

## 16 Language acquisition

Linguists and philosophers are fascinated by the human ability to acquire language. Assuming the relevant input during childhood, language acquisition normally takes place completely effortlessly. Chomsky (1965: 24–25) put forward the requirement that a grammatical theory must provide a plausible model of language acquisition. Only then could it actually explain anything and would otherwise remain descriptive at best. In this section, we will discuss theories of acquisition from a number of theoretical standpoints.

### 16.1 Principles and Parameters

A very influential explanation of language acquisition is Chomsky's Principles and Parameters model (1981a). Chomsky assumes that there is an innate Universal Grammar that contains knowledge that is equally relevant for all languages. Languages can then vary in particular ways. For every difference between languages in the area of core grammar, there is a feature with a specific value. Normally, the value of a parameter is binary, that is, the value is either '+' or '–'. Depending on the setting of a parameter, a language will have certain properties, that is, setting a parameter determines whether a language belongs to a particular class of languages. Parameters are assumed to influence multiple properties of a grammar simultaneously (Chomsky 1981a: 6). For example, Rizzi (1986) claims that the pro-drop parameter affects whether referential subjects can be omitted, the absence of expletives, subject extraction from clauses with complementizers (*that*-t contexts) and interrogatives and finally the possibility of realizing the subject postverbally in VO-languages (see Chomsky 1981a: Section 4.3; Meisel 1995: 12). It has been noted that there are counter-examples to all the correlations assumed.<sup>1</sup> Another example of a parameter is the Head Directionality Parameter discussed in Section 13.1.1. As was shown, there are languages where heads govern in different directions. In his overview article, Haider (2001) still mentions the parameterized Subacency Principle but notes that subacency is no longer assumed as a principle in newer versions of the theory (see Section 13.1.5.2 for more on subacency).

Snyder (2001) discovered a correlation of various phenomena with productive root compounding as it is manifested for instance in compounding of two nouns. He argues

---

<sup>1</sup> See Haider (1994) and Haider (2001: Section 2.2) for an overview. Haider assumes that there is at least a correlation between the absence of expletive subjects and pro-drop. However, Galician is a pro-drop language with expletive subject pronouns (Raposo & Uriagereka 1990: Section 2.5). Franks (1995: 314) cites Upper and Lower Sorbian as pro-drop languages with expletive subjects. Scholz & Pullum (2002: 218) point out that there is an expletive pronoun *ci* in modern Italian although Italian is classed as a pro-drop language.

16 *Language acquisition*

that the acquisition of complex predicate formation is connected to the acquisition of compound structures and that there is a parameter that is responsible for this type of compounding and simultaneously for the following set of phenomena:

- |     |  |                           |
|-----|--|---------------------------|
| (1) | a. John painted the house red.                 | (resultative)             |
|     | b. Mary picked the book up/picked up the book. | (verb-particle)           |
|     | c. Fred made Jeff leave.                       | ( <i>make</i> -causative) |
|     | d. Fred saw Jeff leave.                        | (perceptual report)       |
|     | e. Bob put the book on the table.              | ( <i>put</i> -locative)   |
|     | f. Alice sent the letter to Sue.               | ( <i>to</i> -dative)      |
|     | g. Alice sent Sue the letter.                  | (double-object dative)    |

Snyder examined languages from various language groups (Afroasiatic, Austroasiatic, Austronesian, Finno-Ugric, Indo-European (Germanic, Romance, Slavic), Japanese-Korean, Niger-Kordofanian (Bantu), and Sino-Tibetan, as well as American Sign Language and the language isolate Basque). The languages that were examined either had all of these phenomena or none. This was tested with native speakers of the respective languages. In addition the claim that these phenomena are acquired once noun-noun compounds are used productively was tested for English using CHILDES data. The result was positive with the exception of the double object construction, for which an explanation was provided. The correlation of the phenomena in (1) is interesting and was interpreted as proof of the existence of a parameter that correlates several phenomena in a language. However, Son (2007) and Son & Svenonius (2008) showed that Snyder's claims for Japanese were wrong and that there are further languages like Korean, Hebrew, Czech, Malayalam, Javanese in which some of the phenomena show no correlations.

Gibson & Wexler (1994) discuss the acquisition of constituent order and assume three parameters that concern the position of the verb relative to the subject (SV vs. VS) and relative to the object (VO vs. OV) as well as the V2-property. There is no consensus in the literature about which parameters determine the make-up of languages (see Newmeyer (2005: Section 3.2) and Haspelmath (2008) for an overview and critical discussion). Fodor (1998a: 346–347) assumes that there are 20 to 30 parameters, Gibson & Wexler (1994: 408) mention the number 40, Baker (2003: 349) talks of 10 to 20 and Roberts & Holmberg (2005: 541) of 50 to 100. There is no consensus in the literature as to which parameters one should assume, how they interact and what they predict. However, it is nevertheless possible to contemplate how a grammar of an individual language could be derived from a UG with parameters that need to be set. Chomsky's original idea (1986b: Section 3.5.1) was that the child sets the value of a parameter based on the language input as soon as the relevant evidence is present from the input (see also Gibson & Wexler 1994; Nowak, Komarova & Niyogi 2001). At a given point in time, the learner has a grammar with certain parameter settings that correspond to the input seen so far. In order to fully acquire a grammar, all parameters must be assigned a value. In theory, thirty utterances should be enough to acquire a grammar with thirty parameters if these utterances provide unambiguous evidence for a particular parameter value.

## 16.1 Principles and Parameters

This approach has often been criticized. If setting a parameter leads to a learner using a different grammar, one would expect sudden changes in linguistic behavior. This is, however, not the case (Bloom 1993: 731). Fodor (1998a: 343–344) also notes the following three problems: 1) Parameters can affect things that are not visible from the perceptible constituent order. 2) Many sentences are ambiguous with regard to the setting of a particular parameter, that is, there are sometimes multiple combinations of parameters compatible with one utterance. Therefore, the respective utterances cannot be used to set any parameters (Berwick & Niyogi 1996; Fodor 1998b). 3) There is a problem with the interaction of parameters. Normally multiple parameters play a role in an utterance such that it can be difficult to determine which parameter contributes what and thus how the values should be determined.

Points 1) and 2) can be explained using the constituent order parameters of Gibson and Wexler: imagine a child hears a sentences such as (2):

- (2) a. Daddy drinks juice.  
       b. Pappa trinkt Saft.  
            daddy drinks juice

These sentences look exactly the same, even though radically different structures are assumed for each. According to the theories under discussion, the English sentence has the structure shown in Figure 3.8 on page 103 given in abbreviated form in (3a). The German sentence, on the other hand, has the structure in Figure 3.13 on page 110 corresponding to (3b):

- (3) a. [<sub>IP</sub> [Daddy [<sub>I'</sub> <sub>–k</sub> [<sub>VP</sub> drinks<sub>k</sub> juice]]]].  
       b. [<sub>CP</sub> Pappa<sub>i</sub> [<sub>C'</sub> trinkt<sub>k</sub> [<sub>IP</sub> <sub>–i</sub> [<sub>I'</sub> [<sub>VP</sub> Saft <sub>–k</sub>] <sub>–k</sub>]]]].

English has the basic constituent order SVO. The verb forms a constituent with the object (VP) and this is combined with the subject. The parameter setting must therefore be SV, VO and –V2. German, on the other hand, is analyzed as a verb-final and verb-second language and the parameter values would therefore have to be SV, OV and +V2. If we consider the sentences in (2), we see that both sentences do not differ from one another with regard to the order of the verb and its arguments.

Fodor (1998a,b) concludes from this that one first has to build a structure in order to see what grammatical class the grammar licensing the structure belongs to since one first needs the structure in (3b) in order to be able to see that the verb in the partial constituent occurs after its argument in the VP (Saft <sub>–k</sub>). The question is now how one achieves this structure. A UG with 30 parameters corresponds to  $2^{30} = 1,073,741,824$  fully instantiated grammars. It is an unrealistic assumption that children try out these grammars successively or simultaneously.

Gibson & Wexler (1994) discuss a number of solutions for this problem: parameters have a default value and the learner can only change a parameter value if a sentence that could previously not be analyzed can then be analyzed with the new parameter setting (*Greediness Constraint*). In this kind of procedure, only one parameter can be changed at a time (*Single Value Constraint*), which aims at ruling out great leaps leading

16 *Language acquisition*

to extremely different grammars (see Berwick & Niyogi 1996: 612–613, however). This reduces the processing demands, however with 40 parameters, the worst case could still be that one has to test 40 parameter values separately, that is, try to parse a sentence with 40 different grammars. This processing feat is still unrealistic, which is why Gibson & Wexler (1994: 442) additionally assume that one hypothesis is tested per input sentence. A further modification of the model is the assumption that certain parameters only begin to play a role during the maturation of the child. At a given point in time, there could be only a few accessible parameters that also need to be set. After setting these parameters, new parameters could become available.

In their article, Gibson and Wexler show that the interaction between input and parameter setting is in no way trivial. In their example scenario with three parameters, a situation can arise in which a learner sets a parameter in order to analyze a new sentence, however setting this parameter leads to the fact that the target grammar cannot be acquired because only one value can be changed at a time and changes can only be made if more sentences can be analyzed than before. The learner reaches a so-called local maximum in these problematic cases.<sup>2</sup> Gibson and Wexler then suggest assigning a default value to particular parameters, whereby the default value is the one that will cause the learner to avoid problematic situations. For the V2 parameter, they assume ‘–’ as the default value.

Berwick & Niyogi (1996) show that Gibson and Wexler calculated the problematic conditions incorrectly and that, if one shares their assumptions, it is even more frequently possible to arrive at parameter combinations from which it is not possible to reach the target grammar by changing individual parameter values. They show that one of the problematic cases not addressed by Gibson and Wexler is –V2 (p. 609) and that the assumption of a default value for a parameter does not solve the problem as both ‘+’ and ‘–’ can lead to problematic combinations of parameters.<sup>3</sup> In their article, Berwick and Niyogi show that learners in the example scenario above (with three parameters) learn the target grammar faster if one abandons the Greediness or else the Single Value Constraint. They suggest a process that simply randomly changes one parameter if a sentence cannot be analyzed (*Random Step*, p. 615–616). The authors note that this approach does not share the problems with the local maxima that Gibson and Wexler had in their example and that it also reaches its goal faster than theirs. However, the fact that *Random Step* converges more quickly has to do with the quality of the parameter space (p. 618). Since there is no consensus about parameters in the literature, it is not possible to assess how the entire system works.

Yang (2004: 453) has criticized the classic Principles & Parameters model since abrupt switching between grammars after setting a parameter cannot be observed. Instead, he proposes the following learning mechanism:

<sup>2</sup> If one imagines the acquisition process as climbing a hill, then the Greediness Constraint ensures that one can only go uphill. It could be the case, however, that one begins to climb the wrong hill and can no longer get back down.

<sup>3</sup> Kohl (1999, 2000) has investigated this acquisition model in a case with twelve parameters. Of the 4096 possible grammars, 2336 (57%) are unlearnable if one assumes the best initial values for the parameters.

## 16.1 Principles and Parameters

- (4) For an input sentence,  $s$ , the child: (i) with probability  $P_i$  selects a grammar  $G_i$ ,  
 (ii) analyzes  $s$  with  $G_i$ ,  
 (iii) if successful, reward  $G_i$  by increasing  $P_i$ , otherwise punish  $G_i$  by decreasing  $P_i$ .

Yang discusses the example of the pro-drop and topic drop parameters. In pro-drop languages (e. g. Italian), it is possible to omit the subject and in topic drop languages (e. g. Mandarin Chinese), it possible to omit both the subject and the object if it is a topic. Yang compares English-speaking and Chinese-speaking children noting that English children omit both subjects and objects in an early linguistic stage. He claims that the reason for this is that English-speaking children start off using the Chinese grammar.

The pro-drop parameter is one of the most widely discussed parameters in the context of Principles & Parameters theory and it will therefore be discussed in more detail here. It is assumed that speakers of English have to learn that all sentences in English require a subject, whereas speakers of Italian learn that subjects can be omitted. One can observe that children learning both English and Italian omit subjects (German children too in fact). Objects are also omitted notably more often than subjects. There are two possible explanations for this: a competence-based one and a performance-based one. In competence-based approaches, it is assumed that children use a grammar that allows them to omit subjects and then only later acquire the correct grammar (by setting parameters or increasing the rule apparatus). In performance-based approaches, by contrast, the omission of subjects is traced back to the fact that children are not yet capable of planning and producing long utterances due to their limited brain capacity. Since the cognitive demands are greatest at the beginning of an utterance, this leads to subjects being increasingly left out. Valian (1991) investigated these various hypotheses and showed that the frequency with which children learning English and Italian respectively omit subjects is not the same. Subjects are omitted more often than objects. She therefore concludes that competence-based explanations are not empirically adequate. The omission of subjects should then be viewed more as a performance phenomenon (also, see Bloom 1993). Another argument for the influence of performance factors is the fact that articles of subjects are left out more often than articles of objects (31% vs. 18%, see Gerken 1991: 440). As Bloom notes, no subject article-drop parameter has been proposed so far. If we explain this phenomenon as a performance phenomenon, then it is also plausible to assume that the omission of complete subjects is due to performance issues.

Gerken (1991) shows that the metrical properties of utterances also play a role: in experiments where children had to repeat sentences, they omitted the subject/article of the subject more often than the object/article of the object. Here, it made a difference whether the intonation pattern was iambic (weak-strong) or trochaic (strong-weak). It can even be observed with individual words that children leave out weak syllables at the beginning of words more often than at the end of the word. Thus, it is more probable that “giRAFFE” is reduced to “RAFFE” than “MONkey” to “MON”. Gerken assumes the following for the metrical structure of utterances:

1. Every metrical foot contains exactly one strong syllable.

16 *Language acquisition*

2. Maximally binary feet are created from left to right.
3. Metrical structure is independent of syntactic structure.

Subject pronouns in English are sentence-initial and form a iambic foot with the following strongly emphasized verb as in (5a). Object pronouns, however, can form the weak syllable of a trochaic foot as in (5b).

- (5) a. she KISSED + the DOG
- b. the DOG + KISSED her
- c. PETE + KISSED the + DOG

Furthermore, articles in iambic feet as in the object of (5a) and the subject of (5b) are omitted more often than in trochaic feet such as with the object of (5c).

It follows from this that there are multiple factors that influence the omission of elements and that one cannot simply take the behavior of children as evidence for switching between two grammars.

Apart from what has been discussed so far, the pro-drop parameter is of interest for another reason: there is a problem when it comes to setting parameters. The standard explanation is that learners identify that a subject must occur in all English sentences, which is suggested by the appearance of expletive pronouns in the input.

As discussed on page 519, there is no relation between the pro-drop property and the presence of expletives in a language. Since the pro-drop property does not correlate with any of the other putative properties either, only the existence of subject-less sentences in the input constitutes decisive evidence for setting a parameter. The problem is that there are grammatical utterances where there is no visible subject. Examples of this are imperatives such as (6), declaratives with a dropped subject as in (7a) and even declarative sentences without an expletive such as the example in (7b) found by Valian (1991: 32) in the New York Times.

- (6) a. Give me the teddy bear!
- b. Show me your toy!
- (7) a. She'll be a big hit. Sings like a dream.
- b. Seems like she always has something twin-related perking.

The following title of a Nirvana song also comes from the same year as Valian's article:

- (8) Smells like Teen Spirit.

Teen Spirit refers to a deodorant and *smell* is a verb that, both in German and English, requires a referential subject but can also be used with an expletive *it* as subject. The usage that Kurt Cobain had in mind cannot be reconstructed<sup>4</sup>, independent of the intended

<sup>4</sup> See [http://de.wikipedia.org/wiki/Smells\\_Like\\_Teen\\_Spirit](http://de.wikipedia.org/wiki/Smells_Like_Teen_Spirit). 18.04.2010.

## 16.1 Principles and Parameters

meaning, however, the subject in (8) is missing. Imperatives do occur in the input children have and are therefore relevant for acquisition. Valian (1991: 33) says the following about them:

What is acceptable in the adult community forms part of the child's input, and is also part of what children must master. The utterances that I have termed "acceptable" are not grammatical in English (since English does not have pro subjects, and also cannot be characterized as a simple VP). They lack subjects and therefore violate the extended projection principle (Chomsky 1981a), which we are assuming. Children are exposed to fully grammatical utterances without subjects, in the form of imperatives. They are also exposed to acceptable utterances which are not fully grammatical, such as [(7a)], as well as forms like, "Want lunch now?" The American child must grow into an adult who not only knows that overt subjects are grammatically required, but also knows when subjects can acceptably be omitted. The child must not only acquire the correct grammar, but also master the discourse conditions that allow relaxation of the grammar. (Valian 1991: 33)

This passage turns the relations on their head: we cannot conclude from the fact that a particular grammatical theory is not compatible with certain data, that these data should not be described by this theory, instead we should modify the incompatible grammar or, if this is not possible, we should reject it. Since utterances with imperatives are entirely regular, there is no reason to categorize them as utterances that do not follow grammatical rules. The quotation above represents a situation where a learner has to acquire two grammars: one that corresponds to the innate grammar and a second that partially suppresses the rules of innate grammar and also adds some additional rules.

The question we can pose at this point is: how does a child distinguish which of the data it hears are relevant for which of the two grammars?

Fodor (1998a: 347) pursues a different analysis that does not suffer from many of the aforementioned problems. Rather than assuming that learners try to find a correct grammar among a billion others, she instead assumes that children work with a single grammar that contains all possibilities. She suggests using parts of trees (*treelets*) rather than parameters. These treelets can also be underspecified and in extreme cases, a treelet can consist of a single feature (Fodor 1998b: 6). A language learner can deduce whether a language has a given property from the usage of a particular treelet. As an example, she provides a VP treelet consisting of a verb and a prepositional phrase. This treelet must be used for the analysis of the VP occurring in *Look at the frog*. Similarly, the analysis of an interrogative clause with a fronted *who* would make use of a treelet with a *wh*-NP in the specifier of a complementizer phrase (see Figure 3.7 on page 102). In Fodor's version of Principles and Parameters Theory, this treelet would be the parameter that licenses *wh*-movement in (overt) syntax. Fodor assumes that there are defaults that allow a learner to parse a sentence even when no or very few parameters have been set. This allows one to learn from utterances that one would have not otherwise been able to use since there would have been multiple possible analyses for them. Assuming a default can lead to misanalyses, however: due to a default value, a second parameter could be set because



16 *Language acquisition*

an utterance was analyzed with a treelet  $t_1$  and  $t_3$ , for example, but  $t_1$  was not suited to the particular language in question and the utterance should have instead been analyzed with the non-default treelet  $t_2$  and the treelet  $t_{17}$ . In this acquisition model, there must therefore be the possibility to correct wrong decisions in the parameter setting process. Fodor therefore assumes that there is a frequency-based degree of activation for parameters (p. 365): treelets that are often used in analyses have a high degree of activation, whereas those used less often have a lower degree of activation. In this way, it is not necessary to assume a particular parameter value while excluding others.

Furthermore, Fodor proposes that parameters should be structured hierarchically, that is, only if a parameter has a particular value does it then make sense to think about specific other parameter values.

Fodor's analysis is – as she herself notes (Fodor 2001: 385) – compatible with theories such as HPSG and TAG. Pollard & Sag (1987: 147) characterize UG as the conjunction of all universally applicable principles:

$$(9) \quad \text{UG} = P_1 \wedge P_2 \wedge \dots \wedge P_n$$

As well as principles that hold universally, there are other principles that are specific to a particular language or a class of languages. Pollard and Sag give the example of the constituent ordering principle that only holds for English. English can be characterized as follows if one assumes that  $P_{n+1}$ – $P_m$  are language-specific principles,  $L_1$ – $L_p$  a complete list of lexical entries and  $R_1$ – $R_q$  a list of dominance schemata relevant for English.

$$(10) \quad \text{English} = P_1 \wedge P_2 \wedge \dots \wedge P_m \wedge (L_1 \vee \dots \vee L_p \vee R_1 \vee \dots \vee R_q)$$

In Pollard and Sag's conception, only those properties of language that equally hold for all languages are part of UG. Pollard and Sag do not count the dominance schemata as part of this. However, one can indeed also describe UG as follows:

$$(11) \quad \text{UG} = P_1 \wedge P_2 \wedge \dots \wedge P_n \wedge (R_{\text{en-1}} \vee \dots \vee R_{\text{en-q}} \vee R_{\text{de-1}} \vee \dots \vee R_{\text{de-r}} \vee \dots)$$

$P_1$ – $P_n$  are, as before, universally applicable principles and  $R_{\text{en-1}}$ – $R_{\text{en-q}}$  are the (core) dominance schemata of English and  $R_{\text{de-1}}$ – $R_{\text{de-r}}$  are the dominance schemata in German. The dominance schemata in (11) are combined by means of disjunctions, that is, not every disjunct needs to have a realization in a specific language. Principles can make reference to particular properties of lexical entries and rule out certain phrasal configurations. If a language only contains heads that are marked for final-position in the lexicon, then grammatical rules that require a head in initial position as their daughter can never be combined with these heads or their projections. Furthermore, theories with a type system are compatible with Fodor's approach to language acquisition because constraints can easily be underspecified. As such, constraints in UG do not have to make reference to all properties of grammatical rules: principles can refer to feature values, the language-specific values themselves do not have to already be contained in UG. Similarly, a supertype describing multiple dominance schemata that have similar but language-specific instantiations can also be part of UG, however the language-specific details remain open and are then deduced by the learner upon parsing (see Ackerman & Webelhuth



1998: Section 9.2). The differences in activation assumed by Fodor can be captured by weighting the constraints: the dominance schemata  $R_{en-1} - R_{en-q}$  etc. are sets of feature-value pairs as well as path equations. As explained in Chapter 15, weights can be added to such constraints and also to sets of constraints. In Fodor's acquisition model, given a German input, the weights for the rules of English would be reduced and those for the German rules would be increased. Note that in Pollard & Sag's acquisition scenario, there are no triggers for parameter setting unlike in Fodor's model. Furthermore, properties that were previously disjunctively specified as part of UG will now be acquired directly. Using the treelet  $t_{17}$  (or rather a possibly underspecified dominance schema), it is not the case that the value '+' is set for a parameter  $P_5$  but rather the activation potential of  $t_{17}$  is increased such that  $t_{17}$  will be prioritized for future analyses.

## 16.2 Principles and the lexicon

A variant of the UG-driven theory of language acquisition would be to assume that principles are so general that they hold for all languages and individual languages simply differ with regard to their lexicon. Principles then refer to properties of combined entities. Parameters therefore migrate from principles into the lexicon (Chomsky 1999: 2). See Mensching & Remberger (2011) for a study of Romance languages in this model and Son & Svenonius (2008: 395) for an analysis of Snyder's examples that were discussed in the previous subsection.

At this point, one can observe an interesting convergence in these approaches: most of the theories discussed here assume a very general structure for the combination of heads with their arguments. For example, in Categorical Grammar and the Minimalist Program, these are always binary functor-argument combinations. The way in which constituents can be ordered in a particular language depends on the lexical properties of the combined elements.

The question that is being discussed controversially at present is whether the spectrum of lexical properties is determined by UG (Chomsky 2007: 6–7) and whether all areas of the language can be described with the same general combinatorial possibilities (see Section 21.10 on phrasal constructions).

In Section 16.1, I have shown what theories of acquisition assuming innate language specific knowledge can look like and also that variants of such acquisition theories are compatible with all the theories of grammar we have discussed. During this discussion, one should bear in mind the question of whether it makes sense at all to assume that English children use parts of a Chinese grammar during some stages of their acquisition process (as suggested by Yang (2004: 453)), or whether the relevant phenomena can be explained in different ways. In the following, I will present some alternative approaches that do not presuppose innate language specific knowledge, but instead assume that language can simply be acquired from the input. The following section will deal with pattern-based approaches and Section 16.4 will discuss the lexically-oriented variant of input-based language acquisition.

### 16.3 Pattern-based approaches

Chomsky (1981a: 7–8) proposed that languages can be divided into a core area and a periphery. The core contains all regular aspects of language. The core grammar of a language is seen as an instantiation of UG. Idioms and other irregular parts of language are then part of the periphery. Critics of the Principles & Parameters model have pointed out that idiomatic and irregular constructions constitute a relatively large part of our language and that the distinction, both fluid and somewhat arbitrary, is only motivated theory-internally (Jackendoff 1997: Chapter 7; Culicover 1999; Ginzburg & Sag 2000: 5; Newmeyer 2005: 48; Kuhn 2007: 619). For example, it is possible to note that there are interactions between various idioms and syntax (Nunberg, Sag & Wasow 1994). Most idioms in German with a verbal component allow the verb to be moved to initial position (12b), some allow that parts of idioms can be fronted (12c) and some can undergo passivization (12d).

- (12) a. dass er ihm den Garaus macht  
that he him the GARAUS makes  
'that he finishes him off (kills him)'
- b. Er macht ihm den Garaus.  
he makes him the GARAUS  
'He finishes him off.'
- c. In Amerika sagte man der Kamera nach, die größte Kleinbildkamera der Welt zu sein. Sie war laut Schleiffer am Ende der Sargnagel der Mühlheimer Kameraproduktion.  
*Den Garaus machte* ihr die Diskussion um die Standardisierung des  
the GARAUS made her the discussion around the standardization of the  
16-Millimeter-Filmformats, an dessen Ende die DIN-Norm 19022 (Patrone  
16-millimeter-film.format at whose end the DIN-norm 19022 cartridge  
mit Spule für 16-Millimeter-Film) stand, die im März 1963 zur Norm  
with coil for 16-millimeter-film stood that in March 1963 to the norm  
wurde.<sup>5</sup>  
became  
'In America, one says that this camera was the biggest compact camera in the world. According to Schleiffer, it was the last nail in the coffin for camera production in Mühlheim. What finished it off was the discussion about standardizing the 16 millimeter format, which resulted in the DIN-Norm 19022 (cartridge with coil for 16 millimeter film) that became the norm in March 1963.'
- d. in Heidelberg wird „parasitären Elementen“ unter den Professoren *der*  
in Heidelberg are parasitic elements among the professors the

---

<sup>5</sup> Frankfurter Rundschau, 28.06.1997, p. 2.

## 16.3 Pattern-based approaches

*Garaus gemacht*<sup>6</sup>

GARAUS made

‘In Heidelberg, “parasitic elements” among professors are being killed off.’

It is assumed that the periphery and lexicon are not components of UG (Chomsky 1986b: 150–151; Fodor 1998a: 343) but rather are acquired using other learning methods – namely inductively directly from the input. The question posed by critics is now why these methods should not work for regular aspects of the language as well (Abney 1996: 20; Goldberg 2003a: 222; Newmeyer 2005: 100; Tomasello: 2006c: 36; 2006b: 20): The areas of the so-called ‘core’ are by definition more regular than components of the periphery, which is why they should be easier to learn.

Tomasello (2000, 2003) has pointed out that a Principles & Parameters model of language acquisition is not compatible with the observable facts. The Principles and Parameters Theory predicts that children should no longer make mistakes in a particular area of grammar once they have set a particular parameter correctly (see Chomsky 1986b: 146, Radford 1990: 21–22 and Lightfoot 1997: 175). Furthermore, it is assumed that a parameter is responsible for very different areas of grammar (see the discussion of the pro-drop parameter in Section 16.1). When a parameter value is set, then there should be sudden developments with regard to a number of phenomena (Lightfoot 1997: 174). This is, however, not the case. Instead, children acquire language from utterances in their input and begin to generalize from a certain age. Depending on the input, they can re-order certain auxiliaries and not others, although movement of auxiliaries is obligatory in English.<sup>7</sup> One argument put forward against these kinds of input-based theories is that children produce utterances that cannot be observed to a significant frequency in the input. One much discussed phenomenon of this kind are so called *root infinitives* (RI) or *optional infinitives* (OI) (Wexler 1998). These are infinitive forms that can be used in non-embedded clauses (*root sentences*) instead of a finite verb. Optional infinitives are those where children use both a finite (13a) and non-finite (13b) form (Wexler 1998: 59):

- (13) a. Mary likes ice cream.  
b. Mary like ice cream.

Wijnen, Kempen & Gillis (2001: 656) showed that Dutch children use the order object infinitive 90 % of the time during the two-word phase although these orders occur in less than 10 % of their mother’s utterances that contained a verb. Compound verb forms, e. g. with a modal in initial position as in (14) that contain another instance of this pattern only occurred in 30 % of the input containing a verb (Wijnen, Kempen & Gillis 2001: 647).

- (14) Willst du Brei essen?  
want you porridge eat  
‘Do you want to eat porridge?’

<sup>6</sup> Mannheimer Morgen, 28.06.1999, Sport; Schrauben allein genügen nicht.

<sup>7</sup> Here, Yang’s suggestion to combine grammars with a particular probability does not help since one would have to assume that the child uses different grammars for different auxiliaries, which is highly unlikely.

16 *Language acquisition*

At first glance, there seems to be a discrepancy between the input and the child's utterances. However, this deviation could also be explained by an utterance-final bias in learning (Wijnen et al. 2001; Freudenthal, Pine & Gobet 2006). A number of factors can be made responsible for the salience of verbs at the end of an utterance: 1) restrictions of the infant brain. It has been shown that humans (both children and adults) forget words during the course of an utterance, that is, the activation potential decreases. Since the cognitive capabilities of small children are restricted, it is clear why elements at the end of an utterance have an important status. 2) Easier segmentation at the end of an utterance. At the end of an utterance, part of the segmentation problem for hearers disappears: the hearer first has to divide a sequence of phonemes into individual words before he can understand them and combine them to create larger syntactic entities. This segmentation is easier at the end of an utterance since the word boundary is already given by the end of the utterance. Furthermore according to Wijnen, Kempen & Gillis (2001: 637), utterance-final words have an above average length and do bear a pitch accent. This effect occurs more often in language directed at children.

Freudenthal, Pine, Aguado-Orea & Gobet (2007) have modeled language acquisition for English, German, Dutch and Spanish. The computer model could reproduce differences between these languages based on input. At first glance, it is surprising that there are even differences between German and Dutch and between English and Spanish with regard to the use of infinitives as German and Dutch have a very similar syntax (SOV+V2). Similarly, English and Spanish are both languages with SVO order. Nevertheless, children learning English make OI mistakes, whereas this is hardly ever the case for children learning Spanish.

Freudenthal, Pine, Aguado-Orea & Gobet (2007) trace the differences in error frequencies back to the distributional differences in each language: the authors note that 75 % of verb final utterances<sup>8</sup> in English consist of compound verbs (finite verb + dependent verb, e.g. *Can he go?*), whereas this is only the case 30 % of the time in Dutch.

German also differs from Dutch with regard to the number of utterance-final infinitives. Dutch has a progressive form that does not exist in Standard German:

- (15) Wat ben je aan het doen?  
 what are you on it do.INF  
 'What are you doing?'

Furthermore, verbs such as *zitten* 'to sit', *lopen* 'to run' and *staan* 'to stand' can be used in conjunction with the infinitive to describe events happening in that moment:

- (16) Zit je te spelen?  
 sit you to play  
 'Are you sitting and playing?'

Furthermore, there is a future form in Dutch that is formed with *ga* 'go'. These factors contribute to the fact that Dutch has 20 % more utterance-final infinitives than German.

<sup>8</sup> For English, the authors only count utterances with a subject in third person singular since it is only in these cases that a morphological difference between the finite and infinitive form becomes clear.

## 16.3 Pattern-based approaches

Spanish differs from English in that it has object clitics:

- (17) (Yo) Lo quiero.  
       I    it   want  
       ‘I want it.’

Short pronouns such as *lo* in (17) are realized in front of the finite verb so that the verb appears in final position. In English, the object follows the verb, however. Furthermore, there are a greater number of compound verb forms in the English input (70 %) than in Spanish (25 %). This is due to the higher frequency of the progressive in English and the presence of *do*-support in question formation.

The relevant differences in the distribution of infinitives are captured correctly by the proposed acquisition model, whereas alternative approaches that assume that children possess an adult grammar but use infinitives instead of the finite forms cannot explain the gradual nature of this phenomenon.

Freudenthal, Pine & Gobet (2009) could even show that input-based learning is superior to other explanations for the distribution of NPs and infinitives. They can explain why this order is often used with a modal meaning (e. g. *to want*) in German and Dutch (Ingram & Thompson 1996). In these languages, infinitives occur with modal verbs in the corresponding interrogative clauses. Alternative approaches that assume that the linguistic structures in question correspond to those of adults and only differ from them in that a modal verb is not pronounced cannot explain why not all utterances of object and verb done by children learning German and Dutch do have a modal meaning. Furthermore, the main difference to English cannot be accounted for: in English, the number of modal meanings is considerably less. Input-based models predict this exactly since English can use the dummy verb *do* to form questions:

- (18) a. Did he help you?  
       b. Can he help you?

If larger entities are acquired from the end of an utterance, then there would be both a modal and non-modal context for *he help you*. Since German and Dutch normally do not use the auxiliary *tun* ‘do’, the relevant endings of utterances are always associated with modals contexts. One can thereby explain why infinitival expressions have a modal meaning significantly more often in German and Dutch than in English.

Following this discussion of the arguments against input-based theories of acquisition, I will turn to Tomasello’s pattern-based approach. According to Tomasello (2003: Section 4.2.1), a child hears a sentence such as (19) and realizes that particular slots can be filled freely (also see Dąbrowska (2001) for analogous suggestions in the framework of Cognitive Grammar).

- (19) a. Do you want more juice/milk?  
       b. Mommy is gone.

From these utterances, it is possible to derive so-called pivot schemata such as those in (20) into which words can then be inserted:

16 *Language acquisition*

- (20) a. more \_\_\_\_ → more juice/milk  
 b. \_\_\_\_ gone → mommy/juice gone

In this stage of development (22 months), children do not generalize using these schemata, these schemata are instead construction islands and do not yet have any syntax (Tomasello et al. 1997). The ability to use previously unknown verbs with a subject and an object in an SVO order is acquired slowly between the age of three and four (Tomasello 2003: 128–129). More abstract syntactic and semantic relations only emerge with time: when confronted with multiple instantiations of the transitive construction, the child is then able to generalize:

- (21) a. [<sub>S</sub> [<sub>NP</sub> The man/the woman] sees [<sub>NP</sub> the dog/the rabbit/it]].  
 b. [<sub>S</sub> [<sub>NP</sub> The man/the woman] likes [<sub>NP</sub> the dog/the rabbit/it]].  
 c. [<sub>S</sub> [<sub>NP</sub> The man/the woman] kicks [<sub>NP</sub> the dog/the rabbit/it]].

According to Tomasello (2003: 107), this abstraction takes the form [Sbj TrVerb Obj]. Tomasello's approach is immediately plausible since one can recognize how abstraction works: it is a generalization about reoccurring patterns. Each pattern is then assigned a semantic contribution. These generalizations can be captured in inheritance hierarchies (see page 208) (Croft 2001: 26). The problem with this kind of approach, however, is that it cannot explain the interaction between different areas of phenomena in the language: it is possible to represent simple patterns such as the use of transitive verbs in (21), but transitive verbs interact with other areas of the grammar such as negation. If one wishes to connect the construction one assumes for the negation of transitive verbs with the transitive construction, then one arrives at a problem since this is not possible in inheritance hierarchies.

- (22) The woman did not kick the dog.

The problem is that the transitive construction has a particular semantic contribution but that negated transitive construction has the opposite meaning. The values of SEM features would therefore be contradictory. There are technical tricks to avoid this problem, however, since there are a vast number of these kinds of interactions between syntax and semantics, this kind of technical solution will result in something highly implausible from a cognitive perspective (Müller 2006, 2007c,b, 2010b; Müller & Wechsler 2014a). For discussion of Croft's analysis, see Section 21.4.1.

At this point, proponents of pattern-based analyses might try and argue that these kinds of problems are only the result of a poor/inadequate formalization and would rather do without a formalization (Goldberg 2009: Section 5). However, this does not help here as the problem is not the formalization itself, rather the formalization allows one to see the problem more clearly.

An alternative to an approach built entirely on inheritance is a TAG-like approach that allows one to insert syntactic material into phrasal constructions. Such a proposal was discussed in Section 10.6.3. Bergen & Chang (2005: 170) working in Embodied Construction Grammar suggest an Active-Ditransitive Construction with the form

### 16.3 Pattern-based approaches

[RefExpr Verb RefExpr RefExpr], where RefExpr stands for a referential expression and the first RefExpr and the verb may be non-adjacent. In this way, it is possible to analyze (23a,b), while ruling out (23c):

- (23) a. Mary tossed me a drink.
- b. Mary happily tossed me a drink.
- c. \* Mary tossed happily me a drink.

While the compulsory adjacency of the verb and the object correctly predicts that (23c) is ruled out, the respective constraint also rules out coordinate structures such as (24):

- (24) Mary tossed me a juice and Peter a water.

Part of the meaning of this sentence corresponds to what the ditransitive construction contributes to *Mary tossed Peter a water*. There is, however, a gap between *tossed* and *Peter*. Similarly, one can create examples where there is a gap between both objects of a ditransitive construction:

- (25) He showed me and bought for Mary the book that was recommended in the Guardian last week.

In (25), *me* is not adjacent to *the book* .... It is not my aim here to request a coordination analysis. Coordination is a very complex phenomenon for which most theories do not have a straightforward analysis (see Section 21.6.2). Instead, I would simply like to point out that the fact that constructions can be realized discontinuously poses a problem for approaches that claim that language acquisition is exclusively pattern-based. The point is the following: in order to understand coordination data in a language, a speaker must learn that a verb which has its arguments somewhere in the sentence has a particular meaning together with these arguments. The actual pattern [Sbj V Obj1 Obj2] can, however, be interrupted in all positions. In addition to the coordination examples, there is also the possibility of moving elements out of the pattern either to the left or the right. In sum, we can say that language learners have to learn that there is a relation between functors and their arguments. This is all that is left of pattern-based approaches but this insight is also covered by the selection-based approaches that we will discuss in the following section.

A defender of pattern-based approaches could perhaps object that there is a relevant construction for (25) that combines all material. This means that one would have a construction with the form [Sbj V Obj1 Conj V PP Obj2]. It would then have to be determined experimentally or with corpus studies whether this actually makes sense. The generalization that linguists have found is that categories with the same syntactic properties can be coordinated (N,  $\bar{N}$ , NP, V,  $\bar{V}$ , VP, ...). For the coordination of verbs or verbal projections, it must hold that the coordinated phrases require the same arguments:

- (26) a. Er [arbeitet] und [liest viele Bücher].  
          he works   and   reads many books



16 *Language acquisition*

- b. Er [kennt und liebt] diese Schallplatte.  
he knows and loves this record
- c. Er [zeigt dem Jungen] und [gibt der Frau] die Punk-Rock-CD.  
he shows the boy and gives the woman the punk rock CD
- d. Er [liebt diese Schallplatte] und [schenkt ihr ein Buch].  
he loves this record and gives her a book

In an approach containing only patterns, one would have to assume an incredibly large number of constructions and so far we are only considering coordinations that consist of exactly two conjuncts. However, the phenomenon discussed above is not only restricted to coordination of two elements. If we do not wish to abandon the distinction between competence and performance (see Chapter 15), then the number of conjuncts is not constrained at all (by the competence grammar):

- (27) Er [kennt, liebt und verborgt] diese Schallplatte.  
he knows loves and hides this record

It is therefore extremely unlikely that learners have patterns for all possible cases in their input. It is much more likely that they draw the same kind of generalizations as linguists from the data occurring in their input: words and phrases with the same syntactic properties can be coordinated. If this turns out to be true, then all that is left for pattern-based approaches is the assumption of discontinuously realized constructions and thus a dependency between parts of constructions that states that they do not have to be immediately adjacent to one another. The acquisition problem is then the same as for selection-based approaches that will be the topic of the following section: what ultimately has to be learned are dependencies between elements or valences (see Behrens (2009: 439), the author reaches the same conclusion following different considerations).

## 16.4 Selection-based approaches

I will call the alternative to pattern-based approaches *selection-based*. A selection-based approach has been proposed by Green (2011).

The generalizations about the pattern in (21) pertain to the valence class of the verb. In Categorical Grammar, the pattern [Sbj TrVerb Obj] corresponds to the lexical entry (snp)/np (for the derivation of a sentence with this kind of lexical entry, see Figure 8.3 on page 245). A TAG tree for *likes* was given on page 414. Here, one can see quite clearly that lexical entries determine the structure of sentences in these models. Unlike pattern-based approaches, these analyses allow enough room for semantic embedding: the lexical entries in Categorical Grammar can be combined with adjuncts, and elementary trees in TAG also allow for adjunction to the relevant nodes.

Now, we face the question of how the jump from a pivot schema to a lexical entry with an argument structure takes place. In Tomasello's approach, there is no break between them. Pivot schemata are phrasal patterns and [Sbj TrVerb Obj] is also a phrasal pattern. Both schemata have open slots into which certain elements can be inserted. In selection-

based approaches, the situation is similar: the elements that are fixed in the pivot schema are functors in the selection-based approach. Green (2011) proposes a theory of acquisition in HPSG that can do without UG. For the two-word phase, she assumes that *where's* is the head of an utterance such as (28) and that *where's* selects *Robin* as its argument.

(28) Where's Robin?

This means that, rather than assuming that there is a phrasal pattern *Where's* X? with an empty slot X for a person or thing, she assumes that there is a lexical entry *where's*, which contains the information that it needs to be combined with another constituent. What needs to be acquired is the same in each case: there is particular material that has to be combined with other material in order to yield a complete utterance.

In her article, Green shows how long-distance dependencies and the position of English auxiliaries can be acquired in later stages of development. The acquisition of grammar proceeds in a monotone fashion, that is, knowledge is added – for example, knowledge about the fact that material can be realized outside of the local context – and previous knowledge does not have to be revised. In her model, mistakes in the acquisition process are in fact mistakes in the assignment of lexical entries to valence classes. These mistakes have to be correctable.

In sum, one can say that all of Tomasello's insights can be applied directly to selection-based approaches and the problems with pattern-based approaches do not surface with selection-based approaches. It is important to point out explicitly once again here that the selection-based approach discussed here also is a construction-based approach. Constructions are just lexical and not phrasal. The important point is that, in both approaches, words and also more complex phrases are pairs of form and meaning and can be acquired as such.

In Chapter 21, we will discuss pattern-based approaches further and we will also explore areas of the grammar where phrasal patterns should be assumed.

## 16.5 Summary

We should take from the preceding discussion that models of language acquisition that assume that a grammar is chosen from a large set of grammars by setting binary parameters are in fact inadequate. All theories that make reference to parameters have in common that they are purely hypothetical since there is no non-trivial set of parameters that all proponents of the model equally agree on. In fact there is not even a trivial one.

In a number of experiments, Tomasello and his colleagues have shown that, in its original form, the Principles & Parameters model makes incorrect predictions and that language acquisition is much more pattern-based than assumed by proponents of P&P analyses. Syntactic competence develops starting from verb islands. Depending on the frequency of the input, certain verbal constructions can be mastered even though the same construction has not yet been acquired with less frequent verbs.

The interaction with other areas of grammar still remains problematic for pattern-based approaches: in a number of publications, it has been shown that the interac-

## 16 *Language acquisition*

tion of phenomena that one can observe in complex utterances can in fact not be explained with phrasal patterns since embedding cannot be captured in an inheritance hierarchy. This problem is not shared by selection-based approaches. All experimental results and insights of Tomasello can, however, be successfully extended to selection-based approaches.

### Further reading

Meisel (1995) gives a very good overview of theories of acquisition in the Principles & Parameters model.

Adele Goldberg and Michael Tomasello are the most prominent proponents of Construction Grammar, a theory that explicitly tries to do without the assumption of innate linguistic knowledge. They published many papers and books about topics related to Construction Grammar and acquisition. The most important books probably are Goldberg (2006) and Tomasello (2003).

An overview of different theories of acquisition in German can be found in Klann-Delius (2008) an English overview is Ambridge & Lieven (2011).