well formed, whereas (1b) is not. To understand what precisely this difference is is to give 'a rational account of this behavior, i.e., a theory of the speaker's linguistic intuition . . . the goal of linguistic theory' (Chomsky 1955/1975: 95)—in other words, a psychology, and ultimately, biology of human language.

Once this position—lately dubbed the *biolinguistic* approach (Jenkins 1999, Chomsky 2000)—is accepted, it follows that language, just like all other biological objects, ought to be studied following the standard methodology in natural sciences (Chomsky 1975, 1980, 1986, 1995a). The postulation of innate linguistic knowledge, the Universal Grammar (UG), is a case in point.

One of the major motivations for innateness of linguistic knowledge comes from the Argument from the Poverty of Stimulus (APS) (Chomsky, 1980: 35). A well-known example concerns the *structure dependency* in language syntax and children's knowledge of it in the absence of learning experience (Chomsky 1975, Crain & Nakayama 1987). Forming an interrogative question in English involves inversion of the auxiliary verb and the subject:

(2) a. Is Alex *e* singing a song?
b. Has Robin *e* finished reading?

It is important to realize that exposure to such sentences underdetermines the correct operation for question formation. There are many possible hypotheses compatible with the language acquisition data in (2):

(3) a. front the first auxiliary verb in the sentence
b. front the auxiliary verb that most closely follows a noun
c. front the last auxiliary verb
d. front the auxiliary verb whose position in the sentence is a prime number
e. . . .

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 first noun phrase*, which can be arbitrarily long:

(6) $\mathcal{L}: (S_o, E) \rightarrow S_T$

A learning function or algorithm $\mathcal{L}$ maps the initial state of the learner, S_o, to the terminal state S_T, on the basis of experience E in the environment. Language acquisition research attempts to give an explicit account of this process.

1.2.1 Formal sufficiency

The acquisition model must be *causal* and *concrete*. Explanation of language acquisition is not complete with a mere description of child language, no matter how accurate or insightful, without an explicit account of the mechanism responsible for how language develops over time, the learning function $\mathcal{L}$. It is often claimed in the literature that children just 'pick up' their language, or that children's linguistic competence is identical to adults. Such statements, if devoid of a serious effort at some learning-theoretic account of *how* this is achieved, reveal irresponsibility rather than ignorance.

The model must also be *correct*. Given reasonable assumptions about the linguistic data, the duration of learning, the learner's cognitive and computational capacities, and so on, the model must be able to attain the *terminal* state of linguistic knowledge S_T comparable to that of a normal human learner. The correctness of the model must be confirmed by mathematical proof, computer simulation, or other forms of rigorous demonstration. This requirement has traditionally been referred to as the *learnability condition*, which unfortunately carries some misleading connotations. For example, the influential Gold (1967) paradigm of identification in the limit requires that the learner converge onto the 'target' grammar in the linguistic environment. However, this position has little empirical content.[3]

First, language acquisition is the process in which the learner forms an *internalized* knowledge (in his mind), an I-language

[3] I am indebted to Noam Chomsky for many discussions on the issue of learnability.

learner hears just before language acquisition stops happens to be noise, the learning experience during the entire period of language acquisition is wasted. This scenario is by no means an exaggeration when a realistic learning environment is taken into account. Actual linguistic environments are hardly uniform with respect to a single idealized grammar. For example, Weinreich et al. (1968: 101) observe that it is unrealistic to study language as a 'homogeneous object', and that the 'nativelike command of heterogeneous structures is not a matter of multidialectalism or "mere" performance, but is part of unilingual linguistic competence'. To take a concrete example, consider again the acquisition of subject use. English speakers, who in general use overt subjects, do occasionally omit them in informal speech, e.g. *Seems good to me*. This pattern, of course, is compatible with an optional subject grammar. Now recall that a triggering learner can alter its hypothesis on the basic of a *single* sentence. Consequently, variability in linguistic evidence, however sparse, may still lead a triggering learner to swing back and forth between grammars like a pendulum.

2.1.2 *Developmental incompatibility of the triggering model*

While it might be possible to salvage the triggering model to meet the formal sufficiency condition (e.g. via a random-walk algorithm of Niyogi & Berwick 1996; but cf. Sakas & Fodor 2001), the difficulty posed by the developmental compatibility condition is far more serious. In the triggering model, and in fact in *all* TL models, the learner at any one time is identified with a single grammar. If such models are at all relevant to the explanation of child language, the following predictions are inevitable:

(16) a. The learner's linguistic production ought to be consistent with respect to the grammar that is currently assumed.

b. As the learner moves from grammar to grammar, abrupt changes in linguistic expressions should be observed.

in child language as the learner switches from one grammar to another. However, Bloom (1993) found no sharp changes in the frequency of subject use throughout the NS stage of Adam and Eve, two American children studied by Brown (1973). Behrens (1993) reports similar findings in a large longitudinal study of German children's NS stage. Hence, there is no evidence for a radical reorganization—parameter resetting (Hyams & Wexler 1993)—of the learner's grammar. In section 4.1 we will show that for Dutch acquisition, the percentage of V2 use in matrix sentences also rises gradually, from about 50% at 2;4 to 85% at 3;0. Again, there is no indication of a radical change in the child's grammar, contrary to what the triggering model entails. Overall, the gradualness of language development is unexpected in the view of all-or-none parameter setting, and has been a major argument against the parameter-setting model of language acquisition (Valian 1990, 1991, Bloom 1990, 1993), forcing many researchers to the conclusion that child and adult language differ not in competence but in performance.

2.1.3 *Imperfection in child language?*

So the challenge remains: what explains the differences between child and adult languages? As summarized in Chapter 1 and repeated below, two approaches have been advanced to account for the differences between child and adult languages:

(17) a. Children and adults differ in linguistic performance.
b. Children and adults differ in grammatical competence.

The performance deficit approach (17a) is often stated under the Continuity Hypothesis (Macnamara 1982, Pinker 1984). It assumes an identity relation between child and adult competence, while attributing differences between child and adult linguistic forms to performance factors inherent in production, and (nonlinguistic) perceptual and cognitive capacities that are still underdeveloped at a young age (e.g. Pinker 1984, Bloom 1990, 1993, Gerken 1991, Valian 1991).

conditioned on external stimulus; in the grammar competition model, the hypothesis space consists of Universal Grammar, a highly constrained and finite range of possibilities. In addition, as discussed in Chapter 1, it seems unlikely that language acquisition can be equated to data-driven learning without prior knowledge. And, as will be discussed in later chapters in addition to numerous other studies in language acquisition, in order adequately to account for child language development, one needs to make reference to specific characterization of UG supplied by linguistic theories.

There is yet another reason for having an explicit account of the learning process: because language *is* acquired, and thus the composition, distribution, and other properties of the input evidence, in principle, matter. The landmark study of Newport et al. (1977) is best remembered for debunking the necessity of the so-called 'Motherese' for language acquisition, but it also shows that the development of *some* aspects of language does correlate with the abundance of linguistic data. Specifically, children who are exposed to more yes/no questions tend to use auxiliary verbs faster and better. An explicit model of learning that incorporates the role of input evidence may tell us why such correlations exist in some cases, but not others (e.g. the null subject phenomenon). The reason, as we shall see, lies in the Universal Grammar.

Hence, our emphasis on $\mathcal{L}$ is simply a plea to pay attention to the actual mechanism of language development, and a concrete proposal of what it might be.

2.3 The dynamics of variational learning

We now turn to the computational properties of the variational model in (22).

2.3.1 Asymptotic behaviors

In any competition process, some measure of fitness is required. Adapting the formulation of Bush & Mosteller (1958), we may offer the following definition:

details. We now turn to the more difficult issue of learning parameters that are subject to the interference problem.

Fitness distribution

In what follows, we will suggest that (some variant) of the NPL may be a plausible model of learning that distangles the interference effects from parameter interaction.

First, our conclusion is based on results from computer simulation. This is not the preferred move, for the obvious reason that one cannot simulate all possibilities that may arise in parameter learning. Analytical results—proofs—are much better, but so far they have been elusive.

Second, as far as feasible, we will study the behavior of the model in an *actual* learning environment. As the example of the *Wh* and V2 learning (Fig. 2.2) shows, the relative fitness values of the four composite grammars will determine the outcome of parameter learning. In that example, if the three competitors have high penalty probabilities, intuition tells us that the two parameters rise to target values quickly.[20] So the actual behavior of the model can be understood only if we have a good handle on the fitness distribution of actual grammars.

This is a departure from the traditional linguistic learnability study, and we believe it is a necessary one. Learnability models, in general, do not consider convergence in relation to the actual (statistical) distribution of the learning data.[21] Rather, learning is studied 'in the limit' (Gold 1967), with the assumption that learning can take an arbitrary amount of data as long as it converges on the correct grammar in the end: hence, no sample complexity considerations. However, it is clear that learning data is not infinite. In Chapter 4 we show that it is possible to establish bounds on the amount of linguistic data needed for actual acquisition: if

[20] Although intuition fades rapidly as more and more parameters combine and interact.

[21] A notable exception is Berwick & Niyogi's (1996) elegant Markov model of triggering, where the expected amount of evidence required for convergence can be precisely worked out.

the distributions of w_is across languages, no matter how these languages put them together. It does not seem unreasonable to assume, say, that the frequencies of transitive verbs are more or less uniform across languages, because transitive verbs are used in certain life contexts, which perhaps do not vary greatly across languages. Practically, such assumptions are necessary if there is any hope of estimating the distribution of sentences in many grammars, without reliable parsers or comprehensive corpora. Furthermore, some grammars, i.e. parameter settings, may not be attested in the world.

Given these assumptions, let us see how we may estimate the string distributions for eight grammars in Table 2.2, extrapolating from the grammars for which we do have some statistical results. For the English grammar (SVO–V2), we estimate, using sources like the CHILDES corpus, that about 10% of declarative sentences have an sentence-initial XP; thus 90% of the probability mass will be distributed among SV, SVO, SAV, SAVO. Roughly 50% of all sentences contain an auxiliary, and 50% of verbs are transitives. Assuming that the selection of Auxiliary and Verb is independent, and that the selection of the XP adjunct is independent of the rest of the sentence. We then obtain:

(36) a. P(SV) = P(SVO) = P(SAV) = P(SAVO) = 9/40
b. P(XSV) = P(XSVO) = P(SAV) = P(XSAVO) = 1/40

(36) will be carried over to the other three non-V2 grammars, and assigned to their respective canonical word orders.

For the four V2 grammars, we assume that (36) will carry over to the canonical patterns due to the Spec-Head and Comp-Head parameters. In addition, we must consider the effect of V2: raising S, O, or X to the sentence-initial position. It is known from (Lightfoot 1997: 265) as well as from our own analysis of a Dutch adult-to-child corpus, that in V2 languages, S occupies the initial position 70% of time, X, 28%, and O, 2%. These probability masses (0.7, 0.28, and 0.02) will be distributed among the canonical patterns.

Berko's (1958) classic work that in general, children (and adults) inflect novel verbs with the -*d* suffix, as in *rick-ricked*. Second, young children sometimes *overregularize*: for example, they produce *take-taked* instead of *take-took*, where the suffix -*d* for regular verbs is used for an irregular verb. On average, overregularization occurs in about 10% of all instances of irregular verbs, according to the most extensive study of past tense acquisition (Marcus et al. 1992). Third, errors such as *bring-brang* and *wipe-wope*, mis-irregularization errors where children misapply and overapply irregular past tense forms, are exceeding rare, accounting for about 0.2% of all instances of irregular verb uses (Xu & Pinker 1995).

One leading approach to the problem of past tense, following the influential work of Rumelhart and McClelland (1986), claims that the systematic patterns noted above emerge from the statistical properties of the input data presented to connectionist networks. A number of problems with the connectionist approach have been identified (e.g. Fodor & Pylyshyn 1988, Lachter & Bever 1988, Pinker & Prince 1988, Marcus et al. 1992). To give just one example (from Prasada & Pinker 1993), connectionist models have difficulty with the *Wug*-test, the hallmark of past tense knowledge. When novel verbs such as *slace* and *smeeb* are presented to a trained connectionist model, *fraced* and *imin* are produced as their respective past tense forms, a behavior hopelessly incompatible with human performance.

In this chapter, we will critically assess another leading approach to the problem of past tense, the Words and Rule (WR) model developed by Pinker and his associates (Pinker 1995, 1999). The WR model claims that the computational system for past tense consists of two components. In the 'rule' component, following the tradition of generative linguistics, regular verbs are inflected by making use of a default phonological rule, which adds -*d* to the root (stem). This explains the productivity of -*d* suffixation to novel verbs. Equally important to the WR model is the Blocking Principle, a traditional idea dating back to Pāṇini. In past tense formation, the Blocking Principle has the effect of forcing the use of a more specific form over a more general form: for example, *sang* is a more specific realization of the past tense of *sing* than *singed*, and is therefore used. Irregular verbs are learned in the 'word' component, which works like a connectionist network, by direct association/memorization of the pairing between a stem and its past tense. The strength of association is conditioned upon the frequencies of irregular verbs that children hear; thus, memorization of irregular verbs takes time and experience to be perfected. When the child's memory for an irregular form fails, the default *-d* form is used. This accounts for the second salient pattern of past tense acquisition: overregularization errors in child language.

Here we will put forward an alternative approach, the Rules and Competition (RC) model. The RC model treats both irregular and regular verbs within a single component of the cognitive system: generative phonology. Like the WR model, we assume the presence of a default rule, which attaches the *-d* suffix to the stem and in principle applies to all verbs. In contrast to the WR model, we claim that irregular past tense is also formed by phonological rules. That is, errors such as overregularization are not memory lapses, but result from failures to apply appropriate *irregular* phonological rules over the default rule.

The RC model derives from the variational approach to language acquisition, which holds that systematic errors in child language are reflections of coexisting hypotheses in competition. These hypotheses are associated with weights, and it is the weights, or the distribution of the grammars, that change during learning from data. For the problem of past tense, the hypothesis space for each irregular verb x includes an irregular rule R, defined over a verb class S of verbs of which x is a member. For example, the rule [*-t* suffixation & Vowel Shortening] applies to irregular verbs such as *lose*, *deal*, and *dream*. The acquisition of x involves a process of competition between R and the default *-d* rule, the latter of which in principle could apply to all verbs, regular and irregular. The child learns from experience that for irregular verbs, irregular rules must apply, and thus the default *-d* rule rather, in (44b), R applies with the probability $P_{\mathcal{R}}$, its weight. Only when both decisions are taken correctly will the correct past tense be produced—a match with the input X_{past}. When either of the two steps fails, the overregularized form will be produced, resulting in a mismatch with the input form, X_{past}.

Thus, for each verb, learning involves updating the two probabilities $P(x \in \mathcal{S})$ and $P_{\mathcal{R}}$. Learning is successful when $\forall x$, $P(x \in \mathcal{S})P_{\mathcal{R}} = 1$: the learner can reliably associate an irregular verb with its matching irregular rule, and reliably apply the rule over the default *-d* rule. As remarked in section 2.3, many models for updating probabilities (weights) are in principle applicable. For our purpose, let us assume a learner who increases the probabilities of the decisions he has made when they lead to a match between the input form and the analyzed form.

Under the null hypothesis, we assume that the grammar system the child uses for production is the same one he uses for comprehension/learning, the two-step procedures in (44). As a result, overregularization of an irregular verb x occurs when either $P(x \in \mathcal{S}) < 1$ or $P_{\mathcal{R}} < 1$.

The RC model makes direct and quantitative predictions about the performance of both irregular verbs and irregular verb classes. Write $C(x)$ to denote the correct usage rate of an irregular verb x; clearly $C(x) = P(x \in \mathcal{S})P_{\mathcal{R}}$. While $P(x \in \mathcal{S})$ may increase when the past tense of x is encountered, $P_{\mathcal{R}}$ may increases whenever *any* member of $\mathcal{S}$ is encountered. These two probabilities, and hence the correct usage of an irregular verb x, is positively correlated with $f_x \times f_{\mathcal{S}}$. Hence, if we hold f_x or $f_{\mathcal{S}}$ constant, the RC model makes two directions about the performance of irregular verbs:

(45) a. For two verbs x_1 and x_2 *within* a verb class, $C(x_1) > C(x_2)$ if $f_{x1} > f_{x2}$.
b. For two verbs x_1 and x_2 such that $x_1 \in \mathcal{S}_1$, $x_2 \in \mathcal{S}_2$, and $f_{x1} = f_{x2}$, $C(x_1) > C(x_2)$ if $f_{S1} > f_{S2}$.

In section 3.3 we will systematically evaluate these predictions with children's production data, and demonstrate that irregular verbs are indeed organized into classes.

units. Furthermore each of the possible contents of each representation would be implemented once as a single hardware 'type'; particular words would be representation in separate 'token' units with pointers to the types it contains. Links between stems and pasts would be set up during learning between their representations at two levels: between the token representations of each pair member, and their type representations at the level of representation that is ordinarily accessed by morphology: syllables, onsets, rhymes, feet (specifically, the structures manipulated in reduplicative and templatic systems, as shown in the ongoing work of McCarthy and Prince and others). Ordinary correct retrieval results from successful traversal of token-token links; this would exhaust the process for pairs like *go-went* but would be reinforced by type-type links for members of consistent and high-frequencies families like *sing-sang*. On occasions where token-token links are noisy or inaccessible and retrieval fails, the type-type links would yield an output that has some probability of being correct, and some probability of being an analogical extension (e.g., *brang*). Because the representation of input and output are each highly structured, such extensions would nonetheless be precise and follow constrained patterns, e.g., preserving portions of the stem such as onsets while substituting the appropriate rhymes, and avoiding the chimeras and fuzzy approximations that we do not see among real irregulars but that pure feature-to-feature networks are prone to making. (Pinker & Prince 1994: 334)

It is difficult to evaluate statements like these. The token level association is clear enough: the strength of brute force linking between a stem and its past, hence the retrieval rate of the corresponding verb, can be measured by estimating the frequency of the verb's occurrences in past tense. However, it is not clear how the type-level linkings between phonological structures (syllables, onsets, etc.) are established. But far worse is the vagueness concerning how the two levels interact. For example, while the token-level frequency effect is an important factor in past tense acquisition,[10] it is not clear when the type-level analogy becomes the operative force. Such imprecise formulations are not amenable to analytical results such as (45).

However, we believe that the evidence presented here is strong enough to rule out *any* model that does not use (irregular) phonological rules to describe irregular verbs. The data clearly point to an organization of irregular verbs by rules and classes.

[10] In fact, all 10 pieces of evidence offered by Pinker (1995) in support of the WR model, which we shall review in section 3.5, are frequency based, although section 3.3 has shown that frequency affects performance in a fairly subtle way, unexpected in the WR model.

(when applicable) or the default *-d* rule, to generate the expected past tense form from the extracted root. Now if an overregularized form such as *goed* is repeated several times, the chance of a mismatch (i.e. the child generating *went*) is consequently enhanced—the probability of generating *went* at least once in several consecutive tries—much to children's annoyance, it appears.

3.5.7 *Adult overregularization*

Adult do occasionally overregularize. Pinker claims that the rarity entails that adult overregularization is the result of performance, not the result of a grammatical system. However, this is not the only interpretation of adult overregularization: rule-based grammatical system approaches account for the data equally well. Under the RC model, for an irregular verb (e.g. *smite-smote*) that appears very sparsely, the learner may not be sure which class it belongs to, i.e. the probability of class membership association is considerably below 1. Overregularization thus results, even if the weight of the irregular rule for its corresponding class is very close to 1.

Pinker also notes that since memory fades when people get older, more overregularization patterns have been observed during experiments with older people (Ullman et al. 1993). This interesting finding is consistent with every theory that treats the irregulars as different—cognitively, and ultimately neurologically—from the regulars: in the RC model, it is the class membership that is memorized.

3.5.8 *Indecisive verbs*

Adults are unsure about the past tense of certain verbs that they hear infrequently. *Dreamed* or *dreamt*? *Dived* or *dove*? *Leapt* or *leaped*? *Strided* or *strode*?[21]

[21] Some of those forms are doublets, so both forms are heard. As noted in section 3.2.4, they pose a problem for the Absolute Blocking Principle, which the WR model adopts.

but is also obviously compatible with the RC model, following the discussion in 3.5.8 and 3.5.9.

3.6 Conclusion

We have proposed a rule competition model for the acquisition of past tense in English. A list of irregular rules, defined over classes of irregular verbs, compete with the default *-d* rule for past tense inflection. Hence, the learning of an irregular verb is determined by the probability with which the verb is associated with the corresponding irregular rule, as well as the probability of the rule applying over the default *-d* rule. We have also given justifications for, and explored the consequences of, a stochastic and learning-theoretic version of the Blocking Principle.

The RC model is completely general, and applicable to the acquisition of phonology in other languages. Complemented by the Yip–Sussman model of rule learning, our model makes very precise predictions about verb learning: any two verbs can be directly compared (45), based on quantifiable frequency measures drawn from linguistic corpora. Such quantitative predictions are strongly confirmed by the acquisition data. We view the findings here as a strong challenge to any phonological theory that rejects rules.

Scrutiny over past tense 'errors' revealed much about the organization and learning of phonology. In Chapter 4, we turn to their syntactic counterparts.

Appendix B: The rule system for English past tense

This list is loosely based on Halle & Mohanan (1985: appendix) and Pinker & Prince (1988: appendix). Very rare verbs are not listed.

- [-ø & umlaut]
  fall 266/334
  hold 0/5
  come 109/174
- [-ø & Rime → u]
  blow 5/15, *grow* 4/12, *know* 17/23, *throw* 11/34, *draw* 2/12, *fly* 8/15
- [-*d* & Vowel Shortening]
  say 522/525

by 1.2% of the input, then its target value should be set relatively late; more specifically, as late as the consistent use of subjects in child English.

4.1.2 V1 in V2 learners

Consider then the acquisition of the V2 parameter in Dutch. As noted in (26), there appears to be no direct signature for the V2 parameter: the four competitor grammars together provide a complete covering of the V2 expressions. However, three competitors, namely, the English, Irish, and Hixkaryana type grammars, while compatible with SVO, XVSO, and OVS patterns respectively, nevertheless have very high penalty probabilities: 35.3%, 66%, and 98.7%, according to our corpus analysis. As a result, these grammars are eliminated quite early on; see Fig. 2.1.

A Hebrew grammar, or a similar Semitic grammar such as Arabic, fares considerably better in the competition. By the virtue of allowing SVO and XVSO alternations (Fassi-Fehri 1993, Shlonsky 1997), it is compatible with an overwhelming majority of V2 patterns (98.7% in all). However, it is not compatible with OVS sentences, which therefore are in effect unambiguous signatures for the target V2 parameter after the other three competitors have been eliminated very rapidly. The rarity of OVS sentences (1.3%) implies that the V2 grammar is a relatively late acquisition, with a Hebrew-type non-V2 grammar in coexistence with the target V2 grammar for an extended period of time.

A Hebrew type grammar, then, allows verb-initial (V1) sentences, which are ungrammatical for the target V2 grammar, but will nevertheless constitute a significant portion of Dutch child language, if the variational model is correct. This prediction is confirmed based on the statistics from a Dutch child, Hein

[4] As remarked earlier, Valian nevertheless claims that the subject parameter is set correctly, and attributes the missing subjects to performance limitations; we will return to this in section 4.3.2.

(Haegeman 1995), one of the largest longitudinal studies in the acquisition of V2 languages.[5] The data concern the position of the finite verb in matrix sentences, and are reported in Haegeman's tables 5 and 6, which we combine in Table 4.1.

Based on these, we can compute the ratio of V1 sentences over all sentences. The number of V1 sentences is the number of postverbal subject sentences minus those with overt material left of V; that is, column 3 minus column 2 in Table 4.1. The number of all sentences is the sum of column 2 and column 3 in Table 4.1. The results are shown in Table 4.2.

Some of the V1 patterns are given below (from Haegeman 1995: n. 21):

(60) a. Week ik neit.
know I not
b. Zie ik nog niet.
see I yet not
c. Schijnt de zon.
shines the sun
d. Kan ik niet lopen.
can I not run

Now we have to be sure the V1 patterns in (60) are 'real', i.e. are induced due to the presence of a competing Semitic-type grammar. First, it must be stressed that all the sentences contain overt subjects, hence ruling out the possibility that the superficial V1 patterns are due to subject drop, which Germanic children are known to use. Another compounding factor is the precise location of the (finite) verb. According to Shlonsky (1997), finite verbs in Hebrew move to a position above Tense, presumably an Agreement node. Thus, if the V1 patterns are genuinely Hebrew-like, the finite verb must reside in a position higher than Tense. The presence of an overt subject again confirms this. Stromswold & Zimmerman's (1999) large quantitative study shows, contrary to the earlier claims of Deprez & Pierce (1993), that the subject is consistently placed above Negation,

[5] I should point out that Haegeman's paper does not directly deal with the V2 phenomenon, but with the nature of Optional Infinitives instead; it happens to contain a large body of quantitative data needed by our study.

(Clahsen 1986). Under the present model, it is no coincidence that the timing of the acquisition of English subject use and that of Dutch/German V2 are comparable.

4.2 Quantifying the stimulus poverty argument

Based on the acquisition model and the findings in section 4.1, we can give a quantitative evaluation of the Argument from the Poverty of Stimulus (APS).

Recall from section 1.1 that at the heart of APS lies the question: why do human children unequivocally settle on the correct (structure-dependent) rules for question formation, while the input evidence does not rule out the incorrect, structure-independent, inductive generalization?

(62) a. Front the first auxiliary verb in the sentence.
b Front the auxiliary verb that is most closely follows a noun.
c. Front the last auxiliary verb in the sentence.
d. Front the auxiliary verb whose position in the sentence is a prime number.
e. . . .

for which the relevant evidence is in many ways ambiguous:

(63) a. Is Alex *e* singing a song?
b. Has Robin *e* finished reading?

Recently, the argument for innate knowledge based on structure dependency has been challenged by Sampson (1989), Pullum (1996), and Cowie (1998), among others. They claim that the learner is actually exposed to the relevant evidence to rule out the incorrect, structure-independent hypotheses. Here we will focus on Pullum's objections and show that they are not valid.

First, Pullum (implicitly) assumes that there is only *one* alternative hypothesis to be ruled out, namely, that of (62a), the inversion of the first auxiliary in the sentence. This assumption is incorrect: the learner in fact has to rule out *all*, in principle infinitely many, hypotheses compatible with (63); cf. Freidin (1989). But for the sake of argument, suppose it were the case that the

Let us now turn to Italian children. Recall that Chinese does not allow subject drop when an argument assumes the topic position (71b), and Italian does (with a fronted argument *Wh* phrase). This means that every subjectless question with an argument (object) *Wh* question punishes a Chinese grammar, and of course an English grammar as well.

It is known that approximately 70% of adult utterances have dropped subjects (Bates 1976, cited in Caselli et al. 1995). We also know that *Wh* questions are one of the most frequent constructions children are exposed to. We estimate that about 10% of all sentences are object questions involving empty subjects: again, the lower bound of 7% then warrants an early acquisition. This prediction is confirmed by Valian's findings (1991): at both of the developmental stages investigated (1;6–1;10 and 2;0–2;5), Italian children drop subjects in about 70% of sentences, roughly the same as the figures in adult speech reported in the references cited above.

4.3.2 *English children speak Chinese*

Finally, we consider how English children come to learn that their language uses an obligatory subject grammar, ruling out the Chinese and Italian grammars that are also made available by UG.

We first claim that the Italian grammar can very rapidly be eliminated by English children on the basis of their knowledge of agreement morphology. In Chapter 3 we reviewed the very strong evidence that young children's agreement morphology is near-perfect. Phillips (1995: 327), reviewing a number of crosslinguistic studies, observes that 'in languages with overt agreement morphology, children almost always use the agreement morphemes appropriate to the argument being agreed with'. Again, Guasti (1992) found that three young Italian children used agreement morphology correctly in more than 95% of all contexts; see e.g. Clahsen & Penke (1992) for similar findings in German, Torrens (1995) for Catalan, Levy & Vainikka (1999) for Hebrew.

and independently motivated model which details how such factors affect language acquisition. It is also common to find notions such as 'diachronic reanalysis', which claims that the learner under certain conditions will opt for a radical change in his grammar. Again, these claims can be substantiated only when supporting evidence is found in synchronic child language development.

This chapter extends the acquisition model to a study of language change that satisfies these requirements. It characterizes the dynamic interaction between the internal Universal Grammar and the external linguistic evidence, as mediated by language acquisition. We will again borrow insights from the study of biological evolution, where internal and external forces—genetic endowment and environmental conditions—interact in a similar fashion. Section 5.1 spells out the model and derives a number of formal properties, including a sufficient and necessary condition under which one grammar replaces another. In sections 5.2 and 5.3 we apply the model to explain the loss of V2 in Old French and the erosion of V2 in Old English.

5.1 Grammar competition and language change

5.1.1 The role of linguistic evidence

Given the dynamics of language change in Fig. 5.1, the fundamental question in language change is to identify the causal forces that

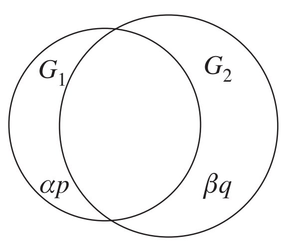

FIGURE 5.2. Two mutually incompatible grammars constitute a heterogeneous linguistic environment

languages,[2] it leaves mysterious the relative stability in other languages, say, the rigidity of word order in Western Germanic languages.

We therefore reject mislearning (under sufficiently similar linguistic evidence) as a possible mechanism of language change. A question immediately arises: what makes the linguistic evidence for generation $n + 1$ different from that of the previous generation? There are many possibilities. For example, migration of foreign speakers might introduce novel expressions; social and cultural factors might also influence the distributional patterns of linguistic expressions used in a population. These are interesting and important topics of research, but are not relevant for a formal model of language change. This situation has a perfect parallel in the mathematical theory of natural selection, which concerns the predictable changes in the population once some new genotypes are introduced. The precise manner in which new genes arise, which could be mutation, migration, etc., is a separate question, which is often affected by too many contingencies to command a firm answer. After all, the world would have looked very different if the comet that led to the demise of dinosaurs had been off target. Similarly, the factors that alter the composition of linguistic evidence from generation to generation may also be generally unpredictable: the linguistic landscape, and indeed the world, might have looked very different had Napoleon's winter in Russia been a lot warmer.

Hence, we are chiefly concerned with the predictable consequences of such changes: what happens to language learners *after* the linguistic evidence is altered, and how does it affect the composition of the linguistic population as a result?

5.1.2 A variational model of language change

Suppose that, as a result of migration, genuine innovation, and other sociological and historical factors, a linguistic environment is

[2] A desirable feature of a competence theory but by no means a necessary one: see Yang (1996) for discussion in relation to the issue of 'psychological reality'.

established for a generation of language learners that is substantially different from that for the previous generation.

The expressions used in such an environment—call it $E_{G1, G2}$—can formally be viewed as a mixture of expressions generated by two independent sources: the two grammars G_1 and G_2. Further, suppose that a proportion α of G_1 expressions are incompatible with G_2, and a proportion β of G_2 expressions are incompatible with G_1. Call α (β) the *advantage* of G_1 (G_2). Fig. 5.2 illustrates.

The variational approach views language acquisition as competition and selection among grammars. Recall from Chapter 2 that the fitness of individual grammars is defined in terms of their penalty probabilities:

(79) The penalty probability of a grammar G_i in a linguistic environment E is
$c_i = \Pr(G_i \not\rightarrow s \mid s \in E)$

The penalty probabilities ultimately determine the outcome of language acquisition:

(80) $$\lim_{t \to \infty} p_1(t) = \frac{c_2}{c_1 + c_2}$$

$$\lim_{t \to \infty} p_2(t) = \frac{c_1}{c_1 + c_2}$$

Suppose that at generation n, the linguistic environment $E_{G1, G2} = pG_1 + qG_2$, where $p + q = 1$. That is, in $E_{G1, G2}$, a proportion p of expressions are generated by G_1, and a proportion q of expressions are generated by G_2, and they collectively constitute the linguistic evidence to the learners in generation $n + 1$. The penalty probabilities of G_1 and G_2, c_1 and c_2, are thus βq and αp. The results in (80) allow us to compute p' and q', the weights of G_1 and G_2 respectively, that are internalized in the learners of generation $n + 1$:

(81) The dynamics of a two grammar system:

$$p' = \frac{\alpha p}{\alpha p + \beta p}$$

constraint is very strongly manifested. Matrix V > 2 patterns are restricted to a small number of adverbs and other specific lexical items, and are quite rare in distribution:

(89) Rare V > 2 patterns in modern German
. . . denn Johann hat gestern das Buch gelesen.
. . . so Johann had yesterday the book read.

Statistical analysis of Dutch, German, Norwegian, and Swedish (cited in Lightfoot 1997: 265) shows that about 70% of all sentences in V2 languages are SVO, and about 30% are VS patterns, which include XVSO and OVS. Our own counts, based on a Dutch sample of adult-to-child speech reported in section 4.1.2, are similar: 66.8% SVO, 23% XVSO, and 1.2% OVS. In contrast, based on the Penn Treebank (Marcus et al. 1993), a corpus of modern English, we found that only about 10% of all sentences have V > 2 word order:

(90) V > 2 patterns in modern English
a. He always reads newspapers in the morning.
b. Every evening Charles and Emma Darwin played backgammon.

Therefore, the 10% advantage of SVO grammar, expressed in V > 2 patterns, cannot throw off a V2 grammar, which has 30% of VS patterns to counter.

If the V2 constraint is so resilient, why did V2 succumb to SVO in French? The reason, in our view, is that OF was also a null subject language.

Recall that the advantage of V2 grammar over SVO grammar is expressed in VS patterns. However, this advantage would be considerably diminished if the subject were dropped to yield [X V *pro*] patterns: a null subject SVO grammar (like modern Italian) can analyze such patterns as [X (*pro*) V].(91) shows the prevalence of subject drop in early Middle French:

(91)

Text	*SV* %	*VS* %	*NullS* %	
Froissart, *Chroniques* (c.1390)	40	18	42	
15 Joyes (*14esme Joye*) (*c*.1400)	52.5	5	42.5	
Chartier, *Quadrilogue* (1422)	51	7	42	(R: 155)

TABLE 5.2. V2 (after languages contact) in the Northern MS (Thornton) of the Mirror of St Edmund

Preposed XP	NP subjects % inverted		Pronoun subjects % inverted	
NP complements	100	(8/8)	64	(16/25)
PP complements	88	(21/24)	70	(48/69)
Adj. complements	100	(10/10)	25	(2/8)
then	86	(6/7)	51	(24/47)
now	100	(4/4)	82	(14/17)
Adverbs	80	(20/25)	57	(35/61)

Source: Kroch et al. (2000)

The effect of language contact is clear. Recall that prior to contact the northern dialect was much like Germanic languages, in which V2 is strongly enforced: Kroch et al. (2000) found subject–verb inversion in 93.3% of all sentences containing subjects. After contact (shown in Table 5.2), while NP subjects still in general follow subjects, the overall subject–verb inversion rate has dropped to 68.2% (208/305). This indicates that as a result of language contact and mixing, the V2 constraint in the northern dialect was considerably weakened. When the V2 constraint is sufficiently weakened, and if the morphological case system of the mixed language got lost, then an SVO grammar would have gradually taken over, in the manner described earlier for the loss of V2 in OE.

For the northern dialect, the initial contact with the southern dialect was crucial in the loss of V2.[5] That is, a West Germanic V2 language similar to the northern dialect would not lose V2 without language contact, even if its morphological case system was lost. Northern Germanic languages such as Swedish, Danish, and Norwegian, with an impoverished morphological case system but nevertheless strongly V2, presumably fall into this category. Once language contact was made, the homogeneity of linguistic evidence was broken, and two distinct grammars were formed by the learners. The loss of the morphological case system resulted in the loss of the clitics system, which further favored the SVO grammar and eventually drove it to complete dominance. K&T's thesis yields new insights on traditional problems as well as suggesting new problems. The investigations reported here are no doubt preliminary; I only hope that they have convinced the reader that this line of research is worth pursuing.

A long way to the vineyard after all.

—— (1993). Grammatical Continuity in Language Development: The Case of Subjectless Sentences. *Linguistic Inquiry* 24: 721–34.

—— Barss, A., Nicol, J., & Conway, L. (1994). Children's Knowledge of Binding and Coreference: Evidence from Spontaneous Speech. *Language* 70: 53–71.

Bloomfield, L. (1927). Review of J. O. H. Jespersen, *The Philosophy of Grammar*. *Journal Of English and Germanic Philology* 26: 244–6.

Borer, H., & Wexler, K. (1987). The Maturation of Syntax. In T. Roeper & E. Williams (eds.), *Parameter Setting*. Dordrecht: Reidel, 123–72.

Boser, K. (1992). Early Knowledge of Verb Position in Children's Acquisition of German: An Argument for Continuity of Universal Grammar. Master's thesis, Cornell University.

de Boysson-Bardies, B. (1999). *How Language Comes to Children: From Birth to Two Years*. Cambridge, Mass.: MIT Press.

Bromberg, H., & Wexler, K. (1995). Null Subjects in Child Wh Questions. In C. Schütze, J. Ganger, & K. Broihier (eds.), *MIT Working Papers in Linguistics* 26: 221–47.

Brown, R. (1973). *A First Language*. Cambridge, Mass.: Harvard University Press.

—— & Hanlon, C. (1970). Derivational Complexity and Order of Acquisition in Child Speech. In J. R. Hayes (ed.), *Cognition and the Development of Language*. New York: Wiley, 155–207.

Burzio, L. (1991). The Morphological Basis of Anaphora. *Journal of Linguistics* 27: 81–105.

Bush, R., & Mosteller, F. (1951). A Mathematical Model for Simple Learning. *Psychological Review* 58: 313–23.

—— & —— (1958). *Stochastic Models for Learning*. New York: Wiley.

Bybee, J., & Moder, C. (1983). Morphological Classes as Natural Categories. *Language* 59: 251–70.

—— & Slobin, D. (1982). Rules and Schemas in the Development and Use of the English Past Tense. *Language* 58: 265–89.

Caselli, M., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A Cross-Linguistic Study of Early Lexical Development. *Cognitive Development* 10: 159–99.

Changeux, J.-P. (1986). *The Neuronal Man*. Oxford: Oxford University Press.

Charniak, E. (1996). *Statistical Natural Language Learning*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1955/75). The Logical Structure of Linguistic Theory. MS, Harvard University and Massachusetts Institute of Technology. Published in 1975, New York: Plenum.

—— (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

—— (1975). *Reflections on Language*. New York: Pantheon.

HYAMS, N. (1986). *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.

—— (1991). A Reanalysis of Null Subjects in Child Language. In J. Weissenborn, H. Goodluck, & T. Roeper (eds.), *Theoretical Issues in Language Acquisition: Continuity and Change in Development*. Hillsdale, NJ: Erlbaum, 249–67.

—— (1996). The Underspecification of Functional Categories in Early Grammar. In H. Clahsen (ed.), *Generative Perspectives on Language Acquisition*. Amsterdam: John Benjamins, 91–128.

—— & WEXLER, K. (1993). On the Grammatical Basis of Null Subjects in Child Language. *Linguistic Inquiry* 24: 421–59.

JAEGGLI, O., & SAFIR, K. (1989). Introduction. In O. Jaeggli & K. Safir (eds.), *The Null Subject Parameter*. Dordrecht: Kluwer, 1–44.

JAKOBSON, ROMAN. (1941/68). *Child Language, Aphasia and Phonological Universals*. The Hague: Mouton.

JENKINS, L. (1999). *Biolinguistics*. Cambridge: Cambridge University Press.

KAYNE, R. (1975). *French Syntax: The Transformational Cycle*. Cambridge, Mass.: MIT Press.

—— (1991). Romance Clitics, Verb Movement, and PRO. *Linguistic Inquiry* 22: 647–86.

—— (1994). *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.

—— (2000). Microparametric Syntax: Some Introductory Remarks. In R. Kayne, *Parameters and Universals*. Oxford: Oxford University Press, 3–9.

KEENAN, E. (2001). The Historical Creation of Reflexive Pronouns in English. MS, University of California at Los Angeles.

VAN KEMENADE, A. (1987). *Syntactic Case and Morphological Case in the History of English*. Dordrecht: Foris.

KIPARSKY, P. (1973). 'Elsewhere' in Phonology. In S. Anderson & P. Kiparsky (eds.), *A Festschrift for Morris Halle*. New York: Holt, Rinehart, & Winston, 93–106.

—— (1997). The Rise of Positional Licensing. In A. van Kemenade & N. Vincent (eds.), *Parameters of Morphosyntactic Change*. Cambridge: Cambridge University Press, 460–94.

KOHL, K. (1999). An Analysis of Finite Parameter Learning in Linguistic Spaces. Master's thesis. Massachusetts Institute of Technology.

KROCH, A. (1989). Reflexes of Grammar in Patterns of Language Change. *Language Variation and Change* 1: 199–244.

—— (2001). Syntactic Change. In M. Baltin & C. Collins (eds.), *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell, 699–729.

—— & TAYLOR, A. (1997). Verb Movement in Old and Middle English: Dialect

J. H. (1993). Neural Modularity in Language: Evidence from Alzheimer's and Parkinson's Disease. Abstract presented at the 23rd Annual Meeting of the Society for Neuroscience, Washington, DC.

Valian, V. (1990). Null Subjects: A Problem for Parameter-Setting Models of Language Acquisition. *Cognition* 35: 105–22.

—— (1991). Syntactic Subjects in the Early Speech of American and Italian Children. *Cognition* 40: 21–82.

Waller, B. (1997). Against a Metrical Basis for Subject Drop in Child Language. MS, Massachusetts Institute of Technology.

Wang, Q., Lillo-Martin, D., Best, C., & Levitt, A. (1992). Null Subject vs. Null Object: Some Evidence from the Acquisition of Chinese and English. *Language Acquisition* 2: 221–54.

Weinberg, A. (1990). Markedness vs. Maturation: The Case of Subject–Auxiliary Inversion. *Language Acquisition* 1: 169–94.

Weinreich, U., Labov, W., & Herzog, M. (1968). Empirical Foundations for a Theory of Language Change. In W. Lehman & Y. Malkiel (eds.), *Directions for Historical Linguistics: A Symposium*. Austin: University of Texas Press, 95–188.

Weverink, M. (1989). The Subject in Relation to Inflection in Child Language. MA thesis, University of Utrecht.

Wexler, K. (1994). Optional Infinitives, Head Movement, and the Economy of Derivation in Child Language. In D. Lightfoot & N. Hornstein (eds.), *Verb Movement*. Cambridge: Cambridge University Press, 305–50.

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

—— & Culicover, P. (1980). *Formal Principles of Language Acquisition*. Cambridge, Mass.: MIT Press.

Wijnen, F. (1999). Verb Placement in Dutch Child Language: A Longitudinal Analysis. MS, University of Utrecht.

Xu, F., & Pinker, S. (1995). Weird Past Tense Forms. *Journal of Child Language* 22: 531–56.

Yang, C. D. (1996). Psychological Reality, Type Transparency, and Minimalist Sentence Processing. MS, Massachusetts Institute of Technology.

—— (1999). Two Grammars Are Better than One. *Bilingualism: Language and Cognition* 2: 201–3.

—— (2000). Dig–Dug, Think–Thunk. Review of Steven Pinker's *Words and Rules*. *London Review Of Books* 22 (16).

—— (2002). Panda's Thumbs and Irregular Verbs. MS, Yale University.

—— (in press) Parametric Learning and Parsing. In R. Berwick (ed.), *Principle-Based Parsing II: From Theory to Practice*. Boston, Mass.: Kluwer.