A Discourse-Pragmatic Analysis of Subject Omission in Child English

Shanley Allen, mary hughes

A Discourse-Pragmatic Analysis of Subject Omission in Child English

Mary Hughes and Shanley Allen
Boston University

1. Introduction

Young children (approximately 1;6 to 4;0 years) acquiring a first language often omit arguments (subjects and objects) whether the target language allows it or not. This phenomenon has been investigated from three main theoretical perspectives: the grammatical account, the processing account, and the discourse-pragmatic account. The present study investigates the role that discourse-pragmatics plays in subject omission in a non-null subject language and tests the hypothesis that children produce or omit arguments as a result of their awareness of the complexities of information flow in discourse.

At present, much of the research on subject omission has been conducted from the grammatical perspective, which assumes that child syntax differs intrinsically from adult syntax (e.g., Radford, 1990; Hyams & Wexler, 1993; Hoekstra & Hyams, 1998; Hyams, 2001). The grammatical account argues that the child's developing language system is incomplete in terms of syntactic knowledge and therefore allows null subjects as an option. Research from the processing account suggests that children have the same internal grammatical system as adults, but due to limited processing abilities, they often omit subjects, especially in longer utterances (e.g., Bloom, 1990; Gerken, 1991; Valian, 1991; Valian, 2005). As children's cognitive resources develop, they produce overt subjects more frequently. Both of these accounts explain many aspects of the phenomenon; however, neither account is able to adequately address the variation that exists in children's production of subjects, something which the discourse-pragmatic account handles very well.

The discourse-pragmatic account proposes that children are aware of information flow to some extent, and their production or omission of arguments reflects this awareness. At some level, children understand that producing an argument is more informative than not doing so – the goal being to communicate more effectively. Children acquiring a language tend to omit arguments more often than adults, but they eventually achieve the same rate of omission. Research from the discourse-pragmatic perspective has demonstrated that discourse-pragmatic features do play a role in argument realization; however, most of this research centers on languages that allow null arguments (e.g., Korean: Clancy, 1997; Inuktitut: Allen, 2000; Italian: Serratrice, 2005).

To date, no thorough or systematic research has focused solely on subject omission for monolinguals acquiring a non-null subject language investigating a full set of discourse-pragmatic features.[1]

Although previous studies have demonstrated that certain discourse-pragmatic features, such as NEWNESS and ABSENCE, play a role in determining whether subjects are realized in null subject languages, it is yet to be determined whether the same features are relevant in non-null subject languages. Moreover, investigating discourse-pragmatics in a non-null subject language offers a different challenge due to a typological constraint requiring subjects. Because children have a greater overall tendency to realize subjects in non-null subject languages, it is more difficult to determine which features, if any, have an effect on subject realization.

In null subject languages, extremely high rates of subject omission have been reported for young children. For example, in Inuktitut, the rate of null subjects has been reported to be as high as 90% (Allen & Schröder, 2003), and in Italian, as high as 70% (Valian, 1991). Although these rates are somewhat higher than reported omission rates in Adult Inuktitut and Adult Italian,[2] it is clear that children in these two languages are sensitive to adult input and to the grammatical constraints of the language that they are acquiring; therefore, they omit subjects quite freely.

On the other hand, subject omission rates for children acquiring English are much lower, about 20%-30%. Again, this is somewhat higher than the adult rate,[3] but it is evident that as children are acquiring the grammar of English, they are attempting to adhere to a grammatical constraint that requires subjects. Therefore, it seems that children acquiring a non-null subject language, such as English, will produce a high proportion of overt subjects. How then can it be determined which subjects are produced because of a typological constraint requiring subjects and which are produced because of discourse-pragmatic constraints?

2. The Present Study

The present study investigates the role of discourse-pragmatics in subject omission in a non-null subject language by analyzing the utterances of a two-year-old monolingual English speaker. The central research questions are:

1. Do the same discourse-pragmatic features that affect subject realization in a null subject language have an effect in a non-null subject language?

2. How do these features interact with one another?

In order to determine the effect that discourse-pragmatic features have on subject realization in a non-null subject language, it is necessary to establish a baseline. This baseline will reveal the percentage of overt subjects that occur because of the grammatical constraint requiring subjects without regard to

discourse-pragmatics. The process of establishing this baseline will be outlined in detail throughout the remainder of this paper.

3. Methodology

This study analyzes the utterances of Annie, a two-year-old monolingual English speaker from Manchester, England. The transcripts and videotapes for Annie were made available by Elena Lieven of the Max Planck Child Study Center at the University of Manchester (Lieven, et al., 2003). Annie was videotaped in weekly hour-long sessions in her home interacting with her mother for approximately seven weeks. These sessions began when she was two years and 4 days (2;0.4) and ended at two years, 1 month and 22 days (2;1.22). Her MLU range was from 2.15 to 2.45.

All fully intelligible non-imperative utterances containing a verb were coded for sentence-level, grammatical, and discourse-pragmatic features. Subject arguments only were chosen for this analysis. There were very few null objects in our data sample.[4] Therefore, in order to get the most accurate picture, objects were excluded from the analysis. Subject arguments occurring in imperative utterances, exact self-repetition, imitation, recitation of poems or songs, frozen forms, and routines were also excluded.

The remaining 613 subject arguments were coded for discourse-pragmatic information using six **accessibility features** (see Table 1).[5] Each argument was assigned a binary value for each feature based on how accessible a referent was in the discourse context (Clancy 1993, 1997; Allen, 2000, 2005; Skarabela, 2005).[6] An **accessible value** indicates that the referent is easily identified in the discourse context, so the speaker may be *less* explicit, meaning that he or she may be more likely to use a null subject. An **inaccessible value** indicates that

Table 1: Accessibility Features

Accessibility Feature	Accessible Value	Inaccessible Value
THIRD PERSON	First and Second Person	Third Person
INANIMACY	Animate	Inanimate
ABSENCE	Physically present	Physically absent
NEWNESS	Given; mentioned within 20 utterances	New; not mentioned within 20 utterances
DIFFERENTIATION IN CONTEXT	Only referent in physical context	Multiple referents in physical context
DIFFERENTIATION IN DISCOURSE	No other possible referents in preceding 5 utterances	Other possible referents in preceding 5 utterances

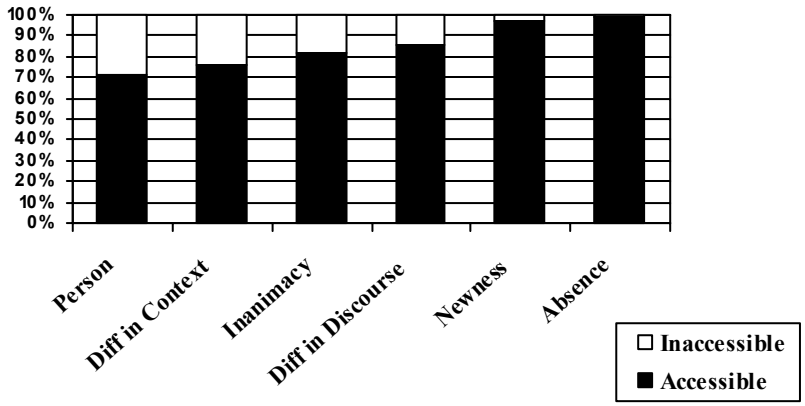
the referent is *not* easily identified in the discourse context, so the speaker must be *more* explicit, meaning that he or she may be more likely to produce an overt argument.

Each subject was coded as either accessible or inaccessible for each of the six features. For example, a child might produce the sentence "ø falling over" about a toy that she is playing with. Because it is more difficult to determine the referent of a third person argument (for which the search space is potentially large) than it is to determine the referent of a first or second person argument (for which the search space is restricted to the speaker and the listener), the null third subject in this example is coded as inaccessible for the feature THIRD PERSON. However, the toy that is the target referent for the argument is physically present and therefore more easily identified than if it were absent from the physical context. The null subject is then coded as accessible for the feature ABSENCE. All arguments were coded in this manner for each accessibility feature.

4. Findings

After the 613 subject arguments were recoded into binary values (accessible, inaccessible), two-way contingency table analyses were conducted for each of the six predictors with *form* as the dependent variable. Chi-square tests of independence were then performed in order to determine whether there was a statistically significant relationship between each of the predictors (the six discourse-pragmatic features) and the form of the subject.

Figure 1 demonstrates the results of the analyses for the 148 null subjects in terms of each feature. In other words, each column represents all 148 null subjects for that particular feature. In the first column for person, 71% of the 148 null subjects are accessible, while 29% are inaccessible. This result supports the hypothesis that discourse-pragmatic features influence subject realization because a much higher proportion of the null arguments are accessible than are inaccessible.[7] Since accessible referents are easily identifiable in the discourse context, their overt realization adds relatively little information to the discourse. This result is repeated for the other five features with percentages as high as 97% and 99% accessible for NEWNESS and ABSENCE. In fact, all 148 null subjects are accessible for each feature at a rate of 71% or better. These results are promising and indicate that the discourse-pragmatic features under investigation do influence subject realization in this data set.



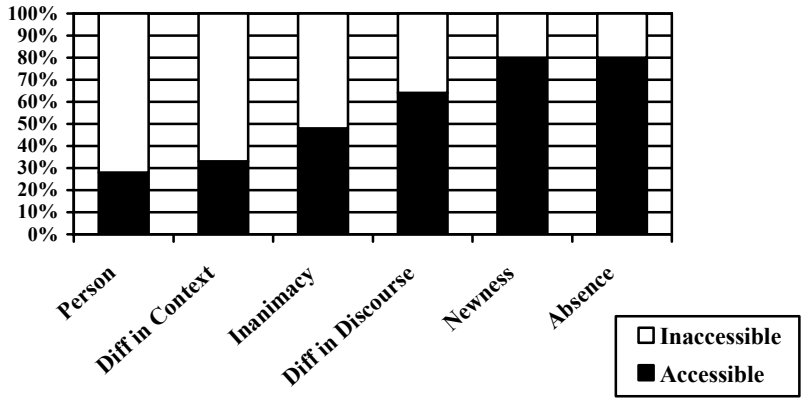
**Chi-square, p <.01 for each feature*

Figure 1: Percentage of 148 null subjects that are accessible vs. inaccessible for each feature

However, results for the 465 overt subjects in Figure 2 look quite different. Now, each column represents all 465 overt subjects for each feature. In this case, the hypothesis predicts that overt subjects are more likely to be inaccessible for one or more features, meaning that their referents are not easily identifiable in the discourse context. For example, in the first column for PERSON, 72% of the 465 overt subjects are inaccessible, while 28% are accessible. In column two, the results for the feature DIFFERENTIATION IN CONTEXT show that 67% of the overt subjects are inaccessible for this feature, and 33% are accessible. The results for these two features are generally in line with the hypothesis that discourse-pragmatic features influence subject realization. However, the remaining four features show a different pattern. For the feature, INANIMACY, the 465 overt subjects are almost equally accessible and inaccessible. Then, for the last three features, DIFFERENTIATION IN DISCOURSE, NEWNESS, and ABSENCE, the number of accessible overt subjects increases over inaccessible overt subjects.

It is interesting to note that these six features overlap; therefore, the results displayed in this chart do not necessarily refute the prediction that overt subjects will have inaccessible values for one or more features. Although for three of the six features there are more overt accessible subjects than overt inaccessible subjects, this reveals nothing about whether these subjects have inaccessible values for any other feature. For example, subjects that are accessible for ABSENCE in the last column may, in fact, be inaccessible for one or more of the other five features. In order to tease out the influence that discourse-pragmatic features actually have, the cumulative effect of the six overlapping

features must be determined. This will also help to establish an accurate baseline for overt subjects with no inaccessible values.



**Chi-square, p <.01 for each feature*

Figure 2: Percentage of 465 overt subjects that are accessible vs. inaccessible for each feature

In order to do this, the six predictor variables were summed for all 613 subjects, assigning each a value of 0 through 6. Figure 3 represents all 613 subjects grouped by the sum of the accessibility values for each feature. The first column represents the 226 subject arguments that are accessible for all six features; the referents for these arguments are highly salient and identifiable in the discourse context so that their overt realization adds virtually no information.

In this context, 57% of these 226 accessible arguments are overtly realized. This then could be called the baseline for overt subject realization for this English-acquiring child, meaning that at least 57% of all subjects will be realized with no regard to discourse-pragmatics. If this is the baseline, then most other summed categories clearly surpass it, indicating that the addition of two or more inaccessible values for a feature, regardless of what that feature is, will have an effect on subject realization. Figure 3 gives evidence of discourse-pragmatic effects: the greater the number of inaccessible features, the greater the effect, so that in the last column, when all six discourse-pragmatic features have an inaccessible value, subjects are overtly realized at 100%.

However, there is a problem inherent in this representation and the baseline assumed here. A closer look at the 226 subjects that have 0 inaccessible features reveals that they are all first and second person. This is quite predictable since THIRD PERSON is an inaccessibility feature itself. Furthermore, 10 of the 23 subjects that have one inaccessible feature are first and second person. All of the other 377 subjects are third person.

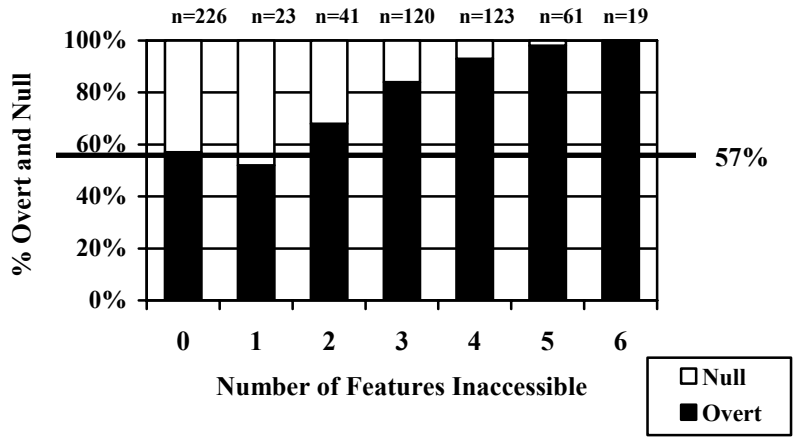


Figure 3: Cumulative effect of accessibility values for six features for 613 subject arguments

A closer look at first and second person reveals important distinctions between first/second and third person subjects. While third person subjects can have inaccessible values for all six features, first person cannot have inaccessible values for any, and second person can have inaccessible values for only two of the six features: DIFFERENTIATION IN CONTEXT and DIFFERENTIATION IN DISCOURSE. Calling THIRD PERSON a feature, in and of itself, may have skewed the results somewhat. Establishing a baseline including first and second person subjects in order to set the threshold for third person subjects may be misleading. Therefore, first and second person were removed from the analysis in order to establish a more accurate baseline.

The remaining 377 third person subjects were then reanalyzed. This time the five remaining predictor variables were summed for all 377 arguments, assigning each a value of 0 through 5. Figure 4 represents all 377 subjects grouped by the sum of the accessibility values for each feature. The first column represents 13 subjects that are accessible for all five features. In this context, 85% of the 13 accessible subject arguments are overtly realized. The baseline established earlier when first and second person were included has risen substantially for third person only. It appears that for this child the threshold for producing overt third person subjects – whether they are accessible or not – is 85%. The second category, representing third person subjects with one inaccessible feature active, does not approach this baseline, and the third category, with two inaccessible features, just approaches it at 84%. It is not until three, four, and five features are inaccessible that this baseline is surpassed.

Again, the more features that have an inaccessible value, the greater the effect, so that subjects with five inaccessible features are realized overtly at 100%.

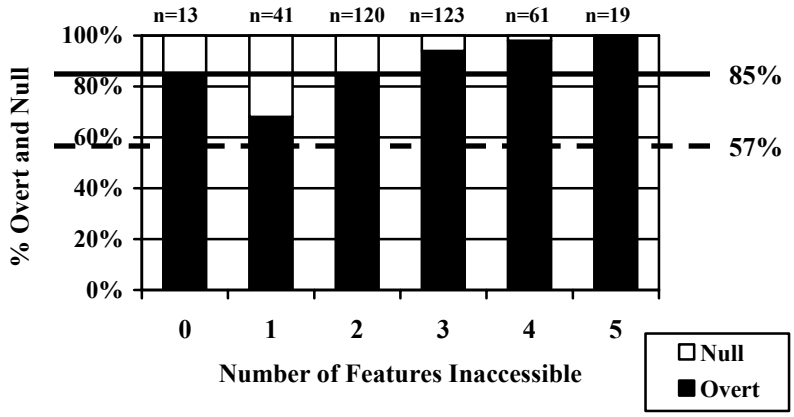


Figure 4: Cumulative effect of informative values for five features for 377 third person subject arguments

5. Discussion

With the exclusion of first and second person, the established baseline demonstrates that certain discourse-pragmatic features do have an effect on subject realization, but how these five features interact with one another is not completely clear. It is often difficult to tell which features have the greatest effect on subject realization because they overlap and appear to work in combination. Interestingly, they are not evenly distributed throughout the combinations of features.

Two in particular, ABSENCE and NEWNESS, do not appear in significant numbers until they are combined with two, three, or four other features (see columns 3, 4, and 5 in Figure 4). Referents that are physically absent and referents that are new to the discourse only begin to appear in substantial numbers when the percentage of overt subjects is over 90%. In fact, all 43 third person null subjects are physically present and only five of them are new to the discourse.

Therefore, it seems that these features are much stronger predictors than the others. ABSENCE and NEWNESS are so powerful that when referents were marked as inaccessible for these features (new and/or absent), the subject arguments were almost always overt. The other three features – DIFFERENTIATION IN CONTEXT, DIFFERENTIATION IN DISCOURSE,

and INANIMACY – are less predictive, but may have a combined effect. Finally, the feature PERSON does not appear to operate in the same way as the other five discourse-pragmatic features, which indicates that first/second person subjects should be analyzed separately from third person subjects.

Certain versions of both the grammatical account and the processing account suggest that the target for null subjects in the adult language would be a pronoun. What evidence does this study offer to support or refute that claim? First, 105 out of 148 null subjects are first and second person, and so would be realized overtly as pronouns. All first and second person referents are considered given and physically present. Moreover, all of the remaining 43 null third person subjects are physically present, and 38 represent given information. The five that represent new information are jointly attended to by the child and her interlocutor. Therefore, it is very likely that the 43 third person null subjects would be realized as pronouns or demonstratives in the adult language. This child appears to be overgeneralizing the use of null subjects when the adult target form would be an overt pronoun or a demonstrative. Ongoing research also suggests that this kind of overgeneralization occurs for this child in the use of pronouns when a demonstrative would be the adult target.[8]

6. Conclusions

The attempt to establish a baseline for this data set indicates first and second person should be analyzed separately from third person. With this baseline established, it becomes clear that certain discourse-pragmatic features do have predictive value in determining subject realization in a non-null subject language, especially in predicting which subjects will be omitted. However, predicting which subjects will be overt is more problematic because accessible subjects are often overt due to a typological constraint requiring subjects in English. Children appear to adhere to this constraint very early on in the language acquisition process.

In order to test these findings, this process needs to be applied to a larger data set with multiple children in order to see whether the baselines established here hold true. Moreover, further statistical analyses, such as logistic regression, should be performed on the larger data set to determine which discourse features have the most powerful effect on argument realization in English. We are now in the process of conducting such a study.

Endnotes:

* This research was funded by NSF grant #BCS-0346841. Special thanks to Elena Lieven of the Max Planck Child Study Center (MPCSC) at the University of Manchester who provided the data and to Annie and her family. Also many thanks to colleagues and students who have contributed to this research: Cathy O'Connor, Paul Hagstrom, Barbora Skarabela, Marina Rodichenko, Nicole Acrey, Joseph Blass, Quinn Duffy, Alessandra Melotti, Maria LaMendola,

Heather Caunt, and Sandy Lord. Thanks also to James Leonard for his help in editing the manuscript.

1. Guerriero et al. (2001) investigated Preferred Argument Structure in English, but only looked at one discourse-pragmatic feature (NEWNESS).

2. Although adult rates for Inuktitut and Italian have not yet been studied systematically, the data in Allen & Schröder (2003) suggest that while children omit subjects in Inuktitut at a rate of about 90%, adults tend to omit subjects at a lower rate of about 80%. Valian (1991) demonstrates similar findings for adult Italian: children omit subjects at a rate of around 70%, while adults omit subjects at a rate of approximately 50%.

3. Although English has a constraint against subject omission, it is widely observed that adults do drop subjects in certain contexts. Data collected by Valian (1991) suggest that English-speaking children omit subjects at a rate of 20% to 30%, while adults omit subjects at a rate of 10% to 15%.

4. Out of 265 object arguments, only 13 were null (5%). These results are in line with previous research demonstrating a subject-object asymmetry terms of null arguments for English-acquiring children.

5. Two other features, LINGUISTIC CONTRAST and QUERY, were found to have an effect on argument realization in previous work in discourse-pragmatics (e.g., Clancy 1997; Allen 2000). However, in this data set, these features appear to operate as categorical variables. The features LINGUISTIC CONTRAST and QUERY are activated in combination with other discourse-pragmatic features or alone, but whenever they are, the subject argument is always realized. In other words, an **accessible** value for these two features overrides all of the other six features. Researchers Vallduví and Vilkuna (1998) label the type of contrast that we are investigating here as **k-kontrast** and claim that in English, contrasts are usually "accentually prominent." They go on to state that WH-words are essentially contrastive because they generate and quantify over a potential WH-set, and the answer to a WH-question picks out a member of that set in the same way that contrastive arguments do. Because LINGUISTIC CONTRAST and QUERY appear to operate as categorical variables in this data set and because this is essentially a study of variation, all utterances with positive values for contrast or query have been removed from the analysis.

6. The **accessibility features** described here are a reformulation of the **discourse prominence** or **informativeness features** described in previous work on discourse-pragmatics and argument realization (e.g., Clancy, 1993, 1997; Allen, 2000). The term *accessibility* is based on the work of Ariel (2001) and was first used to describe discourse-pragmatic features by Skarabela (2005).

7. There are two potential ways to read these data. In this discussion, we are pointing out the fact that there are more accessible than inaccessible referents for null subjects. However, it is equally true that a significantly higher proportion of null than overt subjects have accessible referents, which is indicated by statistical significance (**Chi-square, p <.01 for each feature*) when considering the results in Figures 1 and 2 together.

8. The idea that children may be overgeneralizing referential forms is supported by other findings in the data. Examples occur in which the child points at a new object and utters something like, 'I want it.' In this case, a demonstrative would be the more appropriate referential choice in the adult language. This could suggest that children do have the beginnings of a referential system in place quite early on, but that this system is not yet adult-like, causing children to overgeneralize referential categories.

References:

Allen, Shanley E.M. (2000). A discourse-pragmatic explanation for argument representation in child Inuktitut. *Linguistics, 38,* 483-521.

Allen, Shanley E.M. (2005). Is grammar all in the mind? The case of omitted subjects in child speech. Presentation for the Department of Psychology at the University of Waterloo, Ontario, Canada.

Allen, Shanley E.M. & Schröder, Heike. (2003). Preferred argument structure in early Inuktitut spontaneous speech data. In J.W. Du Bois, L.E. Kumpf and W.J. Ashby (Eds.), *Preferred argument structure: Grammar as architecture for function* (pp. 301-338). Amsterdam: John Benjamins.

Ariel, Mira. (2001). Accessibility theory: An overview. In T.J.M. Sanders, J. Schilperoord and W. Spooren (Eds.), *Text Representation* (pp. 29-87). Amsterdam: John Benjamins.

Bloom, Paul. (1990). Subjectless sentences in child language. *Linguistic Inquiry, 21,* 491-504.

Clancy, Patricia M. (1993). Preferred argument structure in Korean acquisition. In E.V. Clark (Ed.), *Proceedings of 25th annual Child Language Research Forum* (pp. 307-314). Stanford, CA: CSLI Publications.

Clancy, Patricia M. (1997). Discourse motivations for referential choice in Korean acquisition. In H. Sohn & J. Haig (Eds.), *Japanese/Korean Linguistics, 6* (pp. 639-659). Stanford, CA: CSLI Publications.

Gerken, LouAnn. (1991). The metrical basis for children's subjectless sentences. *Journal of Memory and Language, 30,* 431-451.

Guerriero, A.M. Sonia, Cooper, Amy, Oshima-Takane, Yuriko & Kuriyama, Yoko. (2001). A discourse-pragmatic explanation for argument realization and omission in English and Japanese children's speech. In A. Do, L. Dominguez, A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 319-330). Somerville, MA: Cascadilla Press.

Hoekstra, Teun & Hyams, Nina. (1998). Aspects of root infinitives. *Lingua 106,* 91-112.

Hyams, Nina. (2001). Now you hear it, now you don't: The nature of optionality in child grammars. In A. Do, L. Dominguez, A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 34-58). Somerville, MA: Cascadilla Press.

Hyams, Nina & Wexler, Kenneth. (1993). On the grammatical basis of null subjects in child language. *Linguistic Inquiry, 24,* 421-459.

Lieven, Elena, Behrens, Heike, Speares, Jennifer & Tomasello, Michael. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language, 30,* 333-70.

Radford, Andrew. (1990). *Syntactic theory and the acquisition of English syntax: The nature of early child grammars in English*. Oxford: Blackwell.

Skarabela, Barbora. (2005). *The role of* s*ocial cognition in early language: The case of joint attention in argument realization in child Inuktitut.* Unpublished doctoral dissertation, Boston University, Boston, MA.

Serratrice, Ludovica. (2005). The role of discourse pragmatics in the acquisition of subjects in Italian. *Applied Psycholinguistics*, *26,* 437-462.

Valian, Virginia. (1991). Syntactic subjects in the early speech of American and Italian Children. *Cognition, 40,* 21-81.

Valian, Virginia & Aubry, Stephanie. (2005). When opportunity knocks twice: Two-year-olds' repetition of sentence subjects. *Journal of Child Language, 32,* 617-41.

Vallduví, Enric & Vilkuna, Maria. (1998). On rheme and kontrast. In P. Culicover and L. McNally (Eds.), *Syntax and Semantics 29: The Limits of Syntax* (pp. 79-108). New York: Academic Press.