Quantifiers in TIME and SPACE: computational complexity of generalized quantifiers in natural language
Szymanik, J.K.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Chapter 6

Hintikka’s Thesis Revisited

This chapter is devoted to an investigation of potentially branched combinations of quantifiers in natural language. However, as opposed to the previous chapters we do not stay in the safe domain of theoretical discussion but confront our claims with empirical reality. In this way we are aiming to establish the referential meaning of some controversial sentences. As we have shown in Section 3.2, branching interpretations of some natural language sentences are intractable. Therefore, our quest in this chapter is to check whether in everyday language the intractable meaning of branching sentences is in fact realized.

We start by discussing the thesis formulated by Hintikka (1973) that certain natural language sentences require non-linear quantification to express their meaning. We also sum up arguments in favor and against this thesis which have appeared in the literature.

Then, we propose a novel alternative reading for Hintikka-like sentences, the so-called conjunctional reading. This reading is expressible by linear formulae and tractable. We compare the conjunctional reading to other possible interpretations and argue that it is the best representation for the meaning of Hintikka-like sentences.

Next, we describe the empirical support for the conjunctional reading. The basic assumption here is that a criterion for adequacy of a meaning representation is compatibility with sentence truth-conditions. This can be established by observing the linguistic behavior of language users. We report on our experiments showing that people tend to interpret sentences similar to Hintikka’s sentence in a way consistent with the conjunctional interpretation.

This chapter is based on joint work with Nina Gierasimczuk (see Gierasimczuk and Szymanik, 2006, 2007, 2008).
6.1 Hintikka’s Thesis

Jaakko Hintikka (1973) claims that the following sentences essentially require non-linear quantification for expressing their meaning.

(1) Some relative of each villager and some relative of each townsman hate each other.

(2) Some book by every author is referred to in some essay by every critic.

(3) Every writer likes a book of his almost as much as every critic dislikes some book he has reviewed.

Throughout the chapter we will refer to sentence (1) as Hintikka’s sentence. According to Hintikka its interpretation is expressed using Henkin’s quantifier (see Section 3.2 for an exposition on branching quantifiers) as follows:

\[
(\forall x \exists y \forall z \exists w) \left[ (V(x) \land T(z)) \implies (R(x, y) \land R(z, w) \land H(y, w)) \right],
\]

where unary predicates $V$ and $T$ denote the set of villagers and the set of townsmen, respectively. The binary predicate symbol $R(x, y)$ denotes the symmetric relation “$x$ and $y$ are relatives” and $H(x, y)$ the relation “$x$ and $y$ hate each other”. Informally speaking, the idea of such constructions is that for different rows the values of the quantified variables are chosen independently. According to Henkin’s semantics for branching quantifiers, formula (4) is equivalent to the following existential second-order sentence:

\[
\exists f \exists g \forall x \forall z \left[ (V(x) \land T(z)) \implies (R(x, f(x)) \land R(z, g(z)) \land H(f(x), g(z))) \right].
\]

Functions $f$ and $g$ (so-called Skolem functions) choose relatives for every villager and every townsman, respectively. As you can see, the value of $f$ ($g$) is determined only by the choice of a certain villager (townsman). In other words, to satisfy the formula relatives have to be chosen independently.\(^1\) This second-order formula is equivalent to the following sentence with quantification over sets:

\[
\exists A \exists B \forall x \forall z \left[ (V(x) \land T(z)) \implies (\exists y \in A R(x, y) \land \exists w \in B R(z, w) \land \forall y \in A \forall w \in B H(y, w)) \right].
\]

The existential second-order sentence is not equivalent to any first-order sentence (see the Barwise-Kunen Theorem in (Barwise, 1979)). Not only universal and existential quantifiers can be branched; the procedure of branching works in

\(^1\)The idea of branching is more visible in the case of simpler quantifier prefixes, like in sentence (5) discussed in Section 6.3.2.
a very similar way for other quantifiers (see Definition 3.2.2). Some examples — motivated by linguistics — are discussed in the next section of this chapter.

The reading of Hintikka’s sentence given by formula (4) is called the strong reading. However, it can also be assigned weak readings, i.e., linear representations which are expressible in elementary logic. Let us consider the following candidates:

\[(5) \forall x \exists y \forall z \exists w [(V(x) \land T(z)) \Rightarrow (R(x, y) \land R(z, w) \land H(y, w))],\]
\[\land \forall z \exists w \forall x \exists y [(V(x) \land T(z)) \Rightarrow (R(x, y) \land R(z, w) \land H(y, w))].\]

\[(6) \forall x \exists y \forall z \exists w [(V(x) \land T(z)) \Rightarrow (R(x, y) \land R(z, w) \land H(y, w))].\]

\[(7) \forall x \forall z \exists y \exists w [(V(x) \land T(z)) \Rightarrow (R(x, y) \land R(z, w) \land H(y, w))].\]

In all these formulae the choice of the second relative depends on the one that has been previously selected. To see the difference between the above readings and the branching reading consider a second-order formula equivalent to sentence (6):

\[\exists f \exists g \forall x \forall z [(V(x) \land T(z)) \Rightarrow (R(x, f(x)) \land R(z, g(x, z)) \land H(f(x), g(x, z))].\]

It is enough to compare the choice functions in this formula with those in the existential second-order formula corresponding to sentence (4) to see the difference in the structure of dependencies required in both readings. Of course, the dependencies in sentences (5) and (7) are analogous to (6). As a result all the weak readings are implied by the strong reading, (4) (where both relatives have to be chosen independently), which is of course the reason for the names. Formulae (5)-(7) are also ordered according to the entailment relation which holds between them. Obviously, formula (5) implies formula (6), which implies formula (7). Therefore, formula (5) is the strongest among the weak readings.

By Hintikka’s Thesis we mean the following statement:

**Hintikka’s Thesis** Sentences like Hintikka’s sentence have no adequate linear reading. In particular, Hintikka’s sentence should be assigned the strong reading and not any of the weak readings.

**6.1.1. Remark.** Let us stress one point here. Of course, every branching quantifier can be expressed by some single generalized quantifier, so in the sense of definability Hintikka’s thesis cannot be right. However, the syntax of branching quantification has a particular simplicity and elegance that gets lost when translated into the language of generalized quantifiers. The procedure of branching does not employ new quantifiers. Instead it enriches the accessible syntactic means of arranging existing quantifiers, at the same time increasing their expressive power. Therefore, the general question is as follows: are there sentences with...
simple determiners such that non-linear combinations of quantifiers corresponding to the determiners are essential to account for the meanings of those sentences? The affirmative answer to this question — suggested by Hintikka — claims the existence of sentences with quantified noun phrases which are interpreted scope independently. We show that for sentences similar to those proposed by Hintikka the claim is not true.

Because of its many philosophical and linguistic consequences Hintikka’s claim has sparked lively controversy (see e.g. Jackendo, 1972; Gabbay and Moravcsik, 1974; Guenthner and Hoepelman, 1976; Hintikka, 1976; Stenius, 1976; Barwise, 1979; Bellert, 1989; May, 1989; Sher, 1990; Mostowski, 1994; Liu, 1996; Beghelli et al., 1997; Janssen, 2002; Mostowski and Wojtyniak, 2004; Szymanik, 2005; Schlenker, 2006; Janssen and Dechesne, 2006; Gierasimczuk and Szymanik, 2006, 2007, 2008). Related discussion on the ambiguity of sentences with multiple quantifiers has been vivid in the more philosophically oriented tradition (see Kempson and Cormack, 1981a; Tennant, 1981; Kempson and Cormack, 1981b; Bach, 1982; Kempson and Cormack, 1982; May, 1985; Jaszczolt, 2002). In the present study some of the arguments presented in the discussion are analyzed and critically discussed. In particular, we propose to allow interpreting Hintikka’s sentence by the first-order formula (5):

\[
\forall x \exists y \forall z \exists w \left[ \left( V(x) \land T(z) \right) \implies (R(x, y) \land R(z, w) \land H(y, w)) \right]
\]

\[\land \forall z \exists w \forall x \exists y \left[ \left( V(x) \land T(z) \right) \implies (R(x, y) \land R(z, w) \land H(y, w)) \right].\]

In the rest of this chapter we will refer to this reading as the conjunctional reading of Hintikka’s sentence.

Our proposal seems to be very intuitive — as we show in the next section — and it is also consistent with human linguistic behavior. The latter fact is supported by empirical data, which we will present in Section 6.4.

Last but not least, it is worth noticing that our proposition is reminiscent of the linguistic representation of reciprocals. For example, according to the seminal paper on “each other” by Heim et al. (1991), Hintikka’s sentence has the following structure: EACH[[QP and QP] [V the other]; that is, “each” simply quantifies over the two conjuncts, which turns the sentence into [QP1 V the other and QP2 V the other], where “the other” picks up the other quantifiers anaphorically. This already resembles our conjunctional representation. See Chapter 4 for more discussion of reciprocity.

Our conclusion is that Hintikka-like sentences allow linear reading. This of course clearly contradicts Hintikka’s thesis.

### 6.2 Other Hintikka-like Sentences

Before we move on to the central problem let us consider more sentences with combinations of at least two determiners such that their branching interpreta-
6.3. Theoretical Discussion

6.3.1 A Remark on Possible Readings

Let us start with the following remark.

It was observed by Mostowski (1994) that from Hintikka’s sentence we can infer that:

(13) Each villager has a relative.

This sentence has obviously the following reading: $\forall x[V(x) \Rightarrow \exists y R(x, y)]$. It can be false in a model with an empty town, if there is a villager without a relative. However, the strong reading of Hintikka’s sentence (see formula (4)) — having the form of an implication with a universally quantified antecedent — is true in every model with an empty town. Hence, this reading of (13) is not logically implied by the proposed readings of Hintikka’s sentence.

Therefore, the branching meaning of Hintikka’s sentence should be corrected to the following formula with restricted quantifiers:

(14) $$(\forall x : V(x))(\exists y : R(x, y)) \land (\forall z : T(z))(\exists w : R(z, w)) H(y, w),$$
which is equivalent to:
\[
\exists A \exists B \left[ \forall x (V(x) \implies \exists y \in A \ R(x, y)) \land \forall z (T(z) \implies \exists w \in B \ R(z, w)) \land \forall y \in A \forall w \in B \ H(y, w) \right].
\]

Observe that similar reasoning can be used to argue for restricting the quantifiers in formulae expressing the different possible meanings of all our sentences. However, applying these corrections uniformly would not change the main point of our discussion. We still would have to choose between the same number of possible readings, the only difference being the restricted quantifiers. Therefore, for simplicity we will forego these corrections. From now on we will assume that all predicates in our formulae have non-empty denotation.

6.3.2 Hintikka-like Sentences are Symmetric

It has been observed that we have the strong linguistic intuition that the two following versions of Hintikka’s sentence are equivalent (Hintikka, 1973):

(1) Some relative of each villager and some relative of each townsman hate each other.

(15) Some relative of each townsman and some relative of each villager hate each other.

However, if we assume that formula (6), repeated here:
\[
\forall x \exists y \forall z \exists w [(V(x) \land T(z)) \implies (R(x, y) \land R(z, w) \land H(y, w))].
\]
is an adequate reading of sentence (1), then we have to analogously assume that an adequate reading of sentence (15) is represented by the formula:

(16) \[
\forall z \exists w \forall x \exists y [(V(x) \land T(z)) \implies (R(x, y) \land R(z, w) \land H(y, w))].
\]

However, (6) and (16) are not logically equivalent, therefore it would be wrong to treat them as correct interpretations of sentences (1) or (15). Therefore, we have to reject readings (6) and (16) from the set of possible alternatives.

Notice that a similar argument works when we consider other Hintikka-like sentences. For instance, it is enough to observe that the following sentences are also equivalent:

(8) Most villagers and most townsmen hate each other.

(17) Most townsmen and most villagers hate each other.

However, the possible linear reading of (8):
6.3. Theoretical Discussion

(18) Most $x [V(x), \text{Most } y (T(y), H(x, y))]$

is not equivalent to an analogous reading of (17). Hence, the linear reading (18) cannot be the proper interpretation.

In general, we are dealing here with the fact that the iteration operator (recall Definition 3.1.2) is not symmetric (see Section 3.1.1).

One of the empirical tests we conducted was aimed at checking whether people really perceive such pairs of sentences as equivalent. The results that we describe strongly suggest that this is the case. Therefore, the argument from symmetry is also cognitively convincing (see Section 6.4.4 for a description of the experiment and Section 6.4.4 for our empirical results).

In spite of this observation we cannot conclude the validity of Hintikka’s Thesis so easily. First we have to consider the remaining weak candidates, formulae (5) and (7):

$$(5) \forall x \exists y \exists w \left[ (V(x) \land T(z)) \implies (R(x, y) \land R(z, w) \land H(y, w)) \right]$$

$$\land \forall z \forall w \forall x \exists y \left[ (V(x) \land T(z)) \implies (R(x, y) \land R(z, w) \land H(y, w)) \right],$$

$$(7) \forall x \forall z \exists y \exists w \left[ (V(x) \land T(z)) \implies (R(x, y) \land R(z, w) \land H(y, w)) \right].$$

Hintikka does not consider them at all, and other authors focus only on formula (7) (see e.g. Barwise, 1979; Mostowski and Wojtyniak, 2004).

Also for different Hintikka-like sentences we still have to differentiate between some possibilities. As an alternative for formula (18) we can consider not only the branching reading (19) (equivalent to (20)):

$$(19) \left( \begin{array}{c} \text{Most } x : V(x) \\ \text{Most } y : T(y) \end{array} \right) H(x, y).$$

$$(20) \exists A \exists B \left[ \text{Most } x (V(x), A(x)) \land \text{Most } y (T(y), B(y)) \land \forall x \in A \forall y \in B \ H(x, y) \right].$$

but also the conjunctional meaning:

$$(21) \left( \begin{array}{c} \text{Most } x [V(x), \text{Most } y (T(y), H(x, y))] \land \text{Most } y [T(y), \text{Most } x (V(x), H(y, x))] \right].$$

Notice that for proportional sentences, e.g., (8), there is no interpretation corresponding to the weakest reading of Hintikka’s sentence, formula (7), as proportional sentences contain only two simple determiners, and not four as the original Hintikka’s sentence. This very fact already indicates that the conjunctional form — as a uniform representation of all Hintikka-like sentences — should be preferred over the weakest reading. Moreover, we report on another argument against the weakest reading (7) in the next section.

To sum up, the symmetricity argument rules out readings with asymmetric scope dependencies. At this point the adequacy of the weakest reading is also controversial since it is not uniform - it cannot be extended to proportional
sentences. Our space of possibilities contains now: the branching and the conjunctional reading. In the next section we give additional possible reasons to reject the weakest reading of Hintikka’s sentence. However, we do not take this argument as deciding. We tend to reject the weakest reading as a non-uniform interpretation.

6.3.3 Inferential Arguments

Let us move now to Mostowski’s (1994) argument against the weakest reading of Hintikka-like sentences.

Let us consider the following reasoning:

Some relative of each villager and some relative of each townsman hate each other.
Mark is a villager.
Some relative of Mark and some relative of each townsman hate each other.

In other words, if we assume that Mark is a villager, then we have to agree that Hintikka’s sentence implies that some relative of Mark and some relative of each townsman hate each other.

If we interpret Hintikka’s sentence as having the weakest meaning (7):

\[ \forall x \forall z \exists y \exists w [(V(x) \land T(z)) \Rightarrow (R(x, y) \land R(z, w) \land H(y, w))], \]

then we have to agree that the following sentence is true in Figure 6.1.

(22) Some relative of Mark and some relative of each townsman hate each other.

\[ \exists x [R(Mark, x) \land \forall y (T(y) \Rightarrow \exists z (R(y, z) \land H(x, z))]]. \]
The above formula \((23)\) is false in the model illustrated in Figure 6.1. Therefore, it cannot be implied by the weakest reading of Hintikka’s sentence which is true in the model. However, it is implied by the strong reading which is also false in the model. Hence, Mostowski concludes that Hintikka’s sentence cannot have the weakest reading \((7)\).

If we share Mostowski’s intuition, then we can conclude from this argument that the weakest reading, \((7)\), should be eliminated from the set of possible alternatives. Otherwise, we can refer to the non-uniformity problem. Then we are left with two propositions: the branching and the conjunctional interpretation. Both of them have the desired inference properties.

### 6.3.4 Negation Normality

Jon Barwise (1979) in his paper on Hintikka’s Thesis refers to the notion of negation normality in a defense of the statement that the proper interpretation of Hintikka’s sentence is an elementary formula. He observes that the negations of some simple quantifier sentences, i.e., sentences without sentential connectives other than “not” before a verb, can easily be formulated as simple quantifier sentences. In some other cases this is impossible. Namely, the only way to negate some simple sentences is by prefixing them with the phrase “it is not the case that” or an equivalent expression of a theoretical character.

Sentences of the first kind are called *negation normal*. For example, sentence:

\[
(24) \text{Everyone owns a car.}
\]

can be negated normally as follows:

\[
(25) \text{Someone doesn’t own a car.}
\]

Therefore, this sentence is negation normal. As an example of a statement which is not negation normal consider the following (see Barwise, 1979):

\[
(26) \text{The richer the country, the more powerful its ruler.}
\]

It seems that the most efficient way to negate it is as follows:

\[
(27) \text{It is not the case that the richer the country, the more powerful its ruler.}
\]

Barwise proposed to treat negation normality as a test for first-order definability with respect to sentences with combinations of elementary quantifiers. This proposal is based on the following theorem.

**6.3.1. Theorem.** If \(\varphi\) is a sentence definable in \(\Sigma^1_1\), the existential fragment of second-order logic, and its negation is logically equivalent to a \(\Sigma^1_1\)-sentence, then \(\varphi\) is logically equivalent to some first-order sentence.
Barwise claims that the results of the negation normality test suggest that people tend to find Hintikka’s sentence to be negation normal, and hence definable in elementary logic. According to him people tend to agree that the negation of Hintikka’s sentence can be formulated as follows:

\[(28) \text{There is a villager and a townsman that have no relatives that hate each other.}\]

6.3.2. REMARK. Notice that the test works only on the assumption that simple everyday sentences are semantically bounded by existential second-order properties; that is, exactly when we accept the $\Sigma_1^1$-Thesis formulated in Section 1.8.

Barwise’s claim excludes the branching reading of Hintikka’s sentence but is consistent with the conjunctional interpretation. Therefore, in the case of Hintikka’s sentence we are left with only one possible reading: the conjunctional reading — at least as far as we are convinced by the non-uniformity argument Mostowski’s and Barwise’s arguments. However, in the case of proportional sentences we still have to choose between the branching and the conjunctional interpretation.

6.3.5 Complexity Arguments

Mostowski and Wojtyniak (2004) claim that native speakers’ inclination toward a first-order reading of Hintikka’s sentence can be explained by means of computational complexity theory. The authors prove that the problem of recognizing the truth-value of the branching reading of Hintikka’s sentence in finite models is an NPTIME-complete problem. It can also be shown that proportional branching sentences define an NP-complete class of finite models (see Sevenster, 2006). See Section 3.2 for a discussion of the computational complexity of branching quantifiers.

Assuming that the class of practically computable problems is identical with the PTIME class (i.e., the tractable version of the Church-Turing Thesis; see Edmonds, 1965) they claim that the human mind is not equipped with mechanisms for recognizing NP-complete problems.\(^2\) In other words, in many situations an algorithm for checking the truth-value of the strong reading of Hintikka’s sentence is intractable. According to Mostowski and Wojtyniak (2004) native speakers can only choose between meanings which are practically computable.

6.3.3. REMARK. Notice that the argument is similar to the discussion from Chapter 4 where we were considering different interpretations of reciprocal sentences. According to the Strong Meaning Hypothesis the reading associated with

\(^2\)This statement can be given independent psychological support (see e.g. Frixione, 2001). We discuss an influence of computational complexity on cognitive abilities in Chapter 1.
the reciprocal in a given sentence is the strongest available reading which is consistent with the context (see Section 4.1.1). However, we have shown that in some cases the strong interpretation of reciprocal sentences is intractable and suggested in Section 4.5 that in such situations interpretation will shift to a weaker (tractable) inferential meaning.

The conjunctional reading is PTIME computable and therefore — even taking into account computational restrictions — can reasonably be proposed as a meaning representation.

6.3.6 Theoretical Conclusions

We discussed possible obstacles against various interpretations of Hintikka-like sentences. Our conjunctional version for Hintikka-like sentences seems to be very acceptable according to all mentioned properties. It is the only reading satisfying all the following properties:

- symmetry;
- uniformity for all Hintikka-like sentences;
- passing Mostowski’s inferential test;
- being negation normal for Hintikka’s sentence;
- having truth-value practically computable in finite models.

In the case of Hintikka’s sentence the conjunctional reading is arguably the only possibility. In general, for proportional sentences it competes only with the intractable branching reading. The next section is devoted to empirical arguments that the conjunctional reading is consistent with the interpretations people most often assign to Hintikka-like sentences.

6.4 Empirical Evidence

Many authors — taking part in the dispute on the proper logical interpretation of Hintikka-like sentences — have argued not only from their own linguistic intuitions but also from the universal agreement of native speakers. For instance, Barwise claims that:

In our experience, there is almost universal agreement rejecting Hintikka’s claim for a branching reading. \textit{(Barwise, 1979, p. 51)}

However, none of the authors have given real empirical data to support their claims. We confronted this abstract discussion with linguistic reality through experiments.
6.4.1 Experimental Hypotheses

Our hypotheses are as follows:

1. **Hypothesis.** *People treat Hintikka-like sentences as symmetric sentences.*

   This was theoretically justified in the paper of Hintikka (1973) and discussed in Section 6.3.2. To be more precise we predict that subjects will treat sentences like (29) and (30) as equivalent.

   (29) More than 3 villagers and more than 5 townsmen hate each other.

   (30) More than 5 townsmen and more than 3 villagers hate each other.

2. **Hypothesis.** *In an experimental context people assign to Hintikka-like sentences meanings which are best represented by the conjunctional formulae.*

   We predict that subjects will tend to interpret Hintikka-like sentences in the conjunctional way, i.e. they will accept the sentence when confronted with a model that satisfies its conjunctional interpretation. Arguments for that were given throughout the previous section and were summed up in Section 6.3.6.

   In addition we conduct a cross-linguistic comparison. We predict that the comprehension of Hintikka-like sentences is similar in English and Polish — in both languages native speakers allow the conjunctional reading.

6.4.2 Subjects

Subjects were native speakers of English and native speakers of Polish who volunteered to take part in the experiment. They were undergraduate students in computer science at Stanford University and in philosophy at Warsaw University. They all had elementary training in logic so they could understand the instructions and knew the simple logical quantifiers. The last version of the experiment, the one we are reporting on here, was conducted on thirty-two computer science students and ninety philosophy students. However, in the process of devising the experiment we tested fragments of it on many more subjects, getting partial results on which we reported for example at the Logic Colloquium 2006 (see Gierasimczuk and Szymanik, 2006, 2007).

The choice of students with some background in logic was made so that our subjects could be trusted to understand instructions which assume some familiarity with concepts of validity and truth. In that manner, we could formulate the task using such phrases as “one sentence implies the other”, “inference pattern is valid”, and “sentence is a true description of the picture”.
6.4.3 Materials

It was suggested by Barwise and Cooper (1981) and empirically verified by Geurts and van der Silk (2005) that the monotonicity of quantifiers influences how difficult they are to comprehend. Moreover, our results from Chapter 7 also suggest this dependency (see Section 7.4). In particular, sentences containing downward monotone quantifiers are more difficult to reason with than sentences containing only upward monotone quantifiers. Therefore, in the experiment — as we are interested rather in combinations of quantifiers than in simple determiners — we used only monotone increasing quantifiers of the form “More than” and their combinations in simple grammatical sentences. We used determiners, that are relatively easy to process, because we wanted our subjects to focus on combinations of quantifiers and not on individual ones.

In our tasks the quantifiers referred to the shape of geometrical objects (circles and squares). The sentences were Hintikka-like sentences (for the whole test in English see Appendix A and in Polish Appendix B). All sentences were checked for grammaticality by native speakers.

6.4.4 Experiments

The study was conducted in two languages and consists of two parts. It was a “pen and paper” study. There were no time limits and it took 20 minutes on average for all students to finish the test. Below we present descriptions of each part of the English version of the test. The Polish test was analogous (compare Appendices A and B).

Experiment I: Are Hintikka-like Sentences Symmetric?

The first part of the test was designed to check whether subjects treat Hintikka-like sentences as symmetric (see Section 6.3.2 for a discussion).

Let us recall the notion of symmetry for our sentences. Let $Q_1, Q_2$ be quantifiers and $\psi$ a quantifier-free formula. We will say that sentence $\varphi := Q_1 x \ Q_2 y \ \psi(x, y)$ is symmetric if and only if it is equivalent to $\varphi' := Q_2 y \ Q_1 x \ \psi(x, y)$. In other words, switching the whole quantifier phrase (determiner + noun) does not change its meaning.

We checked whether subjects treat sentences with switched quantifier prefixes as equivalent. We presented subjects with sentence pairs $\varphi, \varphi'$ and asked whether the first sentence implies the second sentence. There were 20 tasks. Ten of them were valid inference patterns based on the symmetry. The rest were fillers. Six were invalid patterns similar to the symmetric case. In three we changed the

---

3A quantifier $Q_M$ is upward monotone (increasing) iff the following holds: if $Q_M(A)$ and moreover $A \subseteq B \subseteq M$ then $Q_M(B)$. The downward monotone (decreasing) quantifiers are defined analogously as being closed on taking subsets. See Section 2.2.5.
Chapter 6. Hintikka’s Thesis Revisited

nouns following the quantifiers, i.e., we had $\varphi := Q_1 x \; Q_2 y \; \psi(x, y)$ and $\varphi' := Q_1 y \; Q_2 x \; \psi(x, y)$. In the second three we switched the determiners and not the whole quantifier phrases, i.e. $\varphi := Q_1 x \; Q_2 y \; \psi(x, y)$ and $\varphi' := Q_2 x \; Q_1 y \; \psi(x, y)$. Four of the tasks were simple valid and invalid inferences with the quantifiers “more than”, “all”, and “some”.

We constructed our sentences using nonexistent nouns to eliminate any pragmatic influence on subjects’ answers. For example, in the English version of the test we quantified over non-existing nouns proposed by Soja et al. (1991): mells, stads, blickets, frobs, wozzles, fleems, coodles, doffs, tannins, fitches, and tulvers. In Polish we came up with the following nouns: strzew, memniak, balbasz, protorozec, melarek, kretowiec, stular, wachlacz, fisut, bubrak, wypsztyk. Our subjects were informed during testing that they are not supposed to know the meaning of the common nouns occurring in the sentences. Therefore, subjects were aware that they have to judge an inference according to its logical form and not by semantic content or pragmatic knowledge.

Figure 6.2 gives examples of each type of task in English.

<table>
<thead>
<tr>
<th>More than 12 fleems and more than 13 coodles hate each other.</th>
<th>More than 13 coodles and more than 12 fleems hate each other.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VALID</td>
<td>NOT VALID</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>More than 20 wozzles and more than 35 fitches hate each other.</th>
<th>More than 20 fitches and more than 35 wozzles hate each other.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VALID</td>
<td>NOT VALID</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>More than 105 wozzles and more than 68 coodles hate each other.</th>
<th>More than 68 wozzles and more than 105 coodles hate each other.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VALID</td>
<td>NOT VALID</td>
</tr>
</tbody>
</table>

Some tulvers are mells.
Some mells are tulvers.

VALID                                                   NOT VALID

Figure 6.2: 4 tasks from the first experiment: symmetry pattern, two invalid patterns and a simple inference.

We excluded the possibility of interpreting the sentences as being about the
relations between objects of the same kind (e.g., “68 coo dles hate each other”) by explicitly telling the subjects that in this setting the relation can occur only between objects from two different groups.

Results  We got 90% correct answers in the group consisting of philosophy undergraduates at Warsaw University and 93% correct answers among Stanford University computer science students, where by “correct” we mean here “correct according to our prediction about symmetricity”. In simple inferences we got 83% ($\chi^2=153.4$, df=1, $p<0.001$) and 97% ($\chi^2=110.63$, df=1, $p<0.001$), respectively. In valid symmetricity tasks the result was: 94% ($\chi^2=709.33$, df=1, $p<0.001$) and 98% ($\chi^2=286.90$, df=1, $p<0.001$), and in invalid symmetricity inferences: 86% ($\chi^2=286.02$, df=1, $p<0.001$) and 93% ($\chi^2=138.38$, df=1, $p<0.001$) (see Figure 6.3). This is a statistically significant result for both groups. Therefore, our first hypothesis — that people treat Hintikka-like sentences as symmetric sentences — was confirmed.

Moreover, additional analysis reveals that with respect to the simple inferences 45 philosophy (50%) and 28 computer science (88%) students answered correctly all questions. Focusing on the proper symmetricity tasks, 71 subjects among the philosophers (79%, $\chi^2 = 30.04$, $p<0.001$, df=1) and 29 computer scientists (91%, $\chi^2 = 21.13$, $p<0.001$, df=1) recognized correctly all valid and invalid reasoning with a combination of two quantifiers (see Table 6.1 for summary of the results).

Discussion  Let us shortly justify our statistical analysis of the results. We were only interested in the frequency of correct answers among all answers to the tasks based on the valid symmetric inference pattern (simple inferences and inferences based on the logically invalid schema were treated as fillers) and that is why we

![Figure 6.3: Percentage of correct answers in the first test.](image-url)
used $\chi^2$ to analyze our data and not a statistical model, like MANOVA, in which the observed variance is partitioned into components due to different independent (explanatory) variables (e.g. 2 groups of subjects, 4 types of tasks). We did not analyze the data with MANOVA because the assumptions were violated (see e.g. Ferguson and Takane, 1990). According to our hypothesis we had expected that the number of answers “valid” will dominate. In other words, the normality assumption of MANOVA was not satisfied, i.e., the distribution of the answers is not normal but skewed (-4.728) towards validity. That is another reason to use non-parametric test like $\chi^2$. Additionally, the conditions (within-subject) for each kind of tasks were different (the number of problems varied between 10, 4, 3, 3) and the groups were not equal (90 philosophers, 32 computer scientists) what also indicates the use of non-parametric statistical model.

However, we did compare between-subjects the performance of two unequal groups (philosophers vs computer scientists) with respect to the three tests and found no statistically significant differences. To be more precise, there was no difference neither in the task based on valid symmetric inference schema ($\chi^2=6.583$, df=6, p=0.361), in the simple inferences ($\chi^2=8.214$, df=4, p=0.084), nor in invalid inference patterns ($\chi^2=3.888$, df=4, p=0.421).

**Experiment II: Branching vs. Conjunctural Interpretation**

The second questionnaire was the main part of the experiment, designed to discover whether people agree with the conjunctural reading of Hintikka-like sentences. Subjects were presented with nine non-equivalent Hintikka-like sentences. Every sentence was paired with a model. All but two sentences were accompanied by a picture satisfying the conjunctural reading but not the branching reading. The remaining two control tasks consisted of pictures in which the associated sentences were false, regardless of which of the possible interpretations was chosen.

Every illustration was black and white and showed irregularly distributed squares and circles. Some objects of different shapes were connected with each
other by lines. The number of objects in the pictures varied between 9 and 13 and the number of lines was between 3 and 15. Moreover, the number of objects in pictures where the sentences were false was similar to the number of objects in the rest of the test items. Almost all subjects solved these tasks according to our predictions (90% correct answers).

The sentences were of the following form, where $1 \leq n, m \leq 3$:

(31) More than $n$ squares and more than $m$ circles are connected by lines.

(32) Więcej niż $n$ kwadraty i więcej niż $m$ koła są połączone liniami.

Notice that the Hintikka-like sentences discussed in Chapter 6.1 as well as the items in the symmetricity test contain the phrase “each other”. However, we decided not to use this phrase in the sentences tested in the main part of the experiments. This was because our previous experiments (Gierasimczuk and Szymanik, 2006, 2007) indicated that the occurrence of reciprocal expressions in these sentences made people interpret them as statements about the existence of lines between figures of the same geometrical shape. This surely is not the interpretation we wanted to test.

In the first the usage of the phrase “each other” was meant for making “hating” relation symmetric. In the case of this experiment the relation “being connected by a line” is already symmetric in itself. Moreover, interviews with native speakers suggest that in the context of the relation “being connected by lines” omitting “each other” leads to more natural sentences. Additionally, in the Polish version of the sentences there is no possible phrase corresponding to “each other”. This is a grammatical difference between Polish and English Hintikka-like sentences. Even though we assert the possibility of the influence that reciprocals can have on the interpretation of the Hintikka-like sentences (see e.g. Dalrymple et al., 1998) this discussion falls outside the scope of the chapter.

Figures 6.4 and 6.5 show two examples of our tasks. In the first picture the conjunctional reading is true and the branching reading is false. In the second picture the associated sentence is false, regardless of interpretation. The subjects were asked to decide if the sentence is a true description of the picture.

6.4.1. Remark. Let us give here a short explanation why we did not show pictures with a branching interpretation. The theoretical arguments given in Section 6.1 justify the following opposition: either Hintikka-like sentences are interpreted in the conjunctional or in the branching way. We want empirical evidence for acceptability of the conjunctional reading. In principle we have to compare this with the branching meaning. Notice however, that the branching reading implies the conjunctional reading so it is impossible to achieve consistent results rejecting branching readings and confirming conjunctional reading — at least as long as subjects recognize the inference relations between branching and conjunctional readings, and in our experience most of them do (see Gierasimczuk and Szymanik,
More than 1 square and more than 2 circles are connected by lines.

Figure 6.4: Conjunctinal task from the second part of the experiment.

2006, 2007). Therefore, we want to prove that people accept the conjunctional reading and not that they reject the branching one. In other words, we are looking for the weakest (theoretically justified) meaning people are ready to accept. To do this it is sufficient to have tasks with pictures for which the conjunctional reading is true, but the branching reading is false. As long as subjects accept them we know that they agree with the conjunctional reading and there is no need to confront them with the branching pictures. Of course this does not mean that people in principle reject the branching reading (but see the computational complexity argument from Section 6.3.5).

Results We got the following results.\(^4\) 94% ($\chi^2=444.19$, df=1, $p < 0.001$) of the answers of the philosophy students and 96% ($\chi^2=187.61$, df=1, $p < 0.001$) of the answers of the computer science students were conjunctional, i.e., "true" when the picture represented a model for a conjunctional reading of the sentence. When it comes to 2 sentences that were false in the pictures no matter how subjects interpreted them we got the following results 92% ($\chi^2=136.94$, df=1, $p < 0.001$) and 96% ($\chi^2=50.77$, df=1, $p < 0.001$) (see Figure 6.6). All these differences are statistically significant. Therefore, our second hypothesis — that in an empirical

\(^4\)We used non-parametric statistical test $\chi^2$ because of the analogous reasons like those explained when discussing the first experiment.
More than 3 circles and more than 2 squares are connected by lines.

Figure 6.5: An example of a false task from the second part of the experiment.

case people can assign to Hintikka-like sentences meanings which are best represented by the conjunctional formulae — was confirmed.

Additionally, analysis of the individual subjects’ preferences revealed what follows. 85 (94%, $\chi^2 = 71.11$, $p<0.001$, df=1) philosophers and 31 (97%, $\chi^2 = 28.12$, $p<0.001$, df=1) computer scientists agreed on the conjunctional reading in more than half of the cases. Moreover, 67 (74%, $\chi^2 = 21.51$, $p<0.001$, df=1) philosophers and 28 (88%, $\chi^2 = 18$, $p<0.001$, df=1) computer scientists chose conjunctional readings in all tasks (see Table 6.2 for presentation of all data).

<table>
<thead>
<tr>
<th>Groups</th>
<th>Polish philosophers</th>
<th>American computer scientists</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of subjects</td>
<td>90</td>
<td>32</td>
</tr>
<tr>
<td><strong>conjunctional answers</strong></td>
<td>94%</td>
<td>95%</td>
</tr>
<tr>
<td>recognized falsity</td>
<td>92%</td>
<td>96%</td>
</tr>
<tr>
<td>most conjunctional answers</td>
<td>85 (94%)</td>
<td>31 (97%)</td>
</tr>
<tr>
<td>only conjunctional answers</td>
<td>67 (74%)</td>
<td>28 (88%)</td>
</tr>
</tbody>
</table>

Table 6.2: Results of the second test.
Moreover, we did not observe any differences between our two subject groups neither in judging obviously false situations ($\chi^2 = 0.188$, df = 1, $p = 0.664$) nor in the conjunctional preferences ($\chi^2 = 3.900$, df = 7, $p = 0.791$). Therefore, we conclude that with respect to interpretation of quantifier combinations in Hintikka-like sentences there is no difference between English and Polish.

### 6.5 Summary

**Conclusions**

We argue that Hintikka-like sentences have readings expressible by linear formulae that satisfy all conditions which caused the introduction of branching interpretation, despite what Hintikka (1973) and many of his followers have claimed. The reasons for treating such natural language sentences as having Fregean (linear) readings are twofold.

In Section 6.1 we discussed theoretical arguments. We can sum up them as follows.

- For Hintikka’s sentence we should focus on four possibilities: a branching reading (4), and three weak readings: (5), (6), (7).

- Hintikka’s argument from symmetricity given in Section 6.3.2, together with the results of our first experiment, allows us to reject asymmetric formulae. A similar argument leads to rejecting the linear readings of other Hintikka-like sentences.

- What about the weakest reading? It does not exist for some Hintikka-like sentences so it cannot be viewed as a universal reading for all of them.
Moreover, the inferential argument from Section 6.3.3 suggests that the weakest meaning is also not an appropriate reading of Hintikka’s sentence.

- Therefore, there are only two alternatives — we have to choose between the conjunctional (5) and the branching readings (4).

In section 6.4 we discussed our empirical results. They indicate that people interpret Hintikka-like sentences in accordance with the conjunctional reading, at least in an experimental context. Moreover, we observed no statistically significant differences in preferences of native English and native Polish subjects.

Additionally, our experimental arguments can be supported by the following observations.

- The argument by Barwise from negation normality, discussed in Section 6.3.4, agrees with our empirical results.

- Branching readings — being NP-complete — can be too difficult for language users. Conjunctional readings being PTIME computable are much easier in this sense.

Hence, even though we in principle agree that Hintikka-like sentences are ambiguous between all proposed readings, our experiments and theoretical considerations convince us that in some situations the proper reading of Hintikka-like sentences can be given by conjunctional formulae. This clearly contradicts Hintikka’s thesis.

Perspectives

We have tested one of the best known among non-Fregean combinations of quantifiers, the so-called Hintikka-like sentences. As a result we came up with arguments that those sentences can be interpreted in natural language by Fregean combinations of quantifiers. However, there is still some research to be done here.

6.5.1. QUESTION. One can find and describe linguistic situations in which Hintikka-like sentences demand a branching analysis. For example, the work of Schlenker (2006) goes in this direction (recall example 12).

6.5.2. QUESTION. Moreover, it is interesting to ask which determiners allow a branching interpretation at all (see e.g. Beghelli et al., 1997).

6.5.3. QUESTION. Finally, we did not discuss the interplay of our proposition with a collective reading of noun phrases (see e.g. Lønning, 1997, and Chapter 5) and different interpretations of reciprocal expressions (see Dalrymple et al., 1998, and Chapter 4).
As to the empirical work, we find a continuation toward covering other quantifier combinations exciting and challenging. Some ideas we discussed in the context of Hintikka-like sentences, such as inferential meaning, negation normality, and computational complexity perspective, seem universal and potentially useful for studying other quantifier combinations. For instance, they can be used to investigate the empirical reality of the influence of computational complexity on the Strong Meaning Hypothesis in the domain of reciprocal expressions discussed in Chapter 4.