1 *Logic and human languages*

1.1 Introduction

For thousands of years, everyone believed that the human mind was logical. This was so obvious that no one thought to question it, or to verify it. Forty years ago some scientists decided to investigate, and they found that people failed some simple tests of logical reasoning. Scientists concluded that our minds use rules that conflict with logic. New models of human reasoning have flourished, with a common theme – that human languages and classical logic involve different basic concepts, and different principles of interpretation. We think these conclusions were unwarranted, and that the new models of logical reasoning are misguided. In our view, the erroneous conclusions were reached by using experimental techniques that people find difficult, such as making abstract judgments about the entailments of sentences. We have chosen to take a different approach. We have advocated the research strategy of finding out how logical people are simply by looking at the languages they speak.

In particular, we have chosen to investigate whether certain expressions and structures of typologically different languages, English and Mandarin Chinese, mean the same as the corresponding expressions and structures in classical logic. When we reason in English, there are particular words that we regularly use, words like *if*, *and*, *not*, *or*, *all*, and *some*. Mandarin Chinese has words that correspond to these English words, and it turns out that the words in Mandarin have pretty much the same meaning as the English words. In Mandarin Chinese, the word *ruguo* corresponds to English *if*, the word *he* corresponds to English *and*, *mei* means *not*, *huozhe* means *or*, and so on.

In this book we will investigate the relationship between the logical words of English and Mandarin and the corresponding logical expressions of classical logic. In logic the symbols $\rightarrow$, $\wedge$, $\neg$, $\vee$, $\forall$, and $\exists$ are represented by the English words *if*, *and*, *not*, *or*, *all*, and *some*. Logic has strict rules for combining these and other symbols to form sentences, and it also has strict rules that transform older sentences that have been formed using these symbols into new sentences that contain them. Logic also has rules that state logical equivalences between sentences that have been formed and transformed using the basic rules for combining symbols. These rules are known as de Morgan's laws of propositional logic. One of the logical equivalences in de Morgan's laws states that a negated disjunction, $\neg(A \vee B)$, is equivalent to a conjunction of the negated disjuncts, $\neg A \wedge \neg B$. Let the letter A represent the English sentence *Amy ate sushi* and let B represent the sentence *Bob ate sushi*. Then $\neg(A \vee B)$ can be paraphrased as *Neither Amy nor Bob ate sushi*. The translation of $\neg A \wedge \neg B$ can be paraphrased as *Amy didn't eat sushi and Bob didn't eat sushi*. So, if English adheres to this one of de Morgan's laws, then *Neither Amy nor Bob ate sushi* should be logically equivalent to *Amy didn't eat sushi and Bob didn't eat sushi*. Intuitively, these two sentences are true in exactly the same circumstances, which means that they are logically equivalent. So English appears to conform to this one of de Morgan's laws.

One question we will attempt to answer is whether all of the world's languages obey the basic laws of logic, such as this one of de Morgan's laws. We obviously cannot study all the world's languages, so our approach will be to examine two historically unrelated languages in detail, English and Mandarin Chinese. If these typologically distinct languages share a common logical foundation, then this is at least circumstantial evidence that all human languages adhere to these basic laws of logic. And if this is so, a second question arises. How do children acquiring different human languages discover that these laws hold? This leads to other questions. If children do manage to discover these laws, at what age do they know them, and do they make any mistakes in the course of acquiring them? Again, we will only investigate English-speaking children and Mandarin-speaking children, but the methodological techniques we apply, and the conclusions we reach about these children, can be applied to studies of children acquiring any other human language.

1.2 The disconnect between logic and language

The current literature on language and logic is rife with controversy about the nature of human reasoning. Some researchers maintain that humans reason using a *mental logic* (e.g., Rips 1994). The majority of researchers have reached the opposite conclusion, however, and have abandoned a logic-based approach to human reasoning in favor of non-logical approaches (e.g., Cosmides and Tooby 1992; Johnson-Laird and Byrne 2002). These researchers have offered several reasons for supposing that there is a disconnect between logic and human reasoning. First, it is well documented that humans perform poorly in applying inference rules in formal proofs. For example, the inference rule for introducing the disjunction operator $\vee$ (corresponding to English *or*) holds that if a statement of the form P is true, then a statement of the form $P \vee Q$ is also true. But few English speakers judge it to be valid to infer from the truth of the English statement *snow is white* to the truth of the statement that *snow is white or Julia Gillard is the current Prime Minister of Australia.*

$$\frac{\mathrm{P}}{\mathrm{P} \vee \mathrm{Q}} \qquad \frac{\text{snow is white}}{\text{snow is white or Julia Gillard is PM}}$$

The force of this argument against there being a *mental logic* is not compelling, in our view. Inference rules are governed not only by logic, but also by conversational principles, such as Be Cooperative, i.e., make the strongest statement that you are in a position to make. This conversational principle is relevant even in making basic inferences. To take the example under consideration, $P \vee Q$ makes a weaker statement than P alone, in information-theoretic terms, so people are fully justified in rejecting the direct inference from P to $P \vee Q$, though there are also circumstances in which this inference is judged to be valid (cf. Crain 2008; McCawley 1981).

A second reason has been given for supposing that there is a disconnect between formal logic and logical reasoning in human languages. This is the observation that valid arguments hold in virtue of their form, not their content; yet children younger than 12 (Inhelder and Piaget 1964) and even adults are often swayed by the content of an argument in making judgments of validity, leading to errors in many cases. Unless the premises of an argument are about familiar topics, most people experience difficulty, for example, in interpreting conditional statements (English *If … then …)*, as attested by their poor performance on the Wason card selection task (e.g., Wason 1968). The force of this argument is blunted by noting that, as soon as people are told the solution to the Wason card selection task, they immediately accept the solution as correct. This is a clear demonstration of logic guiding human reasoning (van Benthem 2008).

A third reason for supposing that there is a disconnect between logic and human reasoning is the observation that both children and adults perform poorly in judging entailment relations between sentences. For example, people often stumble over the following questions:

Question 1: If *every cow is a brown cow,* then *is every cow brown?* (Answer: yes)

Question 2: If *only cows are brown cows* then *are only cows brown?* (Answer: no)

Question 3: If *someone brought wine and beer*, then *did someone bring wine and did someone bring beer?* (Answer: yes)

Although we acknowledge that such judgments cannot be made easily, there is more to logic than proof theory, and judgments of validity and entailment relations do not exhaust the logical apparatus that people utilize in reasoning. There is an entire area of research that has remained virtually unexplored in evaluating the role of logic in human reasoning, namely the interpretation of logical expressions in human languages, including child language.

For the past decade, several colleagues and I have been investigating how logical expressions are interpreted across typologically different human languages, and how the meanings of these expressions are acquired by children. For the most part, we have focused on the meanings that language users assign to individual sentences, rather than asking them to judge entailment relations between sentences. Based on cross-linguistic research, and based on the findings of experimental research with children, we have come up with a series of principles of interpretation in human language that adhere to the precepts of classical logic (Crain 2008; Crain, Thornton, and Khlentzos 2008). The fact that different languages adhere to the same logical principles is circumstantial evidence that human languages draw upon an innate set of logical primitives that are used in speaking and in reasoning. As Aboriginal poet Neidjie (2002) put it, *Language is different, like skin. Skin can be different, but blood same*.

This book presents a series of arguments for logical nativism, which is the proposal that humans are biologically equipped with the tools for logical reasoning. The purpose of the book is to show that human languages are logical, in the sense that the meanings of linguistic expressions corresponding to quantificational expressions (e.g., the universal quantifier) and logical connectives (e.g., negation, the disjunction and conjunction operators) are the same as the meanings of the corresponding expressions in classical logic. A number of putatively universal linguistic principles will be advanced, all conforming to first-order logic.[1] Although these principles are manifested in typologically different languages, of which English and Mandarin represent our sample, they are only operative in complex structures in which disjunction combines with negation, or with the universal quantifier. It is highly unlikely that children have sufficient evidence from experience to learn the meanings of the logical expressions of human languages by observing how adult speakers use these expressions, contrary to usage-based accounts of learning. This invites us to consider the alternative possibility, that the meanings of logical expressions are innately specified, as part of the language faculty.

We should make it clear what we take to be the relationship between logic and language. Our topic is the sentence-level language abilities of children and adults, as they comprehend sentences that contain expressions that correspond to those of first-order logic.[2] We will not be discussing the broader range of abilities that are required to calculate entailment relations between sentences. Accessing the meanings of logical expressions is a necessary prerequisite for making such inferences, but it is not the whole story. In addition to knowledge of the meanings of the logical expressions, making inferences requires the resources provided by the recursive computational system (syntax).

In Chomsky (2000a, p. 94), the following "evolutionary fable" was formulated to clarify the need for a computational system that interfaces with the cognitive and perceptual systems in order to use linguistic expressions to convey thoughts:

> Imagine some primate with the human mental architecture and sensorimotor apparatus in place, but no language organ. It has our modes of perceptual organization, our propositional attitudes (beliefs, desires, hopes, fears, etc.) insofar as these are not mentioned by language, perhaps a "language of thought" in Jerry Fodor's sense, but no way to express its thoughts by means of linguistic expressions, so that they remain largely inaccessible to it, and to others.

The thought experiment by Chomsky (2000a) was further fleshed out by Reinhart (2006, p. 2), who endows the primate with an innate system of logical concepts that is the equivalent to that of humans. The primate would still lack the means to express its thoughts, including logical inferences, to other primates:

> Let us assume, further, that he has an innate system of logic, an abstract formal system, which contains an inventory of abstract symbols, connective functions, and definitions necessary for inference. What would he be able to do with these systems? Not much. Based on the rich concept system of humans, his inference system should in principle allow him to construct sophisticated theories and communicate them to his fellow primates. However, the inference system operates on propositions, not on concepts, so it is unusable for the primate in our thought experiment. Possibly he could code concepts in sounds, but not the propositions needed for inference.
>
> Pursuing this thought experiment, the goal of linguistic theory can be described as reconstructing the system the primate lacks, which consists of

> whatever is needed to facilitate the interface of his various cognitive systems. In other words, the goal is to construct the computational system (CS) (syntax in a broad sense) that defines language…

As Reinhart argues, a primate armed only with knowledge of the concepts of first-order logic would be unable to apply this knowledge in the production and comprehension of propositions, or in discerning the entailment relations between sentences that express these propositions. To put the logical vocabulary to use, a computational system is required.

Following Chomsky and Reinhart, we will assume that a single computational system underpins all of the ways in which language is put to use by humans, including assigning truth conditions to sentences with logical expressions, and making logical inferences based on these assignments. In Chapter 3, we substantiate the claim that the same structural principles used to determine the reference of pronouns in human languages are also used (a) for the interpretation of logical expressions in sentences involving quantifier–variable binding, (b) to decide on the interpretations of sentences with disjunction, conjunction, and focus expressions, and (c) to compute the meanings of sentences with multiple logical expressions. As a final note, recent brain imaging research indicates that, in processing sentences with logical expressions, 4-year-old English-speaking and Mandarin-speaking children have been found to recruit the same brain regions that support language in adults (Heschl's Gyrus, BA40 and BA44). Interestingly, the brain regions active in children are only a subset of those regions that have been found to distinguish adult Mandarin speakers and adult English speakers. The brain asymmetries that have been uncovered for adults, across languages, appear to develop only gradually in children. In the early stages of language acquisition, children acquiring different languages appear to engage a specialized language processing system (i.e., Language Acquisition Device) for processing sentences, both with and without logical expressions.

1.3 Principles of logic and language

We begin with an example of a statement that can be expressed in many human languages. This kind of statement invokes two of the logical operators of first-order logic. The statement we have in mind contains disjunction in the scope of negation. Consider, as an example, the English sentence *John didn't see Ted order pasta or sushi*. For English speakers, this sentence is true if (a) John didn't see Ted order pasta *and* (b) John didn't see Ted order sushi. We call this the *conjunctive interpretation* of disjunction under negation. In first-order logic, the conjunctive entailment of disjunction in the scope of negation is stated as one of de Morgan's laws:

$$\neg(A \vee B) \Rightarrow \neg A \wedge \neg B$$

where $\vee$ is inclusive-*or*, $\wedge$ is conjunction, $\neg$ is negation, and $\Rightarrow$ indicates entailment.

Examples (2–5) are translations of (1) *John didn't see Ted order pasta or sushi* into Mandarin in (2), Japanese in (3), Dutch in (4), Russian in (5), Norwegian in (6), and Hungarian in (7). The interpretation of the corresponding sentence in each language is the same. Negation and disjunction are boldfaced in the examples.

(1) John did*n't* see Ted order pasta *or* sushi.
(2) *Yuehan mei kanjian Ted dian yidalimianshi* ***huozhe*** *shousi.*
(3) *John-wa Ted-ga sushi* ***ka*** *pasuta-o tanomu-no-o mi-**nakat**-ta.*
(4) *John zag Ted* ***niet*** *pasta* ***of*** *sushi bestellen.*
(5) *Dzhon* ***ne*** *videl/uvidel chto/kak Borja zakazal/zakazyval pastu* ***ili*** *sushi.*
(6) *Jon så* ***ikke*** *Ted bestille pasta* ***eller*** *sushi.*
(7) *János* ***nem*** *látta Edvardot tésztát* ***vagy*** *szusit rendelni.*

The conjunctive entailment of disjunction in the scope of negation holds only if disjunction is interpreted as inclusive disjunction (inclusive-*or*), as in first-order logic. To see this, consider the truth conditions of inclusive-*or*. A statement of the form $(A \vee B)$ is true in three cases: (i) if A is true but not B, (ii) if B is true but not A, and (iii) if both A and B are true. A statement of the form $(A \vee B)$ is false, therefore, only if both A and B are false. This means that the negated disjunction $\neg(A \vee B)$ is true in cases where both A and B are false. It follows that $\neg(A \vee B)$ logically entails $(\neg A \wedge \neg B)$.[3]

Despite examples like (1–7), many linguists and philosophers have concluded that the expressions corresponding to disjunction in human languages do not have the meaning of disjunction in first-order logic. Many researchers have reached the conclusion that the expressions for disjunction in human languages have the meaning associated with exclusive-*or*. Others are convinced that *or* is ambiguous between inclusive disjunction and exclusive disjunction. For example, Braine and Rumain (1983, p. 291) acknowledge the view that "equates *or* with standard logic," yet reject this view on the grounds that "coherent judgments of the truth of *or*-statements emerge relatively late and are not universal in adults." They conclude that disjunction is exclusive-*or* more often than not, even for adults (e.g., Kegley and Kegley 1978; Richards 1978).

This conclusion is unwarranted, as Grice and others have argued. Consider the following situation – John watches Ted order sushi and then go back to order pasta. Now consider the following alternative descriptions of the situation.

(8) John saw Ted order sushi or pasta.
(9) John saw Ted order sushi and pasta.

From a logical point of view, both statements are true, at least if the English word for disjunction, *or*, is assigned an inclusive-*or* interpretation in (8). But the statement with *or* in (8) is pragmatically odd as a description of the situation. Here is why. In judging the truth or falsity of sentences, people are influenced not only by the semantic properties of these expressions but also by pragmatic principles. One of the most basic principles of pragmatics is Grice's *Principle of Cooperation*. This principle of conversation entreats speakers to be as informative as possible, and it can be invoked to explain why the use of *or* typically implies exclusivity (*not both*), at least for adults. Returning to (8) *John saw Ted order sushi or pasta,* this statement is judged by most adult English speakers to imply that John saw Ted order sushi, or John saw Ted order pasta, but not both. This implicature of exclusivity (*not both*) derives from the availability of another statement that is more informative, namely (9) *John saw Ted order sushi and pasta*. If John had seen Ted order both sushi and pasta, then (9) directly conveys what John saw. Example (8) also conveys this, but less directly – in the sense that (8) would also be true if John saw Ted order sushi but not pasta, or if John saw Ted order pasta but not sushi. Because (9) is true in only the one circumstance (in which John saw Ted order both), it is said to be more *informative*.

More generally, a statement of the form *A and B* is more informative than one of the form *A or B* (where both statements are true descriptions of the events that took place). The statement *A and B* is more informative because a statement of this form is true only if both A and B are true, whereas a statement of the form *A or B* is true in the same circumstances, but it is also true in other circumstances as well, where only A is true, or where only B is true. So, the truth conditions that verify *A and B* constitute a subset of the circumstances that verify *A or B*. Consequently, the expressions *or* and *and* form a scale, based on information strength, with *and* being more informative than *or*. This is where the Principle of Cooperation comes in. Upon hearing someone use the less informative term *or*, listeners assume that the speaker was being cooperative, so they infer that the speaker was not in a position to use the more informative term *and*. Therefore, the speaker's use of *or* is taken to imply the negation of the stronger statement, so the use of *or* is taken to imply *not both A and B*.

The basic meaning of disjunction is inclusive-*or*. In positive statements, however, there is an implicature of *exclusivity*. Disjunction is still inclusive-*or* in statements of the form *A or B*, but hearers are invited to make a pragmatic inference that excludes the possibility of both *A* and *B* being true. So statements of the form *A or B* are interpreted as generating a "derived" meaning in human languages: *A or B, but not both A and B*. This derived meaning of disjunction is not generated in negated statements, so negated disjunctions retain the basic inclusive-*or* interpretation. This was illustrated in (1), repeated here as (10): *John didn't see Ted order pasta or sushi.* This example entails that John didn't see Ted order pasta *and* it entails that John didn't see Ted order sushi. This conjunctive entailment of negated disjunctions follows only if disjunction is assigned the inclusive-*or* interpretation, as in first-order logic.

Why is there no implicature of exclusivity in interpreting sentences of this form? To answer this, compare (10) and (11), where *or* has been replaced by *and*.

(10) John didn't see Ted order pasta or sushi.
(11) John didn't see Ted order pasta and sushi.
  a. He just saw Ted order sushi.
  b. He just saw Ted order pasta.
  c. In fact, he didn't see Ted order either.

Adults judge (11) to be true in various different circumstances, as indicated by the possible follow-up comments in (11a–c). By contrast, (10) is judged to be true in only one set of circumstances, where Ted ordered neither sushi nor pasta. This means that (10) is true in a subset of the circumstances that verify (11). The upshot is that negation reverses the subset/superset relations of truth conditions that hold for positive statements with *or* and *and*. In the scope of negation, the use of *or* makes a stronger statement than the corresponding statement with *and*. Under negation, therefore, the implicature of exclusivity for *or* is not generated.[4]

The moral is that the implicature of exclusivity (*not both*) must be removed to unveil the meaning that human languages assign to statements with disjunction. One way to remove the implicature is to reverse the scale of information strength, by introducing specific linguistic expressions, such as negation. By reversing the entailments that give rise to such implicatures, we clear the decks for a clearer look at the meaning of disjunction in human languages. Once this is done, we immediately see that the expressions for disjunction in many human languages conform to first-order logic. In previous work, we have adopted this research strategy to investigate whether or not typologically different languages assign the conjunctive entailment to negated disjunctions, as in one of de Morgan's laws. Some of the findings are illustrated in the languages surveyed using examples (1–7). They reveal that human languages typically assign the conjunctive entailment to disjunction in negative sentences.

We will use the same strategy, beginning in the next section, to demonstrate that the parallels between logic and language run much deeper than just adherence to one of de Morgan's laws. In order to accomplish our goal, we will extend the range of linguistic expressions that reverse entailment relations. In fact, negation is just the tip of a very large iceberg. Other linguistic expressions that reverse entailments include (the subject phrase of) the universal quantifier and (the antecedent of) conditionals, among many others. The class of linguistic expressions that reverse entailments is called *downward-entailing* expressions. The definition and some of the diagnostic properties of downward entailment is the topic of the next section. Using examples from English and Mandarin, we will invoke downward-entailing linguistic expressions to argue that human languages draw extensively on the resources of first-order logic. Although we will limit the examples to just these two languages for the most part, we want it understood that the same case could have been made using many other languages. In addition to English and Mandarin Chinese, we have verified the logical principles that we describe in the present book across a broad spectrum of human languages, including Russian, Japanese, European Portuguese, Italian, and Dutch.

1.4 Downward entailment

To continue the discussion of the overlap between logic and language, this section will describe several principles that involve the semantic property of downward entailment. At the broadest cut, a linguistic expression is downward entailing if and only if it licenses inferences from general to specific reference. More precisely, a linguistic expression is downward entailing if it licenses inferences from expressions referring to sets (e.g., *car*) to expressions referring to their subsets (e.g., *Prius*). Formally, downward-entailing linguistic contexts reverse entailment relations, so a linguistic context C is downward entailing if it follows from the fact that A entails B that C(B) entails C(A) (e.g., Ladusaw 1980; Landman 1991). A context with the English negative determiner *nobody* (Mandarin *meiyouren*) is downward entailing, according to this definition, since it is logically valid to infer from the statement *Nobody at this table ate* ***fruit*** (*Zhe zhuo meiyouren chi* ***shuiguo***) to the statement *Nobody at this table ate* ***apple*****s** (*Zhe zhuo meiyouren chi* ***pingguo***). Where ⇒ indicates logical entailment, the following entailments hold in both English and Mandarin.