INTELLIGENT REASONING ON NATURAL LANGUAGE DATA:
A NON-AXIOMATIC REASONING SYSTEM APPROACH

A Thesis
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
MASTER OF SCIENCE

---

by
Ozkan Kilic
May 2015

Thesis Approvals:

Pei WANG, Advisory Chair, Dept. of Computer and Information Sciences
Yuhong GUO, Dept. of Computer and Information Sciences
Jamie REILLY, Dept. of Communication Sciences and Disorders

ABSTRACT

Research on Artificial General Intelligence has re-gained attention since the 2000s with a range of feedback from other disciplines, such as neurology, cognitive science, linguistics, psychology, philosophy, and such. NARS, a non-axiomatic reasoning system, is a general-purpose intelligent system able to work with insufficient knowledge and resources, and to adapt to its environment by learning from experience. It treats intelligence as a domain-independent capability with no domain-specific sub-module. Since the human mind evolved under the same restriction, this normative model displays many human-like properties.

NARS is used to reinterpret several well-known results in cognitive science, such as Wason's selection task, the Linda problem, and U-shaped learning, which cannot be explained by traditional normative models, but can now be handled by NARS in a unified way. This study specifically investigates the reasoning capabilities of NARS, a non-axiomatic reasoning system, on natural language data. NARS is used to mimic U-shaped learning of passive voice in English, subjective pronoun resolution, and contextual dependency of concepts. For this purpose, logical form from WordNet is translated to NARS. Furthermore, a convolutional neural network, which is available online and trained with images from ImageNet, is used to recognize possible noun categories of a given image.

The results have shown that a general-purpose system can simulate human-level behavior on language data without a built-in linguistic module.

To Timothy P. Benell.

ACKNOWLEDGMENTS

I am deeply grateful to my advisor Dr. Pei Wang for introducing me to the world of AGI, and for his guidance, suggestions, belief in my ability, and friendly attitude throughout the development of this thesis. He is an inspiration for what academic life should be about in Artificial Intelligence Research.

I would like to thank Dr. Yuhong Guo and Dr. Jaime Reilly for serving on my examining committee. In addition, I am thankful to my professors in the Department of Computer Science at Temple University, especially to Dr. Eugene Kwatny, Dr. James Korsh, Dr. Eduard Dragut and Sally Kyvernitis for their instruction, support, advice and friendly attitude.

I would also like to express my sincere thanks to my best friends, İlgar Seyidov, Yahya Hürol Sipahioğlu, Gökben Kaynar and Dr. Serkan Kükrer in Turkey, and also to my best friends Matthew Matula, Joyce Doman, Aimee Benell in the USA. Finally, my family and Timothy Patrick Benell deserve very special thanks for their eternal encouragement and altruistic support.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

For more than 60 years, Artificial Intelligence (AI) researchers have been working on creating intelligent machines or programs that can perform the intellectual tasks that a human being can. These tasks can vary greatly from playing chess, counting animals in a picture, performing commands given in a natural language to making jokes, having a conversation with a human, producing novel inventions and art. Weak-AI aims to apply human-like intelligence to problems, while the ultimate goal of strong-AI is to have systems that have their own minds. NARS (Non-Axiomatic Reasoning System) is an ongoing open-source project with the ultimate aim to become an Artificial General Intelligence (AGI). It is a general-purpose intelligent system that can function with insufficient knowledge and resources, and adapt to its environment by learning from experience (Wang, 2006, 2013a). Since it treats intelligence as a domain-independent capability, and attempts to cover all cognitive functions, NARS belongs to the emerging field of Artificial General Intelligence (Wang & Goertzel, 2007). This study specifically investigates the reasoning capabilities of NARS on natural language data.

1.1 Motivations

It is well-documented that the traditional logic, as represented by first-order predicate logic (FOPL), is not a proper descriptive theory of human reasoning. A famous example is Wason's selection task: A group of subjects are given a set of cards with letters on one side and numbers on the other side, and asked to evaluate the truthfulness of the statement “If a card has a vowel on one side, then it has an even number on the other side”. The statement can be represented as proposition *(for all x) (Vowel(x) → Even(x))* in FOPL. For example, if the cards display ***E***, ***K***, ***4***, *7*, respectively, subjects tend to choose only ***E*** or ***E*** and ***4*** whereas the correct answer would be ***E*** and *7* (Wason & Johnson-Laird, 1972). In other words, the subjects tend to *verify* to rule; however, they should have checked the cards that may *falsify* the rule according to the classical logic, since checking individual instances cannot verify a universally quantified statement, but can only falsify it.

Another established model of reasoning, probabilistic theory, also fails to provide a descriptive model for human reasoning. A well-known case here is the “conjunction fallacy”: In certain situations, the subjects tend to take the conjunction of statements as having a higher probability than a component statement of the conjunction. Take the “Linda problem” as an example. Given a description of Linda that fits people’s impression of feminist, many subjects take the statement “*Linda is a bank teller and is active in the feminist movement*” as more probable than “Linda is a bank teller”, though according to probability theory, the probability of “P and Q” cannot be higher than the probability of P, for any statements P and Q. (Tversky & Kahneman, 1983).

The fundamental disagreements between existing normative models and psychological observations are not limited to reasoning. For example, similar issues appear in learning, too. Various learning models (Bower & Hilgard, 1981; Sutton & Barto, 1997) turn to treat “learning” as a process in which cognitive systems gradually approach the correct answers or optimal solutions, while the reality of human cognition does not always work in this way.

Such mismatches between human cognition and classical logic are the main motivations for proposing NARS as a Cognitive Framework. Language data is selected to test the plausibility of this proposal.

1.2 Research Questions

The main hypothesis of this study is that NARS can be used as a Cognitive Framework. The study specifically focuses on using language data for reasoning with NARS because language is the uttermost product of human mind, and language data offer a challenge for any artificial intelligence system. The three research questions are: Despite having no language-specific module,

1. Can NARS simulate U-shaped learning in language learning?
2. Can NARS achieve pronoun resolution?
3. Can NARS display a change of belief which corresponds to contextual dependency of concepts?

In the following chapter, NARS and some other AGI systems are briefly introduced. U-shaped learning and pronoun resolution are discussed in the next chapter. Using WordNet and ImageNet to simulate contextual dependency of concepts in NARS is explained before the conclusion chapter.

CHAPTER 2

AGI SYSTEMS AND NON-AXIOMATIC REASONING SYSTEM (NARS)

Searle's (1980) Chinese Room Argument[i] was one of the early and famous reactions to strong-AI. Such reactions, as well as the complexity of strong-AI, have caused many researchers to work on solving various sub-problems with weak-AI approaches, which have resulted in uncombinable systems. On the contrary, humans can learn different problem solving approaches or algorithms and even combine and alter them to come up with novel ones. Furthermore, recent studies have claimed that everybody may be born synesthetic, which means humans' knowledge representations are initially so flexible that they can represent and combine different data modalities (shapes, colors, sounds, meanings) within the same ontology when they are infants. Though human knowledge representations may be specialized as we grow up, our analogy making and even creativity skills seem related to our minds' ability to unify different data modalities.

Research on general purpose systems returned stronger in 2000s. This time, the proponents of strong-AI have been better equipped with knowledge from other disciplines, such as neurology, cognitive science, linguistics, psychology, philosophy, and such. The progress of strong-AI is very central to the mind-body problem stated by Descartes because it may prove that there is no dualism and mind is a product of substance. This study favors strong-AI but not soon because of two reasons: First, strong-

---

[i] See http://plato.stanford.edu/entries/chinese-room/ for a discussion of the Chinese Room Argument

AI research heavily depends on relatively young fields like neurology and cognitive science. Therefore, it requires updates and knowledge transfers from these young fields. Second, there is a partially circular dependency: strong-AI researchers use their minds to deal with the mind problem and to create an artificial mind; and then, understanding mind will help the researchers to create an artificial mind. These make the strong-AI field improve incrementally and slowly, just as the human mind does. Thrun (1997) used the term “lifelong machine learning” for a system that is able to acquire knowledge through learning, retain or consolidate such knowledge, and use it for inductive transfer when learning new tasks. An AGI system needs to be able to transfer its experience to other domains in order to be a lifelong learning system. Thus, knowledge acquisition, representation, reasoning and knowledge transfer are the key components of an intelligent system. Most referred and ongoing AI systems with NLP or related features are reviewed in the following sub sections before NARS.

2.1 AGI Systems

2.1.1 SNePS

Stuart Shapiro’s SNePS[ii] currently has the most developed NLP feature. His research group released the version 2.8. They have been working on a new version, 3.0. The SNePS agent was written in LISP. Every proposition is parsed into a FOL-like representation and then they are represented as nodes in a network. Relations among propositions are represented by arcs in the network.

[ii] http://www.cse.buffalo.edu/sneps/

For example, “*Any robot that talks is intelligent”* is represented as in Figure 2.1 (Shapiro, 2000). SNePS uses many redundant nodes to represent semantics of agent, time, categories, action, act, event, and proposition-denoting terms. They also make use of a lexicon to parse sentences. The lexicon is not automatically incremental. A user needs to update the lexicon.

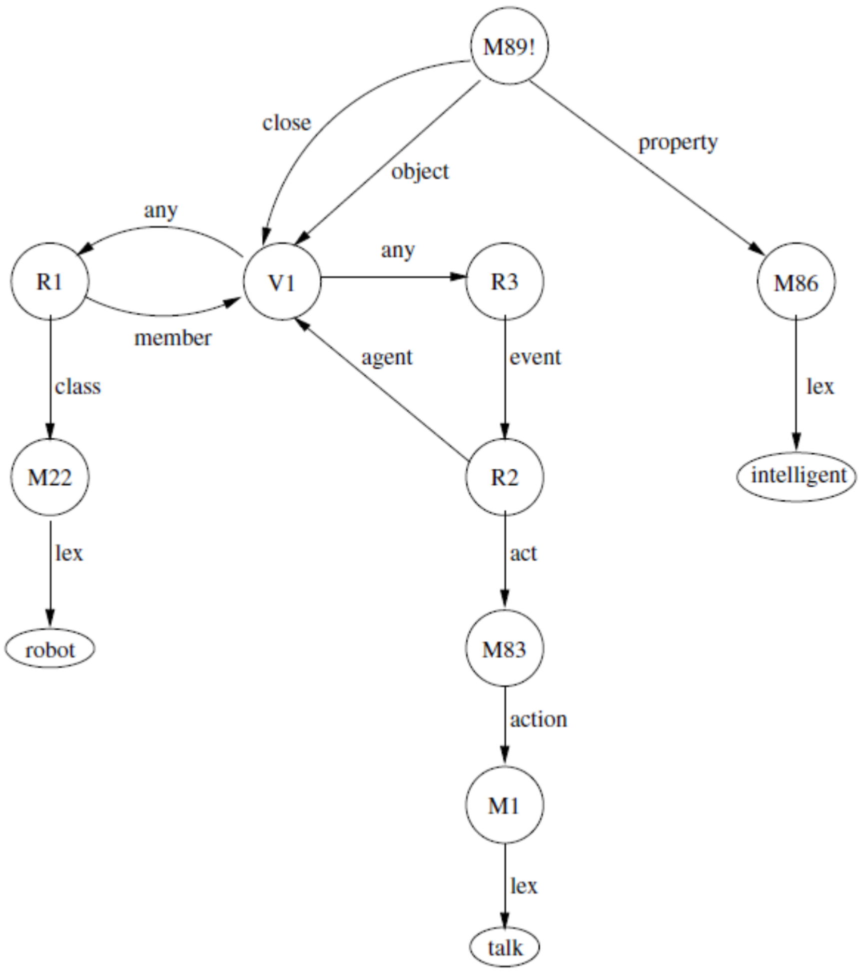

Figure 2.1: A propositional semantic network in SNePS

The network based representation provides flexibility: The SNePS agent may have different networks depending on the previous experience. Arcs can be used as flexible relations among propositions. However, everything depends on the lexicon, and the system cannot build up its own lexicon. Another problematic issue is the belief revision in SNePS. Assume that the following propositions are input to the system, "*All pirates are uneducated*", "*John is a pirate*", "*John is educated*". When the system comes across such a contradiction, it reports the contradiction, and then, either the system asks users to revise the contradicting propositions, or an automated belief revision method recently added to the system runs to determine *the least used proposition* and deletes it. However, this is not what a human being would probably do. Since the lexicon is handmade, it is easy to represent degrees quantification or levels or certainty. A better solution would be, revise the propositions to either one of these:

- John may be a pirate.
- John might be educated.
- Some pirates are uneducated.

Rapaport (2013) states that such semantic networks built online by SNePS agents have the ability to represent Meinongian Semantics (i.e., semantics of impossible objects in real life) because every SNePS term represents an inten*s*ional (mental) entity. For example, "*unicorns have big horns*", and "*A round square is round*" are Meinongian type propositions. Semantic networks build by an agent corresponds to a human's mental repository, and the tree-like structure makes reasoning faster. However, a SNePS framework requires use of an ANSI Common LISP interpreter, which makes it a less user-friendly environment.

2.1.2 LIDA

LIDA[iii] is a Java based cognitive framework using Global Workspace Theory[iv]. The architecture ties in well with both neuroscience and cognitive psychology, but it deals most thoroughly with “lower level” aspects of intelligence, handling more advanced aspects like language and reasoning only somewhat sketchily. The LIDA agent can run in a virtual world, sense the world (colors, numbers, obstacles), and achieves some tasks according to *cognitive cycles* as shown in Figure 2.2 (Franklin, Madl, D'Mello & Snaider, 2014). The LIDA framework mimics three types of learning: Perceptual, episodic and procedural. Perceptual learning concerns the learning of new objects, categories, relations, etc., represented as nodes in the perceptual memory. Episodic learning, on the other hand, involves learning to memorize specific events (i.e., the what, where, and when).

[iii] http://ccrg.cs.memphis.edu/

[iv] GWT resembles the concept of Working Memory, and is proposed to correspond to a "momentarily active, subjectively experienced" event in working memory. It closely related to conscious experiences.

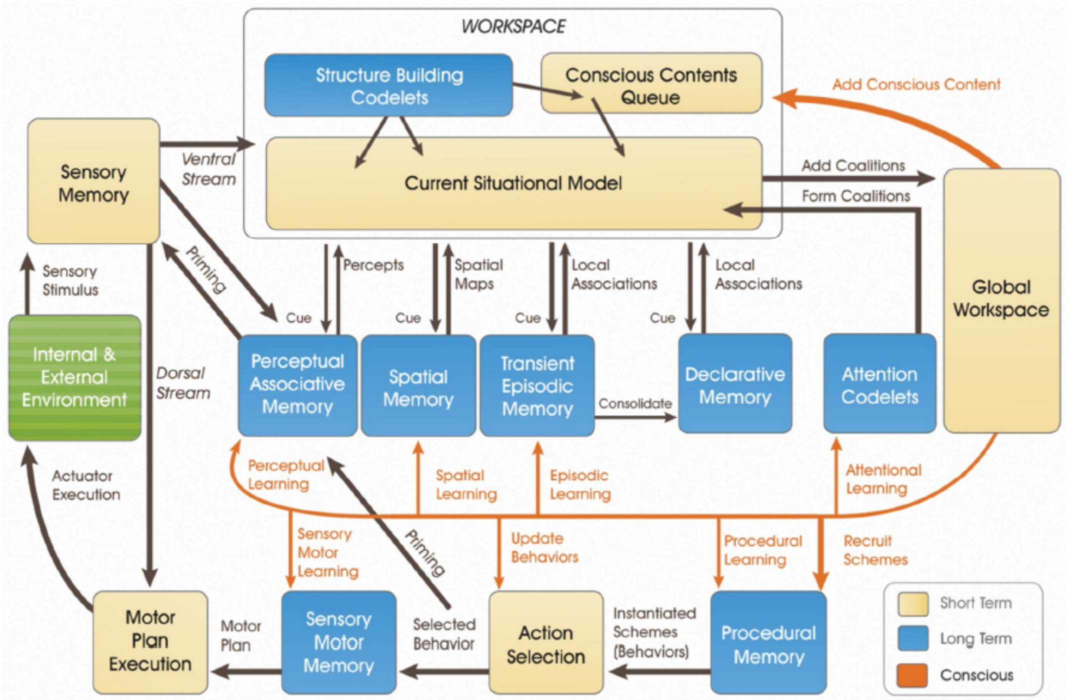

Figure 2.2: LIDA's Cognitive Cycle

Finally, procedural learning concerns the learning of new actions and action sequences with which to accomplish new tasks (e.g., "turn left, walk, grab, walk, turn right" might be a successful solution for the command "fetch my cup"). This architecture may explain many features of mind; however, it remains far away from understanding language, vision, and such. The developers claim that LIDA is the best implementation of "consciousness" because its agent is aware of time, space, itself and its experience. Despite dealing with lower level abilities, LIDA is a good at implementation of different memory types, and dealing with self-awareness, grounded knowledge and reasoning under uncertainty.

2.1.3 OpenCog (CogPrime, OpenCogPrime)

Ben Goertzel and his collaborators have been working on OpenCog Framework[v], which also uses the concept of cognitive cycle as in LIDA. It is a developed and detailed framework. The knowledge units of the system are called atoms, which can be nodes or links in a network representation. Atoms have truth and attention (which sets priority of the node) values. OpenCog has a probabilistic reasoning ability of the network representation, which is actually a hypergraph. Nodes can be conceptual, perceptual, procedural, psyche (goal and feeling of the agent) nodes while links can be logical, member, associative, execution, action, etc. It has a very complex architecture with intermediate modules. OpenCog uses hand-coded rules for Natural Language Processing, which are part of a module called RelEx. The semantic parses are then mapped to *atom* structures with links. The system also has a language generation module: SegSim, which is also rule-based. Here is how it works:

1. The NL generation system stores a large set of pairs of the form (semantic structure, syntactic/morphological realization).
2. When it is given a new semantic structure to express, it first breaks this semantic structure into natural parts, using a set of simple syntactic-semantic rules.
3. For each of these parts, it then matches the parts against its memory to find relevant pairs and uses these pairs to generate a set of syntactic realizations (which may be sentences or sentence fragments).

---

[v] http://opencog.org/

4. If the matching fails, then

   a. It returns to Step 2 and carries out the breakdown into parts again. But if this happens too many times, then

   b. It recourses to a different algorithm

5. If the above step generates multiple fragments, they are pieced together, and a certain rating function is used to judge if this has been done adequately (using criteria of grammaticality and expected comprehensibility, among others). If this fails, then Step 3 is tried again on one or more of the parts; or Step 2 is tried again.

6. Finally, a "cleanup" phase is conducted, in which correct morphological forms, articles and function words are inserted.

This method works for simple cases but would probably fail frequently (according to some examples given in Goertzel, Pennachin & Geisweiller, 2014). Moreover, it is not explained clearly enough.

2.1.4 ACT-R

ACT-R[vi] is a theoretical framework for emulating and understanding human cognition inspired by human brain. It aims at building a system performing the full range of human cognitive tasks. ACT-R aims to describe the mechanisms underlying perception, thinking, and action. It is composed of a set of perceptual-motor modules, different memory modules, buffers, and a pattern matcher. The perceptual motor modules basically serve as an interface between the system and the world. ACT-R has declarative

[vi] http://act-r.psy.cmu.edu/

memory (for factual information) and procedural memory (for procedures). ACT-R uses symbolic constructs (i.e., chunks or productions) created to describe the results of a complex operation. Such chunks may be available without re-computing the next time a similar task occurs (similar to memoization). ACT-R is a general purpose framework for psychological operations abstract reasoning, but creativity and transfer learning are still missing. A probabilistic natural language parsing model was implemented in ACT-R. Declarative memory was used as a lexicon while production rules (or grammar) were saved in procedural memory. However, ACT-R does not have a built-in NLP module. It provides reasoning and processing power given lexicon and grammar (see Lewis & Vasishth, 2005 for some examples). The advantage of using ACT-R for NLP is that intermediate phrasal constructions are automatically kept as chunks for later use, which makes parsing fast, incremental and even probabilistic.

2.1.4 SOAR

SOAR[vii] is a rule-based intelligent system. SOAR stores its knowledge in form of production rules, arranged in terms of operators that act in the problem space, which is composed of a set of states that represent the task at hand. It has a production memory (like a long-term memory) and a working memory (like a short-term memory). SOAR transforms every input into representations (symbols) and works on these representations. It has a production rule system (*if…then rules*) to modify representations or make decisions. SOAR assumes that intelligent action can be formulated as the selection and application of operators to a state, to achieve some goal. SOAR is not strong in areas such

---

[vii] http://sitemaker.umich.edu/soar/home

as episodic memory, creativity, handling uncertain knowledge, or reinforcement learning. Similar to ACT-R, SOAR is originally not designed for NLP but there are some implementations. For example, Mohan, Mininger, Kirk and Laird (2013) used SOAR to model situated language comprehension to make an agent perform commands such as "move the large red cylinder to right of the blue triangle". They indexed and defined all objects, properties and actions before the agent performed actions. Moreover, some independent researchers worked on NL-SOAR which integrates English grammar with WordNet data but it is no longer an active study.

2.2 NARS and Its Architecture

The AGI model used in this study is NARS (Non-Axiomatic Reasoning System) (Wang, 2006, 2013a). NARS is a general-purpose intelligent system with ability to work with insufficient knowledge and resources, and to adapt to its environment by learning from experience.

Wang (2011) defines three properties for insufficient knowledge and resources:

- Finite: The system's hardware includes a constant number of processors (each with a fixed maximum processing speed), and a constant amount of memory space.
- Real-time: New knowledge and problems may come to the system at any moment, and a problem usually has time requirement for its solution.
- Open: The system is open to knowledge and problem of any content, as far as they can be represented in the format acceptable by the system.

NARS is inspired from nature because the system's knowledge is its internal connections that link its internal drives and external sensations to its actions and reactions. These connections are either innate or acquired from its experience. The insufficiency of knowledge and resources is a biological constraint by nature. The system has finite information-processing capability, though it has to be open to novel situations in its environment. These situations are not under the system's control. Furthermore, the system has to respond to them in real time. Otherwise it cannot survive. In this interpretation, the insufficient knowledge and resources drive intelligence because an intelligent animal (or a system) has to find *the good enough solution* to overcome natural selection.

NARS has a special language for knowledge representation, an experience-grounded semantics of the language, a set of inference rules, a memory structure, and a control mechanism. The non-axiomatic logic is used for adaptation with insufficient knowledge and resources, operating on patterns that have the "truth-value" evaluated according to the system's "experience" by using these patterns. This approach allows for emergence of experience-grounded semantics, and inferences defined by judgments. Since it treats intelligence as a domain-independent capability, and attempts to cover all cognitive functions, NARS belongs to the emerging field of Artificial General Intelligence (Wang & Goertzel, 2007). More information about NARS can be found online[viii], and in this paper the system is only briefly described. The overall architecture and procedure of NARS are given in Figure 2.3 (Wang, 2013a). As a reasoning system, the major components of NARS includes

[viii] https://sites.google.com/site/narswang/

- an inference engine with routines that implement the inference rules
- a memory that stores tasks and beliefs, as well as other information
- a few buffers that hold the data items under processing
- a few input/output channels that serve as the system's interface with the environment

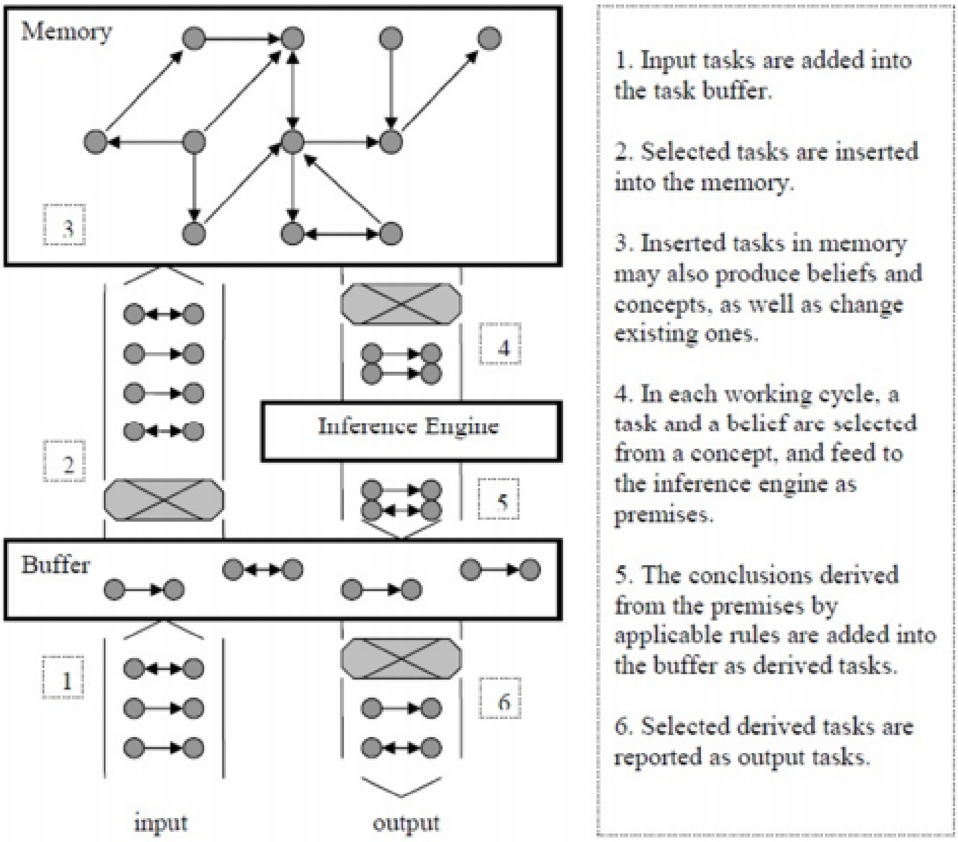

Figure 2.3: NARS Architecture.

In NARS, the basic form of knowledge is an *inheritance relation*, $S \rightarrow P$, from a *subject term S* to a *predicate term P*. Intuitively, it means that "S is a specialization of P, and P is a generalization of S". For example, *bird* → *animal* roughly means "Bird is a kind of animal" in English. Given the assumption of insufficient knowledge, $S \rightarrow P$ is usually uncertain, and the uncertainty can be measured according to available evidence. For a statement $S \rightarrow P$ and a term M, if M is in the extensions of both S and P (i.e., being a common instance of the two), it is positive evidence; if it is the extensions of S but not the extension of P, it is negative evidence. Symmetrically, if M is in the intensions of both P and S (i.e., being a common property of the two), it is positive evidence; if it is in the intension of P but not the intension of S, it is negative evidence. For a given concept, the system's beliefs about it may be either *extensional* (indicating its instances or *special cases*) or *intensional* (indicating its properties or *general cases*). All these beliefs together determine the meaning of the concept.

Beliefs in NARS are not "absolute truth" but summaries of the system's experience. The *truth-value* of a statement measures its *evidential support*, and the evidence can be either *positive* or *negative*, depending on whether it agrees with the statement, as mentioned above. A truth-value consisting of two factors: *frequency* and *confidence*. Wang (2013a) simply formulates *frequency* as $f = w^{+}/w$, and *confidence*, as $c = w/(w+k)$, where k is a positive constant, which is the system parameter for "evidential horizon". While the uncertainty within *frequency* is caused by *negative evidence*, the uncertainty about *frequency* is caused by *future evidence*. The second measurement, *confidence*, is introduced for the latter.

The frequency is the percentage of positive evidence for the statement among all evidence, whereas the confidence is the percentage of current evidence among all evidence at a constant future (such as after the next piece of evidence comes), so it indicates how strong or stable the frequency is. In other words, it is an indicator for how easy it is for the system to change its mind on this matter. A higher *confidence* value does not mean that the *frequency* is closer to the "objective probability", but that the *frequency* is harder to be changed by new evidence. For example, the statement "With 0.9 confidence, I 100% believe that *broken* is the passive form *break*." can be represented as "*("break" × "broken") → passive <1, 0.9>*" in NARS's grammar. This representation states that *broken* and *break* has a relation *passive* between the two. Since it is the only evidence, the *frequency* is 1. However, 0.9 confidence is set by default to state that the system is confident about its belief but open to change it in future.

Using the given experience as premises, the inference rules of NARS carry out various types of inference, including *deduction, induction, abduction, revision, choice, analogy*, etc. Each rule has a truth-value function that determines the truth-value of the conclusion according to those of the premises and the type of the inference.

The following subsection discusses a NARS solution to the two problems introduced in Chapter 1, Wason's Selection Task and Linda Problem in order to show that NARS can provide a better simulation of human reasoning than classical approaches.

2.1.4 NARS Replies to Wason's Selection Task and Linda Problem

Since the treatment of these two problems in NARS have been discussed in previous publications (Wang, 2001; Abdel-Fattah et al., 2012), the arguments are just summarized here. When Wason's selection task is represented in NARS, the truth-value of a statement is determined both by positive evidence and negative evidence, as introduced previously. In this way, the statement "If a card has a vowel on one side, then it has an even number on the other side" is not interpreted as a universally quantified proposition, but more as a statistical statement where the probability is to be determined by the given evidence.

If the subject has considered all the possibilities and there is no other factor, then the correct decision is to turn over the ***E***, ***4***, and *7* cards, but not the ***K*** card. This is the case, because the ***E*** card and the ***4*** card may provide positive evidence, and the ***E*** card and the *7* card may provide genitive evidence. However, NARS assumes that the subject works with insufficient knowledge and resources, therefore may fail to consider certain possibility. Since NARS runs with limited resources and knowledge, we cannot expect such a system to make perfect decisions. In general, the ***E*** card should be selected most often, because it corresponds to two possible ways to provide evidence. On the contrary, the ***K*** card should be selected least often, because it is not evidence, and its selection is only caused by the random factors. Between the ***4*** card and the *7* card, though they are logically symmetric, there are at least two factors that in favor of the former: one is the extra step of negation needed by the *7* card to be recognized as a piece of negative evidence, and the other is the priming effect on even numbers caused by the mentioning of "even number" in the description of the task.

Here the key point is that in FOPL, truth-value is determined by the existence of negative evidence only, while in NARS both positive and negative evidence matter.

If the statements in Linda problem are input to NARS, the system will consider all available evidence to decide the extent to which *Linda* is an instance of another concept, *bank-teller* or *feminist bank-teller*. Given the design of the experiment, the most accessible information about all three concepts are *intensional* (i.e., about their properties), so the system reaches its conclusion by checking if *Linda* has the properties usually associated with *bank-teller* and *feminist bank-teller*, respectively. Since according to the given information the concept *Linda* has more common properties with *feminist bank-teller* than with *bank-teller*, its degree of membership is higher to the former than to the latter.

Comparing NARS and probabilistic model, here there are two key issues: first, what the statements to be compared are. According to the traditional interpretation of the experiment, they are the statement *Bank-teller(Linda)* and the statement *Bank-teller(Linda) ∧ Feminist(Linda)*, while in NARS, they are the statement *{Linda} → bank-teller* and the statement *{Linda} → (feminist ∩ bank-teller).* The latter is not the conjunction of two statements, but a statement whose predicate term is an intersection of two terms. The second issue is how the degree of belief is defined. According to the traditional interpretation, it is the probability for the event described by the statement to occur, or the probability for Linda to fall into the involved set. According to this interpretation, it is clearly less probable for Linda to fall into two sets than into one of the two. On the contrary, in NARS concept membership evaluation can be based on extensional evidence, intensional evidence, or a mixture of the two (Wang & Hofstadter 2006). If all evidence are intensional and one concept is a special case of the other one, then if the instance has most of the properties involved, its relationship with the specific concept will be stronger, since it shares more properties with it. For example, it can be assumed that most human beings will consider themselves more as a human than as an animal, though "by definition" humans are animals.

Now whether these experiments reveal human fallacies or leads into the following questions: Should we accept the FOPL assumption that truth-value depends on negative evidence only or the NARS assumption that truth-value depends on both positive and negative evidence? Should we accept the probability theory assumption that membership depends on extensional evidence only or the NARS assumption that membership depends on both extensional and intensional evidence? The traditional models (FOPL and probability theory) choose their assumptions mainly for mathematical reasons, while the assumptions of NARS are chosen mainly based the considerations of AI and cognitive science.

CHAPTER 3

NARS FOR U-SHAPED LEARNING AND PRONOUN RESOLUTION

In this chapter, U-shaped learning and pronoun resolution in language learning and production are studied via NARS. In cognitive development, U-shaped learning, or U-shaped development, is a type of learning in which the learner first learns the correct behavior, then leaves it, and finally reacquires it. Pronoun resolution is simply finding the noun being referred by a pronoun. Since both are experience- and memory-related phenomena, they are appropriate to be tested in NARS. In the following sections, firstly, U-shaped learning is simulated in the acquisition of passive voice construction in English. Then, subjective pronouns are used as examples for the resolution task.

3.1 U-shaped Learning

U-shaped learning can be observed in various phenomena, such as language acquisition (Marcus et al., 1992), object permanence (Bowerman, 1982), and face recognition (Carey, 1982). A famous example is learning English passive voice: children first learn that *added*, *worked*, *hanged* and *eaten* are passive voices. When children realize that there is a rule which is the addition of *-ed* to a verb in order to make it passive, they produce ungrammatical construction such as *eated* and *telled* due to overgeneralization of the rule. Finally, they correct their errors and acquire that although the *-ed* rule is valid, there are some verbs that do not obey it, such as *eaten* and *told*.

Carlucci and Case (2013) claim that without U-shaped learning fewer classes of languages would be learnable; and U-shaped learning is necessary for full learning power within Gold's formal model of language learning from positive data (Gold, 1967). Gold's theorem states that "no negative evidence" results in the Logical Problem of the Language Acquisition (LPLA). LPLA can be stated that since children are not exposed negative evidence, then it is possible that they would acquire language the target language, plus some more, which means a super set of the target language. In other words, people need knowledge of the structure of their language for which insufficient evidence is available in the data to which they are exposed as children (Hornstein & Lightfoot, 1981). LPLA discussion is one of the central debates between connectionist and nativist scholars (see Pinker, 1979; Chomsky, 1980; Cowie, 1999; Smith, 1999; Pullum & Scholz, 2002; Johnson, 2004; Heinz, 2010; Clark & Lappin, 2010 for some pro and con arguments and implications). Carlucci and Case (2013) suggest that humans might exhibit U-shaped and other non-monotonic learning patterns because otherwise it would be impossible for them to learn what they need to learn to be competitive due to the limitations of human memory and computational power. Therefore, a computational model demonstrating U-shaped learning and non-monotonic learning patterns by nature can be employed as a test-bed in Cognitive Science to study human learning.

3.1.1 U-shaped Learning of Passive Voice by NARS

As mentioned in Chapter 2, a truth-value in NARS consists of two factors: *frequency* and *confidence*. The frequency is the percentage of positive evidence for the statement among all evidence, whereas the confidence indicates how strong or stable this degree of statement is, according to the amount of evidence supporting it. For example, the statement "With 0.9 confidence, I 100% believe that *eaten* is the passive form *eat*."

can be represented as “*(“eat” × “eaten”) → passive <1, 0.9>*” in NARS’s grammar. This representation states that *eat* and *eaten* has a relation *passive* between the two.

Since its truth-values are determined by available evidence, they can be revised when new evidence becomes available. NARS can innately simulate U-shaped learning in case of conflicting/competing information. Table 3.1 is a simple simulation of U-shaped learning of English passive voice. Note that the italic sentences are translation of the NARS statements.

Table 3.1: Simulation of U-shaped learning of English

Steps
1.Initial input as experience
(“eat” × “eaten”) → passive
Eaten is the passive form of eat
(“break” × “broken”) → passive
Broke is the passive form of break
(“add” × “added”) → passive
Added is the passive form of add
(“ask” × “asked”) → passive
Asked is the passive form of ask
2. Output
(“eat” × “eaten”) → passive <1, 0.9>
With 90% confidence, I 100% believe that eaten is the passive form eat

Table 3.1: Simulation of U-shaped learning of English (continued).

Steps
3. Append-ed operations is defined to the system as a simple concatenation
(“add” × “added”) → append-ed
(“ask” × “asked”) → append-ed
(“need” × “needed”) → append-ed
(“eat” × “eated”) → append-ed
4. Inductive conclusion outputted by the system
($x → append-ed) ⇒ ($x → passive) <1, 0.45>
With 45% confidence, I 100% believe that an instance of the passive relation is also an instance of the append-ed relation.
5. Revision by evidence accumulation
($x → append-ed) ⇒ ($x → passive) <1, 0.62>
6. This causes the system to take a wrong step
(“eat” × “eated”) → passive <1, 0.56>
With 0.56 confidence, I still 100% believe that eated is the passive form eat.

Table 3.1: Simulation of U-shaped learning of English (continued).

Steps
7. Counter evidence to the inductive conclusion
(“eat” × “eaten”) → passive
¬ ((“eat” × “eated”) → passive)
eat and eated have no relation of passive.
8. Inductive conclusion, revised again
($x → append-ed) ⇒ ($x → passive) <0.67, 0.72>

When the initial knowledge shown in Step 1 is given to the system as input, the system acknowledges that *eat* and *eated* have a relation: passive, and assigns a default truth-value to the belief, as shown in Step 2. The same happens to the other input lines, though the corresponding beliefs are not listed here. In Step 3, the morphological relation *append-ed* is established between the corresponding word-pairs, which indicates that the first word can become the second in the pair after concatenation of *-ed* as a suffix. The acknowledgements of these inputs are omitted.

When noticing a common instance of the *append-ed* relation and the *passive* relation, the induction rule of NARS generalizes this relation to instances of these two relations. Since this conclusion is only based on a single piece of evidence, its confidence is relatively low, as shown in Step 4. However, another common instance of the two relations will produce the same inductive conclusion, and the revision rule of the system will merge the two into a more confident conclusion, as shown in Step 5. Similarly, the other given positive evidence will increase the confidence of the inductive conclusion, though we will not go through all the details here.

At this moment, if the system sees that "*eat*" becomes "*eated*" by adding the suffix "*-ed*", it will do a deduction using this information and the previously derived conclusion, and take the "*eated*" to be the passive form of "*eat*", as shown in Step 6. The user can correct this error by providing a negative judgment of this conclusion, together with the correct passive form of the word, as in Step 7. Please note that in NARS a single counter example will not "falsify" a general statement, but decrease its frequency value. Now the inductive conclusion is revised again, into the conclusion shown in Step 8. If such counter examples are met again and again, the frequency value of this "general rule" will become lower, and the system will not depend on it to find the passive form of words, but depend on the specific word-pairs that have been remembered.

Finally, when more and more evidence is collected, the general rule of adding "-*ed*" will be judged as valid for most verbs, and it will co-exist with the known counter examples. For a concrete question about the passive form of a given word, the remembered answer directly matching the question $(word \times ?x) \rightarrow passive$ is usually favored over the conclusion produced by the "statistical rule", since the former has a higher confidence value. It is only when there is no counter example recalled, will the general rule to be applied to answer the question. In this way, the U-Shaped learning process is reproduced.

Counter evidence against system's beliefs are tricky. Yet, they do exist and cause U-shaped learning curves which should be captured by any system aiming to model human cognition. For example, Perfors et al. (2010) have been proposing Bayesian models as a cognitive framework; but they state that the *odds* such as U-shaped acquisition should be modeled as the memorization of individual data points in Bayesian models. NARS and Bayesian approach are based on different assumptions. In Bayesian model, whether an event will happen, or whether a proposition is true, is uncertain, but their probability, or degree of uncertainty, is usually certain. On the contrary, in NARS the insufficiency of knowledge and resources is consistently and completely assumed. NARS may contain (explicitly or implicitly) conflicting judgments. To handle them, NARS has both updating rule and revision rule, whereas the latter is not available in a Bayesian model, because the information about confidence is absent there (Wang, 2006).

As for the related issue of LPLA, or the "poverty of negative evidence", NARS does not always depend on the user to explicitly provide negative evidence as shown in the above example (though such evidence is acknowledged and used in belief revision). A major source of negative evidence is failed anticipation. When the system uses wrong conclusion (like "*eated*" as the passive form of "*eat*") in practice, it will cause problems so the system cannot achieve its goal as anticipated. Such feedback is used by NARS as negative evidence of the beliefs responsible.

3.2 Subjective Pronoun Resolution by NARS

There are ambiguities in word-, and even, suffix-level constituents requiring resolution in all world languages, mainly because of the conflict between contrast and efficiency; a user of a language needs to convey meaning as clearly as possible and in an *efficient* and *economical* way (Siddiqi, 2009). The optimization between the two gives rise to some constituents to be resolved. Pronoun resolution is mainly due to the economy constraint because pronouns are generally short words (or *null* elements in some languages) replacing other nouns, phrases, or even complete sentences. For example, *he* in (1) refers to *Tim* while *which* stands for the complementizer phrase *that he would be on time*.

(1) [*Tim*]$_i$ told Gina that [[*he*]$_i$ would be on time]$_k$, [*which*]$_k$ was actually a lie.

Psycholinguistic studies have shown that adults coordinate semantic and syntactic information, contextual clues, and statistical regularities in order to parse and resolve linguistic input (Altmann & Steedman, 1988; Trueswell, 1996; Papadopoulou, Peristeri, Plemenou , Marinis & Tsimpli, 2015) while the situation is not very clear for children (Clahsen & Felser, 2006; Papadopoulou & Tsimpli, 2005). Infants usually construct initial grammatical categories by mapping words to conceptual categories by noticing that words denoting objects, words denoting actions, words denoting properties belong to separate classes, and such (Gentner, 1982). Pronouns are one of them.

Children start to use pronouns around the age of two. Repeating the same sentence with an appropriate pronoun or using verb pairs that logically entail each other are heuristics or some methods to teach children how to use and resolve pronouns. In other words, regularities, mainly with respect to verbs, are main indicators for children to acquire pronouns (Laakso & Smith, 2006). These regularities are due to broad semantic categories of nouns and verbs. Verbs are important heuristics for pronouns resolution. Verbs are particularly relational entities whose meanings are usually not directly perceptible (see Hirsh-Pasek & Golinkoff, 2006). An exception is psychological state verbs like *look*, *think*, *want*, *believe* and *know*. Yet, meaning maps between most of the verbs and the world are not transparent. Children may learn verb meanings through their relations to other words in the input or use default relations of verbs to understand other words, such as pronouns.

Further details, such as gender and number, for example in child-directed speech (CDS), help children to master the pronoun resolution (see Laakso & Smith, 2006 for the relation between regularities in CDS and pronoun resolution). Children show distinct resolution success rates usually with U-shaped development even at the age of 10-to-11 years (Papadopoulou et al., 2015). McCloskey (2006) calls pronouns "elsewhere" elements because they refer to some other elements that are already in or will be memory. Sichel (2014) states that some elements in memory compete to be referent, specifically, for gaps while heuristics and clues determine the winner.

Studies have shown that pronoun comprehension depends on the accessibility of potential referents (see Arnold, 1996). For example, *he* in "*Tim helped Oscar when he was at home*" is generally resolved as *Tim* because it comes first in the sentence. In addition, pronoun resolution is also guided by gender information. For example, it is very unlikely for native speakers of English to resolve *he* in (1) as *Gina*. Although some studies following the Minimalist Hypothesis (McKoon & Ratcliff, 1992) claim that the accessibility factors are prior, Arnold, Eisenband, Brown-Schmidt and Trueswell (2000) experimentally showed that gender and accessibility clues are used almost in parallel. Several other studies (Garvey, Caramazza & Yates, 1976; Pickering & Majid, 2007) emphasized the verb-bias in pronoun resolution. For example, when asked to complete the sentences given in (2), some people choose *he* as *father* for the verb *confess* while they select *son* for the verb *scold*.

(2) The *father* confessed to his son because *he*. . .”

“The father scolded his *son* because *he*. . .”

In a neuropsychological study, McMillan, Clark, Gunawardena, Ryant and Grossmana (2012) found fMRI evidence that pronoun resolution cannot be a core linguistic mechanism problem but must involve decision making task which takes several heuristics and even risks associated with the competing referents. Some researchers claim that infants’ initial knowledge of language is not syntactic, but that abstract grammatical categories are established around the age of 2-3 (Tomasello, 2000). Other researchers claim that category learning is about mapping the input to preexisting abstract syntactic categories (Valian, 2009). In the current study, the resolution of English subjective pronouns is simulated in NARS parallel to human.

3.2.1 Subjective Pronoun Resolution by NARS

The current study is neutral about the preexistence of categories for pronoun resolution. However, it claims that pronouns can be resolved by an artificial general intelligence system which has no language-specific logic. In a toy experiment in NARS, pronoun resolution is simulated. Reasoning with respect to verbs together with some aspects of words, such as gender and number, are main indicators for the system to resolve pronouns as shown below where italic sentences are translations in English. Note that the system is reset in each experiment; thus, they are independent of each other. Furthermore, the system is told that nouns and pronouns are similar to some extent. Although a pronoun can be semantically identical to a noun, there are indeed (morpho)syntactic differences: A pronoun cannot be preceded by an adjective to form an adjectival phrase or cannot be pluralized by simply adding *-s* as a suffix.

(3) **Input:**

{“John”, “Mary”, “cats”, “meat”, “water”} → nouns.

John, Mary, cats, meat and water are instances of nouns

{“he”, “she”, “it”, “they”} → pronouns.

He, she, it and they are instances of pronouns

pronouns ↔ nouns <0.8, 0.9>.

Pronouns and nouns are similar to some extent.

(“John” × “meat”) → eat

John eats meat.

(“he” × “meat”) → eat

He eats meat.

“he” ↔ ?x ?

Who is he?

Output:

(“John” × “meat”) ↔ (“he” × “meat”) <0.8, 0.45>.

With 45% confidence, the system 80% believes that the relation between ‘he’ and ‘meat’ is identical to the relation between ‘John’ and ‘meat’.

“he” ↔ “John” <0.8, 0.45> .

With 45% confidence, the system 80% believes that ‘he’ is ‘John’.

The experiment in (3) represents the condition in which the sentence "*John eats meat*" with the overt subject "*John*" is repeated with the covert subject "*he*" as "*He eats meat.*" The default frequency and confidence values, which are 1 and 0.9 respectively, are used in the input. Due to the immediate repetition of two statements, the system concludes first that the relation *like* is identical in two statements, and then that "*he"* can be resolved as "*John"*. Another way is to see pronoun resolution as a question answering activity, that is, to handle "("he" × "meat") → eat" as a question "(?1 × meat) → eat" (Who eat meat?), and the answer *"John"* will replace *"he"*. The toy experiment (4), the logical entailment "if someone is hungry, he eats meat" is used to resolve the pronoun *"he"* as *"John"*.

(4) **<u>Input:</u>**

{"John", "Mary", "cats", "meat", "water"} → nouns.

John, Mary, cats, meat and water are instances of nouns

{"he", "she", "it", "they"} → pronouns.

He, she, it and they are instances of pronouns

pronouns ↔ nouns <0.8, 0.9>.

Pronouns and nouns are similar to some extent.

($x → [hungry]) ⇒ ($x × "meat") → eat).

If someone is hungry, that person eats meat.

"John" → [hungry]

John is hungry.

("he" × "meat") → eat.

He eats meat.

"he" ↔ ?x ?

Who is he?

Output:

"he" → [hungry]. <0.8, 0.81>.

With 81% confidence, the system 80% believes that he is hungry.

"he" ↔ "John". <0.8, 0.42>.

With 42% confidence, the system 80% believes that 'he' is 'John'.

The system's confidence in (4) is lower than the one in (3) because more intermediate logical mechanism are involved in the reasoning process. The example (5) simulates system's learning gender.

(5) **Input:**

{"John", "Mary", "cats", "meat", "water"} → nouns.

John, Mary, cats, meat and water are instances of nouns

{"he", "she", "it", "they"} → pronouns.

He, she, it and they are instances of pronouns

pronouns ↔ nouns <0.8, 0.9>.

Pronouns and nouns are similar to some extent.

"Mary" → [female].

Mary is female.

($x → [thirsty]) ⇒ ($x × "water") → drink).

If someone is thirsty, that person drinks water.

"Mary" → [thirsty].

Mary is thirsty.

(“she” × “water”) → drink).

She drinks water.

“she” ↔ ?x ?

Who is she?

<u>Output</u>:

“she” → [thirsty]. <0.8, 0.81>.

With 81% confidence, the system 80% believes that she is thirsty.

“she” ↔ “Mary”. <0.8, 0.42>.

With 42% confidence, the system 80% believes that she is Mary.

“she” → [female]. <0.8, 0.38>.

With 38% confidence, the system 80% believes that she is female.

The final experiment in (6) represents the learning of number for pronouns.

(6) **<u>Input:</u>**

{“John”, “Mary”, “cats”, “meat”, “water”} → nouns.

John, Mary, cats, meat and water are instances of nouns

{“he”, “she”, “it”, “they”} → pronouns.

He, she, it and they are instances of pronouns

pronouns ↔ nouns <0.8, 0.9>.

Pronouns and nouns are similar to some extent.

(“John” × “meat”) → eat

John eats meat.

“cats” → [plural].

Cats are plural.

“cats” → [hungry].

Cats are hungry.

($x → [hungry]) ⇒ ($x × “meat”) → eat).

If someone is hungry, that person eats meat.

(“they” × “meat”) → eat).

They eat meat.

“they” ↔ ?x ?

Who are they?

<u>Output</u>:

“they” → [hungry]. <0.8, 0.81>.

With 81% confidence, the system 80% believes that they are hungry.

“they” ↔ “cats”. <0.8, 0.42>.

With 42% confidence, the system 80% believes that they are cats.

“they” → [plural]. <0.8, 0.36>.

With 36% confidence, the system 80% believes that they are plural.

NARS treats pronoun resolution as a reasoning process, by which a pronoun is recognized as representing another noun or noun phrase. The reasoning is based on the conceptual relations provided by the words in the sentence, mainly the verbs. As the system receives more input proving its conclusion about pronouns, it is going to gain more confidence. The toy experiments above show that a general reasoning mechanism is enough to learn pronouns provided that pronouns' semantic content is grounded as relations among the concepts.

CHAPTER 4

WORDNET AND IMAGENET MEET NARS

In this chapter, NARS is combined with data from WordNet and ImageNet. In a set of toy experiments, WordNet plays a reliable source of definitions for concepts. The concepts are translated to NARS. Similarly, ImageNet is used as a reliable source of visuals for concepts. A convolutional neural network trained by ImageNet images, which is available online, is used to retrieve possible conceptual definitions of an input image. These definitions are used in a few experiments to show that NARS can simulate contextual dependency of concepts.

4.1 WordNet

WordNet is an on-line lexical reference system inspired by psycholinguistic theories of human lexical memory (Fellbaum, 2005; Miller, 1995). English nouns, verbs, adjectives, and adverbs are organized into synonym sets, each representing one underlying lexical concept. WordNet contains 155,287 words organized in 117,659 synsets (synonym sets) for a total of 206,941 word-sense pairs. Synsets actually correspond to abstract concepts. An example adjective synset is "*good, right, ripe.*"

Miller and Johnson-Laird (1976) proposed that research on lexicons should be called psycholexicology: The information a lexicon must contain in order for the phonological, syntactic, and lexical components to work together in the everyday production and comprehension of linguistic messages. Standard dictionaries neglect synchronic organization of lexical knowledge. Still, WordNet contains a net of words that are semantically related as shown. In the following section, the semantic relations in WordNet and their corresponding equivalences in NARS are given as examples.

4.1.1 From WordNet Semantics to NARS

Human languages use words to represents concepts despite numerous ambiguities: Sometimes distinct concepts have to share same word as their labels. For example, the word *bed* represents different concepts as nouns and verb given in (7).

(7) **Word**: bed

Nouns: a piece of furniture to sleep

a plot of ground in which plants are growing

a stratum of rock

…

Verbs: to furnish with a bed

to place plants in a bed of soil

to put to bed

…

If we assume that BED_i is the concept for "a piece of furniture to sleep" and BED_k stands for the action "to put to bed", the word *bed* represents the two distinct concepts in English. These two concepts are generally represented with distinct words in different languages: For example, Turkish uses the word *yatak* for BED_i while *yatırmak* stands for BED_k as shown in (8). This allows NARS to be adapted to different languages sharing similar concepts.

(8) ("bed" $\times$ $\mathrm{BED_i}$) $\rightarrow$ represent.

("bed" × BED_k) → represent.

("yatak" × BED_i) → represent.

("yatırmak" × BED_k) → represent.

Although natural language translation is out of the scope of the current study, NARS provides a broad mechanism which allows the representations of the same concepts with different labels. The synonym sets (synsets) in WordNet actually stand for semantic level logical form. Most synonym sets are connected to other synsets via a number of semantic relations based on the type as given in (9).

(9) **Nouns**

hypernyms: Y is a hypernym of X if every X is a (kind of) Y (*canine* is a hypernym of *dog*)

hyponyms: Y is a hyponym of X if every Y is a (kind of) X (*dog* is a hyponym of *canine*)

coordinate terms: Y is a coordinate term of X if X and Y share a hypernym (*wolf* is a coordinate term of *dog*, and *dog* is a coordinate term of *wolf*)

meronym: Y is a meronym of X if Y is a part of X (*window* is a meronym of *building*)

holonym: Y is a holonym of X if X is a part of Y (*building* is a holonym of *window*)

Verbs

hypernym: the verb Y is a hypernym of the verb X if the activity X is a (kind of) Y (*to perceive* is an hypernym of *to listen*) *entailment*: the verb Y is entailed by X if by doing X you must be doing Y (*to sleep* is entailed by to *snore*)

Adjectives

Antonym: The adjective Y has an opposite meaning to adjective X (*healthy* versus *sick* are antonyms).

Similar to: The adjective Y has a similar meaning to adjective X (*costly* and *expensive* are similar).

Pertainym (pertains to noun): An adjective that classifies its noun (*musical* classifies *instrument* in *musical instrument*).

In WordNet, both nouns and verbs are organized into hierarchies, defined by a hypernym or IS-A relationship. Each set of synonyms (synset) has a unique index and shares its properties, such as a gloss (or dictionary) definition for *dog* as shown in (10).

(10) dog, domestic dog, Canis familiaris

=> canine, canid

=> carnivore

=> placental, placental mammal, eutherian, eutherian mammal

=> mammal

=> vertebrate, craniate

=> chordate

=> animal, animate being, beast, brute, creature, fauna

=> ...

The adjectives are organized into clusters. Most adjective clusters contain two antonymous parts. They are connected via synonym relations. Some special symbols, called pointers, are used to represent the relations among the words across the synsets. Although there are many pointer types, only certain types of relations are permitted, for example, in (11).

(11) **The pointers for nouns**

!	Antonym
@	Hypernym
@i	Instance Hypernym
~	Hyponym
~I	Instance Hyponym
#m	Member holonym
#s	Substance holonym
#p	Part holonym
%m	Member meronym
%s	Substance meronym
%p	Part meronym

The pointers for verbs

!	Antonym
@	Hypernym
~	Hyponym
.	Entailment

The pointers for adjectives

! Antonym

& Similar to

\   Pertainym (pertains to noun)

Hypernyms are basic-level category consisting of hierarchical relations. For example, when the following representations retrieved from WordNet is input to NARS, it can easily output the simple reasoning shown in (12).

(12) **Input**

dog → animal

A dog is an animal

collies → dog

Collies are dogs

collies → animal?

Are collies animal?

Output

collies → animal. <1, 0.81>.

With 81% confidence, the system 100% believes that collies are animal.

Antonyms are for the words that are incompatible to be in the same class. They are in the form that "A is X" entails "A is not Y." For example, "if it is a cat, then it entails that it cannot be a dog" can be represented in NARS as in (13) where *$* represents a variable.

(13) $x → cat ⇒ (--, $x → dog).

-- *represents negation in NARS*

One concept covered for the adjectives in WordNet is gradable antonymy: Some adjectives, such as *fat* and *thin*, can be better expressed in a comparative way because they contain some degree of fuzziness, or subjectivity depending on personal experience. The comparative adjectival “fatter than” is represented in NARS as shown in (14). In this example, John's degree of membership to fuzzy concept "tall boy" depends on the extent to which he is taller than the other boys. The system’s belief is determined according to available evidence/counter evidence. Note that “/” represents extensional image while “◊” is a special symbol indicating the location a term in the product (see Wang, 2013a for details).

(14) **Input**

{John} → boy.

John is a boy.

{Tom} → (/ taller_than ◊ {John}).

John is taller than Tom.

{Tom} → (/ taller_than ◊ boy)?

Is Tom is taller than boys?

Output

{Tom} → (/ taller_than ◊ boy). <1, 0.45>

With 45% confidence, the system totally believes that Tom is taller than boys.

Input

{David} → boy.

David is a boy.

(--, {Tom} → (/ taller_than ◊, {David})).

Tom is not taller than David.

{Tom} → (/ taller_than ◊ boy)?

Is Tom is taller than boys?

<u>Output</u>

{Tom} → (/ taller_than ◊ boy). <0.5, 0.62>

With 62% confidence, the system 50% believes that Tom is taller than boys.

Moreover, NARS can easily represent the holonymy, which is the relationship between a term denoting the whole and a term denoting a part of, or a member of, the whole, and the meronymy, which denotes a constituent part of, or a member of something. These two semantic relations are opposite of each other. For example, a limb and a trunk are holonyms for a tree while a tree is a meronym for the two. NARS uses compound terms to represent such relationships. For example, *tree → (bark* × *trunk* × *limb* × *roots* × *leaves)* represents that a *tree* is composed of *bark*, *trunk*, *limbs*, *roots* and *leaves*.

WordNet has entailments to represent semantic relations among the concepts. For example, if it is given that "All ***A*** are ***B***. All ***B*** are ***C***," then the entailment will be "All **A** are **C.**" Entailment is an automatic reasoning process performed by NARS as shown in (15).

(15) **<u>Input</u>**

A → B.

A is B.

B → C.

B is C.

Output

A → C. <1, 0.81>.

With 81% confidence, the system is sure that A is C.

Finally, pertainymy and antonymy for adjectives should be discussed. An adjective that classifies its noun is called a pertainym. For example, *musical* classifies *instrument* as in *musical instrument*. In other words, *musical instrument* is an extensional intersection of the concept *instrument* with intensional property [*musical*] as shown in *musical_instrument* → *([musical] ∩ instrument)*. Antonynmy refers to concepts with opposite meaning. For example, *healthy* and *sick* have opposite meanings because a person cannot be both in a healthy and sick state as simulated in (16).

(16) **Input**

<<$x → [sick]> ⇒ (--, <$x → [healthy]>) >.

If the person X is sick, then X is not healthy.

<{Joseph} → [sick]>.

Joseph is sick.

Output

<{Joseph} → [healthy]>. <0, 0.81>.

It is very unlikely that Joseph is healthy.

As a part of the current study, an interface that searches WordNet database and converts the corresponding results into Narsese (NARS's language). In order to understand the process, WordNet's logical structure is briefly explained below.

4.1.2 Logical forms in WordNet

The concepts in WordNet are also represented as logical forms (see Clark et al., 2008; Fellbaum 1998; Fellbaum, 2005) developed by University of California, Information Sciences Institute. For example, the second most common sense of the noun *ambition* which is defined as "*ambition#n2: A strong drive for success,*" is represented as in (17). This example is re-written by omitting eventuality argument in the original WordNet notation. Eventuality argument is used to represent co-occurrence or temporal order of events. Since the eventuality notations in WordNet are not fully correct, and NARS represents eventuality not as an argument but different operators, they are omitted in the current study for simplicity. The other arguments starting with *x* represent the subject/actor of the concepts while the other arguments are object/direct/indirect objects of the concepts.

(17) $\text{ambition}(x1) \rightarrow \text{a}(x1)\ \&\ \text{strong}(x1)\ \&\ \text{drive}(x1)\ \&\ \text{for}(x1,x6)\ \&\ \text{success}(x6)$

In this representation, if something represented by *x1* is *ambition*, then it is a strong drive for another thing represented by *x6*. The corresponding sentence in NARS is given in (18).

(18) $$<(\$x1 \rightarrow \text{ambition})> \;\Rightarrow\; (\$x1 \rightarrow [\text{strong}] \wedge (\$x1 \times \#x6) \rightarrow \text{drive\_for} \wedge (\#x6 \rightarrow \text{success})$$

The example in (18) basically states that if something is *ambition*, that thing is *strong* and it has a *drive_for* relation with a certain type of *success*. Here is another example re-written from WordNet given in (19) for the most common sense of *politician* as "*politician#n1: a leader engages in civil administration*."

(19) politician(x1) → leader(x1) & engage_in(x1,x2) & civil(x2) & administration(x2).

The representation in (19) stands for that if someone is a politician, that person is a leader, and engaged in something, which is both civil and administration. It is represented in NARS as in (20).

(20) $<(\$x1 \rightarrow politician)> \;\Rightarrow\; (\$x1 \rightarrow leader \wedge$
$(\$x1 \times \#x2) \rightarrow engage\_in \wedge$
$\#x2 \rightarrow [civil] \wedge$
$\#x2 \rightarrow administration).$

In WordNet logical forms, the concepts are followed by a letter either *n*, *v*, *a*, or *r* with a number. The letter corresponds to *noun*, *verb*, *adjective*, and *adverb* categories while the number shows the most common sense. The logical form of a concept is generally a compound concept formed by sub-concepts. For the conversion task, the simple copula "→" is used to represent the *noun* sub-concepts. The *verbs* are represented as relations among concepts. Finally, *adjectives* and *adverbs* are used as properties of succeeding concepts. Therefore, the argument structure and parameters are very important for a translation task from WordNet to Narsese.

In the current study, a graphical user interface is prepared in Java as shown in Figure 4.1.

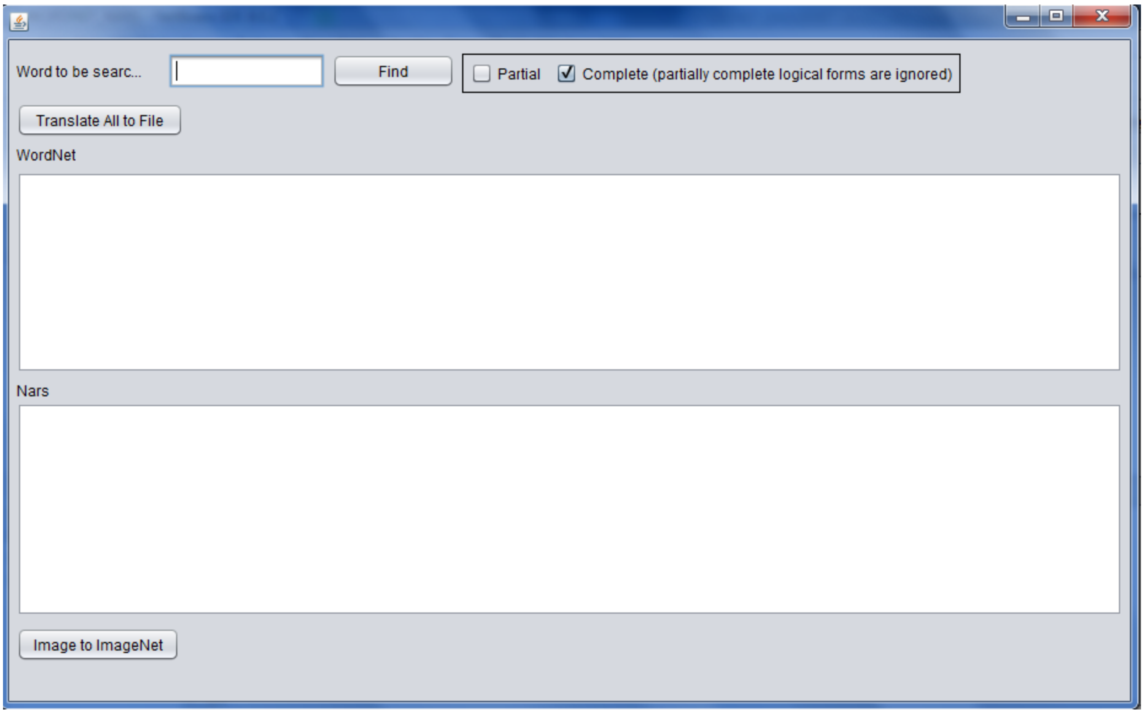

Figure 4.1: WordNet to NARS interface

This interface allows researchers to search words in WordNet, and convert the logical forms from WordNet to Narsese. Unfortunately, the WordNet logical forms are semi-automatically created and incomplete., The names of the variables shown by the arguments starting with *x* are especially important for the automatic translation task because these variables determine the order of sub-concepts and creating extensional images. The errors and incomplete logical forms in WordNet cause erroneous Narsese translation. However, the interface is still much useful for NARS researchers to retrieve a sketch of semantic representations from a trusted source, WordNet.

4.2 ImageNet

ImageNet is an ongoing research effort to provide researchers around the world an easily accessible image database. It is an image dataset organized according to the

WordNet hierarchy. Each meaningful concept in WordNet, possibly described by multiple words or word phrases, is called a "synonym set" or "synset". ImageNet aims to provide on average 1000 images to illustrate each synset. Images of each concept are quality-controlled and human-annotated. The following sections summarize statistics from image net and deep neural networks for image categorization via ImageNet.

4.2.1 ImageNet Statistics

ImageNet has a total number of 21,841 non-empty synsets which are equivalent to WordNet synsets. It contains 14,197,122 images. It contains only nouns because they are visually representable. On the other hand, verbs can be considered as relations among these nouns while adjectives contain specifying information about nouns, such as their colors and sizes. Adverbs modify adjectives or verbs or a whole premise. Table 4.1 shows statistics for noun superclasses.

Table 4.1: Statistics of high level categories.

High level category	**# synset**	**Avg # images per synset**	**Total # images**
amphibian	94	591	56K
animal	3822	732	2799K
appliance	51	1164	59K
bird	856	949	812K
covering	946	819	774K
device	2385	675	1610K
fabric	262	690	181K
fish	566	494	280K
flower	462	735	339K
food	1495	670	1001K
fruit	309	607	188K
fungus	303	453	137K
furniture	187	1043	195K
geological formation	151	838	127K
invertebrate	728	573	417K

Table 4.1: Statistics of high level categories (continued).

High level category	# synset	Avg # images per synset	Total # images
mammal	1138	821	934K
musical instrument	157	891	140K
plant	1666	600	999K
reptile	268	707	190K
sport	166	1207	200K
structure	1239	763	946K
tool	316	551	174K
tree	993	568	564K
utensil	86	912	78K
vegetable	176	764	135K
vehicle	481	778	374K
person	2035	468	952K

ImageNet provides an online interface to query their image database[ix], Table 4.2 shows the summary of results when the word *truck* is queried.

[ix] ImageNet is available online at http://image-net.org/index

Table 4.2: Some results for querying *truck*.

Images	Synset	Definition
	fire engine, fire truck	any of various large trucks that carry firemen and equipment to the site of a fire
	truck, motortruck	an automotive vehicle suitable for hauling.
	pickup, pickup truck	a light truck with an open body and low sides and a tailboard
	garbage truck, dustcart	a truck for collecting domestic refuse; "in Britain a garbage truck is called a dustcart".

It is also possible to see the results of a query in the form of a tree in ImageNet as shown in Figure 4.2.

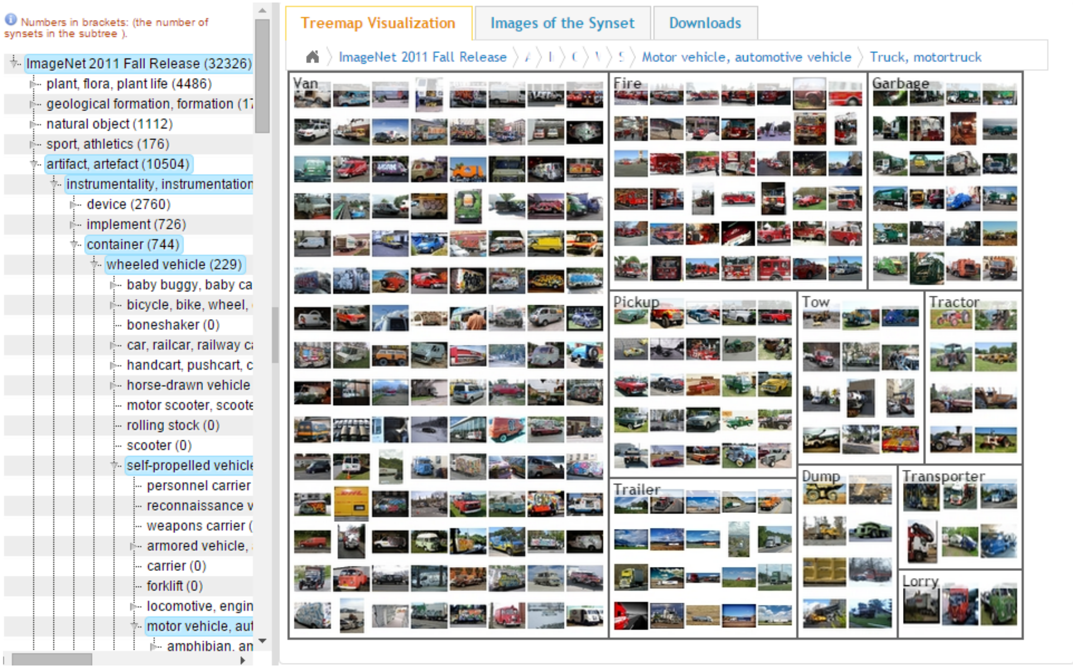

Figure 4.2: Tree view of a query

ImageNet also has 1,034,908 bounding box annotations. The bounding boxes are annotated and verified through Amazon Mechanical Turk, which is a crowdsourcing Internet marketplace that enables individuals and to coordinate the use of human intelligence to perform tasks that computers are currently unable to do. An example with some boxes showing foxes are given in Figure 4.3[x]. The boxes are used to train image processing tools more accurately.

[x] Retrieved from http://image-net.org/index

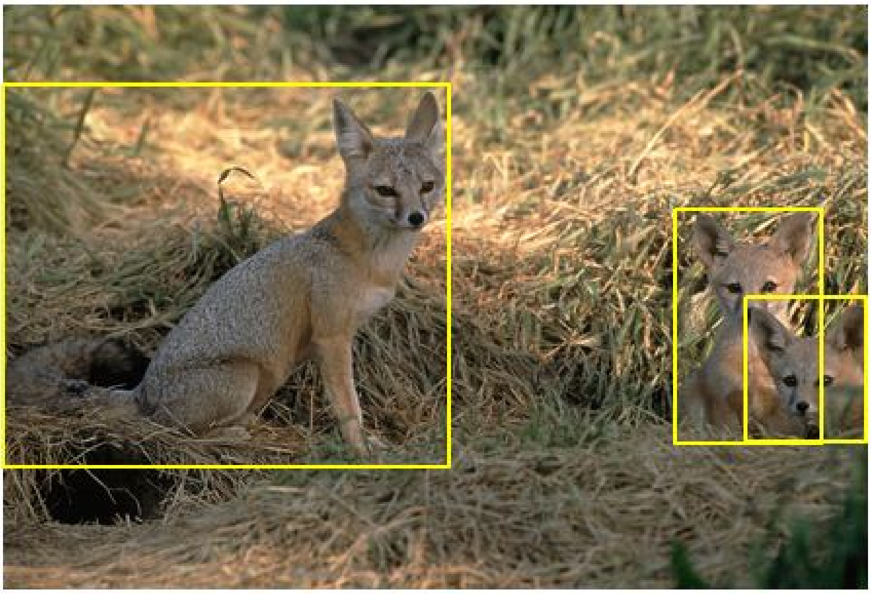

Figure 4.3: Boxes around kit foxes

Currently, ImageNet has bounding boxes for over 3000 popular synsets available. For each synset, there are on average 150 images with bounding boxes. ImageNet also provides scale-invariant feature transform values (SIFT) which are derived by an algorithm to detect and describe local features in images (Lowe, 2004). For any object in a training image, interesting points on the object, especially the ones in the boxes, can be extracted by this algorithm to provide a "feature description" of the object. This description, extracted from a training image, can then be used to identify the object when attempting to locate the object in a test image containing many other objects. It is important that the features extracted from the training image be detectable even under changes in image scale, noise and illumination. These points are usually on high-contrast regions of the image, such as object edges. SIFT key points of the objects are first extracted from a set of reference images. These features are saved to recognize an object in a new image by comparing each feature from the new image to the saved features. The method uses Euclidean distance between the features from the training images and the features from the test image. A full list of matches are used to determine good matches as subsets of key points that agree on the objects and its location, scale, orientation, and such (see Lowe, 1999 for details). ImageNet provides 1000 synsets with SIFT features and 1.2 million images with extracted SIFT features for image processing researchers.

One very successful implementation of image recognition using ImageNet is Caffe[xi]. Caffe is a deep learning framework developed by the Berkeley Vision and Learning Center (Jia et al., 2014). The framework is a licensed C++ library with Python and MATLAB bindings for training and deploying general purpose convolutional neural networks and other deep models.

In deep learning, each layer in deep architecture uses weights from preceding layer to have a higher level representation as exemplified in Figure 4.4.

[xi] Caffe is available online at http://demo.caffe.berkeleyvision.org/

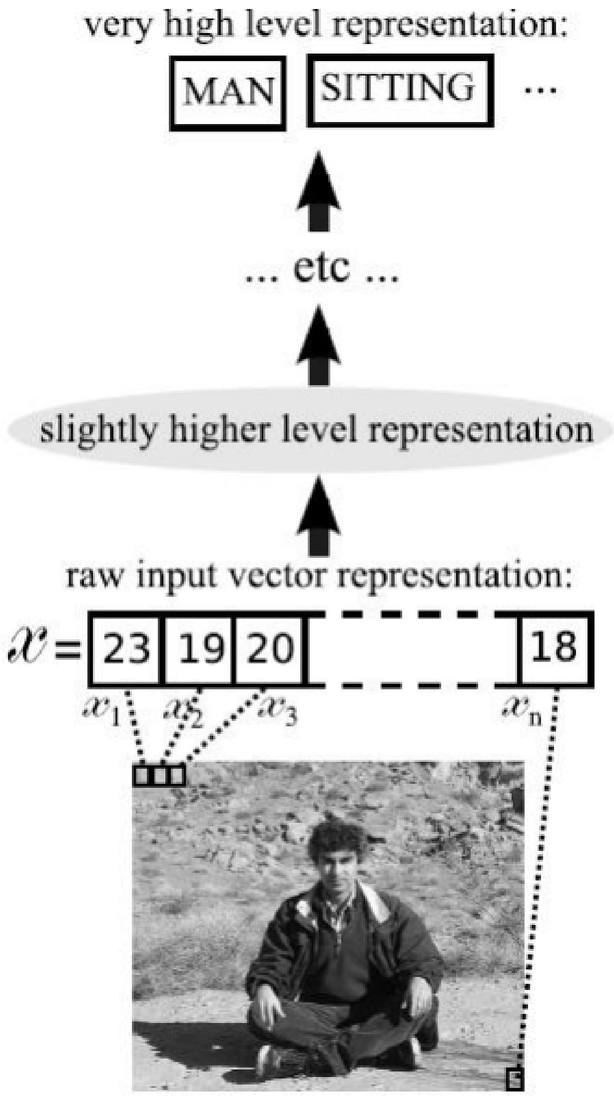

Figure 4.4: DNN example

The online demonstration of the framework is based on convolutional neural networks (for the details of the framework, algorithms, and improvements see Deng, Berg, Li & Fei-Fei, 2010; Fei-Fei, Ferrgus & Perona, 2006; Griffin, Holub & Perona, 2007; Lowe, 2004). A convolutional neural network (or CNN) is a type of feed-forward artificial neural network where the individual neurons are tiled in such a way that they respond to overlapping regions in the visual field. In other words, A CNN consists of multiple layers of small neuron collections which look at small portions of the input image as shown in Figure 4.5. It is first introduced by Kunihiko Fukushima in 1980 (Fukushima, 1980). It is inspired by biological processes: It consist of multiple layers of small neuron collections which look at small portions of the input image, called receptive fields. The results are combined so that they overlap to get a better representation of the original image (Ciresan, Meier, & Schmidhuber, 2012).

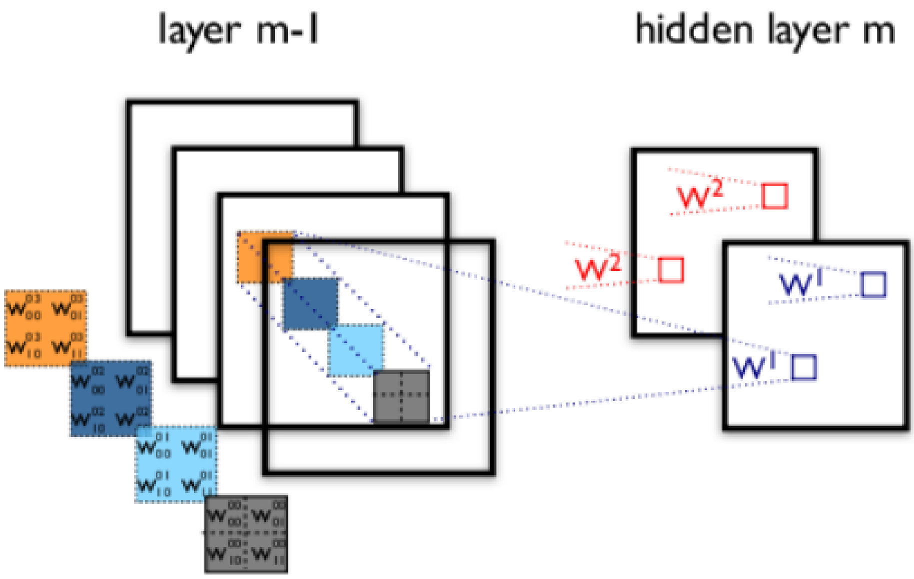

Figure 4.5: CNN example

An example representation for a CNN implementation is taken from Krizhevsky et al. (2004) as in. It is a CNN with 8 layers (5 convolutional versus 3 fully connected layers). The convolutional layers are called as pooling layers. They summarize the outputs of neighboring groups of neurons in smaller maps as shown in Figure 4.6 (Krizhevsky et al., 2004).

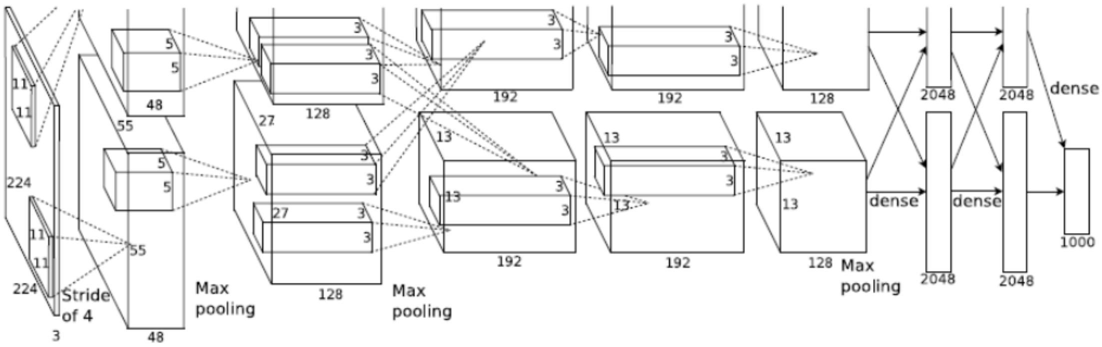

Figure 4.6: CNN max pooling

The resulting pool of the extracted features actually represents WordNet categories. When an image is input, its local features are extracted. For example, the features of different images of *dog* show more resemblance than those of *bird* as visualized in Figure 4.7 (Jia et al., 2014).

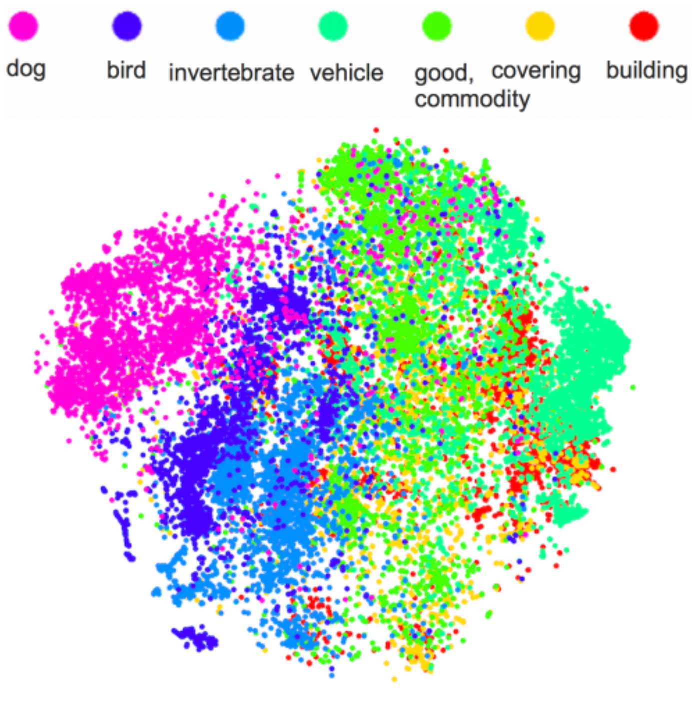

Figure 4.7: Features extracted for CNN visualized in a 2-dimensional space

The pre-trained CNN and a Support-Vector Machine are used to classify the input image given the previously extracted set of features from ImageNet to train the CNN as explained by Girshick, Donahue, Darrel and Malik (2014) and Guadarrama et al. (2014). The online demonstration of Caffe provides the top 5 possible category of an object from WordNet as exemplified in Figure 4.8 (Jia et al., 2014) in which an SVN is used in the step 4.

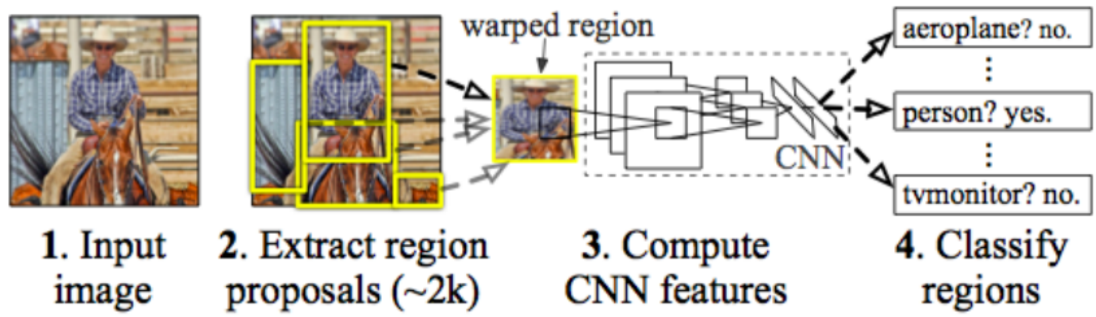

Figure 4.8: Caffe categorization example

4.2.2 WordNet and ImageNet Cooperate with NARS

The interface mentioned in Section 4.1 has an option to input an image to be queried in ImageNet. An example input for a toy cat's picture is shown below. Since the picture is not very representative of the cat class, the CNN's top guess is not a cat; and it also returns *dog* among the possible categories.

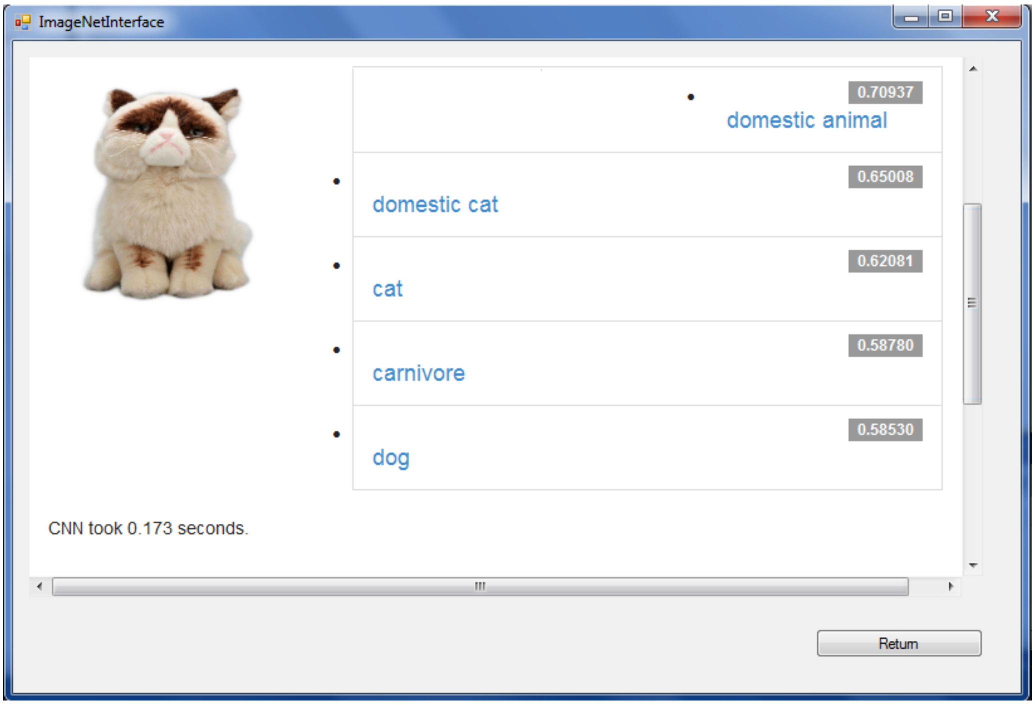

Figure 4.9: ImageNet to NARS interface

The interface returns all possible noun senses for the top 5 categories from WordNet; and translates them into Narsese whose simplified versions are given in (21). The top 5 categories are *domestic animal, domestic cat, cat, carnivore,* and *dog*.

(21) <$x1 → domestic_animal> ⇒ ($x1 → [various] ∧
$x1 → [domestic] ∧
$x1 → animal).

<$x1 → domestic_cata> ⇒ ($x1 → [various] ∧
($x1 × Felis)→ descend_from).

<$x1 → cat> ⇒ ($x1 → feline ∧
$x1 → mammal ∧
$x1 → [domestic] ∧
($x1 × fur) → has ∧

($x1 × claw) → has ∧

($x1 × roar) → unable).

<$x1 → carnivore> ⇒ ($x1 → animal ∧

($x1 × meat) → eat)

<$x1 → dog> ⇒ ($x1 → canis ∧

$x1 → [domestic] ∧

($x1 × Wolf) → descend_from).

The possible noun categories for the image have been retrieved from WordNet by the CNN using the ImageNet. When the Narsese form of the categories are input to NARS, the system is assumed to know these categories. The next step is to name this picture and assume that it can be any of these categories. As the system gains more information, it can consider that one of the concepts is more likely than the other. Moreover, the system can revise its decision depending on future experience as explained in the following section.

4.3 NARS and Contextual Dependency of Concepts and Terms

One area of research in human memory and learning is about contextual-variability: People distinguish two seemingly identical events or items that happened or were experienced at different times. Van Dantzig, Raffone and Hommel (2010) state that conceptual knowledge is acquired through recurrent experiences, by extracting regularities at different levels of granularity: Patterns of feature co-occurrence are categorized into objects, and patterns of concept co-occurrence are categorized into contexts. In other words, features determine the object's category while concept co-occurrence creates a context.

When humans interact with an object, the object automatically activates related knowledge about similar objects previously encountered. Therefore, conceptual knowledge is mostly experience-related (O'Connor, Cree, & McRae, 2009; Rogers & McClelland, 2004). This suggests that besides the perception-related aspects of the concepts (such as visual and auditory properties), context itself plays a role in the mental representation of the concepts. The interaction between the two provides a degree of variability in the concept's mental representation from person to person. Indeed, as social animals, humans share similar conceptual knowledge as a result of continuous communication, but it is dynamic and subjective. Many theories of memory address this issue by making a distinction between item and context information (Dennis & Humphreys, 2001; Shiffrin & Steyvers, 1997): Item information represents a stimulus, and context information represents the environment in which the stimulus was encountered.

Surprisingly, functional and contextual aspects of an object sometimes become more distinguishing than its perceptual properties: For example, it is possible to hear a professor asking a volunteer student to come to the *blackboard* although the boards in classrooms are white. Similarly, people still use the term "*dialing a phone*" although it is specific to archaic rotary phones. These examples show that *blackboard* has evolved to represent whatever is used in the classroom, hanging on the wall, facing students and used for teaching, and that *dialing a phone* still means calling someone on the phone. Although every English speaker today can understand the term *dialing a phone*, there was surely a time when a rotary phone, pushbutton phone, and even smartphones co-existed. The term evolved from one concept to another in context. Actually, the term now represents a different concept, which is *to compose a number* while the literal concept of dialing a rotary phone exists. These examples urge us to make a clear distinction between a *concept* and a *term* representing the concept, and to allow a transfer between the two.

NARS allows us to make a distinction, for example, between the term/word *cat* and the concept *CAT* as {“cat” $\times$ cat} $\rightarrow$ represent where *represent* is a relation between a sign (or symbol) “cat” and an internal concept cat. In Narsese, an English sentence *Cats eat fish* can be represented as {“cat” $\times$ “eat” $\times$ “fish”} $\rightarrow$ sentence (see Wang, 2013b for a broader discussion). The example in (22) is taken from Wang (2013b) to show that the system can make distinction between *concepts* and *words* while reasoning on them.

(22) {“cat” $\times$ cat} $\rightarrow$ represent <1, 0.9>

{“fish” $\times$ fish} $\rightarrow$ represent <1, 0.9>

{“food” $\times$ food} $\rightarrow$ represent <1, 0.9>

{{“cat” $\times$ “eat” $\times$ “fish”} $\times$ ((cat $\times$ fish) $\rightarrow$ food} $\rightarrow$ represent <1, 0.9>

NARS outputs the following conclusion through induction in (23).

(23) ({\$1 $\times$ \$2} $\rightarrow$ represent)) $\Rightarrow$

({{\$1 $\times$ “eat” $\times$ “fish”} $\times$ ((\$2 $\times$ fish) $\rightarrow$ food)} $\rightarrow$ represent) <1, 0.45>

(({\$1 $\times$ \$2} $\rightarrow$ represent) $\cap$ ({\$3 $\times$ \$4} $\rightarrow$ represent)) $\Rightarrow$

({{\$1 $\times$ “eat” $\times$ \$3} $\times$ ((\$2 $\times$ \$4) $\rightarrow$ food)} $\rightarrow$ represent) <1, 0.29>

The terms in (23) with the *\$* prefix are independent variables, and each indicates an arbitrary term. The above conclusions will contribute to the meaning of the phrase “eat fish” and the word “eat" respectively. The example in (24) shows that when the terms *dog* and *meat* input to the system, NARS produces its new understanding of the relations which are not seen by the system before.

(24) **<u>Input</u>**

{“dog” × dog} → represent <1, 0.9>

{“meat” × meat} → represent <1, 0.9>

<u>Output</u>

{{“dog” × “eat” × “fish”} × ((dog × fish) → food} → represent <1, 0.41>

{{“dog” × “eat” × “meat”} × ((dog × meat) → food} → represent <1, 0.26>

The example shows that the repeated patterns in the language will produce linguistic knowledge with higher frequency values, though all the knowledge remains revisable by new experience. By using *words* to represent *concepts,* NARS can differentiate syntactic relations from semantic ones because the former one is about the word and the latter one is about conceptual representation itself.

To simulate the contextual dependency of concepts and words, the top 5 results from the example (21) was input to NARS with various degrees. Here are the definitions for these concepts for the picture in Figure 4.9:

- *Domestic animal*: various domestic animals.
- *Domestic cat*: various animals that descend from Felis.
- *Cat*: a feline mammal that is domestic, unable to road and has fur and claws.
- *Carnivore*: an animal that eats meat.

- *Dog*: a domestic Canis that descends from a wolf.

When the concepts in (21) are active in NARS memory, a word, *kedi*, is chosen to represent all of them with a decreasing order of confidence values as shown in (25). These values are in the same order with the one returned by ImageNet. When a context can be defined as collection of concepts that co-occur in a certain window of time and space, it is plausible to activate all related concepts from WordNed and their equivalents in Narsese returned by ImageNet.

(25) ("kedi" × domestic_animal) → represent <1, 0.9>.

("kedi" × domestic_cat) → represent <1, 0.85>.

("kedi" × cat) → represent <1, 0.8>.

("kedi" × carnivore) → represent <1, 0.75>.

("kedi" × dog) → represent <1, 0.7>.

When NARS is asked what *kedi* represents, it gives the same order in (19). In other words, the system mostly believes that it is a domestic animal, and with a decreasing confidence, it can represent a domestic cat, a cat, a carnivore, and a dog. When (--, "kedi" → canis) is input to the system to tell *kedi* is not a canis, the system's frequency on that *kedi* represents *dog* decreases to 0.24. On the other hand, when ("kedi" × claw) → has which means *kedi has claws* is input to system, NARS' belief that ("kedi" × cat) → represent is restored to 0.8 frequency level. The frequency on that *kedi* represents *domestic_animal* or *domestic_cat* displays a decreasing pattern too. Therefore, depending on system experience, the ranking of what *kedi* represents will change to *cat, domestic_animal, domestic_cat, carnivore* and *dog*. *Kedi* actually means *cat* in Turkish.

CHAPTER 5

CONCLUSIONS

This thesis hypothesized that NARS can be used as a Cognitive Framework. In order to test this hypothesis, three research questions were put forward:

1. Can NARS simulate U-shaped learning in language learning?
2. Can NARS achieve pronoun resolution?
3. Can NARS display a change of belief which corresponds to contextual dependency of concepts?

Learning English passive voice, resolving pronouns, and evolving context-dependent concepts have been successfully simulated by NARS. An interface was developed to transfer conceptual definition from a reliable source, WordNet, to NARS. Moreover, ImageNet and an online convolutional neural network were used as a separate module which provides visual information to NARS. This study is about discovering NARS' capacity on natural language processing. Since there is no similar study, it is not possible to compare the current results with any other literature.

Formal models can be used descriptively (to describe human cognition) or normatively (artificial production of human rational) in Cognitive Science. NARS is designed as a normative model, the system shows some behaviors similar to those happens in human thinking, which are usually explained in terms of heuristics. For example, the system imitates human behavior in truth-value evaluation, membership estimation, and incremental learning. The system realizes a "relative rationality", that is, the solutions are the *best* the system can get given current history and limitations on resources (Wang, 2011).

Humans are better at reasoning in domains that they are familiar with (Johnson-Laird et al., 1972; Griggs & Cox, 1982). Therefore, individual history is crucial and influenced by limitations on computational resources as heavily emphasized and employed by NARS. Obviously, NARS is "less idealized" than traditional normative models, such as FOPL, Bayesian approach, and so on, because it assumes stronger knowledge and resource constraints. It represents a "relative rationality" (Wang, 2011), which is similar to, though not equal to, Simon's "bounded rationality" (Simon, 1957).

Compared to other normative models, the behavior of NARS is more similar to those of people; therefore, we have reason to believe that its assumptions are more "realistic" that is, more similar to the regularities behind the human cognitive mechanism. This result can be explained by the observation that the human mind evolved, and still works, in an environment where knowledge and resources are usually insufficient to solve its problems. Indeed, NARS is not proposed as a replacement of other models. Instead, it underlines that its approach is more appropriate when the system must be open to novel situations and problems, and to make reasonable responses in real time

NARS is neutral about innateness hypothesis (Chomsky, 1972; Chomsky, 1988; Chomsky, 2012). It allows the system to have innate domain-specific knowledge, though such knowledge can be learned and revised, given proper experience. Nevertheless, it is closer to Cognitive Linguistics, which has the following assumptions (Croft & Cruse, 2004):

- Language is not an autonomous cognitive faculty,

- Grammar is conceptualization,
- Knowledge of language emerges from language use.

Furthermore, NARS' approaches to NLP can be summarized as follows: Language-specific knowledge is learned from experience and prone to change according to the system's experience. The learning process can follow different rules: *deduction, induction, abduction, analogy, revision*, and so on. A natural language is treated as a conceptual system that changes over time. New words, phrases, and sentences are introduced from time to time, and the existing ones may change their meaning gradually as shown in Chapter 4. Syntax, semantics, and pragmatics are not processed separately in the system. The processing of natural languages is unified with other cognitive processes, such as reasoning, learning, and categorizing.

NARS proposes that some NLP task can be done without assuming a built-in linguistic competence. Indeed, reducing all NLP tasks to the current version of NARS is not possible. Yet, this thesis emphasizes that considering NLP within general cognitive capacities *is* possible because NARS displays a degree of success in some NLP tasks without any language-specific module.

5.1 Future Directions

The interaction between NARS and ImageNet in this study is one-way. It is a promising direction that some results from NARS can be used to tune the parameters of the convolutional neural network. Another future direction is to add to NARS linguistic functions, such as simple text concatenation. Although NARS is not a descriptive model of the human mind, and has no NLP-specific module, it can still be modified to simulate human-level linguistic skills. For example, a speech-recognition and production module, or perceptual information sources can be added to NARS to allow an artificial general intelligence system to evolve some skill-specific sub-modules.

BIBLIOGRAPHY

Abdel-Fattah, A., Besold, T., Gust, H., Krumnack, U., Schmidt, M., Kuhnberger, K. & Wang, P. (2012). Rationality-guided AGI as cognitive systems, In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, pp 1242–1247.

Altmann, G., & Steedman, M., (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191–238.

Arnold, J. E. (1998). Reference form and discourse patterns. *Ph.D. Dissertation*, Stanford University.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The immediate use of gender information: Eyetracking evidence of the time-course of pronoun resolution. *Cognition, 76*, B13–B26.

Bower, G. H., & Hilgard, E. R. (1981). *Theories of learning*. Englewood Cliffs, NJ: Prentice-Hall.

Bowerman, M. (1982). Starting to talk worse: Clues to language development from children's late speech errors. In S. Strauss & R. Stavy (Eds.), *U-shaped behavioral growth, Developmental Psychology Series* (pp. 101–145). New York: Academic Press.

Carey, S. (1982).Face perception: Anomalies of development. In S. Strauss & R. Stavy (Eds.), *U-shaped behavioral growth, Developmental Psychology Series* (pp. 169–190). New York: Academic Press.

Carlucci, L., & Case, J. (2013). On the Necessity of U-Shaped Learning. *Topics in Cognitive Science, 5*, 56–88

Chomsky, N. (1980). *Rules and Representations*. Oxford: Basil Blackwell.

Ciresan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3642–3649). New York, NY: Institute of Electrical and Electronics Engineers(IEEE).

Clahsen, H., & Felser, C., (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3–42.

Clark, A., & Lappin, S. (2010). Computational learning theory and language acquisition. In R. Kempson, N. Asher & T. Fernando (Eds.), *Philosophy of linguistics* (pp. 445–475). New York: Elsevier.

Clark, P., Murray, W., Thompson, J., Harrison, P., Hobbs, J., & Fellbaum, C. (2008). Augmenting WordNet for deep understanding of text. In J. Bos & R. Delmonte (Eds.), *Semantics in text processing* (Proceedings of STEP 2008).

Cowie, F. (1999). *What's within? Nativism reconsidered*. New York: Oxford University Press.

Croft, W., & Cruse, D.A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

Deng, J., Berg, A., Li, K., & Fei-Fei, L. (2010). What does classifying more than 10,000 image categories tell us? In K. Daniilidis, P. Maragos, N. Paragios (Eds.), *ECCV 2010, Part V, LNCS 6315* (pp.71–84). Berlin Heidelberg: Springer-Verlag.

Dennis, S., & Humphries, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*(2), 452–478.

Fei-Fei, L., Fergus, R., & Perona, P. (2006). One-shot learning of object categories. *PAMI, 28*, 594–611.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Fellbaum, C. (2005). WordNet and wordnets. In K. Brown et al. (Eds.), *Encyclopedia of language and linguistics,* second edition (pp. 665–670). Oxford: Elsevier.

Franklin, S., Madl, T., D'Mello, S., & Snaider, J. (2014). LIDA: A systems-level architecture for cognition, emotion, and learning. *IEEE Transactions on Autonomous Mental Development*, *6*(1), 19-41.

Fukishima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics, 36* (4), 193–202.

Garvey, C., Caramazza, A., & Yates, J. (1976). Factors influencing assignment of pronoun antecedents. *Cognition, 3*, 227–243.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj II (Ed.), *Language development, Volume 2: Language, thought and culture* (pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum.

Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world*. New York: Oxford University Press.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of CVPR*.

Gold, E. (1967). Language identification in the limit. *Information and Control, 10*, 447–474.

Griffin, G., Holub, A., & Perona, P. (2007). Caltech-256 object category dataset. *TechnicalReport 7694*, Caltech.

Guadarrama, S., Rodner, E., Saenko, K., Zhang, N., Farrell, R., Donahue, J., & Darrell, T. (2014). Open-vocabulary object retrieval. *Proceedings of the Robotics: Science and Systems Conference*. Berkeley, California.

Heinz, J. (2010). Computational theories of learning and developmental psycholinguistics. In J. Lidz & J. Pater (Eds.), *The Cambridge handbook of developmental linguistics*. Cambridge, England: Cambridg University Press.

Hirsh-Pasek, K., & Golinkoff, R. M. (Eds.) (2006). *Action meets word: How children learn verbs*. Oxford: Oxford University Press.

Hornstein, D. and Lightfoot, N. (1981). *Explanation in linguistics*. Longman, London.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., & Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. *Proceedings of the ACM International Conference on Multimedia.*

Johnson, K. (2004). Gold's theorem and cognitive science. *Philosophy of Science, 71*, 571–592

Laakso, A., & Smith, L. B. (2007). Pronouns and verbs in adult speech to children: A corpus analysis. *Journal of Child Language, 34*, 725–763.

Lewis, R.L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375– 419

Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Proceedings of the International Conference on Computer Vision 2* (pp. 1150–1157).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91–110.

Marcus, G., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the society for research in child development, 57*(4), 1–182: University of Chicago Press.

McCloskey, J. (2006). Resumption. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell companion to syntax*, (pp. 94–117). Oxford: Blackwell.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*, 440–466.

McMillan, C. T., Clark, R., Gunawardena, D., Ryant, N., & Grossman, M. (2012). fMRI evidence for strategic decision-making during resolution of pronoun reference. *Neuropsychologia, 50*, 674–687.

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM, 38*(11), 39–41.

Miller, G. A., & Johnson-Laird. P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

Mohan, S., Mininger, A., Kirk, J., & Laird, J. (2012). Acquiring grounded representations of words with situated interactive instruction. *Advances in Cognitive Systems*.

O'Connor, C. M., Cree, G. S., & McRae, K. (2009). Conceptual hierarchies in a flat attractor network: Dynamics of learning and computations. *Cognitive Science, 33*, 665–708.

Papadopoulou, D. Peristeri, E., Plemmenou, L., Marinis, T., & Tsimpli, I.M. (2015). Pronoun ambiguity resolution: Evidence from monolingual adults and children. *Lingua, 155*, 98–120.

Papadopoulou, D., & Tsimpli, I. (2005). Morphological cues in children's processing of ambiguous sentences: A study of subject/object ambiguities in Greek. In A. Brugos, M. R. Clark-Cetton, and S. Ha (Eds.), *Proceedings of the BUCLD 29*, (pp. 471–481). Somerville, MA: Cascadilla Press.

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2010). A tutorial introduction to Bayesian models of cognitive development. *Cognition, 120*, 302–321.

Pickering, M. J., & Majid, A. (2007). What are implicit causality and consequentiality? *Language and Cognitive Processes, 22*(5), 780–788.

Pinker, S. (1979). Formal models of language learning. *Cognition, 7*, 217–283.

Pullum, G. K., & Scholz, B. C. (2003). Linguistic models. In M. T. Banich and M. Mack (Eds.), *Mind, brain, and language* (pp. 113-141). Mahwah, NJ: Lawrence Erlbaum Associates.

Rapaport, W. J, (2013). Meinongian semantics and artificial intelligence *Humana.Mente Journal of Philosophical Studies*, *25*, 25-52.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Shapiro, C. T. (2000). An introduction to SNePS 3. In *Proceedings of International Conference on Conceptual Structures* (pp. 510-524). Darmstadt, Germany

Shiffrin, R.M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*, 145-166.

Sichel, I. (2014). Resumptive pronouns and competition. *Linguistic Inquiry, 45* (4), 655–693.

Siddiqi, D. (2009). *Syntax within words: Economy, allomorphy, and argument selection in distributed morphology.* Amsterdam, the Netherlands: John Benjamins.

Simon, H. (1957). *Models of man: Social and rational*. New York: John Wiley.

Smith, N. (1999). *Chomsky: Ideas and ideals*. Cambridge: Cambridge University Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* MIT Press.

Thrun, S. (1997). Lifelong learning algorithms. In S. Thrun & L. Y. Pratt (Eds.), *Learning to learn* (pp. 181-209). Boston, MA: Kluwer Academic Publishers.

Tomasello, M. (2000). The item based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4*, 156–163.

Trueswell, J.C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language, 35*, 566–585.,

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293–315.

Valian, V. (2009). Innateness and learnability. In E. Bavin (Ed.), *Handbook of child language* (pp. 15–34). Cambridge: Cambridge University Press.

Van Dantzig, S., Raffone, A., & Hommel, B. (2011). Acquiring contextualized concepts: A connectionist approach. *Cognitive Science, 35*, 1162–1189.

Wang, P. (2001). Wason's cards: what is wrong? In *Proceedings of the Third International Conference on Cognitive Science*, pp 371-375. Hefei: Press of USTC.

Wang, P. (2006). *Rigid flexibility: The logic of intelligence*. Dordrecht: Springer.

Wang, P. (2011). The assumption on knowledge and resources in models of rationality. *International Journal of Machine Consciousness (IJMC)*, *3*, 193–218.

Wang, P. (2013a). *Non-axiomatic logic: A model of intelligent reasoning*. Singapore: World Scientific.

Wang, P. (2013b). Natural language processing by reasoning and learning. *Lecture Notes in Computer Science, 7999*, 160–169.

Wang, P. & Goertzel B., (2007). Aspects of artificial general intelligence. In B. Goertzel & P. Wang (Eds.), *Advance of artificial general intelligence* (pp 1-16). Amsterdam: IOS Press.

Wang, P., & Hofstadter, D. (2006). A logic of categorization. *Journal of Experimental & Theoretical Artificial Intelligence*, 18(2), 193-213.

Wason, P., & Johnson-Laird, P. (1972). *Psychology of reasoning: Structure and content*. Harvard University Press.