Nature or nurture: structured information knowledge*

Cécile F. van der Weert

School of Linguistics & Applied Language Studies, The University of Reading

Abstract. This paper investigates the innateness of discourse knowledge, or structured information knowledge. Building on the Innateness Hypothesis of Crain and Thornton (1998), it is shown that evidence for a poverty of stimulus exists at the discourse level. It is therefore proposed that linguistic principles operating at this level could be innate and part of Universal Grammar. Three discourse rules from Avrutin (1999) are tested on three-year-old children to investigate their knowledge of the principles which function across a sentence boundary. The first one controls reference of an indefinite DP to an indefinite antecedent DP in a previous sentence. The other two rules control reference of definite DPs to indefinite antecedents in a previously mentioned sentence. It is proposed that instead of these three rules of discourse reference, only one innate principle is involved, the *Principle of Definiteness*. This principle controls the interpretation of definite DPs and could be innate and hard-wired linguistic knowledge. This innate principle is complemented by a pragmatic rule, which is not genetically determined and accounts for the interpretation of indefinite DPs.

1. Introduction

This paper investigates whether principles controlling inter-sentential reference are part of the innate linguistic knowledge available to human beings. When linguistic information is interpreted, it is arranged on 'file cards' (Heim 1982; Avrutin 1999; 2001). File cards are constituents of information structures. The internal organization of information structures is planned by three rules (Avrutin 1999).

(1) RULE 1. Instantiate the variable index of an indefinite DP with a number of a new file card.

* I am greatly indebted to Spyridoula Varlokosta and David Lightfoot for their invaluable support. I thank all the children and staff at the Forbury Gardens Day Nursery and Northumberland Day Nursery, especially deputy-nursery manager Michelle Edmunds. Naturally, I am responsible for this final product. The NUFFIC, the ESRC and the University of Reading have funded this research.

Reading Working Papers in Linguistics 6. 119-132.

(2) RULE 2. Instantiate the variable index of a definite DP with a number of an old file card.

(3) RULE 3. Instantiate the variable index of a definite DP with a number of a new file card only if this card can be bridged to another one.

These three rules control reference between on the one hand file cards containing definite Determiner Phrases (DPs) (such as pronouns and possessives) and, on the other hand, indefinite DPs. Rule 1 entails that indefinite DPs cannot refer back to a previously-mentioned DP, whereas Rule 2 demands that definiteness needs a previously mentioned DP referent in the discourse. Rule 3 controls 'bridging', where a definite DP is allowed to refer back to a DP containing a different noun, in a previously mentioned sentence.

For example, in an information sequence such as (4), it is required that the pronoun 'he' refers to the man in the preceding sentence rather than any other man in the mind of the speaker/hearer (Rule 2). In the example in (5), however, the indefinite DP in the second sentence is not allowed to refer back to the previously mentioned DP (Rule 1). The third example shows bridging of a different DP, 'the sweater', to a previously mentioned DP, in this case 'a man'. This is allowed because of Rule 3.

(4) A man is sitting by the window. He is wearing a pink sweater.

(5) A man is sitting by the window. A man is wearing a pink sweater.

(6) A man is sitting by the window. The sweater is pink.

The aim of this study is to see whether these rules of (in-) definiteness could be innate principles operating in the linguistic module in the human mind. Adults and children are tested to investigate whether their knowledge regarding the referential properties of definite and indefinite DPs corresponds to these rules.

2. Background

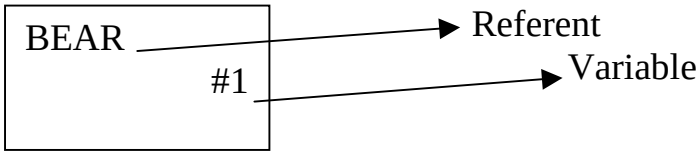
When linguistic information is interpreted, it is arranged following the file change semantics model, as it was first presented by Heim (1982) and has

since been further developed by Avrutin (1999; 2001). This model uses file cards, which are constituents of information structure and serve to keep track of information. File Cards are needed for the interpretation of language since without them, hearers cannot keep track of the different objects in the discourse. The information of the file cards is structured so that one can keep track of referents.

When an information sequence such as (7) is interpreted, a new file card with a new variable is created when the DP 'a bear' is encountered, as in (8).

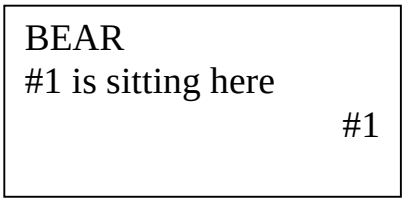
(7) A bear is sitting here. He is happy. The bear is eating a banana. (…)

(8)



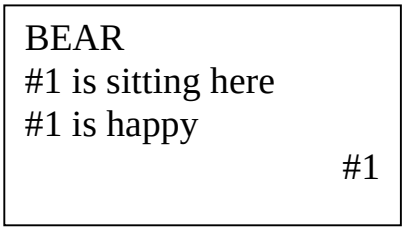
The variable – in this case #1 – is then linked to the created referent, which is the 'bear' on this file card, and this variable is then linked to the first predicate, resulting in the information '#1 is sitting here' being put on the first file card, as in (9).

(9)



Thus, a new file card with a variable was created because an indefinite DP was encountered – following Rule 1. However, when 'he' is heard, Rule 2 says that because this DP is definite, it has to be incorporated in an existing file card. This is not problematic here, because a file card with the needed referent exists, and the process of incorporation follows, as in (10).

(10)



At this point, when the third sentence ‘the bear is eating a banana’ in this discourse is heard, the information is incorporated into the existing file card. When a new indefinite DP is encountered, (in 11), the process of interpreting this indefinite DP follows Rule 1 again and thus a new file card is created, resulting in (12).

(11) (…) And a duck is swimming.

(12)

BEAR
#1 is sitting here
#1 is happy
#1 is eating a banana
#2

DUCK
#2 is swimming
#2

What the organisation on different file cards achieves is that the object referred to on the #2 file card – the duck – is not the same object as the one referred to on the #1 file card – the bear. Along the same lines, when two of the *same* indefinite DPs are used, the indefinite rule makes sure that we interpret them as two different objects. For example:

(13) A girl is skipping. And a girl is watching.

In this information structure, Rule 1 ensures that the second girl is a different one from the first girl. That is, the hearer will interpret this information as referring to two girls, one who is skipping, and another one who is watching. And crucially, the speaker did not intend this discourse to mean that there is one girl who is skipping *and* watching. This is the effect that Rule 1 has on the interpretation of this discourse.

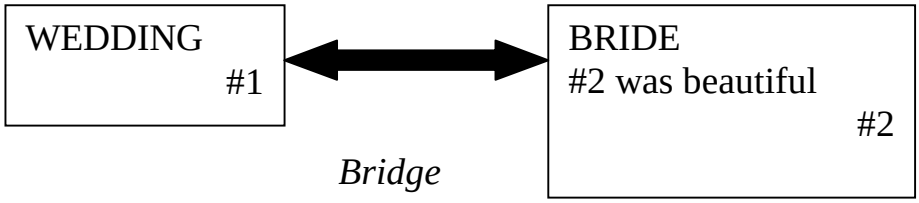
However, there are some other problematic discourses. For example, when a definite DP is encountered, but a file card with the needed referent does not exist, as in (14).

(14) There was a wedding. The bride was beautiful.

First, a new file card with wedding is created. But when ‘the bride’ is encountered, Rule 2 requires the definite DP to be incorporated in an existing file card. However, since there is no file card available with the same referent as ‘the bride’, Rule 3 comes into play. This rule says that a

variable can only be created on a new file card when this file card can then be bridged to an existing file card. This is shown in (11).

(15)



A new file card is created but a link is established to an existing file card. This is how the rules of file change semantics control the processes of interpretation (Avrutin 1999).

3. Data

Even though it is now established what the theory behind the interpretation of definite and indefinite DPs in inter-sentential structures entails, it is not clear whether this is actually what language users do when they interpret linguistic information. To determine this, linguistic knowledge in adults and children will be tested in an experimental set up. The two different types of data, i.e. indefinite and definite DPs, will be used to test the three rules. The template structures for indefinite DPs are presented in (16) and (17).

(16) A tree is standing on grass. And a tree has yellow leaves.

(17) A girl is playing. And a dress is blue.

The sentence sequence in (16) is the one discussed before. In this one, the secondly-mentioned tree is a different one from the previously mentioned one, following Rule 1. There is, however, another type of indefinite, as in (17), where in fact a definite DP in a bridging construction has been changed into an indefinite DP. This change has the effect that, following Rule 1, the dress should not refer to the girl's, but to another person's dress in the physical environment or the minds of the hearer and speaker.

The following constructions in (18), (19), (20) and (21) were used to test the definite DPs.

(18) A man is driving a car. *It is a beautiful day.* And the man is waving.

(19) A girl is playing. And the dress is blue.

(20) A blue car is driving by. And it has black windows.

(21) A cat is sleeping. And its tail is black.

Number (18) is a sentence with a definite DP referring back to an indefinite DP. The sentence 'it is a beautiful day', was inserted to make the use of the definite DP more natural here. If the structure would have been "a man is driving a car. And the man is waving", relevance knowledge would interfere (Sperber and Wilson 1986/1995). Subjects would wonder why the speaker had not used a pronoun 'he', since this is more economical. Their relevance knowledge would make them think that the reason for not using the pronoun was that the speaker was in fact talking about another man. In pilot experiments, it became evident that this interfered with the interpretation the definite DP received, leading to a much lower percentage of expected responses (van der Weert, 2001a). When a sentence like 'it is a beautiful day' is inserted between the two sentences, this creates a distance between the two DPs which makes the use of the definite article natural. Care was taken that there was no possible referent in the inserted sentence.

Number (19) is a bridging condition, where a different noun is used in the DP, but because it is definite, it is still interpreted as referring back to the previously mentioned girl. Number (20) represents the pronoun version of the definite DP. Following Rule 2, 'it' is interpreted as referring to the previously mentioned "blue car". Number (21), finally, is the definite DP version with a possessive DP. Rule 2 works in the same way here: 'its tail' is interpreted as being the tail of the previously mentioned cat rather than any other cat, or even any animal.

Note that the insertion of the word 'and' is used to make the two sentences sound more natural. A small-scale experiment with adults was conducted which showed that they have a strong preference to interpret the word 'and' in a structure like this as being contrastive. This should then provide a bias against the definite test sequences. This is important because the results of the pilot experiments showed that both children and adults scored very high on the definite test structures, but poorly on the indefinite ones.

The hypothesis is based on the requirements of innateness set forth in Crain and Thornton (1998). First, if the rules operating at the inter-sentential level are innate, then it should be investigated whether cross-linguistic evidence is present. That is to say, whether the property or rule is found in different languages. Secondly, evidence should be found that the principle or property is early emerging because if it is, this gives us more evidence for innateness. More importantly, the principles at the discourse level should be subject to the poverty of stimulus argument (Crain and Thornton 1998; Lightfoot 2000). In this paper, evidence for the second and third clause is presented.

In order for a poverty of stimulus argument to exist, it has to be true that there is no negative evidence, for a principle/property.[1] Negative data is data which informs a learner that a certain structure is *not* possible (Crain and Thornton 1998). In a discourse like (18), the pronoun 'she' is understood as referring to the girl mentioned (one with black hair) not to any other girl in the mind of the speaker and hearer.

(22) There was a girl with black hair. And she was wearing a pink ballet suit.

Any hearer of this utterance will know that the pronoun 'she' only refers to the girl with black hair and a pink ballet suit. This is a constraint on the interpretation of the pronoun: it can only refer back to a previously mentioned DP. When children learn as their native language, they have to learn all the properties of a language including constraints such as the above-mentioned.

For children to learn that there is such a constraint, they would have to use negative evidence. That is, evidence that 'she' in this discourse cannot refer to anyone but the girl mentioned before. People in the environment of the child may give the child unsystematic corrections on their linguistic performance (even though the child might not use them) but this will rarely involve telling a child that a certain interpretation (which the child does not express) in the child's mind is incorrect. Even that claim that parents correct the 'falseness' of their children's statements is redundant when not every child receives *systematic* feedback in such manner.

[1] Positive data has been shown to have very little or no effect on the language acquisition by the child (Brown and Hanlon 1970; Bowerman 1987; 1988; Morgan and Travis 1989; Marcus 1993) and is therefore ignored here.

Crucially, the chances of this negative input being given to the child are small. Moreover, the input is not systematic. Therefore, it is not obvious how it could aid children in learning this constraint. Therefore, the child is left without negative evidence in the input for that constraint, and crucially, nothing prevents the child from interpreting 'she' as referring to someone else than the girl with black hair. Nothing informs the child that any other interpretation (but the target-one) is *excluded*. Thus, because there is a lack of negative data, a poverty of stimulus exists at the information structure level (see also van der Weert 2001b). Therefore, I argue that there are principles at this level which are necessarily innate. The task now is to find out what the principles are.

4. Experimental evidence: adults

In the first study, the truth value judgement task (Crain and Thornton 1998) with pictures was used. In this task, subjects judge a description of a picture as being right or wrong, thereby revealing the interpretation they have given to the linguistic utterance. The adults tested, performed 90% correct or higher in the conditions which tested reference of the definite DPs.

However, an unexpected result came up with conditions with indefinite DPs which are not allowed to refer back to a previously mentioned DP. More specifically, when looking at a picture like (23), the subject accepted the sentence sequence (24) 96.5% of the time. Essentially, adults interpreted (24) as two separate statements, which are unrelated and can therefore individually refer to the same man in the picture.

(23)

(Left: red top)
(Right: blue top)

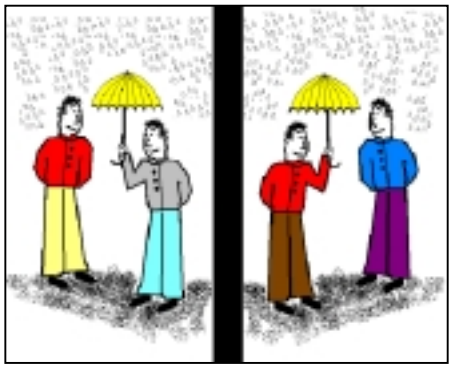
(24) A man is holding an umbrella. A man is wearing a red coat.

This is a "side effect" of the method. The presentation of only one picture may have allowed the subjects to interpret the information at the syntactic

level, leading them to interpret the discourse as two separate statements. This then resulted in subjects accepting utterances which they would have rejected had they not had the option of interpreting them as two separate statements. It then follows that interpreting the utterance as two separate statements is preferred because it makes the utterance true for this picture. (Note that in this first set of experiments, the conjunction 'and' was not used).

Because of these unexpected results, the same experiment was done using a different method, namely the picture selection task. In this task, subjects were presented with two pictures and had to select one as the appropriate description for the sentence sequence they were hearing. For example, subjects heard utterance (24) when presented with the pictures in (25).

(25)



(Far left jumper=red; left jumper=grey; far right jumper=blue; right jumper=red)

The picture on the left represents the target-interpretation where the indefinite DP in the second sentence is not allowed to refer back to the previously mentioned DP. The other picture, on the right, shows the non-target interpretation, where the "incorrect" referent can be chosen.

The advantage of the picture selection task is that the subjects are not forced to accept a certain interpretation that they might not find acceptable in a natural conversational situation. The disadvantage of this method, however, is that when this task is used, the results may show a preference rather than an innate principle.

Sixteen adult subjects were tested individually, using a laptop which showed the pictures. When the subjects would press the space bar, the sentences would be heard on their headphones. If the subject thought the left picture was the one which suited the information best, then he or she would press the z - button, which is on the left side of the keyboard. And if the subject though the right one was the matching one, they would press the / - button, which is on the right side of the keyboard. There were two

trials for every condition. One trial of each condition is presented in Table 1. For every trial, three fillers were used.

RULE 1	A castle has a flag. And a castle has a window.	98%
” ”	A woman is sleeping. And a hat has fallen off.	46%
RULE 2	A boy is playing. And he is kicking a ball.	97%
” ”	A man is driving a car. And the man is wearing a hat.	91%
” ”	A cat is sleeping. And its pillow is red.	100%
RULE 3	A tree is yellow. And the leaves have fallen off.	88%

Table 1 Correct responses for adults in percentages

The first two are the ones where Rule 1 was tested, using indefinite DPs. Whereas the adults only performed correctly 3.5% of the time on this condition in the truth value judgement task, they now scored almost 100%. This means they attend to Rule 1.

In the next condition however, their performance was only 46% correct. This may imply that the rule does not have any effect when two different nouns are used. The rest of the conditions tested Rule 2, using the definite DPs, and the results are as expected. Note that the results of correct responses for the bridging condition are slightly on the low side, but this can be explained by processing complexity, as proposed in Avrutin and Coopmans (2000) and van der Weert (1999).

5. Experimental evidence: the children

Since one of the conditions for innateness is that there is early emergence of the property, 33 children at two nurseries were tested. Their age range was between 2 years and 11 months and 4 years and 11 months, with an average age of 3 years and 11 months. This time, there were 7 to 9 trials per condition, and all trials were divided over 6 sessions. As in the previous test with adults, for every trial, three fillers were used. The task was exactly the same, except that I would be present pressing the buttons for the children. They did not have to wear headphones because the sound came from the speakers on the laptop.

When looking at the correct results of the children in Table 2, a strong discrepancy between the trials which tested Rule 1, and the ones which tested Rule 2 and 3 can be observed.

RULE 1	A castle has a flag. And a castle has a window.	24%
” ”	A woman is sleeping. And a hat has fallen off.	15%
RULE 2	A boy is playing. And he is kicking a ball.	87%
” ”	A man is driving a car. *It is a beautiful day.* And the man is wearing a hat.	79%
” ”	A cat is sleeping. And its pillow is red.	86%
RULE 3	A tree is yellow. And the leaves have fallen off.	91%

Table 2 Correct responses for child subjects in percentages

The first test condition was the one in which the adults scored 98% correct. Children here have a strong preference for the opposite answer than Rule 1 prescribes. In the second indefinite version, this is even lower: children only scored 15% correct here. However, it was shown that adults only scored 46% correct on that condition too. The overall results of the definite versions were close to target-result except for the interference of noise.[2]

6. Discussion

The average results of the correct answers in the conditions testing the three rules was 19.5% for Rule 1, 84% for Rule 2 and 91% for Rule 3. Clearly there is a factor which is relevant for Rule 2 and 3 but not for Rule 1. That factor is definiteness. There is no other division between the three rules, except the one between the definite and indefinite versions. It appears that children know ‘definiteness’, but not ‘indefiniteness’. Moreover, even though the picture selection task helped getting expected results for all adult results, the truth value judgement task showed that there was even a difference for adults between the indefinite and definite conditions.

I therefore propose that instead of three rules, there is only one principle, namely one which defines the interpretation of definiteness. This, I propose, is complemented by a pragmatic rule which describes what the default process of interpretation is in the case of indefiniteness.

(26) Principle of Definiteness:
A definite DP refers to a previously mentioned DP.

[2] In this research framework, noise is caused by temporary lack of attention and should not exceed 10-15% (Crain and Thornton, 1998).

(27) Pragmatic rule:
An indefinite DP introduces a new File Card and the DP on this new File Card does not refer to a previously created File Card.

I propose, based on the evidence presented here, that the principle of definiteness is a principle which both adults and children know, and therefore it may be innate. The pragmatic rule defines the way reference is interpreted for indefinite DPs and its status is different than the status of the definiteness principle. The pragmatic rule does not have the status of a principle which is part of the grammatical hard-wired knowledge we have, but rather it is a rule that we abide by when we use language. In Chomsky's terms, this is part of our performance rather than our competence. Thus, this is the part which adults know, but children have to learn.

However, an immediate question arises now: what about the poverty of stimulus argument? Was it not argued that there was no negative data for children to learn a discourse principle? Let us return to that argument. In a discourse like (28), the pronoun 'she' is understood as referring to the girl mentioned – one with black hair, not to any other girl in the mind of the speaker and/or hearer.

(28) There is a girl with black hair. And she is wearing a pink ballet suit.

However, if we substitute the definite DP for an indefinite one, the rule doesn't necessarily apply, see (29).

(29) There is a girl with black hair. And there is a girl who is wearing a pink ballet suit.

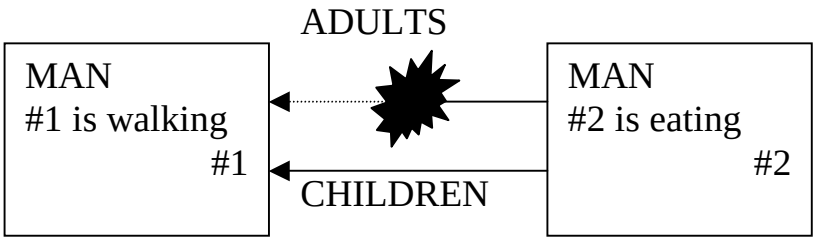
Perhaps this sequence has a preferred interpretation in which there are two girls, but the interpretation where there is one girl with both black hair and a pink ballet suit is not excluded. As was also shown by the experimental data from the truth value judgement test, there may not be a rule which says that the second DP cannot refer back to the first one because sometimes it can. Therefore, the question in the case of indefinite DPs is not, is it innate, but rather, what happens when subjects interpret the indefinite DP.

If we return to the rules in File Change Semantics, and how Avrutin (1999) formulated them, perhaps we can explain what happens in the children's interpretations. Perhaps children know that the indefinite DP

creates a new file card, but there are no restrictions for them on where to find the referent for the new file card. Therefore, they can find the referent deictically, that is to say, in either of the pictures.

Accordingly, adults know that after they create a new file card for an indefinite DP, the file card does not need external reference. However, children do not know this and after they have created a file card for the indefinite DP, they try to find a referent. This process is described in (30).

(30)



What this means is that to learn the process of the pragmatic rule is not a matter of learning when the creation of a new file card is appropriate, but of when the creation of links – or perhaps 'inferences' – between file cards is appropriate.

7. Conclusion

This paper shows that principles which control inter-sentential reference are part of the internal and possibly innate linguistic knowledge in human beings. It has been shown that one principle of definiteness is part of adults' and children's linguistic knowledge and could be innate, whereas another rule controlling indefinite reference has to be acquired by the child.

Evidence for a poverty of stimulus exists at the inter-sentential or discourse level and the data presented here shows that the principle of definiteness is known at an early age (3 years). Further research may find cross-linguistic data to show universal knowledge of this principle, but it is clear at this point that these results provide evidence that the principle of definiteness could be a genetically determined property of humans.

References

Avrutin, S. (1999). *Development of the syntax-discourse interface*. Dordrecht: Kluwer.

Avrutin, S. and P. Coopmans (2000). Children who build bridges. In S. Catherine Howell et al. (eds.) *BUCLD 24 Proceedings*. Somerville, MA: Cascadilla Press. 80-91.

Avrutin, S. (2001). Development of the syntax-discourse interface. Paper presented at LOTS Summer School.

Bowerman, M. F. (1987). Commentary: mechanisms of language acquisition. In B. MacWhinney (ed.) *Mechanisms of language acquisition*. New York: Springer-Verlag.

Bowerman, M. F. (1988). The no negative evidence problem: how do children avoid constructing an overly general grammar? In J. A. Hawkins (ed.) *Explaining language universals.* Oxford: Blackwell.

Brown, R. and C. Hanlon (1970). Derivational complexity and order of acquisition in child speech. In J. Hayes (ed.) *Cognition and the development of language*. New York: Wiley.

Crain, S. and R. Thornton (1998). *Investigations into universal grammar: a guide to experiments on the acquisition of syntax and semantics.* Cambridge, Massachusetts: MIT Press.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases*, Ph.D. dissertation, University of Massachusetts, Amherst. Published: New York: Garland Press.

Lightfoot, D. W. (2000). Plato's Problem, UG and the Language Organ. *University of Maryland Working Papers in Linguistics* **9**. 154-171.

Marcus, G. (1993). Negative evidence in language acquisition. *Cognition* **46**. 53-85.

Morgan, J. L. and L. L. Travis (1989). Limits on negative information in language input. *Journal of Child Language* **16**. 531-552.

Sperber, D. and D. Wilson (1986, 1995). *Relevance: communication and cognition.* Blackwell Publishers.

van der Weert, C. F. (1999). *Syntax-discourse interface: an investigation into incorporation and accommodation in child language development*, unpublished MA thesis. Utrecht University.

van der Weert, C. F. (2001a). Native elements of discourse. Paper presented at Boston University, Linguistic Colloquium.

van der Weert, C. F. (2001b). Native elements of discourse knowledge. In Y. Otsu (ed.) *The Proceedings of the Second Tokyo Conference on Psycholinguistics*. Hituzi Syobo Publishing Ltd, Tokyo.