the (original-n.a.) language, being determined entirely by the assumptions of the learner.

A central problem for Bayesian iterative learning is what constitutes the source of the prior. I propose that the prior is linked to the universal constraints identified in the cross-linguistic analysis of word order, i.e. it is assigned to each language according to the constraints it violates. This can be done in different ways depending on the linking assumptions that will be discussed later. However, the main idea is simple: languages that satisfy more constraints start off with a higher prior. This proposal effectively transforms the predictions of the OT analysis in Chapter 4 from categorical to probabilistic. Rather than completely ruling out suboptimal languages, constraints in the form of learning biases would simply stack the odds against them in iterative learning. Thus, we would expect to find some, albeit few, examples of object-first languages (OVS and OSV).

In Griffiths and Kalish's model, iterative learning is described as a stochastic process over three variables - meaning, form and a mapping hypothesis. Thus, acquiring a language is conceptualized as a function learning task, where the learner must find the most plausible way of mapping inputs to outputs based on a sample of observed meaning-form pairs. They assume that each learner is a rational Bayesian agent equipped with a finite, discrete [1] hypothesis space $H$ and a prior probability distribution $p(h)$ for each hypothesis $h \in H$. Each hypothesis specifies a conditional probability distribution over the set of forms given the set of meanings, or $p(f|m,h)$. For example, if the hypothesis is SOV, the expected probability distribution over focus-free declaratives peaks over the SOV order, but some probability mass is allocated to other orders as well, to account for the expected noise in production and comprehension. It is possible for the speaker to select an incorrect order, or for the learner to classify the meaning incorrectly. Each form $f$ is considered independent given the meaning and hypothesis.

Figure 5.2 represents the graphical model corresponding to this setup. A graphical model is a representation of the dependencies among the variables in a probabilistic model. Each node denotes a variable, and each arch - a dependency. As seen in the diagram,

$$\begin{tikzcd}
& M_k \arrow[d] \arrow[dr] & & M_{k+1} \arrow[d] \arrow[dr] & \\
H_k \arrow[r] & F_k \arrow[r] & H_{k+1} \arrow[r] & F \arrow[r] & H_{k+2}
\end{tikzcd}$$

Figure 5.2: Griffiths and Kalish's Model: Graphical model of iterated Bayesian learning

the k-generation leaner observes the set of meaning-form pairs produced by the $k-1$-th predecessor and computes the posterior over $H$ via the Bayes rule (eq. 5.1)

$$p(h_{k+1}|m_k, f_k) = \frac{p(f_k|m_k, h_{k+1})p(h_{k+1})}{p(f_k|m_k)} \tag{5.1}$$

[1] the assumption of discreteness is not indispensable, but I will maintain it since it is well-suited to our purposes. Obviously, the set of possible grammars given a finite set of universal constraints is finite.

where

$$p(f_k|m_k) = \sum_{h \in H} p(f_k|m_k, h)p(h) \tag{5.2}$$

In the production step, a vector of meanings **m** is generated by a fully independent distribution $p(m)$, which represents events in the world, or, perhaps more accurately, perceived events in the world. Given a meaning-utterance pair at time $k$, the learner samples a hypothesis from $p(h_{k+1}|m_k, f_k)$, and generates an utterance vector **f** according to $p(f_k|m_k, h_{k+1})p(h_{k+1})$.

The model is equivalent to a Markov chain with state space $H$ and transition matrix from $h_k$ to $h_{k+1}$ obtained by summing over the data variables $m$,$f$(eq. 5.4)

$$T(h_k, h_{k+1}) = p(h_{k+1}|h_k) \tag{5.3}$$

$$p(h_{k+1}|h_k) = \sum_m \sum_f p(h_{k+1}|m, f) \times p(f|m, h_k) \times p(m) \tag{5.4}$$

Provided that the transition matrix $T$ is regular [2], it is guaranteed to converge to a stationary distribution, according to the stochastic matrix theorem.

**Stochastic Matrix Theorem:** If A is a regular stochastic matrix, then A has a steady-state vector t so that if $x_o$ is any initial state and $x_{k+1} = A \times k$ for $k = 0, 1, 2, .....$ then the Markov chain $\{x_k\}$ converges to $t$ as $k \to$ infinity.

This is true because any such matrix will have an eigenvalue equal to one and all other eigenvalues, including the second eigenvalue, will be between zero and one. Since the setup requires non-zero transition probabilities, the regularity (or ergodicity cf. [Norris, 1997]) condition is satisfied. Thus, we expect the stochastic process to come to a stationary probability distribution. As Griffiths and Kalish show, the stationary distribution for Bayesian iterative learning is exactly the prior distribution (eq. 5.5).

$$\begin{aligned} p(h_{k+1}) &= \sum_{h_k \in H} \sum_m \sum_f p(h_{k+1}|m, f)p(f|m, h_k)p(m) \times prior(h_k) \\ &= \sum_m \sum_f p(h_{k+1}|m, f) \times \sum_{h_k \in H} p(f|m, h_k)prior(h_k) \times p(m) \\ &= \sum_m \sum_f \frac{p(f|m, h_{k+1})p(h_{k+1})}{p(f|m)} p(f|m)p(m) \end{aligned} \tag{5.5}$$

[2] A stochastic matrix $P$ is regular if for some matrix power $k$, $P^k$ contains only strictly positive entries

$$= p(h_{k+1}) \times \sum_m \sum_f p(f|m, h_{k+1})p(m)$$

$$\sum_m \sum_f p(f|m, h_{k+1})p(m) = 1$$

Let us consider the possible implications of this result for cross-linguistic frequency of word order. First, examine the distribution on $m$ in the context of word order. Word order is largely independent of meaning, with the important exception of discourse features. Thus, we could treat $m$ as a binary variable, where one value corresponds to discourse-neutral meaning representations, and another to some type of discourse-marking. Now we have two choices: a) we can assume that the learner is able to classify most utterances correctly by meaning category, and uses only discourse-neutral trials for learning, or b)that the learner assumes that all utterances are discourse neutral, but the assumption is harmless because an overwhelming majority of utterances are in fact discourse-neutral. Either way, we can eliminate the variable $m$ from the equation, by assuming that $m$ is always a neutral context proposition. Thus, equation (5.4), can be restated as (5.7)

$$T(h_k, h_{k+1}) = p(h_{k+1}|h_k) \tag{5.6}$$

$$p(h_{k+1}|h_k) = \sum_f p(h_{k+1}|f) \times p(f|h_k) \tag{5.7}$$

The current distribution of word orders is the outcome of the iterative learning process at some iteration $t$. We know that the process is bound to converge to a stationary distribution at some iteration $t_s$. The crucial question is whether $t$ is before or after $t_s$. If $t$ is after $t_s$, the process has already converged. Hence, the distribution we observe today would be a precise reflection of the learning bias. If, on the other hand, $t$ precedes $t_c$, we would expect the distribution to reflect the learning bias alongside other variables influencing the convergence rate.

An obvious way to address this issue is to ask: When did humans first begin speaking fully developed languages? Obviously, speaking a fully developed language leaves little and at best, indirect physical evidence. While we know that writing began about 6,000 years ago, estimates of when humans began to speak vary widely - from 9,000 to 1-3 million years ago. Among the most convincing estimates is that of [Krantz, 1980], who argues on the basis of the fossil record that language emerged 50,000 years ago. He ascribes the rapid technological advancement accompanied by population expansion to the emergence of full language and a new cognitive competence in humans. If Krantz is right, at 4-5 generations per century, the estimated number of iterations in human language evolution is somewhere between 2,000 and 2,500.

Whether this number of iterations is sufficient for convergence depends on the transition matrix [3]. In particular, it is determined by those eigenvalues of the matrix $\lambda_i < 1$. The smaller the values are, the quicker the convergence. Griffiths and Kalish investigated the behavior of the second eigenvalue of the matrix for a binomial hypothesis distribution and found it to depend on the size of the learning sample, the non-uniformity factor of the prior distribution and the "error" factor in productions. In particular, the bigger the size of the learning sample, the less uniform the prior, and the smaller the error rate, the more difficult it is for the process to converge. It remains for these results to be generalized to multinomial distributions.

Given the multitude of assumptions necessary to establish both the evolutionary age of human language, and the convergence rate of the process, it is unlikely that the relationship of the current evolutionary iteration $t$ to the stationary $t_c$ would ever be conclusively established. Despite the many uncertainties involved, let us suppose for a moment that convergence has already occurred (i.e. that $t > t_c$). If so, the relative frequency of basic word orders must be perfectly informative of the learning bias. But, according to our hypothesis, so are the linearization constraints. Hence, constraints violated by more frequent word orders must matter less for the formation of the learning bias than constraints violated by rarer orders.

One way of relating the constraints to the current distribution (i.e. the hypothetical prior) is through statistical regression analysis. The type of regression suitable for our data is multiple logistic regression. To perform logistic regression, we assume that the independent variable "Percent languages with this violation profile"[4] is binomially distributed and dependent on the categorial variables Hd>Cmp, HdEdge, HdCmp, Spc>Cmp. Logistic

[3] the influence of initial conditions is for the most part, negligible

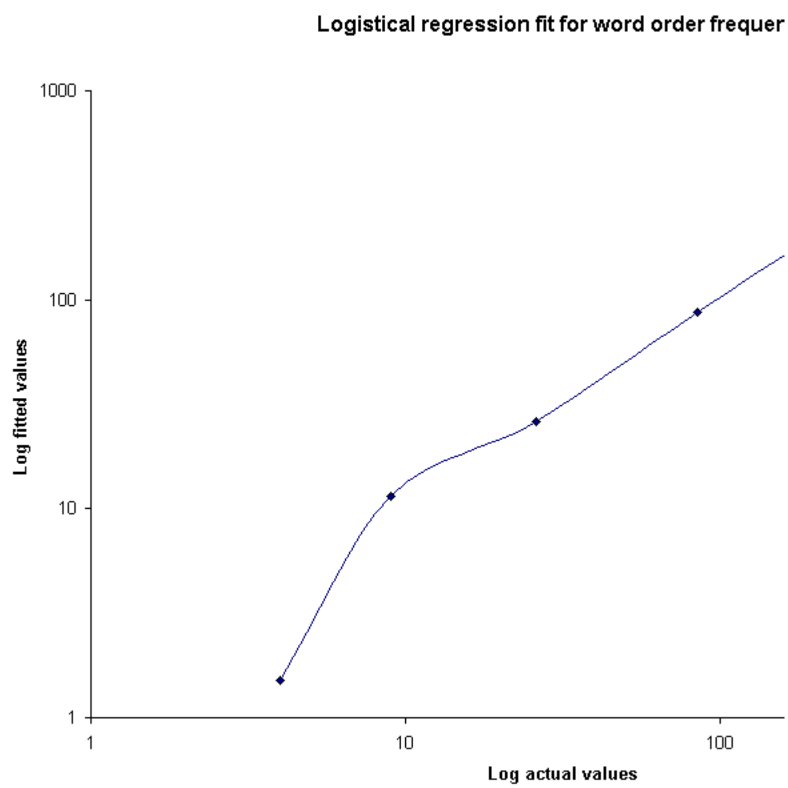
[4] Here by violation profile of a language I mean the violation profile of a basic word order sentence in that language.

regression fits a curve of the form:

$$\frac{e^{\alpha+\sum_i \beta_i x_i}}{1+e^{\alpha+\sum_i \beta_i x_i}}, \tag{5.8}$$

where each $x_i$ is a variable, $\beta_i$ is a multiplication coefficient for that variable, and $\alpha$ is the intercept.

Logistic regression provides a close fit for the word order data at hand. Figure 5.3 is a plot of the fitted against the actual values. The vector of coefficients $\beta$ and the constant term $\alpha$ are given in Table 5.8. The $\beta$ coefficients in logistic regression are interpreted as




an indicator of the odds ratio (R). In precise terms:

$$R_i = e^{\beta_i}$$

The odds ratio is a measure of the probability of "success" given that the variable in question is "on". It is equal to the ratio of the fraction of successful outcomes when the variable is "on," to the fraction of successful outcomes when it is "off." Thus, if a language violates Spc>Cmp, its odds ratio is $R = e^{-3.8529}$, or $1 : 50$, while if it violates Hd>Cmp, its chances are much better $R = e^{-}0.2912$, or $75 : 100$. Thus, the $\beta$ coefficients rank the constraints in order of importance (from least to most): Hd>Cmp $<<$ HdEdge $<<$ HdCmp $<<$ Spc>Cmp.

One disadvantage of accepting this composition of the prior is that there is no independent reason for it. It does not reflect any property of the constraints and is motivated solely by our desire to link it to the actual linguistic frequencies. An alternative is to assume a linear dependency between the prior and the number of inviolated constraints. Since SOV, SVO, VSO and VOS violates only one constraint, they share the same high prior. OVS and OSV on the other hand violate three constraints. To maintain the exact linear relationship, the high prior must be exactly four times the low prior.

$$\begin{aligned} 4 * high + 2 * low &= 100\% \\ low &\approx 6\% \\ high &\approx 22\% \end{aligned} \tag{5.9}$$

As the calculations in equation (5.9) show, this translates into a high prior of $22\%$ and a low prior of $6\%$.

Clearly, this prior is not equal to the current distribution, as we would expect under the assumption that $t > t_c$. But what if $t_c > t$? Is it possible to obtain the current distribution from this prior?

Suppose a discrete finite hypothesis space of word orders, where each hypothesis corresponds to a particular basic word order and defines an expected distribution of word orders in production. If the hypothesis is SOV, the learner expects the majority of sentences to be SOV, with the rest distributed among the non-SOV types. If the remaining probability mass is uniformly distributed among the non-basic orders, the hypothesis space is parameterized by a single parameter. The basic word order is hypothesized with probability $p$, and each non-basic order is hypothesized as $\frac{1-p}{5}$. Obviously, the probability of the basic word order must be greater than that of the non-basic orders (eq. 5.10). Hence, $p$ belongs to the interval

$(\frac{1}{6}; 1)$

$$p > \frac{1-p}{5} \qquad (5.10)$$
$$\frac{1}{6} > p < 1$$

An example of such a hypothesis space is given in Table 5.9. We can now superimpose the constraint-based prior from equation (5.9) on the hypothesis space, and define two other simulation parameters- size of sample *s* and the initial distribution.

Regardless of the parameters the process will eventually converge to the prior. However, it is important to qualitatively assess the behavior of the system before convergence. For this purpose, let us examine the behavior of the process from two very skewed initial distributions and uniform distribution. One initial distribution is heavily skewed toward a low prior language (LOW), the other toward a high prior language (HIGH). Obviously, the uniform distribution asigns equal initial probabilities for all orders. Figures 5.3 and 5.4 shows the evolution of the posterior over two hundred iterations.

We can see that in all cases, while the general trend of high versus low prior languages is reflected in the data relatively early in the iterative process, there is considerable noise among languages with equal prior. In fact, the plots at many pre-convergence iterations is not that different from the word order distribution we observe in the world today. For example, compare Figure 5.5 to Figure 5.6. Figure 5.5 represents the distribution of word orders on the hundred-and-fiftieth iteration of the learning process starting from the LOW skewed distribution. Figure 5.6 shows the word order distribution currently observed in the world. Tn the simulation, as well as the real data, the high-prior languages do not correspond to identical proportions, but are uniformly higher than the low prior languages. Thus, it is possible that the frequency of word orders today reflects a Bayesian iterative learning process pre-convergence, with prior related to constraints.

Of course, it is fair to ask how typical is the distribution shown here of the evolutionary process. The answer depends on what characteristics of the distribution we consider relevant. The chance of an exact match occuring is obviously rare. However, the chance that high prior languages outnumber low prior languages over the course of the process is high. Figure 5.7 shows, for a given number of iterations, percent trials with outcomes such that

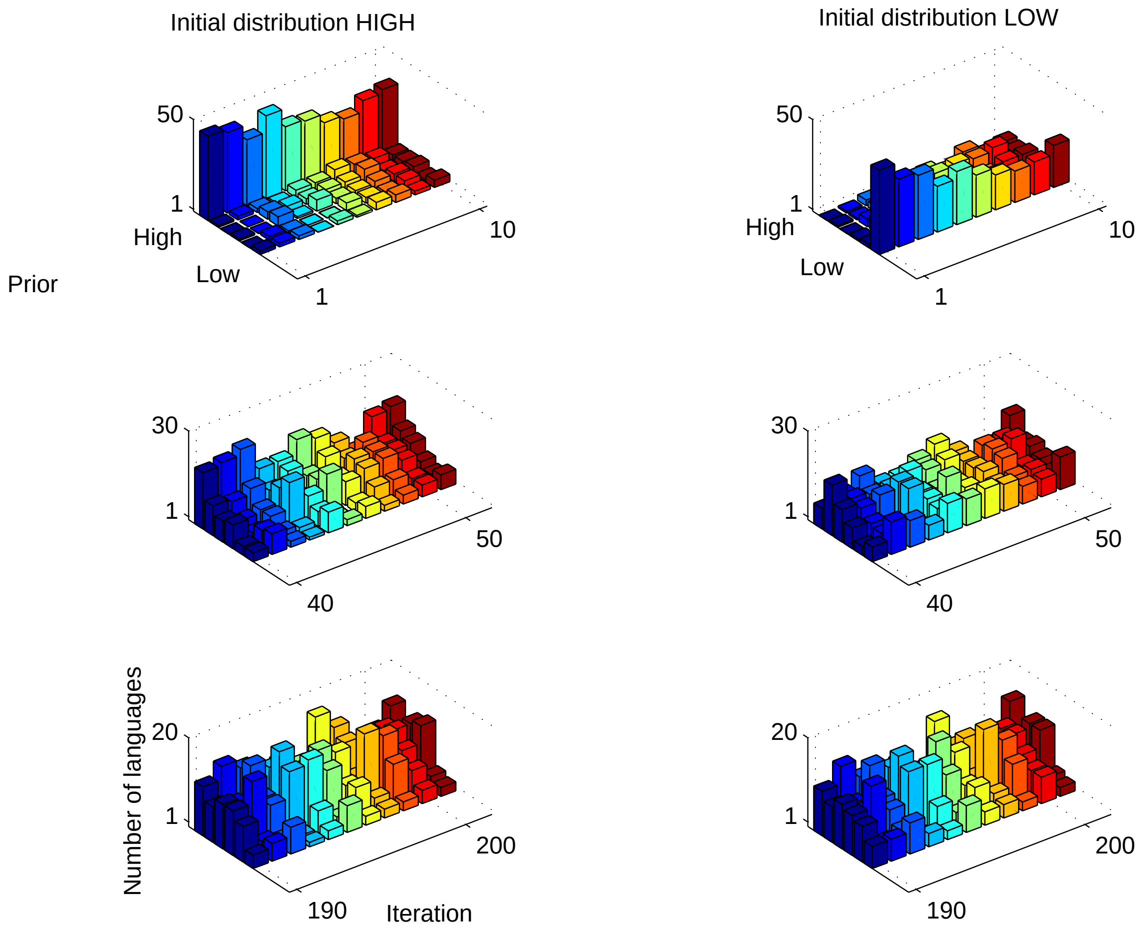



Figure 5.3: Evolution of the posterior over 200 iterations: HIGH and LOW initial distibution

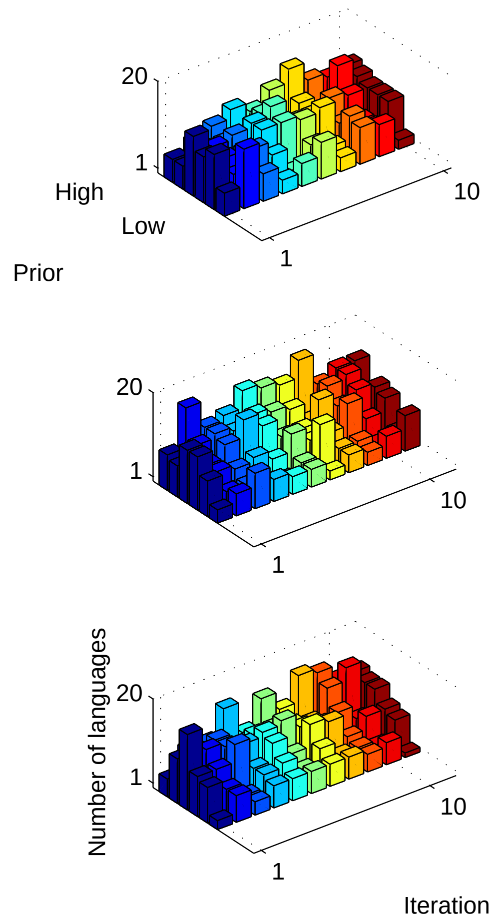



Figure 5.4: Evolution of the posterior over 200 iterations: uniform initial distribution

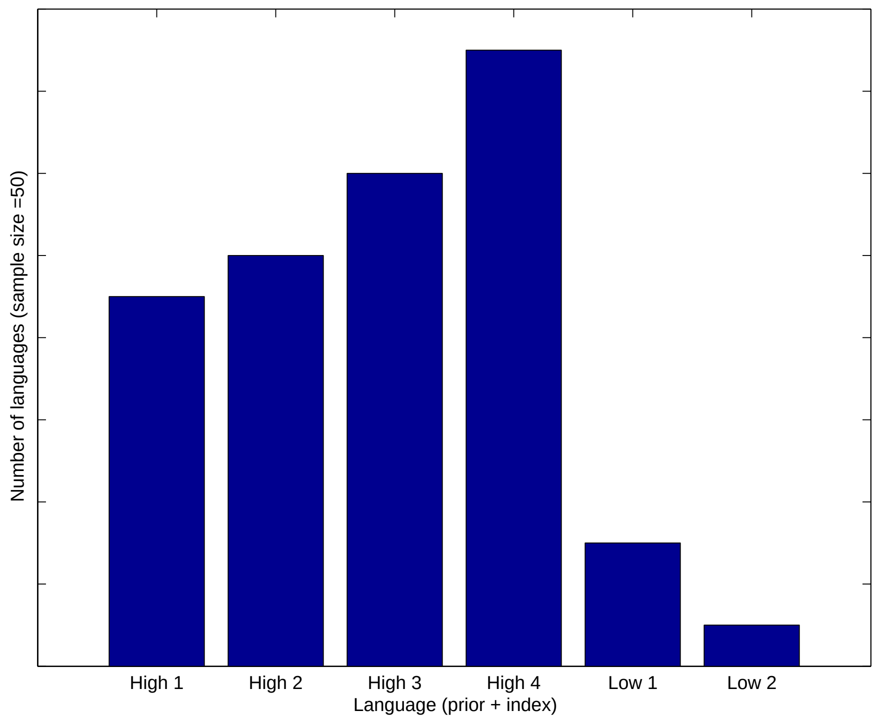



Figure 5.5: Distribution of word orders on the 150th iteration (LOW skewed distribution)

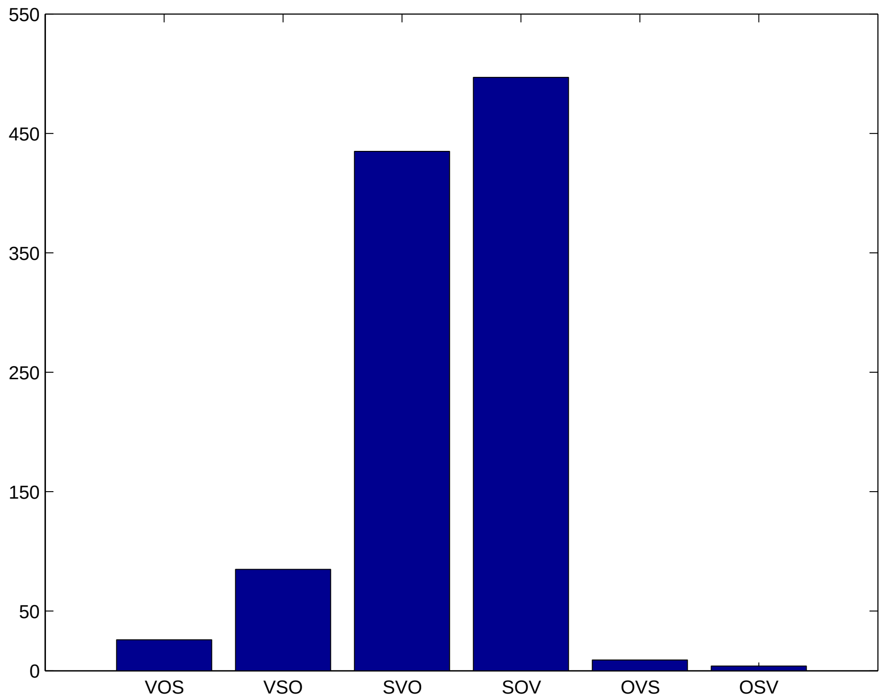



Figure 5.6: Actual current word order distribution

the rarest of the high prior languages is more frequent than the most common of the low prior languages, i.e. in which $\{SOV, SVO, VSO, VOS\} > \{OVS, OSV\}$.

## 5.4 Alternative hypothesis spaces

It is possible to generalize the experimental setup discussed here to a hierarchically structured hypothesis space, which contains many sub-hypotheses for each word order. While the expected distribution under each sub-hypothesis would differ, the probability of basic word order sentences would always be higher than that of any other single sentence type. Obviously, such hypothesis space would correspond to a hierarchical prior. It is also possible to include hypotheses in which the probability mass is not uniformly distributed among the non-basic orders. In this case, the hypothesis space is parameterized by two (vector) parameters. Let $\alpha$ be a single-element vector denoting the probability of the basic order, and $\beta$ - a five element vector of probabilities of the non-basic orders. Then the following holds:

$$\alpha + \sum_i \beta_i = 1 \tag{5.11}$$

$$\alpha = 1 - \sum_i \beta_i \tag{5.12}$$

Furthermore,for all elements $i \in \beta$:

$$\alpha > \beta_i \tag{5.13}$$

An example of a hierarchically structured hypothesis space is given in Table 5.10. The results would not be qualitatively changed by this more realistic extension, since they would be obtained simply by averaging sub-hypotheses over hypotheses.

In general, the learning bias envisioned here is quite crude, because it depends only on the number of observed constraints for each language type. An alternative model would rely on the relative order of constraints in a ranking to determine the bias of the corresponding language. In particular, we could assign a cost to each pairwise comparison of constraint ranks, and determine the bias as a function of these costs. For example, suppose that the bias is penalized by $2$ points for each constraint ranked higher than $Spc > Cmp$, but only $1$ point for each constraint ranked higher than $Hd > Cmp$. All else being equal,

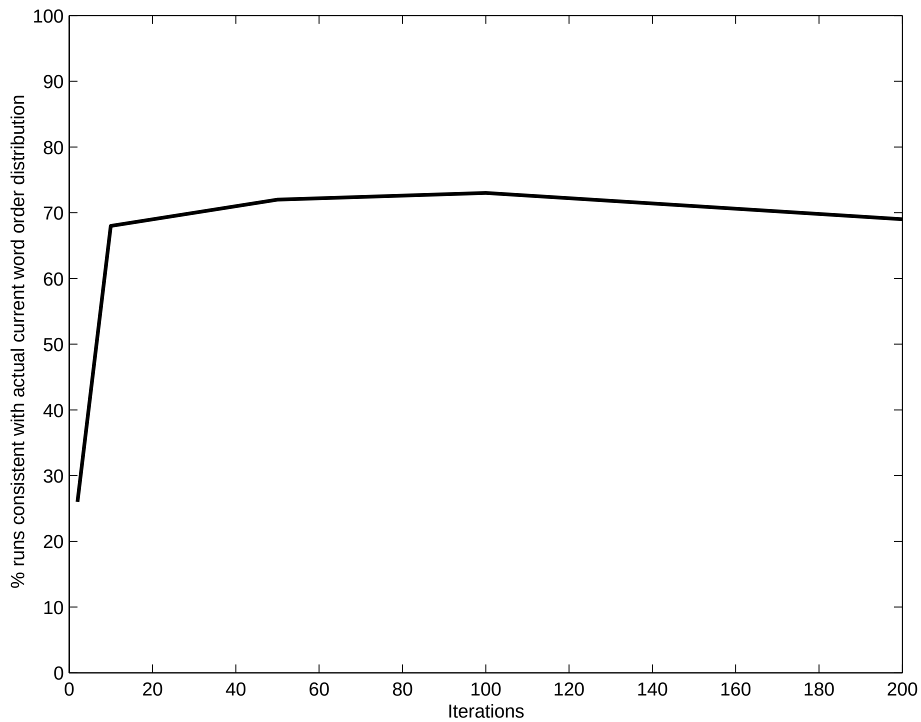



Figure 5.7: Percent trials consistent with current word order distribution plotted against iteration (100 trials per iteration, uniformly initialized)

the bias for pure VOS language type (which ranks $Spc > Cmp$ low) will be lower than the bias for pure SOV language type (which ranks $Hd > Cmp$ low).

There are a couple of reasons why I decided against pursuing such a strategy at this time. First, unlike the simple bias assignment, this type of bias assignment would only apply to languages that are possible winners under some ranking. Consequently, it would not be able to explain the marginal, but in all likelihood real, existence of OVS and OSV word order languages. Second, a model of this kind has a much higher number of free parameters than the one presented above. While a blind commitment to model simplicity is not always warrented, it certainly is when data is scarce, such as the case here happens to be. It is worth remembering that all we have in terms of frequency data are five datapoints, representing the percentage that each of the six basic word orders constitutes in Dryer's sample. It will undoubtedly be interesting to explore the predictions of Bayesian iterative learning for the full set of word order typologies, which includes auxiliary placement and subordinate clause orders. Unfortunately however, there is no frequency data of this type at present, which severely limits the scientific value of extending the analysis in this direction.

In summary, it appears that - given the set of constraints proposed in Chapter 4. the cross-linguistic frequency of basic word orders cannot be directly attributed to the number of rankings that yield each order. However, it may be traced to an innate bias (prior) in an iterative learning process. Each generation of rational Bayesian agents chooses a grammar based on the language sample it receives from the previous generation, and the prior. If the process has converged, then the frequencies are an exact reflection of the prior. Logistic regression allows us to establish a particular function which closely maps the constraint violations to the frequency data. However, it is also possible that the process has not yet converged, in which case the distribution reflects broad characteristics of the prior, along with noise.

| Main | Main+Aux | Sub | Sub+Aux | Language | Rankings |
|---|---|---|---|---|---|
| SxVO | SAuxVO | ScxVO | ScAuxVO | English | 30 |
| SxVO | SAuxVO | cxVSO | cAuxSVO | Breton | 6 |
| SxVO | SAuxOV | cxVSO | cAuxSOV | Dinka | 3 |
| ScxVO | ScAuxOV | SOVx | SOVAux | German | 12 |
| ScxVO | ScAuxVO | xVSO | AuxVSO | Pred. 1 | 2 |
| ScxVO | ScAuxOV | xVSO | AuxVSO | Pred. 2 | 1 |
| ScxVO | ScAuxVO | xVOS | AuxVOS | Miya | 8 |
| ScxVO | ScAuxOV | xVOS | AuxVOS | Pred. 3 | 4 |
| **SxVO** | - | - | - | - | **51** |
| SOVx | SOVAux | SOVxc | SOVcAux | Latin | 18 |
| **SOVx** | - | - | - | - | **30** |
| xVSO | AuxSOV | cxVSO | cAuxSOV | S. Gaelic | 7 |
| xVSO | AuxVSO | cxVSO | cAuxVSO | Niuean | 7 |
| xVSO | AuxSVO | cxVSO | cAuxSVO | Welsh | 4 |
| **xVSO** | - | - | - | - | **18** |
| xVOS | AuxVOS | cxVOS | cAuxVOS | Tsou | 18 |
| **xVOS** | - | - | - | - | **18** |

Table 5.3: Number of winner rankings for each word order predicted in Chapter 4, grouping by simple main order

| Main | Main+Aux | Sub | Sub+Aux | Language | Rankings |
|---|---|---|---|---|---|
| SxVO | SAuxVO | ScxVO | ScAuxVO | English | 30 |
| **SxVO** | - | - | - | - | **30** |
| SOVx | SOVAux | SOVxc | SOVcAux | Latin | 18 |
| ScxVO | ScAuxOV | SOVx | SOVAux | German | 12 |
| **SOVx** | - | - | - | - | **30** |
| xVSO | AuxSOV | cxVSO | cAuxSOV | S. Gaelic | 7 |
| xVSO | AuxVSO | cxVSO | cAuxVSO | Niuean | 7 |
| SxVO | SAuxVO | cxVSO | cAuxSVO | Breton | 6 |
| xVSO | AuxSVO | cxVSO | cAuxSVO | Welsh | 4 |
| SxVO | SAuxOV | cxVSO | cAuxSOV | Dinka | 3 |
| ScxVO | ScAuxVO | xVSO | AuxVSO | Pred. 1 | 2 |
| ScxVO | ScAuxOV | xVSO | AuxVSO | Pred. 2 | 1 |
| **xVSO** | - | - | - | - | **30** |
| ScxVO | ScAuxVO | xVOS | AuxVOS | Miya | 8 |
| ScxVO | ScAuxOV | xVOS | AuxVOS | Pred. 3 | 4 |
| xVOS | AuxVOS | cxVOS | cAuxVOS | Tsou | 18 |
| **xVOS** | - | - | - | - | **30** |

Table 5.4: Number of winner rankings for each word order predicted in Chapter 4, grouping by simple subordinate order

| Main | Main+Aux | Sub | Sub+Aux | Language | Rankings |
|---|---|---|---|---|---|
| ScxVO | ScAuxVO | SxVO | SAuxVO | English | 30 |
| ScxVO | ScAuxVO | xVOS | AuxVOS | Miya | 8 |
| ScxVO | ScAuxVO | xVSO | AuxSVO | Breton | 6 |
| cxVSO | cAuxSVO | xVSO | AuxSVO | Welsh | 4 |
| ScxVO | ScAuxVO | xVSO | AuxVSO | Pred. 1 | 2 |
| **SxVO** | - | - | - | - | **50** |
| SOVxc | SOVAuxc | SOVx | SOVAux | Latin | 18 |
| ScxVO | ScAuxOV | SOVx | SOVAux | German | 12 |
| cxVSO | cAuxSOV | xVSO | AuxSOV | S. Gaelic | 7 |
| ScxVO | ScAuxOV | xVOS | AuxVOS | Pred. 3 | 4 |
| ScxVO | ScAuxOV | xVSO | AuxSOV | Dinka | 3 |
| ScxVO | ScAuxOV | xVSO | AuxVSO | Pred. 2 | 1 |
| **SOVx** | - | - | - | - | **45** |
| cxVOS | cAuxVOS | xVOS | AuxVOS | Tsou | 18 |
| **xVOS** | - | - | - | - | **18** |
| cxVSO | cAuxVSO | xVSO | AuxVSO | Niuean | 7 |
| **xVSO** | - | - | - | - | **7** |

Table 5.5: Number of winner rankings for each word order predicted in Chapter 4, grouping by complex main order

| Main | Main+Aux | Sub | Sub+Aux | Language | Rankings |
|---|---|---|---|---|---|
| SOVxc | SOVAuxc | SOVx | SOVAux | Latin | 18 |
| ScxVO | ScAuxOV | SOVx | SOVAux | German | 12 |
| cxVSO | cAuxSOV | xVSO | AuxSOV | S. Gaelic | 7 |
| ScxVO | ScAuxOV | xVSO | AuxSOV | Dinka | 3 |
| - | - | - | - | **-SOVx** | **40** |
| ScxVO | ScAuxVO | SxVO | SAuxVO | English | 30 |
| ScxVO | ScAuxVO | xVSO | AuxSVO | Breton | 6 |
| cxVSO | cAuxSVO | xVSO | AuxSVO | Welsh | 4 |
| - | - | - | - | **-SVO** | **40** |
| ScxVO | ScAuxVO | xVOS | AuxVOS | Miya | 8 |
| ScxVO | ScAuxOV | xVOS | AuxVOS | Pred. 3 | 4 |
| cxVOS | cAuxVOS | xVOS | AuxVOS | Tsou | 18 |
| - | - | - | - | **-VOS** | **30** |
| cxVSO | cAuxVSO | xVSO | AuxVSO | Niuean | 7 |
| ScxVO | ScAuxVO | xVSO | AuxVSO | Pred. 1 | 2 |
| ScxVO | ScAuxOV | xVSO | AuxVSO | Pred. 2 | 1 |
| - | - | - | - | **-VSO** | **10** |

Table 5.6: Number of winner rankings for each word order predicted in Chapter 4, grouping by complex subordinate order

| Candidate | Hd>Cmp | Spc>Cmp | HdEdge | HdCmp | Frequency |
|---|---|---|---|---|---|
| SOV | * | | | | 47.0 |
| SVO | | | * | | 41.2 |
| VSO | | | | * | 8.0 |
| VOS | | * | | | 2.5 |
| OVS | * | * | * | | 0.9 |
| OSV | * | * | | * | 0.4 |

Table 5.7: Basic word order typology

| | |
|---|---|
| Alpha | 0.1737 |
| Hd>Cmp | -0.2912 |
| HdEdge | -0.5394 |
| HdCmp | -2.5779 |
| Spc>Cmp | -3.8529 |

Table 5.8: Logistic regression fitted parameters

| Hypothesis | SOV | SVO | VSO | VOS | OVS | OSV |
|---|---|---|---|---|---|---|
| hSOV | 25% | 15% | 15% | 15% | 15% | 15% |
| hSVO | 15% | 25% | 15% | 15% | 15% | 15% |
| hVSO | 15% | 15% | 25% | 15% | 15% | 15% |
| hVOS | 15% | 15% | 15% | 25% | 15% | 15% |
| hOVS | 15% | 15% | 15% | 15% | 25% | 15% |
| hOSV | 15% | 15% | 15% | 15% | 15% | 25% |

Table 5.9: Hypothesis space with $p = .25$

| Hypothesis | Sub-hypothesis | Prior | SOV | SVO | VSO | VOS | OVS | OSV |
|---|---|---|---|---|---|---|---|---|
| hSOV | 1 | 7% | 25% | 15% | 15% | 15% | 15% | 15% |
| | 2 | 8% | 50% | 10% | 10% | 10% | 10% | 10% |
| | 3 | 7% | 40% | 5% | 5% | 25% | 15% | 10% |
| hSVO | 1 | 7% | 15% | 25% | 15% | 15% | 15% | 15% |
| | 2 | 8% | 10% | 50% | 10% | 10% | 10% | 10% |
| | 3 | 7% | 5% | 40% | 5% | 25% | 15% | 10% |
| hVSO | 1 | 7% | 15% | 15% | 25% | 15% | 15% | 15% |
| | 2 | 8% | 10% | 10% | 50% | 10% | 10% | 10% |
| | 3 | 7% | 5% | 5% | 40% | 25% | 15% | 10% |
| hVOS | 1 | 7% | 15% | 15% | 15% | 25% | 15% | 15% |
| | 2 | 8% | 10% | 10% | 10% | 50% | 10% | 10% |
| | 3 | 7% | 25% | 5% | 5% | 40% | 15% | 10% |
| hOVS | 1 | 2% | 15% | 15% | 15% | 15% | 25% | 15% |
| | 2 | 2% | 10% | 10% | 10% | 50% | 10% | 10% |
| | 3 | 2% | 25% | 5% | 5% | 40% | 15% | 10% |
| hOSV | 1 | 2% | 15% | 15% | 15% | 15% | 15% | 25% |
| | 2 | 2% | 10% | 10% | 10% | 10% | 10% | 50% |
| | 3 | 2% | 10% | 5% | 5% | 25% | 15% | 40% |

Table 5.10: Hierarchical hypothesis space

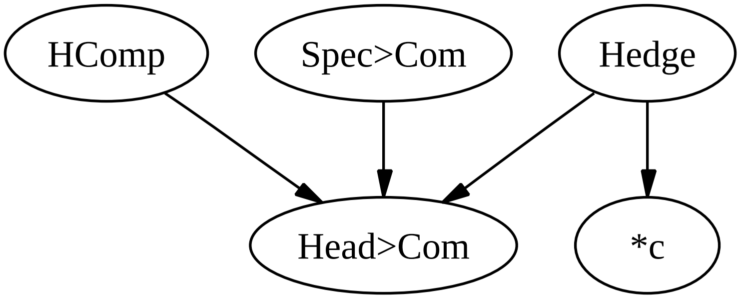



Figure 5.8: Hasse diagram of pure SOV word order

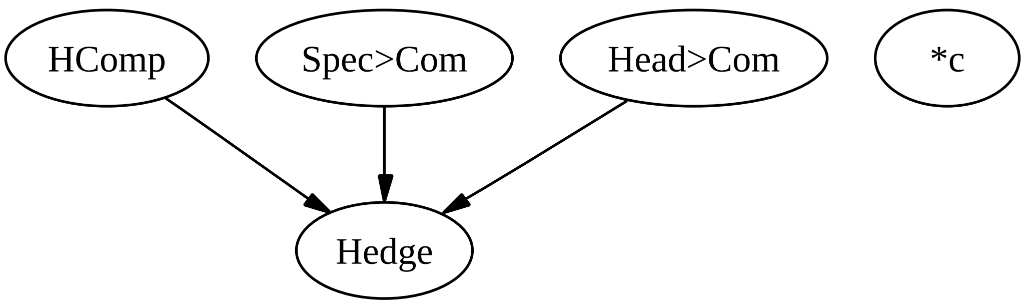



Figure 5.9: Hasse diagram of pure SVO word order

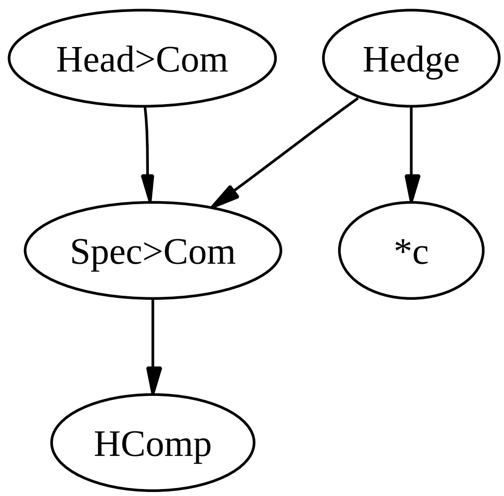



Figure 5.10: Hasse diagram of pure VSO word order

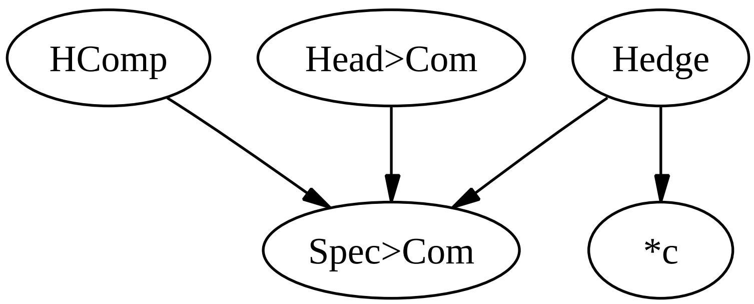



Figure 5.11: Hasse diagram of pure VOS word order

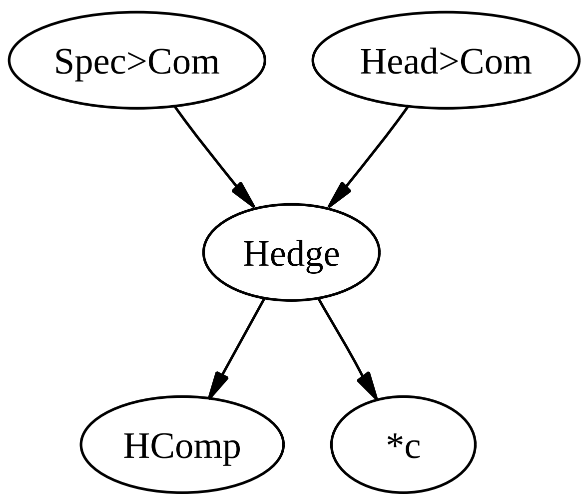



Figure 5.12: Hasse diagram of Welsh

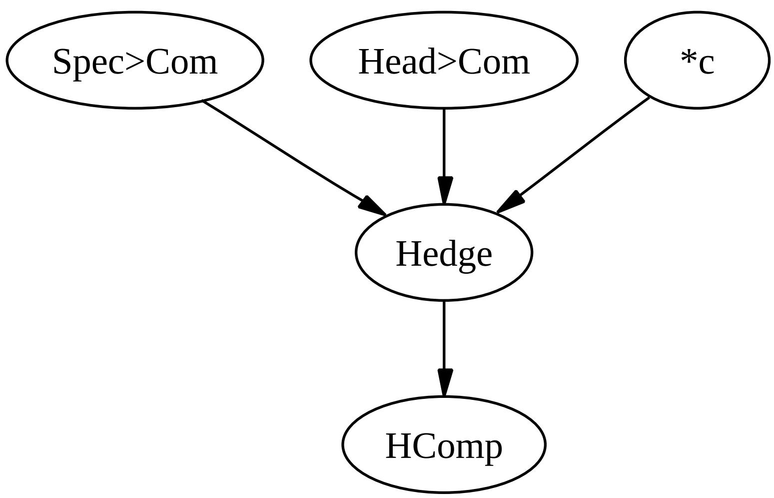



Figure 5.13: Hasse diagram of Breton

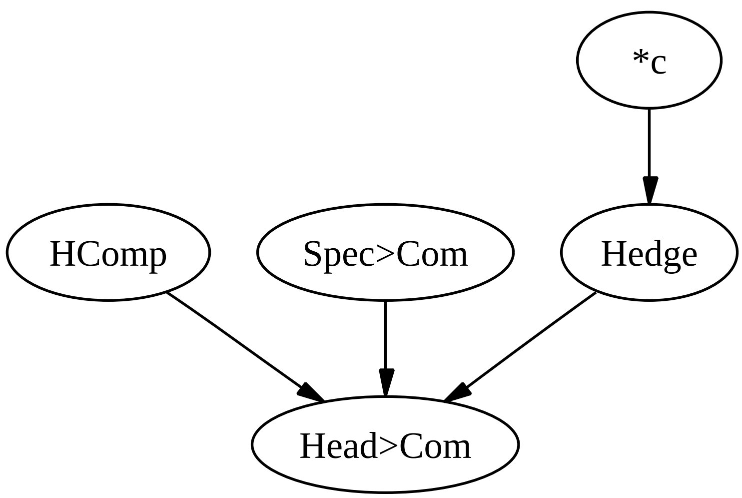



Figure 5.14: Hasse diagram of German

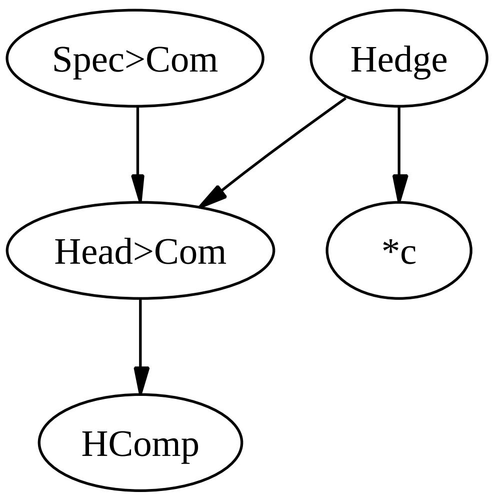



Figure 5.15: Hasse diagram of Gaelic

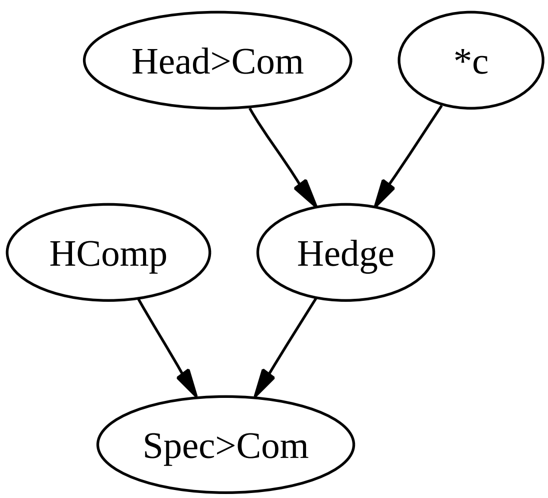



Figure 5.16: Hasse diagram of Miya

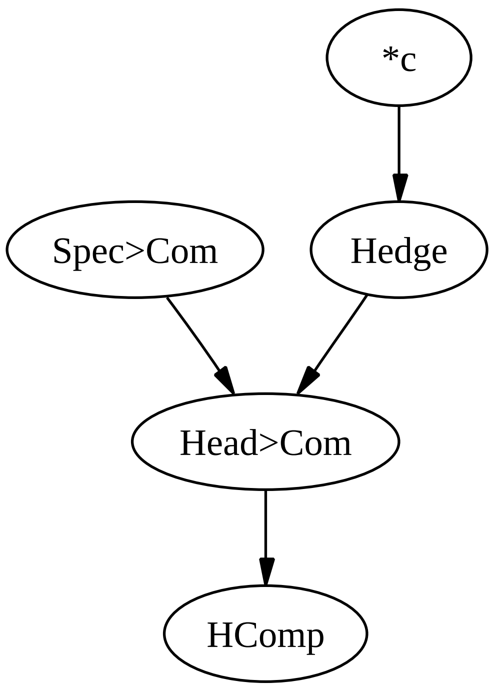



Figure 5.17: Hasse diagram of Dinka

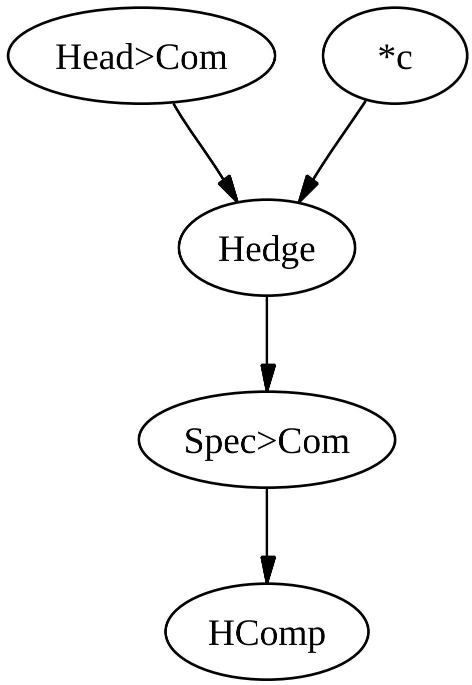



Figure 5.18: Hasse diagram of Predicted language 1

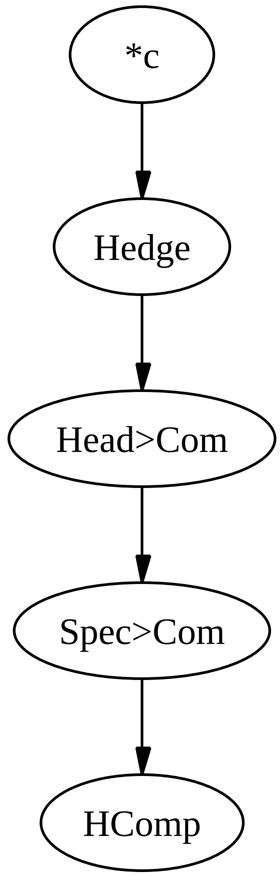



Figure 5.19: Hasse diagram of Predicted language 2

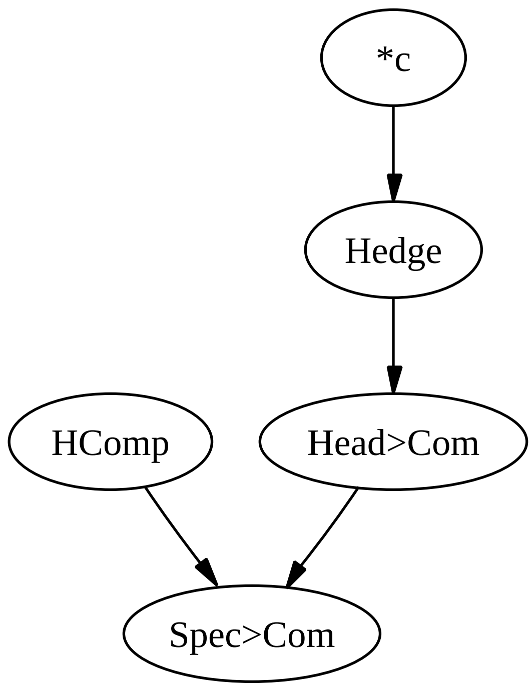



Figure 5.20: Hasse diagram of Predicted language 3

# Part V

# Word order variation within a language

# Chapter 6

# Discourse and morpho-phonological effects on linearization

In this chapter I will address the question of language-internal word order variation driven by factors external to syntax: discourse and morpho-phonology. While any OT analysis unavoidably raises the question of cross-linguistic typology, the primary goal of the chapter is to account for the availability of different word orders within the same language. Thus, only obvious implications for typology will be discussed.

## 6.1 Formation of yes-no interrogatives

I will first turn to the formation of interrogatives in SVO languages to address the issue of clause typing, which I will view as a specific case of discourse-marking. As the reader of this dissertation undoubtedly knows, English yes-no interrogatives are formed by inverting the order of the subject and the tensed verb (auxiliary or main) (81). In contrast, other languages like Russian and French, do not employ inversion. How would the proposed theoretical framework account for such variation?

Instead of resorting to differences at the level of the dependency structure, let us assume that the underlying representation of yes-no interrogatives does not differ structurally from the representation of the corresponding declarative, but differs from it in terms of a clause-typing Q-feature, which is realized on the highest head in the dependency structure. Thus, the representation of the examples in (81) is shown in Figure 6.1.

(80) George bought a camel.

(81) Did George buy a camel?

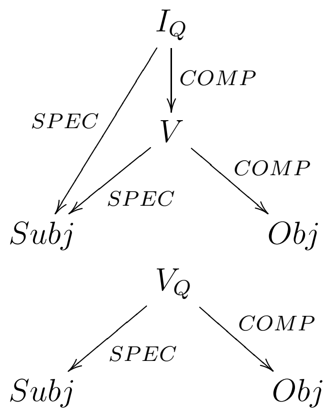



Figure 6.1: Underlying representation of yes-no interrogatives

Recall from Chapter 5 that pure SVO order (SVO, SAuxVO in main as well as subordinate) requires that $HdEdge$ be dominated by $Spc > Cmp$, $Hd > Cmp$ and $HdCmp$. Thus, in order to derive the yes-no question typology within those languages, I will keep the order of these constraints fixed.

Now let us suppose that inversion is affected by a clause typing constraint QLeft, which has the effect of requiring elements carrying the q-feature to align to the left edge of the utterance. When the Q-feature is stamped on the tensed verb, the result is tensed verb inversion. In complex-tense clauses, the result is auxiliary inversion. The predicted typology (Tableau 6.1) includes three cases. In the first case, the tensed verb alone inverts with the subject. In the second case, the "VP" complex, which includes both verb and object inverts past the subject; and in the third case, no inversion is observed. Thus, English is an example of the first case, while Russian is an example of the third case (83).

(82) Georgi kupil verbljuda.
George bought camel.
George bought a camel.

(83) Georgi kupil verbljuda?
George bought camel.
Did George buy camel.

A different picture emerges if the question is introduced by a complementizer. In this case, tensed verb inversion is not expected, since the Q-feature is carried by the complementizer, which is capable to remain in a sentence initial position on its own (Tableau 6.2).

While English does not have an independent Q-complementizer in the main clause, it does have one - “whether” or if, in the subordinate clause (85).

(84) Did George buy a camel?

(85) Laura asked whether/if George bought a camel

Note that, just as predicted, inversion is not observed in English subordinate yes-no questions.

However, if the presence of inversion is conditioned purely on the presence of a (phonologically independent) complementizer, it is difficult to explain why wh-questions in some languages invert in the main, but not in the subordinate clause [1]. For example, Bulgarian wh-questions behave as expected with respect to tensed verb inversion. Since they lack overt complementizers, they utilize inversion in both main and subordinate clauses. But unexpectedly, the pattern of tensed verb inversion in English is the same in wh-questions and yes-no questions despite the absence of a complementizer. Thus, inversion appears to be driven by three factors: type of interrogative, the status of the clause, and the presence of a complementizer (Table 6.1).

The lack of inversion in English subordinate wh-questions is problematic regardless of the structural status of the wh-word. There are three possible structures of a wh-question: the wh-word is itself a complementizer; the wh-word is a dependent of the Q-bearing element; the wh-word is discourse-marked, but structurally undistinguishable from a non-wh counterpart (Figures 6.2). It appears that the analysis of inversion would run into problems regardless of which possibility is chosen. If the wh-word is the complementizer, inversion

[1] By wh-questions here I will mean non-subject wh-questions, unless otherwise noted. The structure of the subject wh-questions is controversial, and in many ways orthogonal to the present discussion.

| Candidate | HdCmp | Spc>Cmp | Hd>Cmp | HdEdge | QEdge |
|---|---|---|---|---|---|
| Aux$_Q$SVO | 1 | | | 1 | |
| SAux$_Q$VO | | | | 2 | 1 |
| Aux$_Q$VOS | | 2 | | | |
| *Aux$_Q$ VSO | 1 | 1 | | | |
| V$_Q$SO | 1 | | | | |
| SV$_Q$O | | | | 1 | 1 |
| V$_Q$OS | | 1 | | | |

Table 6.1: Order in complex-tense, main, yes-no questions with no overt complementizer (q-particle)

| | HdCmp | Spec>Com | Hd>Com | HdEdge | QEdge |
|---|---|---|---|---|---|
| *C$_Q$AuxSVO | 1 | | | 1 | |
| C$_Q$SAuxVO | | | | 2 | |
| *C$_Q$AuxVOS | | 2 | | | |
| *C$_Q$AuxVSO | 1 | 1 | | | |
| *C$_Q$VSO | 1 | | | | |
| C$_Q$SVO | | | | 1 | |
| *C$_Q$ VOS | | 1 | | | |

Table 6.2: Order in complex-tense yes-no questions with complementizer (q-particle)

| | WH | Questions | Yes-no | Questions |
|---|---|---|---|---|
| Language | Main | Subordinate | Main | Subordinate |
| English | + | - | + | - |
| Bulgarian-dali | + | + | - | - |
| Russian | - | - | - | - |

Table 6.3: Tensed verb inversion in questions

would not be expected in the main clause. If it is not a complementizer, inversion would be expected in both clauses.

Obviously, the absence of inversion in subordinate wh-questions cannot be caused by the presence of a complementizer. However, QEdge may be inactive in the subordinate clause for other reasons. It is possible that languages vary with respect to available clause types at the subordinate level. For example, Bulgarian allows interrogative, but not imperative clause typing at that level (87). Perhaps, English is more restrictive than Bulgarian in requiring only declarative clause typing in subordinates.

(86) Georgi, kupì kamila!
George, buy camel
George, buy a camel!

(87) *Laura porucha na Georgi kupì kamila.
Laura ordered to George buy camel

Another possibility is that QEdge is a clause-typing constraint consisting of two subconstraints with fixed mutual ranking. The two subconstraints, QEdgeMain and QEdgeSub, govern clause typing of main and subordinate clauses respectively, and QEdgeMain must always rank higher that QEdgeSub. The justification for the ranking may come from the fact that that the subordinate clause is backgrounded with respect to information structure.

Note that, under this general approach, wh-movement and auxiliary inversion are additive processes. Thus, we would expect to find wh-in-situ languages with auxiliary inversion. An entirely different approach to inversion in wh-questions would be to link the placement

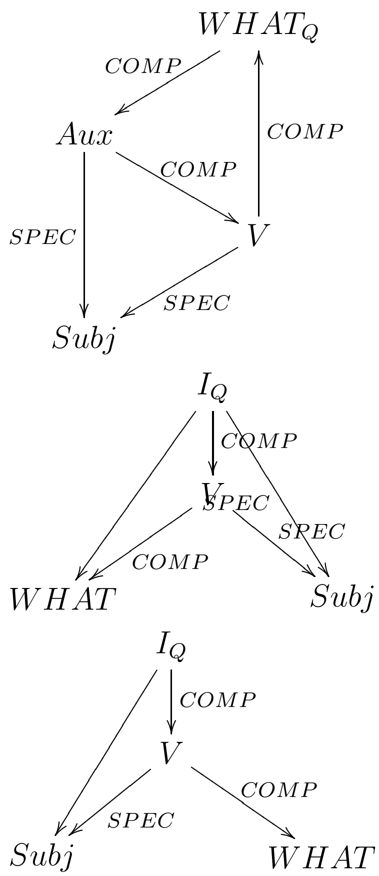



Figure 6.2: Possble underlying representations of object wh-questions

of the auxiliary directly to alignment driven by the position of the wh-phrase. In this case, we would expect inversion to be mostly absent in wh-in-situ languages, unless of course we assume wh-questions are simply treated as yes-no questions.

## 6.2 Wh-movement

We now turn to the explanation of wh-movement. We will pursue the hypothesis that it is partially a clause-typing, and partially a discourse-driven phenomenon. First, let us consider the possible dependency structures for an object wh-question with overt auxiliary. The traditional movement analysis is based on the tree representation in Figure 6.3. It involves head movement from I to C and A' movement of the complement of V to the specifier of CP.

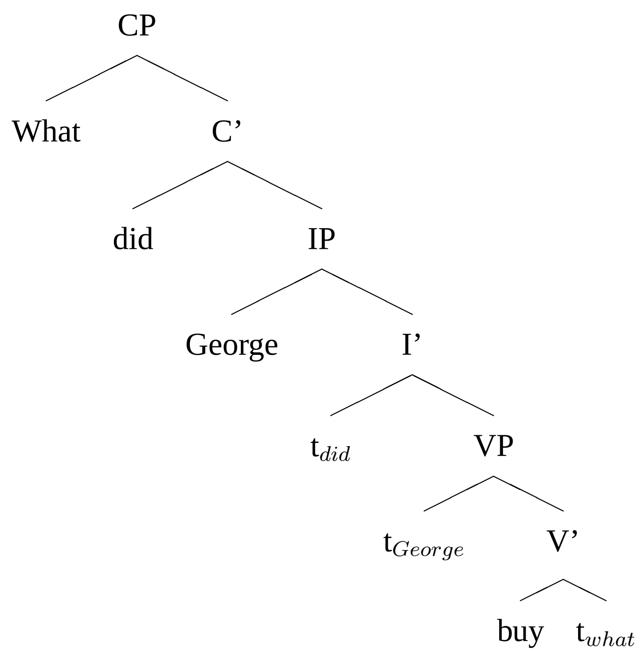
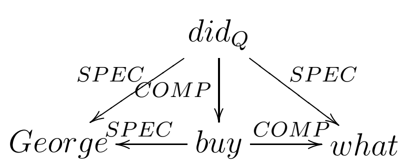



Figure 6.3: Wh-movement account of object wh-questions

In the previous section, we eliminated head movement in favor of a clause-typing Q-feature. The question now is whether A'-movement should be captured in structural terms. The "literal translation" of Figure 6.3 would require *what* and the Q-marked head to be linked by a *Spec*-type dependency Figure 6.2.




However, whether or not this link actually exists is open to debate. As we discussed in Chapter 2, the behavior of wh-phrases with respect to coreference suggest that the wh-phrase is interpreted at its position of origin, rather than at its landing site. For example, the original position of the wh-phrase before displacement authorizes coindexation between the wh-phrase and the pronoun in (88), compared to (89):

(88) Which picture of $himself_i$ does $Bill_i$ like best?

(89) *This picture of $himself_i$ is important to $Bill_i$?

This obligatory reconstruction to be one of the most puzzling facts about wh-movement and A-bar movement in general. If indeed the wh-phrase is in multiple positions, why is pronoun coreference calculated based on the lower position alone? Note that this is not a general fact about pronoun coreference, since it is well within the realm of possibility to establish a "new" coreference relationship based on the higher structural position in A-movement (91)

(90) $George_i$ appears to $himself_i$ to be in need of a camel.

(91) *It appears to $himself_i$ that $George_i$ is in need of a camel.

Since A-bar movement is movement from a thematic position to a non-thematic position, we would expect that, as [Heim and Kratzer, 1998] sensibly proposed, non-thematic aspects of the interpretation should be handled (perhaps exclusively) by the higher position. If this is not the case, are we justified in assuming multidominance in the case of A-bar movement? It may be better to eliminate the multi-headed dependency of the wh-phrase, and attribute movement to a linearization preference based on information structure (Figure 6.2). Let us examine this hypothesis. Suppose that a close-typing constraint requires

$$\begin{tikzcd} & C+did \arrow[dl, "SPEC"'] \arrow[d, "COMP"] & \\ George & buy \arrow[l, "SPEC"'] \arrow[r, "COMP"] & what \end{tikzcd}$$

the wh-marked head to be fronted - WhFront. By adding this constraint to the analysis of yes-no questions, we can examine the predicted linearization for the structure in Figure 6.2. The possible languages predicted by Tableaux 6.4 and 6.5, are given in Table 6.6

The resulting typology derives the wh-driven word order of all widely known types of SVO languages. Languages which support both active and inactive QEdge have different word orders associated with different complementizers and/or clause types. The entries containing empty fields represent additional languages derived from the assumption that only one version of QEdge is available. The first four languages represent languages with overt wh-movement. The next five languages are wh-in-situ languages with or without

| Sub. clause | HdCmp | Spc>Cmp | Hd>Cmp | HdE | *QEdge* | WhFr |
|---|---|---|---|---|---|---|
| AuxSVO | 1 | | | 1 | | 1 |
| SAuxVO | | | | 2 | | 1 |
| OVAuxS | | 2 | 2 | 2 | | |
| OSAuxV | 1 | 1 | 1 | 1 | | |

Table 6.4: Order in complex-tense object wh-questions: QEdge inactive

| Main clause | HdCmp | Spc>Cmp | Hd>Com | HdE | QEdge | WhFr |
|---|---|---|---|---|---|---|
| Aux$_Q$SVO | 1 | | | 1 | | 1 |
| SAux$_Q$VO | | | | 2 | 1 | 1 |
| Aux$_Q$VOS | | 2 | | | | 1 |
| OVAux$_Q$S | | 2 | 2 | 2 | 2 | |
| OAux$_Q$SV | 2 | 1 | 1 | 1 | | |
| OSAux$_Q$V | 1 | 1 | 1 | 1 | 2 | |
| OAux$_Q$VS | 1 | 2 | 1 | 2 | | |

Table 6.5: Order in complex-tense object wh-questions: QEdge active

| -QEdge | +QEdge’ | Language |
|---|---|---|
| OSAuxV | OAuxSV | English |
| OSAuxV | OSAuxV | Russian |
| - | OAuxSV | Bulgarian |
| SAuxVO | SAuxVO | Mandarin |
| SAuxVO | AuxSVO | |
| - | AuxSVO | |
| SAuxVO | AuxVOS | |
| - | AuxVOS | |
| OVAuxS | AuxVOS | |
| OVAuxS | OSAuxV | |
| OVAuxS | OVAuxS | |

Table 6.6: Predicted word orders for complex-tense, main, object wh-questions in SVO languages

auxiliary inversion. Finally, the last three languages exhibit a special type of wh-movement which pied-pipes the verb along with the direct object during movement. The reason for this behavior is the relatively high rank of HdCmp with respect to Hd $>$ Cmp. Thus, this pied-piping is expected to be peculiar to object wh-questions and therefore may remain relatively under the radar.

However, this approach does not address one of the most studied properties of wh-question typology - multiple wh-movement. The phenomenon has received a lot of attention in recent linguistic theory [Rudin, 1988, Richards, 1997, Bokovic, 2002], and is illustrated by the Bulgarian examples in (92, 93)

(92) Georgi kupi kamila na Laura.
George bought camel for Laura .
George bought Laura a camel.

(93) Kakvo na kogo kupi Georgi?
What to whom bought George
What did George buy for whom?

In order to capture the behavior of wh-phrases in languages like Bulgarian, we would have to assume that the constraint WhFront is violated unless all wh-phrases are fronted. But if so, we would expect all languages with wh-movement to be multiple wh-movement languages.

Note that while languages have the choice of exhibiting multiple wh-movement or not, they do not vary with respect to the exact number of wh-movements allowed. Thus, there are no languages that move two wh-phrases but not three, or three but not four etc. This requires the introduction of two constraints - the first to deal with the obligatory movement of one wh-phrase, the second to move the rest.

Obviously, it is possible to present an analysis of multiple wh-movement in other ways. For example, one could ban multiple wh-movement with one constraint and favor it with another. Such strategy is undertaken by [Legendre et al., 1998]. While the type of OT analysis they propose differs widely from the one presented here in terms of underlying and surface representations, they utilize two constraints to capture the difference between Bulgarian and English. *Adjoin penalizes multiple wh's adjoined to the highest clausal specifier, while *Absorb penalizes single wh-movement in the presence of many wh's. Thus, if a language respects *Absorb, it exhibits multiple wh-movement. If, on the contrary, it respects *Adjoin, it allows only one wh-movement. Both *Absorb and *Adjoin are in conflict with a third constraint which prohibits wh-movement altogether.

Instead of setting up a direct conflict over multiple wh-movement, let us assume that WhFront is satisfied whenever there is a wh-phrase at the left edge of the sentence. In addition, let us consider a constraint Wh$>$NonWh, which is violated whenever any wh-phrase remains in situ. When a question contains a single wh-phrase, the effects of both constraints completely overlap. In the presence of multiple wh-phrases however, languages fall alongside of the English/Bulgarian divide (Tableau 6.7). This entails an asymmetric relationship between WhFront and Wh$>$NonWh: while Wh$>$NonWh can be violated if WhFront is not,

| Main clause | HdCmp | Spc> Cmp | Hd> Cmp | HdEd | QEd | WhFr | +Wh> –Wh | Hd> Adj |
|---|---|---|---|---|---|---|---|---|
| $\text{Aux}_Q\text{SVO}_{wh}\text{D}_{wh}$ | 1 | | | 1 | | 1 | 1 | |
| $\text{SAux}_Q\text{VO}_{wh}\text{D}_{wh}$ | | | | 2 | 1 | 1 | 1 | |
| $\text{Aux}_Q\text{VO}_{wh}\text{D}_{wh}\text{S}$ | | 2 | | | | 1 | 1 | |
| $\text{O}_{wh}\text{VAux}_Q\text{SD}_{wh}$ | | 2 | 2 | 2 | 2 | | 1 | |
| $\text{D}_{wh}\text{O}_{wh}\text{VAux}_Q\text{S}$ | | 2 | 2 | 2 | 2 | | | 1 |
| $\text{O}_{wh}\text{Aux}_Q\text{SVD}_{wh}$ | 2 | 1 | 1 | 1 | | 1 | 1 | |
| $\text{O}_{wh}\text{D}_{wh}\text{Aux}_Q\text{SV}$ | 2 | 1 | 1 | 1 | | | | 1 |
| $\text{O}_{wh}\text{SAux}_Q\text{VD}_{wh}$ | 1 | 1 | 1 | 1 | 2 | 1 | 1 | |
| $\text{O}_{wh}\text{D}_{wh}\text{SAux}_Q\text{V}$ | 1 | 1 | 1 | 1 | 2 | | | 1 |
| $\text{O}_{wh}\text{Aux}_Q\text{VSD}_{wh}$ | 1 | 2 | 1 | 2 | | 1 | 1 | |
| $\text{O}_{wh}\text{D}_{wh}\text{Aux}_Q\text{VS}$ | 1 | 2 | 1 | 2 | | | | 1 |

Table 6.7: Order in complex-tense object wh-questions: QEdge active

a violation of Wh>NonWh necessarily involves a violation of WhFront. But what constraint interacts with the two above in order to prevent wh-movement in some languages? Instead of postulating an across-the-board constraint against movement, we will rely on the local ordering relationship which must hold among heads and their dependents. Thus, the movement of a complement is prevented by the previously mentioned constraints HdCmp, Hd>Cmp and Spc>Cmp. Similarly, the movement of an adjunct would be prevented by an independently motivated constraint which enforces the order of the head and its adjunct, e.g. Hd>Adj.

An important typological consequence of this is the possibility that some language allow wh-movement of adjuncts, but not arguments, or vice versa. Some evidence of the availability of such distinction comes from French. While French normally allows wh-heads to remain in-situ (95), it prefers not to do so with certain wh-adjuncts (97) [Mathieu, 2004].

(94) Qui a tu vu?
who have you seen

(95) Tu a vu qui?
you have seen who
Who have you seen?

(96) Ou a tu vu un piano?
where you have seen a piano

(97) ??Tu a vu un piano où?
you have seen a piano where
Where have you seen a piano?

While the optionality of wh-movement in French is problematic as a whole, the data above seems to suggest that whatever constraints encourage the wh-phrases to remain in situ apply differentially to different types of wh-phrases. What is the independent, non-technical motivation for Wh>NonWh? I assume that a constraint of this type has its origins in discourse. It has been argued elsewhere that wh-movement is a type of focus movement. Meinunger claims that “the apparent difference with respect to the unexpected strength of weak islands is explained by showing that the algebraic operations that focus computation is associated with are different from those associated with individual denoting wh-phrases" [Meinunger, 2003]. He assumes that

> "wh-phrases denote into a Boolean algebra defined for all operations including complement and intersection, which makes them good island escapers. On the other hand, focus phrases have as denotation domain a free choice semi-lattice, which is not defined for the mentioned Boolean operations. This traps them within weak islands, hence the apparent asymmetry."

Furthermore, it has long been suggested that languages that lack overt focus movement rule raise a focal constituent to a scope position at the level of logical form (LF) [Chomsky, 1977a]. Finally, many languages exhibit overt focus movement akin to overt wh-movement, e.g. Hungarian [Kiss, 1998]. I interpret these arguments in support of the idea that wh-movement results from a discourse constraint.

## 6.3 The linearization of clitics

The last question we will address is to what extent linearization constraints are driven by lexical morpho-phonological properties. The debate on this issue exists in virtually all frameworks, and revolves around the distributional properties of clitics. Clitics do not share any immediately obvious syntactic properties. They belong to different syntactic categories: pronouns, auxiliaries, and sentential particles with a range of functions. In contrast, all clitics fall into the phonological category of non-stressed elements. That is why their ordering has been traditionally viewed as a phonological phenomenon. Wackernagel's law of second position claims that non-stressed elements agglomerate after the first prosodic constituent of the sentence, as in (98), where clitics are underlined. As we will see in this section, Wackernagel's law is at best a crude approximation, rather than an accurate description of Bulgarian clitic distribution.

(98) Georgi <u>li si e</u> kupil kamila?
Georgi *li self* is bought camel

Was George the one who bought himself a camel?

While it is true that clitics are not members of a single syntactic category, their distributional properties are not independent from their syntactic specification. For example, the sequence of clitics in (1) cannot be randomly rearranged, even though their position would still satisfy Wackernagel's law (99).

(99) *Georgi e li si kupil kamila?

Wackernagel's law, by virtue of its phonological character, has nothing to say about why the auxiliary *e* is not allowed to precede the question particle *li*. These observations give rise to the currently prevalent view: that clitic position is decided by a non-modular interaction of syntax and phonology.

On one view, clitics are allowed to violate certain syntactic principles in order to satisfy phonological requirements on their position [Legendre, 1998]. Alternatively, the phonological features of a clitic force another constituent to move to a position supporting the clitic [Montapanyane, 1997].

Elsewhere I have argued that the second-position behavior of clitics can be explained in syntactic terms in the X-bar framework, if it is assumed that they are heads of functional projections with an inviolable requirement for a full specifier, which can sometimes be satisfied by a remnant constituent [Savova, 2002]. I will begin this section by fleshing out a version of this analysis in the new theoretical framework advanced in this dissertation. I will then proceed to discuss an interesting case of clitic incompatibility which is best explained by conflicting ordering preferences.

### 6.3.1 The position of the clitic interrogative marker in Bulgarian

Yes-no questions in Bulgarian can be formed either with the clitic interrogative marker *li* or with one of its non-clitic counterparts: *dali* ('is it the case that') or *nali* ('isn't it the case that') (cf. Section 6.1). Under neutral prosody, the non-clitic markers question the truthfulness of the proposition as a whole (98), while the clitic marker may focus the query on the constituent preceding it (98).

(100) (7) Dali/nali Georgi e kupil kamila?

Dali/nali Georgi bought from market-Def

Is it/isn't it the case that George bought the market?

Unlike its non-clitic counterparts, the clitic *li* never appears in sentence-initial position 101) and, in most cases, it is adjacent to the main verb. The characterization of *li* as a second-position clitic is based on the neutral *li* questions. Even in the absence of specific focus interpretation, *li* cannot remain the first word of the sentence but must attract some constituent, namely the tensed verb (103).

(101) *Li kupi Georgi kamila

(102) Kupi li Georgi kamila?

bought li Georgi camel

Did George buy a camel?

(103) Beshe li kupil Georgi kamila?

was li Georgi bought camel

Has George bought a camel?

To begin our analysis, let us assume a set of two constraints, *ClEdge, and Foc > Cl, where Foc is a focus-marked element, Cl is a clitic, and *ClEdge penalizes clauses beginning with a clitic. I choose Foc > Cl over Foc > li in order to preserve the universality of the constraint set. Foc > li is obviously not directly universal, since many languages do not have li in their lexicon. When a focus-marked element is present and precedes the clitic, it satisfies both *ClEdge, and Foc > Cl. When no focus-marked elements are available, Foc> Cl is vacuously satisfied, which allows *ClEdge to kick in. In addition, let us also assume that all previous constraints constraints concerning complementizer- particularly HdCmp, Hd>Cmp, and Spc>Cmp), do not apply to clitic heads.

These assumptions give rise to the following effects:

1. In focus-free complex-tense clauses with a full auxiliary, the auxiliary is aligned with the left edge of li. This is because aligning any other head with li will result in unnecessary violations of Hd>Cmp. For example, if the verb is fronted and the

auxiliary remains in situ, the auxiliary will end up following its complement- the verb.

2. By the same principle, the verb aligns with the left edge of li in focus-free simple-tense clauses or complex-tense clauses with a clitic auxiliary.

3. In complex-tense clauses with focus, the focused element aligns with li.

However, one important issue is still outstanding. When Foc > Cl is active, the dependents of the focused head can intervene between it and the clitic, but when *ClEdge is in charge, the clitic must be aligned with the tense-bearing head-verb or auxiliary. In other words, *li* is always adjacent to a verbal element when that element precedes it (106).What is the source of this difference?

(104) Kupil li si beshe Georgi kamila?
bought li himself was George camel
Had George bought himself a camel?

(105) *Kupil kamila li si beshe Georgi?
bought camel was li himself George

(106) Kamila kupil li si beshe Georgi?
bought camel was li himself George

I suggest that there is a third constraint - StrAdj(V, Cl), which prefers clitics to be immediately adjacent to a verbal element. Unlike the aforementioned adjacency constraints, this constraint is only satisfied if nothing- including dependents, intervenes between the verbal element and a clitic.

There is ample independent support for StrAdj(V, Cl). Even when the main verb follows, very few elements are allowed to intervene. For example, neither the subject nor adverbs are permitted between li and the verb (107), 108).

(107) ??/*Beshe li Georgi kupil kamila?
Had li George bought camel

(108) ??/*Beshe li vchera kupil kamila? Had li yesterday bought camel

Among the exceptions are other clitics (109), which intervene both because the consideration satisfies StrAdj(V, Cl), and because of a clitic ordering preference.

(109) Georgi li si kupi kamila? George li si bought camel Did George buy himself camel?

However, it is also the case that non-clitic auxiliaries are allowed (110).

(110) Georgi li beshe kupil kamila? George li was bought camel Was it George who bought himself a camel?

This would be surprising, unless we consider the possibility that *li* is equally happy being adjacent to the auxiliary on either side, which allows for other factors to determine its exact placement. That is exactly what we would expect if StrAdj(V, Cl) is satisfied whenever the clitic is adjacent to any one of the potentially multiple verbal elements-auxiliaries or main verb.

The result of an optimization under these constraints is the Tableau (6.8) In the focus-free case, *ClEdge-inviolable for Bulgarian, forces a choice between fronting the auxiliary, the verb or a verbal dependent. But fronting the auxiliary only violates HdCmp by interrupting the adjacency of Aux and V. Any other move would cause an additional violation of Hd>Cmp. In simple-tense clauses, the same principle prefers the verb over the object. Fronting both the verb and the object is excluded by StrAdj(V, Cl).

However, if the auxiliary is a clitic, fronting it does not resolve the violation of *ClEdge. Consequently, to equally good options remain: fronting either the verb or the object causes violations of HdCmp and Hd>Cmp. One possibility is to leave the resolution of this conflict to an interpretative optimization, of the type suggested by [Buchwald et al., 2002]. Another option is to reconceptualize Foc>Cl as follows:

**Foc>Cl** Cl must be preceded by all and only focused elements.

Thus, whenever a non-focused noun precedes li, Foc>Cl is violated (Tableau 6.9).

The case of sentences with focus-marked elements is given in Tableau 6.10. The correct linearization is derived for any ranking which places Foc>Cl above Hd>Cmp.

The characterization of *li* as second-position is a tad misleading, in that it is not always preceded by a single word, or even a single whole constituent. Instead, it appears to follow

| Input | Candidate | *ClEdge | Foc>Cl | HdCmp | Hd>Cmp | StrAdj(V,Cl) |
|---|---|---|---|---|---|---|
| full Aux | li AuxVO | 1 | | | | |
| ☞ | Aux li VO | | | 1 | | |
| | V li AuxO | | | 1 | 1 | |
| | O li AuxV | | | 1 | 1 | |
| | VO li Aux | | | 1 | 1 | 1 |
| no Aux | li VO | 1 | | | | |
| ☞ | V li O | | | 1 | | |
| | O li V | | | 1 | 1 | |
| | VO li | | | | | 1 |

Table 6.8: The position of interrogative clitic in Bulgarian yes-no questions: focus-free

| Input | Candidate | *ClEdge | Foc>Cl | HdCmp | Hd>Cmp | StrAdj(V,Cl) |
|---|---|---|---|---|---|---|
| cl Aux | li AuxVO | 1 | | | | |
| | Aux li VO | 1 | | 1 | | |
| ☞ | V li AuxO | | | 1 | 1 | |
| | O li AuxV | | 1 | 1 | 1 | |
| | VO li Aux | | | 1 | 1 | 1 |

Table 6.9: The position of interrogative clitic in Bulgarian yes-no questions: focus-free with clitic auxiliary

more than one constituent, or a partial constituent. Consider for example multiple wh-questions. While wh-questions do not normally include the question particle li, it may surface in marked contexts (such as rhetorical questions) for emphasis (111,112).

(111) Koj kakvo na kogo e dal?
Who what to whom is given
'Who bought what to whom?'

(112) Koj kakvo na kogo li e dal?
Who what to whom li is given
'Who bought what to whom?'

(113) *Koj li kakvo na kogo e dal?
Who what to whom li is given
'Who bought what to whom?'

If , as expected from Wackernagel's law, *li* cliticizes to the first stressed element, the data in (113) is surprising. Since the first stressed element is *koj*, we would expect ((113) to be grammatical and (112) ungrammatical. However, we observe the opposite pattern, which is in agreement with our analysis. Since all wh-words are focused, we would expect Foc>Cl to front all of them.

Another type of *li* questions present a similar problem for Wackernagel's law and other analyses (e.g. [Montapanyane, 1997]). *Li* is sometimes preceded by one or more verb dependents together with the verb (114).

(114) Kamila kupi li?
camel bought li
Did he buy a CAMEL in particular?

Here, all preceding verb dependents can be interpreted as foci. For example, an appropriate answer to a question refers to its focus (Swart and Hoop, 1995). That the NPs in (114) are foci is evident by the fact that (115 and 116) are grammatical answers to (114) .

(115) Ne, samo krava.
No, just cow.

No, just a cow.

(116) Ne, samo Laura.

No, just Laura.

Furthermore, the verb dependents can be subject to contrastive focusing 117.

(117) Kamila poluchi li Georgi ili samo krava?

camel was li got Georgi or just cow

Did George get a camel or just cow?

Elsewhere I have argued that this type of focusing is due to the existence of preverbal focus position which allows the focused elements to pied-pipe the remnant focus phrase [Savova, 2002]. The current analysis can also be easily modified to include a pied-piping constraint for the particular semantic subtype of focus which in constituent-based frameworks is referred to as preverbal. Let us call this constraint Align(vFoc, V), where vFoc is verbal focus, and "Align" stands for aligning the right edge of the first item with the left edge of the second item. Assuming that Align(vFoc, V) is ranked higher that Hd>Cmp, the order of (114) is derived (Tableau 6.11).

### 6.3.2 Pronominal clitics and ineffability

Most constraints formulated so far are also involved in the positioning of pronominal clitics, whose behavior poses some interesting questions about the make up of the candidate set in the current framework. In particular, there appear to be cases in which clitics are in direct competitions with the full noun phrases to which they correspond. The question then is: Is pronominalization and cliticization decided at the level of structural linearization?

To shed light on this issue, let us turn to dative and accusative pronominal elements in standard English, which have a curious inability to cooccur. While the double object construction permits any combination of full NPs, or full accusative NP plus dative clitic (118, 119), it disallows sentences like (120, 121).

(118) Show George the actress.

(119) Show him the actress.

(120) *Show George her.

(121) *Show him her.

Bulgarian pronominal clitics exhibit a similar kind of ineffability. Interestingly however, it is limited to a subset of case/number combinations in the paradigm (123, cf 122). If both clitics are not permitted, the meaning can be realized with a single clitic or full NP.

(122) Pokazaha mu go.
showed(3rd pers pl) him(Dat) him(Acc)
They showed him to him

(123) *Pokazaha mu me.
showed(3rd pers pl) him(Dat) me(Acc)
They showed him to him

It is plausible that such ineffability is caused by the interaction of highly ranked linearization constraints. The Bulgarian examples point to two culprits - a constraint on the ordering of case-marked elements and and a constraint on the ordering of person-marked elements. Indeed, both are cross-linguistically motivated. In the English double object construction, the dative argument must precede the accusative (118, 124).

(124) *Show the actress George.(=Show George to the actress)

Even though Bulgarian does not have a double object construction, the fact that the dative clitic argument must precede the accusative is a reflection of the same preference (125-126).

(125) *Pokazaha go mu.
Showed(3rd pers pl) him(Acc) him(Dat)

(126) *Te go mu pokazaha.
they him(Dat) him(Acc) showed (3rd pers pl)

(127) *Te mu me pokazaha.
they him(Dat) me(Acc) showed(3rd pers pl)

(128) Te mu go pokazaha.
they him(Dat) him(Acc) showed(3rd pers pl)
They showed him to him.

This is true irrespective of the position of the verb. On this basis, we can postulate a linearization constraint Dat>Acc, which requires double objects to be ordered according to their case. While Italian and French generally follow the Bulgarian pattern, the latter requires the accusative clitic to precede the dative in the third person. This has prompted other researchers to conclude that there is no universal preferred order of accusative and dative ([Legendre, 1996]). Instead, I suggest that $lui$ should not be considered a true clitic, due to the fact that it belongs to the paradigm of strong pronominal forms [2].

If this is the only constraint responsible for clitic linearization, the ungrammaticality of (127) and (123) is unexpected, since in both cases the constraint is obeyed. The grammaticality of the minimal pairs in (122) and (128) suggests that the crucial factor for (127) is personhood.

It has been noted before that many languages obey a markedness hierarchy of personhood, which influences the realization of lexical items as arguments [Aissen, 1999, Dixon, 1979, Silverstein, 1976].

**Person-hierarchy** 1st>2nd>3rd Pronoun>Proper Noun>Human>Animate>Inanimate

The higher a person is on the person hierarchy, the more likely it is that this person will function as a transitive agent, or as a subject. While the person hierarchy does not restrict the personhood transitive agents in Bulgarian, I will argue that it indirectly governs the placement of clitics, by ruling out cases in which a pronoun on a lower personhood scale precedes a higher pronoun. This is especially true in the cases where first and third person are involved (123, 127) , but is also responsible for the degraded acceptability of some first and second person combinations (129).

(129) ??Te ti me pokazaha.
they me(Acc) him(Dat) showed(3rd pers pl)

[2] $moi, toi, lui, elle, nous, vous, eux, elles, a$

In other words, the case ordering constraint rules out any linearization in which the accusative precedes the dative. Similarly, the person ordering constraint rules out any linearization in which the third person precedes the first. If both the case and the person constraint are sufficiently highly ranked, the clitic combination is disallowed.

A comparable phenomenon in English leads to ineffability of double object clitic constructions in general. In both Bulgarian and English, clitics must be adjacent to their heads. For example, it is not possible to insert a Dative object between the verb and its accusative clitic, even though the Dat>Acc order is preferred (130).

(130) *They showed George her.

Similarly, while Bulgarian generally allows the verb-object adjacency to be interrupted by adverbs for discourse purposes, this is not the case if the object happens to be a clitic.

(131) Georgi kupi vchera kamila.
George bought yesterday camel George bought a camel yesterday.

(132) *Georgi kupi vchera ja.
George bought yesterday her George bought it yesterday.

We have already formulated a constraint with precisely such effects in our analysis of the placement of interrogative *li*- StrAdj(V, Cl). Presumably, if StrAdj(V, Cl) is sufficiently highly ranked, total ineffability of double pronominals results.

However, this line of reasoning poses an interesting question: what exactly is the candidate set for linearization? In order for double clitic combinations to become ineffable, they have to be in direct competition with linearizations containing full arguments. In our setup, this can only be accomplished if the phonological form of arguments is not determined until the level of linearization. Thus, the input to the linearization must contain merely the semantic, syntactic and discourse features of the argument heads, not the whole lexical entries. If so, we can rely on a discourse-motivated constraint (*FullNoun) which prefers clitic forms to full noun forms in cases wherever discourse and syntax allow it. The factorial typology of the analysis includes English, Bulgarian and two additional languages (Tableau 6.3.2). One language allows double clitic construction regardless of person and orders them according to case. The other also allows double clitics but orders 1st pers. acc. before 3rd pers. dat., and 3rd pers. dat. before 3rd pers. acc.

This concludes our examination of language-internal word order variation. We have shown several cases in which linearization is influenced by factors external to syntax and originating at the level of discourse or morpho-phonology. Since the framework proposed here considers syntax only one of the many determiners of word order, it is possible to construct accounts of such phenomena without compromising the theoretical underpinnings of our approach. An important question for further investigation is the precise make up of the candidate set, and consequently, the form of the input.

| Input | Candidate | *ClEdge | HdCmp | Foc>Cl | Hd>Cmp | Align(V,Cl) |
|---|---|---|---|---|---|---|
| full Aux | li AuxVO | 1 | | 1 | | |
| | Aux li VO | 1 | 1 | | | |
| | V li AuxO | 1 | 1 | | 1 | |
| ☞ | O li AuxV | | 1 | | 1 | |
| cl Aux | li AuxVO | 1 | | 1 | | |
| | Aux li VO | 1 | 1 | 1 | | |
| | V li AuxO | 1 | 1 | | 1 | |
| ☞ | O li AuxV | | 1 | | 1 | |
| no Aux | li VO | 1 | | 1 | | |
| | V li O | 1 | 1 | | | |
| ☞ | O li V | | 1 | | 1 | |

Table 6.10: The position of interrogative clitic in Bulgarian yes-no questions: object focus

| Input | Candidate | *ClEdge | Foc> | HdCmp | Align | Hd> | Align |
|---|---|---|---|---|---|---|---|
| | | | Cl | | (vFoc, V) | Cmp | (V,Cl) |
| fullAux | liAuxVO | 1 | 1 | | | | |
| | AuxliVO | | 1 | 1 | 1 | | |
| | VliAuxO | | 1 | 1 | 1 | 1 | |
| | OliAuxV | | | 1 | 1 | 1 | |
| ☞ | OVliAux | | | 1 | | 2 | |
| clAux | liAuxVO | 1 | 1 | | 1 | | |
| | AuxliVO | 1 | 1 | 1 | 1 | | |
| | VliAuxO | | 1 | 1 | 1 | 1 | |
| | OliAuxV | | | 1 | 1 | 1 | |
| ☞ | OVliAux | | | 1 | | 2 | |
| noAux | liVO | 1 | 1 | | 1 | | |
| | VliO | | 1 | 1 | 1 | | |
| | OliV | | | 1 | 1 | 1 | |
| ☞ | OVli | | | 1 | | 2 | |

Table 6.11: The position of interrogative clitic in Bulgarian yes-no questions: verbal focus

| Input | Output | StrAdj(V,Cl) | Dat>Acc | 1 > 3pers. | *FullN |
|---|---|---|---|---|---|
| V(Dat3,Acc1) | Dat3 Acc1 | 1 | | 1 | |
| | Acc1 Dat3 | 1 | 1 | | |
| | Dat3 (NP) | | | | 1 |
| V(Dat3,Acc3) | Dat3 Acc3 | 1 | | | |
| | Acc3 Dat3 | 1 | 1 | | |
| | Dat3 (NP) | | | | 1 |

# Chapter 7

# Relating the current proposal to existing OT syntax literature.

It is important to discuss where the approach advocated in this work fits within the context of existing OT syntax literature. While approaches to OT syntax differ vastly with respect to substantive assumptions on the nature of constraints and representations, one can identify at least two broad areas which are directly relevant to the proposal in this dissertation.

One such area consists of approaches to optimization at the lexical level. These are largely complementary to the linearization formalism presented here, because their goal is to account for lexical inventories, rather than precedence and dominance relations in utterances. Virtually all linguistic frameworks, including this one, accept that languages differ in terms of the concepts they choose to lexicalize. Lexicon-oriented OT proposals should be understood as making claims about the nature of the mechanisms behind these language-specific choices. These claims are independent from theoretical assumptions about the nature of the structure-building component, as well as from assumptions about the relationship of structures to strings.

Simply put, each language might represent a solution to more than one optimization problems. First, a language must decide, given a space of meanings, which of them should be mapped to atomic symbols. These decisions are presumably made on the basis of communicative convenience. The concepts which are (expected to be) referred to more frequently are better candidates for lexicalization. It is well-known, for example, that some Eskimo languages have different words for different kinds of snow, while some languages have no word for snow at all. Similarly, some languages distinguish relatives on the maternal side from relatives on the paternal side, while others do not. It is reasonable to assume that such differences have to do with the environmental and cultural circumstances in which the language is used. It is also reasonable to assume that some concepts are universally more useful in discourse than others, giving rise to a markedness hierarchy in lexicalization. For example, it may be considered universally more important to distinguish one from many than one from two, because the sitations in which the first distinction would apply are much more common that the situations in which the latter would apply (since the latter is a subset of the former). Lexicon-oriented approaches to OT apply to this level of the theory.

Once a language "decides" on its atomic concept inventory, it requires combinatorial rules to build complex concepts out of that inventory. Obviously, the availability of lexical entries determines the structural inventory of the language. However, I have assumed that mechanism through which words are combined is universal. Finally, a language must decide how to flatten the resulting combinatorial structures into a one dimensional speech stream. This is where the linearization optimization fits in the linguistic formalism. Obviously, there is no contradiction between OT approaches to lexicon building and the OT approach to linearization proposed here.

Thus, OT approaches to the lexicon strive to identify the mechanism which enables each individual language to determine what semantic structures should be stored as atomic linguistic units. Consequently, they assume the underlying semantic feature structures compete for realization as lexical items. For example, Bresnan's account of cross-linguistic lexicalization of pronominal elements [Bresnan, 2001] and auxiliaries [Bresnan, 2002] assume that the underlying representation is an f-structure and the surface representation is a partially matching f-structure, accompanied by a c-sturcture, as defined in Lexical Functional Grammar. In brief, f-structure represents the abstract functional features of a lexical entry, including its semantic content and selectional requirements, while a c-structure is their syntactic realization. The constraints differentially penalize the realization of some f-features and f-feature combinations at the surface, which result in the collapse of minimally different underlying f-feature structures into one surface realization. Thus, the underlying f-structure of the first person singular present-tense auxiliary is [BE PRES 1 SG]. In English, its surface f-structure is identical to its underlying f-structure, which results in a unique lexical entry (*am*). In contrast, the underlying f-structure of the second person singular present-tense auxiliary - [BE PRES 2 SG], has no unique surface realization. Instead, it is realized simply as [BE PRES], i.e. the lexical entry *are*. According to Bresnan's analysis, this is because English disprefers the realization of the person feature [2] more than the person feature [1] at the surface. In OT terms, this means there is a constraint *[2], which is ranked higher than the constraint *[1], interleaving a faithfulness constraint in between.

This type of OT account is unambiguously situated at the level of the lexicon. However, there are other accounts in OT literature which are not explicitly positioned at this level, but nevertheless can be interpreted to belong to it. These seek to explain differences in syntactic categorization across languages, which translate into differences in selectional preferences. For example, Legendre's treatment of the classification of unaccusatives and unergatives in Romance links the subcategorization requirements of verbs to their semantic properties. Thus, the underlying representation is the verb's semantic content, while the surface representation is the associated argument structure [Legendre et al., 1991], [Legendre and Sorace, 2003].

Aissen's account of markedness in subjecthood can be attributed to the same line of inquiry [Aissen, 1999]. According to her analysis, the semantic content of nouns determines their cross-linguistic fitness with respect to the subject position. For example, animate entities are better subjects than inanimate entities, and first and second person pronouns are better subjects than third person. While this may appear to be a generalization at the level of structure, this generalization can easily be transferred to the level of the lexicon. To do so, we can appeal to the fairly standard assumption that subjects are specified in the subcategorization frame of verbs. If so, it is also possible to specify the required feature makeup of a subject at that level. Note that it is not necessary to specify this for each individual verb frame. Instead, a syntactic category of potential subjects can specify which lexical items are globally allowed to fulfill this role.

The second area of OT syntax research which is directly relevant to this dissertation consists of approaches to optimization at the interface level between structure and morphophonology. Unlike the aforementioned OT approaches to the lexicon, this area cannot be easily reconciled with the linearization approach, because it relies on the explicit assumption that structural trade-offs can be prompted by precedence considerations. This assumption runs contrary to the central idea of the proposed framework, which advances a strictly modular approach to a structure-building component and a linearization component.

Recall that the strict modularity assumption is largely due to considerations of theoretical parsimony, and it is not inconceivable that some modification of it may be empirically required. However, I am reluctant to take this step before sufficient evidence is amassed of its necessity. Here, I will sketch some potential problems for strict modularity from the interface literature and suggest a solution which tackles them without violating the modularity of linearization.

The first problem is presented to us by German expletive subjects where an apparent interaction of between the linear form of the utterance, and their presence in the structure is observed. Compare the sentences in (133-136):

(133) Schön wurde getanzt.
Beautifully was danced

(134) *Schön wurde es getanzt.
Beautifully was it danced

(135) Es wurde schön getanzt.
It was beautifully danced

(136) *wurde schön getanzt.
was beautifully danced

Strikingly, different word orders are allowed in the presence and in the absence of the expletive *es* which appears to show that the expletive is inserted as a result of the V2 linearization requirement.

Legendre [Legendre, 2001] explains the above pattern with the interaction of three structural constraints, and two linearization constraints ordering information according to novelty and salience. One structural constraint bars expletives (FullInt), while another structural constraint requires the general presence of subjects (Subj). A third structural

| Input | Order | AlignNtw | AlignNewVP | MinProj | Subj | FullInt |
|---|---|---|---|---|---|---|
| *neutral* | es wurde schön getanzt. | | | | | * |
| | wurde schön getanzt. | | | | * | |
| | schön wurde es getanzt. | | * | * | * | |
| | schön wurde getanzt. | | * | | * | |
| *schön=ntw* | es wurde schön getanzt. | * | | | | * |
| | wurde schön getanzt. | * | | | * | |
| | schön wurde es getanzt. | | * | * | * | |
| | schön wurde getanzt. | | * | | * | |

Table 7.1: Apparent interaction of word order and discourse-motivated linear precedence in German

constraint- MinProj, requires that the highest clausal projection must be IP, not CP. A linearization constraint on salience requires the adverb to appear initially in some discourse context (AlignNoteworthy), and a second linearization constraint (AlignNewVP) requires the adverb to appear at the edge of the VP. As a result, neutral contexts prefer 135, while adverb-salient contexts prefer 133 (Tableau 7.1).

However, it is entirely possible to explain this pattern without violating modularity. Let us assume that the auxiliary verb *wurde* is associated with two subcategorization frames in the lexicon — one frame with an expletive *wurde*[N/ProN, V], another without *wurde*[-, V]. Thus, *wurde* can participate in two possible structures (Figure 7.1). Each of these structures enters the linearization optimization separately, and is evaluated with respect to its own set of candidates. In the case when the expletive is present, $Spc > Cmp$ and $HdEdge$, $*EmptyEdge$ and $HdCmp$ forces SAux-Adv-V word order. The adverb is positioned next to its head — V, and V is to the right of it (and the subject) as required by HdEdge. In the case when the expletive is absent, $*EmptyEdge$ forces the adverb to migrate to the left (Tableau 7.2). Note that under no circumstances would we expect (136) to be grammatical, since it will violate a highly ranked $Spc > Cmp$. However, it is possible

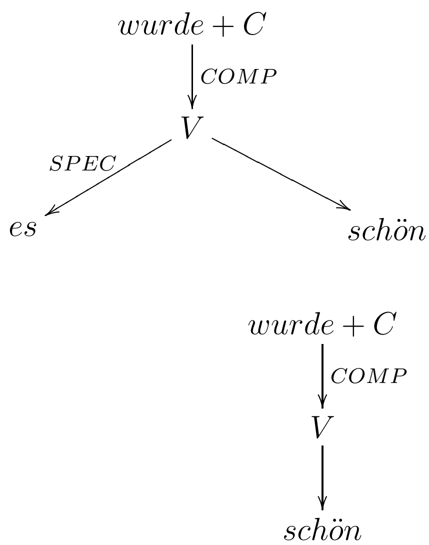



Figure 7.1: Possible structures for *wurde* in German

| Input | Order | *EmptyEdge | HdEdge | HdCmp | Spc>Cmp |
|---|---|---|---|---|---|
| expletive | es wurde schön getanzt. | | | | |
| | schön wurde es getanzt. | | * | | * |
| no expletive | wurde schön getanzt. | * | | | |
| | schön wurde getanzt. | | * | | |

Table 7.2: A modular account of interaction in German

that a full subject could end up sandwiched between the auxiliary and the verb if marked for pre-verbal focus, enforced by a constraint like AlignFocusV. Presumably, an expletive cannot be focus-marked. This analysis differs from Legendre's with respect to at least one prediction: both structures are expected to be able to carry adverb focus, or appear in a neutral context, as long as no relationship between the choice of subcategorization frame and focus-marking is stipulated. . This is definitely true of the expletive structure, which can incorporate focus via intonational emphasis (137).

(137) Es wurde SCHÖN getanzt.
It was BEAUTIFULLY danced

The question is, how true is this in the case of expletive-free structure? We would expect that (133) is a felicitous answer to a neutral question, as in (138),

(138) Was geschah?
What happened?

Indeed, while Legendre assumes this is not true, she points out in a footnote that her informant gives a positive judgement. However, she dismisses it as a "presuppositional", and therefore "unnatural", reading. Without access to more data, and without a thorough examination of the source of presuppositionality, it is difficult to ascertain the validity of the prediction. However, the footnote suggests we might well be on the right track.

A similar case for anti-modularity is provided to us by Bakovic and Keer [Bakovic and Keer, 2001], in their treatment of optional complementizers in English. They point out that contrastive focusing in subordinate clauses can happen only if a complementizer is present (139).

(139) I think that no coat looks good on him.

(140) I think that on him no coat looks good.

(141) I think no coat looks good on him.

(142) *I think on him no coat looks good.

Here again the presence of a structural element appears the be conditioned on linearity.

We can treat this in exactly the same way we treated the German case. Suppose, there are two subcategorization frames for stative verbs in English, one requiring *that*, the other not. Alternatively, there is only one subcategorization frame satisfiable by an empty or full complementizer. The structure of the main plus subordinate clause varies accordingly (Figure 7.2). When contrastive focusing is absent from the subordinate clause, no differences in linearization are apparent. But what happens when contrastive focusing is present?

Suppose the AlignNoteworthy constraint forces the noteworthy element to be aligned with the complementizer (AlignNtwC). When a full complementizer is present and the constrasted element is shifted before the complementizer, $HdEdge$ is violated at least two times — once because of the verb, and once because of the complementizer. If it is shifted after the complementizer, $HdEdge$ is violated once (because of the verb). Thus, contrastive

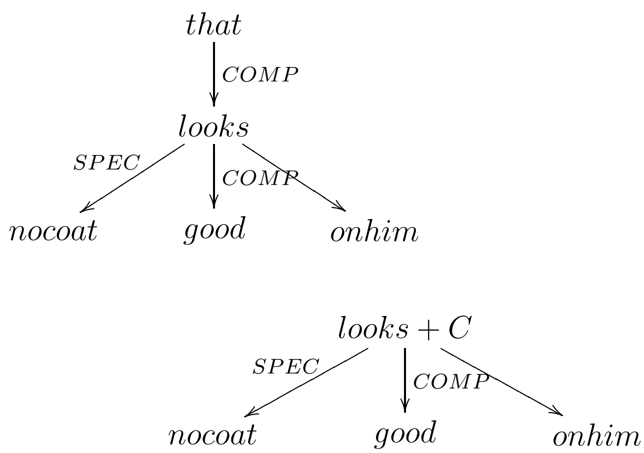



Figure 7.2: Possible structures for stative verbs in English

| Input | Order | Align NtwC | Spc> Cmp | HdCmp | Hd> Cmp | Hd Edge |
|---|---|---|---|---|---|---|
| that | that no coat looks good on him. | * | | | | * |
| | that on him no coat looks good. | | | | *[1] | * |
| | on him that no coat looks good. | | | | * | ** |
| null C | no coat looks+C good on him. | * | | | | ** |
| | on him no coat looks+C good. | * | | | * | ** |

Table 7.3: A modular account of interaction in English subordinate clauses

focusing inside the phrase is preferred. Note however, that when the complementizer is null, this option is not available. Thus, despite tha fact that $HdEdge$ is ranked low in English, as long as it is ranked higher than $AlignNoteworthy$, or a comparable constraint, contrastive focusing will not occur in conjunction with the null complementizer (Tableau 7.3).

However, it is possible for contrastive focusing to occur preverbally, if we include the additional option in Tableau 7.4. The existence of preverbal focus is well known in other languages (cf. Bulgarian, 143).

| Input | Order | *PrevbFoc | Align NtwC | Spc> Cmp | HdCmp Cmp | Hd> Cmp | Hd Edge |
|---|---|---|---|---|---|---|---|
| null C | no coat looks+C good on him. | | * | | | | ** |
| | on him no coat looks+C good. | | * | | | * | ** |
| | no coat on him looks+C good. | | | | | * | ** |

Table 7.4: A modular account of interaction in English subordinates

(143) Georgi kamila kupi (a kon ne).
George camel bought (but horse not). George bought a camel, but not a horse.

What prevents this from happening in English? There are several possibilities. One possibility is that the subject enjoys a priviledged relationship with respect to the verb, with a constraint like SpcCmp. However, not that SpcHd would have to be ranked higher that HdCmp, which goes against our earlier claims for an exclusie relationship between the complement and the verb. A much more attractive possibility is to link the absence of preverbal focus to a latent $V2$ property of English. However, we have so far avoided postulating an explicit $V2$ constraint, relying instead on a combination of $Hd > Cmp$ and $*cEdge$ to derive $V2$-like behavior. Perhaps the best option is to assume that preverbal focus is marked, and consequently languages like English avoid it (Tableau 7.4). Whatever the ultimate explanation is, it appears there are plenty of ways to handle the analysis in modular linearization terms.

In summary, it appears that the approach advocated here is fully compatible with the lexicon-based approaches of previously existing analyses. In so far as conflict exists, it is restricted to the strong interface view of OT which takes phonological movement as a remedy for, or a cause of, structural changes. However, it turns out that at least some arguments for this view do not stand scrutiny. Thus, we can be optimistic about the ability of the modular view to prevail.

# Chapter 8

# Conclusion

The goal of this dissertation was to shed new light on word order variation across and within languages. I hope to have shown that it is both possible and theoretically advantageous to separate completely the structure model from the string it ultimately corresponds to a) because this allows us to preserve the uniformity of structure cross-linguistically and b) because it enables us to model the influence of extra-syntactic factors on word order without reifying them in the syntax itself. It especially is important that my proposal does not require us to abandon the assumption that languages share universal characteristics which are crucial in limiting the hypothesis space of a learner.

Rather than devoting the rest of the conclusion to recaping the details of the argument, I would like to use it as an oportunity to set my work in a broader context. In particular, I will lay out my view of how this work relates to important questions in the cognitive science of language.

During the course of writing this dissertation, I was often called to explain the topic of my research to a non-specialist, and I found that I would invariably start with one thought provoking remark: Have you ever wondered why languages differ?

The fact that languages differ, is obvious, but intellectually surprising in a scientific climate which emphasizes the innate over the learned, and genes over experience. According to the universalist tradition in generative grammar, a language consists of building blocks (words) and recombination rules, partially innate, but otherwise arbitrary.

This view begs the question: if nature took the time to encode some part of grammar into the genome, why did it stop there? Why not encode a full set of grammatical rules and/or lexical entries? Cross-linguistic variation has at least two obvious evolutionary disadvantages: first, a human child must expend resources in order to learn the language of its group, instead of advancing in other areas, like motor skills or navigation; and second, communication with other groups is impeded, often with disasterous effects. No wonder that variation was viewed in the Bible as a punishment bestowed on humanity to impede its progress on the tower of Babel:

> Behold, they are one people, and they have all one language; and this is only the beginning of what they will do; and nothing that they propose to do will now be impossible for them. Come, let us go down, and there confuse their language, that they may not understand one another's speech. (Genesis 11:1-9)

There are two ways of explaning this paradox. One possibility is that variation is an imperfection. In other words, we could assume that evolution hasnt had the time to encode all linguistic knowledge genetically, or that such knowledge cannot be fully innate due to biological design constraints. However, I do not find this line of reasoning particularly appealing on empirical grounds. First, most communication systems which currently exist in other species are fully innate. Vervet monkeys use innate predator calls, corresponding to different sources of danger: the "eagle" cry, "snake" cry and "leopard" cry. [Cheney and Seyfarth, 1990]. Even the communication system with most elaborate syntax-the bee dance, is fully innate. The bee dance is built on combinatorial rules theoretically capable of conveying a non-finite set of meanings [1]. Astonishingly however, bees do not need to learn any part of that system. For example, bees raised in isolation begin to dance as soon as they are united with their tribe. But even more telling is the fact that different lineages of bees have different dances. Italian bees raised in Austrian bee-hives dance off-tune, and neither party learns the others dance. These behavioral experiments strongly suggested a genetic control of bee dance, which was recently discovered [Johnson et al., 2002]. A mutation in a single gene enables an Italian bee to "speak" Austrian bee language. In the

[1] about quantity, direction and distance to food

face of these and numerous other examples from the animal kingdom, it is doubtful that cross-linguistic variation is a result of an evolutionary failure.

Another possibility is that variation constitutes an evolutionary spandrel. In the definition of [Gould and Lewontin, 1979], a spandrel is a feature of an organism that exists as a necessary consequence of other features and is not actually selected for. If a communication system is a mapping of meaning to form, it is possible that particular type of meaning spaces, when paired with paricular types of form necessarily give rise to multiple equally good mappings. Thus, if the role of structures is to represent interpretative dependencies among atomic elements in a high-dimensional space, and the role of strings is to represent structures in a minimally confusing and maximally convenient one-dimensional speech stream, this constitutes a problem with many equally good (or bad) solutions. Hence, cross-linguistic variation is expected in addition to the universal characteristics stemming from the constraints imposed by the problem.

In this general context, it is possible to view dependency structures of the kind presented here as rudimentary structural descriptions. Let us assume that a head with one dependent roughly corresponds to a single-argument function. If human language were designed to express only single-argument functions, and adjacency, along with head-precedence were processing constraints on the speech stream, there is only one, globally optimal solution to linearization: make each head immediately precede its dependent. However, as soon as heads with multiple dependents are introduced, the globally optimal solution to linearization disappears as adjacency enters in conflicts with itself as well as ordering consistency.

Obviously, multiple dependents are crucial for the expressive capacity of human language, as they allow us to efficiently convey an event with multiple participants, or pile up multiple modifiers. The availability of multiple dependents qualifies as a design feature of human language. It is also probably beyond reach for other species, because it implies violations of locality in form, which arguably only humans can process reliably. Thus, in my view, it is the increased complexity of meaning representations that played the role of the tower of Babel in humanity's history by opening the door to word order variation.

# Bibliography


[Abbott, 1991] Abbott, M. (1991). Macushi. In Handbook of Amazonian languages, volume 3, pages 23–160. Mouton de Gruyter, Berlin.

[Ackema et al., ] Ackema, P., Elenbaas, N., Janse, E., Mulders, I., and Visch, J. R. E. Lexicon of linguistics. A project of Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht University - Dutch department, Utrecht University Library.

[Adger, 1996] Adger, D. (1996). Aspect, agreement and measure phrases in scottish gaelic. In Borsley, R. and Roberts, I., editors, The Syntax of the Celtic Languages, pages 200–222. Cambridge University Press.

[Aissen, 1999] Aissen, J. (1999). Markedness and subject choice in optimality theory. Natural Language and Linguistic Theory, 17:673–711.

[Aldridge, 2004] Aldridge, E. (2004). Ergativity and Word Order in Austronesian Languages. PhD thesis, Cornell University.

[Anderson, 1979] Anderson, J. (1979). Syntax and the single mother. Journal of Linguistics, 15:267–287.

[Anderson and Chung, 1977] Anderson, S. and Chung, S. (1977). On grammatical relations and clause structure in verb initial languages. In Cole, P. and Saddock, J., editors, Grammatical Relations. Syntax and Semantics 8. Academic Press, Berlin.

[Anttila, 1997] Anttila, A. (1997). Variation in Finnish Phonology and Morphology. PhD thesis, Stanford University, Stanford, California.

[Awbery, 1976] Awbery, G. (1976). The syntax of Welsh: A transformational study of the passive. Cambridge: Cambridge University Press.

[Bach, 1965] Bach, E. (1965). Control in montague grammar. Linguistic Inquiry, 10:515–531.

[Bach, 1984] Bach, E. (1984). Some generalizations of categorial grammars. In Landman, F. and Veltman, F., editors, Varieties of Formal Semantics. Dordrecht: Foris.

[Baker, 2001] Baker (2001). The atoms of language. New York: Basic Books.

[Bakovic and Keer, 2001] Bakovic, E. and Keer, E. (2001). Optionality and ineffability. In Legendre, G., Grimshaw, J., and Vikner, S., editors, Optimality theoretic syntax. Cambridge, MA: MIT Press.

[Blevins, 1990] Blevins, J. P. (1990). Syntactic complexity: Evidence for discontinuity and multidomination. PhD thesis, University of Massachusetts Amherst. ProQuest document ID: 745125111.

[Bloomfield, 1933] Bloomfield, L. (1933). Language. Holt, New York.

[Bokovic, 2002] Bokovic, . (2002). On multiple wh-fronting. Linguistic Inquiry, 33:351–383.

[Borsley, 2005] Borsley, R. (2005). On the nature of welsh vso clauses. Lingua.

[Bresnan, 2001] Bresnan, J. (2001). The emergence of the unmarked pronoun. In Legendre, G., Grimshaw, J., and Vikner, S., editors, Optimality-theoretic Syntax, pages 113 – 142. Cambridge, MA: The MIT Press.

[Bresnan, 2002] Bresnan, J. (2002). The lexicon in optimality theory. In Stevenson, S. and Merlo, P., editors, The Lexical Basis of Syntactic Processing: Formal,Computational and Experimental Issues, pages 39 – 58. Amsterdam: John Benjamins.

[Brighton, 2002] Brighton, H. (2002). Compositional syntax from cultural transmission. Artificial Life, 8(1):25–54.

[Bromberger and Halle, 1989] Bromberger, S. and Halle, M. (1989). Why phonology is different. Linguistic Inquiry, 20:51–70.

[Buchwald et al., 2002] Buchwald, A., Schwartz, O., Seidl, A., and Smolensky, P. (2002). Recoverability optimality theory: Discourse anaphora and bidirectional optimization: Recoverability optimality theory. In Bos, F. and Matheson, editors, Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002), pages 37–44.

[Bunt, 1996] Bunt, H. (1996). Formal tools for describing and processing discontinuous constituency structure. In Bunt, H. and van Horck, A., editors, Discontinuous Constituency. Natural Language Processing volume 6. Mouton de Gruyter, Berlin.

[Bury, 2000] Bury, D. (2000). Particles, v2 and the ungrammaticality of verb-initial structures. UCL Working Papers in Linguistics, 12:367–386.

[Carnie, 1991] Carnie, A. (1991). Non-Verbal Predication and Head-Movement. PhD thesis, MIT.

[Cheney and Seyfarth, 1990] Cheney, D. L. and Seyfarth, R. M. (1990). How Monkeys See the World: Inside the Mind of Another Species. University of Chicago Press.

[Chomsky, 1955] Chomsky, N. (1955). The logical structure of linguistic theory. Plenum Press, New York, 1975.

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

[Chomsky, 1977a] Chomsky, N. (1977a). Conditions on rules of grammar. In Essays on Form and Interpretation, pages 163–210. Amsterdam.

[Chomsky, 1977b] Chomsky, N. (1977b). On wh-movement. In Formal Syntax, pages 71–132. New York.

[Chomsky, 1995] Chomsky, N. (1995). Minimalist Program. MIT Press.

[Chomsky, 2001] Chomsky, N. (2001). Beyond explanatory adequacy. In Occasional Papers in Linguistics. MIT 20.

[Costa, 1997] Costa, J. (1997). Word order typology in optimality theory. Rutgers Optimality Theory Archive, Number 220-0997.

[Croft, 1990] Croft, W. (1990). Typology and universals. Cambridge: Cambridge University Press.

[Culicover and Nowak, 2003] Culicover, P. and Nowak, A. (2003). Markedness, antisymmetry and complexity of constructions. In Pica, P. and Rooryk, J., editors, Variation Yearbook. John Benjamins, Amsterdam.

[Curry, 1961] Curry, H. B. (1961). Some logical aspects of grammatical structure. In American Mathematical Society Proceedings of the Symposia on Applied Mathematics, volume 12, pages 56–68.

[de Souza, 1994] de Souza, T. (1994). The syntax of an ergative language: Bakairi (carib). In Revista Latinoamericana de Estudios Lingisticos. Peru.

[Derbyshire, 1979] Derbyshire, D. (1979). Hixkaryana syntax. PhD thesis, University of London.

[Derbyshire and Pullum, 1979] Derbyshire, D. and Pullum, G. (1979). Object-initial languages. Work Papers of the Summer Institute of Linguistics, University of North Dakota, 23:1–34.

[Dixon, 1979] Dixon, R. (1979). Ergativity. Language, 55:59–138.

[Dowty, 1995] Dowty, D. (1995). Toward a minimalist theory of syntactic structure. In Bunt, H. and van Horck, A., editors, Discontinuous Constituency. Mouton.

[Dryer, 1997] Dryer, M. (1997). On the 6-way word order typology. Studies in Language, 21:69–103.

[Dryer, 2005a] Dryer, M. (2005a). Order of subject, object, and verb. In Haspelmath, M., Dryer, M., Gil, D., and Comrie, B., editors, The World Atlas of Language Structures. Oxford University Press.

[Dryer, 2005b] Dryer, M. (2005b). Verb-object-negative order in central africa. In Cyffer, N., Ebermann, E., and Ziegelmeyer, G., editors, Negation Patterns in West Africa. Online.

[Eisner, 1996] Eisner, J. (1996). Three new probabilistic models for dependency parsing: An exploration. In COLING, volume 16.

[Epstein et al., 1998] Epstein, S., Groat, E., Kawashima, R., and Kitahara, H. (1998). A Derivational Approach to Syntactic Relations. Oxford University Press: Oxford.

[Erdmann, 1990] Erdmann, P. (1990). German as a sov language. Kwartalnik Neofilologiczny, 37:19–28.

[Fitzpatrick and Groat, 2005] Fitzpatrick, J. and Groat, E. (2005). Movement and c-command as merger: A problem and a solution for multidominance. Presented at the ECO5 Syntax Workshop, March 5-6, 2005, Harvard University, Cambridge, MA.

[Fong and Berwick, 1985] Fong, S. and Berwick, R. C. (1985). New approaches to parsing conjunctions using prolog. In Proceedings of the International Joint Conference on Artificial Intelligence, pages 870–876.

[Frank and Satta, 1998] Frank, R. and Satta, G. (1998). Optimality theory and the generative complexity of constraint violability. Computational Linguistics, 24(2):307–315.

[Gaifman, 1965] Gaifman, H. (1965). Dependency systems and phrase-structure systems. Information and Control, 8:304337.

[Gazdar et al., 1985] Gazdar, G., Klein, E., Pullum, G., and Sag, I. (1985). Generalized Phrase Structure Grammar. Harvard University Press.

[G.Brightwell and P.Winkler, 1991] G.Brightwell and P.Winkler (1991). Counting linear extensions. Order, 8:225–242.

[Gärtner, 1997] Gärtner, H.-M. (1997). General transformations and beyond. PhD thesis, Frankfurt-Main.

[Gärtner, 1999] Gärtner, H.-M. (1999). Phase-linking meets minimalist syntax. In Proceedings of of WCCFL, volume 18.

[Goodall, 1984] Goodall, G. T. (1984). Parallel Structures in Syntax. PhD thesis, University of California San Diego.

[Gordon, 2005] Gordon, Raymond G., J., editor (2005). Ethnologue: Languages of the World. Dallas, Tex.: SIL International, fifteenth edition.

[Gould and Lewontin, 1979] Gould, S. J. and Lewontin, R. C. (1979). The spandrels of san marco and the panglossian paradigm: A critique of the adaptationist programme. In Proceedings Of The Royal Society of London, volume 205 of B, pages 581–598.

[Griffiths and Kalish, 2005] Griffiths, T. L. and Kalish, M. L. (2005). A bayesian view of language evolution by iterated learning. In Proceedings of the 27th Annual Conference of the Cognitive Science Society.

[Grimshaw, 1997] Grimshaw, J. (1997). Projection, heads, and optimality. Linguistic Inquiry, 28:373–422.

[Hale, 1983] Hale, K. (1983). Walpiri and the grammar of non-configurational languages. Natural Language and Linguistic Theory, 1:5–47.

[Hale and Keyser, 1993] Hale, K. and Keyser, J. (1993). On argument structure and the lexical expression of syntactic relations. In Hale, J. and Keyser, S., editors, The View from Building 20. MIT Press, Cambridge MA.

[Harris, 1945] Harris, Z. (1945). Discontinuous morphemes. Language, 21:121–127.

[Hays, 1964] Hays, D. (1964). Dependeny theory: A formalism and some observations. Language, 40:267–287.

[Heim and Kratzer, 1998] Heim, I. and Kratzer, A. (1998). Semantics in Generative Grammar. Oxford: Blackwell.

[Hudson, 1984] Hudson, R. A. (1984). Word Grammar. B. Blackwell, Oxford, UK.

[Hudson, 1991] Hudson, R. A. (1991). English Word Grammar. B. Blackwell, Oxford, UK.

[Hume, 2003] Hume, E. (2003). Language specific markedness: The case of place of articulation. Studies in Phonetics, Phonology and Morphology, 9:295–310.

[Iatridou and Kroch, 1992] Iatridou, S. and Kroch, A. (1992). The licensing of cp-recursion and its relevance to the germanic verb-second phenomenon. Working Papers in Scandinavian Linguistics, 50:1–25.

[Johnson et al., 2002] Johnson, N., Oldroyd, B. P., Barron, A. B., and Crozier, R. H. (2002). Bee dance. Journal of heredity, 93.

[Karlgren, 1976] Karlgren, H. (1976). Why trees in syntax? Studia Linguistica, 8:304337.

[Kathol, 1995] Kathol, A. (1995). Linearization-based German syntax. PhD thesis, Ohio State University. ProQuest document ID: 741166651.

[Kathol, 2000] Kathol, A. (2000). Linear Syntax. Oxford University Press.

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

[Kirby, 2001] Kirby, S. (2001). Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. IEEE Transactions on Evolutionary Computation, 5(2):102–110.

[Kiss, 1998] Kiss, K. (1998). Identificational focus versus information focus. Language, 74:245–273.

[Kracht, 2001] Kracht, M. (2001). Syntax in chains. Linguistics and Philosophy, 24:467529.

[Krantz, 1980] Krantz, G. S. (1980). Sapienization and speech. Current Anthropology, 21:77392.

[Kreps, 1996] Kreps, C. (1996). Dependency, licensing, and the nature of grammatical relations. In UCL Working Papers in Linguistics.

[Lasnik and Kupin, 1977] Lasnik, H. and Kupin, J. (1977). A restrictive theory of transformational grammar. Theoretical Linguistics, 4.

[Legendre, 1996] Legendre, G. (1996). Clitics, verb(non)-movement, and optimality in bulgarian. Technical Report 5, Johns Hopkins University, Department of Cognitive Science.

[Legendre, 1998] Legendre, G. (1998). Morphological and prosodic alignment of bulgarian clitics. In Optimality Theory: Syntax, Phonology and Acquisition. Oxford University Press.

[Legendre, 2001] Legendre, G. (2001). Introduction to optimality theory in syntax. In Legendre, G., Grimshaw, J., and Vikner, S., editors, Optimality theoretic syntax. Cambridge, MA: MIT Press.

[Legendre et al., 1991] Legendre, G., Miyata, Y., and Smolensky, P. (1991). Integrating semantic and syntactic accounts of unaccusativity: A connectionist approach. In Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society. University of California, Berkeley.

[Legendre et al., 1998] Legendre, G., Smolensky, P., and Wilson, C. (1998). When is less more? faithfulness and minimal links in wh- chains. In Barbosa, P., Fox, D., Hagstrom, P., McGinnis, M., and Pesetsky, D., editors, Is the Best Good Enough? Optimality and Competition in Syntax, pages 249–289. MIT Press.

[Legendre and Sorace, 2003] Legendre, G. and Sorace, A. (2003). Split intransitivity in french: An optimality-theoretic perspective. In Godard, D., editor, Les langues romanes: problmes de la phrase simple. Paris: CNRS ditions.

[Lehmann, 1973] Lehmann, W. (1973). A structural principle of language and its implications. Language, 49:47–66.

[Loos et al., 2004] Loos et al. (2004). Glossary of linguistic terms.

[Mathieu, 2004] Mathieu, E. (2004). The mapping of form and interpretation: the case of optional wh-movement in french. Lingua, 114:1090–1132.

[McCawley, 1982] McCawley, J. (1982). Parentheticals and discontinuous constituent structure. Linguistic Inquiry, 13:91–106.

[McCoskey, 1979] McCoskey, J. (1979). Transformational syntax and model theoretic semantics. Dordrecht:Reidel.

[McCoskey, 1980] McCoskey, J. (1980). Is there raising in modern irish? Eriu, 31:59–99.

[Meinunger, 2003] Meinunger, A. (2003). Symmetries and asymmetries of wh- and focus dependencies: the algebraic difference. In Maienborn, C., editor, (A)Symmetrien - (A)Symmetries, Beitrge zu Ehren von E. Lang, page 267 282. Stauffenberg Verlag Linguistik.

[Merlan, 1982] Merlan, F. (1982). Mangarayi. Lingua Descriptive Series, 4.

[Montapanyane, 1997] Montapanyane, V. (1997). Preverbal focus in bulgarian. Journal of Slavic Linguistics, 5:265–299.

[Newmeyer, 2004a] Newmeyer, F. (2004a). Against a parameter-setting approach to language variation. In Pica, P., Rooryck, J., and van Craenenbroek, J., editors, Language Variation Yearbook, volume 4, pages 181–234. Amsterdam: Benjamins.

[Newmeyer, 2004b] Newmeyer, F. J. (2004b). Word order and parameterized grammars: A critical look and explanations of word order. Handout of talks given at the University of Washington and Institut des Sciences Cognitives (Lyon) and the Johns Hopkins University.

[Norris, 1997] Norris, J. R. (1997). Markov Chains. Cambridge University Press.

[Nunes, 2004] Nunes, J. (2004). Linearization of Chains and Sideward Movement. MIT Press.

[Ojeda, 1988] Ojeda, A. (1988). A linear precedence account of cross-serial dependencies. Linguistics and Philosophy, 11:457–492.

[Ojeda, 2005] Ojeda, A. (2005). Discontinuous constituents. In Encyclopedia of language and linguistics. Elsevier. forthcoming.

[Osborn, 1966] Osborn, H. (1966). Warao i: Phonology and morphophonemics. International Journal of American Linguistics, 32.

[Petkevic, 1987] Petkevic, V. (1987). A new dependency based specification. Theoretical Linguistics, 14:143–172.

[Pike, 1943] Pike, K. (1943). Taxemes and immediate constituents. Language, 19:65–82.

[Prince and Smolensky, 1993] Prince, A. and Smolensky, P. (1993). Optimality theory: Constraint interaction in generative grammar. Technical report, Rutgers University and University of Colorado at Boulder. Revised version published by Blackwell in 2004.

[Pullum, 1982] Pullum, G. (1982). Free word order and phrase structure rules. In Proceedings of NELS, volume 12.

[Pullum, 1984] Pullum, G. (1984). On two recent attempts to show that english is not a cfl. Computational Linguistics, 10:182–188.

[Richards, 1997] Richards, N. (1997). What moves where when in which language? PhD thesis, Massachusetts Institute of Technology, Cambridge.

[Robinson, 1970] Robinson, J. (1970). Dependency structures and transformation rules. Language, 46:259–285.

[Romero-Figeroa, 1997] Romero-Figeroa, A. (1997). A Reference Grammar of Warao. Mnchen, Newcastle: Lincom.

[Ross, 1967] Ross, J. (1967). Constraints on variables in syntax. PhD thesis, MIT.

[Rudin, 1988] Rudin, C. (1988). On multiple questions and multiple wh-fronting. Natural Language and Linguistic Theory, 6:445–501.

[Rumsey, 1982] Rumsey, A. (1982). An intra-sentence grammar of Ungarinjin, North-Western Australia. Canberra: Pacific Linguistics B-86.

[Ruskey, 2003] Ruskey, F. (2003). The (combinatorial) object server.

[Sampson, 1975] Sampson, G. (1975). The single mother condition. Journal of Linguistics, 11:1–181.

[Santorini, 1992] Santorini, B. (1992). Variation and change in yiddish subordinate clause word order. Natural Language and Linguistic Theory, 10:595–640.

[Savova, 2002] Savova, V. (2002). Towards a syntactic account of clitics: Remnant movement and sentential clitics in bulgarian. In Proceedings of the West Coast Linguistics Conference.

[Schuh, 1998] Schuh, R. (1998). A Grammar of Miya. University of California Press.

[Schwartz, 1972] Schwartz, A. (1972). The vp constituent of svo languages. In Kimball, J., editor, Tree Adjoining Grammars: Formalisms, Linguistic Analysis and Processing, volume 1 of Syntax and semantics, pages 213–235. New York: Seminar Press.

[Sgall, 1992] Sgall, P. (1992). Underlying structure of sentences and its relations to semantics. In Wiener Slawistischer Almanach, volume 30, pages 349–368. Sonderband.

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

[Siewierska, 1988] Siewierska, A. (1988). Word Order Rules. London: Croom Helm.

[Silverstein, 1976] Silverstein, M. (1976). Hierarchy of features and ergativity. In Dixon, R. M. W., editor, Grammatical Categories in Australian Languages. New Jersey: Humanities Press.

[Sleator and Temperley, 1993] Sleator, D. and Temperley, D. (1993). Parsing english with a link grammar. In Third international workshop on parsing technologies.

[Smith, 2003] Smith, K. (2003). Learning biases and language evolution. In Kirby, S., editor, Proceedings of Language Evolution and Computation Workshop/Course at ESSLLI, pages 22–31, Vienna.

[Starke, 2001] Starke, M. (2001). Move dissolves into merge: A theory of locality. Ms., University of Geneva.

[Stenson, 1981] Stenson, N. (1981). Studies in Irish syntax. Gunter Narr: Tbingen.

[Tallerman, 1990] Tallerman, M. (1990). Vso word order and consonant mutation in welsh. Linguistics, 28:398–416.

[Tesniere, 1959] Tesniere, L. (1959). Elements de Syntaxe Structurals. Paris: Klincksieck.

[Tomlin, 1986] Tomlin, R. S. (1986). Basic Word Order: Functional Principles. Croom Helm, London.

[Wells, 1947] Wells, R. (1947). Immediate constituents. Language, 23:81–117.

[Wheatley, 1973] Wheatley, J. (1973). Knowledge, authority and individualism among the cura (bacairi). Anthropological Linguistics, 15:337–344.

[Wilder, 1999] Wilder, C. (1999). Right node raising and the lca. In Proceedings of WCCFL, volume 18.

[WURM and HATTORI, 1981] WURM, S. A. and HATTORI, S. (1981). Language atlas of the Pacific area - Part 1 (including Polynesian languages). Canberra: Australian Academy of the Humanities.

[Yngve, 1960] Yngve, V. H. (1960). Model and an hypothesis for language structure. In Proceedings of the American Philosophical Society.

[Zepter, 2003] Zepter, A. (2003). Phrase-structure direction: Having a few choices. PhD thesis, Rutgers University.

# Vita

Virginia Savova **Education**:

2003 - Present Johns Hopkins University, Baltimore, MD

Ph.D. candidate, Department of Cognitive Science

2000 - 2003 Johns Hopkins University, Baltimore, MD

M.A., Department of Cognitive Science

1995 - 1999 Harvard University, Cambridge, MA

A.B. cum laude, Linguistics and Classical Archaeology

1990 - 1995 National Classical Lyceum Sofia, Bulgaria

Summa cum laude

**Research experience**:

2000-Present Johns Hopkins University, Baltimore, MD

IGERT Associate researcher

1999 - 2000 Cognitive Neuropsychology Laboratory

Harvard University, Cambridge, MA

Research assistant

Topic: Lexical access.

1995 - 1999 Smithsonian Center for Astrophysics

Harvard University, Cambridge, MA

Research assistant: Solar Astrophysics group.

1997 - 1999 American University in Bulgaria, Blagoevgrad, Bulgaria

Field supervisor of an archeological site.

1998 Biological Labs

Harvard University Cambridge, MA

Laboratory assistant: Visual system of zebra fish.

**Selected publications**:

Bayesian Networks for the Recursive recovery of Syntactic Dependencies. Virginia Savova and Leon Peshkin. In Proceedings of CogSci, 2005.

Dependency Parsing with Dynamic Bayesian Network. Virginia Savova and Leon Peshkin. In Proceedings of AAAI, 2005.

Tree-Adjoining Grammars for Optimality Theory Syntax. Virginia Savova and Robert Frank. Proceedings of TAG+7, 2004.

Bayesian Nets for Syntactic Categorization of Novel Words. Leon Peshkin, Avi Pfeffer and Virginia Savova. Proceedings of the NAACL, 2003.

Part-of-Speech Tagging with Minimal Lexicalization. Virginia Savova and Leon Peshkin. Current Issues in Linguistic Theory: Recent Advances in Natural Language Processing. John Benjamins Publishers.

Towards a Syntactic Account of Clitics: Remnant Movement and Sentential Clitics in Bulgarian. Virginia Savova. Proceedings of the West Coast Linguistics Conference, 2002.

Level of categorization effect: a novel effect in the picture-word interference paradigm. Albert Costa, Bradford Mahon, Virginia Savova, Alfonso Caramazza. Language and Cognitive Processes 2003.

**Teaching Assistantships**:

050.691 Bilinguality (Isabelle Barriere) - Fall 2003

050.381 Language acquisition (Isabelle Barriere) - Spring 2003

050.381 Language development (Isabelle Barriere) - Intersession 2004

050.620 Introduction to Syntax of Natural Languages (Graldine Legendre) - Fall 2002

050.203 Developmental Cognitive Neuroscience (Barbara Landau) - Spring 2001

050.109 Minds, Brains and Computers (Paul Smolensky) - Spring 2001

050.102 Language and Mind (William Badecker) - Fall 2000

**Languages**:

Bulgarian (native), Russian (very good), German (good reading, some conversational), French (reading), Italian (reading), Ancient Greek, Latin (excellent reading).

**Awards**:

1997-98 Harvard College Scholarship

1996-97 John Harvard Scholarship

1996-97 Elizabeth Cary Agassiz Scholarship

1995-96 Harvard College Scholarship

1995 Special Prize at International Latin Competition, Italy

1995 First prize on the National Greek and Latin Olympiad