Competing motivations in children’s omission of subjects? The interaction between verb finiteness and referent accessibility

MARY E. HUGHES AND SHANLEY E. M. ALLEN

9.1 Introduction
Over the last fifty years since child language development has become a focus of scientific study, research and reasoning in the field has largely been divided along theoretical lines (Tomasello 2000; Crain and Pietroski 2002; Pullum and Scholz 2002; MacWhinney 2004; Valian 2009). Nativist accounts claim that children have adult-like grammatical competence from the outset as a result of innate linguistic knowledge. They highlight the central role of formal structures in language development, and acknowledge little relevance of interaction or external input other than to trigger innate linguistic knowledge and provide lexical items for the language at hand (Crain and Thornton 1998). In contrast, emergentist and usage-based accounts assume no innate knowledge specific to language. They rather emphasize the crucial role of external input in language development since they hold that language is learned through interactive cues from caregivers, distributional patterns in the input, frequency information, and the like (Tomasello 2003a). Research from each theoretical perspective has largely been pursued in isolation from the others, with the implicit assumption that other perspectives are wrong, irrelevant, or uninteresting as a suitable explanation of the issues at hand. In other words, discussion in the field has largely been framed as a theory-based competition between explanations or motivations for particular phenomena in language development. Each theory champions one motivation to explain a particular phenomenon, and any inter-theory discussion focuses on which one account provides the best explanation of the data. A similar pattern is found in the literature on explanations for phenomena in linguistic structure, as discussed in other chapters in this volume (see Strunk, this volume, on relative clause extraposition, and Pfeiffer, this volume, on self-repair).
Only in the last fifteen to twenty years have language development researchers seriously attempted to evaluate theoretical positions in the light of each other rather than simply ignoring or dismissing other positions (Russell 2004; Ambridge and Lieven 2011; Dressler, Libben, and Korecky-Kröll, this volume; Krajewski and Lieven, this volume). In ideal cases, the focus has moved from assuming that theories compete with each other—such that one theory explains subset A of the data while another theory explains subset B—or even interact with each other—such that factors relevant to two theories are not independent from each other but rather are both necessary to explain the full set of data. As a result, understanding has advanced considerably in certain areas of language development including how children identify syntactic categories, acquire basic word order, retreat from overgeneralization errors, and acquire complex syntactic constructions like passives and relative clauses. For other key phenomena in language development, however, much progress remains to be made.

The present chapter focuses on one such phenomenon: the early non-target-like omission of sentential subjects. This issue has been central to argumentation in several competing theories of language development for over thirty years (e.g., Bloom 1970; Hyams 1986; Clancy 1993), but little thorough attempt has yet been made to determine whether the explanations posited for it by different theories indeed compete, or whether they might be complementary or even interact (though see Valian 1991). In this chapter, we compare explanations for early non-target-like subject omission from the two theories most current in the literature—nativists claim that subjects are omitted when their accompanying verb is non-finite, while usage-based theorists claim that subjects are omitted when their referents are accessible in discourse. Through analysis of spontaneous speech data from monolingual English-speaking children and their caregivers, we demonstrate that these two motivations interact, rather than competing as commonly claimed in the literature.

9.1.1 Early non-target-like subject omission

It has been widely observed that young children (up to about 3;6) acquiring a first language omit subjects more frequently than adults, thereby producing utterances that are either ungrammatical or highly marked by the norms of the target language as shown in (1).

(1)  Ø fall down [cf. The grape fell down] (Thomas 2:5)
    (telling his mother about a grape that fell on the floor)

This over-omission occurs both in languages like English and German where the target requires overt subjects (e.g. Bloom 1990; Hyams and Wexler 1993) and in languages like Italian and Inuktitut where the target does not require overt subjects (e.g. Valian 1991; Valian and Eisenberg 1996; Skarabela and Allen 2002). However, here we focus on the former, and specifically on English.
A veritable cottage industry of research in the last 30 years has investigated why young children omit subjects in languages where adults do not. Explanations have been put forward from three main theoretical accounts. According to the nativist account, the child grammar differs from the adult grammar in some crucial way such that subject omission is permitted as a grammatical option for the child (see review in Hyams 2011). The cognitivist account rather posits that subject omission results from processing or production limitations due to the child’s limited cognitive capacity (e.g. L. Bloom 1970; P. Bloom 1990; Valian 1991; Gerken 1991). From the perspective of the usage-based account rooted in discourse-pragmatics, children are sensitive to interlocutor knowledge and omit subjects when they think the referent is already known to the interlocutor (see review in Allen, Skarabela, and Hughes 2008).

We focus here on the nativist and usage-based accounts since they are the most salient in the current literature.

Most research on early non-target-like subject omission in English has been from the nativist perspective, which posits that the locus of subject omission lies in the grammar rather than in other aspects of the linguistic system. Research has thus focused on ways in which the child grammar differs from the adult target grammar such that subject omission is permitted for the child but not for the adult. The two currently accepted hypotheses, “underspecification” (Hyams 1996; Wexler 1998) and “truncation” (Rizzi 1993/1994, 2005), both posit child grammars that allow children optionally to produce non-finite verbs in contexts where adults would produce finite verbs—the so-called “root infinitive” stage. (Note that this is different from “infinite” forms which are correctly produced in the adult target without number, person, or tense marking.) An example of a root infinitive is shown in (2) (Brown 1973).

(2)  Papa have it [cf. Papa has it] (Eve 1;6)

These non-finite verbs are missing the number, tense, and agreement features that are required in the equivalent adult target utterance. Because the number, tense, and agreement features are what trigger production of subjects under the nativist account, children assume that the subject can be omitted. Indeed, numerous studies have found that subjects are omitted much more frequently in non-finite clauses than in finite clauses in child English (e.g. Sano and Hyams 1994; Bromberg and Wexler 1995; Wexler 1998; see summary of crosslinguistic results in Guasti 2002: 163). Researchers from the grammatical perspective have thus concluded that the frequent lack of finiteness provides a nearly complete explanation for early subject omission in English. Little attention is given to the omitted subjects of finite verbs that do not fit within this explanation.

Detailed research on child subject omission from the usage-based perspective began some twenty years ago with Clancy (1993), and has posited children’s discourse-pragmatic sensitivity to the complexities of information flow as the locus of early subject omission. Numerous studies have shown that adults choose whether and in what form (e.g. pronoun, demonstrative, lexical NP) to realize subjects in their speech depending on their belief about the cognitive accessibility of the referent (e.g. Givón 1983a; Bock and Warren 1985; Gundel, Hedberg, and Zacharski 1993; Ariel 2001). A variety of discourse factors feed into determining accessibility. For
example, a subject is considered accessible and thus is often omitted when its referent has just been mentioned, is present in the physical context, and/or is the current focus of attention of the speaker and hearer. Both experimental and spontaneous speech studies have revealed that English-speaking children are sensitive to the same discourse-pragmatic factors as adults, and omit subjects in contexts where the referent is cognitively accessible (Campbell, Brooks, and Tomasello 2000; Serratrice, Sorace, and Paoli 2004; Hughes and Allen 2006, 2013; Guerriero, Oshima-Takane, and Kuriyama 2006; Matthews et al. 2006; Graf 2010; see summary of crosslinguistic results in Allen, Skarabela, and Hughes 2008). Thus, researchers from the usage-based perspective have concluded that the cognitive accessibility of the referent provides a solid explanation for early subject omission in English.

9.1.3 Research questions

As just discussed, nativist and usage-based accounts make very different claims about the motivation for children’s early non-target-like omission of subjects and are usually considered orthogonal. To our knowledge, no previous studies from the usage-based perspective have investigated any possible complementarity or interaction with the nativist perspective apart from occasionally acknowledging that the possibility of subject omission at all is controlled by the grammar (Serratrice and Sorace 2003). The few previous studies from the nativist perspective that have investigated a possible role for discourse-pragmatics in explaining subject omission have been limited by a lack of detail or depth in their analysis of the data. Hamann and Plunkett (1998) explored the effect of only one feature of discourse-pragmatics—whether a subject is newly introduced to discourse or already previously introduced. And the oft-cited study by Hyams and Wexler (1993) is based on data for which it is simply assumed that all omitted subjects and pronominal subjects in the data are fully accessible to the listener, rather than on data where each subject is actually coded for referent accessibility. In a review of previous research on subject omission from the nativist perspective, Hyams (2011) acknowledges a potential role for discourse-pragmatic factors in affecting subject omission, but limits this possible effect to subjects of finite verbs without substantiating her assumption with data. She states: “I will assume that the missing subjects in non-finite clauses are licensed by whatever mechanisms license PRO in infinitival contexts in adult grammars” and “I will restrict my discussion [of discourse-pragmatic explanations] to null subjects in finite clauses” (Hyams 2011: 24). She, as well as Wexler and his colleagues (Bromberg and Wexler 1995; Wexler 1998), consider “topic drop” (very similar in practice to the discourse-pragmatic notion of cognitive accessibility) a likely candidate for explaining subject omission, but only in the 10% or so of cases of subject omission in finite contexts.

In the present study, then, we investigate the possible complementarity or interaction of these two “competing motivations” for early non-target-like subject omission in English—non-finiteness from the nativist account and discourse-pragmatic accessibility from the usage-based account. We evaluate Hyams’s (2011) assertion that the two
motivations are complementary—that the nativist account (i.e. non-finiteness) explains the bulk of subject omissions and the usage-based account (i.e. accessibility) plays a role only for a small subset of omissions—and test it against the possibility that the two theories rather crucially interact such that both non-finiteness and accessibility are essential for explaining the full range of data. We do so using a set of naturalistic spontaneous speech data from English-speaking children aged 2 and 3 years (i.e. early and late in the subject omission stage), coded for both finiteness and accessibility.

The nativist account makes two predictions relevant to our study. First, no effect of discourse-pragmatics will be evident for subject omission in non-finite contexts (Prediction N1). Second, the asymmetry of more subject omission in non-finite than finite contexts will not be explainable by a discourse-pragmatic account (Prediction N2). The usage-based perspective also makes two predictions relevant to our study. First, accessible subjects will be realized in the same way regardless of whether they appear in finite or non-finite contexts (Prediction U1). Second, there will be no relationship between referent accessibility and finiteness (Prediction U2).

If the competing theories approach is correct, then both the predictions from one perspective will be met and neither of the predictions from the other perspective.

If the complementary theories approach is correct under the scenario proposed by Hyams (2011), then both the predictions from the nativist perspective will be met (i.e. N1 and N2), and neither of the predictions from the usage-based perspective will be met (i.e. neither U1 nor U2). Further, accessibility will be shown to play a role in explaining subject omissions in finite contexts.

If the interacting theories approach is correct, then none of the four predictions will be met because both non-finiteness and accessibility will play a crucial role in explaining patterns for all of the data. We would then expect effects of accessibility in each of non-finite and finite contexts, and we would expect discourse-pragmatic explanations for the finite/non-finite asymmetry; for example, that non-finite verbs attract subjects with low information content or that accessible referents attract non-finite verb forms.

9.2 Methodology

9.2.1 Participants

Data for this study were taken from The Manchester-Max Planck dense database (Lieven et al. 2003; Lieven, Salomo, and Tomasello 2009). Participants were four monolingual English-speaking children: Annie, Thomas, Eleanor, and Fraser. All the children lived with their families in a large metropolitan area in England and came from middle-class backgrounds. The children’s language was recorded while interacting with their mother and/or a familiar researcher during such everyday activities as playing with toys and having snacks.

1 Thomas is sometimes referred to as Brian, which was the Max Planck pseudonym for the child.
For this study, we used only the videotaped portion of the dense database corpus. Videotapes were collected one hour per week for six weeks at ages 2;0–2;1, and again at 3;0–3;1 (seven weeks were available for Annie at the younger age, and only three weeks at the older age). Thomas was also videotaped once per week from 2;1 to 3;0. All utterances spoken by and to the child were transcribed by research assistants in CHAT format (MacWhinney (2000); see Lieven et al. (2003) for further details about the method of recording and transcription). The transcripts and videotapes were kindly made available to us by Elena Lieven.

Previous studies have shown that both frequency of subject omission and sensitivity to accessibility features change over time (e.g. Valian 1991; Serratrice 2005; Matthews et al. 2006). Therefore, we analyzed utterances separately at the two available time points to assess development. Time 1 for three of the children comprised data from 2;0 to 2;1. One child, Thomas, had a somewhat lower mean length of utterance (MLU) than the other three children during these earliest recordings. Since we also had later recordings for him, we used a slightly later set of data for his Time 1—from 2;4 to 2;7—so that the linguistic level of all four children would be as comparable as possible.2 Time 2 for all four children comprised data from 3;0 to 3;1, by which point the children all had a similar linguistic level as measured by MLU.

We then determined which subjects from the available data to include in our coding and analysis. Only fully intelligible utterances containing a verb were selected for coding. Because a crucial part of our study is determining the effect of discourse-pragmatics on subject omission, we excluded from analysis all utterances in contexts where the omission or overtens of the subject is fixed and thus cannot vary depending on the accessibility of the referent for that subject. This comprised four types of contexts (see Hughes and Allen 2013 for a detailed explanation): all imperative utterances (where subjects are uniformly prohibited), all utterances containing first and second person subjects (where subjects are considered accessible by definition), all utterances containing subjects that are explicitly contrasted with an alternative referent (e.g. using stress or tone of voice—where subjects cannot be omitted by definition), and all utterances containing subjects that themselves are the referent of or the answer to a question (where subjects cannot be omitted by definition).

The final data set for analysis, then, comprised all fully intelligible and non-imperative utterances that included a verb and a third-person subject context that was not explicitly contrasted or queried. The age ranges, MLUs, and number of utterances analyzed for each child is given in Table 9.1.

2 Lieven, Salomo, and Tomasello (2009) analyzed the T1 data from all four children and mention the same problem. They took the approach of analyzing Thomas’s data at four different time points—one at the same age as the other children and then three later ages, one to match the MLU of each of the other children at 2;0.
These utterances were coded for both discourse-pragmatic and grammatical information, including the finiteness of the verb. In order to test intercoder reliability, 9% of the data was blind-coded by two research assistants. These two coders achieved an average agreement of 85%. Moreover, every file was coded by one research assistant, and then subsequently checked by a second research assistant. All files were then reviewed by the first author in order to resolve any intercoder differences.

9.2.3 Accessibility coding

All third-person subjects that fit the above criteria were coded for accessibility. Six accessibility features were selected for analysis, based on those previously investigated in child language studies (Allen et al. 2008): animacy, physical presence, contextual disambiguation, prior mention, linguistic disambiguation, and joint attention. Each feature was given a binary value based on how accessible the referent of that subject was in the discourse context (Clancy 1993; Allen 2000; Ariel 2001; Skarabela 2007).3 The value accessible indicates that the referent can be easily identified in the discourse context on the basis of the feature being coded, while the value inaccessible indicates that the referent is not easily identified in the discourse context on the basis of the feature being coded. The values for each feature are summarized in Table 9.2.

To illustrate the application of this coding, consider the subject of the child utterance in the interaction in (3). At the time of this utterance, the mother and child are playing with multiple blocks on the table. The mother is looking at and touching the block in question (Block B), and the child is also looking at it.

3 In reality, these features are not necessarily binary; instead, they may be more gradient in nature. However, as it has been the practice in the literature to treat these features as binary in order to simplify the analysis and to allow for statistical analysis based on these values, the current study maintains the use of binary values.
**Table 9.2. Accessible and inaccessible values for each accessibility feature**

<table>
<thead>
<tr>
<th>Accessibility feature</th>
<th>Accessible value</th>
<th>Inaccessible value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ANIMACY (AN)</strong></td>
<td>Animate</td>
<td>Inanimate</td>
</tr>
<tr>
<td><strong>CONTEXTUAL</strong></td>
<td>Only referent in physical context</td>
<td>Multiple referents in physical context</td>
</tr>
<tr>
<td><strong>PHYSICAL PRESENCE (PP)</strong></td>
<td>Physically present</td>
<td>Physically absent</td>
</tr>
<tr>
<td><strong>PRIOR MENTION (PM)</strong></td>
<td>Given: mentioned within preceding 5 utterances</td>
<td>New: not mentioned within preceding 5 utterances</td>
</tr>
<tr>
<td><strong>LINGUISTIC DISAMBIGUATION (LD)</strong></td>
<td>No other possible referents in preceding 5 utterances</td>
<td>Other possible referents in preceding 5 utterances</td>
</tr>
<tr>
<td><strong>JOINT ATTENTION (JA)</strong></td>
<td>Referent focus of attention for child and interlocutor</td>
<td>Referent not focus of attention for child and interlocutor</td>
</tr>
</tbody>
</table>

(3) MOT: *That block fell down.* [referring to Block A]
MOT: *Where does this block go?* [referring to Block B]
CHI: *Ø go there.* [referring to Block B] (Annie 2:0)

The subject in the final utterance was coded *accessible* for three features: **physical presence, prior mention, and joint attention**. It was coded *inaccessible* for the other three features.

Following the method outlined in Hughes (2011), we then calculated an **inaccessibility score** for each subject. Each feature coded as *accessible* counted as “0” while each feature coded as *inaccessible* counted as “1.” The possible values for the inaccessibility score in this study range from 0 (meaning that the argument was accessible for all six features) to 6 (meaning that the argument was inaccessible for all six features). For the subject under scrutiny in example (1), three of the six features were coded as inaccessible so its inaccessibility score is 3. As shown in Hughes (2011), a high inaccessibility score predicts that a subject is more likely to be realized using an overt lexical form.

### 9.2.4 Referential form coding

All subjects in the data were coded as one of four possible referential forms: lexical noun phrase (4a), demonstrative (4b), pronoun (4c), and omitted (4d).

(4) a. **Boy** go. (Thomas 2:7)
   b. **This** go round here. (Fraser 2:0)
   c. **It** goes there. (Annie 2:0)
   d. Ø want chips. (Eleanor 2:0)

---

4 It is most likely the case that each feature does not contribute equally to subject choice, but for the purposes of this discussion, the features will be summed without regard to the possible variation in the weight of each feature's contribution.
9.2.5 Verb finiteness coding

All verbs associated with the subjects under study were coded for finiteness. Limiting the analysis to third person subjects meant that the finiteness of the verb was relatively easy to establish. However, because English does not have a true infinitival form that differs morphologically from other verbal forms, it was necessary to establish specific criteria to determine the status of non-finite verb forms. In sum, only correctly conjugated main verbs or correctly conjugated forms of the auxiliaries *be*, *have*, and *do* were considered finite. The detailed possibilities for finite third-person verbs are listed in (5).

(5)  

a. Regular and irregular verbs in the present tense where the agreement marker -s was produced or the correct form of the verb was produced—e.g. *He wants to park* (Annie, 3:0)
b. Present and past progressives for which the auxiliary verb *be* was produced and correctly conjugated—e.g. *My torch is shining* (Thomas, 2:11)
c. Present perfect and present perfect progressive verbs for which the auxiliary verb *have* was produced and correctly conjugated—e.g. *Helga has fall over* (Annie, 2:0)
d. Present and past passive verbs for which the auxiliary verb *be* was produced and correctly conjugated—e.g. *She was sting by a bee* (Fraser, 3:0)
e. Present and past verbs for which *do*-support was required, was provided, and was correctly conjugated—e.g. *Where does it go?* (Annie, 2:1)
f. Correctly conjugated tensed verbs produced with the auxiliary verb *be and the main verb* go with a future interpretation—e.g. *This is going to be very hot* (Annie, 3:0)
g. Present and past tense copular verbs correctly conjugated for third person5—e.g. *He’s in the car* (Fraser, 2:1)

The possibilities for non-finite third-person verbs are listed in (6):

(6)  

a. Regular and irregular verbs in the present tense for which the agreement marker -s was missing—e.g. *He don’t like to go in* (Fraser, 3:0)
b. Present and past progressive verbs for which the auxiliary verb *be* was missing—e.g. *The pig going now* (Eleanor, 3:1)
c. Present perfect verbs for which the auxiliary verb *have* was missing—e.g. *She been crying* (Annie, 3:1)
d. Passive verbs for which the auxiliary was omitted—e.g. *What that called?* (Eleanor, 2:1)
e. Verbs which require *do*-support for which the auxiliary *do* was missing—e.g. *No goes in there* (Fraser, 2:0)
f. Verbs that were produced with the auxiliary verb *be* and the main verb *go* with a future interpretation, for which the auxiliary was missing—e.g. *Big elephant going* (Thomas, 2:10)
g. Constructions where modals are missing—e.g. *Thomas put it on* [meaning: Thomas can or will put it on] (Thomas, 2:7)

5 Utterances with missing copulas were not included in the data set because they do not contain a verb.
Finally, the verb forms listed in (7) were excluded from the analysis because, although they show tense, they do not overtly show agreement in adult English:

(7) a. Modal verbs for which the verb form remains constant in all persons—e.g. She can stay there (Annie 3;0)
   b. Past and future forms in the simple, progressive, and perfect forms for which the verb form remains constant in all persons—e.g. That will be good (Fraser 2;1)

9.2.6 Hypotheses and data analysis

Our analysis of the data comprised four steps. First, we assessed subject form in relation to finiteness by determining the proportion of each of the four subject forms in finite vs. non-finite verbal contexts. Based on previous studies, we expected to find more omitted subjects in non-finite clauses and more of the three other forms in finite clauses. Second, we assessed subject form in relation to discourse-pragmatics by determining the proportion of each of the four subject forms for accessible vs. inaccessible referents. Again based on previous studies, we expected to find more omitted subjects for accessible referents and fewer omitted subjects for inaccessible referents. We also expected to find more lexical noun phrases for inaccessible referents, and pronouns and demonstratives used more for subjects of intermediate accessibility. All of these expectations were confirmed in our data.

Once we established that our data mirrored previous findings from each of the grammatical and discourse-pragmatic perspectives, we then assessed whether there was an interaction between the two accounts. In the third analysis, we investigated whether subjects of non-finite verbs were more likely to be highly accessible than subjects of finite verbs. The nativist perspective would predict that accessibility has no effect on which verbs end up being finite or non-finite (Prediction N2); it is only after finiteness is determined that accessibility might play a role in subject omission, and then only for subjects of finite verbs (i.e. if a complementarity account is assumed). The usage-based perspective would also predict that accessibility has no effect on which verbs end up being finite or non-finite, but rather that accessibility affects subject realization independently from finiteness (Predictions U1 and U2).

In the fourth analysis, we determined whether accessible referents were realized differently in non-finite vs. finite verbal contexts. Based on claims from the nativist perspective, we would expect that the realization of subjects would not be affected by accessibility in non-finite contexts (Prediction N1), but might be affected by accessibility in finite contexts (i.e. if a complementarity account is assumed). In contrast, the usage-based perspective would predict that effects of accessibility should be found equally in non-finite and finite contexts (Predictions U1 and U2).

9.3 Results

As expected on the basis of many previous studies of subject omission in English child language, the children in the present study omitted many subjects at Time 1
and relatively fewer at Time 2 (4%), $\chi^2 = 185.471, p < .0001$. Regardless, enough subjects were omitted at each age to allow us to continue with our analyses. In the following sections, we present the results from the four stages of the analysis as just described.

9.3.1 Effect of finiteness

Our first goal was to determine whether our data set mirrors findings in the literature that subjects are omitted more frequently in non-finite than in finite clauses. Figure 9.1 shows the distribution of the four argument forms in the context of verb finiteness at Time 1 and Time 2. Chi-square results show that the children at both ages omit significantly more subjects in the context of non-finite verbs than in the context of finite verbs: Time 1, 49% non-finite vs. 12% finite, $\chi^2 = 127.292, p < .0001$; Time 2, 25% non-finite vs. 1% finite, $\chi^2 = 145.969, p < .0001$. Thus, our data are consistent with previous findings in the literature from the nativist perspective.

Hyams and Wexler (1993) suggested that omitted subjects are really omitted pronouns and not omitted demonstratives or lexical NPs. If this were the case, we would expect that the number of pronouns plus omissions taken together would be the same in finite and non-finite contexts at both Time 1 and 2. However, this is not the case. There is a significantly higher number of pronouns plus omissions in non-finite than finite contexts at both time points: Time 1, non-finite 55% vs. finite 39%, $\chi^2 = 16.900, p < .0001$; Time 2, non-finite 67% vs. finite 50%, $\chi^2 = 10.834, p < .001$. When pronouns are analyzed alone, however, there is a significant difference in the number of pronouns in finite vs. non-finite contexts only at Time 1: Time 1, non-finite 55% vs. finite 27%, $\chi^2 = 56.368, p < .0001$; Time 2: non-finite 42% vs. finite 6%, $\chi^2 = 2.195, p = .138$.

Significantly more demonstratives occur in finite as opposed to non-finite contexts at both time points: Time 1, non-finite 5% vs. finite 27%, $\chi^2 = 58.757, p < .0001$;
Time 2: non-finite 8% vs. finite 26%, $\chi^2 = 18.688$, $p < .0001$. However, there is no difference in the number of lexical NPs occurring in finite vs. non-finite contexts at either time point: Time 1: non-finite 40% vs. finite 34%, $\chi^2 = 3.638$, $p = .056$; Time 2: non-finite 25% vs. finite 24%, $\chi^2 = 0.205$, $p = .651$.

These results suggest that, while there is obviously a strong relationship between subject omission and non-finite contexts, other factors are also at play.

9.3.2 Effect of accessibility

Next we determine whether subjects are omitted more frequently for referents that are highly accessible, as predicted by the usage-based account. Figure 9.2 shows the children’s use of the four different argument forms in the light of the accessibility of the subject’s referent, at Time 1 and Time 2. Recall that an inaccessibility score of 0 denotes a referent that is maximally accessible, while an inaccessibility score of 6 denotes a referent that is maximally inaccessible. The effect of accessibility is particularly visible at Time 1: subjects are omitted much more when the referent is accessible than when it is not. To test this statistically, we performed a $\chi^2$ analysis comparing the number of omissions for the most accessible subjects (inaccessibility score 0–1) vs. the least accessible subjects (inaccessibility score 5–6). The difference is significant: $\chi^2 = 18.069$, $p < .0001$. The same analysis at Time 2 was not significant ($\chi^2 = 1.274$, $p = .259$) because the children omitted very few subjects. Our data are consistent with expectations from the usage-based perspective.
9.3.3 Occurrence of finiteness with accessible vs. inaccessible referents

In the third analysis, we determine whether there is a relationship between finiteness and accessibility. Figures 9.3 (Time 1) and 9.4 (Time 2) feature the same accessibility categories as in Figure 9.2. However, rather than showing the form in which subjects were realized for each category as in Figure 9.2, Figures 9.3 and 9.4 show the finiteness of the clause in which the subjects were realized. If there is no relationship between finiteness and accessibility, we would expect no difference between the proportion of finite vs. non-finite clauses associated with highly accessible vs. highly inaccessible referents. If there is a relationship, we would expect more non-finite than finite clauses associated with highly accessible referents and, conversely, more finite than non-finite clauses associated with highly inaccessible referents.

![Figure 9.3 Cumulative effect of accessibility for six features by percent of finite and non-finite verb contexts at Time 1](image)

Total $n$ of subjects = 698; 67 subjects excluded due to missing values for JA; total $n$ for each category: 0 = 28, 1 = 50, 2 = 157, 3 = 198, 4 = 158, 5 = 73, 6 = 34

The pattern is very clear at Time 1. Non-finite clauses occur more frequently with highly accessible referents, and finite clauses occur increasingly as referents become less accessible. To test this statistically, we performed a $\chi^2$ analysis comparing the number of uses of finite clauses with the most accessible subjects (inaccessibility score 0–1) vs. the least accessible subjects (inaccessibility score 5–6). The difference is highly significant: $\chi^2 = 32.214, p < .0001$. We conclude that accessibility and verb finiteness are strongly related at Time 1.

At Time 2, however, this pattern does not hold because most of the children’s utterances consist of finite rather than non-finite clauses. A comparison of the number of occurrences of finite clauses with the most accessible subjects (inaccessibility score 0–1) vs. the least accessible subjects (inaccessibility score 5–6) revealed that there was not a significant difference ($\chi^2 = 0.488, p = .485$). Therefore, a relationship between accessibility and finiteness is only visible at Time 1. This relationship is not expected from either the nativist or the usage-based account.
It strongly supports the claim that both finiteness and accessibility play a role in early subject omission, and that neither can offer a full explanation on its own.

9.3.4 Realization of accessible referents in finite and non-finite contexts

Finally, we investigated the relationship between accessibility and finiteness in a slightly different way by determining whether accessible referents were realized differently in finite and non-finite verbal contexts. If there is no relationship, we expect no clear pattern of accessibility affecting subject realization for the subjects of non-finite clauses, because finiteness alone would determine whether subjects are omitted or overt. If there is a relationship, we expect the same pattern of effect of accessibility found in Figure 9.2 to appear separately for each of finite verb contexts and non-finite verb contexts.

Figure 9.5 shows the cumulative effect of accessibility for Time 1, with subjects categorized separately by finite and non-finite contexts. As referents become more inaccessible, the children tend to use fewer omitted subjects across both finite and non-finite verbal contexts. They also use more lexical forms as inaccessibility increases. To test this statistically, we performed a $\chi^2$ analysis comparing the number of omissions for the most accessible subjects (inaccessibility score 0–1) vs. the least accessible subjects (inaccessibility score 5–6), separately for each of the finite contexts and non-finite contexts. The difference is significant for both contexts—finite: $\chi^2 = 10.274, p < .01$; non-finite: $\chi^2 = 13.518, p < .001$. Nonetheless, it is also clear that a greater proportion of omitted subjects realize accessible referents in non-finite contexts, in comparison with a greater proportion of pronouns realizing accessible referents in finite contexts.
Although there is a similar pattern evident at Time 2 (Figure 9.6) demonstrating the same cline of accessibility, the children are now generally using pronouns for accessible referents instead of omitting subjects in both finite and non-finite contexts. Therefore, it is not surprising that a \( \chi^2 \) analysis comparing the number of omissions for the most accessible subjects (INACCESSIBILITY SCORE 0–1) vs. the least accessible subjects (INACCESSIBILITY SCORE 5–6), separately for each of the finite contexts and non-finite contexts, shows no significant difference for either context: finite, \( \chi^2 = 0.915, p = .339 \); non-finite, \( \chi^2 = 0.732, p = .392 \). However, if omissions and pronouns are combined into a single category and compared to the two other overt forms (i.e. demonstratives and lexical NPs), a \( \chi^2 \) analysis now reveals a significant difference—finite: \( \chi^2 = 111.426, p < .0001 \); non-finite: \( \chi^2 = 6.713, p < .01 \). The somewhat more jagged pattern for non-finite contexts at Time 2 is due to the relatively smaller amount of data in this group (only 109 subjects).

Overall, a relationship between accessibility and finiteness is clear at Time 1 and somewhat less so at Time 2. This relationship is not expected from either the nativist or the usage-based perspectives (i.e. it contradicts Predictions N1, N2, U1, and U2). This finding provides additional support for the claim that finiteness and accessibility interact in providing an explanation for early subject omission.
In this chapter, we explored whether two apparently competing explanations for early non-target-like subject omission in English, as posited by two different theories, can rather be seen as interacting. We tested the predominant nativist motivation—verb finiteness—and the predominant usage-based motivation—referent accessibility—in a set of data from four English-speaking children aged 2 to 3 years. As expected, our findings for the individual explanations replicated those of many previous studies: significantly more subjects were omitted in the context of non-finite verbs, and significantly more subjects were omitted when their referents were highly accessible to the interlocutor. However, we also found a clear interaction between finiteness and accessibility of a sort that would not be predicted by either perspective. The nativist account predicts that accessibility will have no influence on subject omission in non-finite contexts (Prediction N1) but may play a role in finite contexts (i.e. if we assume complementarity between the two perspectives). However, we found that a subject is more likely to be omitted when its referent is accessible than when it is inaccessible both in finite and non-finite contexts.

**Figure 9.6** Cumulative effect of accessibility for six features distributed across four forms for children at Time 2 in the context of verb finiteness

### 9.4 Discussion and conclusions

In this chapter, we explored whether two apparently competing explanations for early non-target-like subject omission in English, as posited by two different theories, can rather be seen as interacting. We tested the predominant nativist motivation—verb finiteness—and the predominant usage-based motivation—referent accessibility—in a set of data from four English-speaking children aged 2 to 3 years. As expected, our findings for the individual explanations replicated those of many previous studies: significantly more subjects were omitted in the context of non-finite verbs, and significantly more subjects were omitted when their referents were highly accessible to the interlocutor. However, we also found a clear interaction between finiteness and accessibility of a sort that would not be predicted by either perspective. The nativist account predicts that accessibility will have no influence on subject omission in non-finite contexts (Prediction N1) but may play a role in finite contexts (i.e. if we assume complementarity between the two perspectives). However, we found that a subject is more likely to be omitted when its referent is accessible than when it is inaccessible both in finite and non-finite contexts.
contexts. The usage-based account predicts that finiteness should have no influence on subject omission at all (Prediction U2). However, we found that a subject is more likely to be omitted in a non-finite context than in a finite context both when it is maximally accessible and when it is maximally inaccessible. This means that both accessibility and finiteness play a crucial role in subject omission. These factors should not be seen as competing with each other for which one can provide the “true” explanation. Rather, they are each an important part of the story and must be treated in tandem.

It is also striking that accessible subjects occur more in non-finite contexts and inaccessible subjects occur more in finite contexts, regardless of the form in which the subjects appear. We consider here two possible explanations for this pattern in the data. One explanation is that verbal agreement inflections—found in finite but not non-finite verbs—help to track reference in discourse in the same way that more overt referential forms do. If this is the case, then high accessibility of a referent in the discourse context would more often result in the child omitting the inflection, whereas low accessibility of a referent would more often result in using an overt verbal inflection. A study of Inuktitut child data by Skarabela and Allen (2004) found results consistent with this hypothesis: verbal inflections were omitted significantly more often when the referent in question was jointly attended to by the child and the interlocutor than when it was not. However, this study took into account only one feature of accessibility and was conducted in a language with very different typology from English. This hypothesis remains to be tested empirically in data from English.

Another possible explanation is that some additional factor plays a mediating role—for example, that some property of non-finite verbs particularly attracts accessible subjects, which in turn tend to be omitted because of discourse-pragmatics. Further exploration of the event semantics of non-finite vs. finite clauses, as in studies by Hyams and her colleagues, may provide further insight here (e.g. Hoekstra and Hyams 1998; Hyams 2012). For example, children may omit tense and agreement markers with clauses of certain event types, and these particular event types may also tend to be expressed with accessible subjects. Then both finiteness and accessibility would be caused by a third variable, event type, which in turn would indirectly lead to omission of both tense/agreement and subjects. At any rate, the idea that a particular factor central to one theory is an indirect cause of subject omission mediated by a factor central to another theory is a new way of thinking about explaining subject omission.

Because our study used data from two points in development, it is relevant to mention developmental patterns. The most salient patterns are those already well-known and extensively discussed in the literature: the proportion of omitted subjects and the proportion of non-finite clauses both decrease significantly between 2;0 and 3;0 and disappear soon afterwards (Valian 1991; Wexler 1998; Freudenthal et al. 2007; Hyams 2011). Interestingly, we also find a decrease in the strength of the relationship.

---

6 We thank an anonymous reviewer for this suggestion.
between accessibility, finiteness, and subject omission between Time 1 and Time 2. For instance, we found a relationship between accessibility and subject omission at Time 1 but not at Time 2 (Section 9.3.2), and we found a difference in the proportion of finite vs. non-finite clauses associated with highly accessible vs. highly inaccessible referents at Time 1 but not at Time 2 (Sections 9.3.3 and 9.3.4). However, these results at Time 2 almost certainly occur because there are simply so few omitted subjects and non-finite verbs at Time 2 that there is not enough data to show any effects. We believe that there is no change in children’s sensitivity to the strength of these relationships over time; the relationships are just no longer relevant in the same way.

We turn finally to the roles played by each of the two theories in explaining children’s omitted subjects. As noted earlier, some nativist researchers have suggested a complementary role for the usage-based perspective alongside a main role for a nativist explanation. One clear statement to this effect is in Hyams (2011):

the grammar … allows the occurrence of a null element … under certain structural conditions of licensing and identification. But the syntax does not legislate when a particular subject will be omitted. This is a function of the information structure (IS) of the sentence…. Conversely, discourse conditions alone cannot sanction missing arguments. (Hyams 2011: 40)

However, our study shows that this complementarity explanation does not go far enough. First, as found in many previous studies, subjects are not only omitted in the grammatically licensed context of non-finite verbs; they are regularly omitted in finite contexts as well. In our study, 27% of omitted subjects at Time 1 and 24% at Time 2 are in finite contexts, which are not licensed locations for subject omission in the current grammatical accounts. Second, our results indicate that finiteness and accessibility crucially interact with each other. Accessibility influences subject omission similarly in both the non-finite context that is grammatically licensed for subject omission and in the finite context that is not grammatically licensed for subject omission. And finiteness influences subject omission similarly both when the referent of the subject is accessible and when it is not. Finally, our results show that accessibility may mediate the effect of finiteness, or vice versa, such that the effect of one feature is indirect rather than direct.

In sum, the findings in this chapter show that there is a rich interaction between nativist and usage-based accounts of language acquisition. This is very different from a competition approach that focuses on trying to prove which one account is the best, and also from a complementarity approach that allows a small role for one theory to explain a small subset of data not explainable by another (more preferable) theory. We take the position that, at least in the data we explored here, both theories are essential for an understanding of the patterns in the full set of data. By taking a more holistic perspective of looking together at the patterns predicted and revealed by the different theoretical accounts, we come to a deeper understanding of the intricate interplay between the various factors relevant to language development. This study reveals that this interaction between theoretical perspectives is a rich area of research that will increase a general understanding of the language acquisition process and the connection between grammatical and cognitive development. We hope that this
study will provide a model for further similar studies and serve as a step towards the larger goal of establishing a unified theory of language development.

Acknowledgments

We thank Elena Lieven and Jeannine Goh for providing us with transcripts and videotapes from the Manchester Dense Database. This research was funded by National Science Foundation grant BCS-0346841 to S. Allen.