A discourse-pragmatic explanation for argument representation in child Inuktitut*

SHANLEY E. M. ALLEN

Abstract

This paper assesses discourse pragmatics as a potential explanation for the production and omission of arguments in early child language. It employs a set of features that characterize typical situations of informativeness (Greenfield and Smith 1976; Clancy 1993, 1997) to examine argument status in data from four children aged 2;6 through 3;6 learning Inuktitut as a first language. Results based on logistic regression analyses suggest that a discourse-pragmatic account of argument representation has good explanatory adequacy, and that several of the features characterizing informativeness are good indicators of those arguments that tend to be overtly produced rather than omitted in early child language.

1. Introduction

Children at the early stages of language learning (two-word stage to 4;0; peak before 2;6 in English) tend to omit arguments in their utterances, regardless of whether the target language permits this or not. Examples below from English, which does not normally permit omitted arguments, show utterances with an omitted subject, (1a), an omitted object, (1b), and both subject and object omitted, (1c) (data from Bloom 1970, 1973).

(1) a. (Kathryn 1;10)
   No like celery.
   (telling her mother that she doesn’t like celery)

b. (Allison 1;8)
   Mommy, you wiping.
   (wanting her mother to wipe a doll)

This phenomenon occurs regardless of whether the target language uniformly requires overt arguments as in English (Bloom 1970; Hyams 1986; Bloom 1990; Valian 1991; Wang et al. 1992) and Danish (Hømann and Plunkett i.p.), or permits omitted arguments in topic-drop contexts as in Dutch (de Haan and Tuinman 1988; Krämer 1995) and German (Hømann 1995), or uniformly permits omitted arguments as in Chinese (Wang et al. 1992), Inuktitut (Allen and Schröder i.p.), Japanese (Mazuka et al. 1986; Hirakawa 1993), Korean (Clancy 1993, 1997, i.p.), and Mauritian Creole (Adom 1994). Additionally, subjects tend to be omitted more frequently than objects cross-linguistically.

Three major types of explanation have been put forth to explain argument ellipsis in child language: grammatical, performance, and discourse-pragmatic accounts. Grammatical accounts assume that the child starts out with a grammar that permits (or forces) argument ellipsis in certain well-defined instances and later changes this grammar to one more appropriate to the adult language (if the adult language is different). For instance, Hyams’s (1986) pro-drop parameter account claims that all children are born with the pro-drop parameter in a default setting permitting the omission of subjects, as in Italian and Spanish, and thus subject omission is licensed in early child language. Radford’s (1990) functional-category account claims that children at the early stages of language development have only lexical categories (e.g. NP, VP), and not functional categories (e.g. IP, CP). Thus, subject omission is obligatory at the earliest stages since children do not have the (functional) SPEC-of-IP position to put a subject in. However, the object that appears in the lexical daughter-of-VP position has no reason to be omitted. Topical-drop accounts (de Haan and Tuinman 1988; Hømann 1995; Wang et al. 1992; Hyams and Wexler 1993) assume that children’s early grammars are set to allow omission of an argument when it is the discourse topic, as is allowed in adult Dutch and German. In each of these grammatical accounts, the child’s grammar either matures or develops to the correct adult setting for the language he or she is learning sometime between the ages of 2 and 4, and non-adult-like argument omission disappears from the child’s production. While the first two of these accounts suggest that children will pass through an early stage in which all subjects are omitted, the topic-drop account attempts to explain the robust finding that only a certain percentage of subjects (and objects) are omitted beyond the two-word stage. However, all these grammatical accounts incorrectly
predict a sudden shift between a first stage at which (some or all) subjects are omitted, and a second stage at which no arguments are omitted.

Performance accounts typically assume that the child has adult-like grammatical structure from the earliest stages of language learning but has difficulty in producing all that he or she is capable of due to performance limitations. The most common of the performance accounts appeals to the young child’s limited sentence-processing abilities, which cause him or her to omit certain words in the sentence (Bloom 1970; Mazuka et al. 1986; Bloom 1990; Valian 1991). Since the processing load of a sentence is assumed to be greater at the beginning of the sentence, where the subject tends to be, than at the end of the sentence, where the object tends to be, the subject is more often dropped than the object. A second type of performance account involves prosody andmetrical structure (Gerken 1990). By this account, the subject tends to appear in a prosodically and metrically weaker position than the object, and this is omitted more often by children. Performance accounts tend to predict a gradually decreasing percentage of omission of arguments rather than an early stage of uniform omission followed by a sudden shift to a stage of uniform lack of omission.

Finally, discourse-pragmatic accounts assume that the child’s non-adult-like production of null arguments stems from an early (hyper)sensitivity to the informational structure of events, and thus to pragmatic features of discourse. Greenfield and Smith’s (1976) principle of informativeness (different from the Gricean principle of informativeness) suggests that children tend to encode those aspects of the event that are most informative: “where they see alternatives, where there is uncertainty in terms of the situational structure” (1976: 64). On the other hand, children tend not to encode those aspects of the event that are presupposed, such as the subject and the agent. They state that the subject is “often taken for granted” and typically “less informative than other constituents of the sentence that resist ‘deletion’” (1976: 223), while the agent “is only spoken under unusual conditions, such as a conflict about agency or an actual change of agent” (1976: 184). Following this same line of reasoning, several authors suggest that newness of the referent is the most relevant feature of informativeness, and thus that arguments representing given referents will be omitted (e.g. Bloom 1990; Valian 1991; Hirakawa 1993; though see Hamann and Plunkett i.p.). The subject more often tends to represent given information while the object tends to represent new information, and therefore the subject is omitted more frequently than the object. Clancy (1993, 1997, i.p.) has developed this argumentation much more thoroughly, considering newness as well as three other features of discourse prominence (absence, contrast, query) and person and animacy, in a large database of Korean child speech. Her findings indicate that arguments representing first or second person referents are omitted more frequently than those representing third person referents, that arguments representing animate referents are omitted more frequently than those representing inanimate referents, and that arguments representing referents that do not show absence, contrast, newness, or query are omitted more frequently than their discourse-prominent counterparts.

The present paper uses a discourse-pragmatic approach to analyze argument representation in spontaneous speech data from four Inuit-speaking children. Previous research has shown that Inuit children aged 2.0–3.6 produce a quite low percentage of overt arguments (18.4%), thus omitting most of the arguments in their speech (81.6%) (Allen and Schröder 1984). Since Inuktitut permits argument omission in the target language (presumably due to licensing by the verbal cross-referencing inflection), child argument omission in Inuktitut is not the unusual pattern that it is in English and other languages. Nonetheless, it is interesting to explore the patterns of argument representation in child speech in this language. In fact, in a language that permits rampant argument omission, it is particularly interesting to wonder what causes Inuit children to produce any arguments at all. Here the notion of informativeness provides an intriguing line of investigation, with the hypothesis that children will produce as overt only those arguments that represent informative referents. Thus, the goal of this paper is to determine whether a discourse-pragmatic account can explain the observed patterns of argument representation and ellipsis in Inuktitut. Results indicate that several features of informativeness are indeed good predictors of whether an argument will be overt or not in Inuktitut child language. We expect that similar factors are at work cross-linguistically, and thus that the results found here will be insightful for research on argument ellipsis in other languages, including those in which argument ellipsis is not permitted in the target language.

2. Features of informativeness

This paper tests the hypothesis that children are highly sensitive to the dynamics of information flow in discourse, and that they structure their conversation in order to reduce the potential uncertainty of the listener regarding the referents that they are talking about. Thus, they will tend to omit arguments when the referent of the argument is maximally clear from the discourse and situational context, and they will tend to refrain
from omitting arguments when the referent of the argument is in doubt for any reason. In this paper we consider a set of eight features that have been shown to influence argument representation in other languages (primarily taken from Clancy 1993, 1997; also from Greenfield and Smith 1976; Clancy 1980; Givón 1983; Chafe 1987; and Du Bois 1987). We refer to these features as informativeness features, since they determine how informative the speaker should be in representing in speech the referent at hand. We treat these features as binary, such that they have one value that makes the identity of the referent less certain (e.g. new), and a second value that makes the identity of the referent more certain (e.g. given). We refer to the first value as the informative value since it requires the child to be informative in the way he or she talks about this referent. Similarly, we refer to the second value as the uninformative value since it does not require the child to be as informative. The eight features are detailed in the paragraphs below. Each feature is named for the binary value of the feature that engenders uncertainty of referent and thus requires informativeness. We divide the features into three groups: those that concern the presence in joint knowledge of a referent (knowledge features), those that concern potential confusion about the identity of a referent (confusion features), and those that concern the search space for the referent (search-space features).

2.1. Knowledge features

Knowledge features concern the presence of the referent in the joint knowledge of the speaker and the hearer, whether that knowledge derives from the physical or mental context. They include three features: ABSENCE, NEWNESS, and QUERY.

Absence. The feature ABSENCE characterizes a referent that is not present in the physical context of the conversation. Since the hearer does not have knowledge of the referent from the physical context of the discourse, the identity of the referent is much less certain than it would be were the referent present in the physical context.

Newness. The feature NEWNESS characterizes a referent that has not been previously talked about in the conversation at hand. Since the hearer has no mental knowledge of a new referent, its identity is much less certain than it would be were the referent already given in discourse. We consider an argument to be new if the referent it denotes has not been mentioned in the preceding 20 utterances, and to be non-new (given)

if the referent has been mentioned one or more times in the preceding 20 utterances. First and second person arguments are all classified as non-new, following Chafe (1976) and Du Bois (1987).

Query. The feature QUERY characterizes a referent that is the subject of or response to a question. Since the referent is either not yet identified or newly identified, the listener has little mental knowledge of this referent, and thus its identity is much less certain than it would be were the referent already given in discourse.

2.2. Confusion features

Confusion features concern the resolution of potential confusion about the identity of a referent when various potential referents are present either explicitly or implicitly in the discourse or the physical context. They include three features: CONTRAST, DIFFERENTIATION IN CONTEXT and DIFFERENTIATION IN DISCOURSE.

Contrast. The feature CONTRAST characterizes a referent the speaker is explicitly contrasting against other potential referents in the discourse or in the shared physical or mental context, usually through tone of voice, gesture, or other contextual means. Such a situation typically occurs with contrasting candidate agents when a child wants to prohibit others from doing something he or she is doing, or when a child wants to do something someone else is doing. Since the child is conveying information about change in or maintenance of the current situation, the referent presented is new information to the hearer and thus constitutes a situation in which the child perceives potential uncertainty that must be prevented.

Differentiation in context. The feature DIFFERENTIATION IN CONTEXT characterizes a referent that is one of two or more referents in the immediate physical context (e.g. portion of the room where the child is directing his/her gaze) that could fit the verb semantics and the identifying elements of the argument in question (e.g. person, number). Since there is more than one potential referent in the physical context fitting the characteristics of the argument, there is potential uncertainty on the part of the hearer in identifying the target referent.

Differentiation in discourse. The feature DIFFERENTIATION IN DISCOURSE characterizes a referent that is one of two or more referents already established in the discourse (i.e. in the five preceding utterances,
following Givón (1983) that could fit the verb semantics and identifying elements of the argument in question (e.g., person, number). Since there is more than one potential referent in the discourse context fitting the characteristics of the argument, there is potential uncertainty on the part of the hearer in identifying the target referent.

2.3. Search-space features

Search-space features concern differences in the relative size of the search space one must consider to find the referent in question. They include two features: INANIMACY and THIRD PERSON.

Inanimacy. The feature INANIMACY characterizes referents that are not animate (i.e., not humans or animals). In typical child discourse, the number of animate entities is relatively limited (e.g., child, mother, father, sibling, dog) compared to the vast number of inanimate entities (e.g., table, cup, toy, juice, television, plant, clothes). Thus, the search space for animate referents is relatively small, while the search space for inanimate referents is relatively much larger. If the listener knows nothing about the referent except whether it is animate or inanimate, the identity of an inanimate referent will be much less certain than that of an animate referent.

Third person. The feature THIRD PERSON characterizes a referent that is not first or second person (i.e., not ‘I’, ‘me’, ‘we’, ‘us’, or ‘you’). In typical child discourse (and all other discourse), the number of first and second person entities is relatively limited compared to the vast number of potential third person entities. Thus, the search space for first and second person referents is relatively small, but the search space for third person referents is relatively much larger. If the listener knows nothing about the referent except whether it is first/second or third person, the identity of a third person referent will be much less certain than that of a first or second person referent.

These eight features, then, are employed in the remainder of the paper as features that characterize informativeness. The informative and uninformative values for each of the features (in alphabetical order) are summarized in Table 1.

3. Structure of Inuktitut

Inuktitut is a language of the Eskimo-Aleut family spoken in parts of northeastern Canada; the dialect discussed in this paper (Tarramituq) is spoken in arctic Quebec along Hudson Strait. It is a polysynthetic language that is generally considered morphologically ergative and has basic SOV word order. Various aspects of its grammar encode person and number; four persons (1, 2, 3 coreferent, 3 disjoint) and three numbers (singular, dual, plural) are normally differentiated. Inuktitut also has a rich system of both nominal case-marking affixes and verbal cross-referencing affixes.

Particularly relevant for this paper is the morphological form in which arguments in Inuktitut appear. This involves two components: verbal cross-referencing affixes on the one hand and independent lexical or demonstrative NPs on the other. Verbal cross-referencing affixes are obligatory (except in restricted situations in colloquial speech; see Swift and Allen n.d.), while independent NPs are optional. Each of these is described in detail below.

3.1. Verbal cross-referencing affixes

The system of verbal cross-referencing affixes in Inuktitut is particularly rich (see Doraits 1988 for further information about this and other aspects of the grammar of Tarramituq). Subjects of grammatically intransitive clauses, (2), and both subjects and objects of grammatically transitive clauses, (3), are reflected in portmanteau affixes on the verb, which
provide information about verbal modality as well as about both the
person and the number of the subject and object.

(2) a. (Paul 2,6)6
Arqarama.
arqa-gama
get.down-CSV.1sS7
‘I’m getting down.’
b. (Louisa 3,6)
Qalangamngaviti.
qal-langa-ngiit-gavit
come-FUT-NEG-CSV.2sS
‘You aren’t going to come.’

(3) a. (Paul 3,3)
Qukisigakkit.
qukuq-si-gakkit
shoot-PRSP-CSV.1sS.2sO
‘I’m shooting you.’
b. (Fijlah 2,5)
Aaminaravinga.
aanuug naq gavinga
hurt-such.as.to-CSV.2sS.1sO
‘You’re hurting me.’

While these portmanteau affixes were undoubtedly originally formed
from distinct component parts, these parts are no longer reliably distin-
guishable in a systematic way.

In addition to standard transitive clauses, Inuktut also has an antipassive
construction in which semantically transitive clauses appear as gram-
matically intransitive clauses. In antipassive clauses, the semantic subject
appears in the absolutive case typically used for grammatically intransitive
subjects, and the semantic object appears in the modalis case typically
used for obliques of various sorts. The semantic objects of antipassive
clauses are not reflected in the verbal cross-referencing affix. An example
is shown in (4), note that only the first person subject (and not the third
person semantic object) is marked on the verbal cross-referencing affix.

(4) (Lizzie 3,3)
Piganik aitsilaurlanga?
pig-anik ai-tsi-lauq-langa
thing-MOD.1Sg get-ATP-POL-IMP.1sS
‘Shall I get my thing?’ (referring to her sunglasses)

3.2. Independent NPs for third person arguments

Independent lexical or demonstrative representation of arguments apart
from the verbal cross-referencing affix is not required in Inuktut.
However, third person arguments may be represented in the form of
either independent lexical NPs or independent demonstratives in addition
to verbal cross-referencing affixes. (Note that there are no third person
pronouns in Inuktut, but rather a rich system of demonstratives that
carry information about the number, relative location, and motion of
the referent.) Thus, third person arguments may appear in three possible
morphological forms: lexical (lexical NP plus verbal affix), demonstrative
(demonstrative plus verbal affix), and null (verbal affix only). The differences
in (5) show each of the three possible options of argument represen-
tation with the intransitive verb sinik ‘sleep’; (5a) shows a lexical
argument, (5b) a demonstrative, and (5c) a null argument. Each of the
three options in (5) is possible for a third person argument, while only
(5c) is possible for a first or second person argument.

(5) a. (Paul 3,3)
Panik, piurait sinisijuq,
panik piurait-it sinik-si-juq
daughter baby-ABS.2Sg sleep-PRSP-PAR.3sS
‘Daughter, your baby is sleeping.’
b. (Lizzie 2,10)
Una sinismimat.
u-nu sinik-si-immat
this one ABS.SG sleep-PRSP-CSV.3sS
‘This one is sleeping.’
c. (Lijah 2,9)
Sinimimat.
sinik-liq-immat
sleep-INC-POL-CSV.3sS
‘He/she is sleeping.’

Note that the term argument here is meant to indicate mention of a
referent in either subject or object position, including both independent
and inflectional forms. Overt means that there is an independent lexical
or demonstrative representation of the argument in addition to the verbal
cross-referencing affix, while null means that the argument is represented
only by the verbal cross-referencing affix (or occasionally nothing, if the
verbal cross-referencing affix is omitted).8
3.3. **Independent NPs for first and second person arguments**

As noted above, first and second person arguments are virtually never represented lexically or pronominally, even for emphasis, but rather only through verbal cross-referencing affixes (this is true both in the language in general and in the data discussed in this paper). While first and second person pronouns do exist in Inuktitut, they are used primarily in single-word utterances as arguments of elliptic verbs, (6b), and to indicate possession, (7b), and in equational structures, (8). They may also occur in cases of conjunction with a third person in a plural form, which are very rare in our data. Note that ergative, absolutive, and possessive pronouns are homophonous.

(6) a. (Mother)  
Maqilangangitualu?  
maqaq-langa-ngit-juq-aluk  
hunt-FUT-NEG-PAR.3sS-EMPH  
"He won't go hunting?"

b. (Elijah 2.0)  
Aaa, ivvit.  
aaa ivvit  
yes you/your/yours  
"Right, you (will)."

(7) a. (Mother)  
Kinaup paisikuungua?  
kina-up paisikuq-nga  
who-ERG.SG bicycle-ABS.3sg  
"Whose bicycle is this?"

b. (Paul 3.3)  
Uvangaa.  
uvanga  
l/me/my/mine  
"Mine."

(8) a. (Lizzie 3.3)  
Una ivvit.  
ua-na ivvit  
this one-ABS.SG you/your/yours  
"This one is yours."

b. (Paul 2.11)  
Uvangaunngimat.  
uvanga-u-ngit-mmak  
l/me/my/mine-he-NEG-CSV.3sS  
"It's not me" (referring to a person in a photo)
transcription conventions of the CHILDES project (MacWhinney and Snow 1990). Transcripts of the first, middle, and last months of taping for each child were chosen for intensive analysis. These transcriptions were checked for accuracy by the author in consultation with native speakers of Inuktitut and were then coded for various morphological, syntactic, and discourse features including those described below.

Preliminary data analysis was conducted using the CLAN programs (MacWhinney and Snow 1990). Utterances that were fully intelligible and complete from the point of view of the child’s intonation, and were not comprised solely of routines (e.g. songs, alphabet), were included in the data set to be analyzed. Note that utterances composed of partial and complete self-repetitions and imitations of previous speakers were included in the analysis.

The unit of analysis considered for this paper is the verbal argument, including subjects of intransitive verbal clauses, and subjects and objects of transitive verbal clauses. All verbal clauses contained a verb and/or verbal element (e.g. tense, modality, or negation affix) and its arguments. Examples of intransitive clauses with overt (lexical, [9a]) and null, ([9b), subjects are shown in (9). Examples of transitive clauses with null subjects and overt (demonstrative [10a]) and null, ([10b), objects are shown in (10).

(9) a. (Lizzie 2:11)
Qupanuaarluk qalangammat
qupanuaq-guluk-0 qai-langa-mmatt
bird-DIM-ABS.SG come-FUT-CSV.3sS
‘The little bird is going to come.’

b. (Paul 2:6)
Maanastajunga.
ma-una-aq-si-junga
here-VIA-go-PRSP-PAR.1sS
‘I’m going through here.’

(10) a. (Elijah 2:5)
Una aturtara.
a-na atuq-juna
this-one-ABS.SG use-PAR.1sS.3sO
‘I’m using this one.’

b. (Louise 3:2)
Ijukatilauruk.
ijukkaq-tit-lauq-guk
fall-CAUS-POL-IMP.2sS.3sO
‘Make it fall.’

Verbal clauses lacking either a verb root, ([11a]), or cross-referencing affix, ([11b]), or both, ([11c]), but containing other verbal affixes (e.g. secondary verb, tense, aspect, or negation) were also included in the analysis. Such clauses are considered colloquially appropriate in adult and child Inuktitut, and the root and affix information is recoverable from context (Swift and Allen n.d.).

(11) a. (Lizzie 2:6)
Titauruk.
[- Itsivatilauruk;]
tit-laauq-guk
[- atsiv-tit-laauq-guk;
CAUS-POL-IMP.2sS.3sO
[- sit-CAUS-POL-IMP.2sS.3sO]
‘Make it do X.’
[= ‘Make it sit down’]
(asking her mother to help her make a doll sit down)

b. (Paul 2:6)
Paisikunguaq.
[= Paisikunngutauq;]
paisikuq-nunguaq
[= paisikuq-nunguaq-juq;]
bicycle-pretend
[- bicycle-pretend-PAR.3sS]
‘Pretending to bicycle.’
[= ‘He’s pretending to bicycle.’]
(explaining that a cartoon character is pretending to bicycle)

c. (Elijah 2:6)
Gumangit.
[= Atjiliuqaluangumanginama;]
guua-nangit
[= atjiliujanuq-guma-nangit-gama;]
want-NFG
[- film-PASS-want-NEG-CSV.1sS]
‘Don’t want to.’
[= ‘I don’t want to be filmed.’]
(telling his mother that he doesn’t want to be videotaped)

In addition, arguments in all clauses with overt copular verbs were analyzed, including equational clauses, ([12a]), and location clauses, ([12b]).

(12) a. (Louisa 3:6)
Parajutsutit?
piaraq-u-tusutit
baby-be-C/TM.2sS
‘When you were a baby?’

b. (Elijah 2:6)
Haakirutilu silaaimituuq.
sila-mi-it-juq
hockey-used.for-EMP1-ABS.SG outside-LOC be PAR.3sS
‘The hockey stick is outside.’

Equational and location clauses without overt copular verbs, ([13]), were not included in the analysis, although they are perfectly grammatical in Inuktitut. This was done in order to make the Inuktitut data maximally
comparable with that from other languages, since most other analyses of argument ellipsis include only arguments in clauses with overt verbs.

(13) a. (Paul 3:3)
 Una arnaq,
 u-ná arnaq-th
 this-one-ABS.SG woman-ABS.SG
 'This (is) a woman.'
 b. (Lizzie 2:6)
 Pipili?
 pipi-li
 baby-where
 'Where (is) the baby?'

The final data set used for analysis in this paper, then, contains a total of 3175 arguments: 493 overt and 2682 null. Data are summarized in Table 2.

4.3. Coding

All subject and object arguments were coded for the form in which the argument appears (overt or null), as well as for all eight informativeness features discussed in section 2 above. Below we give an example of an argument (subject of intransitive verb) in its discourse context, (14), and

\[\text{Table 2. Data analyzed}\]

<table>
<thead>
<tr>
<th>Child</th>
<th>Age</th>
<th>No. null arguments</th>
<th>No. overt arguments</th>
<th>Total no. of arguments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elijah</td>
<td>2:0</td>
<td>226</td>
<td>28</td>
<td>254</td>
</tr>
<tr>
<td></td>
<td>2:5</td>
<td>323</td>
<td>73</td>
<td>396</td>
</tr>
<tr>
<td></td>
<td>2:9</td>
<td>341</td>
<td>73</td>
<td>414</td>
</tr>
<tr>
<td>Lizzie</td>
<td>2:6</td>
<td>197</td>
<td>34</td>
<td>231</td>
</tr>
<tr>
<td></td>
<td>2:10</td>
<td>330</td>
<td>59</td>
<td>389</td>
</tr>
<tr>
<td></td>
<td>3:3</td>
<td>117</td>
<td>26</td>
<td>143</td>
</tr>
<tr>
<td>Loura</td>
<td>3:10</td>
<td>139</td>
<td>4</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>3:2</td>
<td>434</td>
<td>62</td>
<td>496</td>
</tr>
<tr>
<td></td>
<td>3:6</td>
<td>281</td>
<td>63</td>
<td>344</td>
</tr>
<tr>
<td>Paul</td>
<td>2:6</td>
<td>109</td>
<td>18</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>2:11</td>
<td>102</td>
<td>14</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>3:3</td>
<td>183</td>
<td>35</td>
<td>218</td>
</tr>
<tr>
<td>Total</td>
<td>2682</td>
<td>493</td>
<td>3175</td>
<td></td>
</tr>
</tbody>
</table>

Note that three utterances in the present data set involved ungrammatical use of first person pronouns in argument position. These are shown in (16).

(14) (Elijah 2:5)
 Atatuq siinitnuq.
 atatu-ga sinik-juq
 father-ABS.SG sleep-PAR.3sS
 'My father is sleeping.'

(setting is a small one-room shack where Elijah and his family are spending the summer; most of the adults are asleep on a sleeping platform since it is raining and thus not good weather for hunting; Elijah tells his mother out of the blue that his father is sleeping)

(15) a. FORM: overt (lexical)
b. ABSENCE: uninformative
c. CONTRAST: uninformative
d. DIFFERENTIATION IN CONTEXT: uninformative
e. DIFFERENTIATION IN DISCOURSE: uninformative
f. INANIMACY: uninformative
g. NEWNESS: informative
h. QUERY: uninformative
i. THIRD PERSON: informative

(16) a. (Lizzie 3:3)
 Uvanga maniikinimatuq!
 uvanga ma-ani-it-kiinaq-juq
 l/me/my/mine here-LOC-be-PAST-PAR.3sS
 'I was here!'

(trying to reclaim the place where she was previously standing, now taken by someone else)
b. (Lizzie 3:3)
 Uvanga inutuulunga.
 uvanga inut-ua-lunga
 l/me/my/mine single.person-be-ICM.1sS
 'I'll be alone.'

(wanting to stand on the bed alone, without her friends who are currently on the bed)
c. (Elijah 2:9)
Uvagut sinilangagunahigut?

uvagut sinik-langa-gunnahag-vugut
we/us/our/ours sleep-FUT-not-be.permitted to-IND 1pS
‘Are we not permitted to sleep?’
(asking his mother whether he and his companions are permitted to go to sleep, probably since he knows he should stay awake while being videotaped)

These utterances were each coded as having an overt subject argument. Utterance (16a) is coded as having a third person subject because of the third person verbal cross-referencing affix, while (16b)-(16c) are coded as having first person subjects.

Finally, there are eight utterances in the data set in which independent NPs seem to be used in a grammatically correct way to represent first and second person arguments. These are utterances with plural arguments in which the non-speaker (for first person) or non-addressee (for second person) is explicitly mentioned. Examples are given in (17).

(17) a. (Elijah 2:5)
Anaana, namungangavisi Maasitul?
aana na-mungag-aq-langa-va
mother where-ALL-go-FUT-INT.2pS matthew-and
‘Mother, where are you going with Matthew?’

b. (Elijah 2:9)
Ataatagalu sikituuqattammialruqung?
atata-qa-lu skituuq-qataq-miqa-liu-iu
father-ERG.1sg-and ride.skidoor-ABS-PRES-FUT-INC-P-1pS
‘Will my father and I ride the skidoor later?’

c. (Elijah 2:9)
Sinilangasivita piararululu?
sinik-langa-si-vita piaraq-guluk-lu
sleep-FUT-PRSP-IND.1pS baby-DIM-and
‘Will the baby and us sleep later?’

These overt NPs could be analyzed as independent representations of (part of the) first and second person arguments. However, they could also be analyzed as phrases conjoined to the argument, and not part of the real argument. The utterances in (17b)-(17c) suggest the latter analysis, since the verbal cross-referencing affix does not seem to include the referent in the conjoined phrase. Thus, we have chosen to code the relevant arguments as null subjects, (17b)-(17c), and objects, (17a).

5. Results

The main question to be answered in this study is whether informativeness could be a reasonable explanation for the overtness of arguments in early childhood Inuktitut (and, by extension, for argument ellipsis). Informativeness features are distributed across the arguments in the data set as shown in Table 3. Note that although there are a total of 3175 arguments in the data set (see Table 2), only 3168 of these have values for all eight informativeness features (e.g. an argument could not be coded for ABSENCE if the videocamera was focused out the window at the time of utterance, and therefore the coder couldn’t determine whether the referent was present in or absent from the physical context). Thus, we include only these 3168 arguments in the analyses from this point on in the paper.

Since we are interested in the relationship between informativeness and overtness of argument, we calculated the number and percentage of arguments with an informative value for each feature that appear as overt rather than ellipsed arguments. These figures are shown in Table 4. For comparative purposes we also calculated the number and percentage of arguments with an uninformative value for each feature that appear as overt rather than ellipsed arguments. These figures are shown in Table 5. The differences in Proportion overt between Tables 4 and 5 is highly indicative that informativeness has a strong effect on the form in which an argument appears.

The result of main importance to this paper is the difference between the figures in the last columns of Table 4 and Table 5, since this difference represents the relationship between informativeness and overtness: the Proportion overt is large if the feature is informative, but smaller and often close to 0 if the feature is uninformative. In order to test the

<table>
<thead>
<tr>
<th>Feature</th>
<th>Number informative</th>
<th>Number noninformative</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSENCE</td>
<td>189</td>
<td>2979</td>
</tr>
<tr>
<td>CONTRAST</td>
<td>349</td>
<td>2819</td>
</tr>
<tr>
<td>DIFFERENTIATION IN CONTEXT</td>
<td>738</td>
<td>2433</td>
</tr>
<tr>
<td>DIFFERENTIATION IN DISCOURSE</td>
<td>261</td>
<td>2907</td>
</tr>
<tr>
<td>INFANIMACY</td>
<td>1089</td>
<td>7079</td>
</tr>
<tr>
<td>NEWNESS</td>
<td>419</td>
<td>2749</td>
</tr>
<tr>
<td>QUERY</td>
<td>4</td>
<td>3164</td>
</tr>
<tr>
<td>THIRD PERSON</td>
<td>1406</td>
<td>1762</td>
</tr>
</tbody>
</table>
Table 4. Number and proportion of informative arguments represented overtly

<table>
<thead>
<tr>
<th>Feature</th>
<th>Number informative</th>
<th>Number overt</th>
<th>Proportion overt</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSENCE</td>
<td>189</td>
<td>69</td>
<td>0.37</td>
</tr>
<tr>
<td>CONTRAST</td>
<td>349</td>
<td>88</td>
<td>0.25</td>
</tr>
<tr>
<td>DIFFERENTIATION IN CONTEXT</td>
<td>735</td>
<td>306</td>
<td>0.42</td>
</tr>
<tr>
<td>DIFFERENTIATION IN DISCOURSE</td>
<td>261</td>
<td>137</td>
<td>0.53</td>
</tr>
<tr>
<td>INANIMACY</td>
<td>1089</td>
<td>380</td>
<td>0.35</td>
</tr>
<tr>
<td>NEWNESS</td>
<td>419</td>
<td>230</td>
<td>0.55</td>
</tr>
<tr>
<td>QUERY</td>
<td>4</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>THIRD PERSON</td>
<td>1406</td>
<td>484</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Table 5. Number and proportion of uninformative arguments represented overtly

<table>
<thead>
<tr>
<th>Feature</th>
<th>Number uninformative</th>
<th>Number overt</th>
<th>Proportion overt</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSENCE</td>
<td>2979</td>
<td>417</td>
<td>0.14</td>
</tr>
<tr>
<td>CONTRAST</td>
<td>2819</td>
<td>391</td>
<td>0.14</td>
</tr>
<tr>
<td>DIFFERENTIATION IN CONTEXT</td>
<td>2433</td>
<td>180</td>
<td>0.07</td>
</tr>
<tr>
<td>DIFFERENTIATION IN DISCOURSE</td>
<td>2907</td>
<td>349</td>
<td>0.12</td>
</tr>
<tr>
<td>INANIMACY</td>
<td>2075</td>
<td>106</td>
<td>0.05</td>
</tr>
<tr>
<td>NEWNESS</td>
<td>2749</td>
<td>226</td>
<td>0.09</td>
</tr>
<tr>
<td>QUERY</td>
<td>3164</td>
<td>485</td>
<td>0.15</td>
</tr>
<tr>
<td>THIRD PERSON</td>
<td>1762</td>
<td>2</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The significance of these results, we performed several logistic regression analyses.

Logistic regression was used since it allows prediction of a discrete (usually binary) outcome such as group membership (e.g. FORM = overt/null) from a set of independent variables. Logistic regression emphasizes the probability of a particular outcome for each case. For example, it evaluates the probability that an argument will be overt given its informativeness properties. The first step in logistic regression is to establish that there is a relationship between the outcome and the set of predictors. If a relationship is found, then one may try to simplify the model by eliminating some predictors while still maintaining strong prediction. Once a reduced set of predictors is found, the equation can be used to predict outcomes for new cases on a probabilistic basis. (See Moore and McCabe [1998] and Tabachnick and Fidell [1996] for more detailed information about logistic regression.)

We performed a direct logistic regression analysis with one binary outcome variable (FORM = null or overt) and eight predictor variables (ABSENCE, CONTRAST, DIFFERENTIATION IN CONTEXT, DIFFERENTIATION IN DISCOURSE, INANIMACY, NEWNESS, QUERY, and THIRD PERSON). All the predictor variables were specified as categorical variables, with value 1 if the argument was informative for that feature, and value 0 if the argument was not informative for that feature. Analysis was performed using SPSS. After deletion of seven cases due to missing values, 3168 cases were available for analysis: 2682 with FORM = null and 486 with FORM = overt.

A test of the full model with all eight predictors against a constant-only model (a model with no predictors) was statistically reliable, \( \chi^2 \) (df = 8, n = 3168) = 1067.697, p < 0.001, indicating that the predictors, as a set, reliably distinguished between overt and null arguments. Prediction success was good, with 95% of null arguments and 47% of overt arguments correctly predicted, for an overall success rate of 87% (an increase of 2% over a model with no predictors, which has the prediction that all arguments are null, so that 0% of the overt arguments are correctly predicted). This result is further substantiated in Table 6, which reports the regression coefficients (B), odds ratios (e^\( \beta \)), and significance (p-value) for each of the eight predictors. The p-values pertain to the null hypothesis that B = 0 or, equivalently, e^\( \beta \) = 1. A p-value close to 0 therefore means that B is reliably different from 0.

The results reported in Table 6 indicate that the probability of an overt argument satisfies the formula in (18):

\[
\text{prob(overt)} = \frac{\text{exp}(\text{const} + \text{absent} + \text{context} + \text{differentiation in context} + \text{differentiation in discourse} + \text{inanimacy} + \text{newness} + \text{query} + \text{third person})}{1 + \text{exp}(\text{const} + \text{absent} + \text{context} + \text{differentiation in context} + \text{differentiation in discourse} + \text{inanimacy} + \text{newness} + \text{query} + \text{third person})}
\]

Table 6. Results of logistic regression analysis of form with informativeness features

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B</th>
<th>e^{\beta}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2.965*</td>
<td>0.05</td>
</tr>
<tr>
<td>ABSENCE</td>
<td>0.418*</td>
<td>1.5195</td>
</tr>
<tr>
<td>CONTRAST</td>
<td>0.713*</td>
<td>2.0367</td>
</tr>
<tr>
<td>DIFFERENTIATION IN CONTEXT</td>
<td>0.305*</td>
<td>1.3566</td>
</tr>
<tr>
<td>DIFFERENTIATION IN DISCOURSE</td>
<td>0.065*</td>
<td>1.0671</td>
</tr>
<tr>
<td>INANIMACY</td>
<td>-0.999*</td>
<td>0.3993</td>
</tr>
<tr>
<td>NEWNESS</td>
<td>0.547*</td>
<td>1.7281</td>
</tr>
<tr>
<td>QUERY</td>
<td>0.258*</td>
<td>1.2952</td>
</tr>
<tr>
<td>THIRD PERSON</td>
<td>2.765*</td>
<td>15.8847</td>
</tr>
</tbody>
</table>

a. p < 0.01.
b. p < 0.001.
c. Not significant.
Here, the individual predictor variables assume the value 1 (in place of $x$) if the argument is informative, and 0 if it is not informative. The coefficients in the formula (e.g., constant, abs $\rightarrow$ ABSENCE; difcon $\rightarrow$ DIFFERENTIATION IN CONTEXT) are the B-weights of Table 6 as they are estimated by logistic regression analysis from the data. A positive B-weight means that increasing the informativeness on that particular feature will lead to a higher probability of an overt argument. A negative B-weight means that increasing the informativeness on that particular feature will lead to a lower probability of an overt argument. A zero B-weight means that there is no effect of that feature. As can be seen, all B-weights are positive except for that of INANIMACY, which is, however, not significantly different from 0. This confirms our theoretical expectation that increasing informativeness will increase the likelihood of using an overt argument.

The next question is whether the B-weights are statistically reliably different from 0, that is, whether they are significant. This is reported in Table 6 in the column with p-values. The p-value of the whole model is 0.0000, meaning that at least one of the B-weights for individual features is significantly different from 0. This allows us to look at the separate p-values for the individual features. These are reported in Table 6, with the null hypothesis that $B = 0$ for that variable.

Of the individual predictor variables, five have a significant effect on the form of the argument: ABSENCE, CONTRAST, DIFFERENTIATION IN CONTEXT, NEWNESS, and THIRD PERSON ($p < 0.001$ for all five variables). The largest effect is that of THIRD PERSON: the odds of an overt argument $^{18}$ (i.e. the ratio $p(\text{overt})/p(\text{null})$) is about 16 times as large if it represents a third person referent than if it represents a first or second person referent ($e^B = 15.88$). This effect will be discussed further below. Of the remaining features of informativeness, CONTRAST is next most important ($e^B = 2.04$), followed by NEWNESS ($e^B = 1.73$), ABSENCE ($e^B = 1.52$), and DIFFERENTIATION IN CONTEXT ($e^B = 1.36$). Although the effect of each of these latter four features is relatively small compared with that of THIRD PERSON, this is not really surprising. Since there are several features involved in informativeness, it is normal that each contributes something and thus that the effect of each is relatively small.

Three variables do not have a significant effect: QUERY ($p = 0.6561$), INANIMACY ($p = 0.909$), and DIFFERENTIATION IN DISCOURSE ($p = 0.4467$). We hypothesize that QUERY has no effect since only four of the 3168 arguments have an informative value for the feature QUERY and thus the numbers in our data set are not large enough to indicate any patterns.

---

We hypothesize that the effect of INANIMACY is not significant because it is too general a feature. Recall from above that knowing that a referent is animate tends to reduce the search space for the referent, while knowing that it is inanimate does not restrict the search space very much. The effect of INANIMACY, then, is to leave the search space wider than it would be if the other binary value for this feature were active. However, search-space increase does not really encourage or force overtness in the same way that the other features do. Thus INANIMACY is not a significant predictor.

We hypothesize that DIFFERENTIATION IN DISCOURSE is not significant as a predictor because its effect is masked by those of DIFFERENTIATION IN CONTEXT and CONTRAST, since these three predictors are confounded (Moore and McCabe 1998). Recall that DIFFERENTIATION IN DISCOURSE applies to those arguments for which one or more additional potential referents has been mentioned in one or more of the preceding five utterances, while DIFFERENTIATION IN CONTEXT applies to those arguments for which one or more additional potential referents is present in the physical context of the discourse, and CONTRAST applies to those arguments for which the child is purposefully contrasting the intended referent with another potential referent. As might be expected, many arguments are informative for either two or three of these features together. These relationships are shown in Tables 7 and 8. To tease apart the effect of these three features as predictors of FORM, we performed a sequential logistic regression analysis. That is, we performed three logistic regression analyses each adding new predictors to the predictors previously tested, and tested the...

---

<table>
<thead>
<tr>
<th>Table 7. Number of arguments when CONTRAST is informative</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIFFERENTIATION IN DISCOURSE</td>
</tr>
<tr>
<td>informative</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>41</td>
</tr>
<tr>
<td>70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 8. Number of arguments when CONTRAST is uninformative</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIFFERENTIATION IN DISCOURSE</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>informative</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>139</td>
</tr>
<tr>
<td>485</td>
</tr>
</tbody>
</table>
improvement after the addition of the new predictors to the model. The
three analyses were as follows:

**Analysis 1:** DIFFERENTIATION IN DISCOURSE
(in addition to ABSENCE, INANIMACY, NEWNESS,
QUERY, THIRD PERSON)

**Analysis 2:** DIFFERENTIATION IN DISCOURSE and
DIFFERENTIATION IN CONTEXT
(in addition to ABSENCE, INANIMACY, NEWNESS,
QUERY, THIRD PERSON)

**Analysis 3:** DIFFERENTIATION IN DISCOURSE,
DIFFERENTIATION IN CONTEXT, and CONTRAST
(in addition to ABSENCE, INANIMACY, NEWNESS,
QUERY, THIRD PERSON)

As in the previous analysis, seven of the original 3175 cases were deleted
due to missing values, resulting in 3168 cases available for this analysis.
A test of the model with the predictors in analysis 1 against a constant-
only model was statistically reliable, \( \chi^2 (df = 6, n = 3168) = 989.877, p < 0.001 \), indicating that these six predictors, as a set, reliably distin-
guished between overt and null arguments. The additions of predictors in
analysis 2 and analysis 3 each significantly improved the model (\( \chi^2 \)
\( df = 1, n = 3168 \) = 31.977, \( p < 0.001 \)) and \( \chi^2 \) (\( df = 1, n = 3168 \) = 45.844,
\( p < 0.001 \)) respectively).

The B-weights and p-values for each predictor in this sequential
logistic regression analysis are reported in Table 9. Note that

| Table 9. B-weights of sequential logistic regression analysis |
|-------------------|-------------------|-------------------|
|                   | Analysis 1        | Analysis 2        | Analysis 3        |
| Constant           | -3.018b           | -2.819b           | -2.065b           |
| ABSENCE            | 0.204a            | 0.412a            | 0.4184a           |
| INANIMACY          | 0.074b            | -0.0092           | -0.0092           |
| NEWNESS            | 0.572a            | 0.535a            | 0.5340a           |
| QUERY              | 0.0017d           | 0.3495d           | 0.7587d           |
| THIRD PERSON       | 2.7665b           | 2.6635b           | 2.7654b           |
| DIFFERENTIATION IN DISCOURSE | 0.1437b | 0.1656b | 0.0659b |
| DIFFERENTIATION IN CONTEXT | 0.3885b | 0.3406b | 0.7123b |

- a. \( p < 0.05 \)
- b. \( p < 0.01 \)
- c. \( p < 0.001 \)
- d. Not significant.

DIFFERENTIATION IN DISCOURSE had a close-to-significant effect on FORM in analysis 1 (\( p = 0.0799 \)), but decreased in significance in
analyses 2 and 3 (\( p = 0.2127 \) and \( p = 0.4467 \) respectively). Thus, we
conclude that DIFFERENTIATION IN DISCOURSE may have an
effect on FORM, but that this effect is confounded by the effects of
DIFFERENTIATION IN CONTEXT and CONTRAST.

5.1. Third person

We noted above that the most significant feature predicting the form of the
argument is THIRD PERSON. Thus, an argument is much more
likely to be overt if it represents a third person referent. Although this
might be expected on independent grounds (DuBois 1987; Clancy i.p.),
recall from section 3.1 above that it is typically considered ungrammatical
in Inuktitut to use an overt form to represent a first or second person
referent (except in cases of conjunction with a third person in a plural
form, which are very rare in our data). Thus, the feature THIRD
PERSON is not really an informativeness feature in Inuktitut, but rather
part of the grammar of how arguments must appear. In order to remove
this confusion from our analysis, we added the above analyses for only
those arguments that represent third person referents. After deletion of
seven cases due to missing values, 1406 cases were available for analysis;
922 with FORM = null and 484 with FORM = overt. Informativeness
features are distributed across these arguments as shown in Table 10.
Since we are interested in the relationship between informativeness and
overtness of argument, we calculated the number and percentage of
arguments with an informative value for each feature that appear as
overt rather than elided arguments. These figures are shown in Table 11.
To compare this, we also calculated the number and percentage of argu-

| Table 10. Number of informative and noninformative third person arguments for each informativeness feature |
|---------------------------------|-----------------|-----------------|
| Feature                        | Number informative | Number noninformative |
| ABSENCE                        | 180             | 1276            |
| CONTRAST                       | 122             | 1284            |
| DIFFERENTIATION IN CONTEXT     | 734             | 672             |
| DIFFERENTIATION IN DISCOURSE   | 261             | 1145            |
| INANIMACY                      | 1076            | 330             |
| NEWNESS                        | 419             | 987             |
| QUERY                          | 4               | 1406            |
Arguments with an uninformative value for each feature that appear as overt rather than ellipsed arguments. These figures are shown in Table 12. The difference in Proportion overt between Tables 11 and 12 is highly indicative that informativeness also has a strong effect on the form in which a third person argument appears.

We performed a logistic regression to test the significance of these results. A test of the full model with all seven predictors against a constant-only model (a model with no predictors) was statistically reliable, $\chi^2 (df = 7, n = 1406) = 192.289, p < 0.001$, indicating that the predictors, as a set, reliably distinguished between overt and null arguments. Prediction success was good, with 84% of null arguments and 47% of overt arguments correctly predicted, for an overall success rate of 71% (an increase of 5% over a model with no predictors, which has the prediction that all arguments are null, so that 0% of the overt arguments are correctly predicted). This result is further substantiated in Table 13, which reports the regression coefficients ($B$s), odds ratios ($e^B$s), and significance (p-value) for each of the eight predictors. As above, the p-values pertain to the null hypothesis that $B = 0$ or, equivalently, $e^B = 1$. A p-value close to 0 therefore means that $B$ is reliably different from 0.

The same variables were significant and not significant as in the analysis including all arguments, with the odds ratios varying little (see Table 6). This is not surprising since logistic regression is meant to measure the effect of each variable assuming the effect of the others is held constant. Only the constant changed, from 2.0652 to 0.6954. Thus we can conclude that the feature THIRD PERSON affects the form in which an argument appears, but it does not affect the relation of FORM with the other predictors.

### Table 11. Number and proportion of third person informative arguments represented overtly

<table>
<thead>
<tr>
<th>Feature</th>
<th>Number informative</th>
<th>Number overt</th>
<th>Proportion overt</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSENCE</td>
<td>180</td>
<td>85</td>
<td>0.48</td>
</tr>
<tr>
<td>CONTRAST</td>
<td>122</td>
<td>33</td>
<td>0.33</td>
</tr>
<tr>
<td>DIFFERENTIATION IN CONTEXT</td>
<td>734</td>
<td>306</td>
<td>0.42</td>
</tr>
<tr>
<td>DIFFERENTIATION IN DISCOURSE</td>
<td>261</td>
<td>137</td>
<td>0.55</td>
</tr>
<tr>
<td>INANIMACY</td>
<td>1076</td>
<td>330</td>
<td>0.30</td>
</tr>
<tr>
<td>NEWNESS</td>
<td>419</td>
<td>230</td>
<td>0.55</td>
</tr>
<tr>
<td>QUERY</td>
<td>4</td>
<td>1</td>
<td>0.25</td>
</tr>
</tbody>
</table>

### Table 12. Number and proportion of uninformative third person arguments represented overtly

<table>
<thead>
<tr>
<th>Feature</th>
<th>Number informative</th>
<th>Number overt</th>
<th>Proportion overt</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSENCE</td>
<td>1226</td>
<td>415</td>
<td>0.34</td>
</tr>
<tr>
<td>CONTRAST</td>
<td>1284</td>
<td>397</td>
<td>0.31</td>
</tr>
<tr>
<td>DIFFERENTIATION IN CONTEXT</td>
<td>672</td>
<td>178</td>
<td>0.27</td>
</tr>
<tr>
<td>DIFFERENTIATION IN DISCOURSE</td>
<td>1145</td>
<td>347</td>
<td>0.30</td>
</tr>
<tr>
<td>INANIMACY</td>
<td>987</td>
<td>254</td>
<td>0.26</td>
</tr>
<tr>
<td>NEWNESS</td>
<td>1406</td>
<td>483</td>
<td>0.35</td>
</tr>
</tbody>
</table>

### Table 13. Results of logistic regression analysis of form with informativeness features, for only third person arguments

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$B$</th>
<th>$e^B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.6954</td>
<td>1.201</td>
</tr>
<tr>
<td>ABSENCE</td>
<td>0.4188</td>
<td>1.5201</td>
</tr>
<tr>
<td>CONTRAST</td>
<td>0.7044</td>
<td>1.9667</td>
</tr>
<tr>
<td>DIFFERENTIATION IN CONTEXT</td>
<td>0.3035</td>
<td>1.3577</td>
</tr>
<tr>
<td>DIFFERENTIATION IN DISCOURSE</td>
<td>0.0653</td>
<td>1.0694</td>
</tr>
<tr>
<td>INANIMACY</td>
<td>0.0009</td>
<td>0.9990</td>
</tr>
<tr>
<td>NEWNESS</td>
<td>0.5470</td>
<td>1.7280</td>
</tr>
<tr>
<td>QUERY</td>
<td>0.2580</td>
<td>1.2952</td>
</tr>
</tbody>
</table>

a. $p < 0.001$.

b. Not significant.

5.2. Child

Since the results of the above analyses are contributed by four separate children, we decided to investigate the feature CHILD as a predictor to see if the children differed from each other on their use of form. The predictor CHILD had a significant effect when inserted in the logistic regressions discussed above ($p < 0.001$). Interestingly, the B and p values for the other predictors did not change much whether CHILD was included in or omitted from the analysis. These results suggest that the children all use the informativeness features in the same way to choose between overt and null arguments (because the B-weights are the same), but that different children may have higher or lower thresholds for actually producing an overt argument (i.e. CHILD affects only the constant, not the other B-weights).
We independently assessed the difference between children in how often they produced overt arguments and arguments with informative values for each of the informativeness features. Table 14 shows these results. This table shows that the four children are significantly different in how often they produce arguments with informative values for all but two of the informativeness features, but that these differences are relatively small. The children also differ significantly in the proportion of their use of overt forms. This finding can be explained by the fact that (a) they have different thresholds for producing overt arguments, and (b) they produce arguments with informative values for given informativeness features in slightly different proportions. It is not in contradiction with our previous conclusion that they apply the same rules of choosing between overt and null arguments depending on the value of the informativeness features.

5.3. Informativeness features together

In addition to looking at each informativeness feature separately, we decided to look at them together. We created a new variable INFORMATIVENESS that has the value 1 if any one (or more) of the features ABSENCE, CONTRAST, DIFFERENTIATION IN CONTEXT, DIFFERENTIATION IN DISCOURSE, NEWNESS, or QUERY is informative, and that has the value 0 if none of these features is informative. (Note that the search space features INANIMACY and THIRD PERSON were not included in this analysis.) We performed a direct logistic regression analysis using the variable INFORMATIVENESS as the only predictor. After deletion of seven cases with missing values, 3168 cases were available for analysis: 1280 informative and 1888 uninformative arguments. A test of the full model with one predictor against a constant-only model was statistically reliable, $\chi^2 (df = 1, n = 3168) = 528.266, p < 0.001$. However, prediction success was unimpressive, with 100% of the null arguments but 0% of the overt arguments correctly predicted, for an overall success rate of 85%. Nonetheless, we found that the odds of an argument being overt were almost four times as large if the argument had an informative value for the feature INFORMATIVENESS than if it had an uninformative value ($e^{\beta} = 3.7074, p < 0.001$).

We then created a second new variable, called INFORMATIVENESS2, for which we coded as 1 all arguments having an informative value for two or more informativeness features ($n = 444$), and for which we coded as 0 all arguments having an informative value for less than two informativeness features ($n = 2724$). Again, a test of the full model with one predictor against a constant-only model was statistically reliable, $\chi^2 (df = 1, n = 3168) = 555.142, p < 0.001$. Prediction success was more impressive than for the previous analysis, with 93% of the null arguments and 54% of the overt arguments correctly predicted, for an overall success rate of 87%. We found that the odds of an informative argument being overt were about the same as in the previous analysis ($e^{\beta} = 3.9519, p < 0.001$).

Next, we repeated both analyses above restricting the analysis of third person arguments only (1406 cases after seven were deleted due to missing values). A test of the full model with the predictor INFORMATIVENESS against a constant-only model was statistically reliable, $\chi^2 (df = 1, n = 1406) = 64.530, p < 0.001$. Prediction success was unimpressive, with 100% of the null arguments but 0% of the overt arguments correctly predicted, for an overall success rate of 66%. Nonetheless, we found that the odds of an argument being overt were almost twice as large if the argument was informative than if it was not informative ($e^{\beta} = 1.7658, p < 0.001$). A test of the full model with the predictor INFORMATIVENESS2 against a constant-only model was also statistically reliable, $\chi^2 (df = 1, n = 1406) = 166.277, p < 0.001$. Prediction success was more impressive, with 80% of the null arguments and 54% of the overt arguments correctly predicted, for an overall success rate of 71%. Again, we found that the odds of an argument being overt were about twice as large if the argument had at least two features of informativeness than if it had fewer than two features of informativeness ($e^{\beta} = 2.1724, p < 0.001$).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Elijah</th>
<th>Lizzie</th>
<th>Louisa</th>
<th>Paul</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overt argument</td>
<td>0.18</td>
<td>0.16</td>
<td>0.13</td>
<td>0.15</td>
<td>a</td>
</tr>
<tr>
<td>ABSENCE</td>
<td>0.09</td>
<td>0.06</td>
<td>0.04</td>
<td>0.03</td>
<td>c</td>
</tr>
<tr>
<td>CONTRAST</td>
<td>0.05</td>
<td>0.06</td>
<td>0.19</td>
<td>0.14</td>
<td>c</td>
</tr>
<tr>
<td>DIFFERENTIATION IN CONTEXT</td>
<td>0.18</td>
<td>0.26</td>
<td>0.23</td>
<td>0.27</td>
<td>c</td>
</tr>
<tr>
<td>DIFFERENTIATION IN DISCOURSE</td>
<td>0.06</td>
<td>0.10</td>
<td>0.09</td>
<td>0.10</td>
<td>b</td>
</tr>
<tr>
<td>INANIMACY</td>
<td>0.29</td>
<td>0.38</td>
<td>0.36</td>
<td>0.37</td>
<td>c</td>
</tr>
<tr>
<td>NEWNESS</td>
<td>0.11</td>
<td>0.15</td>
<td>0.14</td>
<td>0.12</td>
<td>d</td>
</tr>
<tr>
<td>QUERY</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>d</td>
</tr>
<tr>
<td>THIRD PERSON</td>
<td>0.43</td>
<td>0.47</td>
<td>0.41</td>
<td>0.49</td>
<td>a</td>
</tr>
</tbody>
</table>

a. $p < 0.05$.  
b. $p < 0.01$.  
c. $p < 0.001$.  
d. Not significant.
Finally, we may compute from the formula in (18) the probability of an overt form in the two situations in (19).

\[(19) \quad \begin{align*}
\text{a. all informativeness features have the value 0} \\
\text{b. all informativeness features have the value 1}
\end{align*}\]

If we consider all arguments, we get the results in (20).

\[(20) \quad \begin{align*}
\text{a. 11.3\%} \\
\text{b. 95.2\%}
\end{align*}\]

The actual data shows that 0.1\% of arguments \((n = 1514)\) are overt in situation (19a). Since there are no arguments in the actual data that fully conform to situation (19b) — that is, that are informative for all eight informativeness features — we based our calculation on those arguments that are informative for five informativeness features \((n = 32)\). Some 84.4\% of arguments are overt in this situation.\(^5\) Thus, the equation developed here has very good predictive power.

If we restrict our analysis to third person arguments only, we get the results in (21).

\[(21) \quad \begin{align*}
\text{a. 66.7\%} \\
\text{b. 35.3\%}
\end{align*}\]

The actual data show that 15.5\% of third person arguments \((n = 110)\) are overt in situation (19a). Again, there are no arguments in the actual data that fully conform to situation (19b), so we based our calculation on those arguments that are informative for five informativeness features \((n = 32)\). Some 84.4\% of arguments are overt in this situation. Thus, the equation developed here has very good predictive power for those arguments with high informativeness, but substantially less predictive power for those arguments that are not informative at all.

6. Discussion

The results of the analyses presented in section 5 clearly indicate that Inuit children pay attention to discourse pragmatics in choosing whether to represent an argument as overt or null. They produce overt arguments significantly more often if the argument they wish to represent is informative than if it is not. Thus, our initial hypothesis is upheld.

Although there is clearly an effect of informativeness in argument representation, this effect does not initially appear to be as strong as we might have hoped. The improvement of a model with informativeness features as predictors over one with no predictors seems relatively small: 2\% for all arguments (Table 6) and 8\% for third person arguments (Table 13). However, two mitigating factors must be kept in mind when considering this result. First, we consider the rate of improvement in the model overall. Since null arguments are so frequent in Inuktitut, the model with no predictors already accurately predicts 85\% of argument form for all arguments and 66\% for third person arguments, just by assuming that 100\% of the arguments will be null. This means that the possibility to improve the model by adding informativeness predictors is only 15\% and 34\% respectively. If we consider the 2\% and 5\% improvement on the basis of the room there is to improve, about 14\% improvement occurs in each case. This rate of improvement is much better than it would have seemed without this adjustment.

Second, we consider the rate of improvement in the prediction of overt arguments. A model with no predictors accurately predicts 0\% of the overt arguments — a quite unimpressive performance. However, the prediction rate improves substantially when a model with informativeness features as predictors is compared to a model with no predictors: to 47\% both for all arguments taken together and for only third person arguments. It is clear from these percentages that the informativeness features are highly effective in predicting the overtness of arguments.

In sum, the improvement in the model by adding features of informativeness as predictors seems small at first glance. However, when we investigate more closely the improvement in the model’s ability to predict argument form overall, and to predict the form of overt arguments in particular, we see that the actual effect of the informativeness features is much larger than it initially appeared.

Nonetheless, the fact remains that not all informative arguments are represented by overt forms (see Tables 4 and 11), and some uninformative arguments are represented by overt forms (see Tables 5 and 12). For example, 43\% of new arguments are null (Table 4), while 9\% of given arguments are overt (Table 5). A number of potential explanations for both these scenarios are considered below.

One possible explanation for null informative arguments is that other communicative media than speech are being used to indicate the referent, such as pointing or eye-gaze direction. If one or more of these media are indicating the referent, then additional indication through an overt argument may not be necessary. A second possible explanation for null informative arguments is that the speaker makes a mistake in the first utterance containing the argument (in which the argument is null), which is immediately corrected in the following utterance (in which the argument is overt). Although the first utterance would go against our hypothesis that informativeness induces use of the overt form, the immediate
correction in the following utterance would show that the speaker is nonetheless following a discourse-pragmatic strategy in use of argument form in general. Instances of both these situations occur in the data under investigation here, though it is not yet clear how many of the instances of null informative arguments they can explain. Further research in both these areas is certainly warranted.

One possible explanation of overt uninformative arguments is self-repetition and imitation of preceding utterances. For example, if an adult introduces a new referent in one utterance with an overt form, and then the child imitates that utterance in full, the child may be using an overt form for a given referent. A second possible explanation for overt uninformative arguments is that an argument may be uninformative for one feature (e.g. NEWNESS) while being informative for another (e.g. CONTRAST). Thus the agreement of form may come from another feature than that in question. Again, both these situations occur in the data with some frequency and warrant further investigation.

An additional factor relevant to both the above scenarios is that there may be a hierarchical and/or cumulative effect of informativeness features. For instance, some features may be strong enough by themselves to cause an argument to be produced with an overt form, while other features may only be strong enough to cause the effect in combination with additional features. Also, certain combinations of features may cause an overt form, while other combinations may not. Finally, the pure cumulative effect of numbers of features informative for a given argument may have an effect, so that an argument with three informative features may be more likely to be overt than an argument with only one informative feature. We provided one attempt to address the last of these questions by creating the features INFORMATIVENESS and INFORMATIVENESS2 mentioned in section 5.3 above. Recall that INFORMATIVENESS denotes arguments with one or more informative features, while INFORMATIVENESS2 denotes arguments with two or more informative features. We saw in section 5.3 that having at least one informativeness feature, whatever that feature, substantially increased the odds that a given argument would be overt, although the model overall did not correctly predict any of the overt arguments. However, having at least two informativeness features increased the prediction of overt arguments from 0% to 54%. Thus, our very preliminary investigation into this area shows that there is clearly some cumulative effect of informativeness features. All these questions involving hierarchical and cumulative effects of informative features should be examined much more carefully.

In sum, we can say with a fair degree of certainty that Inuit children use discourse-pragmatic information in deciding whether to represent arguments overtly in their speech, or whether to omit them. Arguments that are informative are much more likely to be produced as overt, while arguments that are not informative are much more likely to be omitted. Though a number of other possible influences on this process remain to be investigated (e.g. gesture, cumulative effect of informativeness features), none of these is likely to substantially influence the core finding that informativeness affects argument representation.

Although Inuit children are learning a language that permits argument ellipsis, we believe our findings about the effects of discourse pragmatics on argument form can also be usefully extended to children learning languages in which argument ellipsis is not typically permitted. Thus, it may well be the case that children learning English, for example, initially assume that discourse-pragmatic information provides the main motivation for argument representation. They differ from adults, then, in that they allow uninformative arguments to be omitted altogether instead of being represented as pronouns. One might even hypothesize that children's early production is completely discourse-based, such that children only produce in their early utterances the elements that are necessary for communication (e.g. Greenfield and Smith 1976). As children realize the argument-obligatory nature of the grammar of the language they are learning, they gradually begin to produce more and more uninformative arguments overtly (mostly as pronouns) until they reach adult levels. The discourse-pragmatic account, then, explains which arguments children choose to omit and which they choose to represent overtly.

In this way, the discourse-pragmatic account is an advance over early grammatical accounts (e.g. Hyams 1986; Radford 1990) since the latter cannot explain why some arguments are omitted but others are not. Early grammatical accounts predict that the child would experience two distinct stages — one in which all subjects are omitted and a subsequent one in which all subjects are overt. However, it is now well known that this is not the case: the initial stage of omission is not total, and the change to adult production norms is gradual rather than sudden.

Second, the discourse-pragmatic account is also an improvement over the grammatical topic-drop account (e.g. de Haan and Tuijmman 1988; Hyams and Wexler 1993). Like the other grammatical accounts, the topic-drop account predicts two distinct stages of argument omission: the first in which all and only topics are omitted, and the second in which no arguments are omitted. However, as we noted above, no such distinct stages are evident in the English data. The topic-drop account does, however, predict which of various arguments in the first stage should be
omitted and which should not be an advance over earlier grammatical accounts. In this domain, the discourse-pragmatic account could be seen as a careful elaboration of the topic-drop account. Although the latter states that topics will be omitted, it does not typically specify in much detail what it means in concrete terms to be a topic, nor what factors go into a child's initial understanding of what a topic is. The discourse-pragmatic account outlined here (and in Clancy 1993, 1997) does exactly that. However, the discourse-pragmatic account does not share with the topic-drop account the idea that children's early omission of arguments results from a non-adult-like parameter setting.

Finally, the discourse-pragmatic account is perhaps complementary to processing accounts (e.g. Bloom 1990), since both employ probabilistic explanations of the pattern of argument omission in child language. For instance, it may be that children initially omit subjects in their speech because of processing constraints, but that they make an extra effort to produce subjects in just those cases in which the subject is informative. Thus, in just those cases where the processing account seems to fall short, the discourse-pragmatic account provides an explanation. The interaction of these two accounts requires further research but seems a promising direction to pursue.

7. Conclusions

This paper began with the hypothesis that discourse-pragmatic features of informativeness could provide an explanation for the choice of Inuit children aged 2;0-3;6 to omit arguments (verbal cross-referencing affix only) or to represent them overtly (demonstrative or lexical NP). Eight features of informativeness were investigated: ABSENCE, CONTRAST, DIFFERENTIATION IN CONTEXT, DIFFERENTIATION IN DISCOURSE, INANIMACY, NEWNESS, QUERY, and THIRD PERSON. Through logistic regression analysis, five of these features were shown to have a significant effect: ABSENCE, CONTRAST, DIFFERENTIATION IN CONTEXT, NEWNESS, and THIRD PERSON. Including the informativeness features as predictors in the logistic regression model substantially increased the power of the model to predict which arguments in the child data would be overt. Results were similar whether all arguments were considered in the analysis, or only third person arguments. Various possibilities for future research were outlined to address the question of why the model with informativeness features does not have 100% predictive power. Nevertheless, we conclude that at least five of the informativeness features initially investi-
4. I wish to thank Melissa Bowerman for drawing the idea of search-space-size to my attention.

5. The term verbal cross-referencing affix is used here to denote the affixes on the verb that indicates information about the subject and object of the verb. This affix has been called "verbal inflection," "verbal agreement," and "pronominal argument" among other things, depending on the theoretical persuasion of the writer. One issue relevant to appropriate terminology is whether the verbal cross-referencing affix is actually the argument itself, with independent (lexical, demonstrative, or pronominal) representations of the relevant referent serving as adjectives, or whether the verbal cross-referencing affix is really an agreement marker, while independent representations of the relevant referent are the true arguments. This issue has not been resolved for Inuktitut, and thus a relatively neutral terminology is used here.

6. These and all other names of Inuktit used in this chapter are pseudonyms used to protect the identity of the subjects.

7. The following grammatical abbreviations are used in glosses:

   Nominal case:
   ABV = absolutive
   ALL = allative
   ERG = ergative
   LOC = locative
   MOD = modal
   VIA = violate

   Verbal modality:
   CSV = causative
   CTM = contemporative
   ICM = incontemporative
   IMP = imperative
   IND = indicative
   INT = interrogative
   PAR = participial (functionally equivalent to indicative in Tannantat)

   Word internal morphology:
   ALT = antipassive
   CAUS = causative
   DIM = diminutive
   EMPH = emphatic
   FUT = future
   HAB = habitual
   INCP = inceptive aspect
   NEG = negative
   PASS = passive
   PAST = past
   POL = poleness (preceeding imperative)
   PRSP = prospective aspect

   Verbal inflection (e.g. PAR, 3sS):
   1 = first person
   2 = second person
   3 = third person
   s = singular
   p = plural
   5 = subject
   O = object

   Nominal inflection (e.g. ABS, SG):
   SG = singular
   Possessive nominal inflection (e.g. ERG, 3SG):
   1 = first person possessor
   2 = second person possessor
   3 = third person possessor
   s = singular possessor
   sg = singular possessum

8. It is a matter of some controversy in the literature on Inuit languages whether or not the antipassive clause is semantically equivalent to the standard transitive clause. Billion (1987) shows for West Greenlandic that antipassive objects tend to be indefi- nite, while standard transitive objects tend to be definite, and suggests that the antipas- sive construction in Inuktitut is similar to the English conative construction. Kulmik (1979), on the other hand, claims that the antipassive construction in North Baffin Inuktitut is a nominative-accusative variant of the ergative absolutive transitive con- struction, and that the two structures are semantically equivalent. Allen and Schröder (1980) also suggest that the antipassive construction in the Tannantat dialect of Inuktitut is semantically equivalent to the standard transitive construction (sic perhaps even in the process of becoming the new standard transitive construction), and that it is used frequently in this dialect to avoid having third person arguments in the subject position of a grammatically transitive clause.

9. Although the terms null and overt as used here implicitly assume that the verbal cross- referencing affix is an agreement marker and the independent NP is the actual argument (see note 5 above), these terms are used only to be consistent with the majority of the literature on this topic and are not intended to indicate a particular position on whether or not the verbal cross-referencing affix in Inuktitut is actually a pronominal argument.

10. Only three instances of first person pronouns in argument position (one of second person) were found in the data set examined in this paper (Allen and Schröder, 1983). However, frequent use of first and second person pronouns in argument position has been observed in one specifically language impaired Inuktitut speaking child aged 5.0 (Crookes and Allen 1996). Her family found this practice so unusual that they nicknamed her "reversalopath." "cute little koyuk." Note that Yup'ik, another language in the Eskimo-Aleut family, does permit use of first and second person pronouns in argument position (Jacobson 1984).

11. Included are 1909 subjects of intransitive clauses; 119 subjects of transitive clauses, and 1627 subjects of transitive clauses, for a total of 3175 arguments. The 146 objects of antipassive clauses were not included in this analysis, since they are considered grammatically to be oblique rather than true subject or object arguments.

12. Note that clauses whose main verbal component is a word that begins with a verbal root, but ends as a nominalized construction were not included in the analysis.

13. Included in the analysis are 2555 arguments whose verbal clauses contain both a verb root and a verbal cross-referencing affix, 402 arguments whose verbal clauses lack a verb root, 161 arguments whose verbal clauses lack a verbal cross-referencing affix, and 47 arguments whose verbal clauses lack both a verb root and a verbal cross-referencing affix, for a total of 3175 arguments.
15. The term "overt copular verb" reflects the syntactic function of this morpheme from a generative perspective. Readers from other theoretical perspectives may prefer the term "derivational posibleness" with verbal function.

16. The term "existential clause" follows usage in Clancy (1984). Note that this type of clause focuses on location rather than existence.

17. Note that these variables are not necessarily unrelated to each other. For example, the three variables CONTRAST, DIFFERENTIATION IN CONTEXT, and DIFFERENTIATION IN DISCOURSE often occur together in some combination. In addition, only arguments representing third person referents can have an informative value for any of the other informativeness features except CONTRAST. These two relationships are discussed in subsequent sections. Finally, many arguments have an informative value for more than one feature for other reasons; for example, ABSENCE and NONSENSE often occur together.

18. The odds of an event indicates how often the target event occurs in comparison with its counterpart. For example, the probability of getting a 6 if you roll a die is 1/6, and the odds of getting an even number is 1:5. In Table 4, line 1, 69 out of 189 arguments with an informative value for ABSENCE are overt. The proportion of overt arguments is thus 0.37, and the odds of an overt argument being produced are 69:120, meaning that 69 arguments are overt and 120 are not. The odds ratio by which multiple predictor variables increase the odds increases. For example, in Table 6, the odds of an argument being overt becomes 15.88 times as large if the argument has an informative value for THIRD PERSON than if it has an informative value for THIRD PERSON. Note that coefficients are interpreted in the context of the other predictor variables, such that the probability of an overt argument as a function of an informative value for THIRD PERSON is interpreted after adjusting for all other predictors.

19. Note that this is the effect of the analysis to third person arguments. The part of the constant in Table 13 (0.6454) is simply the result of adding the B-weight of THIRD PERSON in Table 6 (2.7654) to the B-weight of the constant in Table 6 (−2.0057).

20. We also residualized both the sequential logistic regression analysis and the logistic regression analysis restricted to third person arguments discussed in the previous sections, using CHILD as a variable. Again, the results differed little except for the value of the constant.

21. In all but one of these cases, the same two features (QUERY and ABSENCE) are uninformative. If we compute the formula in (18) including these two features as uninformative and the other six features as informative, we get the result that 91.1% of arguments are expected to be overt (compare with 95.2% in [20b]).

References


520 S. E. M. Allen


Swift, Mary; and Allen, Shanley E. M. (n.d.). On the alienability of verb parts in Inuktitut. Unpublished manuscript, Max Planck Institute for Psycholinguistics and Boston University.

