13 The innateness of linguistic knowledge

If we try and compare the theories presented in this book, we notice that there are a number of similarities.[1] In all of the frameworks, there are variants of theories that use feature-value pairs to describe linguistic objects. The syntactic structures assumed are sometimes similar. Nevertheless, there are some differences that have often led to fierce debates between members of the various schools. Theories differ with regard to whether they assume transformations, empty elements, phrasal or lexical analyses, binary branching or flat structures.

Every theory has to not only describe natural language, but also explain it. It is possible to formulate an infinite number of grammars that license structures for a given language (see Exercise 1 on page 80). These grammars are *observationally adequate*. A grammar achieves *descriptive adequacy* if it corresponds to observations and the intuitions of native speakers.[2] A linguistic theory is descriptively adequate if it can be used to formulate a descriptively adequate grammar for every natural language. However, grammars achieving descriptive adequacy do not always necessarily reach *explanatory adequacy*. Grammars that achieve explanatory adequacy are those that are compatible with acquisition data, that is, grammars that could plausibly be acquired by human speakers (Chomsky 1965: 24–25).

Chomsky (1965: 25) assumes that children already have domain-specific knowledge about what grammars could, in principle, look like and then extract information about what a given grammar actually looks like from the linguistic input. The most prominent

[1] The terms *theory* and *framework* may require clarification. A framework is a common set of assumptions and tools that is used when theories are formulated. In this book, I discussed theories of German. These theories were developed in certain frameworks (GB, GPSG, HPSG, LFG, …) and of course there are other theories of other languages that share the same fundamental assumptions. These theories differ from the theories of German presented here but are formulated in the same framework. Haspelmath (2010b) argues for framework-free grammatical theory. If grammatical theories used incompatible tools, it would be difficult to compare languages. So assuming transformations for English nonlocal dependencies and a SLASH mechanism for German would make comparison impossible. I agree with Haspelmath that the availability of formal tools may lead to biases, but in the end the facts have to be described somehow. If nothing is shared between theories, we end up with isolated theories formulated in one man frameworks. If there *is* shared vocabulary and if there are standards for doing framework-free grammatical theory, then the framework is framework-free grammatical theory. See Müller (2015a) and Chapter 22 of this book for further discussion.

[2] This term is not particularly useful as subjective factors play a role. Not everybody finds grammatical theories intuitively correct where it is assumed that every observed order in the languages of the world has to be derived from a common Specifier-Head-Complement configuration, and also only by movement to the left (see Section 4.6.1 for the discussion of such proposals).

variant of acquisition theory in Mainstream Generative Grammar (MGG) is the Principles & Parameters theory, which claims that parameterized principles restrict the grammatical structures possible and children just have to set parameters during language acquisition (see Section 3.1.2).

Over the years, the innateness hypothesis, also known as nativism, has undergone a number of modifications. In particular, assumptions about exactly what forms part of the innate linguistic knowledge, so-called Universal Grammar (UG), have often been subject to change.

Nativism is often rejected by proponents of Construction Grammar, Cognitive Grammar and by many other researchers working in other theories. Other explanations are offered for the facts normally used to argue for the innateness of grammatical categories, syntactic structures or relations between linguistic objects in syntactic structures. Another point of criticism is that the actual complexity of analyses is blurred by the fact that many of the stipulations are simply assumed to be part of UG. The following is a caricature of a certain kind of argumentation in GB/Minimalism analyses:

1. I have developed an analysis for the phenomenon P in the language S.
2. The analysis is elegant/conceptually simple/mine[3].
3. There is no possibility to learn the relevant structures or principles.
4. Therefore, the assumptions A_1 through A_n that are made in this analysis must be part of the innate knowledge of speakers.

By attributing arbitrary assumptions to UG, it is possible to keep the rest of the analysis very simple.

The following section will briefly review some of the arguments for language-specific innate knowledge. We will see that none of these arguments are uncontroversial. In the following chapters, I will discuss fundamental questions about the architecture of grammar, the distinction between competence and performance and how to model performance phenomena, the theory of language acquisition as well as other controversial questions, e. g. whether it is desirable to postulate empty elements in linguistic representations and whether language should be explained primarily based on the properties of words or rather phrasal patterns.

Before we turn to these hotly debated topics, I want to discuss the one that is most fiercely debated, namely the question of innate linguistic knowledge. In the literature, one finds the following arguments for innate knowledge:

- the existence of syntactic universals,
- the speed of acquisition,
- the fact that there is a 'critical period' for language acquisition,

[3] Also, see http://www.youtube.com/watch?v=cAYDiPizDIs. 01.12.2015.

- the fact that all children learn a language, but primates do not,
- the fact that children spontaneously regularize pidgin languages,
- the localization of language processing in particular parts of the brain,
- the alleged dissociation of language and general cognition:
  - Williams Syndrome,
  - the KE family with FoxP2 mutation and
- the Poverty of the Stimulus Argument.

Pinker (1994) offers a nice overview of these arguments. Tomasello (1995) provides a critical review of this book. The individual points will be discussed in what follows.

13.1 Syntactic universals

The existence of syntactic universals has been taken as an argument for the innateness of linguistic knowledge (e. g. Chomsky 1998: 33; Pinker 1994: 237–238). There are varying claims in the literature with regard to what is universal and language-specific. The most prominent candidates for universals are:[4]

- the Head Directionality Parameter
- $\overline{\mathrm{X}}$ structures
- grammatical functions such as subject or object
- binding principles
- properties of long-distance dependencies
- grammatical morphemes for tense, mood and aspect
- parts of speech
- recursion or self-embedding

[4] Frans Plank has an archive of universals in Konstanz (Plank & Filimonova 2000): http://typo.uni-konstanz.de/archive/intro/. On 23.12.2015, it contained 2029 entries. The entries are annotated with regard to their quality, and it turns out that many of the universals are statistical universals, that is, they hold for the overwhelming majority of languages, but there are some exceptions. Some of the universals are marked as almost absolute, that is, very few exceptions are known. 1153 were marked as absolute or absolute with a question mark. 1021 of these are marked as absolute without a question mark. Many of the universals captured are implicational universals, that is, they have the form: if a language has the property X, then it also has the property Y. The universals listed in the archive are, in part, very specific and refer to the diachronic development of particular grammatical properties. For example, the fourth entry states that: *If the exponent of vocative is a prefix, then this prefix has arisen from 1st person possessor or a 2nd person subject.*

These supposed universals will each be discussed briefly in what follows. One should emphasize that there is by no means a consensus that these are universal and that the observed properties actually require postulating innate linguistic knowledge.

13.1.1 Head Directionality Parameter

The Head Directionality Parameter was already introduced in Section 3.1.2. The examples in (7) on page 89, repeated below as (1), show that the structures in Japanese are the mirror image of the English structures:

(1) a. be showing pictures of himself

b. zibun -no syasin-o mise-te iru
himself of picture showing be

In order to capture these facts, a parameter was proposed that is responsible for the position of the head relative to its arguments (e. g. Chomsky 1986b: 146; 1988: 70).

By assuming a Head Directionality Parameter, Radford (1990: 60–61; 1997: 19–22), Pinker (1994: 234, 238), Baker (2003: 350) and other authors claim, either explicitly or implicitly, that there is a correlation between the direction of government of verbs and that of adpositions, that is, languages with verb-final order have postpositions and languages with VO order have prepositions. This claim can be illustrated with the language pair English/Japanese and the examples in (1): the *no* occurs after the pronoun in the prepositional phrase, the noun *syasin-o* 'picture' follows the PP belonging to it, the main verb follows its object and the auxiliary *iru* occurs after the main verb *mise-te*. The individual phrases are the exact mirror image of the respective phrases in English.

A single counter example is enough to disprove a universal claim and in fact, it is possible to find a language that has verb-final order but nevertheless has prepositions. Persian is such a language. An example is given in (2):

(2) man ketâb-â-ro be Sepide dâd-am.
I book-PL-RÂ to Sepide gave-1SG
'I gave the books to Sepide.'

In Section 3.1.4, it was shown that German cannot be easily described with this parameter: German is a verb-final language but has both prepositions and postpositions. The World Atlas of Language Structures lists 41 languages with VO order and postpositions and 14 languages with OV order and prepositions (Dryer 2013b,a).[5]

Furthermore, Gibson & Wexler (1994: 422) point out that a single parameter for the position of heads would not be enough since complementizers in both English and German/Dutch occur before their complements, however, English is a VO language, whereas German and Dutch count as OV languages.

If one wishes to determine the direction of government based on syntactic categories (Gibson & Wexler 1994: 422, Chomsky 2005: 15), then one has to assume that the syn-

[5] http://wals.info/combinations/83A_85A#2/15.0/153.0, 23.12.2015.

tactic categories in question belong to the inventory of Universal Grammar (see Section 13.1.7, for more on this). Difficulties with prepositions and postpositions also arise for this kind of assumption as these are normally assigned to the same category (P). If we were to introduce special categories for both prepositions and postpositions, then a four-way division of parts of speech like the one on page 96 would no longer be possible. One would instead require a binary feature in addition and one would thereby automatically predict eight categories although only five (the four commonly assumed plus an extra one) are actually needed.

One can see that the relation between direction of government that Pinker formulated as a universal claim is in fact correct but rather as a tendency than as a strict rule, that is, there are many languages where there is a correlation between the use of prepositions or postpositions and the position the verb.[6]

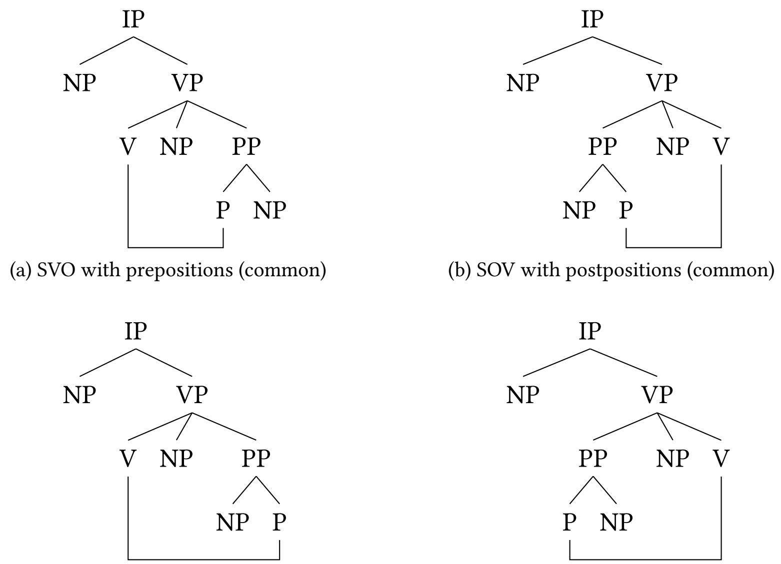
In many languages, adpositions have evolved from verbs. In Chinese grammar, it is commonplace to refer to a particular class of words as coverbs. These are words that can be used both as prepositions and as verbs. If we view languages historically, then we can find explanations for these tendencies that do not have to make reference to innate linguistic knowledge (see Evans & Levinson 2009a: 445).

Furthermore, it is possible to explain the correlations with reference to processing preferences: in languages with the same direction of government, the distance between the verb and the pre-/postposition is less (Figure 13.1a–b) than in languages with differing directions of government (Figure 13.1c–d). From the point of view of processing, languages with the same direction of government should be preferred since they allow the hearer to better identify the parts of the verb phrase (Newmeyer (2004a: 219–221) cites Hawkins (2004: 32) with a relevant general processing preference). This tendency can thus be explained as the grammaticalization of a performance preference (see Chapter 15 for the distinction between competence and performance) and recourse to innate language-specific knowledge is not necessary.

13.1.2 $\overline{\text{X}}$ structures

It is often assumed that all languages have syntactic structures that correspond to the $\overline{\text{X}}$ schema (see Section 2.5) (Pinker 1994: 238; Meisel 1995: 11, 14; Pinker & Jackendoff 2005: 216). There are, however, languages such as Dyirbal (Australia) where it does not seem to make sense to assume hierarchical structure for sentences. Thus, Bresnan (2001: 110) assumes that Tagalog, Hungarian, Malayalam, Warlpiri, Jiwarli, Wambaya, Jakaltek and other corresponding languages do not have a VP node, but rather a rule taking the form of (3):

[6] Pinker (1994: 234) uses the word *usually* in his formulation. He thereby implies that there are exceptions and that the correlation between the ordering of adpositions and the direction of government of verbs is actually a tendency rather than a universally applicable rule. However, in the pages that follow, he argues that the Head Directionality Parameter forms part of innate linguistic knowledge. Travis (1984: 55) discusses data from Mandarin Chinese that do not correspond to the correlations she assumes. She then proposes treating the Head Directionality Parameter as a kind of Default Parameter that can be overridden by other constraints in the language.



(a) SVO with prepositions (common)

(b) SOV with postpositions (common)

(c) SVO with postpositions (rare)

(d) SOV with prepositions (rare)

Figure 13.1: Distance between verb and preposition for various head orders according to Newmeyer (2004a: 221)

(3) S → C*

Here, C* stands for an arbitrary number of constituents and there is no head in the structure. Other examples for structures without heads will be discussed in Section 21.10.

$\overline{\text{X}}$ structure was introduced to restrict the form of possible rules. The assumption was that these restrictions reduce the class of grammars one can formulate and thus – according to the assumption – make the grammars easier to acquire. But as Kornai & Pullum (1990) have shown, the assumption of $\overline{\text{X}}$ structures does not lead to a restriction with regard to the number of possible grammars if one allows for empty heads. In GB, a number of null heads were used and in the Minimalist Program, there has been a significant increase of these. For example, the rule in (3) can be reformulated as follows:

(4) $\text{V}' \rightarrow \text{V}^0\ \text{C}^*$

Here, V^0 is an empty head. Since specifiers are optional, V′ can be projected to VP and we arrive at a structure corresponding to the $\overline{\text{X}}$ schema.

Apart from the problem with languages with very free constituent order, there are further problems with adjunction structures: Chomsky's analysis of adjective structure in $\overline{\text{X}}$ theory (Chomsky 1970: 210; see also Section 2.5 of this book, in particular Figure 2.8 on page 77) is not straightforwardly applicable to German since, unlike English, adjective phrases in German are head-final and degree modifiers must directly precede the adjective:

(5) a. der auf seinen Sohn sehr stolze Mann
the of his son very proud man
'the man very proud of his son'

b. * der sehr auf seinen Sohn stolze Mann
the very of his son proud man

c. * der auf seinen Sohn stolze sehr Mann
the of his son proud very man

Following the $\overline{\text{X}}$ schema, *auf seinen Sohn* has to be combined with *stolze* and only then can the resulting $\overline{\text{A}}$ projection be combined with its specifier (see Figure 2.8 on page 77 for the structure of adjective phrases in English). It is therefore only possible to derive orders such as (5b) or (5c). Neither of these is possible in German. It is only possible to rescue the $\overline{\text{X}}$ schema if one assumes that German is exactly like English and, for some reason, the complements of adjectives must be moved to the left. If we allow these kind of repair approaches, then of course any language can be described using the $\overline{\text{X}}$ schema. The result would be that one would have to postulate a vast number of movement rules for many languages and this would be extremely complex and difficult to motivate from a psycholinguistic perspective. See Chapter 15 for grammars compatible with performance.

A further problem for $\overline{\text{X}}$ theory in its strictest form as presented in Section 2.5 is posed by so-called hydra clauses (Perlmutter & Ross 1970; Link 1984; Kiss 2005):

(6) a. [[der Kater] und [die Katze]], die einander lieben
the tomcat and the cat that each.other love
'the tomcat and the (female) cat that love each other'

b. [[The boy] and [the girl]] who dated each other are friends of mine.

Since the relative clauses in (6) refer to a group of referents, they can only attach to the result of the coordination. The entire coordination is an NP, however, and adjuncts should actually be attached at the $\overline{\text{X}}$ level. The reverse case of relative clauses in German and English is posed by adjectives in Persian: Samvelian (2007) argues for an analysis where adjectives are combined with nouns directly, and only the combination of nouns and adjectives is then combined with a PP argument.

The discussion of German and English shows that the introduction of specifiers and adjuncts cannot be restricted to particular projection levels, and the preceding discussion of non-configurational languages has shown that the assumption of intermediate levels does not make sense for every language.

It should also be noted that Chomsky himself assumed in 1970 that languages can deviate from the $\overline{\text{X}}$ schema (1970: 210).

If one is willing to encode all information about combination in the lexicon, then one could get by with very abstract combinatorial rules that would hold universally. An example of this kind of combinatorial rules is the multiplication rules of Categorial

Grammar (see Chapter 8) as well as Merge in the Minimalist Program (see Section 4). The rules in question simply state that two linguistic objects are combined. These kinds of combination of course exist in every language. With completely lexicalized grammars, however, it is only possible to describe languages if one allows for null heads and makes certain ad hoc assumptions. This will be discussed in Section 21.10.

13.1.3 Grammatical functions such as subject and object

Bresnan & Kaplan (1982: xxv), Pinker (1994: 236–237), Baker (2003: 349) and others assume that all languages have subjects and objects. In order to determine what exactly this claim means, we have to explore the terms themselves. For most European languages, it is easy to say what a subject and an object is (see Section 1.7), however, it has been argued that it is not possible for all languages or that it does not make sense to use these terms at all (Croft 2001: Chapter 4; Evans & Levinson 2009a: Section 4).

In theories such as LFG – the one in which Pinker worked – grammatical functions play a primary role. The fact that it is still controversial whether one should view sentences as subjects, objects or as specially defined sentential arguments (XCOMP) (Dalrymple & Lødrup 2000; Berman 2003b, 2007; Alsina, Mohanan & Mohanan 2005; Forst 2006) serves to show that there is at least some leeway for argumentation when it comes to assigning grammatical functions to arguments. It is therefore likely that one can find an assignment of grammatical functions to the arguments of a functor in all languages.

Unlike LFG, grammatical functions are irrelevant in GB (see Williams 1984; Sternefeld 1985a) and Categorial Grammar. In GB, grammatical functions can only be determined indirectly by making reference to tree positions. Thus, in the approach discussed in Chapter 3, the subject is the phrase in the specifier position of IP.

In later versions of Chomskyan linguistics, there are functional nodes that seem to correspond to grammatical functions (AgrS, AgrO, AgrIO, see page 147). However, Chomsky (1995b: Section 4.10.1) remarks that these functional categories were only assumed for theory internal reasons and should be removed from the inventory of categories that are assumed to be part of UG. See Haider (1997a) and Sternefeld (2006: 509–510) for a description of German that does without functional projections that cannot be motivated in the language in question.

The position taken by HPSG is somewhere in the middle: a special valence feature is used for subjects (in grammars of German, there is a head feature that contains a representation of the subject for non-finite verb forms). However, the value of the SUBJ feature is derived from more general theoretical considerations: in German, the least oblique element with structural case is the subject (Müller 2002a: 153; Müller 2007b: 311).

In GB theory (Extended Projection Principle, EPP, Chomsky (1982: 10)) and also in LFG (Subject Condition), there are principles ensuring that every sentence must have a subject. It is usually assumed that these principles hold universally.[7]

[7] However, Chomsky (1981a: 27) allows for languages not to have a subject. He assumes that this is handled by a parameter. Bresnan (2001: 311) formulates the Subject Condition, but mentions in a footnote that it

As previously mentioned, there are no grammatical functions in GB, but there are structural positions that correspond to grammatical functions. The position corresponding to the subject is the specifier of IP. The EPP states that there must be an element in SpecIP. If we assume universality of this principle, then every language must have an element in this position. As we have already seen, there is a counterexample to this universal claim: German. German has an impersonal passive (7a) and there are also subjectless verbs (7b,c) and adjectives (7d–f).[8]

(7) a. dass noch gearbeitet wird
that still worked is
'that people are still working'

b. Ihm graut vor der Prüfung.
him.DAT dreads before the exam
'He dreads the exam.'

c. Mich friert.
me.ACC freezes
'I am freezing.'

d. weil schulfrei ist
because school.free is
'because there is no school today'

e. weil ihm schlecht ist
because him.DAT ill is
'because he is not feeling well'

f. Für dich ist immer offen.[9]
for you is always open
'We are always open for you.'

Most of the predicates that can be used without subjects can also be used with an expletive subject. An example is given in(8):

(8) dass es ihm vor der Prüfung graut
that EXPL him before the exam dreads
'He dreads the exam.'

However, there are verbs such as *liegen* 'lie' in example (9a) from Reis (1982: 185) that cannot occur with an *es* 'it'.

(9) a. Mir liegt an diesem Plan.
me.DAT lies on this plan
'This plan matters a lot to me.'

might be necessary to parameterize this condition so that it only holds for certain languages.

[8] For further discussion of subjectless verbs in German, see Haider (1993: Sections 6.2.1, 6.5), Fanselow (2000b), Nerbonne (1986b: 912) and Müller (2007b: Section 3.2).

[9] Haider (1986a: 18).

b. * Mir liegt es an diesem Plan.
me.DAT lies it on this plan

Nevertheless, the applicability of the EPP and the Subject Condition is sometimes also assumed for German. Grewendorf (1993: 1311) assumes that there is an empty expletive that fills the subject position of subjectless constructions.

Berman (1999: 11; 2003a: Chapter 4), working in LFG, assumes that verbal morphology can fulfill the subject role in German and therefore even in sentences where no subject is overtly present, the position for the subject is filled in the f-structure. A constraint stating that all f-structures without a PRED value must be third person singular applies to the f-structure of the unexpressed subject. The agreement information in the finite verb has to match the information in the f-structure of the unexpressed subject and hence the verbal inflection in subjectless constructions is restricted to be 3rd person singular (Berman 1999).

As we saw on page 167, some researchers working in the Minimalist Program even assume that there is an object in every sentence (Stabler quoted in Veenstra (1998: 61, 124)). Objects of mono-valent verbs are assumed to be empty elements.

If we allow these kinds of tools, then it is of course easy to maintain the existence of many universals: we claim that a language X has the property Y and then assume that the structural items are invisible and have no meaning. These analyses can only be justified theory-internally with the goal of uniformity (see Culicover & Jackendoff 2005: Section 2.1.2).[10]

13.1.4 Binding principles

The principles governing the binding of pronouns are also assumed to be part of UG (Chomsky 1998: 33; Crain, Thornton & Khlentzos 2009: 146; Rizzi 2009a: 468). Binding Theory in GB theory has three principles: principle A states that reflexives such as *sich* or *himself* refer to an element (antecedent) inside of a certain local domain (binding domain). Simplyfying a bit, one could say that a reflexive has to refer to a co-argument.

(10) Klaus$_i$ sagt, dass Peter$_j$ sich$_{*i/j}$ rasiert hat.
Klaus says that Peter himself shaved has

Principle B holds for personal pronouns and states that these cannot refer to elements inside of their binding domain.

(11) Klaus$_i$ sagt, dass Peter$_j$ ihn$_{i/*j}$ rasiert hat.
Klaus says that Peter him shaved has

Principle C determines what referential expressions can refer to. According to Principle C, an expression A_1 cannot refer to another expression A_2 if A_2 c-commands A_1. c-command is defined with reference to the structure of the utterance. There are various definitions of c-command; a simple version states that A c-commands B if there is a path

[10] For arguments from language acquisition, see Chapter 16.

in the constituent structure that goes upwards from A to the next branching node and then only downwards to B.

For the example in (12a), this means that *Max* and *er* 'he' cannot refer to the same individual since *er* c-commands *Max*.

(12) a. Er sagt, dass Max Brause getrunken hat.
he says that Max soda drunk has
'He said that Max drank soda.'

b. Max sagt, dass er Brause getrunken hat.
Max said that he soda drunk has
'Max said that he drank soda.'

c. Als er hereinkam, trank Max Brause.
as he came.in drank Max soda
'As he came in, Max was drinking soda.'

This is possible in (12b), however, as there is no such c-command relation. For *er* 'he', it must only be the case that it does not refer to another argument of the verb *getrunken* 'drunk' and this is indeed the case in (12b). Similarly, there is no c-command relation between *er* 'he' and *Max* in (12c) since the pronoun *er* is inside a complex structure. *er* 'he' and *Max* can therefore refer to the the same or different individuals in (12b) and (12c).

Crain, Thornton & Khlentzos (2009: 147) point out that (12b,c) and the corresponding English examples are ambiguous, whereas (12a) is not, due to Principle C. This means that one reading is not available. In order to acquire the correct binding principles, the learner would need information about which meanings expressions do not have. The authors note that children already master Principle C at age three and they conclude from this that Principle C is a plausible candidate for innate linguistic knowledge. (This is a classic kind of argumentation. For Poverty of the Stimulus arguments, see Section 13.8 and for more on negative evidence, see Section 13.8.4).

Evans & Levinson (2009b: 483) note that Principle C is a strong cross-linguistic tendency but it nevertheless has some exceptions. As an example, they mention both reciprocal expressions in Abaza, where affixes that correspond to *each other* occur in subject position rather than object position as well as Guugu Yimidhirr, where pronouns in a superordinate clause can be coreferent with full NPs in a subordinate clause.

Furthermore, Fanselow (1992b: 351) refers to the examples in (13) that show that Principle C is a poor candidate for a syntactic principle.

(13) a. Mord ist ein Verbrechen.
murder is a crime

b. Ein gutes Gespräch hilft Probleme überwinden.
a good conversation helps problems overcome
'A good conversation helps to overcome problems.'

(13a) expresses that it is a crime when somebody kills someone else, and (13b) refers to conversations with another person rather than talking to oneself. In these sentences, the nominalizations *Mord* 'murder' and *Gespräch* 'conversation' are used without any arguments of the original verbs. So there aren't any arguments that stand in a syntactic command relation to one another. Nevertheless the arguments of the nominalized verbs cannot be coreferential. Therefore it seems that there is a principle at work that says that the argument slots of a predicate must be interpreted as non-coreferential as long as the identity of the arguments is not explicitly expressed by linguistic means.

In sum, one can say that there are still a number of unsolved problems with Binding Theory. The HPSG variants of Principles A–C in English cannot even be applied to German (Müller 1999a: Chapter 20). Working in LFG, Dalrymple (1993) proposes a variant of Binding Theory where the binding properties of pronominal expressions are determined in the lexicon. In this way, the language-specific properties of pronouns can be accounted for.

13.1.5 Properties of long-distance dependencies

The long-distance dependencies discussed in the preceding chapters are subject to some kind of restrictions. For example, nothing can be extracted out of sentences that are part of a noun phrase in English. Ross (1967: 70) calls the relevant constraint the *Complex NP Constraint*. In later work, the attempt was made to group this, and other constraints such as the *Right Roof Constraint* also formulated by Ross (1967: Section 5.1.2), into a single, more general constraint, namely the Subjacency Principle (Chomsky: 1973: 271; 1986a: 40; Baltin 1981, 2006). Subjacency was assumed to hold universally. The Subjacency Constraint states that movement operations can cross at most one bounding node, whereby what exactly counts as a bounding node depends on the language in question (Baltin: 1981: 262; 2006; Rizzi 1982: 57; Chomsky 1986a: 38–40).[11]

Currently, there are varying opinions in the GB/Minimalism tradition with regard to the question of whether subjacency should be considered as part of innate linguistic knowledge. Hauser, Chomsky & Fitch (2002) assume that subjacency does not form part of language-specific abilities, at least not in the strictest sense, but rather is a linguistically relevant constraint in the broader sense that the constraints in question can be derived from more general cognitive ones (see p. 457). Since subjacency still plays a role as a UG principle in other contemporary works (Newmeyer 2005: 15, 74–75; 2004a: 184; Baltin 2006[12]; Baker 2009; Freidin 2009; Rizzi 2009a,b), the Subjacency Principle will be discussed here in some further detail.

[11] Newmeyer (2004b: 539–540) points out a conceptual problem following from the language-specific determination of bounding nodes: it is argued that subjacency is an innate language-specific principle since it is so abstract that it is impossible for speakers to learn it. However, if parameterization requires that a speaker chooses from a set of categories in the linguistic input, then the corresponding constraints must be derivable from the input at least to the degree that it is possible to determine the categories involved. This raises the question as to whether the original claim of the impossibility of acquisition is actually justified. See Section 13.8 on the *Poverty of the Stimulus* and Section 16.1 on parameter-based theories of language acquisition.

Note also that a parameter that has as the value a part of speech requires the respective part of speech values to be part of UG.

[12] However, see Baltin (2004: 552).

It is possible to distinguish two types of movement: movement to the left (normally called extraction) and movement to the right (normally referred to as extraposition. Both movement types constitute long-distance dependencies. In the following section, I will discuss some of the restrictions on extraposition. Extraction will be discussed in Section 13.1.5.2 following it.

13.1.5.1 Extraposition

Baltin (1981) and Chomsky (1986a: 40) claim that the extraposed relative clauses in (14) have to be interpreted with reference to the embedding NP, that is, the sentences are not equivalent to those where the relative clause would occur in the position marked with t, but rather they correspond to examples where it would occur in the position of the t′.

(14) a. [$_{\text{NP}}$ Many books [$_{\text{PP}}$ with [stories t]] t′] were sold [that I wanted to read].
b. [$_{\text{NP}}$ Many proofs [$_{\text{PP}}$ of [the theorem t]] t′] appeared
[that I wanted to think about].

Here, it is assumed that NP, PP, VP and AP are bounding nodes for rightward movement (at least in English) and the interpretation in question here is thereby ruled out by the Subjacency Principle (Baltin 1981: 262).

If we construct a German example parallel to (14a) and replace the embedding noun so that it is ruled out or dispreferred as a referent, then we arrive at (15):

(15) weil viele Schallplatten mit Geschichten verkauft wurden, die ich noch
because many records with stories sold were that I still
lesen wollte
read wanted
'because many records with stories were sold that I wanted to read'

This sentence can be uttered in a situation where somebody in a record store sees particular records and remembers that he had wanted to read the fairy tales on those records. Since one does not read records, adjunction to the superordinate noun is implausible and thus adjunction to *Geschichten* 'stories' is preferred. By carefully choosing the nouns, it is possible to construct examples such as (16) that show that extraposition can take place across multiple NP nodes:[13]

(16) a. Karl hat mir [ein Bild [einer Frau $\_{i}$]] gegeben, [die schon lange tot
Karl has me a picture a woman given that PRT long dead
ist]$_{i}$.
is
'Karl gave me a picture of a woman that has been dead some time.'

[13] See Müller (1999a: 211) and Müller (2004b; 2007e: Section 3). For parallel examples from Dutch, see Koster (1978: 52).

b. Karl hat mir [eine Fälschung [des Bildes [einer Frau $\_{i}$]]] gegeben,
Karl has me a forgery of.the picture of.a woman given
[die schon lange tot ist]$_{i}$.
that PRT long dead is
'Karl gave me a forgery of the picture of a woman that has been dead for some time.'

c. Karl hat mir [eine Kopie [einer Fälschung [des Bildes [einer Frau $\_{i}$]]]]
Karl has me a copy of.a forgery of.the picture of.a woman
gegeben, [die schon lange tot ist]$_{i}$.
given that PRT long dead is
'Karl gave me a copy of a forgery of the picture of a woman that has been dead for some time.'

This kind of embedding could continue further if one were to not eventually run out of nouns that allow for semantically plausible embedding. NP is viewed as a bounding node in German (Grewendorf: 1988: 81; 2002: 17–18; Haider 2001: 285). These examples show that it is possible for rightward extraposed relative clauses to cross any number of bounding nodes.

Koster (1978: 52–54) discusses some possible explanations for the data in (16), where it is assumed that relative clauses move to the NP/PP border and are then moved on further from there (this movement requires so-called escape hatches or escape routes). He argues that these approaches will also work for the very sentences that should be ruled out by subjacency, that is, for examples such as (14). This means that either data such as (14) can be explained by subjacency and the sentences in (16) are counterexamples, or there are escape hatches and the examples in (14) are irrelevant, deviant sentences that cannot be explained by subjacency.

In the examples in (16), a relative clause was extraposed in each case. These relative clauses are treated as adjuncts and there are analyses that assume that extraposed adjuncts are not moved but rather base-generated in their position, and coreference/coindexation is achieved by special mechanisms (Kiss 2005). For proponents of these kinds of analyses, the examples in (16) would be irrelevant to the subjacency discussion as the Subjacency Principle only constrains movement. However, extraposition across phrase boundaries is not limited to relative clauses; sentential complements can also be extraposed:

(17) a. Ich habe [von [der Vermutung $\_{i}$]] gehört, [dass es Zahlen gibt, die
I have from the conjecture heard that there numbers are that
die folgenden Bedingungen erfüllen]$_{i}$.
the following requirements fulfill
'I have heard of the conjecture that there are numbers that fulfill the following requirements.'

b. Ich habe [von [einem Beweis [der Vermutung $\_{i}$]]] gehört, [dass es
I have from a proof of.the conjecture heard that there

Zahlen gibt, die die folgenden Bedingungen erfüllen]$_i$.
numbers are that the following requirements fulfill

'I have heard of the proof of the conjecture that there are numbers that fulfill the following requirements.'

c. Ich habe [von [dem Versuch [eines Beweises [der Vermutung _$_i$]]]]
I have from the attempt of.a proof of.the conjecture
gehört, [dass es Zahlen gibt, die die folgenden Bedingungen erfüllen]$_i$.
heard that there numbers are that the following requirements fulfill

'I have heard of the attempt to prove the conjecture that there are numbers that fulfill the following requirements.'

Since there are nouns that select *zu* infinitives or prepositional phrases and since these can be extraposed like the sentences above, it must be ensured that the syntactic category of the postposed element corresponds to the category required by the noun. This means that there has to be some kind of relation between the governing noun and the extraposed element. For this reason, the examples in (17) have to be analyzed as instances of extraposition and provide counter evidence to the claims discussed above.

If one wishes to discuss the possibility of recursive embedding, then one is forced to refer to constructed examples as the likelihood of stumbling across groups of sentences such as those in (16) and (17) is very remote. It is, however, possible to find some individual cases of deep embedding: (18) gives some examples of relative clause extraposition and complement extraposition taken from the Tiger corpus[14] (Müller 2007e: 78–79; Meurers & Müller 2009: Section 2.1).

(18) a. Der 43jährige will nach eigener Darstellung damit [$_{NP}$ den Weg
the 43.year.old wants after own depiction there.with the way
[$_{PP}$ für [$_{NP}$ eine Diskussion [$_{PP}$ über [$_{NP}$ den künftigen Kurs [$_{NP}$ der
for a discussion about the future course of.the
stärksten Oppositionsgruppierung]]]]]] freimachen, [die aber mit 10,4
strongest opposition.group free.make that however with 10.4
Prozent der Stimmen bei der Wahl im Oktober weit hinter den Erwartungen
percent of.the votes at the election in October far behind the expectations
zurückgeblieben war]. (s27639)
stayed.back was

'In his own words, the 43-year old wanted to clear the way for a discussion about the future course of the strongest opposition group that had, however, performed well below expectations gaining only 10.4 percent of the votes at the election in October.'

b. […] die Erfindung der Guillotine könnte [$_{NP}$ die Folge [$_{NP}$ eines verzweifelten
the invention of.the guillotine could the result of.a desperate

[14] See Brants et al. (2004) for more information on the Tiger corpus.

Versuches des gleichnamigen Doktors] gewesen sein, [seine Patienten ein
attempt the same.name doctor have been his patients once
für allemal von Kopfschmerzen infolge schlechter Kissen zu befreien].
for all.time of headaches because.of bad pillows to free
(s16977)

'The invention of the guillotine could have been the result of a desperate attempt of the eponymous doctor to rid his patients once and for all of headaches from bad pillows.'

It is also possible to construct sentences for English that violate the Subjacency Condition. Uszkoreit (1990: 2333) provides the following example:

(19) [$_{\text{NP}}$ Only letters [$_{\text{PP}}$ from [$_{\text{NP}}$ those people $\__i$]]] remained unanswered [that had received our earlier reply]$_i$.

Jan Strunk (p. c. 2008) has found examples for extraposition of both restrictive and non-restrictive relative clauses across multiple phrase boundaries:

(20) a. For example, we understand that Ariva buses have won [$_{\text{NP}}$ a number [$_{\text{PP}}$ of [$_{\text{NP}}$ contracts [$_{\text{PP}}$ for [$_{\text{NP}}$ routes in London $\__i$]]]]] recently, [which will not be run by low floor accessible buses]$_i$.[15]

b. I picked up [$_{\text{NP}}$ a copy of [$_{\text{NP}}$ a book $\__i$]] today, by a law professor, about law, [that is not assigned or in any way required to read]$_i$.[16]

c. We drafted [$_{\text{NP}}$ a list of [$_{\text{NP}}$ basic demands $\__i$]] that night [that had to be unconditionally met or we would stop making and delivering pizza and go on strike]$_i$.[17]

(20a) is also published in Strunk & Snider (2013: 111). Further attested examples from German and English can be found in this paper.

The preceding discussion has shown that subjacency constraints on rightward movement do not hold for English or German and thus cannot be viewed as universal. One could simply claim that NP and PP are not bounding nodes in English or German. Then, these extraposition data would no longer be problematic for theories assuming subjacency. However, subjacency constraints are also assumed for leftward movement. This is discussed in more detail in the following section.

13.1.5.2 Extraction

Under certain conditions, leftward movement is not possible from certain constituents (Ross 1967). These constituents are referred to as islands for extraction. Ross (1967: Section 4.1) formulated the *Complex NP Constraint* (CNPC) that states that extraction is not

[15] http://www.publications.parliament.uk/pa/cm199899/cmselect/cmenvtra/32ii/32115.htm, 24.02.2007.
[16] http://greyhame.org/archives/date/2005/09/, 27.09.2008.
[17] http://portland.indymedia.org/en/2005/07/321809.shtml, 27.09.2008.

possible from complex noun phrases. An example of extraction from a relative clause inside a noun phrase is the following:

(21) * Who$_i$ did he just read [$_{NP}$ the report [$_S$ that was about _$_i$]?

Although (21) would be a semantically plausible question, the sentence is still ungrammatical. This is explained by the fact that the question pronoun has been extracted across the sentence boundary of a relative clause and then across the NP boundary and has therefore crossed two bounding nodes. It is assumed that the CNPC holds for all languages. This is not the case, however, as the corresponding structures are possible in Danish (Erteschik-Shir & Lappin 1979: 55), Norwegian, Swedish, Japanese, Korean, Tamil and Akan (see Hawkins (1999: 245, 262) and references therein). Since the restrictions of the CNPC are integrated into the Subjacency Principle, it follows that the Subjacency Principle cannot be universally applicable unless one claims that NP is not a bounding node in the problematic languages. However, it seems indeed to be the case that the majority of languages do not allow extraction from complex noun phrases. Hawkins explains this on the basis of the processing difficulties associated with the structures in question (Section 4.1). He explains the difference between languages that allow this kind of extraction and languages that do not with reference to the differing processing load for structures that stem from the interaction of extraction with other grammatical properties such as verb position and other conventionalized grammatical structures in the respective languages (Section 4.2).

Unlike extraction from complex noun phrases, extraction across a single sentence boundary (22) is not ruled out by the Subjacency Principle.

(22) Who$_i$ did she think that he saw _$_i$?

Movement across multiple sentence boundaries, as discussed in previous chapters, is explained by so-called cyclic movement in transformational theories: a question pronoun is moved to a specifier position and can then be moved further to the next highest specifier. Each of these movement steps is subject to the Subjacency Principle. The Subjacency Principle rules out long-distance movement in one fell swoop.

The Subjacency Principle cannot explain why extraction from sentences embedded under verbs that specify the kind of utterance (23a) or factive verbs (23b) is deviant (Erteschik-Shir & Lappin 1979: 68–69).

(23) a. ?? Who$_i$ did she mumble that he saw _$_i$?
b. ?? Who$_i$ did she realize that he saw _$_i$?

The structure of these sentences seems to be the same as (22). In entirely syntactic approaches, it was also attempted to explain these differences as subjacency violations or as a violation of Ross' constraints. It has therefore been assumed (Stowell 1981: 401–402) that the sentences in (23) have a structure different from those in (22). Stowell treats these sentential arguments of manner of speaking verbs as adjuncts. Since adjunct clauses are islands for extraction by assumption, this would explain why (23a) is marked. The adjunct analysis is compatible with the fact that these sentential arguments can be omitted:

(24) a. She shouted that he left.
b. She shouted.

Ambridge & Goldberg (2008: 352) have pointed out that treating such clauses as adjuncts is not justified as they are only possible with a very restricted class of verbs, namely verbs of saying and thinking. This property is a property of arguments and not of adjuncts. Adjuncts such as place modifiers are possible with a wide number of verb classes. Furthermore, the meaning changes if the sentential argument is omitted as in (24b): whereas (24a) requires that some information is communicated, this does not have to be the case with (24b). It is also possible to replace the sentential argument with an NP as in (25), which one would certainly not want to treat as an adjunct.

(25) She shouted the remark/the question/something I could not understand.

The possibility of classifying sentential arguments as adjuncts cannot be extended to factive verbs as their sentential argument is not optional (Ambridge & Goldberg 2008: 352):

(26) a. She realized that he left.
b. ?? She realized.

Kiparsky & Kiparsky (1970) suggest an analysis of factive verbs that assumes a complex noun phrase with a nominal head. An optional *fact* Deletion-Transformation removes the head noun and the determiner of the NP in sentences such as (27a) to derive sentences such as (27b) (page 159).

(27) a. She realized [$_{\mathrm{NP}}$ the fact [$_{\mathrm{S}}$ that he left]].
b. She realized [$_{\mathrm{NP}}$ [$_{\mathrm{S}}$ that he left]].

The impossibility of extraction out of such sentences can be explained by assuming that two boundary nodes were crossed, which was assumed to be impossible (on the island status of this construction, see Kiparsky & Kiparsky 1970: Section 4). This analysis predicts that extraction from complement clauses of factive verbs should be just as bad as extraction from overt NP arguments since the structure for both is the same. According to Ambridge & Goldberg (2008: 353), this is, however, not the case:

(28) a. * Who did she realize the fact that he saw $\__i$?
b. ?? Who did she realize that he saw $\__i$?

Together with Erteschik-Shir (1981), Erteschik-Shir & Lappin (1979), Takami (1988) and Van Valin (1998), Goldberg (2006: Section 7.2) assumes that the gap must be in a part of the utterance that can potentially form the focus of an utterance (see Cook (2001), De Kuthy (2002) and Fanselow (2003c) for German). This means that this part must not be presupposed.[18] If one considers what this means for the data from the subjacency discussion, then one notices that in each case extraction has taken place out of presupposed material:

[18] Information is presupposed if it is true regardless of whether the utterance is negated or not. Thus, it follows from both (i.a) and (i.b) that there is a king of France.

(29) a. Complex NP
She didn't see the report that was about him. → The report was about him.

b. Complement of a verb of thinking or saying
She didn't whisper that he left. → He left.

c. Factive verb
She didn't realize that he left. → He left.

Goldberg assumes that constituents that belong to backgrounded information are islands (*Backgrounded constructions are islands* (BCI)). Ambridge & Goldberg (2008) have tested this semantic/pragmatic analysis experimentally and compared it to a purely syntactic approach. They were able to confirm that information structural properties play a significant role for the extractability of elements. Along with Erteschik-Shir (1973: Section 3.H), Ambridge & Goldberg (2008: 375) assume that languages differ with regard to how much constituents have to belong to background knowledge in order to rule out extraction. In any case we should not rule out extraction from adjuncts for all languages as there are languages such as Danish where it is possible to extract from relative clauses.[19] Erteschik-Shir (1973: 61) provides the following examples, among others:

(30) a. Det$_i$ er der mange [der kan lide _$_i$].
that are there many that can like
'There are many who like that.' (lit.: 'That, there are many who like.')

b. Det hus$_i$ kender jeg en mand [som har købt _$_i$].
that house know I a man that has bought
'I know a man that has bought that house.' (lit.: 'This house, I know a man that has bought.')

Rizzi's parameterization of the subjacency restriction has been abandoned in many works and the relevant effects have been ascribed to differences in other areas of grammar (Adams 1984; Chung & McCloskey 1983; Grimshaw 1986; Kluender 1992).

We have seen in this subsection that there are reasons other than syntactic properties of structure as to why leftward movement might be blocked. In addition to information structural properties, processing considerations also play a role (Grosu 1973; Ellefson & Christiansen 2000; Gibson 1998; Kluender & Kutas 1993; Hawkins 1999; Sag, Hofmeister & Snider 2007). The length of constituents involved, the distance between filler and

(i) a. The King of France is bald.
b. The King of France is not bald.

[19] Discussing the question of whether UG-based approaches are falsifiable, Crain, Khlentzos & Thornton (2010: 2669) claim that it is not possible to extract from relative clauses and the existence of such languages would call into question the very concept of UG. ("If a child acquiring any language could learn to extract linguistic expressions from a relative clause, then this would seriously cast doubt on one of the basic tenets of UG.") They thereby contradict Evans and Levinson as well as Tomasello, who claim that UG approaches are not falsifiable. If the argumentation of Crain, Khlentzos and Thornton were correct, then (30) would falsify UG and that would be the end of the discussion.

gap, definiteness, complexity of syntactic structure and interference effects between similar discourse referents in the space between the filler and gap are all important factors for the acceptability of utterances. Since languages differ with regard to their syntactic structure, varying effects of performance, such as the ones found for extraposition and extraction, are to be expected.

In sum, we can say that subjacency constraints do not hold for extraposition in either German or English and furthermore that one can better explain constraints on extraction with reference to information structure and processing phenomena than with the Subjacency Principle. Assuming subjacency as a syntactic constraint in a universal competence grammar is therefore unnecessary to explain the facts.

13.1.6 Grammatical morphemes for tense, mood and aspect

Pinker (1994: 238) is correct in claiming that there are morphemes for tense, mood, aspect, case and negation in many languages. However, there is a great deal of variation with regard to which of these grammatical properties are used in a language and how they are expressed.

For examples of differences in the tense system see Dahl & Velupillai (2013b,a). Mandarin Chinese is a clear case: it has next to no morphology. The fact that the same morphemes occur in one form or another in almost every language can be attributed to the fact that certain things need to be expressed repeatedly and then things which are constantly repeated become grammaticalized.

13.1.7 Parts of speech

In Section 4.6, so-called cartographic approaches were mentioned, some of which assume over thirty functional categories (see Table 4.1 on page 149 for Cinque's functional heads) and assume that these categories form part of UG together with corresponding fixed syntactic structures. Cinque & Rizzi (2010: 55, 57) even assume over 400 functional categories that are claimed to play a role in the grammars of all languages.[20] Also, specific parts of speech such as Infl (inflection) and Comp (complementizer) are referred to when formulating principles that are assumed to be universal (Baltin: 1981: 262; 2006; Rizzi 1982; Chomsky 1986a: 38; Hornstein 2013: 397).

Chomsky (1988: 68; 1991; 1995b: 131), Pinker (1994: 284, 286), Briscoe (2000: 270) and Wunderlich (2004: 621) make comparatively fewer assumptions about the innate inventory of parts of speech: Chomsky assumes that all lexical categories (verbs, nouns, adjectives and adpositions) belong to UG and languages have these at their disposal. Pinker, Briscoe and Wunderlich assume that all languages have nouns and verbs. Again critics of UG raised the question as to whether these syntactic categories can be found in other languages in the form known to us from languages such as German and English.

Braine (1987: 72) argues that parts of speech such as verb and noun should be viewed as derived from fundamental concepts like argument and predicate (also see Wunderlich

[20] The question of whether these categories form part of UG is left open.

(2008: 257)). This means that there is an independent explanation for the presence of these categories that is not based on innate language-specific knowledge.

Evans & Levinson (2009a: Section 2.2.4) discuss the typological literature and give examples of languages which lack adverbs and adjectives. The authors cite Straits Salish as a language in which there may be no difference between verbs and nouns (also, see Evans & Levinson 2009b: 481). They remark that it does make sense to assume the additional word classes ideophone, positional, coverb, classifier for the analysis of non Indo-European languages on top of the four or five normally used.[21] This situation is not a problem for UG-based theories if one assumes that languages can choose from an inventory of possibilities (a toolkit) but do not have to exhaust it (Jackendoff 2002: 263; Newmeyer 2005: 11; Fitch, Hauser & Chomsky 2005: 204; Chomsky 2007: 6–7; Cinque & Rizzi 2010: 55, 58, 65). However, if we condone this view, then there is a certain arbitrariness. It is possible to assume any parts of speech that one requires for the analysis of at least one language, attribute them to UG and then claim that most (or maybe even all) languages do not make use of the entire set of parts of speech. This is what is suggested by Villavicencio (2002: 157), working in the framework of Categorial Grammar, for the categories S, NP, N, PP and PRT. This kind of assumption is not falsifiable (see Evans & Levinson 2009a: 436; Tomasello 2009: 471 for a discussion of similar cases and a more general discussion).

Whereas Evans and Levinson assume that one needs additional categories, Haspelmath (2009: 458) and Croft (2009: 453) go so far as to deny the existence of cross-linguistic parts of speech. I consider this to be too extreme and believe that a better research strategy is to try and find commonalities between languages.[22] One should, however, expect to find languages that do not fit into our Indo-European-biased conceptions of grammar.

13.1.8 Recursion and infinitude

In an article in *Science*, Hauser, Chomsky & Fitch (2002) put forward the hypothesis that the only domain-specific universal is recursion, "providing the capacity to generate an infinite range of expressions from a finite set of elements" (see (36b) on p. 66 for an example of a recursive phrase structure rule).[23] This assumption is controversial and there have been both formal and empirical objections to it.

[21] For the opposite view, see Jackendoff & Pinker (2009: 465).

[22] Compare Chomsky (1999: 2): "In the absence of compelling evidence to the contrary, assume languages to be uniform, with variety restricted to easily detectable properties of utterances."

[23] In a discussion article in *Cognition*, Fitch, Hauser & Chomsky (2005) clarify that their claim that recursion is the only language-specific and human-specific property is a hypothesis and it could be the case that are not any language-specific/species-specific properties at all. Then, a particular combination of abilities and properties would be specific to humans (p. 182–201). An alternative they consider is that innate language-specific knowledge has a complexity corresponding to what was assumed in earlier versions of Mainstream Generative Grammar (p. 182).

Chomsky (2007: 7) notes that Merge could be a non language-specific operation but still attributes it to UG.

13.1.8.1 Formal problems

The claim that our linguistic capabilities are infinite is widespread and can already be found in Humboldt's work:[24]

> Das Verfahren der Sprache ist aber nicht bloß ein solches, wodurch eine einzelne Erscheinung zustande kommt; es muss derselben zugleich die Möglichkeit eröffnen, eine unbestimmbare Menge solcher Erscheinungen und unter allen, ihr von dem Gedanken gestellten Bedingungen hervorzubringen. Denn sie steht ganz eigentlich einem unendlichen und wahrhaft grenzenlosen Gebiete, dem Inbegriff alles Denkbaren gegenüber. Sie muss daher von endlichen Mitteln einen unendlichen Gebrauch machen, und vermag dies durch die Identität der gedanken- und spracheerzeugenden Kraft. (Humboldt 1988: 108)

If we just look at the data, we can see that there is an upper bound for the length of utterances. This has to do with the fact that extremely long instances cannot be processed and that speakers have to sleep or will eventually die at some point. If we set a generous maximal sentence length at 100,000 morphemes and then assume a morpheme inventory of X then one can form less than $X^{100,000}$ utterances. We arrive at the number $X^{100,000}$ if we use each of the morphemes at each of the 100,000 positions. Since not all of these sequences will be well-formed, then there are actually less than $X^{100,000}$ possible utterances (see also Weydt 1972 for a similar but more elaborate argument). This number is incredibly large, but still finite. The same is true of thought: we do not have infinitely many possible thoughts (if *infinitely* is used in the mathematical sense of the word), despite claims by Humboldt and Chomsky (2008: 137) to the contrary.[25]

[24] The process of language is not simply one where an individual instantiation is created; at the same time it must allow for an indefinite set of such instantiations and must above all allow the expression of the conditions imposed by thought. Language faces an infinite and truly unbounded subject matter, the epitome of everything one can think of. Therefore, it must make infinite use of finite means and this is possible through the identity of the power that is responsible for the production of thought and language.

[25] Weydt (1972) discusses Chomsky's statements regarding the existence of infinitely many sentences and whether it is legitimate for Chomsky to refer to Humboldt. Chomsky's quote in *Current Issues in Linguistic Theory* (Chomsky 1964a: 17) leaves out the sentence *Denn sie steht ganz eigentlich einem unendlichen und wahrhaft grenzenlosen Gebiete, dem Inbegriff alles Denkbaren gegenüber.* Weydt (1972: 266) argues that Humboldt, Bühler and Martinet claimed that there are infinitely many thoughts that can be expressed. Weydt claims that it does not follow that sentences may be arbitrarily long. Instead he suggests that there is no upper bound on the length of texts. This claim is interesting, but I guess texts are just the next bigger unit and the argument that Weydt put forward against languages without an upper bound for sentence length also applies to texts. A text can be generated by the rather simplified rule in (i) that combines an utterance U with a text T resulting in a larger text T:

(i) $T \rightarrow T\ U$

U can be a sentence or another phrase that can be part of a text. If one is ready to admit that there is no upper bound on the length of texts, it follows that there cannot be an upper bound on the length of sentences either, since one can construct long sentences by joining all phrases of a text with *and*. Such long sentences that are the product of conjoining short sentences are different in nature from very long sentences that are admitted under the Chomskyan view in that they do not include center-self embeddings of an arbitrary depth (see Section 15), but nevertheless the number of sentences that can be produced from arbitrarily long texts is infinite.

In the literature, one sometimes finds the claim that it is possible to produce infinitely long sentences (see for instance Nowak, Komarova & Niyogi (2001: 117) and Kim & Sells (2008: 3) and Dan Everett in O'Neill & Wood (2012) at 25:19). This is most certainly not the case. It is also not the case that the rewrite grammars we encountered in Chapter 2 allow for the creation of infinite sentences as the set of symbols of the right-hand side of the rule has to be finite by definition. While it is possible to derive an infinite number of sentences, the sentences themselves cannot be infinite, since it is always one symbol that is replaced by finitely many other symbols and hence no infinite symbol sequence may result.

Chomsky (1965: Section I.1) follows de Saussure (1916b) and draws a distinction between competence and performance: competence is the knowledge about what kind of linguistic structures are well-formed, and performance is the application of this knowledge (see Section 12.6.3 and Chapter 15). Our restricted brain capacity as well as other constraints are responsible for the fact that we cannot deal with an arbitrary amount of embedding and that we cannot produce utterances longer than 100,000 morphemes. The separation between competence and performance makes sense and allows us to formulate rules for the analysis of sentences such as (31):

(31) a. Richard is sleeping.
  b. Karl suspects that Richard is sleeping.
  c. Otto claims that Karl suspects that Richard is sleeping.
  d. Julius believes that Otto claims that Karl suspects that Richard is sleeping.
  e. Max knows that Julius believes that Otto claims that Karl suspects that Richard is sleeping.

The rule takes the following form: combine a noun phrase with a verb of a certain class and a clause. By applying this rule successively, it is possible to form strings of arbitrary length. Pullum & Scholz (2010) point out that one has to keep two things apart: the question of whether language is a recursive system and whether it is just the case that the best models that we can devise for a particular language happen to be recursive. For more on this point and on processing in the brain, see Luuk & Luuk (2011). When constructing strings of words using the system above, it cannot be shown that (a particular) language is infinite, even if this is often claimed to be the case (Bierwisch 1966: 105–106; Pinker 1994: 86; Hauser, Chomsky & Fitch 2002: 1571; Müller 2007b: 1; Hornstein, Nunes & Grohmann 2005: 7; Kim & Sells 2008: 3).

The "proof" of this infinitude of language is led as an indirect proof parallel to the proof that shows that there is no largest natural number (Bierwisch 1966: 105–106; Pinker 1994:

As for arbitrarily long texts there is an interesting problem: Let us assume that a person produces sentences and keeps adding them to an existing text. This enterprise will be interrupted when the human being dies. One could say that another person could take up the text extension until this one dies and so on. Again the question is whether one can understand the meaning and the structure of a text that is several million pages long. 42. If this is not enough of a problem, one may ask oneself whether the language of the person who keeps adding to the text in the year 2731 is still the same that the person who started the text spoke in 2015. If the answer to this question is no, then the text is not a document containing sentences from one language L but a mix from several languages and hence irrelevant for the debate.

86). In the domain of natural numbers, this works as follows: assume x is the largest natural number. Then form x + 1 and, since this is by definition a natural number, we have now found a natural number that is greater than x. We have therefore shown that the assumption that x is the highest number leads to a contradiction and thus that there cannot be such a thing as the largest natural number.

When transferring this proof into the domain of natural language, the question arises as to whether one would still want to class a string of 1,000,000,000 words as part of the language we want to describe. If we do not want this, then this proof will not work.

If we view language as a biological construct, then one has to accept the fact that it is finite. Otherwise, one is forced to assume that it is infinite, but that an infinitely large part of the biologically real object is not biologically real (Postal 2009: 111). Luuk & Luuk (2011) refer to languages as physically uncountable but finite sets of strings. They point out that a distinction must be made between the ability to imagine extending a sentence indefinitely and the ability to take a sentence from a non-countable set of strings and really extend it. We possess the first ability but not the second.

One possibility to provide arguments for the infinitude of languages is to claim that only generative grammars, which create sets of well-formed utterances, are suited to modeling language and that we need recursive rules to capture the data, which is why mental representations have a recursive procedure that generates infinite numbers of expressions (Chomsky, 1956: 115; 2002: 86–87), which then implies that languages consist of infinitely many expressions. There are two mistakes in this argument that have been pointed out by Pullum & Scholz (2010): even if one assumes generative grammars, it can still be the case that a context-sensitive grammar can still only generate a finite set even with recursive rules. Pullum & Scholz (2010: 120–121) give an interesting example from András Kornai.

The more important mistake is that it is not necessary to assume that grammars generate sets. There are three explicitly formalized alternatives of which only the third is mentioned here, namely the model-theoretic and therefore constraint-based approaches (see Chapter 14). Johnson & Postal's Arc Pair Grammar (1980), LFG in the formalization of Kaplan (1995), GPSG in the reformalization of Rogers (1997) and HPSG with the assumptions of King (1999), Pollard (1999) and Richter (2007) are examples of model-theoretic approaches. In constraint-based theories, one would say for an example like (31) that certain attitude verbs select a nominative NP and a *that* clause and that these can only occur in a certain local configuration where a particular relation holds between the elements involved. One of these relations is subject-verb agreement. In this way, one can represent expressions such as (31) and does not have to say anything about how many sentences can be embedded. This means that constraint-based theories are compatible with both answers to the question of whether there is a finite or infinite number of structures. Using competence grammars formulated in the relevant way, it is possible to develop performance models that explain why certain strings – for instance very long ones – are unacceptable (see Chapter 15).

13.1.8.2 Empirical problems

It is sometimes claimed that all natural languages are recursive and that sentences of an arbitrary length are possible in all languages (Hornstein, Nunes & Grohmann 2005: 7 for an overview, and see Pullum & Scholz (2010: Section 2) for further references). When one speaks of recursion, what is often meant are structures with self-embedding as we saw in the analysis of (31) (Fitch 2010). However, it is possible that there are languages that do not allow self-embedding. Everett (2005) claims that Pirahã is such a language (however, see Nevins, Pesetsky & Rodrigues (2009) and Everett (2009)). A further example of a language without recursion, which is sometimes cited with reference to Hale (1976), is Warlpiri. Hale's rules for the combination of a sentence with a relative clause are recursive, however (page 85). This recursion is made explicit on page 98.[26] Pullum & Scholz (2010: 131) discuss Hixkaryána, an Amazonian language from the Caribbean language family that is not related to Pirahã. This language does have embedding, but the embedded material has a different form to that of the matrix clause. It could be the case that these embeddings cannot be carried out indefinitely. In Hixkaryána, there is also no possibility to coordinate phrases or clauses (Derbyshire (1979: 45) cited by Pullum & Scholz (2010: 131)), which is why this possibility of forming recursive sentence embedding does not exist in this language either. Other languages without self-embedding seem to be Akkadian, Dyirbal and Proto-Uralic.

There is of course a trivial sense in which all languages are recursive: they follow a rule that says that a particular number of symbols can be combined to form another symbol.[27]

(32) X → X … X

In this sense, all natural languages are recursive and the combination of simple symbols to more complex ones is a basic property of language (Hockett 1960: 6). The fact that the debate about Pirahã is so fierce could go to show that this is not the kind of recursion that is meant. Also, see Fitch (2010).

It is also assumed that the combinatorial rules of Categorial Grammar hold universally. It is possible to use these rules to combine a functor with its arguments (X/Y ∗ Y = X). These rules are almost as abstract as the rules in (32). The difference is that one of the elements has to be the functor. There are also corresponding constraints in the Minimalist Program such as selectional features (see Section 4.6.4) and restrictions on the assignment of semantic roles. However, whether or not a Categorial Grammar licenses

[26] However, he does note on page 78 that relative clauses are separated from the sentence containing the head noun by a pause. Relative clauses in Warlpiri are always peripheral, that is, they occur to the left or right of a sentence with the noun they refer to. Similar constructions can be found in German:

(i) Es war einmal ein Mann. Der hatte sieben Söhne.
there was once a man he had seven sons
'There once was a man. He had seven sons.'

It could be the case that we are dealing with linking of sentences at text level and not recursion at sentence level.

[27] Chomsky (2005: 11) assumes that Merge combines n objects. A special instance of this is binary Merge.

recursive structures does not depend on the very general combinatorial schemata, but rather on the lexical entries. Using the lexical entries in (33), it is only possible to analyze two sentences and certainly not to build recursive structures.

(33) a. the: np/n
b. woman: n
c. cat: n
d. sees: (s\np)/np

If we expand the lexicon to include modifiers of the category n/n or conjunctions of the category (X\X)/X, then we arrive at a recursive grammar.

Fitch, Hauser & Chomsky (2005: 203) note that the existence of languages that do not license recursive structures is not a problem for UG-based theories as not all the possibilities in UG have to be utilized by an individual language. With this view, we have actually the same situation as with parts of speech (see Section 13.1.7) that you can posit any number of properties belonging to UG and then decide on a language by language basis whether they play a role or not. An extreme variant of this approach would be that grammars of all languages become part of UG (perhaps with different symbols such as $\mathrm{NP}_{\mathrm{Spanish}}$, $\mathrm{NP}_{\mathrm{German}}$). This variant of a UG-based theory of the human capacity for language would be truly unfalsifiable (Evans & Levinson 2009a: 436, 443; Tomasello 2009: 471).

13.1.8.3 Recursion in other areas of cognition

There are also phenomena in domains outside of language that can be described with recursive rules: Hauser, Chomsky & Fitch (2002: 1571) mention navigation, family relations and counting systems.[28] One could perhaps argue that the relevant abilities are acquired late and that higher mathematics is a matter of individual accomplishments that do not have anything to do with the cognitive capacities of the majority, but even children at the age of 3 years and 9 months are already able to produce recursive structures: In 2008, there were newspaper reports about an indigenous Brazilian tribe that was photographed from a plane. I showed this picture to my son and told him that Native Americans shot at the plane with a bow and arrow. He then asked me what kind of plane it was. I told him that you cannot see that because the people who took the photograph were sitting in the plane. He then answered that you would then need another plane if you wanted to take a photo that contained both the plane and the Native Americans. He was pleased with his idea and said “And then another one. And then another

[28] Pinker & Jackendoff (2005: 230) note, however, that navigation differs from the kind of recursive system described by Chomsky and that recursion is not part of counting systems in all cultures. They assume that those cultures that have developed infinite counting systems could do this because of their linguistic capabilities. This is also assumed by Fitch, Hauser & Chomsky (2005: 203). The latter authors claim that all forms of recursion in other domains depend on language. For more on this point, see Chomsky (2007: 7–8). Luuk & Luuk (2011) note that natural numbers are defined recursively, but the mathematical definition does not necessarily play a role for the kinds of arithmetic operations carried out by humans.

one. One after the other". He was therefore very much able to imagine the consequence of embeddings.

Culicover & Jackendoff (2005: 113–114) discuss visual perception and music as recursive systems that are independent of language. Jackendoff (2011) extends the discussion of visual perception and music and adds the domains of planning (with the example of making coffee) and wordless comic strips. Chomsky (2007: 7–8) claims that examples from visual perception are irrelevant but then admits that the ability to build up recursive structures could belong to general cognitive abilities (p. 8). He still attributes this ability to UG. He views UG as a subset of the Faculty of Language, that is, as a subset of non domain-specific abilities (Faculty of Language in the Broad Sense = FLB) and the domain-specific abilities (Faculty of Language in the Narrow Sense = FLN) required for language.

13.1.9 Summary

In sum, we can say that there are no linguistic universals for which there is a consensus that one has to assume domain-specific innate knowledge to explain them. At the 2008 meeting of the *Deutsche Gesellschaft für Sprachwissenschaft*, Wolfgang Klein promised € 100 to anyone who could name a non-trivial property that all languages share (see Klein 2009). This begs the question of what is meant by 'trivial'. It seems clear that all languages share predicate-argument structures and dependency relations in some sense (Hudson 2010a; Longobardi & Roberts 2010: 2701) and, all languages have complex expressions whose meaning can be determined compositionally (Manfred Krifka was promised 20 € for coming up with compositionality). However, as has been noted at various points, universality by no means implies innateness (Bates 1984: 189; Newmeyer 2005: 205): Newmeyer gives the example that words for sun and moon probably exist in all languages. This has to do with the fact that these celestial bodies play an important role in everyone's lives and thus one needs words to refer to them. It cannot be concluded from this that the corresponding concepts have to be innate. Similarly, a word that is used to express a relation between two objects (e. g. *catch*) has to be connected to the words describing both of these objects (*I*, *elephant*) in a transparent way. However, this does not necessarily entail that this property of language is innate.

Even if we can find structural properties shared by all languages, this is still not proof of innate linguistic knowledge, as these similarities could be traced back to other factors. It is argued that all languages must be made in such a way as to be acquirable with the paucity of resource available to small children (Hurford 2002: Section 10.7.2; Behrens 2009: 433). It follows from this that, in the relevant phases of its development, our brain is a constraining factor. Languages have to fit into our brains and since our brains are similar, languages are also similar in certain respects (see Kluender 1992: 251).

13.2 Speed of language acquisition

It is often argued that children learn language extraordinarily quickly and this can only be explained by assuming that they already possess knowledge about language that does not have to be acquired (e. g. Chomsky 1976b: 144; Hornstein 2013: 395). In order for this argument to hold up to closer scrutiny, it must be demonstrated that other areas of knowledge with a comparable degree of complexity require longer to acquire (Sampson 1989: 214–218). This has not yet been shown. Language acquisition spans several years and it is not possible to simply state that language is acquired following *brief exposure*. Chomsky compares languages to physics and points out that it is considerably more difficult for us to acquire knowledge about physics. Sampson (1989: 215) notes, however, that the knowledge about physics one acquires at school or university is not a basis for comparison and one should instead consider the acquisition of everyday knowledge about the physical world around us. For example, the kind of knowledge we need when we want to pour liquids into a container, skip with a skipping rope or the knowledge we have about the ballistic properties of objects. The complexity in comparing these domains of knowledge in order to be able to make claims about language acquisition may turn out to be far from trivial. For an in-depth discussion of this aspect, see Sampson (1989: 214–218). Müller & Riemer (1998: 1) point out that children at the age of six can understand 23,700 words and use over 5000. It follows from this that, in the space of four and a half years, they learn on average 14 new words every day. This is indeed an impressive feat, but cannot be used as an argument for innate linguistic knowledge as all theories of acquisition assume that words have to be learned from data rather than being predetermined by a genetically-determined Universal Grammar. In any case the assumption of genetic encoding would be highly implausible for newly created words such as *fax*, *iPod*, *e-mail*, *Tamagotchi*.

Furthermore, the claim that first language acquisition is effortless and rapid when compared to second language acquisition is a myth as has been shown by estimations by Klein (1986: 9): if we assume that children hear linguistic utterances for five hours a day (as a conservative estimate), then in the first five years of their lives, they have 9100 hours of linguistic training. But at the age of five, they have still not acquired all complex constructions. In comparison, second-language learners, assuming the necessary motivation, can learn the grammar of a language rather well in a six-week crash course with twelve hours a day (500 hours in total).

13.3 Critical period for acquisition

Among ducks, there is a critical phase in which their behavior towards parent figures is influenced significantly. Normally, baby ducks follow their mother. If, however, a human is present rather than the mother during a particular time span, the ducks will follow the human. After the critical period, this influence on their behavior can no longer be identified (Lorenz 1970). This kind of critical period can also be identified in other animals and in other areas of cognition, for example the acquisition of visual abilities among primates. Certain abilities are acquired in a given time frame, whereby the presence of the relevant input is important for determining the start of this time frame.

Lenneberg (1964) claims that language acquisition is only possible up to the age of twelve and concludes from the fact that children can learn language much better than adults that this is also due to a critical period and that language acquisition must have similar properties to the behavior of ducks and hence, the predisposition for language acquisition must be innate (Lenneberg 1967: Chapter 4).

The assumptions about the length of the critical period for language acquisition vary considerably. It is possible to find suggestions for 5, 6, 12 and even 15 years (Hakuta et al. 2003: 31). An alternative assumption to a critical period would be to assume that the ability to acquire languages decreases continuously over time. Johnson & Newport (1989) tried to determine a critical period for second-language acquisition and they claim that a second language is learned significantly worse from the age of 15. Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett (1996) have, however, pointed out that there is a different curve for Johnson and Newport's data that fits the individual data better. The alternative curve shows no abrupt change but rather a steady decrease in the ability to learn language and therefore offers no proof of an effect created by a critical period.

Hakuta, Bialystok & Wiley (2003) evaluate data from a questionnaire of 2,016,317 Spanish speakers and 324,444 speakers of Mandarin Chinese that immigrated to the United States. They investigated which correlations there were between age, the point at immigration, the general level of education of the speakers and the level of English they acquired. They could not identify a critical point in time after which language acquisition was severely restricted. Instead, there is a steady decline in the ability to learn as age increases. This can also be observed in other domains: for example, learning to drive at an older age is much harder.

Summing up, it seems to be relatively clear that a critical period cannot be proven to exist for second-language acquisition. Sometimes, it is assumed anyway that second-language acquisition is not driven by an innate UG, but is in fact a learning process that accesses knowledge already acquired during the critical period (Lenneberg 1967: 176). One would therefore have to show that there is a critical period for first-language acquisition. This is, however, not straightforward as, for ethical reasons, one cannot experimentally manipulate the point at which the input is available. We cannot, say, take 20 children and let them grow up without linguistic input to the age of 3, 4, 5, 6, … or 15 and then compare the results. This kind of research is dependent on thankfully very rare cases of neglect. For example, Curtiss (1977) studied a girl called Genie. At the time, Genie was 13 years old and had grown up in isolation. She is a so-called feral child. As Curtiss showed, she was no longer able to learn certain linguistic rules. For an objective comparison, one would need other test subjects that had not grown up in complete isolation and in inhumane conditions. The only possibility of gaining relevant experimental data is to study deaf subjects that did not receive any input from a sign language up to a certain age. Johnson & Newport (1989: 63) carried out relevant experiments with learners of American Sign Language. It was also shown here that there is a linear decline in the ability to learn, however nothing like a sudden drop after a certain age or even a complete loss of the ability to acquire language.

13.4 Lack of acquisition among non-human primates

The fact that non-human primates cannot learn natural language is viewed as evidence for the genetic determination of our linguistic ability. All scientists agree on the fact that there are genetically-determined differences between humans and primates and that these are relevant for linguistic ability. Friederici (2009) offers an overview of the literature that claims that in chimpanzees and macaques (and small children), the connections between parts of the brain are not as developed as in adult humans. The connected regions of the brain are together responsible for the processing of lexical-semantic knowledge and could constitute an important prerequisite for the development of language (p. 179).

The question is, however, whether we differ from other primates in having special cognitive capabilities that are specific to language or whether our capability to acquire languages is due to domain-general differences in cognition. Fanselow (1992b: Section 2) speaks of a human-specific formal competence that does not necessarily have to be specific to language, however. Similarly, Chomsky (2007: 7–8) has considered whether Merge (the only structure-building operation, in his opinion), does not belong to language-specific innate abilities, but rather to general human-specific competence (see, however, Section 13.1.8, in particular footnote 28).

One can ascertain that non-human primates do not understand particular pointing gestures. Humans like to imitate things. Other primates also imitate, however, not for social reasons (Tomasello 2006b: 9–10). According to Tomasello et al. (2005: 676), only humans have the ability and motivation to carry out coordinated activities with common goals and socially-coordinated action plans. Primates do understand intentional actions, however, only humans act with a common goal in mind (*shared intentionality*). Only humans use and understand hand gestures (Tomasello et al. 2005: 685, 724, 726). Language is collaborative to a high degree: symbols are used to refer to objects and sometimes also to the speaker or hearer. In order to be able to use this kind of communication system, one has to be able to put oneself in the shoes of the interlocutor and develop common expectations and goals (Tomasello et al. 2005: 683). Non-human primates could thus lack the social and cognitive prerequisites for language, that is, the difference between humans and other primates does not have to be explained by innate linguistic knowledge (Tomasello 2003: Section 8.1.2; Tomasello et al. 2005).

13.5 Creole and sign languages

When speakers that do not share a common language wish to communicate with each other, they develop so-called pidgin languages. These are languages that use parts of the vocabularies of the languages involved but have a very rudimentary grammar. It has been noted that children of pidgin speakers regularize these languages. The next generation of speakers creates a new language with an independent grammar. These languages are referred to as *creole languages*. One hypothesis is that the form of languages that develop from creolization is restricted by an innate UG (Bickerton 1984b). It is assumed that the parameter setting of creole languages corresponds to the default values of parameters (Bickerton: 1984a: 217; 1984b: 178), that is, parameters already have values at birth and these correspond to the values that creole languages have. These default values would have to be modified when learning other languages.[29] Bickerton claims that creole languages contain elements that language learners could not have acquired from the input, that is from the pidgin languages. His argumentation is a variant of the classic Poverty of the Stimulus Argument that will be discussed in more detail in Section 13.8.

Bickerton's claims have been criticized as it cannot be verified whether children had input in the individual languages of the adults (Samarin 1984: 207; Seuren 1984: 209). All that can be said considering this lack of evidence is that there are a number of demographic facts that suggest that this was the case for at least some creole languages. This means that children did not only have the strings from the pidgin languages as an input but also sentences from the individual languages spoken by parents and others around them. Many creolists assume that adults contribute specific grammatical forms to the emerging language. For example, in the case of Hawaiian Creole English one can observe that there are influences from the mother tongues of the speakers involved: Japanese speakers use SOV order as well as SVO and Philippinos use VOS order as well as SVO order. In total, there is quite a lot of variation in the language that can be traced back to the various native languages of the individual speakers.

It is also possible to explain the effects observed for creolization without the assumption of innate language-specific knowledge: the fact that children regularize language can be attributed to a phenomenon independent of language. In experiments, participants were shown two light bulbs and the test subjects had to predict which of the light bulbs would be turned on next. If one of the bulbs was switched on 70% of the time, the participants also picked this one 70% of the time (although they would have actually had a higher success rate if they had always chosen the bulb turned on with 70% probability). This behavior is known as *Probability Matching*. If we add another light bulb to this scenario and then turn this lamp on in 70% of cases and the other two each 15% of the time, then participants choose the more frequently lit one 80–90% of the time, that is, they regularize in the direction of the most frequent occurrence (Gardner 1957; Weir 1964).

Children regularize more than adults (Hudson & Newport 1999; Hudson Kam & Newport 2005), a fact that can be traced back to their limited brain capacity ("less is more"-hypothesis, Newport 1990; Elman 1993).

Like creolization, a similar situation can be found in certain social contexts with the acquisition of sign language: Singleton & Newport (2004) have shown that a child (Simon) that learned American Sign Language (ASL) makes considerably less mistakes than his parents. The parents first learned ASL at the age of 15 or 16 and performed particular obligatory movements only 70% of the time. Simon made these movements 90%

[29] For problems that can arise from the assumption of defaults values, see Meisel (1995: 17). Bickerton (1997: 56, fn. 13) distances himself from the claim that creole languages have the default values of parameters.

of the time. He regularized the input from his parents, whereby the consistent use of form-meaning pairs plays an important role, that is, he does not simply use Probability Matching, but learns selectively. Singleton & Newport (2004: 401) suspect that these kinds of regularizations also play a role for the emergence of creole and sign languages. However, the relevant statistical data that one would need to confirm this hypothesis are not available.

13.6 Localization in special parts of the brain

By measuring brain activity during speech production/processing and also by investigating patients with brain damage, one can identify special parts of the brain (Broca's area and Wernicke's area) that play an important role for language production and processing (see Friederici (2009) for a current overview). Chomsky talks about there being a center of language and even calls this (metaphorically) an *organ* (Chomsky 1977: 164; Chomsky 2005: 1; Chomsky 2008: 133). This localization was seen as evidence for the innate basis for our linguistic knowledge (also see Pinker 1994: 297–314).

However, it is the case that if these parts are damaged, other areas of the brain can take over the relevant functions. If the damage occurs in early childhood, language can also be learned without these special areas of the brain (for sources, see Dąbrowska 2004: Section 4.1).

Apart from that, it can also be observed that a particular area of the brain is activated when reading. If the conclusion about the localization of processing in a particular part of the brain leading to the innateness of linguistic knowledge were valid, then the activation of certain areas of the brain during reading should also lead us to conclude that the ability to read is innate (Elman et al. 1996; Bishop 2002: 57). This is, however, not assumed (also see Fitch, Hauser & Chomsky 2005: 196).

It should also be noted that language processing affects several areas of the brain and not just Broca's and Wernicke's areas (Fisher & Marcus 2005: 11; Friederici 2009). On the other hand, Broca's and Wernicke's areas are also active during non-linguistic tasks such as imitation, motoric coordination and processing of music (Maess et al. 2001). For an overview and further sources, see Fisher & Marcus (2005).

Musso et al. (2003) investigated brain activity during second-language acquisition. They gave German native speakers data from Italian and Japanese and noticed that there was activation in Broca's area. They then compared this to artificial languages that used Italian and Japanese words but did not correspond to the principles of Universal Grammar as assumed by the authors. An example of the processes assumed in their artificial language is the formation of questions by reversing of word order as shown in (34).

(34) a. This is a statement.
 b. Statement a is this?

The authors then observed that different areas of the brain were activated when learning this artificial language. This is an interesting result, but does not show that we have innate linguistic knowledge. It only shows that the areas that are active when processing our native languages are also active when we learn other languages and that playing around with words such as reversing the order of words in a sentence affects other areas of the brain.

A detailed discussion of localization of languages in particular parts of the brain can be found in Dąbrowska (2004: Chapter 4).

13.7 Differences between language and general cognition

Researchers who believe that there is no such thing as innate linguistic knowledge assume that language can be acquired with general cognitive means. If it can be shown that humans with severely impaired cognition can still acquire normal linguistic abilities or that there are people of normal intelligence whose linguistic ability is restricted, then one can show that language and general cognition are independent.

13.7.1 Williams Syndrome

There are people with a relatively low IQ, who can nevertheless produce grammatical utterances. Among these are people with Williams Syndrome (see Bellugi, Lichtenberger, Jones, Lai & George (2000) for a discussion of the abilities of people with Williams Syndrome). Yamada (1981) takes the existence of such cases as evidence for a separate module of grammar, independent of the remaining intelligence.

IQ is determined by dividing a score in an intelligence test (the mental age) by chronological age. The teenagers that were studied all had a mental age corresponding to that of a four to six year-old child. Yet children at this age already boast impressive linguistic ability that comes close to that of adults in many respects. Gosch, Städing & Pankau (1994: 295) have shown that children with Williams Syndrome do show a linguistic deficit and that their language ability corresponds to what would be expected from their mental age. For problems of sufferers of Williams Syndrome in the area of morphosyntax, see Karmiloff-Smith et al. (1997). The discussion about Williams Syndrome is summarized nicely in Karmiloff-Smith (1998).

13.7.2 KE family with FoxP2 mutation

There is a British family – the so-called KE family – that has problems with language. The members of this family who suffer from these linguistic problems have a genetic defect. Fisher et al. (1998) and Lai et al. (2001) discovered that this is due to a mutation of the FoxP2 gene (FoxP2 stands for *Forkhead-Box P2*). Gopnik & Cargo (1991) conclude from the fact that deficits in the realm of morphology are inherited with genetic defects that there must be a gene that is responsible for a particular module of grammar (morphology). Vargha-Khadem et al. (1995: 930) have demonstrated, however, that the KE family did not just have problems with morphosyntax: the affected family members have intellectual and linguistic problems together with motoric problems with facial muscles. Due to the considerably restricted motion in their facial muscles, it would make sense to assume that their linguistic difficulties also stem from motory problems (Tomasello 2003: 285). The linguistic problems in the KE family are not just limited to production problems, however, but also comprehension problems (Bishop 2002: 58). Nevertheless, one cannot associate linguistic deficiencies directly with FoxP2 as there are a number of other abilities that are affected by the FoxP2 mutation: as well as hindering pronunciation, morphology and syntax, it also has an effect on non-verbal IQ and motory problems with the facial muscles, dealing with non-linguistic tasks, too (Vargha-Khadem et al. 1995).

Furthermore, FoxP2 also occurs in other body tissues: it is also responsible for the development of the lungs, the heart, the intestine and various regions of the brain (Marcus & Fisher 2003). Marcus & Fisher (2003: 260–261) point out that FoxP2 is probably not directly responsible for the development of organs or areas of organs but rather regulates a cascade of different genes. FoxP2 can therefore not be referred to as the language gene, it is just a gene that interacts with other genes in complex ways. It is, among other things, important for our language ability, however, in the same way that it does not make sense to call FoxP2 a language gene, nobody would connect a hereditary muscle disorder with a 'walking gene' just because this myopathy prevents upright walking (Bishop 2002: 58). A similar argument can be found in Karmiloff-Smith (1998: 392): there is a genetic defect that leads some people to begin to lose their hearing from the age of ten and become completely deaf by age thirty. This genetic defect causes changes in the hairs inside the ear that one requires for hearing. In this case, one would also not want to talk about a 'hearing gene'.

Fitch, Hauser & Chomsky (2005: 190) are also of the opinion that FoxP2 cannot be responsible for linguistic knowledge. For an overview of this topic, see Bishop (2002) and Dąbrowska (2004: Section 6.4.2.2) and for genetic questions in general, see Fisher & Marcus (2005).

13.8 Poverty of the Stimulus

An important argument for the innateness of the linguistic knowledge is the so-called Poverty of the Stimulus Argument (PSA) (Chomsky 1980: 34). Different versions of it can be found in the literature and have been carefully discussed by Pullum & Scholz (2002). After discussing these variants, they summarize the logical structure of the argument as follows (p. 18):

(35) a. Human children learn their first language either by data-driven learning or by learning supported by innate knowledge (a disjunctive premise by assumption)
b. If children learn their first language by data-driven learning, then they could not acquire anything for which they did not have the necessary evidence (the definition of data-driven learning)
c. However, children do in fact learn things that they do not seem to have decisive evidence for (empirical prerequisite)

d. Therefore, children do not learn their first language by data-driven learning. (*modus tollens* of b and c)

e. Conclusion: children learn language through a learning process supported by innate knowledge. (disjunctive syllogism of a and d)

Pullum and Scholz then discuss four phenomena that have been claimed to constitute evidence for there being innate linguistic knowledge. These are plurals as initial parts of compounds in English (Gordon 1986), sequences of auxiliaries in English (Kimball 1973), anaphoric *one* in English (Baker 1978) and the position of auxiliaries in English (Chomsky 1971: 29–33). Before I turn to these cases in Section 13.8.2, I will discuss a variant of the PSA that refers to the formal properties of phrase structure grammars.

13.8.1 Gold's Theorem

In theories of formal languages, a language is viewed as a set containing all the expressions belonging to a particular language. This kind of set can be captured using various complex rewrite grammars. A kind of rewrite grammar – so-called context-free grammars – was presented in Chapter 2. In context-free grammars, there is always exactly one symbol on the left-hand side of the rule (a so-called non-terminal symbol) and there can be more of these on the right-hand side of the rule. On the right side there can be symbols (so-called non-terminal symbols) or words/morphemes of the language in question (so-called terminal symbols). The words in a grammar are also referred to as vocabulary (V). Part of a formal grammar is a start symbol, which is usually S. In the literature, this has been criticized since not all expressions are sentences (see Deppermann 2006: 44). It is, however, not necessary to assume this. It is possible to use Utterance as the start symbol and define rules that derive S, NP, VP or whatever else one wishes to class as an utterance from Utterance.[30]

Beginning with the start symbol, one can keep applying phrase structure rules in a grammar until one arrives at sequences that only contain words (terminal symbols). The set of all sequences that one can generate are the expressions that belong to the language that is licensed by the grammar. This set is a subset of all sequences of words or morphemes that can be created by arbitrary combination. The set that contains all possible sequences is referred to as V^*.

Gold (1967) has shown that in an environment E, it is not possible to solve the identification problem for any language from particular languages classes, given a finite amount of linguistic input, without additional knowledge. Gold is concerned with the identification of a language from a given class of languages. A language L counts as identified if at a given point in time t_n, a learner can determine that L is the language in question and does not change this hypothesis. This point in time is not determined in advance, however, identification has to take place at some point. Gold calls this *identification in the limit*. The environments are arbitrary infinite sequences of sentences $\langle a_1, a_2, a_3, ...\rangle$,

[30] On page 275, I discussed a description that corresponds to the S symbol in phrase structure grammars. If one omits the specification of head features in this description, then one gets a description of all complete phrases, that is, also *the man* or *now*.

whereby each sentence in the language must occur at least once in this sequence. In order to show that the identification problem cannot be solved for even very simple language classes, Gold considers the class of languages that contain all possible sequences of words from the vocabulary V expect for one sequence: let V be the vocabulary and x_1, x_2, x_3, … the sequences of words from this vocabulary. The set of all strings from this vocabulary is V*. For the class of languages in (36), which consist of all possible sequences of elements in V with the exception of one sequence, it is possible to state a process of how one could learn these languages from a text.

(36) $L_1 = V^* - x_1, L_2 = V^* - x_2, L_3 = V^* - x_3, \ldots$

After every input, one can guess that the language is $V^* - \sigma$, where σ stands for the alphabetically first sequence with the shortest length that has not yet been seen. If the sequence in question occurs later, then this hypothesis is revised accordingly. In this way, one will eventually arrive at the correct language.

If we expand the set of languages from which we have to choose by V*, then our learning process will no longer work since, if V* is the target language, then the guessing will perpetually yield incorrect results. If there were a procedure capable of learning this language class, then it would have to correctly identify V* after a certain number of inputs. Let us assume that this input is x_k. How can the learning procedure tell us at this point that the language we are looking for is not $V^* - x_j$ for $j \neq k$? If x_k causes one to guess the wrong grammar V*, then every input that comes after that will be compatible with both the correct ($V^* - x_j$) and incorrect (V*) result. Since we only have positive data, no input allows us to distinguish between either of the hypotheses and provide the information that we have found a superset of the language we are looking for. Gold has shown that none of the classes of grammars assumed in the theory of formal languages (for example, regular, context-free and context-sensitive languages) can be identified after a finite amount of steps given the input of a text with example utterances. This is true for all classes of languages that contain all finite languages and at least one infinite language. The situation is different if positive and negative data are used for learning instead of text.

The conclusion that has been drawn from Gold's results is that, for language acquisition, one requires knowledge that helps to avoid particular hypotheses from the start. Pullum (2003) criticizes the use of Gold's findings as evidence for the fact that linguistic knowledge must be innate. He lists a number of assumptions that have to be made in order for Gold's results to be relevant for the acquisition of natural languages. He then shows that each of these is not uncontroversial.

1. Natural languages could belong to the class of text-learnable languages as opposed to the class of context-free grammars mentioned above.

2. Learners could have information about which sequences of words are not grammatical (see p. 453–454 of Gold's essay for a similar conjecture). As has been shown since then, children do have direct negative evidence and there is also indirect negative evidence (see Section 13.8.4).

3. It is not clear whether learners really restrict themselves to exactly one grammar. Feldman (1972) has developed a learning procedure that eliminates all incorrect grammars at some point and is infinitely many times correct, however, it does not have to always choose one correct grammar and stick to the corresponding hypothesis. Using this procedure, it is possible to learn all recursively enumerable languages, that is, all languages for which there is a generative grammar. Pullum notes that even Feldman's learning procedure could prove to be too restrictive. It could take an entire lifetime for a learner to reach the correct grammar and they could have incorrect yet increasingly better hypotheses along the way.

4. Learners could work in terms of improvements. If one allows for a certain degree of tolerance, then acquisition is easier and it even becomes possible to learn the class of recursively enumerable languages (Wharton 1974).

5. Language acquisition does not necessarily constitute the acquisition of knowledge about a particular set of sequences, that is, the acquisition of a generative grammar capable of creating this set. The situation is completely different if grammars are viewed as a set of constraints that partially describe linguistic structures, but not necessarily a unique set of linguistic structures (for more on this point, see Section 6.7 and Chapter 14).

Furthermore, Pullum notes that it is also possible to learn the class of context-sensitive grammars with Gold's procedure with positive input only in a finite number of steps, if there is an upper bound k for the number of rules, where k is an arbitrary number. It is possible to make k so big that the cognitive abilities of the human brain would not be able to use a grammar with more rules than this. Since it is normally assumed that natural languages can be described by context-sensitive grammars, it can therefore be shown that the syntax of natural languages in Gold's sense can be learned from texts (also, see Scholz & Pullum 2002: 195–196).

Johnson (2004) adds that there is another important point that has been overlooked in the discussion about language acquisition. Gold's problem of identifiability is different from the problem of language acquisition that has played an important role in the nativism debate. In order to make the difference clear, Johnson differentiates between identifiability (in the Goldian sense) and learnability in the sense of language acquisition. Identifiability for a language class C means that there must be a function f that for each environment E for each language L in C permanently converges on hypothesis L as the target language in a finite amount of time.

Johnson proposes the following as the definition of *learnability* (p. 585): *A class C of natural languages is learnable iff, given almost any normal human child and almost any normal linguistic environment for any language L in C, the child will acquire L (or something sufficiently similar to L) as a native language between the ages of one and five.* Johnson adds the caveat that this definition does not correspond to any theory of learnability in psycholinguistics, but rather it is a hint in the direction of a realistic conception of acquisition.

Johnson notes that in most interpretations of Gold's theorem, identifiability and learnability are viewed as one and the same and shows that this is not logically correct: the main difference between the two depends on the use of two quantifiers. Identifiability of *one* language L from a class C requires that the learner converges on L in *every* environment after a finite amount of time. This time can differ greatly from environment to environment. There is not even an upper bound for the time in question. It is straightforward to construct a sequence of environments E_1, E_2, … for L, so that a learner in the environment E_i will not guess L earlier than the time t_i. Unlike identifiability, learnability means that there is a point in time after which in every normal environment, *every* normal child has converged on the correct language. This means that children acquire their language after a particular time span. Johnson quotes Morgan (1989: 352) claiming that children learn their native language after they have heard approximately 4,280,000 sentences. If we assume that the concept of learnability has a finite upper-bound for available time, then very few language classes can be identified in the limit. Johnson has shown this as follows: let C be a class of languages containing L and L', where L and L' have some elements in common. It is possible to construct a text such that the first n sentences are contained both in L and in L'. If the learner has L as its working hypothesis then continue the text with sentences from L', if he has L' as his hypothesis, then continue with sentences from L. In each case, the learner has entertained a false hypothesis after n steps. This means that identifiability is not a plausible model for language acquisition.

Aside from the fact that identifiability is psychologically unrealistic, it is not compatible with learnability (Johnson 2004: 586). For identifiability, only one learner has to be found (the function f mentioned above), learnability, however, quantifies over (almost) all normal children. If one keeps all factors constant, then it is easier to show the identifiability of a language class rather than its learnability. On the one hand, identifiability quantifies universally over all environments, regardless of whether these may seem odd or of how many repetitions these may contain. Learnability, on the other hand, has (almost) universal quantification exclusively over normal environments. Therefore, learnability refers to fewer environments than identifiability, such that there are less possibilities for problematic texts that could occur as an input and render a language unlearnable. Furthermore, learnability is defined in such a way that the learner does not have to learn L exactly, but rather learn something sufficiently similar to L. With respect to this aspect, learnability is a weaker property of a language class than identifiability. Therefore, learnability does not follow from identifiability nor the reverse.

Finally, Gold is dealing with the acquisition of syntactic knowledge without taking semantic knowledge into consideration. However, children possess a vast amount of information from the context that they employ when acquiring a language (Tomasello et al. 2005). As pointed out by Klein (1986: 44), humans do not learn anything if they are placed in a room and sentences in Mandarin Chinese are played to them. Language is acquired in a social and cultural context.

In sum, one should note that the existence of innate linguistic knowledge cannot be derived from mathematical findings about the learnability of languages.

13.8.2 Four case studies

Pullum & Scholz (2002) have investigated four prominent instances of the Poverty of the Stimulus Argument in more detail. These will be discussed in what follows. Pullum and Scholz's article appeared in a discussion volume. Arguments against their article are addressed by Scholz & Pullum (2002) in the same volume. Further PoS arguments from Chomsky (1986b) and from literature in German have been disproved by Eisenberg (1992).

13.8.2.1 Plurals in noun-noun compounding

Gordon (1986) claims that compounds in English only allow irregular plurals in compounds, that is, *mice-eater* but ostensibly not * *rats-eater*. Gordon claims that compounds with irregular plurals as first element are so rare that children could not have learned the fact that such compounds are possible purely from data.

On pages 25–26, Pullum and Scholz discuss data from English that show that regular plurals can indeed occur as the first element of a compound (*chemicals-maker*, *forms-reader*, *generics-maker*, *securities-dealer*, *drinks trolley*, *rules committee*, *publications catalogue*).[31] This shows that what could have allegedly not been learned from data is in fact not linguistically adequate and one therefore does not have to explain its acquisition.

13.8.2.2 Position of auxiliaries

The second study deals with the position of modal and auxiliary verbs. Kimball (1973: 73–75) discusses the data in (37) and the rule in (38) that is similar to one of the rules suggested by Chomsky (1957: 39) and is designed to capture the following data:

(37) a. It rains.
 b. It may rain.
 c. It may have rained.
 d. It may be raining.
 e. It has rained.
 f. It has been raining.
 g. It is raining.
 h. It may have been raining.

(38) Aux → T(M)(have+en)(be+ing)

T stands for tense, M for a modal verb and *en* stands for the participle morpheme (*-en* in *been*/*seen*/… and *-ed* in *rained*). The brackets here indicate the optionality of the expressions. Kimball notes that it is only possible to formulate this rule if (37h) is well-formed. If this were not the case, then one would have to reorganize the material in rules such that the three cases (M)(have+en), (M)(be+ing) and (have+en)(be+ing) would be covered.

[31] Also, see Abney (1996: 7) for examples from the Wall Street Journal.

Kimball assumes that children master the complex rule since they know that sentences such as (37h) are well-formed and since they know the order in which modal and auxiliary verbs must occur. Kimball assumes that children do not have positive evidence for the order in (37h) and concludes from this that the knowledge about the rule in (38) must be innate.

Pullum and Scholz note two problems with this PSA: first, they have found hundreds of examples, among them some from children's stories, so that the Kimball's claim that sentences such as (37h) are "vanishingly rare" should be called into question. For PSA arguments, one should at least specify how many occurrences there are allowed to be if one still wants to claim that nothing can be learned from them (Pullum & Scholz 2002: 29).

The second problem is that it does not make sense to assume that the rule in (37h) plays a role in our linguistic knowledge. Empirical findings have shown that this rule is not descriptively adequate. If the rule in (38) is not descriptively adequate, then it cannot achieve explanatory adequacy and therefore, one no longer has to explain how it can be acquired.

Instead of a rule such as (38), all theories discussed here currently assume that auxiliary or modal verbs embed a phrase, that is, one does not have an Aux node containing all auxiliary and modal verbs, but rather a structure for (37h) that looks as follows:

(39) It [may [have [been raining]]].

Here, the auxiliary or modal verb always selects the embedded phrase. The acquisition problem now looks completely different: a speaker has to learn the form of the head verb in the verbal projection selected by the auxiliary or modal verb. If this information has been learned, then it is irrelevant how complex the embedded verbal projections are: *may* can be combined with a non-finite lexical verb (37b) or a non-finite auxiliary (37c,d).

13.8.2.3 Reference of *one*

The third case study investigated by Pullum and Scholz deals with the pronoun *one* in English. Baker (1978: 413–425, 327–340) claims that children cannot learn that *one* can refer to constituents larger than a single word as in (40).

(40) a. I would like to tell you another funny story, but I've already told you the only *one* I know.
 b. The old man from France was more erudite than the young *one*.

Baker (416–417) claims that *one* can never refer to single nouns inside of NPs and supports this with examples such as (41):

(41) * The student of chemistry was more thoroughly prepared than the one of physics.

According to Baker, learners would require negative data in order to acquire this knowledge about ungrammaticality. Since learners – following his argumentation – never have access to negative evidence, they cannot possibly have learned the relevant knowledge and must therefore already possess it.

Pullum & Scholz (2002: 33) point out that there are acceptable examples with the same structure as the examples in (41):

(42) a. I'd rather teach linguistics to a student of mathematics than to one of any discipline in the humanities.

b. An advocate of Linux got into a heated discussion with one of Windows NT and the rest of the evening was nerd talk.

This means that there is nothing to learn with regard to the well-formedness of the structure in (41). Furthermore, the available data for acquiring the fact that *one* can refer to larger constituents is not as hopeless as Baker (p. 416) claims: there are examples that only allow an interpretation where *one* refers to a larger string of words. Pullum and Scholz offer examples from various corpora. They also provide examples from the CHILDES corpus, a corpus that contains communication with children (MacWhinney 1995). The following example is from a daytime TV show:

(43) A: "Do you think you will ever remarry again? I don't."

B: "Maybe I will, someday. But he'd have to be somebody very special. Sensitive and supportive, giving. Hey, wait a minute, where do they make guys like this?"

A: "I don't know. I've never seen one up close."

Here, it is clear that *one* cannot refer to *guys* since A has certainly already seen *guys*. Instead, it refers to *guys like this*, that is, men who are sensitive and supportive.

Once again, the question arises here as to how many instances a learner has to hear for it to count as evidence in the eyes of proponents of the PSA.

13.8.2.4 Position of auxiliaries in polar questions

The fourth PoS argument discussed by Pullum and Scholz comes from Chomsky and pertains to the position of the auxiliary in polar interrogatives in English. As shown on page 99, it was assumed in GB theory that a polar question is derived by movement of the auxiliary from the I position to the initial position C of the sentence. In early versions of Transformational Grammar, the exact analyses were different, but the main point was that the highest auxiliary is moved to the beginning of the clause. Chomsky (1971: 29–33) discusses the sentences in (44) and claims that children know that they have to move the highest auxiliary verb even without having positive evidence for this.[32] If, for example, they entertained the hypothesis that one simply places the first auxiliary at the beginning of the sentence, then this hypothesis would deliver the correct result (44b) for (44a), but not for (44c) since the polar question should be (44d) and not (44e).

(44) a. The dog in the corner is hungry.

[32] Examples with auxiliary movement are used in more recent PoS arguments too, for example in Berwick, Pietroski, Yankama & Chomsky (2011) and Chomsky (2013: 39). Work by Bod (2009b) is not discussed by the authors. For more on Bod's approach, see Section 13.8.3.

b. Is the dog in the corner hungry?
c. The dog that is in the corner is hungry.
d. Is the dog that is in the corner hungry?
e. * Is the dog that in the corner is hungry?

Chomsky claims that children do not have any evidence for the fact that the hypothesis that one simply fronts the linearly first auxiliary is wrong, which is why they could pursue this hypothesis in a data-driven learning process. He even goes so far as to claim that speakers of English only rarely or even never produce examples such as (44d) (Chomsky in Piattelli-Palmarini (1980: 114–115)). With the help of corpus data and plausibly constructed examples, Pullum (1996) has shown that this claim is clearly wrong. Pullum (1996) provides examples from the Wall Street Journal and Pullum & Scholz (2002) discuss the relevant examples in more detail and add to them with examples from the CHILDES corpus showing that adult speakers cannot only produce the relevant kinds of sentences, but also that these occur in the child's input.[33] Examples from CHILDES that disprove the hypothesis that the first auxiliary has to be fronted are given in (45):[34]

(45) a. Is the ball you were speaking of in the box with the bowling pin?
b. Where's this little boy who's full of smiles?
c. While you're sleeping, shall I make the breakfast?

Pullum and Scholz point out that *wh*-questions such as (45b) are also relevant if one assumes that these are derived from polar questions (see page 99 in this book) and if one wishes to show how the child can learn the structure-dependent hypothesis. This can be explained with the examples in (46): the base form from which (46a) is derived is (46b). If we were to front the first auxiliary in (46b), we would produce (46c).

(46) a. Where's the application Mark promised to fill out?[35]
b. the application Mark [$_{\text{AUX}}$ PAST] promised to fill out [$_{\text{AUX}}$ is] there
c. * Where did the application Mark promised to fill out is?

Evidence for the fact that (46c) is not correct can, however, also be found in language addressed to children. Pullum and Scholz provide the examples in (47):[36]

(47) a. Where's the little blue crib that was in the house before?
b. Where's the other dolly that was in here?
c. Where's the other doll that goes in there?

[33] For more on this point, see Sampson (1989: 223). Sampson cites part of a poem by William Blake, that is studied in English schools, as well as a children's encyclopedia. These examples surely do not play a role in acquisition of auxiliary position since this order is learned at the age of 3;2, that is, it has already been learned by the time children reach school age.

[34] See Lewis & Elman (2001). Researchers on language acquisition agree that the frequency of these kind of examples in communication with children is in fact very low. See Ambridge et al. (2008: 223).

[35] From the transcription of a TV program in the CHILDES corpus.

[36] These sentences are taken from NINA05.CHA in DATABASE/ENG/SUPPES/.

These questions have the form *Where's NP?*, where NP contains a relative clause.

In (45c), there is another clause preceding the actual interrogative, an adjunct clause containing an auxiliary as well. This sentence therefore provides evidence for falsehood of the hypothesis that the linearly first auxiliary must be fronted (Sampson 1989: 223).

In total, there are a number of attested sentence types in the input of children that would allow them to choose between the two hypotheses. Once again, the question arises as to how much evidence should be viewed as sufficient.

Pullum und Scholz's article has been criticized by Lasnik & Uriagereka (2002) and Legate & Yang (2002). Lasnik and Uriagereka argue that the acquisition problem is much bigger than presented by Pullum and Scholz since a learner without any knowledge about the language he was going to acquire could not just have the hypothesis in (48) that were discussed already but also the additional hypotheses in (49):

(48) a. Place the first auxiliary at the front of the clause.
     b. Place the first auxiliary in matrix-Infl at the front of the clause.

(49) a. Place any auxiliary at the front of the clause.
     b. Place any finite auxiliary at the front of the clause.

Both hypotheses in (49) would be permitted by the sentences in (50):

(50) a. Is the dog in the corner hungry?
     b. Is the dog that is in the corner hungry?

They would, however, also allow sentences such as (51):

(51) * Is the dog that in the corner is hungry?

The question that must now be addressed is why all hypotheses that allow (51) should be discarded since the learners do not have any information in their natural-linguistic input about the fact that (51) is not possible. They are lacking negative evidence. If (50b) is present as positive evidence, then this by no means implies that the hypothesis in (48b) has to be the correct one. Lasnik and Uriagereka present the following hypotheses that would also be compatible with (50b):

(52) a. Place the first auxiliary in initial position (that follows a change in intonation).
     b. Place the first auxiliary in initial position (that follows the first complete constituent).
     c. Place the first auxiliary in initial position (that follows the first parsed semantic unit).

These hypotheses do not hold for sentences such as (53) that contain a conjunction:

(53) Will those who are coming and those who are not coming raise their hands?

The hypotheses in (52) would also allow for sentences such as (54):

(54) * Are those who are coming and those who not coming will raise their hands?

Speakers hearing sentences such as (53) can reject the hypotheses (52) and thereby rule out (54), however, it is still possible to think of analogous implausible hypotheses that are compatible with all data previously discussed.

Legate & Yang (2002) take up the challenge of Pullum and Scholz and explicitly state how many occurrences one needs to acquire a particular phenomenon. They write the following:

> Suppose we have two independent problems of acquisition, P_1 and P_2, each of which involves a binary decision. For P_1, let F_1 be the frequency of the data that can settle P_1 one way or another, and for P_2, F_2. Suppose further that children successfully acquire P_1 and P_2 at roughly the same developmental stage. Then, under any theory that makes quantitative predictions of language development, we expect F_1 and F_2 to be roughly the same. Conversely, if F_1 and F_2 turn out significantly different, then P_1 and P_2 must represent qualitatively different learning problems.
>
> Now let P_1 be the auxiliary inversion problem. The two choices are the structure-dependent hypothesis (3b-i) and the first auxiliary hypothesis (3a-i). (Legate & Yang 2002: 155)

The position of auxiliaries in English is learned by children at the age of 3;2. According to Legate and Yang, another acquisition phenomenon that is learned at the age of 3;2 is needed for comparison. The authors focus on subject drop[37], that is learned at 36 months (two months earlier than auxiliary movement). According to the authors, acquisition problems involve a binary decision: in the first case, one has to choose between the two hypotheses in (48). In the second case, the learner has to determine whether a language uses overt subjects. The authors assume that the use of expletives such as *there* serves as evidence for learners that the language they are learning is not one with optional subjects. They then count the sentences in the CHILDES corpus that contain *there*-subjects and estimate F_2 at 1,2 % of the sentences heard by the learner. Since, in their opinion, we are dealing with equally difficult phenomena here, sentences such as (44d) and (47) should constitute 1.2 % of the input in order for auxiliary movement to be learnable.

The authors then searched in the Nina and Adam corpora (both part of CHILDES) and note that 0,068 to 0,045 % of utterances have the form of (47) and none have the form of (44d). They conclude that this number is not sufficient as positive evidence.

Legate and Yang are right in pointing out that Pullum and Scholz's data from the Wall Street Journal are not necessarily relevant for language acquisition and also in pointing out that examples with complex subject noun phrases do not occur in the data or at least to a negligible degree. There are, however, three serious problems with their argumentation: first, there is no correlation between the occurrence of expletive subjects

[37] This phenomenon is also called *pro-drop*. For a detailed discussion of the pro-drop parameter see Section 16.1.

and the property of being a pro-drop language: Galician (Raposo & Uriagereka 1990: Section 2.5) is a pro-drop language with subject expletive pronouns, in Italian there is an existential expletive *ci*,[38] even though Italian counts as a pro-drop language, Franks (1995) lists Upper and Lower Sorbian as pro-drop languages that have expletives in subject position. Since therefore expletive pronouns have nothing to do with the pro-drop parameter, their frequency is irrelevant for the acquisition of a parameter value. If there were a correlation between the possibility of omitting subjects and the occurrence of subject expletives, then Norwegian and Danish children should learn that there has to be a subject in their languages earlier than children learning English since expletives occur a higher percentage of the time in Danish and Norwegian (Scholz & Pullum 2002: 220). In Danish, the constructions corresponding to *there*-constructions in English are twice as frequent. It is still unclear whether there are actually differences in rate of acquisition (Pullum 2009: 246).

Second, in constructing their Poverty of the Stimulus argument, Legate and Yang assume that there is innate linguistic knowledge (the pro-drop parameter). Therefore their argument is circular since it is supposed to show that the assumption of innate linguistic knowledge is indispensable (Scholz & Pullum 2002: 220).

The third problem in Legate and Yang's argumentation is that they assume that a transformational analysis is the only possibility. This becomes clear from the following citation (Legate & Yang 2002: 153):

> The correct operation for question formation is, of course, structure dependent: it involves parsing the sentence into structurally organized phrases, and fronting the auxiliary that follows the subject NP, which can be arbitrarily long:
>
> (4) a. Is [the woman who is singing] e happy?
>   b. Has [the man that is reading a book] e eaten supper?

The analysis put forward by Chomsky (see page 99) is a transformation-based one, that is, a learner has to learn exactly what Legate and Yang describe: the auxiliary must move in front of the subject noun phrase. There are, however, alternative analyses that do not require transformations or equivalent mechanisms. If our linguistic knowledge does not contain any information about transformations, then their claim about what has to be learned is wrong. For example, one can assume, as in Categorial Grammar, that auxiliaries form a word class with particular distributional properties. One possible placement for them is initial positions as observed in questions, the alternative is after the subject (Villavicencio 2002: 104). There would then be the need to acquire information about whether the subject is realized to the left or to the right of its head. As an alternative to this lexicon-based analysis, one could pursue a Construction Grammar (Fillmore 1988: 44; 1999; Kay & Fillmore 1999: 18), Cognitive Grammar (Dąbrowska 2004: Chapter 9), or HPSG (Ginzburg & Sag 2000) approach. In these frameworks, there are

[38] However, *ci* is not treated as an expletive by all authors. See Remberger (2009) for an overview.

simply two[39] schemata for the two sequences that assign different meanings according to the order of verb and subject. The acquisition problem is then that the learners have to identify the corresponding phrasal patterns in the input. They have to realize that Aux NP VP is a well-formed structure in English that has interrogative semantics. The relevant theories of acquisition in the Construction Grammar-oriented literature have been very well worked out (see Section 16.3 and 16.4). Construction-based theories of acquisition are also supported by the fact that one can see that there are frequency effects, that is, auxiliary inversion is first produced by children for just a few auxiliaries and only in later phases of development is it then extended to all auxiliaries. If speakers have learned that auxiliary constructions have the pattern Aux NP VP, then the coordination data provided by Lasnik and Uriagereka in (53) no longer pose a problem since, if we only assign the first conjunct to the NP in the pattern Aux NP VP, then the rest of the coordinate structure (*and those who are not coming*) remains unanalyzed and cannot be incorporated into the entire sentence. The hearer is thereby forced to revise his assumption that *will those who are coming* corresponds to the sequence Aux NP in Aux NP VP and instead to use the entire NP *those who are coming and those who are not coming*. For acquisition, it is therefore enough to simply learn the pattern Aux NP VP first for some and then eventually for all auxiliaries in English. This has also been shown by Lewis & Elman (2001), who trained a neural network exclusively with data that did not contain NPs with relative clauses in auxiliary constructions. Relative clauses were, however, present in other structures. The complexity of the training material was increased bit by bit just as is the case for the linguistic input that children receive (Elman 1993).[40] The neural network can predict the next symbol after a sequence of words. For sentences with interrogative word order, the predictions are correct. Even the relative pronoun in (55) is predicted despite the sequence Aux Det N Relp never occurring in the training material.

(55) Is the boy who is smoking crazy?

Furthermore, the system signals an error if the network is presented with the ungrammatical sentence (56):

(56) * Is the boy who smoking is crazy?

A gerund is not expected after the relative pronoun, but rather a finite verb. The constructed neural network is of course not yet an adequate model of what is going on in our heads during acquisition and speech production.[41] The experiment shows, however, that the input that the learner receives contains rich statistical information that can be used

[39] Fillmore (1999) assumes subtypes of the Subject Auxiliary Inversion Construction since this kind of inversion does not only occur in questions.

[40] There are cultural differences. In some cultures, adults do not talk to children that have not attained full linguistic competence (Ochs 1982; Ochs & Schieffelin 1984) (also see Section 13.8.4). Children have to therefore learn the language from their environment, that is, the sentences that they hear reflect the full complexity of the language.

[41] See Hurford (2002: 324) and Jackendoff (2007: Section 6.2) for problems that arise for certain kinds of neural networks and Pulvermüller (2003, 2010) for an alternative architecture that does not have these problems.

when acquiring language. Lewis and Elman point out that the statistical information about the distribution of words in the input is not the only information that speakers have. In addition to information about distribution, they are also exposed to information about the context and can make use of phonological similarities in words.

In connection to the ungrammatical sentences in (56), it has been claimed that the fact that such sentences can never be produced shows that children already know that grammatical operations are structure-dependent and this is why they do not entertain the hypothesis that it is simply the linearly first verb that is moved (Crain & Nakayama 1987). The claim simply cannot be verified since children do not normally form the relevant complex utterances. It is therefore only possible to experimentally illicit utterances where they could make the relevant mistakes. Crain & Nakayama (1987) have carried out such experiments. Their study has been criticized by Ambridge, Rowland & Pine (2008) since these authors could show that children do really make mistakes when fronting auxiliaries. The authors put the difference to the results of the first study by Crain and Nakayama down to unfortunate choice of auxiliary in Crain and Nakayama's study. Due to the use of the auxiliary *is*, the ungrammatical examples had pairs of words that never or only very rarely occur next to each other (*who running* in (57a)).

(57) a. The boy who is running fast can jump high. →
* Is the boy who running fast can jump high?

b. The boy who can run fast can jump high. →
* Can the boy who run fast can jump high?

If one uses the auxiliary *can*, this problem disappears since *who* and *run* certainly do appear together. This then leads to the children actually making mistakes that they should not have, as the incorrect utterances actually violate a constraint that is supposed to be part of innate linguistic knowledge.

Estigarribia (2009) investigated English polar questions in particular. He shows that not even half of the polar questions in children's input have the form Aux NP VP (p. 74). Instead, parents communicated with their children in a simplified form and used sentences such as:

(58) a. That your tablet?
b. He talking?
c. That taste pretty good?

Estigarribia divides the various patterns into complexity classes of the following kind: FRAG (*fragmentary*), SPRED (*subject predicate*) and AUX-IN (*auxiliary inversion*). (59) shows corresponding examples:

(59) a. coming tomorrow? (FRAG)
b. you coming tomorrow? (SPRED)
c. Are you coming tomorrow? (AUX-IN)

What we see is that the complexity increases from class to class. Estigarribia suggests a system of language acquisition where simpler classes are acquired before more complex ones and the latter ones develop from peripheral modifications of more simple classes (p. 76). He assumes that question forms are learned from right to left (*right to left elaboration*), that is, (59a) is learned first, then the pattern in (59b) containing a subject in addition to the material in (59a), and then in a third step, the pattern (59c) in which an additional auxiliary occurs (p. 82). In this kind of learning procedure, no auxiliary movement is involved. This view is compatible with constraint-based analyses such as that of Ginzburg & Sag (2000). A similar approach to acquisition by Freudenthal, Pine, Aguado-Orea & Gobet (2007) will be discussed in Section 16.3.

A further interesting study has been carried out by Bod (2009b). He shows that it is possible to learn auxiliary inversion assuming trees with any kind of branching even if there is no auxiliary inversion with complex noun phrases present in the input. The procedure he uses as well as the results he gains are very interesting and will be discussed in Section 13.8.3 in more detail.

In conclusion, we can say that children do make mistakes with regard to the position of auxiliaries that they probably should not make if the relevant knowledge were innate. Information about the statistical distribution of words in the input is enough to learn the structures of complex sentences without actually having this kind of complex sentences in the input.

13.8.2.5 Summary

Pullum & Scholz (2002: 19) show what an Argument from Poverty of the Stimulus (APS) would have to look like if it were constructed correctly:

(60) APS specification schema:
   a. ACQUIRENDUM CHARACTERIZATION: describe in detail what is alleged to be known.
   b. LACUNA SPECIFICATION: identify a set of sentences such that if the learner had access to them, the claim of data-driven learning of the acquirendum would be supported.
   c. INDISPENSABILITY ARGUMENT: give reason to think that if learning were data-driven, then the acquirendum could not be learned without access to sentences in the lacuna.
   d. INACCESSIBILITY EVIDENCE: support the claim that tokens of sentences in the lacuna were not available to the learner during the acquisition process.
   e. ACQUISITION EVIDENCE: give reason to believe that the acquirendum does in fact become known to learners during childhood.

As the four case studies have shown, there can be reasons for rejecting the acquirendum. If the acquirendum does not have to be acquired, than there is no longer any evidence for innate linguistic knowledge. The acquirendum must at least be descriptively adequate. This is an empirical question that can be answered by linguists. In three of the four PoS arguments discussed by Pullum and Scholz, there were parts which were not descriptively adequate. In previous sections, we already encountered other PoS arguments that involve claims regarding linguistic data that cannot be upheld empirically (for example, the Subjacency Principle). For the remaining points in (60), interdisciplinary work is required: the specification of the lacuna falls into the theory of formal language (the specification of a set of utterances), the argument of indispensability is a mathematical task from the realm of learning theory, the evidence for inaccessibility is an empirical question that can be approached by using corpora, and finally the evidence for acquisition is a question for experimental developmental psychologists (Pullum & Scholz 2002: 19–20).

Pullum & Scholz (2002: 46) point out an interesting paradox with regard to (60c): without results from mathematical theories of learning, one cannot achieve (60c). If one wishes to provide a valid Poverty of the Stimulus Argument, then this should automatically lead to improvements in theories of learning, that is, it is possible to learn more than was previously assumed.

13.8.3 Unsupervized Data-Oriented Parsing (U-DOP)

Bod (2009b) has developed a procedure that does not require any information about word classes or relations between words contained in utterances. The only assumption that one has to make is that there is some kind of structure. The procedure consists of three steps:

1. Compute all possible (binary-branching) trees (without category symbols) for a set of given sentences.
2. Divide these trees into sub-trees.
3. Compute the ideal tree for each sentence.

This process will be explained using the sentences in (61):

(61) a. Watch the dog.
b. The dog barks.

The trees that are assigned to these utterances only use the category symbol X since the categories for the relevant phrases are not (yet) known. In order to keep the example readable, the words themselves will not be given the category X, although one can of course do this. Figure 13.2 on the following page shows the trees for (61). In the next step, the trees are divided into subtrees. The trees in Figure 13.2 have the subtrees that can be seen in Figure 13.3 on page 489. In the third step, we now have to compute the best tree for each utterance. For *The dog barks.*, there are two trees in the set of the subtrees that correspond exactly to this utterance. But it is also possible to build structures out of subtrees. There are therefore multiple derivations possible for *The dog barks.* all of which use the trees in Figure 13.3: one the one hand, trivial derivations that use the entire tree, and on the other, derivations that build trees from smaller subtrees. Figure 13.4 on page 490 gives an impression of how this construction of subtrees happens. If we now want to decide which of the analyses in (62) is the best, then we have to compute the probability of each tree.

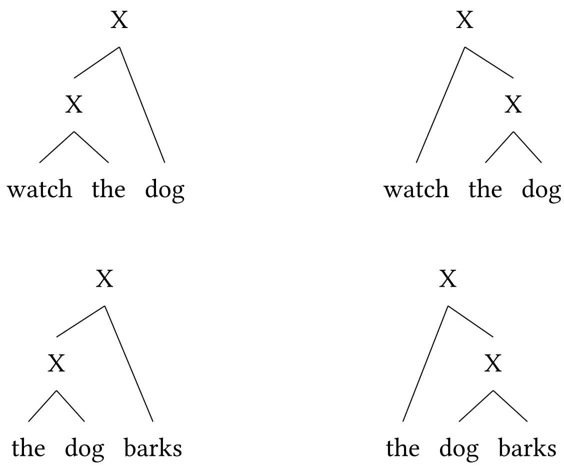


Figure 13.2: Possible binary-branching structures for *Watch the dog* and *The dog barks*.

(62) a. [[the dog] barks]
 b. [the [dog barks]]

The probability of a tree is the sum of the probabilities of all its analyses. There are two analyses for (62b), which can be found in Figure 13.4. The probability of the first analysis of (62b) corresponds to the probability of choosing exactly the complete tree for [the [dog barks]] from the set of all subtrees. Since there are twelve subtrees, the probability of choosing that one is 1/12. The probability of the second analysis is the product of the probabilities of the subtrees that are combined and is therefore $1/12 \times 1/12 = 1/144$. The probability of the analysis in (62b) is therefore $1/12 + (1/12 \times 1/12) = 13/144$. One can then calculate the probability of the tree in (62a) in the same way. The only difference here is that the tree for [the dog] occurs twice in the set of subtrees. Its probability is therefore 2/12. The probability of the tree [[the dog] barks] is therefore: $1/12 + (1/12 \times 2/12) = 14/144$. We have thus extracted knowledge about plausible structures from the corpus. This knowledge can also be applied whenever one hears a new utterance for which there is no complete tree. It is then possible to use already known subtrees to calculate the probabilities of possible analyses of the new utterance. Bod's model can also be combined with weights: those sentences that were heard longer ago by the speaker, will receive a lower weight. One can thereby also account for the fact that children do not have all sentences that they have ever heard available simultaneously. This extension makes the UDOP model more plausible for language acquisition.

In the example above, we did not assign categories to the words. If we were to do this, then we would get the tree in Figure 13.5 on page 490 as a possible subtree. These kinds of discontinuous subtrees are important if one wants to capture dependencies between elements that occur in different subtrees of a given tree. Some examples are the following

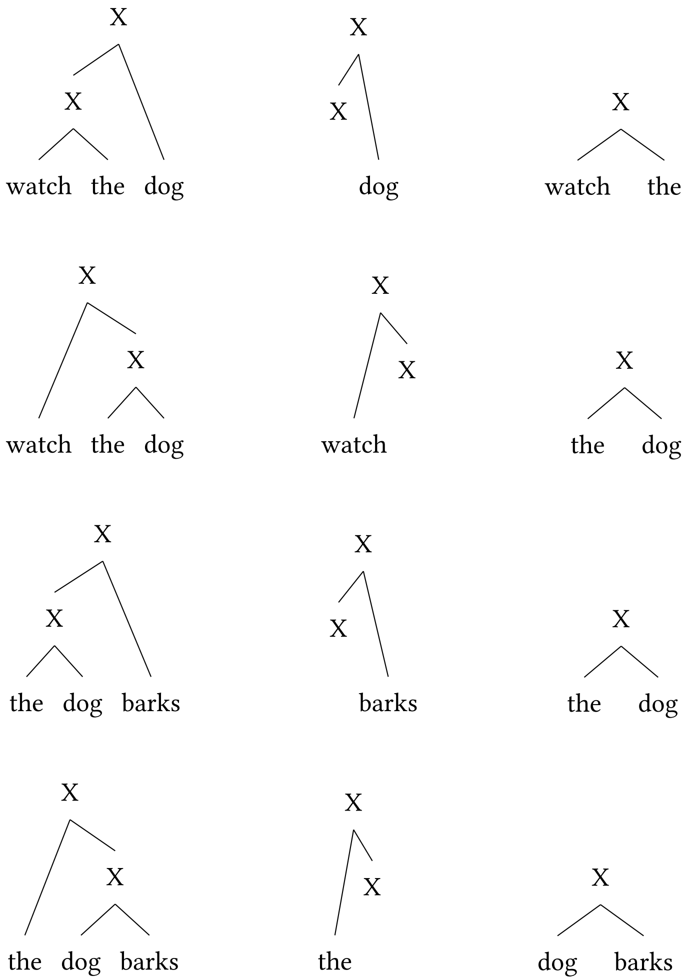


Figure 13.3: Subtrees for the trees in Figure 13.2

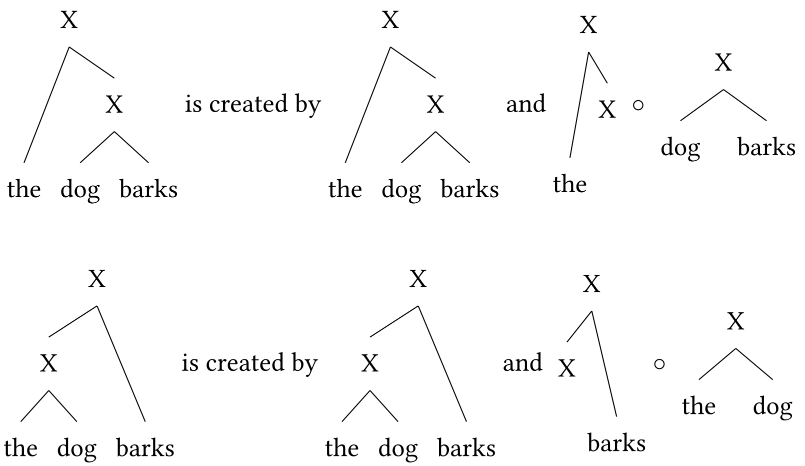


Figure 13.4: Analysis of *The dog barks* using subtrees from Figure 13.3

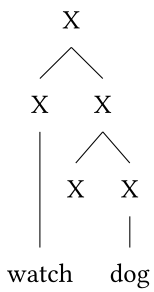


Figure 13.5: Discontinuous partial tree

sentences:

(63) a. BA carried *more* people *than* cargo in 2005.
  b. *What's* this scratch *doing* on the table?
  c. Most software *companies* in Vietnam *are* small sized.

It is then also possible to learn auxiliary inversion in English with these kinds of discontinuous trees. All one needs are tree structures for the two sentences in (64) in order to prefer the correct sentence (65a) over the incorrect one (65b).

(64) a. The man who is eating is hungry.
  b. Is the boy hungry?

(65) a. Is the man who is eating hungry?

b. * Is the man who eating is hungry?

U-DOP can learn the structures for (64) in Figure 13.6 from the sentences in (66):

(66) a. The man who is eating mumbled.
b. The man is hungry.
c. The man mumbled.
d. The boy is eating.

Note that these sentences do not contain any instance of the structure in (65a). With

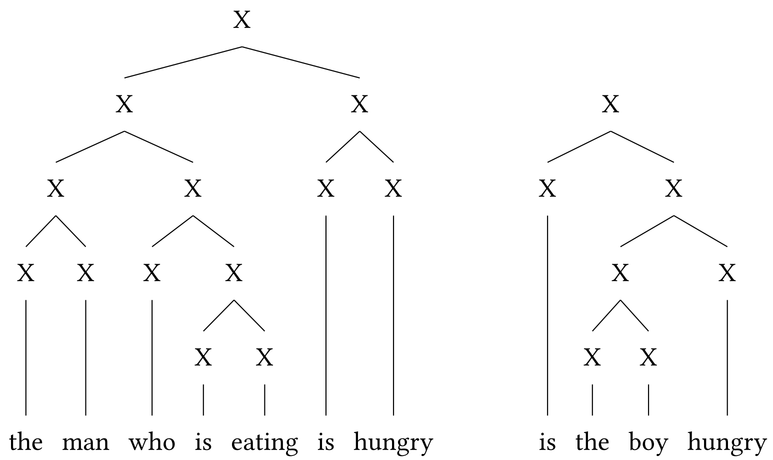


Figure 13.6: Structures that U-DOP learned from the examples in (64) and (66)

the structures learned here, it is possible to show that the shortest possible derivation for the position of the auxiliary is also the correct one: the correct order *Is the man who is eating hungry?* only requires that the fragments in Figure 13.7 on the next page are combined, whereas the structure for * *Is the man who eating is hungry?* requires at least four subtrees from Figure 13.6 to be combined with each other. This is shown by Figure 13.8 on the following page.

The motivation for always taking the derivation that consists of the least subparts is that one maximizes similarity to already known material.

The tree for (67) containing one auxiliary too many can also be created from Figure 13.6 with just two subtrees (with the tree [$_X$ is$_X$ X] and the entire tree for *The man who is eating is hungry*).

(67) * Is the man who is eating is hungry?

Interestingly, children do produce these kind of incorrect sentences (Crain & Nakayama 1987: 530; Ambridge, Rowland & Pine 2008). However, if we consider the probabilities

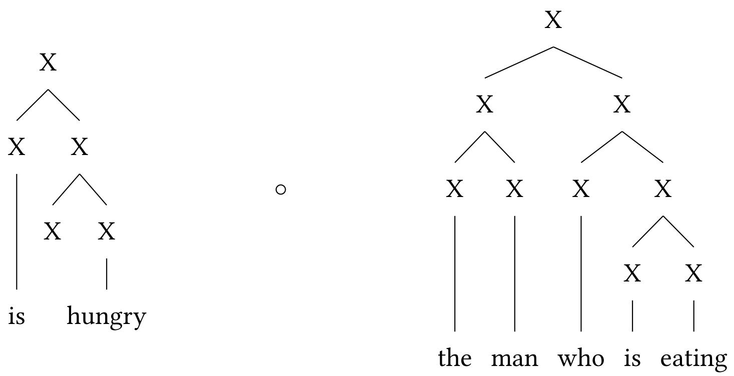


Figure 13.7: Derivation of the correct structure for combination with an auxiliary using two subtrees from Figure 13.6

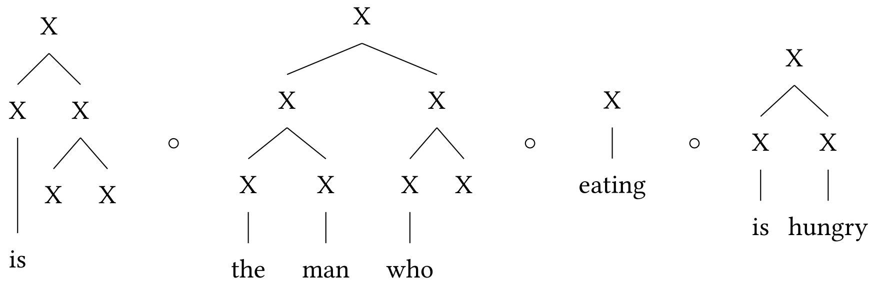


Figure 13.8: Derivation of the incorrect structure for the combination with an auxiliary using two subtrees from Figure 13.6

of the subtrees in addition to the the number of combined subparts, we get the correct result, namely (65a) and not (67). This is due to the fact that *the man who is eating* occurs in the corpus twice, in (65a) and in (66a). Thus, the probability of *the man who is eating* is just as high as the probability of *the man who is eating is hungry* and thus derivation in Figure 13.7 is preferred over the one for (67). This works for the constructed examples here, however one can imagine that in a realistic corpus, sequences of the form *the man who is eating* are more frequent than sequences with further words since *the man who is eating* can also occur in other contexts. Bod has applied this process to corpora of adult language (English, German and Chinese) as well as applying it to the Eve corpus from the CHILDES database in order to see whether analogy formation constitutes a plausible model for human acquisition of language. He was able to show that what we demonstrated for the sentences above also works for a larger corpus of naturally occurring language: although there were no examples for movement of an auxiliary across a complex NP in the Eve corpus, it is possible to learn by analogy that the auxiliary from a complex NP cannot be fronted.

It is therefore possible to learn syntactic structures from a corpus without any prior knowledge about parts of speech or abstract properties of language. The only assumption that Bod makes is that there are (binary-branching) structures. The assumption of binarity is not really necessary. But if one includes flat branching structures into the computation, the set of trees will become considerably bigger. Therefore, Rens Bod only used binary-branching structures in his experiments. In his trees, X consists of two other X's or a word. We are therefore dealing with recursive structures. Therefore, Bod's work proposes a theory of the acquisition of syntactic structures that only requires recursion, something that is viewed by Hauser, Chomsky & Fitch (2002) as a basic property of language.

As shown in Section 13.1.8, there is evidence that recursion is not restricted to language and thus one can conclude that it is not necessary to assume innate linguistic knowledge in order to be able to learn syntactic structures from the given input.

Nevertheless, it is important to point out something here: what Rens Bod shows is that syntactic structures can be learned. The information about the parts of speech of each word involved which are not yet included in his structures can also be derived using statistical methods (Redington et al. 1998; Clark 2000).[42] In all probability, the structures that can be learned correspond to structures that surface-oriented linguistic theories would also assume. However, not all aspects of the linguistic analysis are acquired. In Bod's model, only occurrences of words in structures are evaluated. Nothing is said about whether words stand in a particular regular relationship to one another or not (for example, a lexical rule connecting a passive participle and perfect participle). Furthermore, nothing is said about how the meaning of expressions arise (are they rather holistic in the sense of Construction Grammar or projected from the lexicon?). These are questions that still concern theoretical linguists (see Chapter 21) and cannot straightforwardly be derived from the statistic distribution of words and the structures computed from them (see Section 21.8.1 for more on this point).

A second comment is also needed: we have seen that statistical information can be used to derive the structure of complex linguistic expressions. This now begs the question of how this relates to Chomsky's earlier argumentation against statistical approaches (Chomsky 1957: 16). Abney (1996: Section 4.2) discusses this in detail. The problem with his earlier argumentation is that Chomsky referred to Markov models. These are statistical versions of finite automatons. Finite automatons can only describe type 3 languages and are therefore not appropriate for analyzing natural language. However, Chomsky's criticism cannot be applied to statistical methods in general.

[42] Computational linguistic algorithms for determining parts of speech often look at an entire corpus. But children are always dealing with just a particular part of it. The corresponding learning process must then also include a curve of forgetting. See Braine (1987: 67).

13.8.4 Negative evidence

In a number of works that assume innate linguistic knowledge, it is claimed that children do not have access to negative evidence, that is, nobody tells them that sentences such as (44e) – repeated here as (68) – are ungrammatical (Brown & Hanlon 1970: 42–52; Marcus 1993).

(68) * Is the dog that in the corner is hungry?

It is indeed correct that adults do not wake up their children with the ungrammatical sentence of the day, however, children do in fact have access to negative evidence of various sorts. For example, Chouinard & Clark (2003) have shown that English and French speaking parents correct the utterances of their children that are not well-formed. For example, they repeat utterances where the verb was inflected incorrectly. Children can deduce from the fact that the utterance was repeated and from what was changed in the repetition that they made a mistake and Chouinard and Clark also showed that they actually do this. The authors looked at data from five children whose parents all had an academic qualification. They discuss the parent-child relationship in other cultures, too (see Ochs (1982); Ochs & Schieffelin (1984) and Marcus (1993: 71) for an overview) and refer to studies of America families with lower socio-economic status (page 660).

A further form of negative evidence is indirect negative evidence, which Chomsky (1981a: 9) also assumes could play a role in acquisition. Goldberg (1995: Section 5.2) gives the utterance in (69a) as an example:[43]

(69) a. Look! The magician made the bird disappear.
b. * The magician disappeared the bird.

The child can conclude from the fact that adults use a more involved causative construction with *make* that the verb *disappear*, unlike other verbs such as *melt*, cannot be used transitively. An immediately instructive example for the role played by indirect negative evidence comes from morphology. There are certain productive rules that can however still not be applied if there is a word that blocks the application of the rule. An example is the *-er* nominalization suffix in German. By adding an *-er* to a verb stem, one can derive a noun that refers to someone who carries out a particular action (often habitually) (*Raucher* 'smoker', *Maler* 'painter', *Sänger* 'singer', *Tänzer* 'dancer'). However, *Stehler* 'stealer' is very unusual. The formation of *Stehler* is blocked by the existence of *Dieb* 'thief'. Language learners therefore have to infer from the non-existence of *Stehler* that the nominalization rule does not apply to *stehlen* 'to steal'.

Similarly, a speaker with a grammar of English that does not have any restrictions on the position of manner adverbs would expect that both orders in (70) are possible (Scholz & Pullum 2002: 206):

(70) a. call the police immediately
b. * call immediately the police

[43] Also, see Tomasello (2006a: 277).

Learners can conclude indirectly from the fact that verb phrases such as (70b) (almost) never occur in the input that these are probably not part of the language. This can be modeled using the relevant statistical learning algorithms.

The examples for the existence of negative evidence provided so far are more arguments from plausibility. Stefanowitsch (2008) has combined corpus linguistic studies on the statistical distribution with acceptability experiments and has shown that negative evidence gained from expected frequencies correlates with acceptability judgments of speakers. This process will be discussed now briefly: Stefanowitsch assumes the following principle:

(71) Form expectations about the frequency of co-occurrence of linguistic features or elements on the basis of their individual frequency of occurrence and check these expectations against the actual frequency of co-occurrence. (Stefanowitsch 2008: 518)

Stefanowitsch works with the part of the *International Corpus of English* that contains British English (ICE-GB). In this corpus, the verb *say* occurs 3,333 times and sentences with ditransitive verbs (Subj Verb Obj Obj) occur 1,824 times. The entire total of verbs in the corpus is 136,551. If all verbs occurred in all kinds of sentences with the same frequencies, then we would expect *say* to occur 44.52 times (X / 1,824 = 3,333 / 136,551 and hence X = 1,824 × 3,333 / 136,551) in the ditransitive construction. But the number of actual occurrences is actually 0 since, unlike (72b), sentences such as (72a) are not used by speakers of English.

(72) a. * Dad said Sue something nice.
     b. Dad said something nice to Sue.

Stefanowitsch shows that the non-occurrence of *say* in the ditransitive sentence pattern is significant. Furthermore, he investigated how acceptability judgments compare to the frequent occurrence or non-occurrence of verbs in certain constructions. In a first experiment, he was able to show that the frequent non-occurrence of elements in particular constructions correlates with the acceptability judgments of speakers, whereas this is not the case for the frequent occurrence of a verb in a construction.

In sum, we can say that indirect negative evidence can be derived from linguistic input and that it seems to play an important role in language acquisition.

13.9 Summary

It follows from all this that not a single one of the arguments in favor of innate linguistic knowledge remains uncontroversial. This of course does not rule out there still being innate linguistic knowledge, however, those who wish to incorporate this assumption into their theories have to take more care than was previously the case to prove that what they assume to be innate is actually part of our linguistic knowledge and that it cannot be learned from the linguistic input alone.

Comprehension questions

1. Which arguments are there for the assumption of innate linguistic knowledge?

Further reading

Pinker's book (1994) is the best written book arguing for nativist models of language.

Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett (1996) discuss all the arguments that have been proposed in favor of innate linguistic knowledge and show that the relevant phenomena can be explained differently. The authors adopt a connectionist view. They work with neuronal networks, which are assumed to model what is happening in our brains relatively accurately. The book also contains chapters about the basics of genetics and the structure of the brain, going into detail about why a direct encoding of linguistic knowledge in our genome is implausible.

Certain approaches using neuronal networks have been criticized because they cannot capture certain aspects of human abilities such as recursion or the multiple usage of the same words in an utterance. Pulvermüller (2010) discusses an architecture that has memory and uses this to analyze recursive structures. In his overview article, certain works are cited that show that the existence of more abstract rules or schemata of the kind theoretical linguists take for granted can be demonstrated on the neuronal level. Pulvermüller does not, however, assume that linguistic knowledge is innate (p. 173).

Pullum and Scholz have dealt with the Poverty-of-the-Stimulus argument in detail (Pullum & Scholz 2002; Scholz & Pullum 2002).

Goldberg (2006) and Tomasello (2003) are the most prominent proponents of Construction Grammar, a theory that explicitly tries to do without the assumption of innate linguistic knowledge.