Morphosyntactic Learning and the Development of Tense

Julie Anne Legate  Charles Yang
University of Delaware  Yale University
jlegate@udel.edu  charles.yang@yale.edu

June 30th, 2005

In this paper, we propose that the Root Infinitive (RI) phenomenon in child language is best viewed and explained as a morphological learning problem. We make the following specific suggestions:

• The optionality in RI reflects the presence of a grammar such as Chinese which does not manifest tense marking.

• The gradual elimination of the Chinese-style grammar is facilitated by the learning of the morphosyntactic system of the target language.

• Quantitative differences among morphosyntactic systems result in the cross-linguistic variation in the RI phenomenon.

More broadly, we aim to demonstrate that quantitative aspects of language learning data and concrete mechanisms of the language learning process can play an important role in the generative approach to language acquisition.

In section 1, we give a brief overview of the RI literature along with methodological remarks regarding the explanation of the phenomenon. In section 2, we lay out our approach to morphosyntactic learning, which builds on and complements the ideas from previous analyses. Section 3 presents a quantitative study of child-directed Spanish, English, and French. We show that quantitative differences in the morphosyntactic systems of these three languages explain the brief RI stage in Spanish acquisition, the prolonged RI stage in English acquisition, as well as the intermediate status of the RI stage in French. Section 4 is the conclusion.
1 Preliminary Remarks

1.1 The RI Phenomenon

One needs no reminder that the problem of Root Infinitive (also known as Optional Infinitive) occupies a central place in language acquisition research. Children acquiring a typologically diverse range of languages use non-finite verbs in root clauses, which is not generally grammatical in the corresponding adult languages (Weverink 1989, Platzck 1990, Wexler 1994, Haegeman 1995, Harris and Wexler 1996, Bar-Shalom and Snyder 1997, Schaeffer and Ben Shalom 2004 among others). Some examples are given in (1):

(1) a. Papa have it. (English)
   b. thee drinken. (Dutch)
      tea drink-INF
   c. Dormir petit bébé. (French)
      sleep-INF little baby
   d. mein Kakao hinstelln. (German)
      my chocolate milk put-INF
   e. Malon lauf. (Hebrew)
      balloon fly-INF.
      ’The balloon is flying”

While in some of these languages, adults do use RI patterns in certain limited contexts (Rizzi 1994, Wijnen, Kempen, & Gillis 2001, Lasser 2002), children’s use of RI is robust and found in wider contexts than adults, indicating that a considerably different grammatical system is at play. The RI phenomenon, therefore, adds to the already impressive range of findings that language acquisition is not simply a matter of children repeating what they hear from adults.

It is important to bear in mind that the RI phenomenon is gradient, in two senses. First, there is no evidence that the exit from the RI stage is sudden: rather, the frequency of RI usage drops gradually to adult level, sometimes over the span of 2-3 years or even longer (Behrens 1993, Wijnen and Bol 1993, Haegeman 1995, Phillips 1995). Second, and

---

1. We note, however, that if adult use of root infinitives is statistically significant, this may contribute to extended use of RI in children. Under our proposal elaborated in section 3 and 4, adult use of RI would be giving the child conflicting evidence about the obligatoriness of tense marking in her language. However, we will not pursue a quantitative analysis of the effect of adult RI in the present paper.
crosslinguistically, the distribution of the RI is not categorical but rather falls along a continuum. It appears that children learning languages with “richer” morphologies, as sometimes reflected in the possibility of pro-drop, have shorter RI stages and lower frequency of use (Phillips 1995, Wexler 1998, Guasti 2002). For example, the production of RI ranges from rare in the Italian acquisition (Guasti 1992), to a prolonged stage that may extend beyond the third birthday as in the case of Dutch (Haegeman 1995), German (Poeppel and Wexler 1993) and English (Phillips 1995, Harris and Wexler 1996).

1.2 Approaches to RI

Here we give a (very) brief summary of some leading proposals in the RI literature. It is impossible in the current context to give a comprehensive survey, and we will not dwell on the virtues and problems with these proposals; see Phillips (1995), Wexler (1998), and Guasti (2002) for reviews. As will become evident later, our own proposal builds on the insights of many of the existing accounts.

Some theories of RI locate children’s RI behavior in their core syntax, which is conjectured to be qualitatively different from that of an adult speaker of the target language. According to this view, an RI grammar may involve a tense node that is not interpreted at LF (Wexler 1994), incomplete clausal structures (Rizzi 1994; cf. Radford 1990), or different checking options for the D-feature on nominals (Schütze & Wexler 1996, Wexler 1998). Another class of theories relates RI to a combination of syntactic and non-syntactic factors. For example, it is hypothesized that RI is due to underspecified features in the Tense node, which in turn are connected to the development of pragmatic knowledge (Hoekstra & Hyams 1998), or to a phonologically null auxiliary which marks tense and represents related semantic properties (Boser et al. 1992, Whitman 1994). Still another group of theories links RI to non-syntactic performance limitations. For instance, Avrutin (1999) suggests that RI results from children’s limited resources to carry out utterance planning and other pragmatic computation, while Phillips (1995) attributes RI to the failure of merging the Tense node with the verb, resulting from not-yet-automatic use of morphological knowledge.

All these accounts have led to important empirical findings in the RI phenomenon, and have advanced our understanding of language acquisition and Universal Grammar in general. However, we believe that a number of crucial problems remain, some of which have not been frequently discussed.

In our opinion, a theory of RI, in addition to providing an insightful description of the
RI stage, must also provide a specific model of how the child exits the RI stage (and why there should be an RI stage to begin with). In other words, what kind of learning mechanisms lead the child to abandon an RI grammar, and on the basis of what kind of learning data? Biological maturation, which presumably takes place largely independent of linguistic data, is one possibility (Rizzi 1994, Wexler 1994, 1998). This proposal is not entirely satisfactory, however, as the mechanisms of biological maturation of linguistic ability are not currently well understood. Moreover, biological maturation abandons the Continuity Hypothesis (McNamara 1982, Pinker 1984, Crain 1991, Carey 1995), the hypothesis that children’s competence system is not qualitatively different from adults’, which has served well in the investigation of child language and cognitive development.

More empirically, it is again useful to bear in the mind the gradient distribution of RI across languages as well as within languages. As noted earlier, the extent of RI use across languages varies quantitatively along a broad spectrum, and for languages with relatively long RI stages, the percentage of RI use declines gradually. Even during the RI stage, it is not the case that the child uses RI exclusively (hence the optionality of RI). These facts seem to indicate that the RI phenomenon cannot be a categorical or universal deficiency in children’s grammar. Nor do they support the classic view of grammar development as an on-or-off process of switching parameter values, e.g., triggering (Gibson & Wexler 1994; cf. Yang 2002). Finally, it seems unlikely that RI is entirely due to the limitations on processing and other performance sources. For example, there appears to be no independent reason to suppose that Italian children’s pragmatic capacity is significantly more advanced than Dutch children’s, whereas the former group has at most a very short RI stage and the latter group has a prolonged one. Given the strong correlation between the productivity of RI and the morphological richness of the language (Phillips 1995, Guasti 2002), it appears that morphological learning must play a crucial—and quantitative—role in the explanation of the RI phenomenon.

Building on these insights from earlier work, we propose an alternative approach to the problem of Root Infinitives, one that is equipped with a concrete model of acquisition in which morphological learning across languages is connected to the underlying grammatical system of RI.
2 Optionality, Grammar, and Morphology

2.1 Variational learning

We assume the variational learning approach to language acquisition (Yang 1999, 2002, 2004; see Roeper 2000, Kroch 2001, Crain & Pietroski 2002, for similar approaches). Under variational learning, the child’s language is modeled as a finite population of grammars defined by the innate parameter space of UG. Each grammar, or more specifically each parameter, is associated with a probability. It is this probability distribution that changes adaptively in response to the linguistic data in the environment. (This critically differs from models such as triggering, where a single grammar changes into another grammar on the basis of linguistic evidence.) Schematically, learning proceeds as follows:

(2) For an input sentence $s$, the child

a. with probability $P_i$, selects a grammar $G_i$,

b. analyzes $s$ with $G_i$

c. • if successful, reward $G_i$ by increasing $P_i$

• otherwise punish $G_i$ by decreasing $P_i$

In a parametric approach to syntax, the competing candidates should be interpreted as alternative values of parameters. A formal treatment of the variational learning model in a parametric space is beyond the scope of this paper; see Yang (2002) for details.

The variational framework has a number of features that distinguishes it from the traditional conception of parameter setting. First, the rise of the target grammar is gradual, as its probability gradually approaches 1; this is characteristic of language development in general. Second, the demise of non-target grammars is also gradual: before these are eliminated, they will be accessed (probabilistically), which leads to a principled interpretation of “errors” in child grammar as potential adult grammars made available by UG, in line with the guiding principle of the Continuity Hypothesis. Finally, the speed with which a grammar (or a parameter value) rises to dominance is correlated with how incompatible its competitor is with the input, which can be determined by syntactic investigation of the cross-linguistic variations in the parameters framework. The frequency of such input evidence can be directly estimated from corpora of child-directed speech such as CHILDES (MacWhinney 1995), which in turn allows one to make quantitative assessments of longitudinal trends in syntactic development.

Previous analyses of RI, which include both generative accounts as well as empiricist learning accounts (e.g., Pine et al. 2005), have largely focused on the deviation of the...
RI grammar from the target grammar that the child eventually acquires. The variational framework, by contrast, seeks parallels between how the RI grammar deviates from the target grammar, and how the RI grammar relates to the totality of grammatical options made available by the innate UG. It is therefore useful to turn to the typology of languages in search of possible competing grammars that may form the basis of optionality in the RI phenomenon. In all RI languages, tense is an active morpho-syntax feature of the grammar: call these [+Tense] (tense) languages. However, in many other languages—call these [-Tense] (non-tense) languages—tense is not an active morpho-syntactic feature. A few examples from Mandarin Chinese are given below. Although Mandarin lacks tense, the semantics of temporal distinctions may still be expressed in the language through adverbial adjuncts (3a) (see Enç (1987), Smith (1991) for discussion related to the semantic encoding of temporal relations in [-Tense] languages). Aspect morphology may also be present, (3b); [± Asp] (aspect) is independent from [± Tense].

(3) a. Zhangsan tiantian da qiu.
Zhangsan everyday play ball.
"Zhangsan plays ball everyday."

b. Zhangsan zai da qiu.
Zhangsan ASP play ball.
"Zhangsan is playing ball."

Important for the present discussion is that since UG makes [-Tense] grammars an available option (i.e., Mandarin Chinese is a human language), a child acquiring a [+Tense] language must rule out this option. Under the variational approach of probabilistic learning, the [-Tense] option may take non-zero time to be eliminated, and its elimination is based on the input data in the linguistic environment of [+Tense] languages. If so, the use of non-finite verbs in root clauses of [+Tense] languages is expected, and expected to be optional, as the [-Tense] option is accessed with a decreasing probability.\(^2\) When the [-Tense] option is accessed, children will use verb forms that do not mark tense, i.e., infinitives. And these utterances will exhibit certain properties that are generally associated with infinitive verbs in adult grammars, as the existing literature on RI has amply demonstrated. Thus, the differences between an RI-stage grammar and the full specification of UG are only quantitative, rather than qualitative, contrary to the biological maturation accounts. And it is important to stress that the use of probabilistic learning, which draws insights from learning

\(^2\)So far as we know, Weinberg (1994), in a commentary on Wexler (1994), was the only other researcher that suggested identifying the use of RI with a grammar like Chinese, although this proposal was not developed further in her paper.
problems in other cognitive and perceptual domains (Bush & Mosteller 1951, Herrnstein & Loveland 1975), in no way denies or replaces the critical assumption of UG. The selection of grammars in learning requires a pool of candidates to select from, and that is supplied by parameters and UG; see Yang (2002) for additional discussion.

2.2 Morphological learning drives syntactic learning

On what basis does the child (gradually) eliminate the [-Tense] grammar in favour of the [+Tense] grammar? Most obviously, clauses with overt tense morphology, like English -ed, will reward the [+Tense] grammar and punish the [-Tense] grammar. However, the issue is more complex in two respects. First, certain morphemes that do not express tense directly nevertheless serve as evidence for the [+Tense] grammar. Consider English -s, which marks third singular, but only in the present tense. It is well established that children have determined the correct features realized by tense/agreement morphology well before they use them consistently; for example Brehens (1992), Guasti (1992), Torres (1995) demonstrate that children's morphological errors are errors of omission, rather than use in inappropriate morpho-syntactic contexts. In addition, there is evidence that English children can correctly identify English 3sg -s as expressing 3sg in the environment of present tense. For example, Harris & Wexler (1996) show that even though English children do not use -s consistently in third person singular present tense, they almost never use it in incorrect person/number combinations. Thus, tokens of 3sg -s, although not expressing tense, do require tense for their realization, and therefore also reward the [+Tense] grammar and punish the [-Tense] grammar. When determining the frequency of relevant input to the child in any particular language, morphology that requires tense must be included (for related discussion, see Carstairs 1987 on primary versus secondary exponence, and Harley & Noyer 1999 on primary versus secondary expression of a feature).

Second, consider the impact of verb forms that express overtly neither tense, nor morphology dependent on tense. Such forms are compatible with a [+Tense] grammar, which would entail the additional postulation of zero tense morphology. However, they are also compatible with a [-Tense] grammar: more straightforwardly, one might add, for no covert tense morphology need be postulated. If the child’s language acquisition device attempts to analyse the form with a [-Tense] grammar, the analysis will be successful and the form will (incorrectly) punish the [+Tense] grammar, and reward the [-Tense] grammar. Eventually, the evidence for the [+Tense] grammar must overwhelm this apparent evidence for the

---

Footnote: 3The willingness to accept zero morphemes depends on one’s theoretical persuasion; see, e.g., Anderson (1992) for a different view.
[-Tense] grammar, and zero tense morphology will be consistently posited. However, until that point, the verb forms that exhibit no tense or tense-dependent morphology are liable to punish the target [+Tense] grammar. Thus, it is important to consider the frequency of such forms when examining the relevant input to the child. We turn to this question in section 3.

3 Quantitative Evidence for Tense Across Languages

Under our proposal, the evidence for the grammatical option of [±Tense] is manifested in the verbal morphology of the language. Since the features active in the morphosyntax of a language is language specific, they must be learned. Learning requires data, which in turn takes time to accumulate. And this is where languages diverge with respect to RI: it is our contention that the morphological evidence for [+Tense] in languages with shorter RI stages is far more abundant than the morphological evidence in languages with extended RI use.

In this section, we quantify this claim with an analysis of the inflectional paradigms of three languages sampled from the RI spectrum, specifically, Spanish, French and English. It must be noted that the use of root infinitives, just like many other aspects of grammatical development, shows considerable individual variations. Thus, cross-linguistic generalizations of root infinitive use reflect tendencies rather than absolutes. And it is for this reason that we chose three languages for which children’s quantitative uses of root infinitives are significantly different, and these differences have been consistently found in the research literature. Spanish is a language where children have relatively infrequent use of root infinitives. For example, Grinstead (1994) found that prior to 2;0, about 10% of otherwise finite verbs are infinitives, and by 2;6, this figure drops below 5%. English, by contrast, has a much prolonged RI stage. As noted by Phillips (1995) (see also Hoesktra & Hyams 1998), Adam’s use of infinitives was still extensive at 3;0. Even the linguistically precocious Eve managed to produce approximately 50% of root infinitives by the time her recording sessions stopped at 2;3. In the acquisition of French, the frequency of root infinitives is found to at 15–30% for three children in the age group of approximately 1;8–2;6 (Rasetti 2003), while Pierce’s (1992) study of the same children reported somewhat higher frequencies. Taken together, Spanish, English, and French show significant quantitative differences with respect to children’s use of RI during language acquisition.

We show that the disparity between these three languages can be attributed to their verbal morphologies, which lead to the rapid decline of the [-Tense] grammar in Spanish,
and a more gradual decrease in English, with French situated between these two extremes.

3.1 Morphological Evidence for Tense in Spanish

In order to estimate the amount of crucial data available to the Spanish child, we counted the frequency of tensed verb forms. Of these forms, we counted how many express neither tense morphology, nor tense-dependent morphology (see discussion of primary and secondary expression of features in section 2.2 above), and thus are liable to incorrectly punish the [+Tense] grammar. The present tense of verbs in Spanish do not exhibit an overt tense morpheme; thus, any verbs in the present tense without tense-based agreement or other tense-based morphology would fall into this category of incorrectly punishing [+Tense]. All person/number combinations, excepting the third singular discussed immediately below, exhibit agreement suffixes that do not appear on untensed verb forms—infinitivals and participles (the latter expressing aspect rather than tense). Therefore, these all correctly reward the [+Tense] grammar.

(4) Present cantar “to sing”
   1sg canto
   2sg cantas
   3sg canta
   1pl cantamos
   2pl cantáis
   3pl cantan

(5) cf Infinitive
    cantar

(6) cf Past Participle
    cantado

(7) cf Present Participle
    cantando

Consider the third singular form of the present. For verbs that end in ar in the infinitive, like the illustrated cantar (“to sing”), there is an ambiguity as to whether the final theme vowel a in the third singular should be considered part of the stem, and so not consist of tense-based morphology, or be considered a tense-based agreement affix. We chose to
make the most conservative choice, and treated all such verbs as lacking tense-based agree-
ment morphology. Thus, all figured in the count of forms potentially incorrectly punishing
the [+Tense] grammar. In addition, the second singular imperative, identical to the third
singular present for these verbs, was also included in the count:

(8) Imperative
canta

For a second class of verbs, ending in *er* in the infinitive, identical considerations apply:

(9) Present *comer* “to eat”
   1sg como
   2sg comes
   3sg come
   1pl comemos
   2pl coméis
   3pl comen

(10) cf Infinitive
    comer

(11) cf Past Participle
    comido

(12) cf Present Participle
    comiendo

(13) Imperative
    come

Thus, the third singular forms of such verbs were also counted as potentially punishing the
[+Tense] grammar.

For a third class of verbs, those ending in *ir*, the issue plays out slightly differently. In
these verbs, the theme vowel does not form part of the stem for the infinitive or the partic-
ciples; thus, in these forms the theme vowel does consist of tense-dependent morphology,
and these forms correctly reward the [+Tense] grammar.

(14) Present *vivir* “to live”
Finally, several classes of Spanish verbs require a stem change in the singular and in third plural of the present tense;\(^4\) an example follows:

(19) Stem change in present tense: *poder* “be able to”

<table>
<thead>
<tr>
<th></th>
<th>1sg</th>
<th>2sg</th>
<th>3sg</th>
<th>1pl</th>
<th>2pl</th>
<th>3pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>v</td>
<td>puedo</td>
<td>puedes</td>
<td>puede</td>
<td>podemos</td>
<td>podéis</td>
<td>pueden</td>
</tr>
</tbody>
</table>

(20) cf Infinitive

<table>
<thead>
<tr>
<th></th>
<th>poder</th>
</tr>
</thead>
</table>

(21) cf Past Participle

<table>
<thead>
<tr>
<th></th>
<th>podido</th>
</tr>
</thead>
</table>

(22) cf Present Participle

<table>
<thead>
<tr>
<th></th>
<th>podiendo</th>
</tr>
</thead>
</table>

\(^4\)This stem change is also found in the present subjunctive, and the imperative, both of which arguably involve (present) tense.
This stem change again requires tense for its application, and thus correctly rewards the [+Tense] grammar. The third singular present and second singular imperative forms of such verbs thus were not counted as punishing the [+Tense] grammar. The stem change verbs found in the files included the following:

\[(24)\]
\[
\begin{align*}
&\text{a. } o \rightarrow ue \\
&\text{dormir, poder, contar, acordar, colgar, llover, volver, oler, encontrar, mover} \\
&\text{b. } e \rightarrow ie \\
&\text{querer, tener, cerrar, venir, pensar, despertar, empezar} \\
&\text{c. } e \rightarrow i \\
&\text{decir, seguir, reñir}
\end{align*}
\]

We examined 1194 child-directed utterances from 9 files in the Fernández and Aguado corpus in the CHILDES database. The counts were carried out by hand. Table 1 reports our findings:

<table>
<thead>
<tr>
<th>Morphology</th>
<th>± Tense</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>tensed verbs</td>
<td></td>
<td>877</td>
</tr>
<tr>
<td>zero agr, zero tense, no stem change</td>
<td>-</td>
<td>200</td>
</tr>
</tbody>
</table>

Table 1: Verbal inflections in child-directed Spanish and the quantitative evidence for [±Tense] in grammar acquisition.

There are \(677/877=77.2\%\) of input tensed verbs with verbal morphology that unambiguously implicate the [+Tense] grammar. These are countered by only \(200/877=22.8\%\) of input tensed verbs that, by virtue of being consistent with the [-Tense] grammar, may impede the gradual dominance of the [+Tense] grammar. The [+Tense] grammar has a significant numerical advantage (54.4%). The morphology of Spanish, then, offers the child plenty of opportunity to learn that their language makes use of tense.

If our approach to RI is on the track, then the crosslinguistic distribution of RI is due to the (gradient) quantities in the morphological evidence for the [+Tense] grammar. We would therefore expect that the extended RI languages have a smaller amount of evidence for the marking of Tense than a brief RI language such as Spanish. To this end, we turn to a corpus study of child-directed English.
3.2 Morphological Evidence for Tense in English

In order to estimate the amount of crucial data available to the English child, we counted the frequency of verb forms that express tense, or require tense for their usage that are present in adult speech in English CHILDES. (Again, see discussion of primary and secondary expression of features above).

In English, past tense is overtly marked for most regular and irregular verbs, for example jump-ed and burn-t; all such tokens were counted. In addition, a class of irregular verbs with no tense suffix exhibit stem allomorphy dependent on the past tense, for example sang. These were counted as well. In the present, third person singular (3S) shows the tense-dependent -s; such tokens reward the [+Tense] grammar. In addition, regarding the verb to be, the forms am and are are found only in the present tense, and were counted as evidence for the [+Tense] grammar.

On the other side of the coin, forms in English exhibiting zero expression of tense and no tense-dependent morphology were counted as potential evidence against the [+Tense] grammar. In English these include, obviously, the non-third singular forms of regular verbs in the present tense. A small class of verbs that do not exhibit change in past tense, for example, hit-hit, also punish the [+Tense] grammar.

We took a sample of 58,447 child-directed sentences from Brown’s (1972) Harvard studies. Because English morphology is relatively simple, and the distributional cues for morphological analysis are relatively clear, there are industrial strength part-of-speech tagging tools that are sufficiently accurate and informative, thereby making large scale corpus analysis possible. For the present study, we used the part-of-speech tagger designed by Brill (1995), which achieves an accuracy of 95-97% on unrestricted English texts: this level of performance is acceptable for our purposes.\(^5\)

The results of counts are summarized in Table 2:

<table>
<thead>
<tr>
<th>Morphology</th>
<th>± Tense</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tensed verbs</td>
<td></td>
<td>27198</td>
</tr>
<tr>
<td>present non-3s, past (no change)</td>
<td>–</td>
<td>12405</td>
</tr>
</tbody>
</table>

Table 2: Verbal inflections in child-directed English and their relations to the [±Tense] acquisition.

\(^5\)We should add that the morphological tagging provided as part of the CHILDES database distribution does not appear to be as accurate as the Brill tagger, at least impressionistically. Proper evaluation of the CHILDES tagging ought to be carried out using standard tools and datasets in computational linguistics.
The quantitative analysis of English morphology reveals that there is in fact relatively little unambiguous evidence for the [+Tense] option in the grammar. Specifically, there are \(\frac{14793}{27198} = 54.4\%\) of sentences that implicate a [+Tense] grammar, but these are countered by \(\frac{12405}{27198} = 45.6\%\) of sentences that suggest otherwise. Though [+Tense] still possesses the numerical advantage (8.8\%), it is far less weighty than in the case of Spanish.

### 3.3 Morphological Evidence for Tense in French

Finally, we considered the evidence available to the French child regarding the status of the [+Tense] grammar in her language. Similarly to Spanish, French does not exhibit overt tense marking in the present or the imperative. The agreement morphology found in French is significantly impoverished as compared to Spanish. A paradigm follows; in contrast to Spanish, the only pronounced suffixes are those in the first plural and second plural.

(25) Present *chanter* “to sing”
- 1sg chanté
- 2sg chantes
- 3sg chante
- 1pl chantons
- 2pl chantez
- 3pl chantent

(26) cf Infinitive
- chanter

(27) cf Past Participle
- chanté

(28) cf Present Participle
- chantant

(29) Imperative
- chante

Thus, more present verb forms in French than in Spanish fail to provide unambiguous evidence for the [+Tense] grammar. This is all the more true in that the first plural and second plural are rarely used in spoken French (we found only two instances in our sample).
Like Spanish, many French verb forms exhibit tense-based stem change in the present. These, despite the lack of overt agreement or tense morphology, do provide evidence for the [+Tense] grammar. An example follows:

(30) Present *tenir* “to hold”
1sg tiens
2sg tiens
3sg tient
1pl tenons
2pl tenez
3pl tiennent

(31) cf Infinitive
    tenir

(32) cf Past Participle
    tenu

(33) cf Present Participle
    tenir

(34) Imperative
    tiens

Such verbs found in the selection included *tenir, re(venir), savoir, pouvoir, vouloir, falloir, devoir, sortir*. In addition, the verbs *aller, avoir, and être* (excepting the second plural) show significant stem allomorphy in the present, and thus also provide evidence for the [+Tense] grammar:

(35) Present *aller* “to go”
1sg vais
2sg vas
3sg va
1pl allons
2pl allez
3pl vont

(36) Present *être* “to be”
We examined 1046 child-directed utterances from the Phillippe corpus in the CHILDES database. The counts were carried out by hand. Table 3 reports our findings:

<table>
<thead>
<tr>
<th>Morphology</th>
<th>± Tense</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>tensed verbs</td>
<td></td>
<td>1049</td>
</tr>
<tr>
<td>zero agr, zero tense, no stem change</td>
<td>-</td>
<td>316</td>
</tr>
</tbody>
</table>

Table 3: Verbal inflections in child-directed French and the quantitative evidence for [±Tense] in grammar acquisition.

There are $733/1049 = 69.9\%$ of input tensed verbs with verbal morphology that unambiguously implicate the [+Tense] grammar. These are countered by $316/1049 = 30.1\%$ of input tensed verbs that, by virtue of being consistent with the [-Tense] grammar, may impede the gradual dominance of the [+Tense] grammar. The evidence that favors the [+Tense] grammar constitutes 39.8% of the input. The morphology of French, then, provides far more opportunity than English (8.8%) for the child to learn that the tense is an active morpho-syntactic feature, but not as much as Spanish (54.4%). The status of the French RI stage as intermediate between English and Spanish is thus directly accounted for.

4 Conclusion

The simple proposal outline here provides a learning component for an overall theory of RI that retains many of the insights in previous analyses. We agree with all (generative)
accounts that RI is a genuine grammatical phenomenon rooted in UG and thus constitutes a forceful argument against the empiricist position to language acquisition. We agree with Schütze & Wexler (1996) and Wexler (1998) in noting the connection between tense and agreement marking in both child and adult grammars. We agree with Hoekstra & Hyams (1998) and Schaeffer & Bar Shalom (2004) in recognizing that morphology plays an important and revealing role in the underlying grammatical system of Tense. Our account is closest to that Phillips (1995), which also relates morphological learning to RI in a quantitative way. In Phillips’ account, the underlying grammar system during the RI stage is just like the target adult grammar and the performance of the child’s morphological system is directly responsible for the use of RI. In our account, the underlying RI grammar system is a statistical ensemble of potential adult-like grammars (including but not limited to the target grammar), and that is because the morphological system, still under developed, has yet driven out the [-Tense] option. Note that the morphological learning approach advocated here may offer an explanation for the extended root infinitive stage in a population of SLI children (Rice & Wexler 1996, Rice, Wexler, & Hershberger 1998). For us, the extended RI stage is not the result of underlying grammatical deficits. Rather, as has been well documented in the acquisition of past tense (Leonard et al. 1992, Rice et al. 2000, and the references cited therein), SLI children may simply be slow morphological learners, which in turns slows down the emergence of the [+Tense] grammar.

Note that crucial to our approach is the assumption that the child is a superb morphological learner, capable to penetrating the morphological structures of her language, which in turns guides the development of her syntactic knowledge. While this assumption is empirically justified–here the generative approach (reviewed in, e.g., Phillips 1995, Guasti 2002) confirmed the traditional observations of child language (Slobin 1973, Pinker 1984)–it remains a mystery how children do so. For an approach that relies heavily on the connection between the learning of morphology and the learning of syntax, it is important to develop a concrete theory of morphological acquisition. Future research will lead to more extensive quantitative studies of the morphological data available to children acquiring a diverse range of languages, and how such data lead to quantitatively different patterns in the development of Tense.

Another interesting direction for research may focus on the properties of root infinitives and their interactions with the development of the aspect system; see Becker (2000) for an extensive investigation. If our approach is correct, that root infinitives involve a [-Tense] grammar, then one might expect to find evidence for aspect marking in such structures. For example, the root use of participles (e.g., “puppy gone”), in marking aspect rather than
tense, is perfectly grammatical under a Chinese-like [+Asp] grammar. This invites one to search for syntactic and semantic parallels between, say, adult Chinese and children’s English that marks aspect in RI utterances; cf. Yang (2002) on the parallels between English children’s null arguments and Chinese adults’ topic drops. We leave this for future work.

In conclusion, we would like to suggest that both the input data and the mechanisms by which children internalize a grammar on the basis of the input data be taken seriously in the generative approach to language acquisition. Doing so in no way diminishes the importance of Universal Grammar, but Universal Grammar does not have to do all the work for the success of language acquisition. Under the contemporary idea that much of language variation (and thus acquisition) comes down to the acquisition of the lexicon, it would seem more pressing to develop a theory of experience-dependent language learning. (If anything is learned, it’s words.) For aspects of the grammar that take time to complete (such as the RI phenomenon), it remains a possibility that the child is not receiving the sufficient amount of the relevant linguistic evidence: what “counts” as the relevant linguistic evidence will inevitably turn on UG, for the child does not simply match and replicate what is presented in the input. It might be added that even for aspects of grammars that children successfully acquire very rapidly, it is still desirable to have a specific account of the learning mechanism, and how that mechanism makes use of the input data so spectacularly. UG is a biological miracle: learning needn’t be.

References


21


