Between Logic and Common Sense: The Formal Semantics of Words

Five year research program, supported by an NWO *Vici* grant

February 2010

Principal investigator: Yoad Winter

1. Program goal: incorporating content words into a model of entailment

One of the most important properties of human languages is their ability to convey intricate meanings. The vastness and effectiveness of those meanings for everyday communication and reasoning transcends all known non-human languages, including other animal languages and artificial languages. For studying communication and reasoning in language, an indispensable empirical concept is *entailment:* the relation between premises and valid conclusions expressed as natural language sentences.

Entailment relations may appear between sentences due to the presence of words and expressions like *other*, *either…or*, *not* or *exactly five*, whose meanings have been studied in logical frameworks since antiquity. However, entailments may also appear due to semantic properties of "non-logical" words like *parrot, hug*, *far* or *knowledge*. Such *content words* constitute the bulk of the lexicon in all natural languages. Without considering them, it is simply impossible to understand entailment phenomena and human reasoning in general. However, while content words have played an important role in cognitive psychology and artificial intelligence, their meanings have turned out to be richer and fuzzier than what current logical techniques in semantics can handle. As a result, the study of entailment has lagged behind important advances of other areas in the study of language.

This program will develop a comprehensive model of entailment in language. A new framework of *Content Sensitive Formal Semantics* will treat the interactions between common sense meanings of content words (e.g. *parrot*) and logical operators in language (e.g. *not*). The resulting model of entailment is expected to constitute a breakthrough in our understanding of linguistic reasoning and its practical applications.

2. Main Problem: how does language divide labor between logic and common sense meaning?

What are the interactions between "logical" and "non-logical" aspects of word meaning? How do they affect entailment relations between sentences? Formal accounts of meaning have impressively succeeded in modeling many "logical" aspects of language. Rooted in classical logic and philosophy of language (Geach and Black 1970), formal semantics (Montague 1973, Gamut 1991) has characterized the major processes that allow syntactic structures to be interpreted. Thereby, formal analysis accounts for many classical logic inferences stated as natural language entailments. For instance:

Each feminist is vegetarian and each vegetarian is smart. (premise)

$\subset$ Each feminist is smart. (conclusion)

Treating this entailment in formal semantics relies on elaborating the logical analysis of the words *each* and *and*. However, together with such logical aspects of entailment, also "non-logical" meanings of content words comprehensively affect entailment validity.

Consider the following sentence pairs in (1)-(3). In each pair, the common sense meanings of the underlined content words lead to different sentential interpretations.

(1) a. My family has a red <u>car</u>.
b. My family has red <u>hair</u>.

Here the required color of objects denoted by the modifier *red* differs because of the different semantics of the modified content words *car* and *hair*. The color that is referred to in (1b) as *red* in connection with hair, would normally be called *orange* when associated with cars (Smith et al. 1988, Osherson and Smith 1997).

(2) a. The three girls <u>know</u> each other.

b. The three girls are <u>pinching</u> each other.

The numbers of acquaintances/pinching relations required by the reciprocal expression *each other* differ. In (2a) every girl needs to know every other girl, whereas in the (2b) every girl only needs to pinch one other girl. This difference appears because of common sense meanings of the *relational* verbs *know* and *pinch*: it is common for people to be involved simultaneously in multiple `knowing' relations, but not in multiple `pinching' relations (Winter 2001b).

(3) a. We are <u>far</u> from the lakes.

b. We are <u>near</u> the lakes.

Whether distance is measured from all the lakes or from one particular lake depends on the concepts denoted by the spatial words *far* and *near* (Zwarts and Winter 2000, Mador-Haim and Winter 2007). Given a set of lakes, (3a) reports on being far from all of them, whereas (3b) only requires being near one of them.

The contrasts between the sentence pairs above are manifested in the validity ('$\subset$') or invalidity ('$\not\subset$') of entailments (1c-3c), uttered in the background of (1-3) respectively.

(1c) John's car is the same color as my family's *car*/*hair*. $\subset$ / $\not\subset$ John's car is red.

(2c) The three girls are Ann, Jane and Mary. $\subset$ / $\not\subset$ Ann *knows*/*is pinching* Mary.

(3c) The lakes are lakes Superior, Michigan, Huron, Erie, and Ontario.
$\subset$ / $\not\subset$ We are *far from*/*near* lake Michigan.

Based on the information in (1a/b), whether entailment (1c) is valid or not depends on the content words *car*/*hair*. Given the information in (1a) about my family's red *car*, the conclusion in (1c) that John's car is red necessarily follows from the premise. However, if John's car is the same color as my family's *hair*, the conclusion that John's car is red is invalid: the car might as well be classified as orange. The lexical variations in (2a,b) and (3a,b) with the words *know*/*pinch* and *far*/*near* similarly affect the (in)validity of the entailments in (2c) and (3c). Understanding the influence of content words on entailments as in (1c)-(3c) is critical for transcending the limitations of current formal semantic theories (Heim and Kratzer 1997, Partee 2004, Barker and Jacobson 2007).

The effects of common sense meaning on entailment are furthermore critical for analyzing how humans perform tasks like question-answering, text summarization and translation, and for mimicking these abilities (Dagan et al. 2006, Manning 2006). Consider for example a simple question-answer interaction. Suppose that the question in (4Q) is introduced to a person (or a computer) and that the information in (4I) is found in some authoritative source. A typical answer would then be the one in (4A).

(4) Q: When was commercial whale hunting banned by the IWC?

I: Japan has long sought to overturn a 1986 ban on all commercial whaling, and its proposal to allow the hunting of minke whales in the Antarctic came at the IWC's annual meeting in Italy.

A: Commercial whale hunting was banned by the IWC in 1986.

The question (4Q) and the text (4I) are stated differently, with different focuses and rhetoric structures. When answering (4Q), people, and computers, must perform a non-trivial semantic task and recognize the *entailment* between the given information (4I) and the expected answer (4A) (Groenendijk and Stokhof 1984). This process requires recognizing common sense meaning relations as between *whaling* and *whale hunting,* together with logical effects like those of the word *all* in (4I). Interactions between content words and logic in entailments are therefore an integral part of question-answering, and similarly of other tasks of semantic processing.

By modeling the effects of content words on entailments as in (1c)-(3c) and (4), the program will provide a unique window into the interface between lexical and formal semantics, and more generally – between language meaning, use and computation.

Main Puzzle: How are entailments influenced by “non-logical” aspects of content words and their interactions with “logical” elements of language?

3. Outline of the research program: Content Sensitive Formal Semantics

The program will develop a new semantic framework that is rich enough to model the meanings of content words and their effects on entailment. This new model of *Content Sensitive Formal Semantics* will be informed by experimental evidence and designed to break new grounds in the analysis of entailment in natural language.

A large body of work in philosophy, psychology, linguistics and artificial intelligence has led to the following thesis (Margolis and Laurence 1999):

> *Content word concepts are described using a set of* <u>*semantic features*</u> *that classify instances of the concept.*

Classically, features are used for describing necessary conditions in the meaning of content words. For instance, the “classical” analysis (Katz and Fodor 1963, Jackendoff 1983, Cruse 1986) of the word *bachelor* involves the features *male*, *adult* and *unmarried.* However, content words may also involve features that are *typical* rather than necessary conditions (Wittgenstein 1953, Rosch 1978, Taylor 2003, Boyd-Graber et al. 2006). In order for an entity to be categorized as an instance of *car*, for example, relevant typical features include having four wheels, an engine and a steering wheel. These are not necessary criteria for identifying a car but parameters of family resemblance that affect performance of categorization tasks. The more typical an entity is according to these features, the quicker it is categorized as related to the relevant concept expression (Smith 1990). A new Peugeot is more quickly categorized as an instance of the word *car* than an old Messerschmitt Kabinenroller with three wheels.

Although feature-based models have been very fruitful in the study of content words, the understanding of their influence on entailment models has so far been rather limited (Osherson and Smith 1981,1997; Kamp and Partee 1995). To overcome this critical gap, this program will use the following hypothesis in developing the new framework of Content Sensitive Formal Semantics.

Main Hypothesis: Modeling entailment involves logical semantic manipulation of feature-based representations of content word meanings.

When testing and substantiating this hypothesis, three research areas must be addressed in order to highlight its different aspects: (i) formal semantic meaning *composition,* (ii) linguistic *performance* with word and sentence meanings, and (iii) *acquisition* of entailment using linguistic corpora. These topics will be addressed in the five research projects of the program (section 5). Section 4 elaborates on the core principles of the overall framework.

Innovation: Content Sensitive Formal Semantics will be the first linguistic theory that extends the formal semantic model of entailment into a general framework that is sensitive to the common sense aspects of content words.

4. Principles of Content Sensitive Formal Semantics

The starting point of Content Sensitive Formal Semantics is that of elementary Formal Semantics. Like Montague (1973), this program assumes that real understanding of linguistic faculties involves a *formal* analysis in two different senses of the term "formal". Like all logical studies, the linguistic analysis of entailment must involve rigorous mathematical principles. Furthermore, the study of entailment creates a critical interface between mathematically-informed meanings and linguistic *forms*, hence its centrality for the theory of language. In these days of the information age, it comes without saying that the formal analysis of entailment will have important implications for the computational treatment of electronic information. The program will start to address these more applicative aspects of the new paradigm of Content Sensitive Formal Semantics. However, the rigorously-defined ideas in this program have implications that are more far-reaching than those technological innovations: they aim at the heart of our understanding of one of the major linguistic faculties of the brain.

Starting point: In Content Sensitive Formal Semantics, the mathematical analysis of entailments in ordinary language is a way for bridging logical analysis with performance of actual linguistic tasks.

This section introduces the underpinnings of Content Sensitive Formal Semantics. After giving background in 4.1, sections 4.2–4.3 present the core principles of the program, and section 4.4 introduces its applications for computational linguistics.

4.1 Typicality-based meanings and their limitations

What are typicality effects and how do feature-based meanings model them? Typicality is observed as performance variability in *categorization* assignments. When asked to categorize objects as being *red* or *not red*, subjects' reaction and response time vary considerably. Shades of red are most quickly and reliably categorized as *red* when they are close to a blood-like shade of red, 255-0-0 in the Red-Green-Blue (RGB) representation of colors. Other shades of red (crimson, pink) or of related colors (purple, orange) are categorized less reliably or less quickly, whereas colors like green are never categorized as *red*. To capture such typicality effects, feature-based accounts (Rosch 1978, Boyd-Graber et al. 2006) use a natural generalization of *feature structures* (Shieber 1988). In (5) the concept *red* has one feature, *color,* whose values are mapped to numeric values on the scale 0-6.

(5) **RED**: *COLOR*: BLOOD-RED → 6 CRIMSON → 4 ORANGE → 2 BLACK → 0

The small-cap notations for color values (e.g. "BLOOD-RED") stand for perceptual variables (better represented by the respective RGB values). The numeric values represent the typicality of those colors as reflected in categorization tests. This representation exemplifies the core semantic content of the word *red* in the linguistic lexicon, while abstracting over further details in the representation of the concept *red* in the brain/mind that are not immediately necessary for linguistic modeling.[1]

[1]Such simplifications for expository purposes include the ultimate choice of features, their exact values and representation in the brain/mind. Specifically, numeric values are not assumed to be represented in the mind/brain, but stand for parameters in the cognitive categorization mechanism that interface formal semantic faculties.

The meanings of the content words *car* and *hair* require multiple semantic features:

(6) **CAR**: *WHEELS*: FOUR → 4 THREE → 2 FIVE+ → 2 TWO → 0 ONE → 0 ZERO → 0

ENGINE: YES → 1 NO → 0 *STEERING- WHEEL*: YES → 1 NO → 0

(7) **HAIR**: *TEXTURE*:  →5  →4  →2  →0.5

COLOR: BLOOD-RED → 0.1 CRIMSON → 0.25 ORANGE → 0.5 BLACK → 1

Like (5), the representations (6,7) assign every object a value on the 0-6 scale, describing its typicality as an instance of the relevant concept. Using (6), objects that have four wheels, an engine and a steering wheel get typicality 4+1+1=6 (typical cars); similar objects that only have three wheels get typicality 2+1+1=4 (atypical cars), and objects that have less than three wheels and no engine or steering wheel (e.g. bikes) get typicality 0 (categorically non-cars). Similarly, (7) approximates prominence of the *texture* and *color* features of the word *hair* in social and contextual circumstances where most people have curly black hair and where red hair is a rarity (e.g. in parts of Africa).[2]

The different interpretations of sentences (1a-b) follow from the features of the content words *car* (6) and *hair* (7). Because the concept *car* does not put any restrictions on possible colors, the objects denoted by the expression *red car* in (1a) "inherit" their shades of colors from the concept representation (5) of *red.* By contrast, typical instances of the concept *hair* in (7) are not of typical shades of red, thus require a "compromise" between conflicting typicalities for *red hair* in (1b). Similar compromise effects have been identified (Van Jaarsveld and Draskovic 2003) with other constructions involving color terms (*blue sea*) or other modifiers (*striped apple*).

Previous accounts (Smith et al. 1988, Kamp and Partee 1995) give various treatments of "compromise" phenomena with modified nominal expressions, but without offering a solution to the problem of entailments with content words as in (1c)-(3c) (Margolis and Laurence 1999:41-42). Some researchers (Osherson and Smith 1982:317, Connoly et al. 2007) even doubt that the truth-conditional semantics of language, as reflected in entailments, should be related to categorization processes with content words. However, the fact remains that many entailments like (1c-3c) cannot be understood without taking typicality effects with content words into account. This serious gap between the formal semantics of entailment and lexical semantic accounts of content words is one of the central challenges for this program. A closely related limitation of feature-based accounts is their focus on *unary* predicates (*car, red*), denoting *sets* of entities. However, other content words like *know*, *pinch*, *far* and *near* (cf. 2-3) have a more complex logical structure, involving binary/spatial *relations* between entities, which are not treated in current typicality-based theories.

The logical structure of relational/spatial expressions and its effects on entailment make their treatment a central component of the new interface between lexical and sentential meaning in Content Sensitive Formal Semantics.

[2]The *wheels/texture* features are more diagnostic than the other features for typicality of car/hair instances, which is reflected in their higher influence on the numeric value.

4.2 Typicality and entailment with relational expressions: the case of reciprocals

Meaning representations of relational content words like *pinch* and *know* must describe semantic properties that from a logical point of view are *defeasible* (=cancelable). Whenever a given sentence states that a person *A* is *pinching* a person *B*, the common sense meaning of the relational verb *pinch* tells us that *A* is not simultaneously pinching anybody else. This inference is defeasible since *A might* be pinching other people at the same time, but such a possibility would be atypical for the verb *pinch*. We describe this using the following entailment, where the symbol '$\prec$' highlights its defeasibility.

(8) Matilda is *pinching* Sue $\prec$ Matilda is pinching nobody else

This defeasible entailment is parallel to entailments with other relational expressions, like the relational expression *give birth to* in the following entailment.

(9) Sue *gave birth* to Matilda $\subset$ Nobody else *gave birth* to Matilda

However, unlike (8), entailment (9) is <u>in</u>defeasible. Formally, the relation GIVE-BIRTH-TO must denote a *function* on its object argument: every person has a unique woman who gave birth to him/her. Similarly, the relation PINCH typically denotes a (partial) function on its *subject* argument: typically, everyone pinches *at most* one other person simultaneously. This parallelism between defeasible and indefeasible aspects of meaning has gone unnoticed in previous work, and it is also reflected with unary predicates:

(10) a. Matilda is a *bird* $\prec$ Matilda has feathers (defeasible)

b. Matilda is a *mother* $\subset$ Matilda has a child (indefeasible)

A new conclusion about lexical meaning follows from these observations:

> ***Lexical predicates and (in)defeasibility***: *Defeasible and indefeasible lexical representations share the same logical structure, and apply similarly to unary and binary predicates.*

This conclusion is exemplified by the representations (11) for the concepts *pinch*, *know* and *give birth to*, referring to numbers of agents (entities having control over the action denoted by the verb) or patients (entities being affected by the action).

(11) Matilda **PINCH**: *NO.OF-PATIENTS*: ONE $\rightarrow$ 6 TWO $\rightarrow$ 4 THREE $\rightarrow$ 2

Matilda **KNOW**: *NO.OF-PATIENTS*: ONE $\rightarrow$ 6 TWO $\rightarrow$ 6 THREE $\rightarrow$ 6

GAVE-BIRTH-TO Matilda: *NO.OF-AGENTS*: ONE $\rightarrow$ 6 TWO $\rightarrow$ 0 THREE $\rightarrow$ 0

The *NO.OF-PATIENTS* feature for *pinch* describes typicality of its denotations with different numbers of simultaneous patients. Maximal typicality is attained when Matilda only pinches one patient, and typicality decreases as the number of patients increases.[3] By contrast, the concept for *know* is indifferent to the number of patients Matilda simultaneously knows. The "classical" concept *give birth to* has a representation similar to *pinch*, but typicality is either maximal (=6, when Matilda has one mother) or zero (having more than one mother). This accounts for the indefeasibility of entailment (9).

The representations (11) allow us to account for sentential meanings (2a,b) and entailments (2c). In (2a,b) the expression *each other* typically requires a maximal number of relations between agents and patients. However, because of the difference in the typical number of patients between the relational verbs *know* and *pinch*, the meanings of the expressions *know each other* and *pinch each other* differ. The former fully inherits the maximality of the reciprocal, and sanctions six acquaintance relations between the three

[3]The possibility of pinching different patients successively in time is abstracted over, and may only be addressed in later stages of this program.

girls in (2a). By contrast, the conflicting typicalities in the interpretation of the latter expression are best satisfied in (2b) if every girl only pinches one other girl, amounting to three pinching relations. This analysis explains the entailment in (2c) as a result of the interaction between the logic of the reciprocal expression (*each other*) and the common sense meaning of the relational expression (*know*/*pinch*). Similar effects appear with other reciprocal sentences:

(12) The delegates *follow one another* into the room.
Protective diodes are utilized for connecting *mutually consecutive* solar cells.
The four karatekas are *kicking each other*.

Previous works (Langendoen 1978, Dalrymple et al. 1998, Winter 2001b) have identified this variability of reciprocal interpretations, but without offering a general theory of how the lexical semantics of relational content words affect reciprocity. Ideas about the relation between feature-based accounts and logical semantics are fairly recent (Kerem, Friedmann and Winter 2009). By developing a unified feature-based account the program will provide a novel account of entailments resulting from sentential meanings of reciprocal sentences.

Modeling typicality with relational expressions, and their interactions with the logic of reciprocity, exposes an important parallelism with modification constructions (*red car*/*hair*) and their entailments. The common sense meaning of *car* has no typical color feature. As a result, the expression *red car* can attain a maximally red color. Similarly, the expression *know each other* in (2) attains maximal reciprocity, resulting in six acquaintance relations between the three women. By contrast, the expressions *red hair* and *pinch each other* involve “compromise” effects. Maximal typicality is still attained, but now at lower levels of redness and reciprocity, due to the conflict with the *COLOR*/*NO.OF-PATIENTS* features of the content words *hair*/*pinch.* This parallelism between the interactions of different logical operators and common sense meanings is one of the key elements in this program, which accounts for entailment (in)validity in (1c,2c).

Maximal Typicality Hypothesis: Entailment is derived by a unified interface between logical operators in formal semantics and typicality-based lexical meanings. The logic of modification and reciprocity similarly imposes *maximal typicality* given possibly conflicting typicality preferences.

4.3 Typicality in binary formal semantic models

Are feature-based representations of content words consistent with logical semantics? How should the two theories be integrated into a unified framework of Content Sensitive Formal Semantics? On first blush, graded typicality of objects with respect to content words (Zadeh 1975) may seem to stand in opposition to the binary (true-false) dichotomy of traditional logical semantics. However, the correspondence between graded typicality and binary models follows from the following principle (Kamp 1975):

***Correspondence principle*:** *Suppose that two objects x and y have typicality TC(x) and TC(y) with respect to a concept C, where TC(x) ≤ TC(y). Then in any binary model: if the object x belongs to the concept C then also the object y does: C(x)=true entails C(y)=true.*

Suppose that an object *x* lacks an engine but is still considered a *car* in a binary model M. According to the correspondence principle and the feature-based representation (6), all objects *y* that do have an engine and share with *x* its other car-features, must also be categorized as cars in the same model M.

The correspondence principle does not completely define the denotation of content words in binary models. For instance: both a *pelican* and a *robin* may be equally considered as *birds* in every model, although their typicality is different (Osherson and Smith 1997). Similarly, in mathematical discourse, both 3 and 11 are equal members of the concept *odd number*, although 3 is more "typical" (Armstrong et al. 1983, Connoly et al. 2007). Thus, the *boundaries* of concepts like *bird* or *odd number* are not defined by typicality and the correspondence principle alone. For the main goal of this program the question is: is a full theory of concept boundaries needed for a theory of entailment? By developing a logical semantic system with content word meanings, project 1 of this program (section 5) will address this foundational question.

The correspondence principle specifies an important part of the interface with truth-conditional semantics for concepts whose representation involves typicality. But which natural language meanings are represented by such concepts? The following qualification makes an important caveat.

> ***Composition qualification***: *Typicality is an integral part of the meaning of content <u>words</u>. By contrast, meaning derivation for complex <u>expressions</u> may or may not involve typicality-based features, depending primarily on linguistic factors.*

The use of typicality with content words is an important interface between the linguistic lexicon and non-linguistic cognitive faculties of *categorization* (Smith 1990). However, composing lexical meanings into meanings of complex expressions is a specialized linguistic process, performed in correspondence with natural language syntax (Barker and Jacobson 2007). This process does not necessarily involve typicality. For instance, unlike modifier and reciprocal constructions, where typicality partakes in meaning composition via the Maximal Typicality Hypothesis, relative clauses as in (13) do not have typical instances (Osherson and Smith 1981, Fodor 1981).

(13) *an apple that is not an apple; cars that uncles drive on Sunday in East Tanzania*

This program will define the linguistic factors that allow typicality-based semantic features to partake in the meaning derivation of complex expressions like modified/reciprocal constructions in (1)/(2), but block their influence with other complex expressions like the relative clauses in (13).

Typicality interacts with *contextual parameters*. A well-known example is the behavior of dimensional adjectives like *tall*, where contextual information outside the noun phrase is required for interpreting the adjective (Kamp 1975):

(14) My 2-years old son built a *tall snowman*.

The snowman is expected to be tall relative to heights of small children. This effect is explained if the adjective *tall* has a parameter for preferred degrees of heights (Winter 2005, Kennedy 2007). In (14) this parameter is set by the context of the expression *tall snowman* and not from within the expression itself. By contrast, consider (15):

(15) My 2-years old son climbed a *tall giraffe*.

In (15), unlike (14), typical heights of the head noun *giraffe* are prominent when interpreting the noun phrase *a tall giraffe*, and the child's height has little influence on the understood height of the giraffe. Thus, in (15) the contextual parameter of *tall* is "plugged in" to the height feature of the word *giraffe*, which has of course typical heights. The concept *snowman*, by contrast, does not have a prominent height feature. The adjective *tall* in (14), thus, remains "unplugged", which accounts for its context sensitivity. This principle is more generally stated below.

> ***Plug-in principle***: *When a typicality-based representation T of an expression E1 takes part in the meaning composition of a complex expression E1-E2, contextual parameters of E2 are most sensitive to ("plugged-into") the parallel features in T, if such features exist.*

This principle extends the *head primacy principle* (Kamp and Partee 1995), which explains why typical instances of the expression *midget giant* are taller than typical instances of the expression *giant midget*. The head noun (*giant* and *midget*, respectively) is prominent in determining the relevant heights in a similar way to the primacy of the head *giraffe* in (15). The plug-in principle accounts for the linguistic circumstances where semantic composition with typicality is dependent on contextual parameters. In this way, contextual parameters intervene in the actual composition of linguistic meanings, and not only in the use of such meanings (Stalnaker 1970). Figure 1 summarizes the interaction of the principles introduced above within Content Sensitive Formal Semantics.

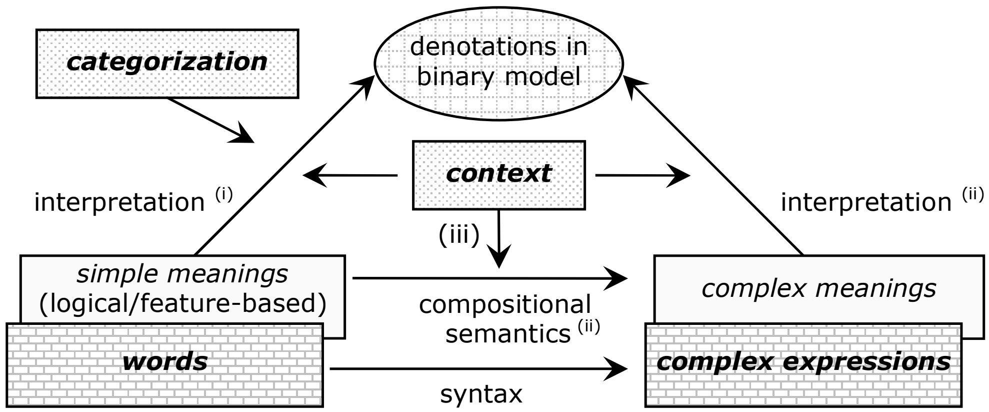

Figure 1 – Content Sensitive Formal Semantics – categorization and context in the binary model: (i) *The correspondence principle fixes interpretation of words*. (ii) *The Maximal Typicality Hypothesis concerns derivation of complex meanings*. (iii) *The context partakes in compositional semantics via the plug-in principle*.

4.4 Annotation and acquisition of linguistic information for textual entailment

As exemplified by (4), understanding entailment relations is crucial for question-answering, and similarly for other practical applications, including summarization, search engines and machine translation. Content Sensitive Formal Semantics will be used in this program for improving semantic processing in applicative computational linguistics. In order to achieve this aim, the following questions must be addressed:

1. What are the entailment relations that are most common in ordinary usages, and thus most useful for semantic processing in computational applications?
2. What are the most important factors that govern these entailment relations, and how can they be automatically recognized?

The PASCAL/NIST challenges (RTE 2004-7, Dagan et al. 2006) for recognizing textual entailment (RTE) have created a corpus of entailments from datasets of practical applications. This corpus is currently used for evaluating competing systems, and thus forms a practical platform for answering the first question. The program will answer the second, more foundational question by exploiting insights of Content Sensitive Formal Semantics (cf. Figure 3) in a computational system for recognizing data entailments. When addressing this question, the program will rely on two assumptions:

Assumption 1 – entailment annotation: In order to obtain a computational model of the most relevant factors affecting entailment, a big sample of exemplary entailments must be manually annotated using a theoretically-informed scheme.

Assumption 2 – entailment acquisition: In order to automatically recognize unseen entailments, machine learning algorithms should use the annotated entailments for

acquiring relevant linguistic/conceptual data from other resources.

Assumption 1 involves manual annotation of relations between content words/expressions. Assumption 2 is about the automatic acquisition of such relations from external resources. The two assumptions will be used for automatically recognizing semantic relationships between sentences on the basis of the developed Content Sensitive Formal Semantics.

For example, the entailment from (4I) to (4A) above can be decomposed into smaller pieces of information, embodied in the following elementary entailments:

(16) 1: *whaling* $\subset$ *whale hunting*

2: If $X \subset Y$ then *commercial X* $\subset$ *commercial Y*

In the example: since *whaling* entails *whale hunting* (cf. 16.1), also *commercial whaling* entails *commercial whale hunting.*

3: if $X \subset Y$ then *all X* $\subset$ *Y*

In the example, *all commercial whaling* in (4I) entails *commercial whale hunting* in (4A).

4: *1986 ban on X* $\subset$ *X was banned in 1986*

5: *proposal to allow X at Y's meeting* & *X was banned* $\prec$ *X was banned by Y*

In the example: given that the proposal to allow whaling came at IWC's meeting, is likely to conclude that also the ban was issued by IWC.

Entailment (16.1) is embedded in WordNet's indefeasible *lexical definition* (LD) of the verb "whaling" as "hunt for whales" (Fellbaum 1998). Entailment (16.2) illustrates the indefeasible upward monotonicity of *intersective modifiers* (IM) like *commercial* (Keenan and Faltz 1985). Entailment (16.3) illustrates the logical strength of the quantifier *all* over ordinary plurals and mass terms (Winter 2002), which can be described as *restrictive modification* (RM), embedded in computational semantic inference systems (Fyodorov, Winter and Francez 2003). Entailment (16.4) is a general property of *nominalization* (N), and reflects familiar semantic relations between English nouns (e.g. *ban*) and related verbs (*to ban*, Chierchia 1984). By contrast to these four indefeasible entailments, the entailment in (16.5) is based on defeasible higher-level world knowledge and pragmatic inference, and not on lexical-structural semantic information. Entailments (16.1-4) are annotated in Figure 2, directly on the respective expressions in the entailment (4I)$\subset$(4A).

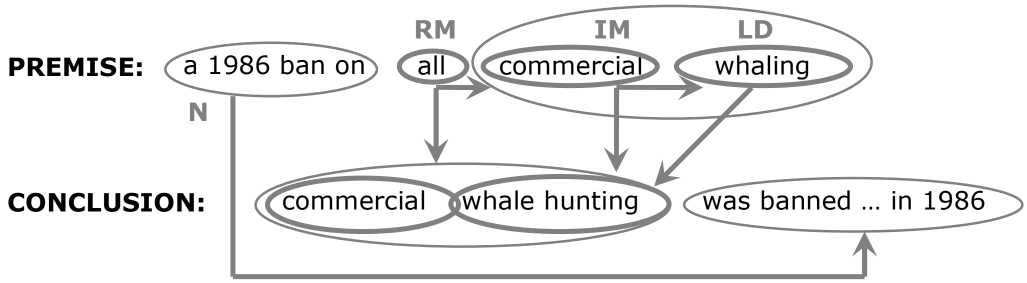

Figure 2: A*nnotation of restrictive modification* (RM), *intersective modification* (IM), *lexical definition* (LD) and *nominalization* (N) *in the entailment* (4I)$\subset$(4A).

In this way, information about useful entailments can be gleaned from existing resources. Assumptions 1 and 2 mean that machine learning techniques (Daelemans and Van den Bosch 2005, Lappin and Shieber 2007) require manual annotations of the entailment corpus for creating an *acquisition model*, which will be employed for automatically deriving a semantic classifier: an *Entailment Parser* that decides whether previously unseen entailments are supported or not by information in linguistic/conceptual resources

(WordNet: Fellbaum 1998; FrameNet: Johnson et al. 2002; Penn Treebank: Marcus et al. 1993; Cyc: Liu and Singh 2004, OpenCyc 2007). This configuration dictates the organization of the program as described in Figure 3.

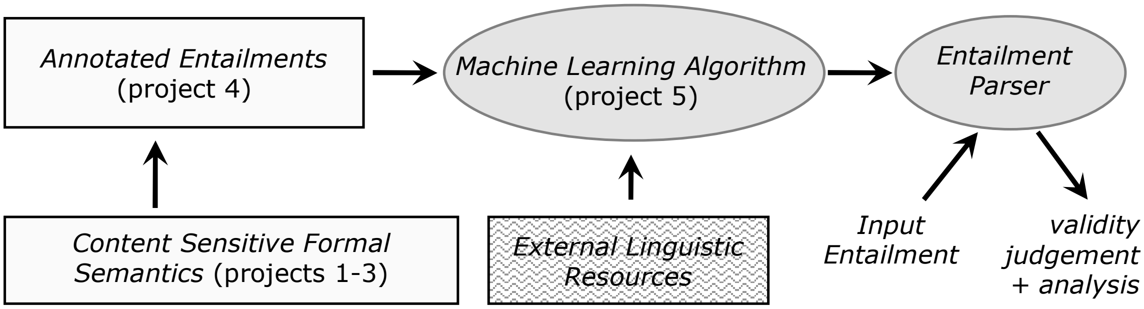

Figure 3: *Content Sensitive Formal Semantics for annotating entailments; Entailment Parser generated from annotated entailments and external resources.*

5. Program set-up

The program will consist of five sub-projects:

1. **PhD/Postdoc project in Formal Semantics – *Logic and the semantics of concepts*:** The project will model the effects of lexical meanings (e.g. of transitive verbs like *wash* or *praise*) on logical meanings (e.g. of reciprocal expressions like *each other* and *mutually*), employing theories of concepts in cognitive psychology, especially Prototype Theory.
2. **PhD/Postdoc project in Psycholinguistics – *Typilcality effects with logical expressions*:** The project will develop and run psycholinguistic experiments that check the effects of typicality on the interpretation of logical expressions, especially in complex reciprocal predicates (e.g. *stand on each other*), and compare these effects to parallel effects with modification constructions (e.g. *red hair*).
3. **PhD/Postdoc project in Formal Semantics – *Quantification and spatial expressions*:** The project will characterize the lexical meanings of spatial and mereological expressions (e.g. locative prepositions, verbs of movement and location) and the ways they affect quantification in natural language.
4. **PhD project in Computational Linguistics – *Formal semantic annotation of textual entailments*:** The project will develop a computational version of formal semantic theory that is suitable for annotating a corpus of textual entailments, and will accompany actual annotation of the RTE corpus. The annotated corpus and annotation principles will be used for the automatic acquisition of textual entailments in Project 5.
5. **PhD/Postdoc project in Computational Linguistics – *Automatic acquisition of textual entailments using semantic corpora*:** The project will use annotated entailments for developing machine learning algorithms that recognize unseen entailments on the basis of lexical/syntactic tools and resources. The learning model will use the annotated corpus of project 4 as a training set for acquiring semantically relevant information from other resources (e.g. *WordNet*, *Penn Treebank*, *OpenCyc*) when parsing entailments.

For more detailed project descriptions, see: http://www.phil.uu.nl/~yoad/vici/projects.html

References

Armstrong, S., L. R. Gleitman and H. Gleitman (1983). "What some prototypes might not be", *Cognition* 13:263-308.

Asher, N. and A. Lascarides (2003). *Logics of Conversation.* Cambridge University Press, Cambridge.

Barker, C. and P. Jacobson, editors (2007). *Direct Compositionality*, Oxford University Press, Oxford.

Bar-Haim, R., K. Sima'an and Y. Winter (2007). "Part-of-Speech tagging of Modern Hebrew text". Natural Language Engineering 14:223-251.

Bayer S., J. Burger, L. Ferro, J. Henderson and A. Yeh (2005). "MITRE's Submissions to the EU Pascal RTE Challenge". In *Proceedings of the first PASCAL Recognising Textual Entailment Challenge (RTE 1).*

Bekkerman, R., R. El-Yaniv, N. Tishby and Y. Winter (2003). "Distributional word-clusters vs. words for text categorization". *Journal of Machine Learning Research*, 3:1183-1208.

Bensley J. and A. Hickl (2008). "Application of LCC's GROUNDHOG System for RTE-4". In *Proceedings of Text Analysis Conference (TAC).*

Bos, J. and K. Markert (2005). "Recognising textual entailment with logical inference". *Proceedings of EMNLP-05*.

Boyd-Graber, J., C. Fellbaum, D. Osherson and R. Schapire (2006). "Adding Dense, Weighted Connections to WORDNET". *Proceedings of the Third Global WordNet Meeting*, Jeju, Korea, Jan. 2006. Seoul: Korean Advanced Institute of Science.
http://wordnet.cs.princeton.edu/papers/wordnetplusintro.pdf

Brewka, G., J. Dix and K. Konolige (1997). *Nonmonotonic Reasoning: An Overview*. CSLI Publications, Stanford.

Carlson, G. N. and F. J. Pelletier, editors (1995). *The Generic Book. University of Chicago Press,* Chicago.

Chierchia, G. (1984). *Topics in the Syntax and Semantics of Infinitives and Gerunds*. Amherst: University of Massachussets dissertation. [Published, New York: Garland Press, 1988).

Connolly, A. C., J. A.. Fodor, L. R. Gleitman and H. Gleitman (2007). Why stereotypes don't even make good defaults. *Cognition* 103:1-22.

Cruse, D. A. (1986). *Lexical Semantics.* Cambridge University Press, Cambridge.

Daelemans, W. and A. van den Bosch (2005). *Memory-Based Language Processing*. Cambridge: Cambridge University Press.

Dagan, I., O. Glickman and B. Magnini (2006). "The PASCAL Recognising Textual Entailment Challenge". *Lecture Notes in Computer Science,* volume 3944: 177-190.

Dalrymple, M., M. Kanazawa, Y. Kim, S. Mchombo and S. Peters (1998). "Reciprocal expressions and the concept of reciprocity". Linguistics and Philosophy 21:159-210.

Dowty, D. (1979). *Word Meaning and Montague Grammar: the semantics of verbs and times in Generative Semantics and in Montague's PTQ*. D.Reidel, Dordrecht.

Farkas, D. and H. de Swart (2003). The Semantics of Incorporation: from argument structure to discourse transparency. CSLI publications, Stanford.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press. Cambridge, Massachusetts. Tool available at http://wordnet.princeton.edu

Fodor, J. A. (1981). The present status of the innateness controversy. In *Representations: Philosophical Essays on the Foundations of Cognitive Science.* MIT Press, Cambrdige, MA.

Fyodorov, Y., Y. Winter and N. Francez (2003). "Order-based inference in natural logic. Logic" *Journal of the IGPL* 11:385-417.

Geach, P. and M. Black, M., editors (1970). *Translations from the Philosophical Writings of Gottlob Frege*. Blackwell, Oxford.

Gamut, L. T. F. (1991). *Logic, Language and Meaning*, volumes 1+2. The University of Chicago Press, Chicago.

Gärdenfors, P. (2000). "Concept combination: a geometrical model", pp. 129-146 in L. Cavedon, P. Blackburn,, N. Braisby and A. Shimojima (eds), *Logic language and Computation* Vol 3, CSLI, Stanford, CA.

Glinos, D. G. (2008). "Recognizing Textual Entailment at RTE4 with CERES". In *Proceedings of Text Analysis Conference (TAC).*

Groenendijk, J. and M. Stokhof (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. PhD dissertation, University of Amsterdam.

Hay, J., C. Kennedy and B. Levin (1999). "Scalar structure underlies telicity in `degree achievements'". In *Proceedings of Proceedings of Semantics and Linguistics Theory* 9.

Heim, I. and A. Kratzer (1997). *Semantics in Generative Grammar*. Blackwell.

Jackendoff, R. S. (1983). *Semantics and Cognition.* MIT Press. Cambridge, Massachusetts.

Johnson, C. R., C. J. Fillmore, M. R. L. Petruck, C. F. Baker, M. Ellsworth, J. Ruppenhofer and E. J. Wood (2002). FrameNet: Theory and Practice. International Computer Science Institute, Technical Report-02009. Berkeley, CA. See also J. Ruppenhofer et al. (2006), FrameNet II: Extended Theory and Practice, at: http://framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=126

Kamp, J. A. W. (1975). "Two theories about adjectives". In Keenan, E., editor, *Formal Semantics of Natural Language*. Cambridge University Press, Cambridge.

Kamp, H. and B. Partee (1995). "Prototype Theory and Compositionality". *Cognition* 57:129-191.

Katz, J. and J. A. Fodor (1963). "The structure of semantic theory". *Language* 39:170-210.

Keenan, E. (2000). "Logical objects", in C. A. Anderson and M. Zeleny (editors), *Logic, Meaning and Computation: Essays in Memory of Alonzo Church*. Kluwer, Amsterdam.

Keenan, E. and L. Faltz (1985). *Boolean Semantics for Natural Language*. D. Reidel, Dordrecht.

Kennedy, C. (2007). "Vagueness and grammar: the semantics of relative and absolute gradable adjectives". *Linguistics and Philosophy* 30:1-45.

Kerem, N., N. Friedmann, and Y. Winter (2009), "Typicality effects and the logic of reciprocity". To appear in *Proceedings of Semantics and Linguistics Theory* (SALT) 19.

Langendoen, D. T. (1978). "The logic of reciprocity". *Linguistic Inquiry* 9:177-197.

Lappin, S. and S. Shieber (2007). "Machine Learning Theory and Practice as a Source of Insight into Universal Grammar". *Journal of Linguistics* 43:393-427.

Link, G. (1983). "The logical analysis of plurals and mass terms: a lattice theoretical approach". In R. Bauerle, C. Schwarze and A. von Stechow (eds.), *Meaning, Use and Interpretation of Language*, De Gruyter, Berlin.

Liu, H. and P. Singh (2004). "Commonsense Reasoning in and over Natural Language". *Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems* (KES'2004). Wellington, New Zealand. Lecture Notes in Artificial Intelligence, Springer. See also the ConceptNet website: http://web.media.mit.edu/~hugo/conceptnet

MacCartney, B. and C. D. Manning (2007). "Natural logic for textual inference". In *Workshop on Textual Entailment and Paraphrasing at ACL 2007.*

Mador-Haim, S. and Y. Winter (2007). "Non-existential indefinites and semantic incorporation of PP complements". To appear in *Proceedings of Semantics and Linguistics Theory* 17.

Manning, C. D. (2006). "Local Textual Inference: It's hard to circumscribe, but you know it when you see it - and NLP needs it". MS, Stanford University. Downloadable from http://nlp.stanford.edu/~manning/papers.

Mansour, S., K. Sima'an and Y. Winter (2007). "Smoothing a lexicon-based POS tagger for Arabic and Hebrew". ACL Workshop on Computational Approaches to Semitic Languages.

Marcus, M., B. Santorini and M. Marcinkiewicz (1993). "Building a large annotated corpus of English: The Penn treebank". *Computational Linguistics* 19:313-330.

Margolis E. and S. Laurence (1999). "Introduction: Concepts and Cognitive Science". In Margolis E. and S. Laurence (editors), *Concepts: Core Readings.* MIT Press. Cambridge, Massachusetts.

McNally, L. and V. van Geenhoven (2005). 'On the property analysis of opaque complements', *Lingua* 115:885–914.

Montague, R. (1973). "The proper treatment of quantification in ordinary English", in J. Hintikka, J. Moravcsik and P. Suppes (editors), *Approaches to Natural Languages: proceedings of the 1970 Stanford workshop on grammar and semantics*, D. Reidel, Dordrecht.

OpenCyc project 2007: See http://opencyc.org/

Osherson, D. and Smith, E. (1981). "On the adequacy of prototype theory as a theory of concepts". *Cognition* 9:35-58.

Osherson, D. and E. E. Smith (1997). "On typicality and vagueness", *Cognition* 64: 189-206.

Partee, B. H. (2004). *Compositionality in Formal Semantics: Selected Papers by Barbara H. Partee*. Blackwell.

Pustejovsky, J. (1995). *The Generative Lexicon.* MIT Press. Cambridge, Massachusetts.

Rips, L. J. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking.* MIT Press. Cambridge, Massachusetts.

Roberts, C. (1987). *Modal Subordination, Anaphora, and Distributivity*. PhD thesis, University of Massachusetts at Amherst.

Rosch, E. (1978). "Principles of categorization". In R. Rosch and B. B. Lloyds (eds.), *Cognition and Categorization*. Erlbaum. Hillsdale, NJ.

RTE (2004-2007). *Recognising Textual Entailment PASCAL Challenge*. At the webpage: http://www.pascal-network.org/Challenges/RTE.

Sànchez-Valencia, V. (1991). *Studies on Natural Logic and Categorial Grammar*. PhD dissertation, University of Amsterdam.

Sabato, S. and Y. Winter (2005). "From Semantic Restrictions To Reciprocal Meanings". Proc. of *Formal Grammar/Mathematics of Language* (FG-MOL), Edinburgh.

Scha, R. (1981). "Distributive, collective and cumulative quantification". In J. Groenendijk, M. Stokhof and T. M. V. Janssen (eds.), *Formal Methods in the Study of Language.* Mathematisch Centrum, Amsterdam.

Schwarzschild, R. (1996). *Pluralities*. Kluwer, Dordrecht.

Shieber, S. M. (1988). *An Introduction to Unification-Based Approaches to Grammar*. CSLI Publications, Stanford.

Sima'an, K., A. Itai, A., Y. Winter, A. Altman, and N. Nativ (2001). "Building a tree-bank of Modern Hebrew text". *Traitment Automatique des Langues*, 42:347-380.

Smith, E. E. (1990). "Categorization", in D. N. Osherson, and E. E. Smith (eds.), *Thinking, An Invitation to Cognitive Science*, volume 3. MIT Press, Cambridge, Massachusetts.

Smith, E. E., D. Osherson, L. Rips and M. Keane (1988). Combining prototypes: A selective modification model. *Cognitive Science* 12:485–527.

Stalnaker, R. (1970), "Pragmatics." *Synthese* 22:272-289.

Stenning, K. and M. van Lambalgen (2007). *Human Reasoning and Cognitive Science*. To appear in MIT Press.

Taylor, J. R. (2003). *Linguistic Categorization*, 3rd edition. Oxford University Press, Oxford.

Van Benthem, J. (1987). "Meaning: interpretation and inference". Synthese 73:451-470.

Van Jaarsveld, H. J. and Draskovic, I. (2003). "Effects of collocational restrictions in the interpretation of adjective-noun combinations". *Language and Cognitive Processes* 18:47-60.

Van Strien, G. (2009). "Semantic Annotations for Automatic Recognition of Textual Entailments", MA Thesis, Utrecht University.

Vanderwende, L., D. Coughlin and B. Dolan (2005). "What Syntax can Contribute in the Entailment Task". In *Proceedings of the first PASCAL Recognising Textual Entailment Challenge (RTE 1)*.

Vanderwende, L., A. Menezes, D. Coughlin and R. Snow (2006). "*Microsoft Research* at RTE-2: Syntactic Contributions in the Entailment Task: an implementation". In *Proceedings of the second PASCAL Recognising Textual Entailment Challenge (RTE 2)*.

Winter, Y. (2001a). *Flexibility Principles in Boolean Semantics*. MIT Press, Cambridge, Massachusetts.

Winter, Y. (2001b). "Plural predication and the Strongest Meaning Hypothesis". *Journal of Semantics* 18:333-365.

Winter, Y. (2002). "Atoms and sets: a characterization of semantic number". *Linguistic Inquiry* 33:493-505.

Winter, Y. (2005). "Cross-categorial restrictions on measure phrase modification". *Linguistics and Philosophy*, 28:233-267.

Winter, Y. (2006). "Closure and telicity across categories", In *Proceedings of Semantics and Linguistic Theory* 16.

Wittgenstein, L. (1953). *Philosophical Invesitigations*, 3rd edition. G.E.M. Anscombe and R. Rhees (eds.), G.E.M. Anscombe (trans.), Oxford: Blackwell.

Zadeh, L. A. (1975). Calculus of fuzzy restrictions, in L. A. Zadeh et al. (eds.), *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, Academic Press.

Zimmermann, T. E. (1993). "On the proper treatment of opacity in certain verbs". *Natural Language Semantics* 1:149-179.

Zimmermann, T. E. (2006). “Monotonicity in opaque verbs”. *Linguistics and Philosophy* 29:715-761.

Zwarts, J. and Y. Winter (2000). "Vector space semantics: a modeltheoretic analysis of locative prepositions". *Journal of Logic, Language and Information* 9:171-213.