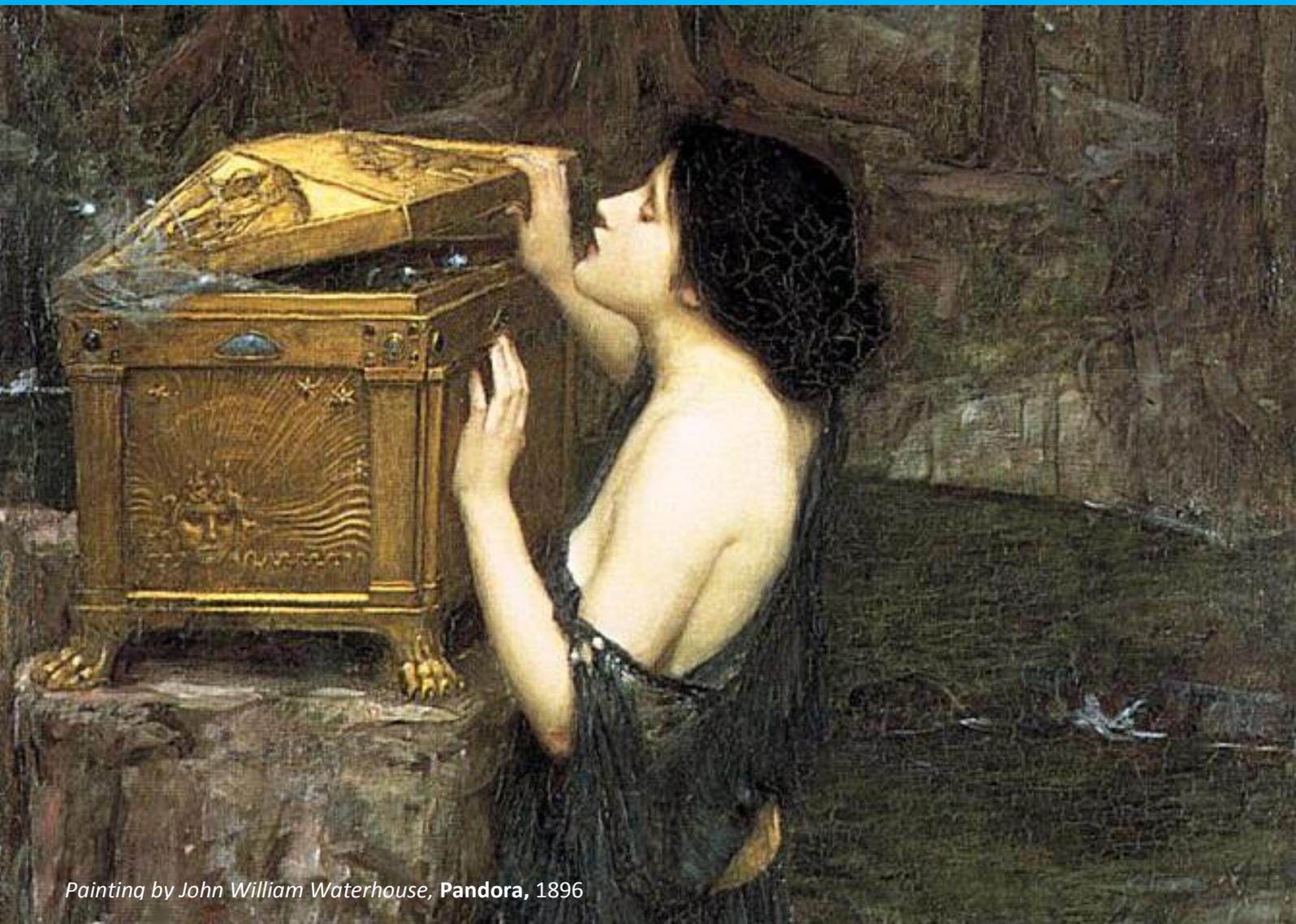


The Natural Language Understanding Problem

A literature review of recent work using an embodied, usage-based, cognitive linguistic approach



Painting by John William Waterhouse, **Pandora**, 1896

Steven Bos, June 2012

Milestone 1 for MSc Project TU Delft

Supervisor: Dr. Ir. Pascal Wiggers

Contents

About this research assignment	5
1 Introduction	6
2 Embodiment and the Neural Theory of Language	8
3 A new philosophical framework for NLU	10
3.1 The current definitions of meaning and understanding	10
3.2 The relation between meaning and understanding	11
3.3 The Embodied definitions of meaning and understanding	13
3.3.1 Defining Meaning	13
3.3.2 Defining Understanding	15
3.4 Epistemology: Study of knowledge	17
3.4.1 Defining Experience	17
3.4.2 Defining Knowledge	20
3.4.3 Defining Logic	22
3.4.4 Defining Truth	24
3.4.5 Defining Belief	25
4 The human NLU system: Early developments	28
4.1 The genesis of cognition	28
4.2 Early cognitive development	29
4.3 Development of language skills	30
4.3.1 Passive language learning	30
4.3.2 Active language learning	31
5 The human NLU system: Components in the brain	33
5.1 Anatomy of the brain	34
5.1.1 The brain at large	34
5.1.2 Cerebral cortex	35
5.1.3 Neurons	35
5.1.4 Training neurons	36

5.1.5	The central nervous system and information flow	37
5.2	Functional structures in the brain	39
5.2.1	Broca's area	40
5.2.2	Wernicke's area	41
5.3	Functionality of the brain: Memory	43
5.3.1	Sensorial memory	44
5.3.2	Short term memory	45
5.3.3	Long term memory	46
5.3.4	Working memory	46
6	The human NLU system: Theories on enabling NLU.....	49
6.1	Embodied theory of language structure	50
6.2	Embodied theory of language aquisition.....	53
6.2.1	Language learning: a usage-based theory.....	54
6.2.2	Grammar learning.....	55
6.2.3	Infant teaching methods of language	57
6.3	Embodied theory of language use.....	58
6.3.1	Modelling motivation and personality.....	58
6.3.2	Modelling human inference	61
7	The artificial NLU system.....	64
7.1	Developing (Embodied) AI systems	65
7.1.1	A brief history of problems with developing AI systems	65
7.1.2	Principles of engineering Embodied AI systems.....	67
7.1.3	Key directions in engineering Embodied AI systems	68
7.2	Developing (Embodied) NLU systems.....	70
7.2.1	General NLU Architecture	70
7.2.2	General Components	71
7.2.3	Problems of traditional NLU systems.....	72
7.2.4	Computational approaches to Language Aquisition.....	73
7.2.5	Computational approaches to Language Structure	74

7.2.6	Computational approaches to Language Usage	75
7.3	Benchmarking and validating artificial NLU systems.....	78
7.4	Existing NLU Systems.....	79
7.4.1	ITALK	80
7.4.2	Watson	81
7.4.3	XPERO.....	83
7.4.4	Protothinker	84
7.4.5	NTL.....	85
8	Conclusion and final remarks	88
	References.....	90
	Appendix 1: Normal language development checklist.....	100
	Appendix 2: The three general learning theories	101
	Appendix 3: Conversations with the Loebner 2011 chatbot	103
	Appendix 4: Encoding of typical communication episode	105
	Appendix 5: Principles of Embodiment	106
	Appendix 6: Hierarchical model of embodied knowledge	107

About this research assignment

The problem of natural language understanding (NLU) has been studied by tens of generations of philosophers, scientists and engineers. Yet it remains illusive to work on, having similarities to opening Pandora's Box (frontpage illustration). No single psycholinguistic publication exists that covers all the phenomena that occur, sometimes uniquely in human language. Examples are the multitude of languages and how an infant develops an understanding of it in the brain. Although some are promising efforts (Harley, 2008), none are available from an Embodied view. This research assignment will not be so bold and try to fill that gap. Rather I engage in a specific approach to the NLU problem using Embodiment and cognitive linguistic theories and both empiricism and rationalism doctrines. I will show that these topics jointly unfold nicely in a theory called the Neural Theory of Language (NTL) (Feldman J., 2006).

Next to the human side, I also review the artificial side of the NLU problem. The main research question of this literature study is "what is needed to build a robot or computer program that is able to process human language such that human-like language understanding can be achieved." To answer that question I will not be limited by just providing an overview, but also provide many novel ideas, viewpoints and interpretations based on that literature. This report is meant to be a thought provoking starting point and by all means not a definitive guide – it is certainly not for me. An unbiased, open mind is required to read it.

In my quest to engineer NLU machines four main questions can be derived:

- What does "understanding" mean and paradoxically, what does "mean" or "meaning" mean? (a working definition including a theory of knowledge - discussed in chapter 3)
- What biological components are involved in the human NLU system? (chapter 4 and 5)
- How does the human NLU system work (and what models this best)? (chapter 6)
- How can human NLU be implemented artificially? (how to bridge from the biological domain to the computational domain and what implementations are readily available – discussed in chapter 7)

By providing and discussing the literature I try to answer these questions. It will no doubt give rise to many more questions going beyond the famous NBIC (Nanotechnology, Biotechnology, Information technology and Cognitive science) fields. To start somewhere many controversial theories are introduced, assumptions are made and even presumptions. I hope they prove to be a solid (and flexible) enough foundation to answer these new questions.

The models, conclusions and other findings of this research assignment will form the basis of a research chapter, part of my MSc Thesis. Reading this research overview will be a rollercoaster *experience*, which by the end I hope enables you to *understand meaning* on all three proposed levels, *believe* (for the full 100%) in its *truth* and find it *logic*. To do so, definitions and examples will be given on the italic words including their intimate relation. Please reread this last paragraph after completing this report and hopefully the italic words will have a clearer definition.

After reading this report the reader should realise that this report contains interpretation errors on both the author as well the reader account – transforming thoughts into language and vice versa is an error prone procedure. Suggestions, request for clarification and discussion are very much welcome at the address beneath.

Last but not least this report took well over a year, but did complete because of the sheer enthusiasm of my environment; many thanks dr. Pascal Wiggers, Jessica, Friends and Family – I owe you one.

Steven Bos, june 2012

"I know that you believe you understand what you think I said, but I'm not sure you realize that what you heard is not what I meant"

Robert McCloskey

1 Introduction

In which the need for researching the natural language understanding problem is motivated. The umbrella research field "natural language processing" is introduced, mentioning the many disciplines working on the problem and the slow progress.

A vast collection of publications exist on language related topics in robots or computer programs such as emotions in speech (Chiriacescu, 2009), topic extraction in a conversation (Wiggers, 2008), grammar acquisition (Gavaldà & Waibel, 1998). Relatively few exist on the topic of artificial language understanding, a research field that eventually integrates all of the above topics. This chapter is divided in two parts; the first part contains general background on the problem while the second discusses reasons why anyone would take interest in it.

Natural language understanding (NLU)¹ is a subfield of natural language processing (NLP)² which in turn is a subfield of psychology or artificial intelligence (AI) depending on the model (human or computer). The field of NLP envelops at least the following topics²:

- Signal processing
- Syntactic analysis
- Semantic analysis
- Pragmatics

The careful reader notices that NLU is not part of this list. The field of NLU is an integrating field that takes results from all of the listed subfields. Unfortunately, it is not *simply* taking the results from those subfields since processing language (to understand it) does not work cascade style, but rather iterative, parallel and seemingly without a fixed flow³. According to James Allen the following types of knowledge are relevant to NLU (Allen J., 1995) (from: ²)

- Phonetic and phonological knowledge: how words are related to sounds.
- Morphological knowledge: how words are built from more primitive morphemes (e. g., how "friendly" comes from "friend." Morphology deals with the different inflections of a word, the forms it can take: a noun can be singular or plural, a verb can have different tenses, and so forth. Morphemes include "run," "laugh," "non-," "-s," and "es." Programs can be written to process tokens of words or even the more basic level of morphemes.
- Syntactic knowledge: how sequences of words form correct sentences. Knowledge of the rules of grammar.
- Semantic knowledge: how words have "meaning"; how words have reference (denotation) and associated concepts (connotations)

¹ the term "(natural) language comprehension" is also used in literature

² the term "computational linguistics" or "computational semantics" is also used in literature. A compact primer is available online by the The Mind Project (2006)

http://www.mind.ilstu.edu/curriculum/protothinker/natural_language_processing.php

³ seemingly, because some scientists consider the notion of no *actual* free will or flow, as discussed in chapter 4

- Pragmatic knowledge: "how sentences are used in different situations and how use affects the interpretation of the sentence," this involves the intentions and context of the conversation.
- Discourse knowledge: how preceding sentences determine the meaning of a sentence, such as in the case of the referent of a pronoun.
- World knowledge: general knowledge about for example, other users' beliefs and goals in a conversation.

This shows that the field of NLU touches many topics and interestingly can be approached from many different angles such as linguistics, (evolutionary) biology, psychology, neurology, philosophy, pedagogy, computer science, etc - making it a true interdisciplinary field.

Artificial NLU, together with its somewhat older brother AI, is a newborn field with only about 40 years of active research. NLU literature is continuously enriched by all the mentioned disciplines. Recently, due to medical imaging advancements many neuroscientists are publishing important results, such as correlations between behavior and fMRI results (Raichle, 2009). Many famous journals and publishers such as Nature and the Journal of Neuroscience seem to give special attention to these new results – and with all reasons, the brain and its workings is a hot topic in the public domain. Still, progress is slow and correlations seem often weak since the brain is highly *plastic* – a term for adaptive neural rewiring. Because of the slow progress many left the field. The work by (Carruthers & Boucher, 1998) starts with one big plea to scientists from different disciplines to return to the NLU field. Their historical overview of similarities and differences of the many views on the NLU problem is excellent. Worth mentioning is the *zeitgeist* during the '70 and '80 where work in NLU was based on the assumption of a language faculty in the brain, on formal models of grammar and semantics with notable contributions from (Chomsky, 1959), (Pinker S., 1994) and (Montague, 1974). Starting from the '90 more pragmatic and pattern recognition approaches were preferred. The work of psychologist (Skinner, 1957) and Ludwig Wittgenstein (Wittgenstein, 1922), (1953) were revisited and (Lakoff & Johnson, 1980) with his metaphor theory went head to head with his former mentor Chomsky, claiming that there is no autonomous language faculty in the brain. Since the '90 many technological advancement and insights are gained, amongst them NLU based on the Embodiment assumption, such that work in this direction is justified. Hopefully, this course will boost the field, attract many investors, such that the famous "AI Winter" will truly end.

The second question of this introduction relates to motivation. Not much words are required to emphasize the huge potential of NLU and the exciting times that lie ahead in this field. As mentioned before, the artificial side it is a relatively new field with much inventions and honor achievable and is highly interdisciplinary, enabling collaboration with other fields. But most importantly the field of NLU has implications for the human society at large. Instant killer applications would be autonomous translation, dictation and a computer that actually does understand you when you order a ticket by phone. Furthermore, systems that can understand language can communicate with humans in utmost natural fashion. Such systems would be in favor to for example Brain Computer Interfacing systems simply because it provides humans the privacy to expose their thoughts. Another application would be that of domestic appliances, a robot that can offer help with cleaning and cooking tasks on a whole new level, to name a few. Hospitals and senior accommodations would also greatly benefit from language understanding robots, they can assist with operations or provide companionship. Lastly, through understanding of artificial models (artificial NLU), insight is gained which could in principle validate and extend current psychological models (human NLU) treating mental traumas or autism.

“The mind is nothing but a plaything of the body”

Friedrich Wilhelm Nietzsche (paraphrase)

2 Embodiment and the Neural Theory of Language

In which Jerome Feldman’s Neural Theory of Language (NTL) is introduced, its core assumptions such as Embodiment and some implications. The NTL model discussed throughout this report by Chang, Bergen and Mok is also introduced.

Jerome Feldman in his publication (Feldman J. , 2006) strives to create a new research field which coordinates and stimulates interdisciplinary research on cognition never seen before. He coined the name Unified Cognitive Science for this umbrella field. At the heart of this field are theories that are open to all kinds of data and models from different fields with regards to cognition. One such theory is his own ambitious work, baptised as The Neural Theory of Language (NTL).

The NTL framework is founded on the assumption of “Embodiment”. This controversial claim is gaining scientific momentum (see (Borreggine & Kaschak, 2006), (Glenberg & Kaschak, 2002) (Clark, 1997)). In a nutshell it predicts that experience, action, language, perception can all, in principle, be linked to the properties and limitations of the body. *Meaning*, as it will be properly defined later, is therefore Embodied. It also predicts that the body is always *situated*, stating that all input gathered by our internal and external senses are always “in a situation”, so always in some social and physical context. According to Lakoff and Johnson’s (Lakoff & Johnson, 1980), (Lakoff, 1987) and also Feldman and Narayanan (Feldman & Narayanan, 2004) using metaphor theory, even the most complex thoughts and ideas can be simplified to the point it can be represented as (a pattern of) direct bodily experience. Metaphors are generic bridges that can link concepts so that these time consuming bodily experiences need not to be displayed to communicate it (explained more elaborately in chapter 3). An example is the metaphor COUNTRY = HUMAN, like in “*Holland needs a financial injection to stay healthy*”. Although a country is not a living organism with bodily needs and problems like illness, mapping it to the body which is understood well and allowing pointing to bodily aspects makes it easier to discuss. One can imagine a country like a gigantic body needing money instead of food to stay healthy. An important implication of the Embodiment assumption is that the knowledge (ie. their lexicons, behavior) of no two people are identical because no two bodies are identical to the biological level (not even twins, as discussed in chapter 5). Another implication is that the body and the way humans learn how to control it and function with it in their cultural environment (eg. language environment) has influence on their efficiency in doing tasks. An easy argument would be “stronger muscles result in easier lifting”, but it goes even deeper with language and mental tasks. The Embodiment thesis supports the weak version of the famous Sapir-Whorf hypothesis (language influences thoughts and non-linguistic behavior). For details on this matter see Kay and Kempton (Kay & Kempton, 1984) and results from Levinson (Levinson, 2003).

NTL is a generic framework that adapts the key principles of cognitive linguistics such as the Cognitive Commitment (see (Evans, Bergen, & Zinken, 2006) for a brief overview of the cognitive linguistic field) that pursues after human cognition theories that are consistent with all other

cognitive and brain sciences. NTL is also open to both empiricism and rationalism doctrines, predicting that knowledge is gained through senses, but at the same time knowledge is gained through inference and mental simulation (such as dreaming).

The questions that occupy NTL are mainly in the corner of “how”, such as “how is language structured in the brain” and “how is a linguistic construction parsed in the brain to some meaning”. Feldman (Feldman J., 2006) tries hard on the “how” question, going as deep as to connect the workings of molecules and neurons to his theory. The “why” question is equally interesting, although mostly for philosophers. Question that arise are often related to the famous *qualia* or *subjective experience* problem. Why does human experience differ (so much) in the event of drinking a glass of ice cold coke or playing a game of korfbal. No theory can accurately explain it, but an answer could solve classical questions about free will and the meaning of life and potentially debunk the fundamentals of NTL theory such as Embodiment. Addressing this question is needed for creating artificial NLU system and will be discussed in chapter 3 and 7.

The main NTL model presented in fig 2-2 will be discussed intensively in this report (chapter six), such that glancing at it at this point is sufficient. This model was made by Nancy Chang and Benjamin Bergen (Bergen & Chang, 2003) (see also fig 2-1) and was probabilistic enhanced by Eva Mok (Chang & Mok, 2006). It displays how language is learned via (parts of) Michael Tomasello’s usage-based language acquisition theory (Tomasello, 2003) and how meaning is extracted from a sentence using simulation and inference methods, connecting *Form* and *Meaning* using so called *Embodied Constructions*.

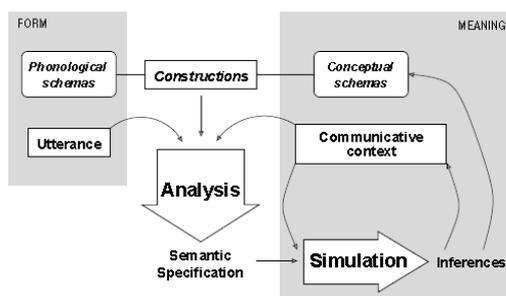


Fig. 2-1 Constructions connect Form and Meaning, image from (Bergen & Chang, 2003)

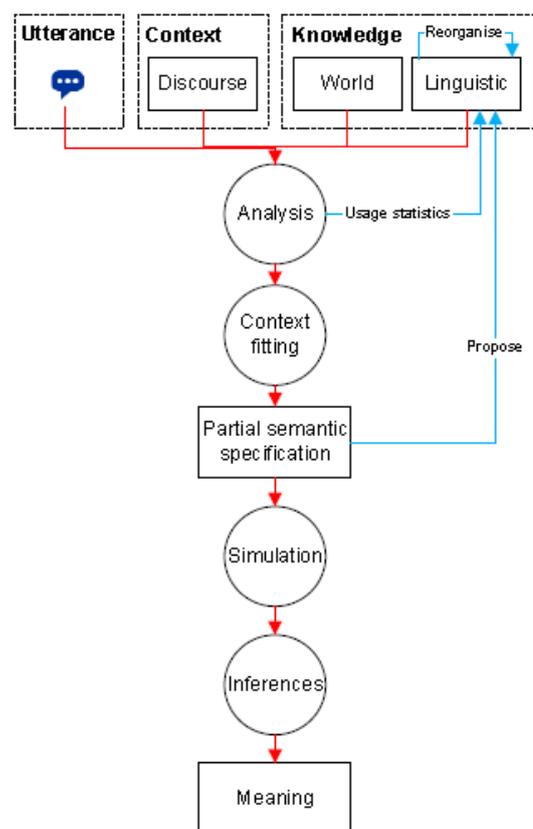


Fig. 2-2 Complete model from sentence (Form) to Meaning, image based on models from (Bergen & Chang, 2003) and (Chang and Mok, 2006)

“A word is not a crystal, transparent and unchanged, it is the skin of a living thought and may vary greatly in color and content according to the circumstances and the time in which it is used.”

Oliver Wendell Holmes

3 A new philosophical framework for NLU

In which is suggested that the current definitions of meaning and understanding are not suitable. To reach new working definitions, Embodiment is used in combination with Shannon and Weavers communication model to explain the importance of the body and environment (“the situated context”). The new definitions are fitted in a new theory of knowledge (epistemology). The chapter is the only one in this report written primarily as a discussion introducing many novel definitions.

3.1 The current definitions of meaning and understanding

A short survey of various dictionaries interestingly produce various definitions of *meaning* and *understanding*, presented in Table 3-1 and 3-2. The work by Carl Bereiter (Bereiter, 2002) provides an extensive overview of many practical and theoretical approaches to defining understanding. This, together with the short survey suggest that there currently is no single (general consensus) definition for understanding. Most of the definitions try to either define understanding using examples of situations (contexts) in which the words are used or can be used (Bereiter, 2002) or don't define it at all by being cyclic and thus crossreferencing a similar definition.

Table 3-1 Definitions of Understanding (noun) / Understand (verb)

Source	Definition
(Dictionary.com, 2012)	Noun: Mental process of a person who comprehends; personal interpretation Verb: 1) To perceive meaning of; 2) to regard as firmly communicated; 3) to interpret
(Penguin, 2006)	Noun: 1) The power to make experience intelligible by applying concepts; 2) mental grasp Verb: 1) To interpret (something) in one of a number of possible ways; 2) to assume, infer, or suppose (something); 3) to have thorough knowledge of or expertise in
(Chaitin, 2006) (paraphrase)	Verb: Being able to figure out a simple set of rules that explain it (“comprehension is compression, the simpler to predict observations, the better understanding of it”)

Table 3-2 Definitions of Meaning (noun)

Source	Definition
(Dictionary.com, 2012)	1) What is intended to be, or actually is , expressed or indicated; 2) the end, purpose or significance of something
(Penguin, 2006)	1) That which is conveyed or which one intends to convey 2) significant quality or value

Why is it so hard to define these two concepts? One of the more trivial reasons is that meaning and understanding are very subjective, their definition depends very much on a persons own interpretation. Dictionaries use clever techniques to deal with subjectivity by using very abstract terms and crossreferences. For example if the definition of *definition* is examined, the Penguin Complete English Dictionary (2006) returns:

Table 3-3 Definition of Definition (noun)

Source	Definition
(Penguin, 2006)	1) Statement of what a word or phrase means 2) a word or phrase expressing the essential nature of a person or thing

The definitions of *meaning* and *definition* are crossreferencing each other in Table 3-3 at 1), so avoiding an accurate definition. In Table 3-4 an abstract definition of a ball is given because the *essential nature* of something is very subjective. If no clear definition can be given of *definition*, how can any definition be correct? What is the *essential nature* of simple words like “ball”, is it really like stated? Looking at the definition of *spherical*, one would find a crossreference to *ball*.

Table 3-4 Definition of Ball (noun)

Source	Definition
(Penguin, 2006)	1) Spherical object

The second interpretation of *definition* in Table 3-3, finding words that express something, clarifies the problem with any definition. Expression using words is bound to the function as well as the limitation of human language. Language has many functions, such as to be informative, declarative or imperative. All functions involve a social setting, ie. a sender, receiver (could be same person) and a message. These functions can be seen as motives for human communication (Tomasello, 2008). The limitations of language are bound to the limitations of the human body and its environment; humans can only express language that is compatible with human speech organs and only process language that is compatible with the human hearing range and processing brain. Furthermore, limitations such as grammar (eg. word length), are not only determined by factors like genetics (attention span, etc) but also on its practical usage in the environment – culture bound.

Human language, in both its function and limitations, are bound to the human body and its environment. Unfortunately, the realisation that language is fundamentally Embodied requires a complete review of how language is structured, used and learned (chapter 6). Fortunately, it allows new and more clear definitions for *meaning* and *understanding* including a framework to study this new definition of meaning in (epistemology), which are discussed next.

3.2 The relation between meaning and understanding

To reach an Embodied definition of *meaning* and *understanding*, it is important to know how both occur during the usage of language, so during communication. This subsection will first discuss the close relationship of *meaning* and *understanding* using the communication model.

The origin of language is often thought to be a function of communication (Tomasello, 2008), which is a functionalist viewpoint opposing (Chomsky’s) formal viewpoint (Butler, 2010). Language *understanding*, how vastly complex it may be, is but one component during social interaction. One of the first to recognize and model the different components in communication were Shannon and Weaver (1948) who came up and popularized the communication model. Initially this model was designed for sending signals (information) between systems, without mentioning human communication. Later, the model was adopted by linguists, speech scientists and others to more realistic models of human communication allowing for instance sender/receiver duality. The reductionist approach by Shannon and Weaver remains the basis for most models, describing four key components; the sender, receiver, the message and the channel. An example of an (preliminary) adopted human communication model is (Bach, 2008) in Fig. 3-1.

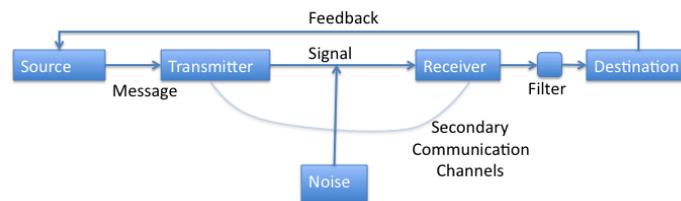


Fig. 3-1 Bach's communication model based on Shannon and Weaver's model (source: (Bach, 2008))

Message

When discussing human senders, messages can be produced with the body using various languages (*forms of communication*), where language is defined as "a structured system to communicate meaning". Language takes the role of the *compression and encoding* of the message, since the content of the message is often not expressible or very cumbersome using bodily primitives (such as fingers to point) for example when discussing concepts like MONEY.

The two main forms of communication are verbal and non-verbal language. Many forms of non-verbal language exist, such as body language, object language (wearing a branded t-shirt) or action language (communicate by events, such as searching for the recipe of Triple Chocolate Brownies means baking them in the future). It is popular belief that the majority of communication is non-verbal. Research by (Pease & Pease, 2004) even claims the information impact of the message on the brain is distributed to vision (83%), hearing (11%), smell (3%), touch (2%), taste (1%). In section 5.3.4 the principle of *sensor saliency* will be discussed stating that information impact is often not fixed and sensor preference depends on content.

Channel

The channel in which a message sent by a human can affect the external sense of the receiver, so one or more of the five primary senses. A message can be sent over a single channel or over multiple (ie. to produce redundant, contradictory and/or complementary information). A channel that is indirectly affected by the message is the internal sense which for instance triggers emotional meaning such as heart rate, transpiration, etc. The environment in which the channel operates such as air or water affects the message by appending noise to it.

Sender (transmitter)

The sender broadcasts meaning, in the form of an intentional or unintentional message. When discussing a human sender, a message is mentally constructed by making assumptions about the knowledge of the receiver. Wrong assumptions can severely hinder the interpretation of the message by the receiver (eg. using a wrong language, talking about an event the receiver has no knowledge about).

The organ responsible for sending verbal messages, the speech organ, has properties to code intentional or unintentional messages (dialect, articulation, lisp, etc).

Receiver

The receiver receives meaning in the form of noisy or noiseless messages and needs to interpret (map) this to understanding. Mostly experience (as will be properly defined later) and also some motivation and mental skills (simulation, reasoning and others as also will be discussed later) will play an important role in the interpretation process. Even a noiseless received message can be interpreted wrongly due to limited interest or unconscious filtering (as many math teachers will acknowledge).

The organ responsible for receiving auditory messages, the hearing organ, has properties to decode the noisy or noiseless message. Unfortunately, messages are often noisified by the channel not mentioning the organ deteriorating with age, resulting in receiving virtually always a noisy message.

With respect to the Natural Language Understanding problem, all four components are relevant. A message encoded in verbal language or non-verbal language, expressed in a channel such as sound or vision and being send or received by actors are all components that solve the NLU problem when studied *together*. In Appendix 4 standard practices are provided on the encoding of a sender sending meaning and a receiver receiving it using various symbols.

3.3 *The Embodied definitions of meaning and understanding*

Undoubtedly, dictionaries write their definitions through careful peer review. These definitions try to cover the objective or culture-free descriptions of definitions – often using disembodied and highly abstract language. This section will discuss the definitions of *meaning* and *understanding* using Embodied language and with that try to reach a lower abstraction level.

3.3.1 Defining Meaning

Meaning is philosophically confined in a spatial and temporal dimension and is further limited by the capabilities (body) of the sender and receiver. In other words, *meaning* is bound to physical context, it depends on time and place. It is also bound to social context, it depends on the sender and receiver(s) including their culture, physical and intellectual capabilities. Given both contexts together, *meaning* is always bound to a situation (see also chapter 2) involving actors and environment, called *situated meaning*. This term will be abbreviated to just *meaning*.

A real life situation is often not discrete, where as the situation in a photograph is discrete. The difference between an identical situation in continuous time compared to discrete time is that more meaning can be conveyed. This seems trivial since a whole dimension is added (time), but if some intended meaning is sent to a receiver and is fully understood by the receiver, then more time will not result in more meaning. The effect of **receiving** more meaning over time is called *spreading (neural) activation* in the brain or *incremental interpretation* by the receiver although as (Schlesewsky & Bornkessel, 2004) mention not all meaning is incrementally interpreted. An example of incremental interpretation is reading a sentence which requires reading the entire sentence before intended meaning can be determined (ie. *garden path sentences* like “*the old man the boat*” – the boat is manned by the old).

The first two paragraphs are properties of meaning – always in *a situation* and the same situation in continuous time has *the same or more* meaning than in discrete time. Meaning itself is packed in a message containing the content (this will be defined in 3.4.1 as *functional meaning*) and the compression and encoding (later defined as *symbolic meaning*) of the content. Both types of meaning are *Embodied experiences* or *experience* for short, which will be properly defined in subsection 3.4.1. For now, examples of experiences are objects that can be seen, heard or felt or activities that can be done. In other words *experiences are any single or multiple sensor value(s) from one or more sensors, be it from the five major senses or from the tiniest muscle sensors that are currently active or simulated to be active.*

As now will become apparent, experience –later defined as the *raw meaning* since it contains both symbolic and functional meaning - is not always easy to express (“*how should a message containing millions or more sensor values in continuous time be expressed?*”). One very efficient way

is to encode functional meaning with symbolic meaning, allowing abstractions such as words to be used. One important tool for encoding abstractions is syntax, so the correct usage of these words. Another efficient tool is to exploit context (discourse, shared experiences, etc) such that messages can be made more compact by leaving out symbolic meaning, expecting -or not- that the receiver understands the decompressed *functional meaning* by assuming the missing parts implicitly. As will be described later, the symbolic meaning is very hollow, so correct usage of words does not imply understanding the functional meaning. To understand the content of the message and not only its encoding and compression, feedback and implicit/explicit mutual agreement from a language community is required.

A distinction should be made between animate and inanimate objects sending meaning. Animate objects such as animals and humans send meaning both intentionally and unintentionally while inanimate objects such as a water or tree send meaning neither intentionally or unintentionally. The lack of determining intentionality of inanimate object is caused by the absence of feedback that can be received and processed by the human receiver. The result is one-way communication which hinders the construction of a mutual understanding between sender and receiver. Note that some meaning **is** send (but one-way) from inanimate objects, for example light reflects from a ball into the eye such that it can be recognized and understood by the receiver. The same ball can evoke additional meaning to another receiver who interacted with that ball. It is important to make the distinction between the objects that are animate or inanimate for the type of meaning they express as defined in the new Embodied definition of meaning:

Meaning /'meening/ noun:

1) Type 1: Symbolic meaning. *Encoding and compression of the experience as an external sensor only embodied experience. A (complex) pattern of values from the five primary sensors which map to (Type 2/3) a possible larger pattern of values from both internal and external sensors. The pattern functions as an interface to Type 2 and 3 meaning, allowing recognition of them. Eg. water (not the word, but the liquid stuff) conveys a symbolic meaning which consist of the properties of water as sensed from the five external senses only, such that it can be recognized as water. The word "water" is decoded and decompresses to the external sensed properties of the concept water. Both are symbolic meaning, however the first is encoded and compressed in a single word while the other is encoded and decompressed consisting of all the (external) embodied properties of water.*

2) Type 2: Intended Functional meaning. *The experience (sensory, both internal and external) consisting of concepts like observations, assumptions, predictions and expectations that are intended to be conveyed. Eg. a person conveys the feeling of his happiness using a smile, the smile is (visual) symbolic encoding, the complex experience (a feeling) named "happiness"⁴ the intended meaning send. Note that Type 2 implies Type 1, since that is its interface.*

3) Type 3: Unintended Functional meaning. *The experience (sensory, both internal and external) consisting of concepts like observations, assumptions, predictions and expectations that are **not** intended to be conveyed. Eg. a person conveys the feeling of his uneasiness using a tic (the tic is symbolic encoding, the complex experience named "uneasiness" the unintended meaning send). Note that Type 3 implies Type 1, since that is its interface.*

⁴ Note that a smile is poor encoding, because "happiness" is subjective and for example the intensity of "happiness" as well as other bodily sensations that activate with *this* person cannot be accurately derived from it. However with additional feedback, that same smile can become strong encoding.

3.3.2 Defining Understanding

Starting on average at the age of two, children are able to correctly use adjectives like “big” and “tall”. However experiments by Tribushinina (Tribushinina, 2008) show that adult-level understanding, the relativeness of adjectives like “big and “tall”, starts at about 7 years. Up until then, Tribushinina hypothesizes that children accurately mimic adult usage of the term - so which objects are referred to as big and how are they combined with other words. As will be discussed in section 6.2, children engage in social settings and experience the effects of language. They finetune their language usage according to feedback from the language community, to reach a usage level that is *normal* (the word *expected* will be used later on) by the community.

The principle of feedback is crucial to develop understanding. For example, understanding the language of a porpoise (related to dolphins) or dog which according to Michio Kaku both have a non-uniform structured and thus “intelligent” language⁵ is at the moment not possible. The link between functional and symbolic meaning a dog sends to a human receiver - so which part of the dog’s experience (eg. seeing a cat) maps to which symbol (in this case a barking sound) - cannot be understood because it is unclear what the dog experiences and how that it is compressed and encoded in a message. The Embodiment hypothesis theorizes that a similar body is needed to understand the dog’s experiences and language behavior.

The first paragraph discusses a distinction between adult-level and child-level language understanding, stating that a child *understands* rather than accurately *copy-paste* adjectives in language around the age of 7. This distinction is too vague to define *understanding*, but does indicate that understanding requires more than simple mimicking competent language users. The second paragraph states that understanding requires feedback and a similar body. Thirdly, the communication model discussed earlier shows that understanding is always subjective, it is an interpretation of the message by the receiver⁶. These three points make clear that to define understanding a clear definition of interpretation and experience is needed.

An interpretation of the message - so both the encoding/compression of the content and the content itself - is defined to be “*the experience that is evoked at the receiver and caused by the sended message*”. Note the subtle difference compared to general definition of understanding the message which is defined to be “*all experiences that can be evoked at the receiver and caused by the message*”. The interpretation of the content of the message (the functional meaning) by the receiver is called the *interpreted (functional) meaning*. Successful communication occurs if the *interpreted (functional) meaning* equals the sended (*functional) meaning*.

It is important to reread the first sentence of the previous paragraph on what makes an interpretation. It was Ludwig Wittgenstein (1953) who noted that “*understanding is not in making the interpretation (in his definition of interpretation) but by abiding to that interpretation*”. The definition of interpretation used here relies on experience which is based on causality, so causes (eg. observations, assumptions) and consequences (eg. expectations, predictions). More importantly, as will be further explained in 3.4.1, these causes and consequences are motivated such that this interpretation **includes abiding** to it. For example a person understands that a ball thrown to the body can either be caught, avoided or simply taken as a hit. One interpretation in a certain situation could be that the person decides to avoid the ball. The motivation to avoid the ball is within this interpretation, leading to all kinds of expectations and predictions regarding the body and environment (including the ball). For

⁵ <http://www.youtube.com/watch?v=9JzXlJyHkks>, last retrieved june 2012

⁶ *Judgement* would be a better word instead of interpretation (emphasizing subjectivity), however Kants *Critique of Judgement* (1790) already gave it a definition. Redefining interpretation is probably less confusing.

instance, the sensors on the body will not expect sensor responses from the ball on the outer skin but do expect sensor responses from the flexes and stresses of muscles because of the activation of avoiding behavior. This example promotes a more pragmatic general definition of understanding which is *forming expectations through interpretation*.

A detailed definition of understanding (see below), depends much on the above definition of interpretation and –like with the definition of meaning – depend on the definition of Embodied experience. The definition of understanding uses the same three types which make a clear distinction between the encoding/compression of the message and the actual content of it – intentionally sent or not. The definition is theoretical since the receiver will have great difficulty determining what sender Embodied experience is intended and what not, so type 2 and 3 understanding will often have identical definitions in practice.

Understanding /und əˈstanding/ noun:

1) Type 1: Symbolic understanding. *Decoding and decompression of the sensed symbols (external sensor only Embodied experience) which map to (Type 2/3) a possible larger experience containing both internal and external sensors at the receiver. Example: an English sender conveys the single word “water” by phone which the English receiver interprets as the noun “water” or verb “(to) water”, with its inherent external properties. In the case of the noun this means the transparent liquid visual pattern (and other) while in the case of the verb it means a visual pattern where an actor is involved in using his body to use some liquid on some object. What object, when and other questions are not part of the symbolic understanding. Symbolic understanding is shallow understanding, it has only direct reference to external sensorial patterns and includes correct usage of the symbols without knowing possible Type 2/3 meanings.*

2) Type 2: Intended Functional understanding. *The experience (sensory, both internal and external) consisting of concepts like observations, assumptions, predictions and expectations which through reasoning match the experience that was intended to be evoked. Example 1: a receiver has two interpretations of the word “water”; the noun or the verb. The sender intended the word “water” to mean the noun, with no further intentions. Example 2: The sender intended the word “water” to mean the verb, and with that some implicit experience as to how long to spray water. The receiver did not understand the full intention and kept spraying until the lawn was submerged.*

Note that Type 2 is a goal state, while Type 1 is an initial state. The goal state has experience associated (“mapped”) to the external sensor experience like assumptions and expectations. However, if it was intended to not have these associations, then it is identical (“initial = goal”), like in the first example. In example 2 the goal state was intended to be spraying for a period of about 10 sec per m². To be clear: when the intended goal state does not equal intended initial state, Type 2 is functional understanding. If it does equal, Type 2 is only referential to the intended external experience and has no assumptions, expectations nor internal sensorial patterns.

3) Type 3: Unintended Functional understanding. *The experience (sensory, both internal and external) consisting of concepts like observations, assumptions, predictions and expectations which through reasoning match the experience that was **not** intended to be evoked. Example: a receiver understands a tic to mean either “excitement” or “uneasiness” (so two interpretations). He does not understand that it can mean a neural defect. The sender conveys an unintentional tic which the receiver interprets as “uneasiness”. The actual meaning of the tic was “excitement”.*

The occurrence of the three types are not uniform. Type 1 understanding is the most common type since a sender with no or unknown intentions always conveys symbolic meaning (seeing it, hearing it, etc). Type 3 is arguably the second most common type. A sender with no or unknown intentions to convey meaning often sends unintentional meaning to the receiver that extends on the symbolic meaning.

The examples from the embodied definitions of Type 1, 2 and 3 meaning and understanding show the problem of subjectivity in communication, so finding the correctness of both encoding/compression and content as desired by the sender. Subjectivity is a direct effect of the Embodiment thesis. Shared experience and validation through feedback is required to minimize the discrepancy (the degree or measure of understanding⁷) between the sender's Embodied experience and the receiver's Embodied experience. This measure of understanding allows a more clear definition of "deeper or better understanding". Since understanding in general equals *forming expectations through interpretation*, deeper understanding equals to have *more accurate or simply more* expectations. For example more accurate or more expectations of events of the past, the near or far future or in more contexts. In terms of pattern recognition (since humans are often referred to as biological pattern recognition machines), deeper understanding can be explained as finding a larger matching or more accurate pattern such that more data can be recognized, leading to better predictions and eventually expectations.

3.4 Epistemology: Study of knowledge

The new definitions of *meaning* and *understanding* rely heavily on the definition of experience. In this subsection experience is defined, including how it closely relates to knowledge. The branch of philosophy that is concerned with questions like these is called *epistemology* or *study of knowledge*.

3.4.1 Defining Experience

"There can be no doubt that all knowledge starts with experience"
Immanuel Kant in *Critique of Pure Reason* (1781)

At his point it is important to shortly (re)state the implications of the Embodiment thesis on the definition of experience (for more assumptions see, (Gallagher, 2006)). *"The position that experience is embodied- that is, structured in part by the nature of the kinds of bodies/neuro-anatomical structure we have- has consequences for conceptual structure. This corollary follows because the concepts we have access to, and the nature of the 'reality' we think and talk about, is a function of our embodiment. In other words, we can only talk about what we can perceive and think about, and the things that we can perceive and think about derive from embodied experience. Hence, the human mind must bear the imprint of embodied experience."* (Evans, 2010, p. 199). In other words, Embodiment assumes a species-specific view of the world. Furthermore, (Lakoff, 1987), (1980)) using his metaphor theory theorizes that embodied experience allows the construction of any (abstract) concept and as such is the foundation of knowledge (see 3.4.2).

⁷ Note that measuring understanding is now seemingly trivial by comparing identical types of meaning and understanding. Ofcourse some "omniscient entity" needs to do the fair comparison of observations, assumptions, predictions and expectations since someone's belief (see 3.4) on these differs from reality.

Experience is a concept that many readers already have a gutfeeling about. For example it can be divided in social experience like courtesy rules, physical experience like kicking a ball or mental experience like simulating scoring a goal. Another way to discuss experience is the difficulty to acquire it, like the experience of pursuing a gold olympic medal or reaching the top of the Mt. Everest. In section 3.2.1 experience was introduced to be *“any single or multiple sensor value(s) from one or more sensors, be it from the five major senses or from the tiniest muscle sensors that are currently active or simulated to be active”*. However, how can these sensor values (the firing of a sensorneuron) be transformed into abstract concepts like “time” or concrete concepts like “a ball”. What does the firing of a single neuron or a network of neurons mean? How can symbolic and functional meaning and interpretations (or understanding in general) be made clear using embodied experience? In this subsection a concrete definition (using Embodiment) is provided which in combination with a definition of knowledge tries to answer these questions. To formulate these definitions, a philosophy of mind and language needs to be chosen.

Throughout the course of history many approaches have been proposed to model the mind in general (beliefs, desires, knowledge, etc) and model behavior (such as language) in particular. For this report a *monoist* and *cognitivist-functional* approach is chosen which in short models the mind as integral from the body, which fits the Embodiment thesis (Evans, 2010, p. 199). *Dualist* and *formalist* approaches argue that the mind can be modelled as distinct from the body. Choosing the above approaches does not imply that other theories and models are wrong or incompatible, they simply do not fit as well with the Embodiment assumption as the other do with only a quick scan through literature. It was never the aim of this report to review the stellar array of publications on philosophy of mind (eg. (Ravenscroft, 2005)) and philosophy of language (eg. (Barber & Stainton, 2010)) that exist.

The functionalist approach has three key principles (Butler, 2010, p. 268):

- Regard communication as the primary function of language, which shapes the forms languages take
- Attribute great importance to external (cognitive and socio-cultural) factors in explaining linguistic phenomena
- Reject the claim that syntax is autonomous from semantics and pragmatics

The distinction *cognitivist-functional* emphasises (Evans, Bergen, & Zinken, 2006, p. 30) that equal interest is put in both the Form and Meaning of language. Based on the principles, many language acquisition theories have been invented (with some unclear overlap⁸). One such theory is the constructivist view (construction grammars) which states that *“enough information required to construct a grammar is available in the linguistic environment”* (Butler, 2010, p. 270). Another view, not necessarily excluding the constructivist view is the socio-pragmatic view which states that *“a child know that a speaker who produces a word whilst looking at an object (a) is attending to that object and (b) intends to label that object for the child”*. (Ambridge & Lieven, 2011). Again another view, also compatible with both other views, is the usage-based view, which states that *“language acquisition is driven by desire to use language to communicate”* (Ambridge & Lieven, 2011). This group of models claim that *“the speakers linguistic system is fundamentally grounded in “usage events”: instances of a speaker’s producing and understanding language”* (Barlow & Kemmer, 2000). Key principles of this view according to (Tomasello, 2009) are:

⁸ As (Butler, 2010, p. 268) mentions, a larger dimensional typology is needed to clearly distinguish functionalist theories from each other – there is much overlap in the different approaches such as emergent grammar, (embodied) construction grammar, functional grammar, semantic grammar and even Chomsky’s universal grammar (Newmeyer, 2003).

- Meaning is use ((Wittgenstein, 1953))
- Structure emerges from use ((Langacker R. , 1987) (2000))

The usage-based view is built upon the “meaning as use” semantic view (Martin, 2010). This philosophical theory on semantics differs from other established theories like *direct reference theory* and *idea theory*. Direct reference theory (Mill, 1843) treats meaning as a direct reference to objects, ie. words get meaning by referring to objects. A few problems arise with this theory, most importantly that sentences like “my name is Santa Clause” mean more than a reference to the name property of the person. Often it means a social introduction or a funny statement – so regardless of what the name or reference was (Martin, 2010). Another theory, idea theory (Martin, 2010, p. 436), which is very similar to classical Greek philosopher Plato’s *theory of Forms/Ideas*, suggest that meaning are ideas, where ideas are referred to as mental imagery or mental activity in general. A problem with this theory is that two people might have identical meaning (eg. a cat), but not identical imagery – for example the prototypical cat differs per person (Martin, 2010). Wittgenstein (1953) suggested a pragmatic view and linked meaning to use, the usage-based view which is the third major and currently popular theory on semantics. However, I argue that not every meaning is usage-based, since direct reference-based with no additional use cases are also possible. For example imagine during a walk the word “anklyrias” painted on a wall. It has no use cases, no intension, no mental imagery, no reference - it is basically an *empty name* (Caplan B. , 2010), a series of symbols . However at the exact moment of observation, I argue that it receives immediately Type 1 meaning, the symbols are placed in context (location, how you feel at the time of observing, etc). After the moment of observation its use (function, Type 2 meaning) is still unclear, however it can be referred to and thus has symbolic meaning or direct reference. A usefull construction to model the two semantic theories is Frege’s “Sinn und Bedeutung” or Sense and Reference (Frege, 1892). Using this construction meaning (“the message” as it is used in the communication model) is split in a Form (the representation of the message, the symbolic dimension, Type 1 meaning) and Meaning (the content of the message, the functional dimension, Type 2/3 meaning). A meaning construction is one holistic Form-Meaning pair. Note that with this construction idea theory can also be fitted in; only the representation (Type 1 meaning) and references and use cases (Type 2/3) needs to be shared by a language community. A “private language” or additional personal symbolic and functional meaning, not shared with any community can easily be added.

Within the group of usage-based, a distinction can be made between speech-act-based as opposed to truth-conditional-based meaning. Speech-act-based models regard “*language as action; a rational agent uses language in an attempt to change the state of the world*” (Vieira, Moreira, Wooldridge, & Bordini, 2007, p. 225). Meaning is dependent on the intention of the agent and not on the truth condition of the used language (see Example 3-1). Speech-act compared to non-speech-act emphasises the focus on attempting to change the mental state of the receiver (Vieira, Moreira, Wooldridge, & Bordini, 2007, p. 225). The intention-dependent view fits well within the Embodied approach to the NLU problem since it requires the permanent inclusion of an agent. Speech-act theorists make a similar semantical distinction as used in this report. They use the terms *locutionary act* (similar to the Type 1 “symbolic” meaning), *illocutionary act* (similar to Type 2 “intended” meaning, except that Type 2 meaning also includes acting upon the intention) and *perlocutionary act* (the effect of the illocutionary act on the receiver, whether it was intended by the sender or not). I argued in 3.2.1 that intention is deeply intertwined with the expectation of the receiver acting upon it, not just understanding the linguistic function of the intention but the “full” intention –whether conscious or unconscious.

- 1) I dropped ten marbles and found all of them, except for one. It is probably under the sofa.
- 2) I dropped ten marbles and found only nine of them. It is probably under the sofa.

Example 3-1 From a truth-conditional point of view both sentences have equivalent meaning, so disregard the hidden communicative differences. (source: (Groenendijk & Stokhof, 2010))

In short, the theoretical framework to deal with the NLU problem in this report is embodied, monoist, functionalist-cognitivist, constructivist, usage-based, socio-pragmatic and speech-act-based. This framework allows generating a definition using a reductionalist (bottom-up) formulation to answer the questions this subsection started.

Experience / *ik-speer-ee-uhns* / noun:

- 1) Any single or multiple sensor value(s) that are sensed (actual or simulated) by one or more sensors belonging to a single actor in the spatial and temporal dimension.

3.4.2 Defining Knowledge

One of the main question in Epistemology is the question “*what is knowledge?*”. For starters, what is knowledge for someone is just senseless data (*empty data*) for someone else. Knowledge is thus very personal and is only knowledge if it has applications (is usable). The psychologist and Nobel laureate Daniel Kahnemann mentions the difference between “the experiencing self” and “the remembering self” is large⁹ and thus the relevance of processing is important in the definition. For example an extended holiday with no new activities seem important but has no huge impact on memories since nothing new happened. What is important according to Kahneman is not just the temporal dimension (the amount of changes) but the spatial dimension as well (the relative size of the changes).

Using these insights, knowledge can roughly be defined as “*stored processed experience waiting to be used just-in-time*”. The just-in-time part makes sense when considering that knowledge needs to be recalled from storage under time constraints to be usable. Also the use statement is included to emphasize that knowledge needs to be applicable for some situation, otherwise it would be a record of senseless (symbolic) data. Furthermore, if knowledge is not stored, so not in memory, it cannot be used to think and reason with. Fourthly, related to procesing, not all aspects (defined to be *features* in Appendix 6) of an experience are stored. Some features are merged (implicit knowledge such as motoric skills) while other features are stored one to one (explicit knowledge such as visual symbols). Also, not all features are selected to be processed for a multitude of reasons, for instance they are not salient enough or the brain is physically not able to connect two brain areas. Although this rough defintion is embodied in nature, it can be made more specific with the question “*when is experience knowledge*”, which forces to define “processed”. The above description on processing experience into knowledge is insufficient, hence the need to include the link between experience and knowledge into the defintion – a rough definition on knowledge aquisition. Consider the current definitions of knowledge in Table 3-5:

⁹ http://www.ted.com/talks/lang/en/daniel_kahneman_the_riddle_of_experience_vs_memory.html

Table 3-5 Definitions of Knowledge

Source	Definition
(Dictionary.com, 2012)	Noun: 1) Acquaintance with facts, truths, principles, as from study or investigation 2) Acquaintance or familiarity gained by sight, experience or report
Plato (+/- 380 B.C.)	Noun: 1) Knowledge is justified, true beliefs
(Collinsdictionary.com, 2012)	Noun: 1) The facts, feelings or experiences known by a person or group of people. 2) Specific information about a subject. 3) Awareness, consciousness or familiarity gained by experience or learning

Broadly defined, there exist two categories of knowledge, *a priori* (“from the earlier”, the past) and a *posteriori* (“from the later”, the future). A priori or prior knowledge is independent of new experiences. It is knowledge which can be verified true or false without - as Galen Strawson phrases - “*having to leave your couch and examine it in the physical world*”¹⁰, for example the sentence “if I can count to ten, I can count to nine”. New knowledge is gained by using the existing knowledge (“stored experiences”) to create new knowledge. A posteriori or later knowledge is dependent on new experiences. It is knowledge that can only be verified by examination, so using the senses, for example learning to count to ten. I argue that a third type can be defined for the sake of easier discussion of knowledge; *a nunciari* (“from the now”, the current). This knowledge is unprocessed, but is **sensed** experience and can be thought of as raw knowledge (see Fig. 3-2).

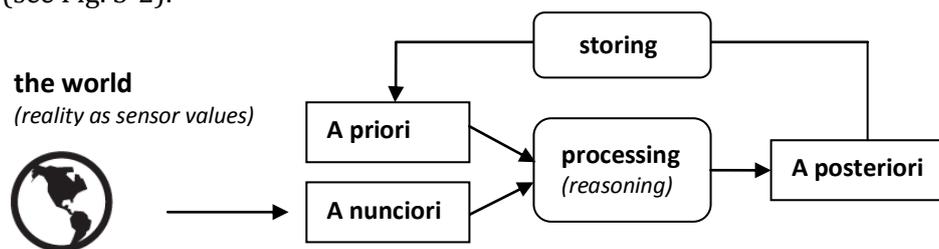


Fig. 3-2 The relation between three types of knowledge

The two doctrines empiricism and rationalism theorize how knowledge is acquired, in other words how the processing to reach a posteriori and eventually a priori knowledge works. In short the two doctrines differ in their claim on the main input for processing (Prinz, 2010):

- **Empiricism.** New knowledge is mostly gained through the senses (a nunciari) using deductive reasoning.
- **Rationalism.** New knowledge is mostly gained through prior knowledge (a priori) using inductive reasoning.

The question of importance is not interesting for the definition of knowledge. Both are important and differ maybe at different stages during life. Although for certain is that a human life is too short to experiences everything first hand, inductive reasoning allow creation of second hand experiences. Furthermore, relying (*believing*, to be discussed in 3.4.5) solely on empirical evidence to form knowledge will not be a good representation of reality; for example I can't see nor otherwise experience oxygen, but I believe, rationally thinking, that it exists around me.

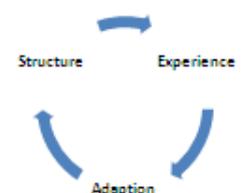


Fig. 3-3 Scaffolding experience (Source: (Feldman J., 2006))

The model in Fig. 3-3 depicts cyclic behavior. The phrase “build upon prior experience” can be

¹⁰ http://www.believmag.com/issues/200303/?read=interview_strawson, last retrieved june 2012

modelled in it. One gets better in understanding an environment through more experience, which can be translated as having more and more accurate expectations in a environment due to the greater base of knowledge. While the feature values from the senses (see Appendix 6) are the fundamental building blocks of knowledge, the smallest unit of knowledge is a pattern of feature values (in spatial and temporal dimension) deemed to be cohesive. It is good to mention that the process of finding patterns in *a nunciari* and *a priori* knowledge is fundamentally embodied and thus both a cognitive and biological one. The brain needs to physically adapt to consolidate new knowledge (detect relevant patterns, since not all experience is relevant) and allow extension on it. Also, what is deemed relevant is not always cognitively determined, but also due to biological mechanisms. (Feldman J. , 2006) describes a cyclic model in which this process is modelled, the Experience-Adaption-Structure (EAS) cycle (Fig. 3-3). In short, from any experience, regularities (patterns) can be detected which can be stored such that new regularities can be created with and upon it. In section 6.2 a theory will be discussed that focusses on the detection and acquisition of linguistic knowledge from experience.

The notion of discovering patterns through inductive and deductive reasoning and through physical development extend the rough embodied definition of knowledge into a slightly more explicit one:

Knowledge / *nol-ij* / noun:

1) *Stored discovered patterns(in spatial and temporal dimension) based on experience waiting to be used just-in-time*

3.4.3 Defining Logic

“Logic will get you from a-z, imagination will get u everywhere”
Albert Einstein

With the new definitions for meaning and understanding a clear separation between symbolic meaning – the hollow message - and functional meaning – the content of the message - is introduced. However on the neurological level, all types of meaning are a pattern of activated neurons (defined as the raw meaning). The whole pattern, from the start of the experience to the end, is coherent, meaning that the neurons responsible are connected and structured.

It now becomes apparent how important connection management is; together with the neurons that fire at some point in time it makes up what a pattern actually is. Connection management is closely tied to logic, the study that investigates the validity of conclusions derived from premises (see the general consensus in Table 3-6) which perfectly fits the Embodiment hypothesis and is therefore left unchanged. A whole proposition is defined to be illogical, invalid inference or incorrect reasoning if¹¹:

- a connection between a premise (eg. a sense neuron or “an observation”) and a conclusion (eg. a motor neuron or “an action”) does not exist. This is error is named *unknown inference validity*, see principle 1 below.
- a wrong or false conclusion is drawn from (the connected) correct or true premises. Note the subjective word “wrong”, meaning that logic is always subjective. This error is named *invalid inference*, see principle 2 below.

¹¹ Note: all terms are intended to mean the same

Table 3-6 Definitions of Logic

Source	Definition
(Dictionary.com, 2012)	Noun: <i>The science that investigates the principles governing correct or reliable inference</i>
Averroes [+/- 1200]	Noun: <i>The tool for distinguishing between the true and the false</i>
(Collinsdictionary.com, 2012)	Noun: <i>The branch of philosophy concerned with analysing the patterns of reasoning by which a conclusion is drawn from a set of premises</i>
(Penguin, 2006)	Noun: <i>a science that deals with the formal principles and structure of thought and reasoning and assesses reasoned arguments on the basis of their validity or invalidity.</i>

Communicating a pattern means communicating the active neurons (sending out positive or negative votes) **and** its activation pattern. This activation pattern consist of a set of connections between neurons and includes for each connection an entailment: “*which input neuron(s) need to fire postively and which negatively (or not fire at all) such that the weighted total sum is positive enough (above threshold) to fire an output neuron*” (paraphrased from (Feldman J. , 2006, p. 53). This view of logic is different from traditional view such as Boolean logic which only considers the entailment to be true or not, while implicitly assuming a connection is available. In my view of logic, the explicit availability of a connection models certain knowledge *to be known* (and thus **can** be reasoned with, so can evaluate the entailments), while inavailability of a connection models certain knowledge to be *unknown* (and thus **cannot** be reasoned with). Using this view, interpretations of propositions such as:

- Example 3-2** *1 + 1 = 1 or Susan*
Example 3-3 *1 + 1 = Koen*
Example 3-4 *if you dive in water, you get wet*

can be analysed and believed to be logical or not based on:

- **Principle 1.** Availability of a connection between premises and conclusion
- **Principle 2.** Correct entailment (if premises are true, conclusion must be true)

In the case of Example 3-2 most receivers will have no connection (so *no prior knowledge*) between an addition between numbers and a result with a name **prior to the question**. The question forces them to create a new connection (“unknown becomes known”) by assuming it right after reading it. Every connection has an entailment attached to it and in the case of a new connection (new knowledge) the attached entailment is derived through reasoning. The result is that some readers might find the new knowledge correct and thus logical. Other people, however, will induce or deduce (processes of reasoning, see section 6.3) it to be invalid entailment, for example due to contradicting prior knowledge that a connection of $1 + 1 = 2$ is true and implicitly causing all other entailments to be false (or differently interpreted: the threshold for other hypotheses to be true is high). If, however, additional symbolic meaning is provided to the existing symbolic meaning by stating that the proposition means 1 father + 1 mother have 1 child or name of child, then future constructions such as Example 3-3 will be believed logical. What is communicated through this explanation is the activated neurons (abstractions of father, mother, child) and its activation pattern (a connection between father, mother and child including syntax). Note that that logic is completely independent of symbolic or functional meaning (and independent of understanding as will be discussed at the end of the chapter). It does not imply one or the other. In the case of Example 3-2 the symbolic meaning can be interpreted as logical, but the functional meaning can be interpreted as illogical, for example since **parents** could have **negative, zero** or more than 1 child. Logic always occurs in a situation (like meaning and understanding) therefore one could define logic - as used by humans - as *situated logic* and like mentioned above, is always subjective.

In the case of Example 3-4 most readers will have a connection (so have prior knowledge) of the premises (dive, in water) and the conclusion (get wet). The processes of reasoning *tries to find proof* or in terms of neurons: *evaluate the weighted positive and negative votes and see if the sum is larger than a threshold, if so a proof is found*. The interesting bit here is the potential amount of uncertainty involved despite finding the example logic. For example, with 20% positive, 10% negative, 70% not voting and a low threshold of 10%, a proposition is still believed logic despite it being highly uncertain¹². In other words some degree of evidence (proof) is believed for the conclusion but no entailment is ensured. It is however enough to be believed a logical proposition. High thresholds model low uncertainty, implying that expectations are frequently met. Note that uncertainty is always present due to the fact that the non-existence of unknown knowledge can never be guaranteed. In the case of Example 3-4 many experience with diving in water will have sufficient proof for it to be logical that one gets wet. However when diving while wearing an atmospheric diving suit one doesn't get wet. Given this additional premise, the example become illogical since one doesn't get wet. This also emphasizes the black and white nature of logic, it either is or is not, similar to (and founded on) a neuron firing or not. For a proposition to be 70% logic for someone means that while the entire proposition is found to be illogical, if chosen just the 70% of the premises that are found to be true according to the prior knowledge of that person, the proposition becomes true.

Another and final situation discussed here using the proposed view of logic, is when someone states "to understand the logic of a proposition but not find it true". This is interpreted as "understanding that the receiver finds the proposition logical if being the other person", so by simulating her prior knowledge and mind reaching a logical inference, which compared to the inference by his prior knowledge and mind reaching a logical inference is illogical. This can happen due to a multitude of reasons, such as drawing the same conclusion but based on different premises.

3.4.4 Defining Truth

In the previous paragraph it was shown that valid reasoning is no guarantee for something to be actually true or believed true. Up until now the terms *true*, *false*, objective and subjective *truth* were used without being clear on what these four concepts mean and how they relate to each other. It is again important to clarify these concepts as they need to be implemented in artificial NLU systems. Belief will be discussed in 3.4.5.

If looking at the definition of truth in Table 3-7, truth is again a historically problematic definition; very general and full with crossreferences. We already saw that valid reasoning is no guarantee for it to be actual or believed true.

Table 3-7 Definitions of Truth

Source	Definition
(Dictionary.com, 2012)	Noun: 1) the true state or actual state of a matter. 2) conformity with fact or reality
(Collinsdictionary.com, 2012)	Noun: 1) the quality of being true, genuine, actual or factual. 2) a proven or verified principle or statement; fact

¹² Note that these percentages are only solid and verifiable when discussing neurons (eg. 20% of the neural circuit fires positively can be verified). When discussing belief of a percentage, this number is not solid, but rather a feeling or reasoned estimate. Since it is a belief (see next section), the use of such a statement can be questionable. In the case of beliefs, which are always subjective, it might be better to discuss the degree of evidence for logic in subjective terms like "substantial" or "few" evidence and use hard numbers only in theoretical and neurological scenarios.

(Penguin, 2006)

Noun: **1)** *the state or quality of being true or factual. 2)* *something that is true or is believed to be true, as distinct from lie*

From these definitions two things can be extracted; firstly that some kind of proof is needed for something to be true (with no explicit evidence something is implicitly false or unknown) and secondly that truth is a property of a state, for example a concept (“fish need water”) or linguistic construction (“tomorrow it will rain”).

Before forming the new definitions of truth, the Embodiment hypothesis is recalled. Since Embodiment is very personal by nature, any truth - be it subjective or believed objective - is always subjective. Any definition of truth should account for this. Using the this and the above, truth can be defined on multiple levels as:

Truth /*ˈtrʊθ*/ noun:

1) Subjective truth. *A state which proof is evaluated by an individual. A state is believed true if and only if proof exist. Reality is made up of believed truths which proofs are provided by direct and indirect Embodied experience.*

2) Objective truth. *A state which proof is evaluated by an individual but is believed to be the consensus of self and other or other individuals. Objective truth can be seen as a collective (eg. culture) truth since its objectivity depends on the collective; another collective might find the discussed truth (so the state) false, since its proofs are believed to be not relevant or not existing.*

The ultimate objective truth is defined to be the universal truth and holds the consensus of the majority of humanity, its proofs are so overwhelmingly convincing, eg. the color of an apple is red or green (color is grounded in direct Embodied experience, a visual primitive).

3) Divine truth. *Distinct from subjective and objective truth is the divine truth, a state which proof is not evaluated by any amount of individuals. It is the absolute truth or simply the truth. Its proofs are not necessarily Embodied, it is independent of what an individual or group of individuals believes and is a truth that can never be proven by humans. Eg. in ancient times of war, history was often written by the victor, a one sided representation of history. However, even if the loser was to write it or even stranger, history written in perfect collaboration, the actual truth might still be different. For example a third party might be involved, who pulled the strings unknown to both. The actual truth can only be known if there are no unknowns and no uncertainty, it thus requires an omniscient being, hence defined to be the divine truth.*

3.4.5 Defining Belief

Deeply intertwined with truth is belief. For something to be believed true (eg. “it is raining now outside”) proof or evidence is needed, as defined in the three definitions of truth. The proof can be gathered empirical (looking outside the window) or rational (weather forecast was sunny for today). This proof is not infallible, for example optical illusions could be framing the eyes or a wrong forecast. Also, the proofs are not necessarily compatible. Multiple proofs might compete (see example in previous section on Logic) for something to be true, and a single additional proof might change the believed truth from true to false or vice versa. In addition, proof about proof is required, making belief cyclic – and causing uncertainty. It leads to the question when something is believed true and how much it is believed true.

Table 3-8 Definitions of Belief

Source	Definition
(Collinsdictionary.com, 2012)	Noun: 1) A principle, proposition, idea, etc. accepted as true 2) trust or confidence
(Schwitzgebel, 2006) from wikipedia.com	Noun: 1) Belief is the psychological state in which an individual holds a proposition or premise to be true
(Penguin, 2006)	Noun: 1) Conviction of the truth of some statement or the reality of some being, thing, or phenomenon, especially when based on examination of evidence.

Before discussing this important cyclic problem, first a definition of belief is needed. The definitions of various sources in Table 3-8 mention acceptance and confidence as a basis for truth. These two concepts introduce a degree of belief in truth since something can be not fully accepted or not fully confident in. Furthermore, the mentioned definitions allow something to be true without proof. In the new Embodied definition of truth, proof is required for example using implicit assumptions. The new definition of belief is required to incorporate degrees of belief for combating uncertainty and be compatible with the Embodiment hypothesis:

Belief /bi'leef/ noun:

1) The availability (existence) of evidence (proof) to determine the truth of something. For something to be believed true is to accept the weighted average of available proofs in favor to be larger than proofs against.

When the weighed average of available proofs in favor is significantly larger than proofs against, a belief is called a believed fact or simply fact. For example referring to the rules of a game written on a paper is a strong proof, thus a fact.

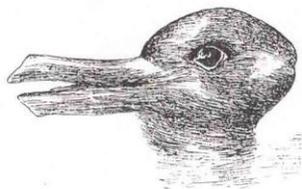


Fig. 3-4 The Duck-Rabbit (Source: (Wittgenstein, 1953))

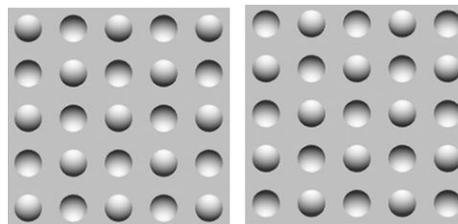


Fig. 3-5 Bumps or Holes? (Source: (Ramachandran, 1988))

The question “when is something believed true?” with the new definition becomes “when is proof accepted? At the embodied level, the physical level, a neuron fires or not. The firing of the neuron can be modelled as proof; something is explicitly true if the neuron fires (the proof), else it is implicitly false or unknown. However humans are not aware of their individual neurons firing and this is where belief comes in. In evaluating the truth of questions like “it is raining now outside” assumptions are made which are beliefs of proof, eg. “I believe I’m not seeing rain outside”. Direct sensory experience (observations) is often felt as strong proof (“seeing is believing”). Empirical proof has difficulty with competing rationalised (indirect sensory experience, the *a priori* knowledge) proof, often overruling it (see Fig. 3-5). (Ramachandran, 1988) shows with this image that *a priori* knowledge changes the perception of an object fundamentally, in this case the *a priori* knowledge that light comes from above changes a bump into a hole when rotated 180 degrees. The same holds for Fig 3-4, where (Wittgenstein, 1953) show that ignoring prior knowledge is impossible when classifying an the image as either a duck or rabbit but not both at the same time. It might very well be true that three or more animals are hidden or that the duck-rabbit itself is a (unknown) species, but the observer has no

prior knowledge of it. In this case these animals do not interfere with the observation and classification of the image since they cannot be recognized.

So when is something true at the believed level? An important problem of belief is the seemingly infinite need for assumptions based on assumptions or proof about proof, the so called *regress argument*. The in-depth analysis of seemingly neverending *why's* will remain undiscussed in this report. For the NLU problem a practical solution will be chosen, the philosophical school of skepticism. This view treats proofs as never air-tight, always leaving (some) uncertainty, since humans are not omniscient and thus encourages to keep investigating. Some other notable school are foundationalism (there exist foundational/elemental proofs), coherentism (there exist a coherent system of mutually supporting beliefs, for example circular proofs) and infinitism (there exist no repeating or converging structure, no foundational proof ever). An interesting book on coherence in natural language (eg. dialogues) is written by (Wolf & Gibson, 2006).

The regress argument shows why it is difficult to determine the truth of something. It is not always easy or possible (limitation of language!) to explicitly express the assumptions which underlie the truth of something. It is often felt as merely a intuitive feeling, which effectively is Embodied experience.

Using the new definitions of logic, truth and belief, statements can be analysed such as "I cannot understand the truth" which means that the entailments cannot be made by this person, either because there is no connection between the form and meaning or because the entailment is believed to produce a false result based on the believed proof. Or statements with numbers (see page 24 for some remarks on usefulness) such as "when the swallows fly low it will rain tomorrow" is believed 70% logic and 40% believed true, meaning that 40% of the proof is believed present as evidence for making 70% of the premises for the conclusion true and the rest false or unknown.

Discussion

"If you can't explain it simply, you don't understand it well enough"

Albert Einstein

Chapter 3 provides the philosophical framework to read this RA as well as build NLU machines. Hopefully the terminology are conveyed as intended by the author and hopefully the proofs are sufficient for it to be believed true as well as being logical. In chapter 6 a model for knowledge, belief, logic, truth and uncertainty is given which make the theory more tangible.

*I argue that the receiver **understanding** (any experience) **is independent of** him/her finding it **logic and/or belief** it to be true, since entailment and evidence can be simulated using the aquired knowledge. By understanding Type 1 and 2 Meaning, the receiver simulates (see chapter 6.3) the logic or beliefs as evoked by the sender and compares this with the logic or beliefs that are stored in his/her memory. For example if I declare that tomorrow the sun will explode, most (sane) people will understand my Type 1 and 2 Meaning but do not belief it to be true and do not find it logic. Another example: if I declare that I ordered 3 pizza's for myself (and show the receipt), people will again understand my Type 1 and 2 Meaning, probably belief it, but not find it logic.*

While epistemology covers many more topics, the discussed topics belong to the core of the field. Because this chapter is written as a discussion with more interpretations and novel ideas than widely cited scientific references, no additional topics are discussed here.

"It is with children that we have the best chance of studying the development of logical knowledge, mathematical knowledge, physical knowledge, and so forth"

Jean Piaget

4 The human NLU system: Early developments

In which various stages including the genesis of cognitive development of infants is discussed using Jean Piaget's theory. Special attention is paid to one cognitive development, that of language understanding.

With a theory of knowledge and definitions on meaning and understanding in mind, next attention is given to the actor, perceiving meaning and processing it to understand it. The next three chapters focus on the human and its NLU system. In this chapter the general development of the body (behavior) and the mind (cognition) is examined as well as some general NLU benchmarks (vocabulary and speech acts) of an English infant. In chapter 5 components of the human language understanding system are discussed in isolation and finally in chapter 6 theories are discussed of how these components might work together to enable NLU.

4.1 The genesis of cognition

One unknown aspect of the human cognition is its genesis. Another is the detailed mechanics of human cognition. Behavioral analysis of infants and young children by (Rizzolatti & Arbib, 1998) show that deficits in social growth (such as Autism) go hand in hand with deficits in language and motoric action growth. Ergo, the hypothesis that language, action and interaction develop in parallel and have impact on each other is born, although this hypothesis could very well be just a correlation. For certain is that all parts of an infant's brain are developing, although some parts might develop faster than others. So, at least initially the assumption of parallel development holds. The **cognitive bootstrap hypothesis** states that *"parallel development of action and social interaction permits the bootstrapping of cognitive development"* (ITALK, 2010), see fig. 4-1. In situ sightings of 1-2 year old children at a Dutch daycare show that when exposed to stackable blocks, the task of stacking blocks is abandoned after around two repetitions. The majority of the children show more interest in interacting with other children by throwing the blocks. Without claiming any solid position due to lack of scientific evidence, it might support the notion that learning social skills through interaction at that age is slightly more important than learning novel tasks, although both are attended to with interest. One reason for this might be the infants' dependency on social skills to survive at that age. Equally interesting is the possibility of a relation between the short attention span typical to young children and the parallel development of many cognitive functions. A positive result would support the bootstrap hypothesis.

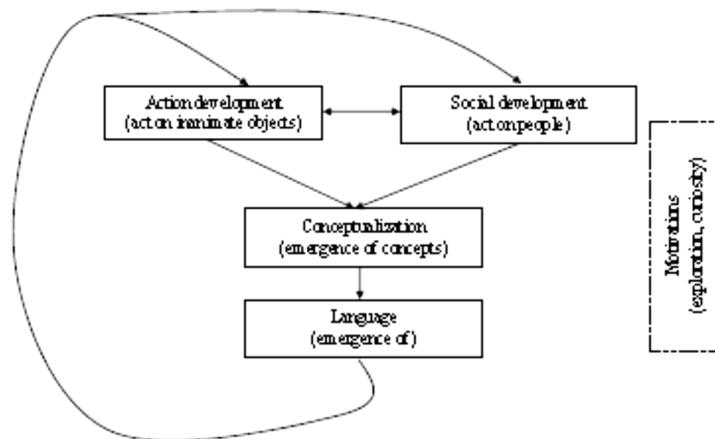


Fig. 4-1 Bootstrapping cognition (source: (ITALK, 2010))

4.2 Early cognitive development

Jean Piaget (Piaget, 1983) found, through generalisation of empirical evidence, that child cognitive development could roughly be divided in four learning stages. The first stage, the sensory-motor stage, occurs from birth to around 24 months.

Piaget's theory is highly controversial since it does not account for contextual development and is therefore a purely "nature" model. However, any model of child development should validate with the empirical evidence Piaget found and is therefore a good reference point. Piaget discovered six substages in the sensory-motor stage (see Table 4-1), which normal healthy infants independent of social context will pass through. They are crucial for acquiring language understanding and other skills:

Table 4-1 The six substages of Piaget's first stage of cognitive development (sensory-motor stage)

1. Simple reflexes	Coordination of sensation and action through reflexive behaviors, such as movement tracking with eyes and grasping.
2. First Habits and Primary Circular reactions phase (CRP)	Reproduction of an event that initially occurred by chance. Classical conditioning starts.
3. Secondary CRP	Development of habits. Repeat actions that bring interesting or pleasurable results. Intentional grasping, differentiate means and end, repetition of actions involving external object).
4. Coordination of secondary CRP	Coordination of vision and touch (hand-eye), schemes and intentionality. Deliberate goal oriented planning starts.
5. Tertiary CRP	Discovery of new means to meet (novelty, curiosity) goals and discovery of many properties and interaction possibilities of objects.
6. Internalization of schemes (insight, creativity)	Use primitive symbols and form related to mental representations such as sound, words, drawings.

4.3 Development of language skills

During Piaget' sensory-motor stage, so from birth to 24 months, language understanding start developing. There is scientific (and empirical) agreement on the two phases an average infant passes to get to the point of adult language understanding, a passive (or receptive) and an active (or expressive) language learning stage (Bowen, 2011). Both stages are accompanied with body language interaction, except the latter allows finer nuance in understanding due to vocal interaction. Since almost nothing develops in serial in the brain, these two stages co-develop as well. A point of discussion is if the biological development seems to focus on passive language learning more initially (from pre-birth to about 1 year) and changes gradually after that with the rapid development of the voicebox.

4.3.1 Passive language learning

In the passive language learning stage an infant is able to learn and understand language (as defined in chapter 3, *so forming expectations through interpretation*) but is not capable of speech production. Already at pre-birth, the infant starts to take an interest in sounds and language and discovers patterns in it, for example hearing rhythms in the language used by its mother. After birth at around 4-5 months, patterns are recognized based on multi-modal sensory fusion, for example to discover that certain sounds are used always in a certain context. Recognition of sound can evoke a response from the infant, such as reacting to its name. The correlations, sound used in situations, are stored in the brain and could be seen in its entirety as Embodied grammar, a term further explained in 6.1 and 6.2. It is arguably the childs'discovery of structural use of sounds (language) by a language community that could mark its start of intention reading.

While Piaget's Sensor-Motor stage in Table 4-1 describes many generic aspects of cognitive development, (Tomasello, 2003) describes aspects specific to that of language development. In Fig. 4-2, a compilation of milestones has been put together that occur in the passive language learning stage.

Passive language learning related development milestones (Tomasello)	
4-5 months	<ul style="list-style-type: none"> • Concept learning starts of simple objects and events • Word-like sound patterns learning
9-12 months	<ul style="list-style-type: none"> • Triadic insight between object, adult and self • Gaze following • Social referencing • Imitative learning • Intention reading of actions
Around 12 months	<ul style="list-style-type: none"> • Joint-attention engagements • Communicative intention reading (special case of intention reading of actions) • Cultural learning (role reversal imitation) • Word learning starts • Learning early gestures: <ul style="list-style-type: none"> ○ Ritualizations. Behavior to get things done (often imperative motive) ○ Deictics. Behavior to direct attention such as showing, pointing (often declarative motive) ○ Symbolic. Behavior which refers to a property such as sniffing like a dog or arms wide like a plane. Symbolic learning is the foundation for signed language used by deaf children (often informative motive)

Fig. 4-2 Milestones of passive language learning related developments [Tomasello, 2003]

4.3.2 Active language learning

To reach a level where speech production is used to learn and understand language, obviously requires the speech organ to develop. After the discovery of its own voice at 2-3 months (Babyinfo.nl, 2010), the infant will deliberately test it, a stage called *cooing*. Other uses are for requesting attention, an evolutionary necessity since the infant is dependent on caretakers. Around 4-5 months the infant reaches a stage called *babbling*, where short repeated sounds (such as “ba da ba”) of the mother language are made, much similar to phonemes. After 7-8 months most babies can pronounce 2 syllable fantasy words (Babyinfo.nl, 2010). The infant tries fanatically with its newly acquired verbal skills, even appearing to use it intentional but in this stage it is often used for practice purposes. Finally, at around 12 months intentionality comes to play and the infant starts to make use of the initial form of Embodied grammar; the holophrase grammar, clear single “words” with meaning. Often accompanied with these phrases is body language such as pointing. It is truly a feat of nature to acquire an arsenal of language at the age of 1 year, even confusing parents with its multiple communicative meanings (Tomasello, 2003). At its first birthday phrases can be understood, but often not produced, such as “grab the ball” and “see hello to grandma”. Soon after its first birthday, grammar evolves and phrases are split in multiple parts. After about 24 months the average infant has acquired a vocabulary of 250 words (Tomasello, 2003), see Table 4-2 for an overview. The rate of learning depends on many factors, but is on average one new word every one to two waking hours during the first 5 years (Tomasello, 2003). One of the external growth stimulators influencing that rate are the amount of dialogues and feedback. After 24 months the average infant is capable of performing several speech acts and verbal intentions¹³ (Owens, 2007) which is presented in Table 4-3. A small checklist of normal language development at that stage can be found in Appendix 1.

Table 4-2 English vocabulary by most 2-year-olds in a play school (Bloom, Tinker, & Margulis, 1993)

Food	Toys	Misc.	People	Sound	Emotion	Action	Prepositions	Demonstratives	Social
apple	cow	shoe	girl	woof	yum	go	down	this	yes
juice	ball	eye	baby	moo	whee	get	up	there	no more
bottle	bead	door	daddy	choochoo	uhoh	sit	out	here	more
spoon	truck		mommy	boom	oh	open	in	that	bye
banana	hammer		boy				on		hi
cookie	box								no
	horse								

Table 4-3 Early verbal intentions and speech acts at 24 months (Owens, 2007)

Broad Pragmatic Categories (Wells, 1985)	Primitive Speech Acts (PSAs) (Dore, 1974)	Early Verbal Intentions (Owens, 1978; Wells, 1985)	Examples
Control	Requesting action Protesting	Wanting demands	Cookie (Reach)
		Direct request/commanding	Help (Hand object to or struggle)
		Protesting	No (Push away or uncooperative)
Representational	Requesting answer Labeling Answering	Content questioning	Wassat? (Point)
		Naming/labeling	Doggie (Point)
		Statement/declaring	Eat(Commenting on dog barking)

¹³ more general: dialogue acts, see (Jurafsky & Martin, 2000)

		Answering	Horsie (in response to question)
		Reply	Eat (in response to "The doggie's hungry")
Expressive		Exclaiming	Squeal when picked up
		Verbal accompaniment to action	Uh-oh (With spill)
		Expressing state or attitude	Tired
Social	Greetings	Greeting/farewell	Hi Bye-bye
Tutorial	Repeating/practicing	Repeating/practicing	Cookie, cookie, cookie
Procedural	Calling	Calling	Mommy

“Because we do not understand the brain very well we are constantly tempted to use the latest technology as a model for trying to understand it. In my childhood we were always assured that the brain was a telephone switchboard. (‘What else could it be?’) I was amused to see that Sherrington, the great British neuroscientist, thought that the brain worked like a telegraph system. Freud often compared the brain to hydraulic and electro-magnetic systems. Leibniz compared it to a mill, and I am told some of the ancient Greeks thought the brain functions like a catapult. At present, obviously, the metaphor is the digital computer.”

John R. Searle ⁵

5 The human NLU system: Components in the brain

In which the human brain is discussed, its general anatomy and its NLU related brain structures. Furthermore attention is paid to the mechanisms of one of the most important functions of the brain, the memory, including how knowledge is stored and retrieved such that language can be learned and used.

In the previous chapters important aspects of the development of mind and body were discussed, the general cognitive development that leads to language understanding. Although only scratching the surface, the first phase and second phase of language understanding consist of passive and active language understanding. This means for a toddler to understand language that the sensors are used as input, the brain as a processing plant and the voice box as output. If assumed that the sensors are a collection of dumb binary input monitors and the voice box as a collection of simple binary output devices then all the processing is done in the brain. Of course these assumptions are a bit simple since some sensors (eg. knee-jerk reflex) have local processing independent of the brain, but in general the same signal is sent from the brain to the output neurons every time it is activated. Variation in intensity is determined in the brain by the amount of neighbouring sensors also activated at the same time frame.

The brain is responsible for many other cognitive functions besides understanding language, so pinpointing and isolating mechanisms only related to understanding language is not trivial. This chapter will start with the general anatomy of the brain, from its mass of neurons and fibres down to individual neurons and their chemical workings (flow of neurotransmitters). The next subsection discusses specific brain structures involved with language; Broca's and Wernicke's area. Special attention is given to trauma's to these areas. Finally, one of the most important functions of the brain for language understanding and many other cognitive functions is discussed, the memory. The discussion ranges from how memory is facilitated by the brain to what is involved with storing and retrieving memories from the brain. In chapter 6 the isolated

¹⁴ And even this metaphor is outdated according to Olaf Sporns (Sporns, 2010), stating that the view of *the brain as a series of components* is too much of an abstraction and viewing *the brain as a dynamic network* puts emphasis on the connections between components since not a single neuron causes cognition but a network of them does. This view also allows better discussion on neural plasticity, the recovery of a damaged network.

knowledge on brain structures and functionality discussed here is connected to form a theory on how human language understanding can be enabled.

5.1 Anatomy of the brain

This section will discuss in general the anatomy of the brain starting from the brain at large and then zoom in until the level of neurons.

5.1.1 The brain at large

The brain is one of the or maybe the most important organ of the body, it regulates all life support functions including sleep and breathing. Without a heart you would die, but without a brain you cannot live either. The brain has roughly two types of cells, neurons and glial cells, both in roughly the same quantity. The distribution of the glial cells are 4 to 1 more present in the cortex (Kandel, Schwartz, & Jessell, 2000). The adult brain consist of 100 billion neurons and has over 150.000 km of myelinated nerve fibres (Pakkenberg & Gundersen, 1997) (Pakkenberg, et al., 2003) connecting all the neurons. A child brain weights about 0.3-0.4 kg while an adult brain weights between 1.0 and 1.5 kg (Pakkenberg & Gundersen, 1997). More interestingly it occupies 2 per cent of the mass of an average 68 kg human but consumes 20% of our oxygen and 25% our our energy needs (Elsevier, 2011). Although the brain consist of only two cell high level celltypes many different brain structures exist such as the Thalamus (involved as the main signal router), the Hypocampus (involved with explicit memory), the Cerebellum (the small brains involved with implicit memory and coordination of fine motor skills), the Corpus Callosum (the bridge between the two brain hemispheres) and others.

Computationally, the brain is one big parallel asynchroneous system. Continously changing its mental state (activated regions) and have some experiences related to behavior in the present while others are related the future (using simulation). Nowadays various imaging techinques exist to track the workings of the brain including fMRI, PET, MEG and DTI.

The brain appears symmetrical, but actually is both functional and physical asymetric. The existance of Broca and Wernicke's area is proof for the functional part. The physical part is proven by amongst others the distribution of gray and white matter, with the left hemisphere containing more gray matter then the right (Carter, 2004). The distribution of neurotransmitters is also asymteric having more dopamine (reward) neurotransmitters on the left and more norepinephrine (excitement, attention) transmitters on the right (Carter, 2004).

The Cerebrum (the large brain) compared to the Cerebellum can be divided in four lobes[ref] (see Fig. 5-1). Each lobe is thought to be reponsible for some core functionality, but boundaries are not clear, are changing due to neural plasticity and functionality overlaps frequently:

- **Frontal lobe.** Responsible for *parts-of-speech* (word categorization, eg. verb, noun, etc.), *reasoning*, planning, problem solving, emotions and movement.
- **Parietal lobe.** Responsible for *recognition*, orientation, *perception of stimuli*, movement.
- **Temporal lobe.** Responsible for *recognition of auditory stimuli*, *explicit memory* and *speech*.
- **Occipital lobe.** Repponsible for visual processing.

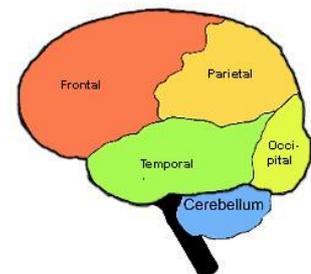


Fig. 5-1 The four lobes of the cerebrum visualised. Note that the cerebellum is a physically separate structure(image wikipedia)

The text in italic mentions a non-exhaustive list of functionality important for the NLU problem.

5.1.2 Cerebral cortex

The cerebral cortex (*cortex* is Latin: *shell*) is the outer layer of the brain see (fig. 5-2). It almost entirely consist of gray matter. The gray matter consist of 6 dense, complexly stacked layers of neurons, each of which has a different composition in terms of neurons and connectivity. The cortex is about 2-4 mm thick. Right under the cortex is white matter that consist of dense layers of myelinated nerve fibres (bundles of *axons and glial cells only, no neurons*) and connects the different gray matter areas. Myelin is a sort of insulationsheet that form around the nerve fibres and enables long distance information (electrical signal) transport. One source (Braitenberg & Schulz, 1991) claims that the subcortex axon connections (direct connection to the neighbours) are rare, most (99%) consist of connections to different cortex areas. On a side note, the thickness of the cortex or neural density is often used as a rough measure of intelligence, but research show that other aspects are just as or more important such as neural efficiency and neural activity depending on the task (Neubauer & Fink, 2009).

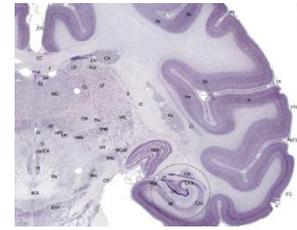


Fig. 5-2 The cortex is composed of gray matter visualized as dark purple. Note that some areas of gray matter exist outside the cortex. (image wikipedia)

The shape of the cortex is peculiar, not only being being visually near symmetrical but also having many folds or grooves, called sulci and wrinkles or ridges called gyri in different patterns. As can be seen in fig. 5-3, other mammals also show this structure, but for example rodents have a smooth brain surface instead of one with sulci and gyri. The folds in the brain can be explained with evolution, they are to increase the cortex surface having more area with gray matter and thus allowing more computations within the limited dimensions of the cranium (Toro, et al., 2008). It is assumed that this enhanced surface (more than 2/3th buried in the grooves) allowed humans to evolve new cognitive skills such as working memory, speech and language (Jarvis, et al., 2005). It is interesting to notice that identical twins at the embryo stage develop different sulci and gyri brain structures, so a physcially and thus functionally different cortex while having the same genetic information (Swaab, 2010). Also interesting is that the occipital and frontal lobe evolved considerably (resulting in more cortex area) compared to other mammals of equivalent body size (Johanson, 1996).



Fig. 5-3 Various brains from left to right: wild pig, bottlenose dolphin, human (image from Museum of Natural History Stuttgart)

5.1.3 Neurons

It is the neurons that play the most important role, together with the signal molecules, the neurotransmitters. Many differences can be found between neurons, for example its sensitivity to stimuli (action potentials, regulated with a threshold). Another difference is the amount of inbound connections, some neurons can receive as many as 10,000+ connections (in the form of dendrites) from other neurons (Feldman J. , 2006). The gap between neurons is called the synaptic gap or synaptic cleft (see fig. 5-4). With all these differences and more, neurons can be classified in at least three broad categories:

- Structural. Classify by shape or by amount of dendrites and axon.
- Functional. Classify by direction, by interaction with other neurons or by electrical discharge patterns.
- Neurotransmitter production. Classify by the type of neurotransmitter they release.

For the purpose of this overview I use the simplest classification; classify by direction. Three types of neurons can then be identified:

- **Sensory neurons** (afferent nerves). They carry information *from* sense receptors (nerves that help see, smell, hear, taste and feel) *to* the CNS
- **Motor neurons** (efferent nerves). They carry information *from* CNS *to* effectors (muscles or glands)
- **Interneurons or association/relay neurons**. They connect sensory and motor neurons.

Each neuron also has a support cell called a (neuro)glial cell discovered by Rudolf Virchow, for example the *Swann cell* in the Peripheral Nervous System and *astrocytes* in the Central Nervous System. The amount of glial cells is roughly equal to neurons in the brain but the distribution varies largely. In grey matter of the cerebral cortex the ratio is about 4 support cells for 1 neuron cells, in the cerebellar cortex about 1 for 5 neurons (Azevedo, et al., 2009). Different type of glial cells exist supporting at least 6 functions include neurotransmission, protection of neurons by surrounding them and keeping them in place, supply nutrients like oxygen, insulate one neuron from another, repair broken axon connections (*synaptic plasticity*) and remove dead neurons. The main difference between neurons is that glial cells don't have dendrites and axons and cannot generate action potentials. With respect to size, glial cells make up 90% of the brain tissue [Allen & Barres, 2009]. It is believed this ratio is at the basis of the myth "we only use 10% of our brain", which has been proven to be false by (Beyerstein, 1999).

5.1.4 Training neurons

At birth the infant brain is basically an incomplete organ, having too many neurons for reasons unknown. One explanation could be that of bootstrapping the entire neural network since neurons also activate their neighbouring neurons and so train them a bit using a spread mechanism. Starting around age 3 the amount of neuron and with that neural connections decrease (a process called *synaptic pruning*), while the neuron support cells (glial cells) increase. An evolutionary explanation could be to increase brain efficiency (less connections to evaluate). The brain has a network with redundant connections and rerouting mechanisms, explaining why connection loss (by natural cause or otherwise) can and sometimes cannot be accommodated - this research area is called neuroplasticity. A fairly recent discovery is the plastic brain, proofing that *neurogenesis* (new neuron creation) is possible at all ages indicating that at least the majority of the brain is not hardwired, not at birth and not after a certain age. This is great news for those wanting to learn a new language or need rerouting - a neural bypass - due to brain damage caused by eg. a TIA, unfortunately the speed of regeneration decreases at higher age.

Neuron training is assumed to rely on two basic learning rules (Bliss & Collingridge, 1993):

- **Strengthen only the synapses from the neurons that are active** (neurons that fire together, wire together" and "neurons that fire apart, wire apart). Especially in the early years connections that are not physically connected but are very near each other and fire together tend to grow to each other. In neuroscience this process of positive feedback is called *long-term potentiation* (LTP).
- **Unused connections weaken over time**. If synapses would only be strengthened (by positive feedback), they would become static at some point and don't react or constantly react when excited. In neuroscience this process of negative feedback is called *long-term depression* (LTD).

On the chemical (neurotransmitter) level much is still unknown how these support these two learning rules. The choice of which connections to activate is the heart of the learning problem. This problem is also known as the *correlation problem*. It explains why so many learning theories (see Appendix 2) have proven to work, since it is highly flexible. One neural mechanism to steer the flow of traffic is the activation level of neurons. Another part of the learning problem is the amount of strength to add to a connection as well as changing the activation level of the neuron. Research is needed to find if connections who have been activated the same amount of time receive identical strength additions.

Special attention is needed for sleeping – a critical activity consuming on average 1/3th of the day– and its relation to learning. Neurological research by (Buccino, et al., 2001) show that mental simulation of an activity and actual doing it use the same neural circuits and involve mirror neurons (Rizzolatti & Craighero, 2004). The brainstem (*locus coeruleus*) blocks execution, for example during sleep. Damage to it results in phenomena like sleepwalking. Recent investigation to other animals like cats show that they suffer from the same phenomena if their brainstem is damaged, chasing imaginary mice during deep sleep. Dreaming (the REM-part during sleep) is not a simple byproduct of sleeping and may very much lie at the core of learning, involving training and connecting neural circuits used earlier during the day.

Solely training the neurons (ie. learning through simulation in the mind) will result in a misaligned mind-body. Learning a skill requires neurological and physical training. There are three types of muscles of which only one type the human can control consciously: the skeletal type of muscle (+/- 40% of all muscles). These muscles are connected to tendons that connects to the bone. The muscle itself contains the end points of the axon and dendrites of two types of motorneurons, alpha and gamma (which are physically located in grey matter at the CNS). Both types of motorneurons (the neurological part) and the muscle fibres (physical part) need to be trained to command precise control of the muscle, ie. acquire motor skill. Some neural circuits like the knee-jerk reflex do not need to be trained, since they are hardwired. For instance, during walking, the knee-jerk reflex is active because incorrect placement of the full body mass above the knee can damage the joint. These circuits evolved to operate independent of the brain.

5.1.5 The central nervous system and information flow

Gray matter areas in the brain are the only parts that are able to calculate, most of it is in the cortex but there are also smaller but important patches under the cortex and in the spinal cord. One such patch is the Thalamus and is believed to be router or switchboard station of signals (Sherman, 2006). Information flows to and from the Thalamus via white matter to the cortex or for instance to the Cerebellum to activate some motor neurons. Using Diffusion Tensor Imaging (DTI) the nerve fibres (white matter) of the central nervous system can be visualised. Fig. 5-5 present the average connections of a group of people (not just a single individual), which interestingly shows similarity in their underlying connections. The author, Owen Philips¹⁵, mentions that while underlying connections are similar, each brain is uniquely wired, probably giving rise to the variation in human behavior.

The body at large has many sensors and parts that need to be instructed and are relatively distant from the brain. Evolution created a “highway” in the form of a spinal cord and together with the brain its called the central nervous system (CNS). The CNS differs from the peripheral nervous system (PNS), in that they only connect the CNS to the limbs and organs and don’t compute much (with of course some exceptions, such as the knee-jerk reflex). At the top of the spinal cord is a structure called the brain stem and on top of that is the Thalamus. The Brain

¹⁵ <https://plus.google.com/101010574579417348339/posts/2BhNg8yD5UL> (last retrieved, june 2012)

Stem is of particular importance, not only is it the oldest part of the brain, it also responsible for vital life function such as sleep and breathing. In addition it is also part of the route (see Fig. 5-4) from the ear towards the auditory cortex (the temporal lobe) which is important for language processing.

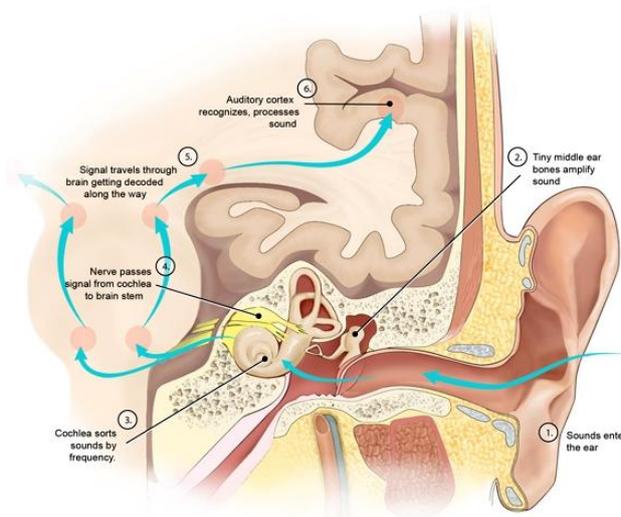


Fig. 5-4 *The information flow of sound entering the ear towards being processed by the auditory cortex [Image source unknown]*

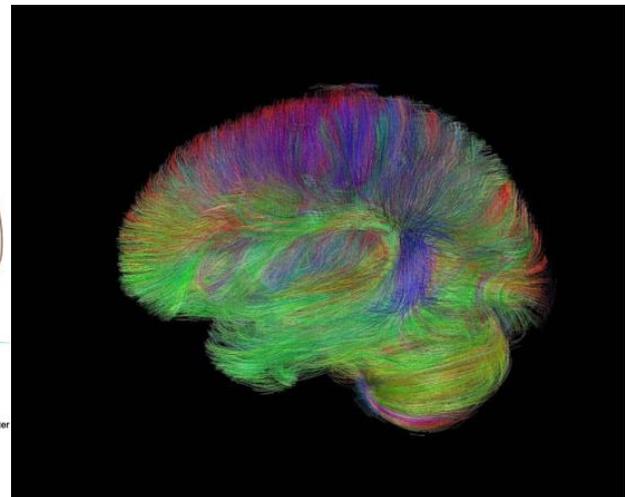


Fig. 5-5 *Fibre connections (white matter) averaged from a group of people [Image from Owen Philips].*

In classical chemical transmission neurons flow in the following way¹⁶:

1. If enough (larger than threshold) positive ions entered the neuron cell body it fires an action potential (electrical signal, so contains the ions) downwards the axon until it reaches the end (presynaptic neuron). This “if above threshold” mechanism is not a simple voting mechanism, but is rather complex weighted (some have larger vote), non-complete (not all synapses “vote”) and time restricted voting mechanism (Feldman J. , 2006).
2. This triggers the release of neurotransmitters (signal molecules) in the synaptic cleft. The amount and type depend on the type of neuron (see Dale’s principle). The neurotransmitters function as keys that bind to the slots (receptors on ion channels) at the postsynaptic neuron across. These neurotransmitters open the ion channels.
3. With the ion channels across open, the ions from the presynaptic neuron can flow across. This ion charge can potentially trigger the other neuron cell body to also fire and repeat these steps.

This process can repeat itself (the refractory period or the calculation speed of the neuron) in about 1 ms for sodium ion action potentials. See Fig. 5-6 for a visual description of the process.

¹⁶ Mind project, 2008

http://www.mind.ilstu.edu/curriculum/neurons_intro/flash_chemical.php?modGUI=232&compGUI=1827&itemGUI=3159 (last retrieved june 2012)

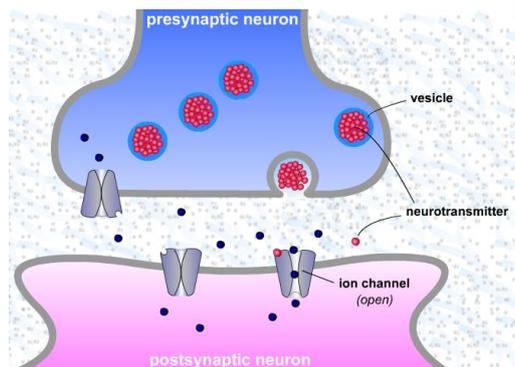


Fig. 5-6 Chemical transmission between two neurons. Visible are the neurotransmitters opening up a channel to let an action potential consisting of ions (blue) pass through from axon (presynaptic neuron) to synapse (postsynaptic neuron).

Discussion

On encoding structure in the synapses

This last part is a discussion on the efficiency of the brain. How is experience or the patterns of stimuli stored in the brain on the neural level? Three theories exist (Feldman J. , 2006): the grandmother cells (no distribution), the hologram cells (full distribution) or partial distribution cells. The grandmothers cells are called after the concept that everything about your grandmother is stored in a single cell. This theory is provable wrong since the human brain doesn't have enough neurons to store all the worlds concepts. The other extreme, all knowledge is distributed over all neurons, is also provable wrong, fMRI scan show that some areas light up when thinking for example about your grandmother, but not all areas. The third theory is more reasonable. Neurons are highly efficient by allowing some degree of overlap, clustering patterns that relate, meaning that although it seems humans have infinite amount of storage space, our brain is actually limited. Quite likely evolution invented the forgetting factor to allow reuse of freed neurons or in terms of software engineering – the garbage collector.

At this last paragraph of this subsection about the structure of the brain I would like to pause. It is simply amazing that neurons, special cells, are basically tiny biological life forms with their own development, desires, needs and problems. It is amazing that biochemistry principles like “matching (neurotransmitters fitting as a key on the lock of a channel)” and “shape changing(channel opening and closing)” (Feldman J. , 2006) allow micro behavior that in turn allow macro behavior: cognition, memory, consciousness and language understanding.

5.2 Functional structures in the brain

There is a long history of scientists trying to pinpoint behavior to specific brain locations. Although Aristoteles already predicted a relation, it was Gall who conceived the first mapping “science” Phrenology (see Fig. 5-9), popular in the early 19th century and mapped various behavior to areas on the skull. His approach was much inspiring for neuroscience and psychiatry at the time. A century later this mapping proved to be wrong and using carefully constructed experiments and better technology new maps (see Fig. 5-7, 5-8, 5-10) where developed. To date these maps have proven never to be 100% accurate or even close to it and due to the brain plasticity principle quite likely never will be. Interestingly the maps do show many similarities, showing empirical proof from different sources. The pursuit is not a definitive map, but one that is the general rule.

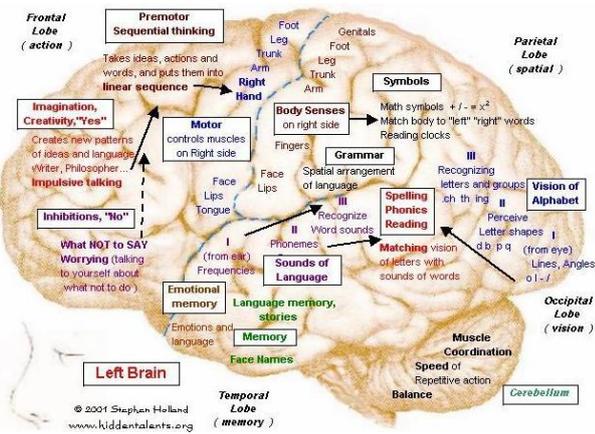


Fig. 5-7 Functional chart in English from 1983 (left hemisphere) (Image from Stephan Holland)

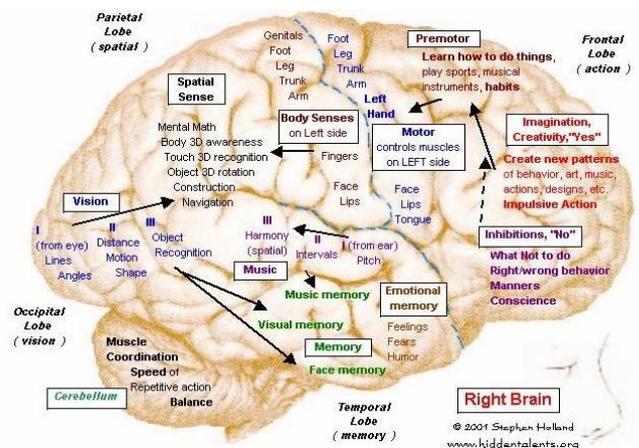


Fig. 5-8 Functional chart in English from 1983 (right hemisphere) (Image from Stephan Holland)



Fig. 5-9 Phrenology chart from 1883 (people's encyclopedia of universal knowledge, 1883)

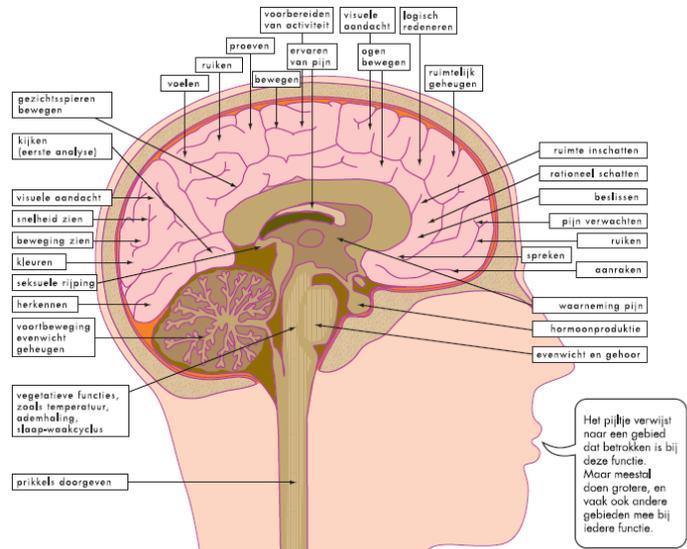


Fig. 5-10 Functional chart in Dutch from 2011 (Hersenstichting Nederland)

In relation to speech and language processing and production two of the general rules (brain structures) are Broca's area and Wernicke's area (see Fig. 5-11).

5.2.1 Broca's area

This area is mostly involved with speech production. Paul Broca discovered after autopsy that two patients having speech problems both had injuries to their *inferior frontal gyrus* of the brain. Since 9 out of 10 people were right handed (and still are (Hardyck & Petrinovich, 1977)) and Broca's two patients were right handed, Broca cautiously spoke the famous words "we speak with our left hemisphere" (Broca, 1865, p. 494).

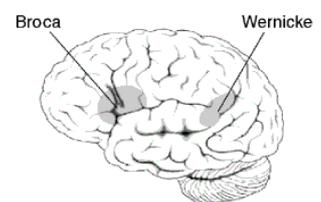


Fig. 5-11 Location of Broca's and Wernicke's area

Speech production was not completely halted, one patient was able to speak one word and the other 5 words. In a recent study by (Dronkers, Plaisant, Iba-Zizen, & Cabanis, 2007) the original dataset of Broca was reinspected using a high resolution volumetric MRI system. They concluded that speech impairment cannot be exclusively caused by lesions to Broca's area, other damaged areas must have contributed as well.

Sudden removal of Broca's area after a slow growing tumor does not result in speech impairment (Plaza, Gatignol, Leroy, & Duffau, 2007). This amazing result again proves the neural plasticity of the brain. Although the patient was unable to produce complex syntactic constructions with more than two subjects, he had a full recovery. (Plaza, Gatignol, Leroy, & Duffau, 2007) suggest that full recovery was possible due to neural functionality that had been transferred to neighbouring regions around Broca's area and that the syntax problems were caused by working memory problems.

It was later discovered that Broca's area is also involved in speech comprehension (Caplan, 2006), but is not as important as Wernicke's area in this task. Some videos of patients can be found at the website¹⁷ from the authors of Auditory Neuroscience (Schnupp, Nelken, & King, 2011). One patient (Sarah Scott) shows that pronouncing words can be supported by writing them down first, indicating a neural linking problem. fMRI and PET scans show that seeing words involve different brain areas than hearing words (Calvin & Ojemann, 1994), suggesting that the word stored in memory itself is not damaged but the route to it might be. Some evidence suggest that ASL speakers (sign language) suffering from damage to Broca's area show the same syntactical disabilities in sign language production (Lebrun, 1985).

5.2.2 Wernicke's area

This area is mostly involved with speech comprehension, understanding written and spoken language. Patients diagnosed with problems with this area show fluentlike speech patterns but often with imaginary words or improper used words, possibly (phonologic or semantic) related to the correct one. One patient shows that his social abilities during speech remain intact, using laughter and body language to support misunderstanding (Schnupp, Nelken, & King, 2011). While patients with Broca's malfunctioning are aware of their handicap, Wernicke's are not. They are mostly completely unaware of their own utterances and consequences such as that other people don't understand them, as demonstrated in the video of Schnupp. Often they get annoyed because of this constant miscommunication.

Interestingly, problems with speech comprehension are not restricted to human speech, but extend to understanding animal and environmental sounds as well (Saygin, Dick, Wilson, Dronkers, & Bates, 2003). This suggests problems with processing the auditory information in the patterns it was originally learned in, since some words can still be understood and used well while most others cannot. Again, this could be due to neural linking (routing) problems.

Most of the auditory problems fall under the umbrella of Auditory Processing Disorder (APD). A more specific term is Aphasia, used for language ability impairment. Wernicke's patients are often diagnosed with *receptive aphasia*, while Broca's patients are often called *expressive aphasia* patients.

Discussion

On innate brain structures

Are humans born with no neural structure at all? Are humans truly born with no experience, a tabula rasa? That mostly depends on the definition. If focussed on behavior, then most scientists agree that this is not the case. For example, from studying seperated twins it shows that their genetic heritage influences various personal traits such as IQ more or less identically (Deary, Spinath, & Bates, 2006) . However if focussed on world knowledge then being born as a blank slate, so having no previous knowledge of this world or reality, is pretty sound – a child has no memory of a possible past life or has a

¹⁷ <https://mustelid.physiol.ox.ac.uk/drupal/?q=vocalization> [last visited: december 2011]

vocabulary with certain words of any kind. (Rakic, 1988) provides evidence that organisms evolve to adapt to the environment. For a human that is able to adapt in a single generation, this means that being born a blank slate is predictable since any fixed information could cause slower or no adaptation. For the human brain this means that a certain neuron layout and type of neurons adapt better to its environment by reacting better to certain stimuli compared to others – the basis of pattern recognition. This notion is not farfetched if considering neurons to be in essence complex signal discretizers, transforming continuous signals from the world to sampled versions of it. The sampled versions are linked both spatially and temporally and can be considered chunks. Hence patterns are born: if chunk A activates, then chunk B does as well since it is linked. Considering the whole, this leads to promoting a blank slate of knowledge AND accepting innate functionality (fixed brain mechanics). Individuals have personal traits and sensitivity for language due to these innate brain mechanics which are basically evolved pattern recognition skills.

*One of the most heated arguments between two schools of language researchers (Chomsky/Pinker vs. Lakoff) discussed earlier was on the innate ability for grammar and the product of that ability, the universal grammar. Since syntax abnormalities arise when Broca's area is damaged, the hypothesized grammar faculty of Chomsky is at least partially related to Broca's area. However the same innate ability used to learn grammar (patterns in the audio stream) is also used to learn other innate skills, for example to recognize faces (patterns in the visual stream). The innate ability is like mentioned above, "just" generic pattern recognition skills. The formation of a cluster of language related functions ("the language faculty"), is not caused solely by human DNA, by nature, it is also caused by the environment, by nurture. The reason is the same as for universal concepts. The embodied view offers an evolutionary explanation for both accounts: the body including the brain evolves to better survive in its environment. The environment has various laws such as Physics (like gravity) and Chemistry (like conservation of mass). Organisms are constantly exposed to these laws, influencing their evolutionary development. Water, a critical component for humans (by mass, human cells consist of 65% water (Freitas, 1999)) is influenced by these laws. In terms of a plant metaphor, human biology provides the seeds (neurons), soil (organs, blood vessels, etc) and bucket (body) from which a cluster of language related functions ("the language faculty") can grow. The seeds are planted **at random**, but due to the environment the seeds at the surface are provided with stimuli that influences the direction of growth and growing conditions such that the seeds can hatch and become the language "plants". Just like universal concepts it is likely that there exist universal language patterns that could be viewed as universal grammar. But again this is not due to nature's innateness, it is due to nature interacting with nurture.*

On free will

The Dutch neuroscientist Dick Swaab states in (Swaab, 2010) that behavior is only the result from the laws of physics and chemistry operating in the brain. This arguments regarding the causal link between human behavior and its environment leaves no space for the concept of "free will" or "soul". The deterministic philosopher Daniel Dennett (Dennett, Freedom Evolves, 2003) agrees with this view stating that free will is indeed not possible on the physical level and strictly it is thus void on all levels. He nuances however, that if a human is able to make a choice without feeling pressure he considers it free will using the terms "inevitable" and "evitable" and the subjective concept of feeling to explain the distinction. It should be noted that the field of quantum mechanics introduced the opposite of this view, stating that quantum indeterminacy paves the way for randomness and thus free will. However complete randomness wouldn't allow humans to plan either, so that wouldn't be possible either. Partial (in)determinacy would allow free will, but theories to

unify such as “global/cosmic consciousness” are still farfetched or not falsifiable. The truth, the existence of free will or not is quite likely only of philosophical interest since humanity lived generations with belief of free will – and is perfectly happy with that.

The existence and/or details of the concepts of function localisation and free will will not be further explored in this report, but was relevant to discuss in short since it influences the design of an NLU system (see chapter 7). Still the idea that the laws of physics and chemistry (nurture, the environment) in combination with our body (nature, the container) are possibly at the heart of function localisation and universal concepts is interesting. Just as interesting is that brainplasticity is likely the bodies’ mechanism for function localisation (“placing the seeds”) and recovery of brain traumas. Various cases (Smith & Sugar, 1975), (McClelland & Maxwell, 2007) report that near full recovery is possibly (including personality, intellect, language and motor skills) in children and more recently also select adults who had half their brain removed (hemispherectomy involving the cutting of the corpus callosum, corpus callosotomy). The brain structures needed for language, personality, etc, “just” grew at other places. Quite possibly this new layout of structures is the optimal CNS->PNS infrastructure given the new brain and its environment. Nature’s will to adapt to any environment with new physical conditions is nothing short of amazing.

5.3 Functionality of the brain: Memory

The brain with its structured layers of neurons enable a multitude of cognitive functions such as language understanding, perception, awareness, creativity, imagination, belief and reasoning – all functions involving thinking - but one of them could be considered the *great tree* from which its three branches enables all these cognitive functions. This concept is *memory* with its three core functions *encoding, storage, and retrieval*. Without memory, input (sense) and output (behavior) would have a direct connection which would make the person or system not much adaptable for changing environments, nor an efficient person/system. The delay between input and output that is caused by a network of interneurons and other chemical processes enables processing opportunities. One hypothesis is that by delaying a signal, the temporal dimension can be exploited such that a concept is not mapped to a single “grandmother” cell or fully distributed as discussed in 5.1, but form a small pattern or unique neural circuit. In a sense that pattern between input and output could be called the memory of the system, which in the case of humans is both highly adaptable to changing environments and efficient as noted in the discussion on neurons. Using the tree analogy, the human memory with storage (= learning) and retrieval (= recall) is placed at the root of human intelligence. It enables prediction in changing environments and thus understanding of the environment. It also enables efficient usage of the knowledge in those environments; human behavior. Much is already known about memory, but far more awaits to be (re)discovered. A complete description enables more efficient access and retrieval strategies, which might for example revolutionalize our educational system. Such a discovery would surely win a Nobel price.

From a distance, memory can be viewed as the classic three tier structure (Atkinson & Shiffrin, 1986) composed of *long term memory, short term memory* and *extremely short term (sensorial) memory*. A separate notion is for *working memory* which is more of a conceptual support framework indicating a changing set of accessible (consciously or subconsciously)

memory (see Fig. 5-12). Reality is sampled by the five senses and transformed from experience into knowledge and stored into memory. The route with the short term memory between sensory and long term

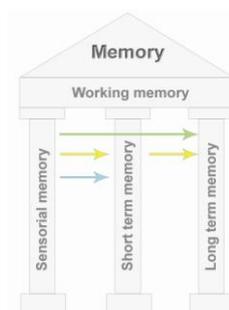


Fig. 5-12 Model of memory with arrows representing flows (always via working memory)

memory is crucial for filtering out irrelevant or redundant information. Some disorders include the shortcut route from sensorial to long term (*hyperthermesia*) or the inability to store knowledge in long term memory. Language related disorders include *nominal aphasia* (a type of expressive aphasia) known for its tip-of-the-tongue phenomenon due to damage to the frontal and/or parietal lobe(s). In this particular disorder research by (Fridriksson, et al., 2010) suggests that the memory of a word (including form/meaning, so the understanding of it) is not damaged but the route to retrieve it is. The damage is not in Broca's or Wernicke's area (grey matter area's) but in white matter area's near it.

It should be noted that the three tier structure is again, not without debate. Research by (Cowan, 1995), (Oberauer, 2002), (Tarnow, 2008) suggest that there is no separate system for short and long term, reducing it to a two tier system; sensory and long term memory. This makes sense since recalled experiences such as seeing a ball recalls the experience *from* long term memory, matching it *with* the sensory experience. Otherwise identical experiences are stored at two memory locations in the brain, which is not the case. Short term memory in this two tier model is just a *shortlived* long term memory structure. Despite the three tier model being or not being neurologically correct, based on the many empirical phenomena the three tier model remains useful to discuss behavior. The disorders discussed above certainly justify the distinction between short and long term memory.

5.3.1 Sensorial memory

Sensorial or sensory memory (SM) is a memory system that stores the parsed raw information acquired from the internal or external environment, for less than a second. It is assumed that each sense has a different memory system although only three have been given a proper nomenclature:

- **Iconic memory (vision):** An example is waving a light during night time, the "afterimage effect" remain very shortly. From a somatic (cell) view an explanation of the decay of memory could be that the excited rod and cone (retinal) receptor cells (continuously sending stimuli) slowly return to their default (unexcited) state.
- **Echoic memory (auditory):** This is much similar to the workings of a tuning fork that decays slowly when hit. Another example is still hearing a sound while the source stopped. From a somatic (cell) view the vibrating hair receptor cells would be the ones that are excited and slowly return to their default state.
- **Haptic memory (kinesthetic):** The sensation of pressure, itching or pain after the actual (physical) touch. From an somatic view the somatosensory system is excited and relaxed.

Since the memory is heavily dependent on the chemical ability to return to its default state, factors that influence the cell workings (such as age) influence the SM. Several similarities and differences have been found between the sensory memory systems, and while still under heavy research, four common features have been identified by (Winkler & Nelson, 2005).

- **Attention-Independence.** This implies that SM is outside of at least conscious cognitive control. The stimuli that are acquired unattentively are possibly input data for implicit learning and has a possible evolutionary function to deal with danger that is not attended to.
- **Modality-Specificity.** Each sensor has a dedicated and exclusive memory system.

- **Resolution.** Sensory memory is far more detailed than normal, categorical memory. Unfortunately, no categorical (symbolic) representation is able to represent all the details.
- **Duration.** The stimuli or memory is utterly lost after a very short period of time, since no learning (no structure change) takes place in the sensory cells after it relaxes. It is “fortunate” that the sensory cell only relays the stimuli, otherwise humans would have permanent image retention, a problem old Plasma TV’s used to have.

5.3.2 Short term memory

The short term memory (STM) stores an experience a longer time than sensory memory, but eventually decays as well. (Tarnow, 2008) theorises (simplified) that short term memory, like long term memory involves firing of neurons which deplete a pool of neurotransmitter while firing. After the firing, another chemical process called *endocytosis* clears any trace of the pool being once depleted (“it being used”). Unlike short term memory, long term memory reaches a metastable pattern because the endocytosis process could not clear all traces because the neuron was activated again before it could finish. In a sense, *structure* is created and at that point *long term memory*. The timescale of short term memory according to his theory matches that of the synaptic endocytosis process and is estimated by Tarnow between 3 seconds and 15 minutes. Using the same theory, the time scale for storing an experience into the short term memory equals the opposite chemical process (synaptic exocytosis) which takes about 0.5 to 3 seconds. Another suggestion from Tarnow is that short term memory is not overwritten (since there are no two concepts of memory), instead, other neurons (from other experiences) are activated while the previous slowly deactivate. He assumes that a catalyst speeds up the endocytosis process, possibly when a certain stresspoint is reached.

There are many interesting things to note about the short term memory; the duration and the disorders discussed earlier are only a few of them. In one famous paper by George Miller (1956), he presented his findings on limitations of the human capacity to process information. Miller’s law states that the average human short term memory can hold up to 7 ± 2 objects or chunks. It is later shown that this highly depends on the representation (encoding), ie. 7 for digits, 6 for letters etc. Quite interesting is that Chinese can hold up to 10 digits, suggesting that their representation for digits is a more efficient encoding (Dehaene, 1999). One explanation is that all Chinese number words are single syllables. Using mnemonic tricks the short term memory can be effectively extended. Early findings by (Van den Berg, Shin, Chou, George, & Ma, 2012) show that quite possibly short term memory limitations could be explained better in terms of encoding precision than with Miller’s “magic numbers” model. Another magic number and long lasting hypothesis¹⁸ is the sampling periods of the present, in other words “the duration of experience”. A multitude of animal and human behavior (Nagy, 2011) show that average behavior patterns last about 3 seconds and would suggest that the present is perceived in three second time frames. This number is clearly an average, with wide tails as context is important.

Compared to sensory memory, the short term memory can be accessed consciously since the stimuli are now linked to categories/symbols that activate, however the resolution of this consciously accessible short term memory is much lower than the subconsciously accessible memory. Recent and early findings by (Sligte, Scholte, & Lamme, 2008) underline this by showing that for storing objects in the visual short term memory, three phases can be distinguished. Each of the phases have a lower resolution (or subsampled) version of the original. The phases proceed from first extracting features, then extracting objects based on the

¹⁸ <http://news.sciencemag.org/sciencenow/2011/01/hugs-follow-a-3-second-rule.html>, retrieved june 2012

features and finally assigning attention to some of the extracted objects to get it into consciously accessible memory. It is possible, but yet unknown if this three phase model also works for other types of short term memory (ie. auditory or kinesthetic). On a side note: it is interesting to see how encoding works in visual short term memory, since apparently at least features and objects are encoded separately from the original sensed and perceived visual experience, which can be seen active in different brain areas.

5.3.3 Long term memory

The long term memory (LTM) and its seemingly infinite capacity is truly fascinating. The biological underpinnings of LTM such as *long term potentiation* (see section 5.1.4) and other principles of synaptic plasticity such as the ability to renew neuron cells without losing the structure are still not understood. Solving it could potentially cure for example the infamous memory disease Alzheimer's. The field of long term memory is divided in conscious (explicit) and unconscious (implicit) memory or a mixture. Each type has several subtypes:

- **Explicit memory**
 - Declarative memory:
 - Episodic memory, context dependent knowledge such as memories about events in time and place. The *when* and *where* questions recall episodic memory. Episodic memory is assumed to be a scaffold of semantic memory.
 - Semantic memory, context "independent" or factual knowledge such as knowledge about the world (function of a chair, persons name). It is here where (abstract) concepts are stored. The *who* and *what* questions recall semantic memory.
- **Implicit memory**
 - Procedural memory or skills memory is a type of memory that records complex repetitive experiences such as learning how to walk. The "feeling" for a language (knowing its rhythm, allowable structure such as word sequences, etc) is also part of the procedural memory. It is assumed that implicit memory is active not only for the obvious motor skills but also for *any* cognitive activity. This could explain why problems with accessing declarative memory ("the tip of the tongue") can be supported by motoric behavior (such as jumping, waving in a repetitive pattern or pushing your eyes into a corner).
- **Mixture**
 - Emotional memory, this type of memory is partly conscious and partly not. Emotional memories can be recalled consciously but have many unconscious physical reactions. Emotions are, like the motoric behavior, a scaffold for accessing declarative memory. Using the embodiment theory, knowledge is always acquired in situated and embodied context, explaining this emotionally enhanced memory retention.

5.3.4 Working memory

While the authors of the three tier model (Atkinson & Shiffrin, 1986, p. 92) used WM as a synonym for STM, theorists now use it as conceptual support framework for the three types of memory. While STM is static activation of experience, WM allows manipulation or processing of that experience before it is discarded or put to another memory tier. Important processes that are involved with WM are learning (size of WM, see 5.3.2), awareness (activated items) and attention (selection of activated items).

Awareness

A person (or animal) becomes aware of *something* when the experience is sensed **and** perceived. Perception does not imply that the person is consciously aware of *something*, rather the brain was able to process and understand the *something*, for example being able to categorize it (eg. the *something* becomes a ball). Reasons for a person not being aware of eg. *a ball in a room*, are various, such a sensing and **not** perceiving experiences (ie. not able to proces and understand *the ball*, both consciously and unconsciously) or a sensed and perceived experience that has been dropped out of working memory due to forgetting or other reasons (eg. *not used the ball for ages, while the other present toys are used every day*).

Self-awareness (the ability to consciously introspect and reflect on ones own behavior, the ego) is something a human child is assumed to gradually learn and is not born with (Demetriou & Kazi, 2001).

Attention

While being aware of a certain context with multiple items, there is simply too much information (stimuli) to attend to, the so called *information overload problem*. The relative limited human sensor capabilities (eyes, ears, etc compared to rest of body) generate huge streams of information. One biological solution is only further processing sensed stimuli that are different compared to the previous state of the sensor neuron, similar states are filtered out[Feldman, X]. However looking at differences still leaves too much information. Of course using prediction of stimuli (see chapter 3.3 – dealing with uncertainty) any amount of information can be processed. Unfortunately, prediction suffers if the WM is full, there is a limit to the amount of information one can be consciously and subconsciously aware of.

One of the tricks biology found to keep important information in WM and thus aid prediction is the principle of *attention*. Attention attacks the information overload problem by assigning relative importance to information, such that only important information remains in WM. The exact working of the weighting of information is unknown, but part is hardcoded (a natural danger reflex) by assigning additional weight when huge changes in stimuli occur(*sensor saliency*). Other parts involve experience, (“I need to watch the first step of an escalator” or “that bear is emotionally important”). Attention can be consciously activated by motivation (“I want to pay attention to that beautiful painting”). Also worth mentioning is that one can be aware of information (due to previous attention to it, eg. my book “Dune” is stored in on my shelf) but not having attention to it currently, with the risk that the information leaves WM or not being accurate (it changed, leading to false believes).

Important tools of attention are *focus and concentration*. Concentration can be considered an effort to keep important information in WM, while focus is an effort to find important information. Focus narrows the information space (“the world”) by consciously and subconsciously considering parts of it. Using focus, the resolution of the information is enhanced (more details become visible) such that better processing (such as recognition, prediction, learning) can be done. The trade-off is of course “tunnel vision”, not being able to see the larger picture/patterns. The biological implementations for visual focus is eye gaze. Tomasello hypothesizes a role for eye movement during language comprehension (Tomasello, 2003, p. 328). Eye gaze is a fine-motoric behavior steered by the oldest part of the brain, a part of the *pons* of the brainstem called the *abducens nucleus*, the sixth cranial nerve. Attention through gaze is one of the three reflexes an infant is born with, next to sucking and the grasping. The brainstem is a set of neurons responsible for vital functions to support survival such as breathing, temperature management and is also responsible for the direction of eye gaze. Besides the use for detecting changes in the environment and possible danger, it is also used for planning ahead often moving just before the rest of the body moves. It is critical for predictions

and goal-driven last minute motoric adjustments (such as escaping using the least risqué route). During communication the eye gaze gives away your direct attention and part of your intention. The eyes are poetically named the “doors to the soul” and thus for a good reason. Eye gaze also plays an important role during deep sleep, the REM (rapid eye movement) sleep and is assumed (controversial) to give away certain communication cues such as lying when right cornering the eye for a right handed persons (accessing the left “logical” part of the brain).

Discussion

This section on memory, awareness and attention is not hard science nor are the definitions generally accepted. With the rise of neuroimaging as hard evidence, solid progress is made. It is interesting to note that there is still much theoretical discussion on the workings of attention such as early selection models vs. late selection models in the “cocktail party problem”. Attention and awareness are closely related since it is not possible for a person to give attention to something he/she is not aware of (ie. predict the unknown is not possible, see Logic in chapter 3). “Something” first needs to be sensed and perceived (consciously or subconsciously) in order to be given attention and thus assign it importance.

One observation that might be important for investigating attention in the cocktail party problem is the robustness in recalling experiences. In a local time reversal and speech comprehension experiment (Saberri & Perrott, 1999) discovered that by changing the temporal dimension of a sound pattern by cutting it up and reversing these chunks, a sound becomes incomprehensible. This might not seem exciting, but what is discovered is that this process is gradual, for example when the temporal dimension is distorted in chunks of 20ms the sound can be comprehended perfectly, see examples¹⁹. A hypothesis why patterns can be recognized even if they are not in correct heard is that sensory memory (this case echoic memory) has a small decay time when activated and until it is inactive, it keeps emitting the stimuli that combined with the other stimuli activate the pattern, regardless of sequence (see voting mechanism, 5.1.5).

Another important subject in the memory framework is (the degree of) belief in a certain memory which is directly related to expecting the memory. Degree of belief could be modelled as activation with a variable decay function. The strength of the decay depends on the available proof in favor or against it, for example observing a gold bar in a safe will lead to the high expectation (strong belief) it will be there when closing the door and opening it again, however after 10 years and then opening it again that belief has shrunk.

¹⁹ https://mustelid.physiol.ox.ac.uk/drupal/?q=vocalization/local_time_reversal (last retrieved, June 2012)

"If a lion could talk, we could not understand him."

Ludwig Wittgenstein

6 The human NLU system: Theories on enabling NLU

In which three embodied theories for enabling NLU are discussed; aquisition, usage and structure. The theories acknowledge and build on the evidence discussed in the previous chapters on cognitive development and brain anatomy, although additional translation steps between the theories and neurological practice is needed. The theory on structure (embodied grammar) is based on the NTL work by (Bergen & Chang, 2003). The theory on aquisition is based on the usage-based theory by (Tomasello M. , 2003). The theory on usage is based on various NTL assumptions by (Feldman J. , 2006).

Historically, many contributions on the workings of the human language processing system including NLU have been made by great linguists such as Chomsky, Pinker, Goldberg, Kay, Fillmore and Lakoff (see references in 6.1). Since explaining the language system from a neurological viewpoint was not possible due to technological limitations (eg. functional brain imaging), the system was explained from a language output (behavioral) point of view (eg. sentence construction analysis, monitoring the reading of difficult grammatical sentences such as *garden path* sentences, etc). The human language system proved to be too hard to capture in a single theory through backwards reasoning, leading to contradicting theories already at an high abstraction level; the existance of the innate language faculty and with that hard wired grammar rules. The NTL theory boldly tries to explain the human language system by approaching the system from both a neurological viewpoint and behavioral viewpoint and working inwards to connect both ends. This is possible due to steady advancements in fields like neuroscience, computer science, evolutionary biology and psychology.

This chapter discusses the NTL theory that hopefully will be flexible enough to translate from a conceptual to a neurological level with future efforts. This fragile hope described by Feldman and his NTL team is made stronger by basing the theory on many abstract assumptions such as Embodiment and concrete assumptions such as "neurons that fire together, wire together" which have wide implications on the formation of language theories. It was the formal linguist Noam Chomsky (1986) who split the linguistic behavior problem between three key areas: its nature (structure), origin (aquisition) and use. Nancy Chang (2008) visualised the *chicken-egg* relations between the areas into a triangle relation (see fig. 6-1). For finding a solution to the NLU problem a theory for each area is needed. Section 6.1 discusses a theory on language structure that is scalable enough to add all the complexities (nuances) of human language including its Form such as emotions, syllable stress, but also view dependences like "from my or his point of view". In section 6.2 the aquisition of language is discussed, the way embodied grammar is formed and how continuous exposure to language leads to more and more understanding of the language used in a community including its ambiguities and correct usage (Type 1 understanding) and effect on others and self (Type 2 and 3 understanding). Finally, in section 6.3 the usage of language is discussed, focussing on language behavior and mentioning the way a human with its creative capabilities and motivation can use the freedom an aquired grammar provides to fulfill and achieve desires and goals.

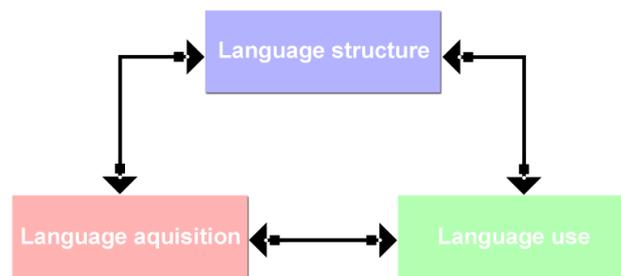


Fig. 6-1 Chomsky's language triangle

6.1 Embodied theory of language structure

How are sensed and perceived stimuli – experience – that is processed into knowledge, stored in the brain? A complex orchestra of neurons fire and leave a permanent impression in the brain. Interestingly this impression does not last forever and has a highly dynamic structure both biological as well as functional. Furthermore, it scales incredibly, continuously building on previous experience to better predict and so survive in our environment.

This section on structure is again not undisputed – the brain is difficult to measure and generalise. The key NTL assumption used is “*language structure emerge from language use*”, implicitly assuming that the neural structure a human is born with is *untrained* and due to exposure and usage becomes *trained*. This assumption is based on the fundamental NTL assumption “neurons that fire together, wire together”, indicating that biology (or in this case synapses and axon growth – “the wiring” - and also neuron behavior) reacts on the environment. Hence language structure follows suit from language use.

In chapter 3 a theory on knowledge decomposition was discussed which in this section will be made more tangible, allowing it to be implemented. Although knowledge decomposition (splitting form and meaning) certainly does not mimic brain mechanics, it is useful for modelling purposes. One of the theories of decomposed knowledge was developed by Charles Fillmore in 1965 (Fillmore, 1988) (1989) together other notable pioneers such as Lakoff (1987), Kay (Kay & Fillmore, 1999), (Langacker R. , 1987), (Goldberg, 1995) and (Croft, 2001), naming this field construction grammar theory or CxG theory. The basic unit of a specific type of knowledge related to language, grammar (although it can be used much more broadly), is named a construction. “*Constructions involve a symbolic relationship between form and meaning*” (Chang, 2008, p. 33). More recently Nancy Chang and Benjamin Bergen (Bergen & Chang, 2003) proposed a version CxG from the position of Embodiment, the Embodied Construction Grammar (ECG). ECGs is a structure that is used both for language acquisition and usage. Eva Mok extended the work by adding probabilistic mechanism for learning ECG (Mok, 2008).

Chang mentions that “*Constructions function as gestalts - that is, each construction is a whole whose behavior is not determined by the behavior of its constituent parts; they thus exemplify the phenomenon studied by the Gestalt school of psychology. Here ‘behavior’ includes both Form and Meaning ...*” (Chang, 2008, p. 33). The connection between Form and Meaning is not consistent as are the patterns that compromise a Form and/or Meaning. This means that the Form, Meaning and the connection are learned and interpreted different for each individual. Chang mentions as an example that “*the sound ‘dog’ has no motivated connection to the concept of dog, conventional in that the relationship is only valid in the language community that it accepts it as such and intentional in that they are used with referential or communicative intent*”.

A nice property of ECGs is its high abstraction level. The Form part is flexible enough to incorporate emotional intonation, word order, word stress, etc. (Feldman J. , 2006, p. 302). These patterns contain information how it is used in context. Although the model allows the full fledged bio inspired system, the model doesn't require it to be incorporated right away, and is able to add extra usage patterns while learning – basically building on existing experience, just like children discovering new usage patterns while developing. With this property Feldman (Feldman J. , 2006, p. 298) states that it might really be possible to capture embodied meaning at the computational level.

The most important structure of the ECG is the schema. Schema theory is widely researched in psychology, most notably by Gestalt psychologists. Mark Johnson, an Embodied philosopher, used several core ideas from Emmanuel Kant's schema theory combined with ideas of cognitive grammars (Langacker R. , 1987) and created *image schemas* (Johnson, 1987). A single image schema is a pattern that is grounded in embodied patterns. How this pattern is found and what bodily features (called *roles* in ECG) and values are selected for a pattern is the main issue of the *correlation problem*. The prefix *image* does not reference to solely visual experience, but to all bodily experience (patterns), so often this prefix is left out. Chang uses schemas for both the Form and Meaning. An example of an construction in ECG is presented in Table 6-1 (Feldman J. , 2006, p. 289).

Table 6-1 Example ECG construction of the word "to"

Construction ID	123
Type	Lexical
Mnemonic	"To", as in going to a place
Subcase of	Spatial Preposition
Evokes	Schema SPG (source-path-goal) referred to as spg
Form	Schema "to"
Meaning	Schema TL (trajector-landmark) having roles (features): <ul style="list-style-type: none"> • landmark <-> spg.goal • trajector <-> spg.trajector

Although the details of the ECG are beyond what can be provided in this small introductory text (see (Bergen & Chang, 2003) or (Feldman J. , 2006) for more details), some details are interesting to emphasize here. First is the "Mnemonic" tag such that humans can easily label a construction and discuss it. Second is the "Subcase of" allowing a hierarchy such that a *child* construction could also be called whenever the *parent* construction is called. Third is the "Evokes" tag, (unique to ECG) and allows related or associated schemas to also activate. This allows deep contextual integration in the form-meaning pair as is visible in the example. Using a best-fit matching mechanism (*the inference engine*) associated schema's can be actually evoked or not and linked or not, leading to different role fillings and thus meanings. Through this mechanism, the actual meaning becomes highly dynamic (=context dependent) since the meaning of a word in context is captured by the joint activity of all of the activated schema's. Activated schema's are schema's that are activated by a single word ("immediates"), evoked ones ("associates") and most importantly by the schema's from the discourse, the neighbouring words in a sentence ("contextuals"). So in short, the final meaning is generated through a linked

network of schema's, called a semantic specification or SemSpec in NTL lingo. The fourth bit of interest visible in the example is the **<-> notation**, used to depict the compatible association between a role from schema A and another role in schema B. In table 6-1 *TL.landmark* is associated with *SPG.goal*, meaning that a value filled in for *TL.landmark*, such as the city Berkeley, is filled in as a goal in *SPG.goal*. Other examples such as end-point ECG constructions, abstract constructions, more complex Form schema's and a complete SemSpec containing all of the above, see (Feldman J. , 2006) for a thorough ECG decomposition of "when Harry strolled to Berkeley".

Up until now, only two of the four abstract building blocks (primitives) of an ECG were discussed; Form and Meaning. The complete set also involves Mappings and Mental Spaces. The full set is essential, for it allows analysis of any statement about other people's thoughts or utterances – a field called linguistic analysis. In short, the building block are defined as:

- **Form.** Splitting the signifier from the signified (Chang, 2008, p. 33). Where the Form part is the signifier. Uses schemas to be captured.
- **Meaning.** "The signified" part and is also captured using schemas. Highly dependent on the form part, since meaning is always context dependent. Therefore, a Form-Meaning pair should be viewed as a single entity. Using the proposed definition in chapter 3, *empty names* can be modelled as F-F pairs, in which the meaning is only symbolic.
- **Mapping.** To help conceptualize (imagine) abstract concepts *metaphors* are used, such as **CHANGE = MOTION**. This causes certain (depending on context, culture and personality) properties to transfer from *change* to *motion*, like in "*i went from happy to super happy in a second when i won the lottery*", referring to a change of state indicated by the "from", "to" and "in a second" properties of Motion. The power of metaphores can be directly seen in games such as "Pictionary" where a player has to evoke abstract concepts using only body language. Another mapping tool between concepts are *metonyms*. Whereas metaphors use similarity between concepts, metonyms use an "overloaded" association between concepts such as **THE WINDY CITY = CHICAGO**. Metonyms are powerful instruments in a language community, but lead to serious misunderstanding when expressed to receivers that are not aware of the associated meaning (ie. understand it *literally* instead of *figuratively*), for example limited people know that "the windy city" is a metonym for Chicago, not some city affected by wind. Furthermore, it is not a simple synonym like **TO BUY = TO PURCHASE**, since the meanings are not identical; its usage originates from political context and is intended to evoke this additional context implicitly (hence the naming "overloaded" association).
- **Mental space.** Initially infants appear to be only capable of perceiving the world egocentrically, but within a few months to a year they learn the benefit of estimating the worldview from the sender to enhance communication. One possible mechanic that could enable this are mental spaces (Fauconnier & Turner, 2003). A mental space is subconsciously constructed to understand the senders embodied worldview and mental state and contains all believed activated schemas that belong to that sender. Using a mental space one can make predictions on the knowledge and belief state of senders and thus their behavior.

In the rest of this report abstract Embodied experiences such as ball and food are discussed. For these abstractions to be used consistently and clearly requires modelling it first. The EAS cycle from chapter 3 shows that knowledge is built on top of previous knowledge, so connections and even hierarchy between knowledge is created, so called ontologies or (*fundamental*) *categories*. An example of an ontology can be found in Fig. 6-2. Although there

is controversy whether fundamental categories actually exist, prototype theory by (Rosch, 1973) empirically show that humans use categories all the time. These categories are most of the time not fundamental categories (have strict boundaries) but often have soft, context dependent boundaries. Modelling knowledge as an ontology is not strange, since neurons also have these context dependent properties. The ontology for knowledge is depicted in Appendix 6. Since knowledge is nothing more than stored (and processed) Embodied experience which in turn is just a pattern of Embodied sensory values, the knowledge tree is in a sense one huge pattern. Starting from the top, the pattern becomes smaller until the level of individual, processed Embodied sensorial values are reached. The subdivision in smaller patterns allow easier discussion of abstract Embodied experience.

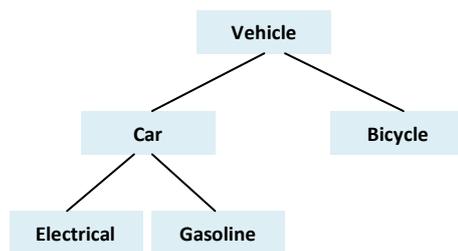


Fig. 6-2 Example ontology of a vehicle with strict boundaries

6.2 Embodied theory of language aquisition

“.. learning a language, like playing tennis, consists in becoming competent at a rule-governed practice. It does not consist in explicitly learning the rules, having them, so to speak, written in one’s mind; being able to play tennis, or being able to speak and understand a language consists in being proficient at doing something according to the rules.”

Wittgenstein in (Marti, 2010)

How does embodied experience gets translated to thoughts (concepts) and eventually language? Learning language, although smaller in scope than learning in general is still unsolved science. Some concepts like “time” and “events” seem to be encoded in our DNA since they arise in every culture. Or, as mentioned in chapter 5, these concepts could be the product of the Earth’s laws of physics and the identical biological body which effects every culture. Clearly, some concepts are undeniably learned such as “cycling” and the word “Tamagotchi”. In this section a theory on language aquisition is discussed that extends on the Embodiment hypothesis. Michael Tomasselo (2003) proposes a theory that is taken from a cognitive linguistic angle. According to his theory, understanding, and with that learning language²⁰, is about discovering its usage. This process already starts in the womb and assumes that the infant is truly a *tabula rasa*, interacting with the environment from the moment a neuroncell is functional and let its functionality (including its wiring) adapt to the input as assumed with “*language structure emerge from language use*”.

In the first subsection the theory is explained in some depth. The grammar part of the theory is highlighted and discussed seperately in the second subsection. In the third subsection child-directed speech is discussed, a special type of interaction with an infant which could (arguably) speed up language learning.

²⁰ Note that in this report no distinction is made between language aquisition and language learning as is proposed by (Krashen, 1982) to indicate *natural* learned languages (mother tongue = L₁) and consciously learned languages (foreign langauges = L₂).

6.2.1 Language learning: a usage-based theory

Learning the sound patterns uttered by the mother and surroundings in the earliest stages of development is no trivial task. *"Language is not simply the association of a sound pattern with an event. Sound becomes language for young children when and only when they understand that the adult is making that sound with the intention that they attend to something"* (Tomasello, 2003, p. 23). This also emphasises the social aspect of language. Based on current knowledge (see chapter 5), the learning mechanisms of the brain allow only physical connections (wiring) and strengthening of that connections and possibly some chemical mechanism to manage it all. Although humans have tens of billion neurons they are too few to connect each possible sound pattern to each possible event. The brain apparently does things more efficient as discussed in chapter 5. Using Tomasello's usage-based theory, part of that efficiency of the brain is explained by storing knowledge that is functional. The view used for his theory is usage-based, functionalist, socio-pragmatic, constructivist and emergentist (see (Ambridge & Lieven, 2011) for an overview of language acquisition theory overview including claims and predictions).

Recent work on infants by Michael Tomasello (Tomasello, 2003) shows that infants might learn language competence by understanding how a competent user uses it. The communication oriented language acquisition theory proposes that humans learn a language because it has a communicative function. Humans according to (Tomasello M. , 2009) are innately motivated to cooperate and collaborate with the social environment. Even before a child can talk, it can communicate by other forms of bodily communication, such as pointing. Tomasello notes the following: *"The interesting thing about pointing, is that it contains almost no information in the gesture itself; it basically says 'look in that direction and you know what I mean'."* (Tomasello, 2003, p. 70). Motives for pointing are imperative (a command), declarative (a statement) or informative (instructive). By using language, communication of these motives can be made much more precise than with pointing. Each of these motives involve social interaction and before the motives can be communicated a mutual context needs to be constructed, for example by making eye contact or by pointing. The result is a so called joint-attentional frame (subset of the world) in which the child assumes that this frame is understood by all communicative parties (the shared understanding principle). It is within this frame communication takes place most conveniently. Often elements from the frame are highlighted or focussed on to be even more precise, for example using pointing or gaze tracking.

Tomasello discusses many interesting observations in his book, for example (Tomasello, 2003, p. 23): *"It is interesting that deaf children acquiring a signed language do so on the same general timetable as hearing children learning a verbal language, thus demonstrating something of the robustness of the symbolic dimensions of human linguistic competence."* One hypothesis based on this observation is that shared interaction is more important in language acquisition than choosing a preferred type of modality. Language is robust through its context dependence, using a fusion of different modalities.

The usage-based language acquisition theory describes two key cognitive processes: *intention reading* (the functional dimension) and *pattern finding* (the grammatical dimension) (Tomasello, 2009). Using the definitions of chapter 3, the intention reading dimension can be named the Meaning dimension, while the pattern recognition dimension can be named the Form dimension. Both are involved in four specific language learning processes (Tomasello, 2003):

- **Intention reading and cultural learning** (learn conventional Form-Meaning pairs, including everything from words to complex constructions)
- **Schematization and Analogy** (learn abstract syntactic constructions)

- **Entrenchment and Pre-emption** (constrain abstractions to what is conventional in a linguistic community)
- **Functionally based distribution analysis** (learn linguistic categories such as verbs, nouns)

As mentioned above, learning language is a social activity. It is learned by being repeatedly exposed to it. Of course the usage is not always correct and it's up to the language community to correct wrong and stimulate fitting usage. This is a normal path to reaching a competent language level. One of the things that should become apparent is that partly due to the corrections of wrong usage both Form and Meaning change often during childhood. Its truth is unstable. Some trivial examples of Form and Meaning change are presented below:

Form change

- **Specialisation:** Learn that the sound pattern "likepancake" is actually a combination of two sounds, 'like' and 'pancake'. The meaning of like and pancake is initially near identical after the split (for example only the order differs) but will diverge with meaning change through addition or decay.
- **Generalisation:** Learn a rule (pattern) that states that the sound patterns "star" and "fish" is part of a larger sound pattern "starfish".

Meaning change

- **Addition:** Learn that a pancake not only affords itself to be eaten but also to be thrown like a frisbee.
- **Decay:** Learn that pancakes are not edible if one is allergic to ingredients, so the probability of activation in context with eating stabilizes or decays while thrown like a frisbee (for example) rises.

6.2.2 Grammar learning

"Functionalist theories, with their communication-in-context view of language, understandably tend to favor a constructivist view of language acquisition, according to which enough information is available in the linguistic environment for the child to be able to construct a grammar, with the help of more general cognitive factors and learning capacities. This is not to say that functionalists totally reject the possibility of innate characteristics: what is at issue is rather the nature of this endowment. For the functionalist, it is more likely that what is innate will be a set of general cognitive principles and predispositions to learn rather than some concrete set of linguistic rules and principles."

Christopher Butler (**Butler, 2010**)

Learning language requires learning the constraints/rules of a language. Grammar is *a set of rules, but not in the sense as Chomsky and other Formalists use them*. Constructivists such as Tomasello prefer to see them as a set of non-static structures or simply a *set of slowly changing patterns* (since language is dynamic and thus difficult to capture in static rules). A grammar in their view is not a collection of rules but rather a collection of constructions (Form-Meaning pairs), since lexical symbols are bound to its usage. The construction is not static, it emerges from the need for it (Tomasello M. , 2006). Besides grammar giving structure to understanding and constructing sentences, it also plays an important role in giving cues for constructing mental spaces (Fauconnier G. , 1994). Learning grammar, finding and learning patterns according to the usage-based language acquisition theory, involves the four processes described above, which are based on **intention reading** and **pattern recognition** skills.

A study by (Saffran, Aslin, & Newport, 1996), show how great the **pattern recognition** skills work for young children (8 months). Already they can recognize word patterns in utterances. Finding smaller and larger patterns, the Form-Meaning constructions used in situated context, is the basis for grammar. In producing language a child may use this grammar as a template to fill in (or connect) new words into previous associated words with same sense (use) and thus the same implicit and explicit functional restrictions.

Research by (Akhtar & Tomasello, 1996) show how great the **intention reading** process works for young children, in this case for finding a linguistic pattern. Children were learned a made up verb “meek” and attached it to a for them new action (“swinging a toy on a toyswing”). A child is shown how to swing a toyswing, but when told how the action is called the toyswing was removed such that the child could not see the action paired with the verb. However when the child was asked “to meek” they immediately searched for the toyswing and when found gave it a swing. This research shows that the infant can, within the attentional frame, find the general intention of the speaker and associate this to the newly heard word. Note that, like with most behavioral research, not all children show this behavior – which fits the Embodied position of a delicate interaction with the body, its mind set (including motivation and attention) and its environment.

There are over 6000 spoken languages each having different grammar and vocabulary. The partitioning of language both structurally and functionally, is deeply related to the cognitive processes of intention reading and pattern finding. According to Tomasello the strategy to partition is either holistic (for languages such as English with many small words) or reductionistic (for polysynthetic languages such as Eskimo containing large single words). The human brain mechanics for pattern finding are generic enough to search for patterns using either strategy, and it is assumed that children use both strategies. Furthermore, Tomasselo assumes that children always “*attempt to understand the entire communicative intention of the adult’s entire utterance*” (Tomasello, 2003, p. 75).

child:	daddy (waving arms to attract attention, neutral tone of voice)
daddy:	yes sweetie
child:	daddy (again waving arms, neutral tone)
daddy:	do you want to play?
child:	daddy! (with pointing, stressed tone)
daddy:	ah, you want your toy (child smiles as daddy passes the toy)

Example 6-1 Example of a dialogue with holophrase usage by the child

Tomasello hypothesized that an (English) child’s grammar, the (Embodied) construction grammar, passes four phases (Tomasello M., 2006):

1. **Holophrasis.** A single word such as “Birdie!” (or Example 6-1), with various possible interpretations (communicative intentions) to be guessed by the receiver. In the holophrasis stage an utterance is accompanied with various body language, context and tone of voice.
2. **Pivot Schemas.** Multiple words to express communicative intention, thus partitioning the experiential scene into two or more words like “Birdy fly” or “where’s-the-X”.
3. **Item-Based Constructions (verb island constructions).** Syntactic markings such as word order are used to express communicative intention. Syntactic markings are not

consistently used and depend on construction. They are often based around verbs such as “X hit Y”, “X broken”.

4. **Abstract Constructions.** Full (adult-like) sentences, with abstractions (levels of substitutions, transitivity) are used to express communicative intent. For example “My name is Steven” is monotransitive (one subject, one object), while “I gave this paper to Pascal” is ditransitive (one subject, two objects), etc.

The usage-based theory of language acquisition is clear enough to produce claims and predict behavioral results. Independent research by Saskia Peels (Peels, 2011) testing the theory on Dutch children show promising results in favor of Tomasello’s theory compared to formalist theories of language acquisition. Further research is needed to see whether the four grammar stages apply to other languages than English.

6.2.3 Infant teaching methods of language

This last subsection is not devoted to the acquisition of language but rather teaching it. Addressing children during a speech act is called *child directed speech* (CDS) and if done by an adult called *adult-to-child speech*. It is interesting that seemingly automatically adults in some cultures such as Western cultures, change their speech and interaction pattern (see Fig. 6-3) when addressing children. This is called *motherese* or *parenthese* [Newport, Gleitman, & Gleitman, 1977] and is hypothesized to support language learning at the early stages. Steven Pinker (Pinker S., 1994) noted that motherese is not essential in learning languages since some cultures don’t use it (while still producing language competent children of course). Instead these children learn the language by being exposed to the full extent of their language by competent language users. Pinker (Pinker, 1994, pp. 39-40) found this not unreasonable, stating: “*After all, young children plainly can’t understand a word you say. So why waste your breath in soliloquies?*”. Although arguable plausible that parents or any explicit teacher are not responsible for the vast majority of a child’s linguistic competence (the language community is responsible), the role of structured exposure by means of a teaching strategy such as Motherese should not be misjudged. It is not unthinkable that learning auditory patterns explicitly by being pointed out to them by adults has some effects in the brain, however further research is needed to discover the extent and polarity (the effects being positive or negative).

Paralinguistic

- Slower speech with longer pauses between utterances and after content words
- Higher overall pitch; greater pitch range
- Exaggerated intonation and stress
- More varied loudness pattern
- Fewer dysfluencies (one dysfluency per 1,000 words versus 4.5 per 1,000 for adult-adult)
- Fewer words per minute

Lexical

- More restricted vocabulary
- Three times as much paraphrasing
- More concrete reference to “here” and “now”

Semantic

- More limited range of semantic functions
- More contextual support

Syntactic

- Fewer broken or run-on sentences
- Shorter, less complex sentences (approximately 50% are single words or short declaratives)
- More well-formed and intelligible sentences

<ul style="list-style-type: none"> • Fewer complex utterances • More imperatives and questions (approximately 60% of utterances) <p><i>Conversational</i></p> <ul style="list-style-type: none"> • Fewer utterances per conversation • More repetitions (approximately 16% of utterances are repeated within three turns)

Fig. 6-3 Characteristics of motherese compared to Adult-to-Adult Speech (Owens, 2007)

6.3 Embodied theory of language use

As mentioned in section 6.1, language structure emerges from language use, but how do we use language? First off, using language requires mental effort – try to speak a foreign language, it is tiresome to search for words. Using “*language use = mental effort*” as (trivial) insight, the NTL and philosophical framework as a toolkit and the knowledge that language is used to communicate (see chapter 3), the how question is approached. This subsection will model key (bio-inspired) mechanisms required to use language. The section starts with a model on motivation and personality to use language, introducing various behavioral models such as the *Belief Desire and Intention (BDI)* model and the more sophisticated *Integrated Change* model. The next part models human reasoning, discussing the *simulation hypothesis*, *indexical hypothesis* and *Bayes’ Rule* to deal with uncertainty and language use.

6.3.1 Modelling motivation and personality

What is it to be human? A question asked by many philosophers and certainly by many AI pioneers. Is it the ability to be creative or is it the never ending human curiosity - explaining phenomena, how little they may be? Are humans defined by their unique personalities generating desires, needs and goals or is it their ability to reason, to infer and act logically? For certain, whatever *it* is, broke with animal instinct and brought humanity to the top of the foodchain.

For this report *creativity* is presumed to be the major differentiator between humans and other species. While other species might be creative as well, the extend to this is comparatively limited. I argue that no other species is as creative to adapt to new environments. The definition of creativity in Table 6-2 can be reinterpreted as “*the ability to produce new knowledge using existing knowledge*”, however an ability does not act on itself, it requires some sort of trigger; the process of motivation.

Table 6-2 Definitions of Creativity

Source	Definition
(Dictionary.com, 2012)	Noun: 1) The ability to transcend traditional idea, rules, patterns, .. or the like and create to create new meaningful ideas, forms, ..
(Penguin, 2006)	Noun: 1)The ability or power to create 2) having the quality of something imaginatively created

Table 6-3 Definitions of Motivation

Source	Definition
(Dictionary.com, 2012)	Noun: 1) the act or an instance of motivating (give incentive to) or providing with a reason (a basis or cause) to act in a certain way

(Penguin, 2006)	Noun: 1) enthusiasm or drive 2) the act or an instance of motivating or being motivated 3) a motivating force, influence or incentive that directs one's action toward achieving a desired goal 4) a motive
-----------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

So what psychological feature arouses a human to get enthusiastic or motivated (see Table 6-3)? Well that could be anything from anything intrinsic stimuli such as “*feeling cold*” and a desire or need “*to not feel cold*”, to anything extrinsic stimuli such as seeing the latest Microsoft Surface (tablet PC) and a desire to have it. But if motivation is the driving force to pursue a desire, what decides it to be an arouser, what makes *anything* a desired goal? This is where *personality* enters the scene (see definitions in Table 6-4).

Table 6-4 Definitions of Personality

Source	Definition
(Dictionary.com, 2012)	Noun: 1) a person as an embodiment of a collection of qualities 2) the organized pattern of behavioral characteristics of an individual
(Penguin, 2006)	Noun: 1) the totality of an individual's behavioral and emotional tendencies; a distinguished complex of individual or group characteristics

In other words personality is a “*complex, (unique) mix of behavioral qualities or characteristics/traits that creates desires, needs and goals*”. A unique personality lies at the basis of creativity and thus makes up the *it* in “*what it takes to be human*”. This answers the initial question, but how does personality relate to language use? A personality consist of traits that cause behavior, especially planned behavior such as language use. Various models try to model human behavior such as the traditional behavioral model *Sense-Think(Plan)-Act*²¹ (Fig. 6-4) and the *Attitude-Social influences-Self efficacy (ASE) behavioral model* [eg. (De Vries, 1989) (Fig. 6-5).

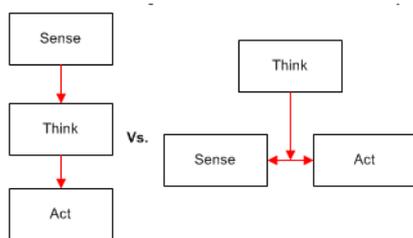


Fig. 6-4 Two takes on the STA behavioral model

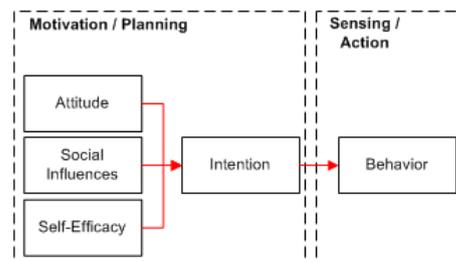


Fig. 6-5 ASE behavioral model

The ASE model models generating intention caused by motivation, but is not always the case. Behavior is not always voluntary, for example the Tourette syndrom or the knee-jerk-reflex force unplanned behavior. Also, the biological underpinnings of motivation and personality are largely unknown, needed for example to model implicit or “innate” motivation. An example of implicit motivation is the hypothesis that most infants have keywords that trigger attention to parts of a verbal message (such as the pause “uuh”) because the words after the pause might be important since it took the sender some mental effort (time) to come up with the word (Kidd, White, & Aslin, 2011).

²¹ A nice body of work on STA approaches including detailed models: http://www.sense-think-act.org/index.php/Introduction_to_the_STA_project, last visited june 2012

A more practical behavioral model is the *Belief-Desire-Intention* (BDI) framework from (Bratman, 1987). This model contains knowledge (all three types, see chapter 3 and 6.3.2) of the world and the self in the form of beliefs. It models personality and motivation (eg. rating *a nunciari* and other types of knowledge such that they can be compared and selected) in the form of desires. And finally model complete and incomplete (reasoned) hierarchical plans of actions as intentions (Bratman, 1987, pp. 2-3). Bratman defines intentions to be not merely predominant desires (blz 18-20), but desires that have been committed to (time for deliberation is over). A plan to pursue the desire is formed and the initial step to execute is taken; the plan is *in transit*, it has become an intention. Note that physical and cognitive processes such as reasoning, perception and action are not explicitly modelled. Modeling reasoning is discussed in the next subsection (6.3.2). Modelling action is slightly discussed in chapter 7 with theorized elementary motor schemas or *motor synergy* and a model for perception is slightly discussed in chapter 5 on sensor saliency, awareness and attention (the three-tier memory model).

None of the above models is detailed enough to accurately predict human behavior. A recent model that is only given as reference in this report is the *Integrated (I) Change* model by (De Vries, Mesters, Van de Steeg, & Honing, 2005), depicted in Fig. 6-6. This model tries to integrate all the concepts of the STA, ASE and BDI models in a psychological explicit form.

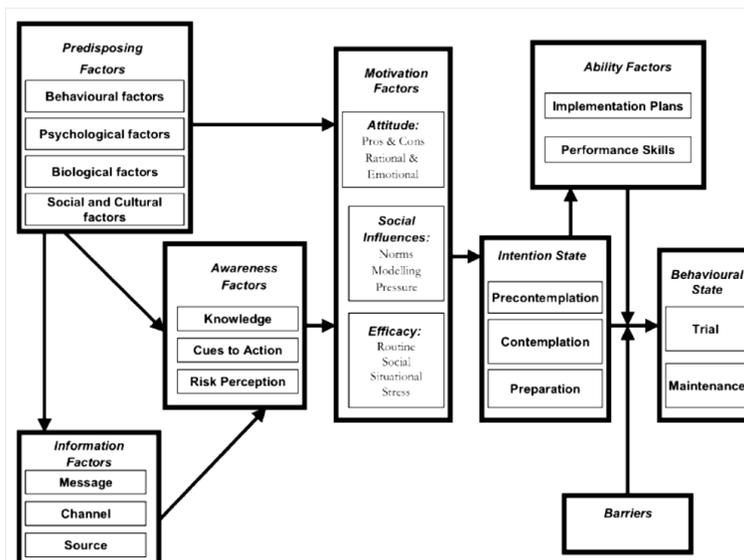


Fig. 6-6 The I-change model how human behavior could arise (source: (De Vries et al., 2005))

Discussion

Some of the human desires and needs are learned such as the possible desire for a tablet PC, but others seem to be hardwired in the human genes such as the need for sex, sleep, air, food and water. How the acquisition, rating (for pursuit) and selection of these desires and needs works is largely a mystery and is at the root of the famous qualia problem discussed earlier (as well as in chapter 7). Why do some people enjoy drinking a coke for the first time and others don't or why does the same force cause different pain experiences for two people? Assumed is that each human personality has a unique mix of mentioned attributes that makes every experience subjective. According to some studies using new brain imaging techniques, some subjective experiences such as love and beauty correlates with specific brain areas (Kawabata & Zeki, 2003), making the qualia problem a measurable science. This still doesn't reveal the mechanisms of personality, but it does validate aspects of it in hindsight. Since the desired artificial NLU system should have (initially a few) human traits, requires to address this problem in the AI phase.

6.3.2 Modelling human inference

Before modelling human inference, a legitimate question to ask is what human inference – human reasoning – actually is. The term inference has already been used several times in chapter 3.4.3 (logic) and 6.1 (language structure) in which examples were given of wrong inference (eg. if knowledge is missing) and a strategy of an inference engine (eg. best-fitting of *a nunciiori* to *a priori* knowledge leading to good expectatations). However that is not the complete picture, and while that is not the aim in this report, important characteristics listed by (Devlin, 2000) allow a more clear definition than in Table 6-5.

Source	Definition
(Dictionary.com, 2012)	Noun: 1) The process of forming conclusion, judgments or inferences from facts or premises.
(Penguin, 2006)	Noun: 1) The faculty of mind by which evidence is weighted, inferences are drawn, truth and falsity are judged and explanations are sought and given 2) an explanation or justification

Fig. 6-5 Definitions of Reasoning

The process of forming thoughts, or strongly simplified as *human, evidence-based reasoning* is summed up by (Devlin, 2000, pp. 5-6) and listed in Fig. 6-7. These features often do not stroke with traditional computational approaches to reasoning, requiring the AI and NLU field to rethink its models (see chapter 7).

- | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> • It is not always linear • It is often holistic • The information on which the reasoning is based is often not known to be true. The reasoner must, as far as possible, ascertain and remember the source of the evidentiary information used and maintain an estimation of its likelihood of being reliable. • Reasoning often involves searching for information to support a particular step. This may involve looking deeper at an existing source or searching for an alternative source. • Reasoners often have to make decisions based on incomplete information • Reasoners sometimes encounter and must decide between conflicting information • Reasoning often involves the formulation of a hypothesis followed by a search for information that either confirms or denies that hypothesis • Reasoning often requires backtracking and examining your assumptions • Reasoners often make unconscious use of tacit knowledge, which they may be unable to articulate |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Fig. 6-7 Features of human, evidence-based reasoning (source: (Devlin, 2000))

Keith Devlin continuous right after this summation (Devlin, 2000, pp. 5-6): *“The factors we have just outlined imply that real-life, evidence-based reasoning is rarely about establishing “the truth” about some state of affairs. Rather it is about marshalling evidence to arrive at a conclusion. If the reasoner wants to attach a reliable degree of confidence to the conclusion, she or he must keep track of the sources of all the evidence used, the nature and reliability of those sources, and the reliability of the reasoning steps used in the process.”*

Using the above insights allow a more appropriate Embodied definition of reasoning that includes tracking the line of reasoning, focusses on using reasoning for both comprehension and expression and the use of logic. It is furthermore important to note that reasoning is independent of rationality. Rationality is a motivational strategy that steers the process of reasoning, such that the process can be executed both rational as well as irrational (mixing emotions with carefull considerations).

Lets start with defining reasoning as “forming a logical (ie. neurological connected and thus verifyable) line of argumentation that reaches a conclusion including calculating its logical and confidence (belief) level using stored Embodied experiences”. Note that logical is explained as neurological connected (only connect what is known, so based on a priori knowledge) and verifyable. As mentioned in chapter 3 regarding “believed logic in percentages”, verifiability can be quite a vague term in terms of belief, but is very clear conceptually and neurologically. With explicit knowledge (see chapter 5) one can trace the reasoning of an conclusion to a “precise” origin (the core beliefs/assumptions) whereas with implicit (tacit) knowledge the trace is fuzzyfied since its is based on weighted and cumulative experiences. The formation of a logical line of argumentation can be done in three ways: *inductive*, *deductive* and *abductive*. Inductive reasoning tries to determine the pattern by learning the pairing of a conclusion with a premise often after observing many examples. For example this brand of cola tastes great, so the next taste of cola of the same brand will also be great. Deductive reasoning tries to determine the conclusion using a learned pattern and filling in the observed or assumed premises. This slot-filler type of reasoning is (seemingly) easy to use but highly depends on the predictive quality of the learned pattern. An example is the pattern “cutting people is a crime”, the premise “surgeons cut people”, which lead to the (wrong) conclusion “surgeons are criminals”. Deductive reasoning has many operational strategies “rules of inference”, such as the *modus ponens*. The third type, abductive reasoning tries to determine the premise by using a learned pattern and assuming a conclusion. This type of reasoning best fits the hypothesis testing strategy and search for evidence to support it. An example is observing a red banana (conclusion), the pattern is “bananas are yellow”, however if it was painted with a red color (the hypothesis) it turns red.

The formation of thoughts, in other words *the process of reasoning*, is “managed” by motivation and personality discussed earlier. Understanding thoughts through reasoning involves finding the best match between newly formed thoughts and the current mental state (a set of active thoughts) (Feldman J. , 2006). Finding the best match (eg. the best possible Type 1, 2 or 3 Understanding) involves combining many factors including evoking metaphores (Feldman J. , 2006) and metonymes - allowing distant connections to be included in the calculations. One of the key tools the brain is assumed to use in the reasoning process is simulation. Simulation allows detailed approximation of situations either already experienced or not yet experienced. Possible neural tools for simulation is propagating a smaller portion of the stimuli to all connected neurons or have the motivation process activate associated stimuli from the input experience and mirror neurons (see chapter 5). According to the *indexical hypothesis* by (Glenberg & Robertson, 2000), understanding for example a sentence means being able to simulate the actions (uses) that underlie them. The symbols in a sentence are indexed first (similar to Type 1 Understanding) and then mapped to uses (similar to Type 2 and 3 Understanding), hence the name *indexical*. Another term of “uses” in literature is *affordances*; humans categorize the situation around the possibilities it affords them. The indexical hypothesis models the intimate relationship between language and action knowledge (see the phenomena “Action-Compatilby Effects (ACE) during language comprehension”, (Glenberg & Kaschak, 2002)) and is further proof for Embodiment. (Chang, 2008) emphasizes the intimate relation by posing the *simulation hypothesis*: “Language exploits many of the same structures used for action, perception, imagination, memory and other neurally grounded processes. Linguistic structures set parameters for simulations that draw on such embodied structures”. The larger perspective on the intimate relation between perception and action in humans is debated in *common coding theory* which finds neurological proof in the mirror neuron system and fully support Embodiment.

According to (Körding & Wolpert, 2004), all the discussed insights on inference can be modelled using the classic Bayes' Rule. They even claim²² that humans are in essence "Bayesian inference machines". Using this model, uncertainty (eg. understanding a meaning) can be combatted by fusing *a priori* knowledge with *a nunciari* knowledge resulting in a belief given the input experience. The following terms and formulas can be used to materialize the theoretical model of knowledge discussed in chapter 3 and else in this report:

- The *a posteriori* knowledge (belief | a nunciari) is defined as **Expectation**
- The *likelihood* (a nunciari | belief) is defined as **Prediction**
- The *a priori* knowledge (belief) is defined as **Assumption**

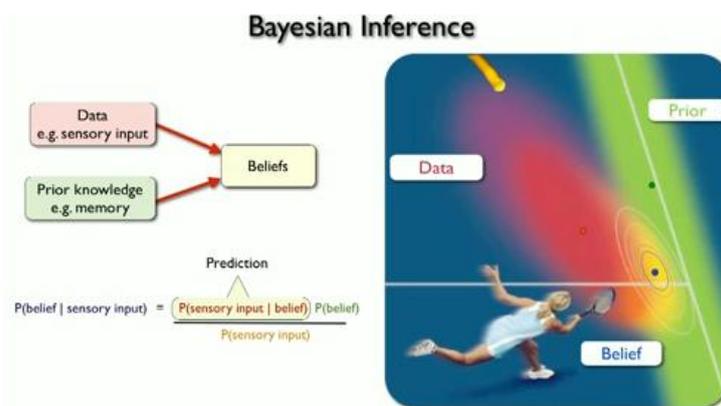


Fig. 6-8 Bayesian inference (source: (Körding & Wolpert, 2004)). Note that Wittgensteins example of playing tennis can be used well as a metaphor to learn language

Using Bayes' model now allow dealing with uncertainty and predicting the consequences of actions (Wolpert, Ghahramani, & Jordan, 1995) which are fundamental for:

- Control with delays (eg. fine motor control such as the voice box)
- Mental simulation (eg. communicative motive estimation)
- Likelyhood estimation (eg. intention reading)
- Sensory filtering (eg. environmental noise reduction)

These subjects are trivially important for language usage and the NLU system in general (as discussed in this chapter) and further mark the potential of this model. It is interesting to note that intuitively the model allows discussing complex statements such as "*wrong assumptions are most dangerous to understanding, since they cloud objective observations (sensory input)*".

The inference model of Knopper and Wolpert (Fig. 6-8) can be integrated in larger models for example by combining insight from game theory. Clark (2011), who favors the social nature of language and use-based models, argues that meanings are conditioned by the sender such that the receiver interprets it as intended by the sender. This is essentially a game of predictions and decisions between multiple agents bound by rules inflicted by Embodiment. With the help of various language games artificial agents can develop coordination, cooperation, common knowledge and other types of knowledge. Most importantly the games (together with a behavioral model and reasoning framework) could help bootstrap advanced language understanding and usage skills with no or few primitive language skills at its basis, since in theory – as discussed in chapter 3.4.2 - knowledge should scaffold.

²² <http://royalsociety.tv/rsPlayer.aspx?presentationid=80>, last visited june 2012

“Cogito ergo sum – I think therefore I am”

René Descartes

7 The artificial NLU system

In which the design guidelines, computational techniques and difficulties of transforming the human NLU system to an artificial one are discussed. Methods for NLU system validation and benchmarking are presented and a small overview of important NLU systems.

At this point of the report, a global understanding of the human NLU system is reached; its development, components and most importantly its dependability on the body. Also a working definition for meaning, understanding and a theory of knowledge is presented such that the body's interactions with its environment lead to reusable experience. Converting this insight to an artificial system is not trivial since computational, sensory and actuator technology is not even near the same level of the human biological counterpart. Again assumptions need to be done, restrictions need to be imposed and worse; abstractions are needed from reality, which wasn't holding stable truth due to its many assumptions to begin with. One might wonder if this day and age is the right time for such a pioneering endeavor. Fortunately, the “Principles of Embodied AI” by Rolf Pfeiffer together with the earlier discussed NTL framework give some support and direction that justifies this next step.

For the bigger picture it is interesting to briefly mention where artificial NLU currently is. Possibly one of the earliest applications was Weizenbaums Question&Answer (Q&A) program “Eliza” in the nineteesixties (Weizenbaum, 1966). The technology behind it was simple, basic scripting (template matching) and some rules for the rewriting of questions like a Rogerian psychotherapist (open ended questions) – “I don't feel well.” becomes “Can you elaborate on not feeling well?”. The program was a shocking succes for Weisenbaum, some people chatting with the program for more then half an hour without noticing it was a computer program. He left the field and became a critic of AI because of its ethical implications. Amazingly, but since then (over 45 years ago) little progress has been made in the development of chat/dialogue programs or “chatbots” (see dialogues with the current best chatbot in Appendix 3). The current most advanced chatbot is voted to be “Rosette” from Bruce Wilcox winning the bronze medal of the 2011 Loebner Prize. The Loebner Prize is an annual contest where several judges need to be convinced their chat partner is human and not a computer program – the famous *Turing Test* (Turing, 1950). Currently a simplified version of the Turing Test is used with textual input/output only, so no visual or audio input/output. Since the price's inception at 1991 only the encouragement award – the bronze medal with a modest price of \$4000,- has been awarded, showing the difficulty of the test. It should be noted that winning the Turing Test – which is essentially a mimicing game – is only the starting point of good NLU systems and intelligent systems, since it is a one-sided test. The inventor of the term AI, John McCarthy mentions on his website²³, that passing the test “*does not really test whether a machine or computer program is intelligent, only that it is able to fool a human*” (paraphrase) – something that Eliza did without much effort, if the human is caught off guard. McCarthy also mentions that AI reaching human level intelligence requires new paradigms and new approaches to escape the current “in-the-

²³ <http://www-formal.stanford.edu/jmc/whatisai/node1.html>, last visited june 2012

box-thinking". The current widely popular "more data" approach is the easy approach, however human intelligence is not about quantity – humans are not omnipotent - but about qualitative usage of data. This means letting go of the scripting of dialogues and sometimes requires a step backwards for two steps forward – in search for the global maximum.

This chapter starts very broad, discussing several important historical problems and directions on the development of generic AI systems. In section 7.2 the development of artificial NLU systems are discussed, such as its general architecture and components, NLU specific historic problems and most importantly practical, computational techniques to build Embodied NLU technology. Section 7.3 discusses validation and benchmarking of NLU systems such that its performance can be compared with other NLU research. The last section, 7.4, present a short overview of key work in the artificial NLU field.

7.1 *Developing (Embodied) AI systems*

7.1.1 A brief history of problems with developing AI systems

To better understand the AI technology implemented in NLU systems it might be interesting to briefly discuss some historic problems of AI technology in general, which in one way or another will have impact on NLU systems. Some of the problems are trivial in large scale systems engineering²⁴ on both hardware and software level, such as complexity and scalability. These are umbrella terms for challenges such as modularity, parallelism, storage and choosing the the right components including their interoperability and model abstraction level. These are in their own right mind boggling. Working with them is done using various problem solving approaches such as extensive research, analogy searching, divide-and-conquer, reduction and trial-and-error. Non-trivial problems also exist and in particular the following three philosophical problems affect human-like NLU systems:

- **The Chinese room**

The philosopher John Searle posed a famous argument "The Chinese Room" (Searle, 1980) in which an English speaking person is in a room with two dictionaries Chinese to English and reverse. He receives Chinese instructions and translates them with the dictionaries and hands out the Chinese instructions just like a native Chinese would do. John Searles argument is that this person or computer for that matter, does not literally understand Chinese, but merely simulates the ability to understand Chinese. He defines a computer that literally understand Chinese "strong AI" and the other "weak AI (or applied AI)"; the first able to have a mind, **is** intelligent, **is** thinking while the other **acts like** it has a mind and **acts like** its intelligent and thinking (Russell & Norvig, 2003, p. 947).

Discussion

From a pragmatic point of view the question rises if it matters if a system has strong or weak AI. Humans act all the time, so why shouldn't systems be the masters in acting – so good that humans can't tell the difference? The proverb "Ignorance is bliss" is appropriate here, so if we are not aware that we deal with machines that perfectly act like they have a mind, we could care less that it has weak AI technology. It is a whole other story when viewing strong vs. weak AI in the light of ethical implications. I hope that long before reaching human-like general AI, independent of it being weak or strong, ethical implications are accepted by the vast majority of the human population.

²⁴ the term cognitive (systems) engineering is maybe better suited since our focus is on humans

- **The Frame problem**

The next problem is related to artificial learning. It was posed by John McCarthy and Patrick Hayes (McCarthy & Hayes, 1969) and put in the wider philosophical context in 1978 by Daniel Dennett (Dennett, 1987) and is called the *frame problem*. This problem discusses *the amount of changes* that occur while performing an action such as kick a ball – some changes occur (the leg and ball move), while most stay the same (the playing field, colors, etc), just like drawing cartoon frames, hence the name. Humans biological systems are highly sensitive for changes but more importantly learning which of them is or is not relevant for performing an action is not fully understood yet –it is related to pattern finding. Artificial systems need to have some sort of information selection mechanism to be able to learn from situations. Learning and reasoning without a frame (boundaries) becomes difficult if the computer systems doesn't know which information is relevant to learn or reason with. Note the great similarities with the *correlation problem* mentioned in chapter 5. As defined by McCarthy and Hayes the frame problem is limited to correlations between actions and the environment originally in computers, but since the concept is the same on the neural abstraction level of humans, I use both as synonyms.

Discussion

Finding the amount of changes in an experience lies at the root of human evolution – our ability to detect patterns. If no patterns were detected, no adaptation/learning in a single generation can take place, decreasing chances of survival. Detecting all patterns is both theoretically impossible and practically inefficient. Clearly an all or nothing strategy for pattern recognition is not how humans operate, instead humans gradually learn more and more patterns and have selection mechanisms that activate them. Theories on how these selection mechanisms could work is discussed in chapter 5.

A short historical analysis of the problem is done by (Viger, 2010). Currently the frame problem has been solved in limited expressive (= non-realistic) situations using several approaches, such as using situation calculus or event calculus or is avoided at all using simple action scripts (eg. whatever the situation, just kick the ball). These solutions are all formal languages to describe a situation, reducing the amount of variables by creating a fixed set of possibilities allowed by the language. However, a consequence of using a formal language is that the system is limited in its expressiveness – basically telling the system what is important in the world and what not – and in all avoids the pattern recognition that makes human so flexible. A new approach is needed that is dynamic and has its language grounded in Embodiment instead of fixed and with formal symbols. This new approach must focus more on implicit learning using probabilistic approaches and allow single examples. I would suggest a solution involving an inference engine based on decaying beliefs (short term memory), positive and negative evidence belief modifiers and salience thresholds.

- **The Qualia problem (subjective experience)**

The last philosophical problem discussed is the famous qualia problem discussed and defined earlier in chapter 2 and 3. Interesting work on this problem is done by Damasio (Damasio, 2003), relating the latest in biological and clinical findings to qualia and Ramachandran and Hubbard (2001) with their four “laws” of qualia .

Discussion

How should two computer systems differentiate from each other while receiving the same input? Is differentiation desired? To start with the last question, differentiation can be

desired, depending on the application. Differentiation can give the machine unpredicted or at least non-linear behavior which is useful in social environments.

The how question is difficult, it is currently unknown how subjective experience manifest in humans so modelling it is largely guess work. One approach is to let the sensors of the machine have different weight functions to modify input experience. Effectively treating the qualia problem as variance in experience caused by variance in sensory input (eg. Decibel sensitivity in ears per frequency) and mental workings (eg. type of inference engine used). This is not odd since brain imaging using DTI and MRI show that each brain is uniquely wired. Also some sort of personality (motivation component) need to be constructed that has traits, habits, needs, desires and goals and is able to select them accordingly leading to "unique" or non-linear behavior. Combined, the variable sensors and developing motivation component provide a practical solution that needs to be tested for usefulness.

Of course many other philosophical problems remain such as *the free will problem* (partly discussed in section 5.2) and *the mind-body problem* (monoism or dualism - is the mind or soul separate from the body, a question affecting the Embodiment assumption). For now the non-discussed problems, philosophical or otherwise, are cast aside hoping to have sufficient ground and proceed to the pragmatic domain of NLU and its specific problem areas.

7.1.2 Principles of engineering Embodied AI systems

Rolf Pfeifer proposes 8 key principles (Pfeifer & Scheier, 1999)²⁵ based on Embodiment (Wilson, 2002) (see Appendix 5). One of the advantages of engineering according to these principles is that some of the problems discussed earlier such as scalability, modularity and parallelism can be approached using them. For example, consider this biological example:

The human body has many reflexes such as the knee-jerk reflex. This simple neural circuit is able to operate autonomously since waiting for instructions from the central nervous system could potentially damage the knee since force on the joint - for example during walking - is near immediate. The idea of many simple reflexes (simple because they only have a few parameters to operate like flexing and relaxing) working together to enable complex motor behavior is not new. (Bernstein, 1967) hypothesized biological design principles to deal with motor complexities quoting (Bailey, 1997, p. 21) "The key idea is the notion of the motor synergy, a subcortical continuous feedback control circuit for stereotype motion, which may be modulated by parameters". This implies (and is still considered controversial) that there is some sort of network of simple reflexes, quite possible build hierarchical with certain command neurons, since the human is only able to consciously control some muscles. (Franz, Zelaznik, & McCabe, 1991) has shown that humans are unable to draw a square and circle at the same time, indicating some sort of coherent fine motoric structure (Sternberg, Monsell, Knoll, & Wright, 1978) furthermore support the idea of the brain sending out sequential plans since there is a delay in starting and stopping certain sequences. One cause could be that motoric plans involve a stream of operations that can be interrupted but the stream that is "in transit" cannot be stopped. Of course the brain also support parallel motion plans (concurrent actions) such as preshaping the hand while moving the hand towards the object to be grasped (Arbib & Lyons, 1987).

²⁵ also see the 5 principles of Developmental Robotics (Stoytchev, 2006) which overlaps with Pfeifers.

Although all 8 principles are valuable to develop a system that is inspired on this example, two of them are especially worth mentioning here:

- **Cheap design involving redundancy**
- **Parallel and loosely coupled processes**

The consequence of the *cheap design principle* is that controlling complex motoric behavior should not be programmed, but discovered through trial and error. Initially a concept such as grasping involves only a limited set of hardcoded parameters (ie. the palm reflex, folding the palm around the object) but while experiencing grasping learn new parameters such as closing the fingers around the object. The physical experience as well as the mental experience (goals reached) that is acquired contribute to the construction of motoric plans (timing, selection of parameters, etc) to repeat the experience in the future under possibly different circumstances (in different contexts).

Quite likely, training time can be saved by exploiting the symmetry in the body, so learning fine motoric action plans for the left hand can be used in mirrored fashion for the right hand although reaching the same outcome is ofcourse not guaranteed (ie. writing with left hand is not equal to writing with the right hand). The muscles are unable to execute the desired level of motoric detail.

Another consequence, that of the *parallel and loosely couple processes principle* is that complex motoric behavior can be achieved with multiple action plans, not necessarily one, although humans clearly have certain preferences for certain plans. This means that various hierarchical architectures such as the Sense-Think-Act and Subsumption architecture need to be abandoned since components in these architectures can occur together in parallel.

7.1.3 Key directions in engineering Embodied AI systems

Many directions along the road of constructing human-like artificial technology have been “talked and walked”, a few of them are important to mention here, especially for near future work in the embodied AI field.

Connectionist Model vs. Constructionist Model

Although connectionist models (ie. Neural Networks) are the closest to artificial biomimicry of the brain, the models and workings of the models are currently too much an abstraction to capture human intelligence and language acquisition in particular. For example the function of neurotransmitters, the growth and decay of a single neuron and connections and the changing structure of a neural circuits are all not incorporated in these models. Current connectionist models also lack correct mechanisms to train the neural net, for example to acquire language as advocated by Tomasello. Mechanisms such as focus and communicative intention and function are needed to train the correct part of the neural nets. Tomasello also argues that current connectionist models “... work only with very small units such as words and grammatical morphemes. But the data and theoretical arguments reviewed in this book suggest that children also work with larger units such as whole linguistic constructions.” (Tomasello, 2003, p. 324).

Contrasted to the connectionist approach is the usage-based approach according to Tomasello which can be seen as a functionalist-constructionist model and uses three pillars as its foundation (Tomasello, 2003, p. 325):

- **Functionalist based.** The model is based on intention reading (segmenting and understanding of communicative intentions in linguistic structures).

- **Construction based.** Focus on whole constructions and focus on the learning and using particular words, phrases and expressions. (Tomasello, 2003, p. 326) *“As children attempt to read the intentions of other persons as expressed in utterances, they extract words and functionally coherent phrases from these utterances, but they also create item-based constructions with open slots on the level of whole utterances.”*
- **Usage based.** Focus on production, because language structure emerge from language use. (Tomasello, 2003, p. 326) *“The hypothesis is that a child hears and stores concrete utterances and find patterns in these stored utterances. This is a gradual and uneven process that depends crucially on the type and token frequency with which certain structures appear in the input ... This also means that over developmental time children’s linguistic representations become stronger and more abstract.”*

Another argument for using a construction model over a connectionist model is higher transparency (not a black box) and reduced complexity (reasoning can be tracked).

Logical teaching methods vs. Empirical teaching methods

As mentioned before, logic is a subjective experience. *“Humans often don’t act logic”* is a common phrase, but that doesn’t rule out that the actor might well find his actions consciously perfectly logical. Any body and any mind is slightly different, in the most surprising ways. The following example using *teaching subject matter* illustrates the importance of this slight difference.

Learning multiplication in primary school can be done using various methods, one of them drill and practice multiplication tables. While being able to answer $2 \times 1 = 2$, the child might not be able to discover the rules of multiplication (eg. multiplication is repeated additions), for example to compute $2 \times 10 = 20$. As much as the teacher tries different methods, the *aha-moment* for a particular child might come from an seemingly illogical method such as learning to map colors to integers ie. a certain red is 2, a certain blue is 10, a certain pink is 20. Thinking about multiplying for that child equals fusing colors. Although this child has a braincondition called color synesthesia – seeing colors for numbers, it just exemplifies that language comprehension is not about using methods that seem logical in the eyes of the teacher, its about what works for the learner (in this case an agent or program) using methods that are designed for a specific artificial body (or brain) just like in the real world. Through trial and error, one can end up at a method to teach agents how to multiply, guiding the pattern discovery process and validating with the desired result, without forcing one particular method.

Since AI machines are “human designed logic machines”, it is only sensible to use a logical teaching method, since the designers knows how the input is processed to output. There are few surprises in the processing and there is full control over the process. Humans on the other hand are somewhat unpredictable and very flexible. Learning subject matter should therefore be possible using various pattern discovery methods, leading sometimes to surprises which pattern work best for the system. Tomasello explains this nicely at the end of his book: “how children become competent users of a natural language is not a logical problem but an empirical problem.” (Tomasello, 2003, p. 328).

Data driven vs. Hypothesis driven

Most approaches to computational intelligence are build on the notion of *more data leads to better results*, for example, IBM’s Watson was being able to best the top human players in a game of Jeopardy! (see 7.4.2) by fully exploiting the strength of a computer and its data storing and processing capabilities. This data driven notion works quite similar in learning language, since patterns are often discovered after being exposed to it several times. Tomasello (2003) proposed a technique called “functional distribution analysis” to discover patterns in the usage of linguistic items.

However this is only half of the cake, or in this case half of learning language. Children are able to learn from a single example and furthermore can only become competent language user by trying out hypotheses – validating discovered language patterns in context Tomasello (2003). This trial and error method is not peculiar, since language is highly ambiguously used. The nice thing about the hypothesis testing techniques is that one can predict an outcome more accurate as opposed to the larger patterns found through statistical (data driven) methods. Another bonus is the ability to discover local effects (smaller patterns) within larger patterns as demonstrated by Jörg Reichardt²⁶. In short both directions are complementary and needed to acquire any language, although slightly more emphasis might be put on hypothesis driven language acquisition models.

7.2 Developing (Embodied) NLU systems

“Nana Korobi Ya Oki – fall 7 times, rise 8”
Japanese proverb

This proverb hits the spot when developing NLU systems. More then often an approach does not work, it might not scale up or its performance is not what expected. Nonetheless many scientists and engineers keep trying to find practical NLU systems. With the succes of Watson, past failures finally paid off and many more succesfull system will certainly arise. This section will discuss the general architecture NLU systems use, its components and most importantly how these components can be implemented using computational techniques.

7.2.1 General NLU Architecture

A NLU system usually processes natural language input to some knowledge structure output. Surprisingly, no general architecture was found in literature. By looking at the architecture of the five NLU systems discussed in this chapter, a general architecture can be found as displayed in Fig. 7-1. A semantic theory is not visable in the architecture but is affecting both the interaction flows (making the flow not neccessarily serial and hierarchical) as well as the actual processing (parser, inference, etc).

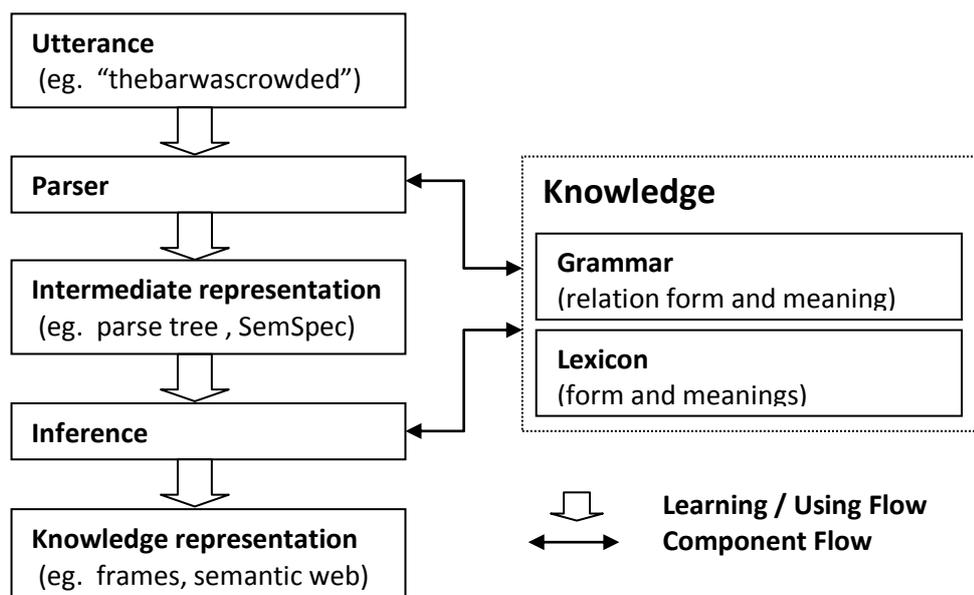


Fig. 7-1 General NLU architecture

²⁶ http://videlectures.net/cvss08_reichardt_ddhd/, last retrieved june 2012

Many derivations have been constructed and links to the general architecture can often only be detected by close inspection, for example Fig. 7-2. This architecture mostly zooms in on the mechanics of the parser except for “Tokenization” which can be thought off as a pre-parser process. Another derivation is the NTL architecture devised by (Chang & Mok, 2006) in Fig. 7-3. Here some different terminology is used but context fitting using “World and Linguistic knowledge” is basically the same as parsing using a lexicon and grammar. In this architecture more components have been added such as discourse and simulation for aiding the inference engine with disambiguation and learning.

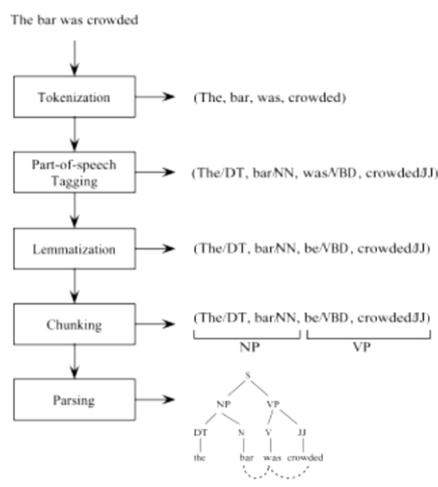


Fig. 6. An example of preprocessing steps of text.

Fig. 7-2 Example of NLU architecture (Source: (Navigli, 2009))

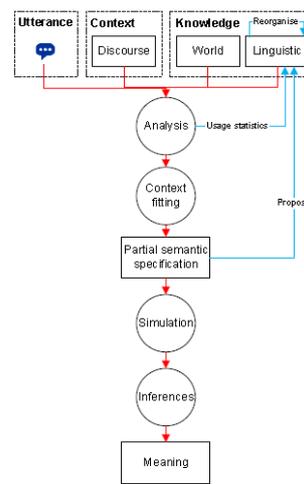


Fig. 7-3 Example of NLU architecture (source: edited from (Bergen & Chang, 2003) and (Chang & Mok, 2006))

7.2.2 General Components

Any NLU system will try to parse a natural language input sentence into some knowledge representation structure, such as the Frame structure, with possibly some intermediate representation such as a parse tree or NTL’s SemSpec. This conversion process often includes the following components:

- **Lexicon.** The dictionary, a collection of words/symbols that can exist in a language. Complex or rich lexicons also refers to the meaning (sense) of the lexeme (word) and allows multiple meanings for a single lexeme, possibly including a suitable ontology (like in the WordNet lexicon). These meanings are “objective” (disembodied) and culture bound (see chapter 3 on truth).
- **Grammar.** Rules (patterns) to break the sentence into an internal representation. Traditional grammars have expandable, non-terminal rules and non-expandable, terminal rules (Jurafsky & Martin, 2000).
- **Parser.** Procedure (often called search strategy) to segment the sentence, such that all possible structures that can be derived from the input sentence are found. The procedure involves matching the lexemes from the lexicon (often called tokens) in a sentence (often called token stream) using the rules from the grammar. Its output is one or more structures (such as a parse tree) since one or more structures (read: meanings, or *possible (even partial) meanings* = hypotheses) can be found. These structures range from abstract meanings to very detailed ones. While human excel at selecting the right

abstraction level of a meaning under time stress, computer typically have a fixed abstraction level. Traditional parsers have two search strategies to segment the input sentence: top-down (eg. Earley parser) and bottom up (eg. CYK parser) (Jurafsky & Martin, 2000).

- **Logical Inference.** While the parser is basically a dumb pattern recognizer or pattern search and matching mechanism, the inference engine is the actual “intelligent” mechanism. Its design is based on the semantic theory. A good engine would have two functions:
 - **Disambiguation (pattern selection).** Make a choice (disambiguate) between intermediate structures (if there are multiple structures) that is the closest to Type 2 Meaning. Disambiguation is a subfield of NLU and has many computational solutions (Navigli, 2009). Advanced disambiguation involves predictions and creation of additional structures not created by the parser.
 - **Pattern finding and learning.** While the disambiguation function would *predict* (and select the believed) Type 2 Meaning, the learning strategy validates that prediction and selection. One of the techniques involves comparing with the actual events that occurred, another technique involves creativity. The result is new rules and endpoints for the grammar and lexicon that are used for future parses and disambiguation. These new rules and end points can be learned using various learning strategies (see Appendix 2).
- **Knowledge representation.** The intermediate or internal representation (analogous to human thoughts) like the parse tree is used to extract patterns from it. These patterns are the actual discovered, learned and stored knowledge. Knowledge such as the lexicon and grammar can be stored in various knowledge structures such as semantic webs, frames, graphs and many others. The type of representation depends on the research question, in other words even if the human neural network representation could be fully replicated artificially, depending on the question, this black box representation might not provide insights. The human knowledge representation is known for its efficiency (reusing neuron circuits for multiple types of knowledge), not for its transparency.
- **Semantic Theory.** Theory on how the knowledge such as the grammar and lexicon are acquired, structured and used. Various theories exist (ie. different procedures) to guide the parser and inference engine to transform the input sentence into meaningful knowledge. The theories vary on what *meaningful* actually is (see chapter 3). Computational theories of semantics range from simple naive semantics and probabilistic enhanced semantic analysis to pragmatics (for example using embodied, usage-based thesis) to derive meaning from context.

7.2.3 Problems of traditional NLU systems

The traditional systems are too simplistic, often only performing acceptably in a very small domains and with limited use. Some key problems, but certainly not all, found in literature:

- **Problem 1: Modelling context.** Problem about what part of the world needs to be included in a knowledge representation and how to structure it, enabling meaning on the human level (*see the Frame problem*).
- **Problem 2: Learning meaning over multiple words.** Problem of learning meaning not mapped to a single symbol (word) but to multiple, such as “Last Christmas” meaning the song title of Wham! and not the past christmas.
- **Problem 3: Scalability of the system / Domain independence.** Problem of limited or no scaffolding on previously learned knowledge when switching to a new domains. New insights in metaphor structures and creativity might (partially) solve this problem.

- **Problem 4: Architecture.** Problem of static flows and serial processing, while stress due to time or other constraints can demand adaptive flows and parallel processing. For example, most NLU systems generate syntax parse trees independently (and often in serial) of the semantic parse tree. Humans do this in parallel, since syntax is an important cue for finding the correct semantics (Hagoort, 2005). Another problem with the architecture is that some components should be made more explicit such as motivation, learning, and the multiple stages of meaning from instinctive to fully analysed (simulated and logically processed).

7.2.4 Computational approaches to Language Aquisition

Learning a language according to the work of Tomasello (see chapter 6.2), is about finding patterns – in other words partitioning the sensed and perceived experience spatially and temporally to the usage customs of a language community. Such a *community constructed slice of experience* is immediately or possibly at a later moment (eg. deaf children that hear after surgery) extended with sound; the correlation with a sequence of phonemes and pauses – words. So learning the Form part of any language involves finding the correct partitioning and correlate it with the correct sounds (correct according to the language community). Learning the Meaning part of any language proceeds in parallel and shares a deep relation to the Form part. As mentioned many times before, words receive different meanings depending on context; some context is present (actual or assumed) in a situation and other times context is present by shared experience in the past or hinted to be received in the predicted and simulated future. This implies that the scope of the partitioning algorithms stretches over the spatial and temporal domain, as predicted by Problem 2 from the previous subsection. Understand the words “Last Christmas” requires understanding the whole of the utterance which is greater than the parts (isolated words). Knowing the correct sense of the loose words Last and Christmas wouldn’t evoke the correct meaning of the combination. The consequence of this realisation is huge, effectively destroying any theories on decomposability.

Many traditional approaches of learning language (such as the syntax-driven pipeline of meaning discovery, see Fig. 7-4) rely on the principle of decomposability, such that learning a language becomes learning two loosely coupled entities; a vocabulary and a grammar (eg. a Context Free Grammar or CFG). Using the rules from the grammar, words from the vocabulary can form sentences with meaning, such that reproduction of learned language and some creativity with generalized rules is possible. Of course these traditional approaches face many problems (eg. CFG are inflexible and only capture meaning on word level, see (Jurafsky & Martin, 2000) for more) and deviate a lot of how humans acquire language. It is important to mention that the definition of “meaning” as is used in traditional approaches differs from the one used here. Their definition is not embodied and is more like an abstraction, such that I prefer to call it “context free meaning” (CFM). For example “cook” is in terms of CFM a verb and a noun, but has no reference to image schema’s of the visual experience of a cook or the physical experience (eg. emotions that arise) of cooking.

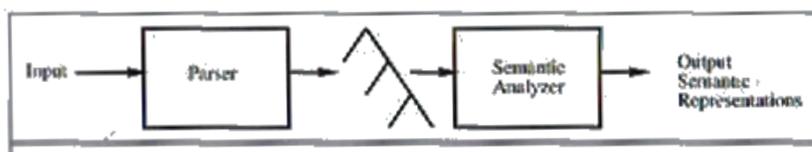


Fig. 7-4 Syntax driven NLU architecture, serialising syntax and semantic analysis (source: (Jurafsky & Martin, 2000))

Others tried approaches using semantic grammars (Brown & Burton, 1975) that capture the non-terminal rules, to aid the parser in its WSD. Since the terminal meaning is still based on the principle of compositionality, Problem 2 is not solved. Related to semantic grammars is the semantic web, a network of associations or semantic relations. While the semantic web as a representation structure is mentioned in 7.2.6, as a language acquisition techniques various approaches are tried (Van der Plas, 2008). Syntax based (eg. *verb* – play), alignment based (eg. synonymes and 1-to-1 translations such as: *pokoe* – music) and proximity based (*party* – evening, dancing, drinking, etc) methods all yield different results for different tasks, because each technique describes a different kind of relation.

Discussion

In short, all of the problems discussed in section 7.2.3 apply on the traditional approaches. A new type of approach is needed that uses all available structure (sensed and perceived) to aquire language. Such an approach would exceed a simple fusion of promixity and syntax based techniques requiring more structures available such as emotions and tense. Again emphasis is put on the usage-based view; language needs to have function for it to be a discriminatory from non-language. Also, the approach needs to allow constructions that can be used for both production as well as recognition. The approach requires solutions for the four key processes and four grammar stages described in section 6.2.

(Han & Kim, 2010)' *intention reading system in Fig. 7-5, might be a good start in the right direction. Their system was inspired on the mirror-neuron system and simulation hypothesis, such that predicted behavior (enhanced by simulations) can be related to certain language behavior. The algorithm involves solutions for key concepts: dealing with uncertainty, motivation (interest), goal creation and selection and learning from human actors through interaction. Still, this is only a start since many of their algorithms are not bio-inspired in its implementation, posing "some" limitations such as scripted behavior.*

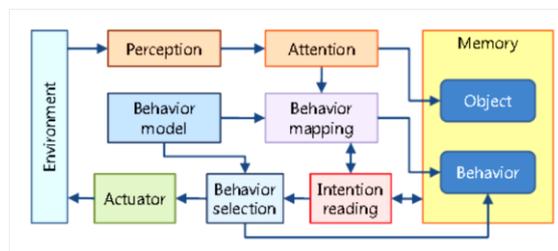


Fig. 7-5 Intention reading system within overall cognitive architecture (Source: (Han & Kim, 2010))

7.2.5 Computational approaches to Language Structure

Many types of structure exist to describe experience including their connections, a field called Knowledge Representations (KR). Connectionist approaches such as neural networks (NN), the computational model of our brain, are not yet sophisticated enough to be usable. Furthermore, NN are a black box and insight in NLU processing is just as interesting as discovering the most realistic structure. One of the more transparent approaches is the frame-based KR devised by the founding father of the NN, Marvin Minsky (Minsky, 1974). Frames contain slots that contain either values, references to other frames or methods for aquiring values. Embodied Construction Grammars, as discussed in 6.1, are based on Frames. Many other KRs exist such as the Semantic Web, Resource Description Framework (RDF) and others.

Knowledge structures have requirements in order to be useful for computational semantic processing, which according to (Jurafsky & Martin, 2000, p. 541) are:

- Determine the truth of propositions to support unambiguous representations
- Represent variables
- Support inference
- Sufficiently expressive

In addition, to better represent human knowledge structures according to the Embodied view, they need to be:

- Self-motivated (Barto, Singh, & Chentanez, 2004)
- Grounded in embodied primitives directly or indirectly (using metaphor or metonymy structures)

Discussion

(Jurafsky & Martin, 2000, pp. 513, 538-539) *mention that the semantic web and the slot filler frames mentioned above are often used knowledge representations by NLU researchers, however First Order Predicate Logic (FOPL) knowledge structures have the advantage of being flexible enough to represent the knowledge from the previous two **and** provide computational basis for verifiability, inference and expressiveness requirements. I think that the benefits of the great readability of FOPL disappears quickly when considering huge patterns. For the purpose of modelling the proposed experience tree I recommend using the frame-based KR since it is well scalable, transparent and user friendly by viewing experience on different abstraction levels, to allow better support for small and large patterns. I also think that connected frames have the optimal bio-inspired brain structure realism vs. transparency trade off.*

The requirement of variable or value representation is interpreted for frame-based KRs to be flexible by allowing expansion in size (its properties and their respective values), thus allowing virtually unlimited connectivity (neurons can have up to 10.000 connections to the most remote brain areas, but that number is likely an efficiency reason). I discussed earlier my dislike for the term rule and instead prefer patterns since rules change slightly due to a multitude of reasons. Therefore I like to call it flexible frame-based KRs or more appropriate pattern-based KRs. Depending on implementation ECG can be a pattern-based KRs.

Instead of "to represent variables" I prefer "to represent patterns, references to patterns or methods for acquiring patterns" since only modelling variables ignores modelling the important relationship between the variables. Furthermore, additional requirements are needed in order to let the same KR be used for both permanent knowledge representation (memory) as well as temporary (thoughts).

7.2.6 Computational approaches to Language Usage

In chapter 6.3 possibilities were discussed of how humans form thoughts and structure them into (mental or spoken) language relying heavily on personality and reasoning and use the indexical and simulation hypothesis as key assumptions. The notion was coined that humans are inherently motivated to pose hypotheses to deal with uncertainty, to understand and thus predict sensory input, interaction and feedback with the world, in order to better survive. To answer or solve these hypotheses evidence from past experience is gathered - learned knowledge. Together with some other constraints this type of reasoning is called *human reasoning* (see 6.3), or *evidence based reasoning*. Language usage fits in as a (very effective) tool to support the interaction with the world, both as a way to scaffold thoughts in the "inner world" and as a way to express thoughts in the "outer world". Reasoning functions as the process for thought manipulation.

The computational counterpart to human reasoning is a field called *case based reasoning*. This field developed by Roger Schanks in 1982 (Schank, 1982), uses past experience (“the cases”) to solve new problems or hypotheses. Technologies such as IBM’s Watson (evidence based reasoning, see 7.4.2) is a type of reasoning within the case based reasoning framework. Many more types of automated/computer reasoning exist, some not case based or even logical at all (eg. imagine modelling human reasoning affected by alcohol).

The case based reasoning framework fits well in the NTL framework and has been formalized for computational reasoning by (Aamodt & Plaza, 1994):

1. **Retrieve.** Given a problem, retrieve from memory cases to solve it. Human use various strategies to retrieve, such as focussing on a narrow range of cases or selecting the first case that is good enough, see (George, 1980) and (Piatelli-Palmerini, 1996). Personality plays a huge role in selecting a retrieve strategy.
2. **Reuse.** Map the solution of the retrieved cases to the problem. One group of (meaning discovery techniques) that do this is grouped under the umbrella of Word Sense Disambiguation (WSD).
3. **Revise.** Simulate (simulation hypothesis) if the mapped solution is an actual solution (“has the desired meaning”) or otherwise revise (affected by personality) until it does or return to step 1 by retrieving more cases, so digging deeper in memory or step 2 by reevaluating new or existing solution-problem mappings.
4. **Retain.** If the mapped solution is thought to be an desired and thus an immediate or eventually usable solution (indexical hypothesis) to the problem, store it as a new case (new experience).

Traditionally much focus has been given to the second step, linking stored data to input data to find a solution or meaning. One of the open research problems in this step is finding the correct amount of word senses in a language community, which is believed to be an AI-complete problem (Mallery, 1988). (Navigli, 2009) presents an WSD overview targeting this problem, discussing many supervised and unsupervised pattern recognition techniques that can potentially be a good starting point. However note that some or most of these techniques are not necessarily modelled after human WSD practices. Two broad categories can be distinguished in WSD, deep and shallow. Since deep approaches require some sort of domain independent knowledge including a reasoning engine (discussed later), most focus is given to shallow ones. Often a window of n words around each word to be disambiguated in the corpus is used, hence the name *shallow approach*. Four general (both shallow and deep) approaches exist ranging from “no truth discovery” to “full truth discovery”:

- **Dictionary- and knowledge-based methods.** These consist of dictionaries, thesauri, and lexical knowledge bases that are all preconstructed without using any corpus evidence (or other types of own “experience”).
- **Supervised methods.** These make use of sense-annotated corpora to train from, meaning that the truth discovery can be traced, but is due to sense annotation from a third party inherently based on experience from others. Methods include Naive Bayes classifiers, decision trees and kernel-based methods (Support Vector Machines, SVM).
- **Semi-supervised or minimally-supervised methods.** These make use of a secondary source of knowledge such as a small annotated corpus as seed data in a bootstrapping process, or a word-aligned bilingual corpus. Truth discovery is large based on own gathered experience.

- **Unsupervised methods.** These consist (almost) completely of external information and work directly from raw unannotated corpora or other types of experience. These methods are also known under the name of word sense discrimination/induction, or the automatic identification of a word sense.

The holy grail in WSD are the “deep” approaches in which a large body of domain independent (= “world”) knowledge is available and some sort of reasoning engine to deal with it, such that no fixed window around a target word or words is needed. The difficulty of modelling world knowledge can be seen in the following example from The Mind Project²⁷

- 1) *I read a story about evolution in ten minutes.*
- 2) *I read a story about evolution in the last million years.*

To understand this example correctly one needs to have general knowledge of human life spans, reading speeds, the theory of evolution and in addition some mental capabilities such as reasoning with simulation to sort this out. Currently, the *common knowledge problem* as this problem is called, is unsolved since it is unknown how to structure, store and process world knowledge such that machines can work with it. Eventually a machine should understand that (2) is very likely to be taken a bit sarcastically (hidden expression of emotions!) since reading a story can only be done within the average 80 years of human life span. Many other interpretations of (2) are also possible depending on context, such as it being a funny remark or actually true if the character saying it is immortal (for example in a film). However the key point of the *common knowledge problem* is that even without explicit or additional context some likely interpretations (including assumptions and expectations) can be extracted simply because common knowledge is implicitly assumed and used as context.

Discussion

“Language usage” compared to “language acquisition” and “language structure” is quite possibly the toughest research area since it integrates all practices of the latter areas. Having an hypothetical embodied grammar-dictionary full of patterns of words, word pairs, triplets or even small or large chunks of sentences with probabilities is great. Having this grammar-dictionary as a machine readable structure would be the next best thing. But combining this body of language experience with modeled mental capabilities such as reasoning (including various levels of creativity) and a personality (with beliefs, desires and goals) that drives the reasoning is in our opinion another league of its own. Traditional computational approaches to reasoning and personality are often static and rule based (“scripted”), meaning that its creative behavior is tied to certain domains (see Watson, Protothinker and XPERO in section 7.4). This limitation also holds for humans, since creativity is only possible within the confines of human embodiment. Reality or fantasy is limited to what our brain can imagine/compute. Furthermore, reality in contrast to fantasy is constrained to the senses and actuators of the body. Despite facing the same limitation as humans, computational models are not up to par with humans, quite possibly because the techniques for mapping learned knowledge (through for example metaphores and metonyms) is not very advanced. Again this is an important pattern recognition problem that has huge implications for creating human-like creative systems.

Much literature related to language usage as shown for example by the overview of (Navigli, 2009) hints for predictable and thus low risk results with the help of external

²⁷ http://www.mind.ilstu.edu/curriculum/protothinker/natural_language_processing.php , last visit june 2012

structured and unstructured knowledge sources such as WordNet (Miller G. , 1995). This direction has room for new and improved results, however at some point the so called “data/knowledge acquisition bottleneck” problem, also called the “automatic acquisition of sense-tagged corpora” problem will arise. My ideal embodied machine should not rely on truth provided by external knowledge sources but by making its own knowledge source, its own truth. Language is highly dynamic, the sensed and perceived (embodied) experience make the correct sense very subjective and unique. The machine should discover the sense and its meaning/function for itself and within the language community. Amongst other, this requires solving the Frame problem discussed earlier, which might contribute to solving the common knowledge problem. In turn, this might help in finding a “deep approach” and promises a new chapter in human-like language learning and usage.

It seems that traditional research focusses much on the receiver part, but as mentioned earlier meaning discovery is a communication problem - something or someone is sending out meaning. Interpreting that meaning by assuming it from ones own perspective is one thing but to discover the intended meaning by something or someone is something else, it requires some sort of feedback. This feedback can be explicit (eg. “no it is ball, not wall” or implicit (eg. data driven, by seeing many correct examples). Just as important as in real life, the tutor is required to provide correct teachings of situations and word usage so that the correct meaning can be learned. Being consistent, accurate, aware of the limitations of the system (eg. Motherese input since the short term memory is not full grown yet), correcting incorrect usage by feedback, provide many correct examples and stimulate correct usage will dramatically increase meaning accuracy and learning rate. It is common for young infants to overlearn (generalization) and underlearn (specialisation) and in general improves to the correct amount of word senses with more experience. This is the power of human learning, there is no need to comprehend the full utterance at once (“ie. comprehend all the possible meanings of the utterance at once), but rather understand the function it its used context or even partially understand the function (understand some aspect of the utterance) .

7.3 Benchmarking and validating artificial NLU systems

Comparing and evaluating results with other researchers is essential to remain critical at the direction NLU is heading. For that, benchmarks and standardized experiments are required. The Loebner Prize (Turing Test) is a benchmark, however it is very much a subjective, qualitative one. Its difficult to say this years winner has improved with some % at some specific area compared to winner of the previous year. Objective, quantitative benchmarks are needed, but are not trivial to set up. First and foremost is the need to breakdown intelligence in specific areas, a question deliberately omitted by the creation of the Turing Test.

Despite the difficulty, serious effort in this quantitative direction is put up by the SemVal (former SenseVal) workshops (see fig. 7-6). Each couple of years new tasks are conceived including data sets and “golden standards”. NLU systems (or parts, like semantic analysers) are given the task and the results are compared to the gold standards (the task done by humans). The results and approaches are discussed in the workshop. Each year the amount of tasks seem to increase, for example for the 2012 workshop in Canada, semantic textual similarity is part of the contest, which is the good step in the direction of contextual word sense analysis – and away from the principle of decomposability.

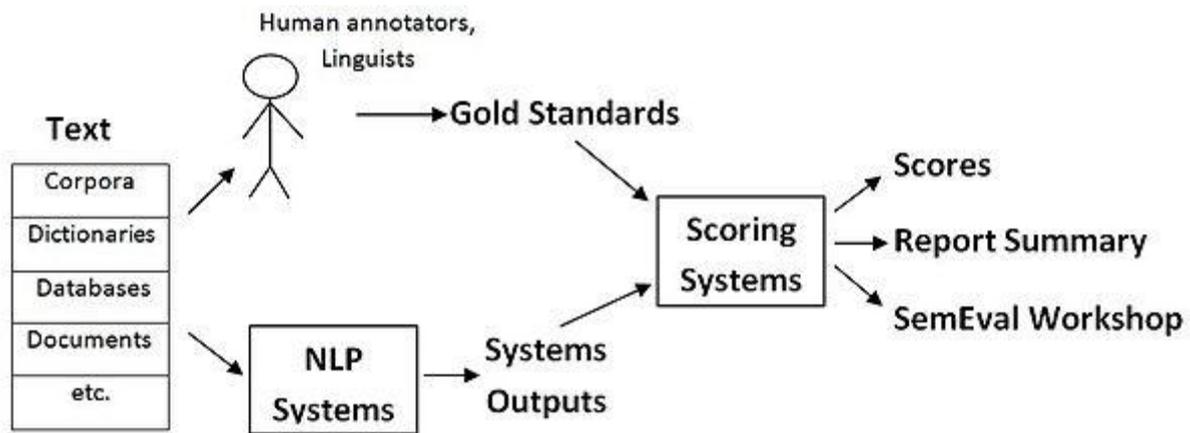


Fig. 7-6 Evaluation of NLU systems at SemEval (source: <http://en.wikipedia.org/wiki/SemEval>)

Another quantitative approach, one that does not rely on the weighted, but subjective gold standards is comparing results with those objectively acquired from large bodies of text. For example from the freely available google corpus²⁸ 1 billion (US: trillion) English words and near 100 million sentences, n-grams and other statistics can be retrieved. Or instead of comparing statistics, ontologies can be compared for example to those of Cognition’s semantic map²⁹ which is based on WordNet and has over half a million basewords.

Extensive quantitative benchmarks are needed –full realism is not needed – to push the NLU field further, possibly bootstrap it to a level where qualitative benchmark (the psychological domain) is justified and usable. The quantitative approach chosen by SemVal is in line with the usage based approach; create a huge set of simple and complex scenarios (the world) and a huge set of tasks (interaction) and use human behavior as the benchmark. Dataset are available for any researcher and for free, although producing a state of the art NLU prototype likely involves license fees for using large corpora’s such as the Brown Corpus, SemCor, WordNet, etc. Each couple of years the datasets are extended to stimulate researchers to produce adaptive prototypes, forcing them to hard code behavior in scenarios only to the level of humans (like reflexes, instinctive behavior).

CHILDES Corpora

A great body of text specific for NLU systems targeting child directed speech is the CHILDES corpora (MacWhinney, 2000). This text resource contains conversations between child and adult (often in so called motherese or parentese format) in many diverse situations such as a block tower construction task. The data set is constructed to be an ideal starting point for realistic input for a child simulation NLU system.

7.4 Existing NLU Systems

Many NLU systems have been built in the past, some more successful in terms of media exposure than others, some only in theory and again others where only parts of a complete NLU system. As mentioned before none of these NLU systems was successful enough to be commercially viable or even practical outside the laboratory. The only exception, is the very recent IBM’s Watson, who has already proven to be an accurate and robust NLU system. In this subsection various NLU systems will be discussed, some complete and some just key parts of the larger system.

²⁸ <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13>

²⁹ http://www.readriteweb.com/archives/cognition_semantic_map.php

Although some structure is applied in analysing the systems, no in-depth analysis and fair comparison (to the human) is done. The systems were primarily chosen for its unique features in the quest to analyse the NLU problem and be inspired for the ideal NLU system.

7.4.1 ITALK

Background

The ITALK³⁰ (Integration and Transfer of Action and Language Knowledge in Robots) project is a EU funded (7th Framework) project and aims to “develop artificial embodied agents able to acquire complex behavioural, cognitive, and linguistic skills through individual and social learning”. Their objectives include the development of new theoretical insights, such as models, explanations and engineering principles and has the Embodiment hypothesis and usage-based language acquisition at its basis. Over 100 publications have been delivered on the subject in three years. One of the earliest papers in this project by (Saunders, Lyon, Nehaniv, Dautenhahn, & Förster, 2008) is used for this subsection on existing NLU systems since it discusses experimental approaches to engineer a complete NLU system that tries to mimic human infant language acquisition.



The system has many NLU components mentioned in this report and based on similar Embodied assumptions and similar methodologies mentioned here. According to (Saunders, Lyon, Nehaniv, Dautenhahn, & Förster, 2008), reaching the holophrasis stage encompasses three stages:

- **Phoneme and holophrasis acquisition.** The program or *agent* starts with knowledge (“the lexicon”) of phonemes with equal probabilities attached. The agent learns these probabilities while hearing speech and tries to segment the utterances by extracting the phonemes. Checks (one of them using human interaction and feedback) are done to see if the sequence of connected phonemes segmented for an utterance form a word.
- **Grounding speech: action and object learning.** The agent looks for clusters of association occurring between the learned words and other modalities (sensor experience), such as sensorimotor readings over a time window and visual patterns. It uses variants of the Crutchfield-Renyi information distance measure to detect these associations across different modalities, which effectively will lead to holophrase constructions (words associations with meaning, defined as actions, objects and interactions).
- **Speaking and interacting.** The agent vocalizes learned words or sequence of phonemes to provoke human feedback which hopefully leads to scaffolding further linguistic competencies.

Technology

The total NLU system is called LESA (Linguistically Enabled Synthetic Agent) and includes (but is not limited to), the following technologies and data sets:

- ARToolkit (vision system for object recognition)
- Esmeralda (speech recognition system)
- CHILDES corpus (motherese input)
- SCRIBE corpus (phonemes)
- ROSSUM (social learning architecture)
- iCub and KASPER (humanoid robots)
- eSpeak (speech synthesis)

³⁰ <http://www.italkproject.org/>, last visited june 2012

Advantages

- Based on the Embodied, usage-based approach
- Subjective truth learning through both motivated actions and probabilistic analysis
- Supervised and unsupervised learning including feedback-enabled learning
- Usable for any language, if appropriate phoneme lexicon
- Offline and mobile (the robots is independent)

Disadvantages

- No logical inference engine (reasoning system), such that Tomasello's Schematization and Analogy and Entrenchment and Pre-emption skills (see chapter 6.2) are not available (eg. no generalisation and metaphor learning)
- Small domain, due to missing of logical inference engine (no abstraction can be made nor new knowledge by combining previous learned knowledge)
- Phoneme lexicon is provided, not learned

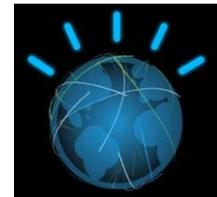
Closing thoughts

THE NLU system by Saunders and others to reach the holophrasis stage is still in its infancy (or as they phrase: the road is very long). The ITALK project at large has good funding, reknowned reseachers and an ambitious roadmap³¹ pursuing a bio-inspired approach. The foundation however is still not solid and a more integrative approach (more disciplines need to cooperate) is needed. Also further advances should include an inference system, more advanced sensors and actuators and less focus on hardcoded knowledge (such as the phoneme lexicon) and more focus on self-discovered flexible pattern based knowledge.

7.4.2 Watson

Background

IBM's Deep Blue shook the world in 1997 by defeating the world champion Gary Kasparov in a game of chess and recently in 2011, IBM's Watson defeated the top two players of *Jeopardy!* a language understanding game. This latest achievement is a milestone in computational intelligence. Watson is able to analyse the question in natural language, decompose it in several hypotheses, find support for each hypothesis and merge and rank it 70% of the questions with 80% precision in under 3 seconds³² (also see Fig. 7-7).



Technology

The technology behind Watson is complex, but not revolutionary. It uses extensively trained models and both machine and human structured sources. In that sense Watson is a *intelli-tuned* machine instead of an *intelligent* machine. The approach used by Watson does show similarities to a human approach, using evidence based reasoning. However its decomposition in a multitude of hypotheses that are not revoked until the final scoring and the serial flow to reach the final answer is not how human inference works. This strategy does exploit the computational prowess of the mainframe and allow unique capabilities for the machine. Its lead architect David Ferruci in a TED talk mentions that "*Watson was not build from the perspective of understanding of how the human mind works ... but is inspired by human cognition, at least the kind we are conscious of*"³³

³¹ <http://www.tech.plym.ac.uk/SoCCE/ITALK/publications/roadmap-year2.pdf>, last visited june 2012

³² <http://www-05.ibm.com/ch/fr/watson/watson.pdf>, last visited june 2012

³³ <http://www.ted.com/webcast/archive/event/ibmwatson>

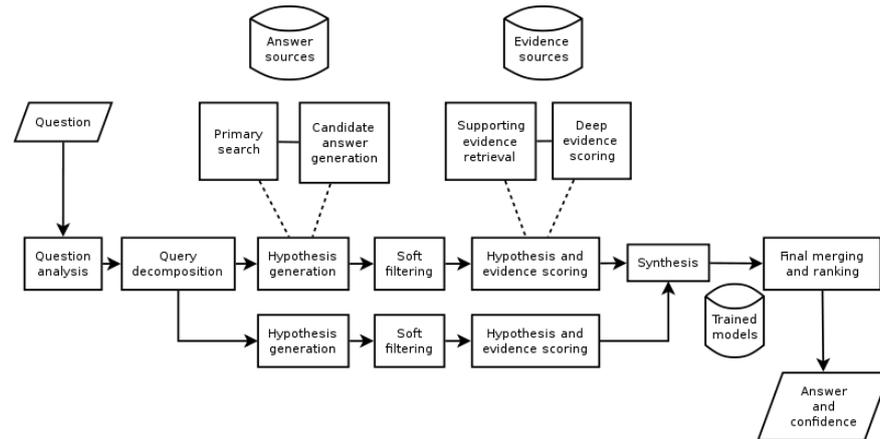


Fig. 7-7 Watson's architecture (source: (Ferrucci, et al., 2010))

Watson is able to access 200 million pages of structured and unstructured content including dictionaries, novels, news stories and the full wikipedia text. It uses techniques from NLP, machine learning and IR. Its “magic” lies in the evidence scoring component or analogous to the human brain, the truth assignment component, which relies heavily on redundancy exploitation for scoring. Also a machine learned, variable confidence threshold is key in Watson's intelligence.

Advantages (for Q&A NLU tasks)

- High Accuracy and Confidence
- Speed (highly scalable with more processors)
- Broad domain using (un)structured sources
- Flexible knowledge representation
- Offline, Mobile (4TB of content and 6 million logic rules are stored in RAM)
- Roughly bio-inspired

Disadvantages

- Questions limited to facts not opinions
- NLU is limited to Q&A domain (only Type 1 and Type 2 understanding)
- Unable to build (extend) on own knowledge unsupervised
- Hardwired inference engine (6 million logic rules)

Closing thoughts

John Searle³⁴ states that Watson is unable to understand its own answer using his famous Chinese Room argument. I refuted this argument earlier by mentioning that if human understanding cannot be distinguished from computational understanding then what does it matter – maybe humans are also advanced computers. I do recon that, in a sense Watson is so much designed, trained and optimized for the Jeopardy! task and knowledge domains that he is unable to discover his own subjective truth. Instead he relies on the redundancy principle which occasionally leads to very unhuman like answers, since humans rely on many other principles in their truth assignments and often pursue personal goals and beliefs instead of (weighted) truths of others.

With its succes over its human opponents, one might wonder if a more bio-inspired NLU approach is desired. If bio-inspired design was the only school of design, humanity would end up in a world for example, without dishwashers and instead with handwashing robots or with ornithopters instead of airplanes. Time will tell which school serves humanity best.

³⁴ <http://online.wsj.com/article/SB10001424052748703407304576154313126987674.html>, last visit june 2012

7.4.3 XPERO

Background

Funded by the 6th Framework Programme for Research and Technological Development, the XPERO project³⁵ is in many ways the predecessor of the ITALK project. The project took place right before ITALK (between 2002 and 2006) and is more generic in its approach, it does not focus on language specifically. According to their website “*the overall objective is to develop an Embodied cognitive system which is able to conduct experiments in the real world with the purpose of gaining new insights about the world and the objects therein and to develop and improve its own cognitive skills and overall performance*”. Their result led to models, techniques and solutions enabling an autonomous Embodied robot and to awareness raising and knowledge sharing initiatives.

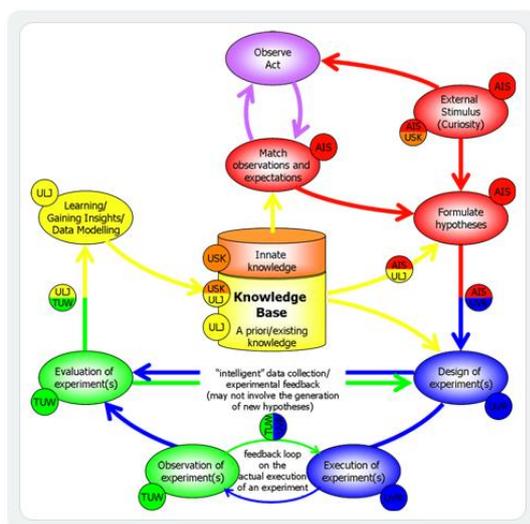


Fig. 7-8 XPERO Human inspired architecture (Source: xpéro.org)

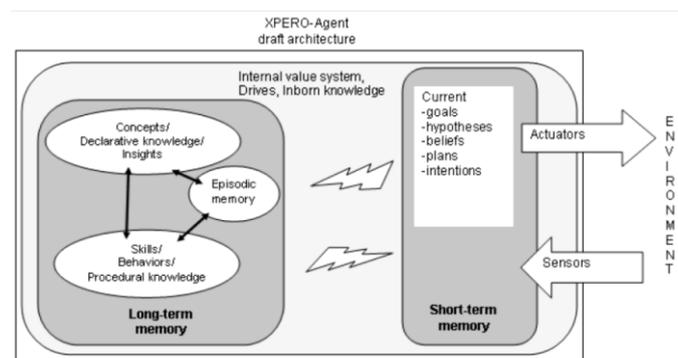


Fig. 7-9 Simplified XPERO human architecture (draft) (Source: xpéro.org)

Technology

The main approach of XPERO to develop an autonomous Embodied robot was to develop a methodology for learning by experimentation and use that knowledge in conjunction with a motivation component to enable goal driven behavior. The knowledge structure consist of Piagetian-like schema’s (they call them “proto-objects”) with both the assimilation and accomodation properties to capture not only a coherent experience but also build/extent on it. The motivation component plays a fundamental role in these schema and are implemented in the schemas to become self-motivated to be executed (Barto, Singh, & Chentanez, 2004).

Again, like the Watson approach, the system is continuously posing hypotheses to explain and test the environment by predicting and matching expectation, see Fig. 7-8. The architecture is not as computationally elaborate compared to Watson but is more human-like. An abstraction of the architecture is presented in Fig. 7-9.

Advantages for NLU tasks

- Dynamic structure (able to bootstrap cognition by assimilating and accomodating experiences including language)

³⁵ http://www.xper0.org/portal/viewpage.php?page_id=23, last visited june 2012

- Autonomous, motivation based (including personality with curiosity and self-preservation traits and a desire to explain unexpected phenomena)
- Able to perform stochastic and planned experimentation (for acquiring knowledge)
- Internal value system (subjective truth discovery)
- Has a simple body to manipulate environment

Disadvantages

- no skills/mechanisms to acquire human language (eg. unable to read intention)
- no skills/mechanisms to use human language (eg. simulation)
- no social skills and body to naturally interact with humans

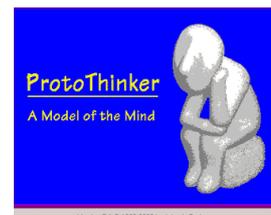
Closing thoughts

The XPERO project is a wonderful milestone for bio-inspired computational intelligence. It is generic, so it is possible to implement NLU technology. Furthermore, it is actually build on Embodied AI principles like cheap design and redundancy. Not all limitations of the developed technology are discussed, since its flaws are rather trivial and not meant to be fully human-like.

7.4.4 Protothinker

Background

The software package Protothinker (PT)³⁶, developed by John Barker, is a teaching tool for classroom purposes designed to stimulate thinking about human thinking. Using interaction with the program – an artificial mind – the user learns more about the mind, thinking, understanding, logic, belief, reasoning, language usage and other related topics.



Technology

PT is an expert (template matching) system build on Prolog, so its knowledge (belief) structures are simple frames and its reasoning or inference capabilities are based on the logic processing of Prolog. Its “behavior” or verbal responses are hardcoded. Prolog limits the system severely since processing is done in 2-valued logic (true or false), while human reasoning allows 2-valued logic + *unknown* when processing illogical premises. For example “will it rain tomorrow?” cannot be processed since the *unification* algorithm that lies at the heart of Prolog is unable to assume premises; the relation between premises must be either *identical or not*(syntactical unification) or *equal or not*(semantic unification) and therefore cannot be *unknown*. Since PT is supposed to be an AI mind, its behavior is programmed to deal with uncertainty in a natural manner by replying “*I don’t know the answer. I could use some help on this.*”.

The PT algorithm decomposes language input by first parsing it syntactically (according to learned grammar rules) then semantically (by doing a dictionary match) and finally a cognitive parse (by glueing them together, which is not actual processing but is useful for finding a proper behavioral response).

Advantages for (for Q&A NLU tasks)

- Prolog is a declarative programming paradigm –covering some parts of human logic processing
- Able to process most types of dialogue acts (statement, command, question, request)
- Learning tool

³⁶ <http://www.mind.ilstu.edu/curriculum/modOverview.php?modGUI=194>, last visited june 2012

Disadvantages

- Questions limited to “facts” (only 1 answer correct)
- NLU is limited to Q&A domain (only Type 1 and Type 2 Understanding)
- Unable to build (extend) on own knowledge unsupervised
- Hardwired inference engine
- No body
- poor WSD
- poor handling of uncertainty

Closing thoughts

The program is quite extensive in its functionality; it is able to show the AI's thoughts, reasoning pattern, memories and so on. As a learning tool it suffices but not for research purposes, due to disembodiment. It should be noted that the spinoffs of this project involve robots, but it appears that the project at large is halted since it is unupdated since 2006.

7.4.5 NTL

Background

While NTL is actually a collection of theories on language (see chapter three for more background), some have been implemented to become (parts of) NLU systems. Important foundational work has been done by Terry Regier (Regier, 1996), whose model in Fig. 7-10 allowed learning spatial relation words (universal concept) independent of language, but with some constraint on the language (see (Feldman J. , 2006, p. 158)). David Bailey (Bailey, 1997) continued on Regier's work, solving some limitations such as learning from a single example. His model for learning action words is presented in Fig. 7-11.

Regier and Bailey avoid the need for a grammar by directly associating a single word with one or more embodied schemas (x-schemas). Understanding language in these systems thus becomes very limited. Implementing solutions for the components *inference* (using metaphores) and *simulation* as done by Srinivas Narayanan (Narayanan, 1997), see Fig. 7-12, allowed for more complex understanding. His initial work used preanalysed sentences to avoid a grammar theory. Nancy Chang (Chang, 2008) and later on Eva Mok (Mok, 2008) who continued on the work of Chang developed a grammar theory, the Embodied Construction Grammar (ECG) as discussed in Chapter 6.1. John Bryant (Bryant, 2008) developed a parser that analyses the ECG including deep meanings and flexible forms. One of the significant features of this parser is that it can analyse ECG that cannot be fully analysed based on the given grammar, for example for understanding less than perfect sentences. This is important because children also don't always hear perfect sentences. With a grammar theory and a parser NTL theory can now be extended to become a complete NLU system. A prototype has been built, but many challenges to scale up to a practical system remain, in the words of Feldman (Feldman J. , 2006, p. 306) *“to scale this program up to a practical system for understanding English would require solving several hard problems in linguistics, knowledge representation, inference and computer science.”*

Technology

Regier's Model

The technology in Regier's Model is based on a neural net with a simple visual system to recognize two objects in spatial dimension. Its learning assumption is based on the principle that if one universal concept (such as above) explicitly correlates positively with a word, the other universal concepts (such as below, left, etc) all implicitly correlates negatively with that word. The neural net is trained using a backpropagation algorithm.

Bailey's Model

The technology in Bailey's Model is like Regier's based on the human body having certain static features; parameterization to such a level that all properties of an action relevant for making distinction can be made. Using a neural abstraction technique called model merging where huge differences result in a new sense and slight differences fine tune the sense, the correct sense or senses for a word are learned even from a single example. Learning is done in supervised fashion, so only labeled data can be used and no hidden structures/patterns can be discovered.

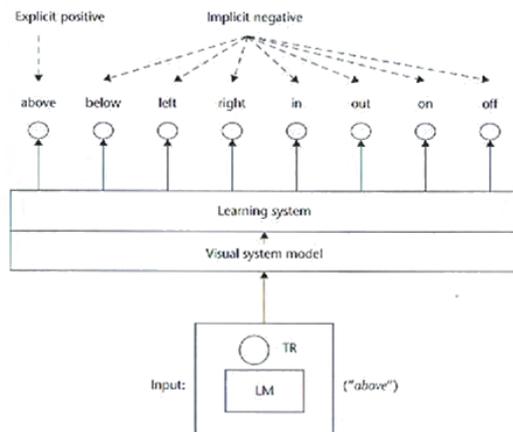


Fig. 7-10 Regier's model (source: (Regier, 1996))

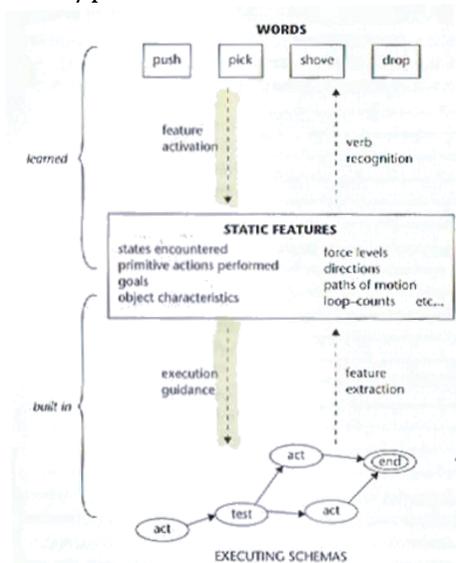


Fig. 7-11 Bailey's Model ((Bailey, 1997))

Narayanan, Bryant, Chang and Mok's Model

The technology in Narayanan's part of the model (see Fig. 7-12) is based on building a temporal belief network (using Bayes) that is updated using direct evidence or indirect via metaphor projections. These metaphor projections are based on some common knowledge (health and well being) and a universal concept (event structure metaphor) which in turn bind to embodied concepts like *forces*, *falling sick*, etc. Bryant's part of the model is based on building a parse chart (actually multiple parse trees) for each found sense. To cope with the large amount of parse trees that not lead to the desired sense, a new measure of semantic cohesion is used, a key feature of Bryant's Parser, using three kinds of estimated parameters to predict three different parts of the ECG. Although not proven with a biological link, the parser, like humans, does require more time for finding the low probable and correct sense of a word (the human heuristic is called *surprisal* – how surprised one is with the outcome = correct parse) measured using an entropy calculation (Bryant, 2008). The part of Chang and Mok is discussed in section 6.1 and 6.3 and in short is based on Fillmore's concept of structuring language using constructions.

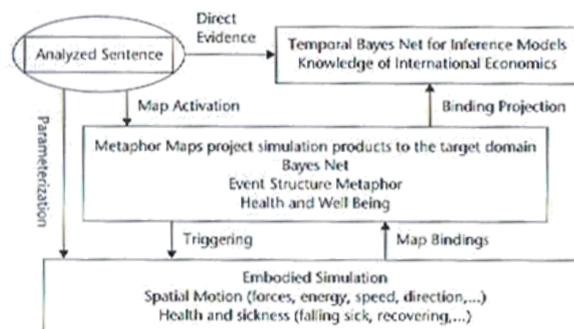


Fig. 7-12 Narayanan's Model (source: (Narayanan, 1997))

Advantages

Regier's Model

- Based on a simplified neural network with backpropagation
- Able to learn any language (although some restriction apply such as only single word concepts)

Bailey's Model

- Able to learn from a single example
- Able to learn any language (with same restrictions as Regier)

Narayanan, Bryant, Chang and Mok's Model

- Able to use simulation and inference using metaphores to achieve deeper meaning
- Able to learn from a single example
- Able to use a grammar and analyse incomplete (grammar wise) sentences

Disadvantages

Regier's Model

- Training of neural network requires many examples
- No grammar theory
- Restricted due to hardcoded concepts (only spatial relation words)

Bailey's Model

- No grammar theory
- Restricted due to hardcoded concepts (only action relation words)

Narayanan, Bryant, Chang and Mok's Model

- No grammar theory
- Restricted due to hardcoded concepts (eg. parse flows and knowledge representations)

Closing thoughts

The discussed implementations all rely on the usage-based Embodiment thesis. All discussed implementations are not yet practical for commercial NLU purposes. This has never been the goal. It does bring NTL theories into practice and the models are an excellent base to further extend to. No doubt they could eventually lead to full scale commercial NLU applications. Since NTL tries to incorporate all empirical evidence on language aquisition, usage and structure, new theories and implementations will emerge. This makes NTL a promising research area.

“Our difficulty lies in that we have built around ourselves conclusions which we call understanding. These conclusions are hindrances to understanding. If you go into this more deeply you will see that there must be complete abandonment of all that has been accumulated for the being of understanding and wisdom. [...] It is comparatively easy to renounce family, property, fame, things of the world; that is only a beginning; but it is extremely difficult to put away all knowledge, all conditioned memory.”

Jidda Krishnamurti, 1976

8 Conclusion and final remarks

In which the main research question is revisited and put in the context of future directions using the content of this report.

Up to this point, a significant overview and many novel ideas, viewpoints and interpretations based on recent literature are presented and discussed. Important questions that are touched are: “does language affect (ie. limit or enhance) understanding” (section 2), “how do meaning and understanding relate” (section 3), “what language development stages does a child pass” (section 4 and 6), and “are robots or computer programs capable of truly understanding human language” (section 7). However several other important ones remain: “to what extend is this report complete and true?” And with respect to the research question: “to what extend is this report helpful in the construction of NLU programs capable of processing human-like language understanding?” And finally for technophobists: “will solving the NLU problem, the holy grail of AI, really bring impending doom?”.

To what extend is this report complete and true?

As “promised” in the about section, this report is not a complete work on psycholinguistic phenomena. This report does unofficially aim for the honorable title of “*the ultimate starting point to build Embodied NLU systems*” and tries to be complete in that narrower sense. The essential tools, theories, approaches (including some novel ones) and supporting NTL framework are described including their biological background to get from data to theory to prototype NLU system and vice versa. With the extensive interdisciplinary reference list to large NLU publications an effort is made to be a good starting point for NLU reseachers from various fields; most importantly philosophy, linguistics, biology, psychology, neuroscience and computer science. The time to work together and integrate all knowledge from these fields is now.

The content of this report is written as objective as within ability with numerous, mostly scientific, references. However, like any other “objective” report, it is with no exception subjective (see chapter 3 on truth). Tracing the line of reasoning to see if similar assumptions and conclusions are reached (by reviewing the original source) is best practice when using this report. The NTL framework tries to fit all available interdisciplinary data and allows contradicting theories if and only if they do not contradict with that data. The framework relies on both weak and strong (falsifiable evidence based) assumptions (discussed throughout the report) and is work in progress. For example, recent work by (Falck-Ytter, Gredebäck, & Von Hofsten, 2006) support Tomasello’s theory (see Fig. 4-2) on intention reading of actions at 9-12 months by verifying that the mirror neuron system is active at that age and not before. Another example is by (Barsalou, 2008) who emphasizes that semantics is grounded in Embodiment.

To what extend is this report helpful in the construction of NLU programs?

Although this report might suggest that human level artificial NLU systems are imminent, this is not the case. Major hurdles still need to be taken ranging from better understanding of the

human brain to neurolinguistic models to computational models. I do believe that enough fundamentals have been uncovered for a solid (NLU) framework to start incremental work. Initiatives in physical and simulation NLU research for example using the Aldebaran's NAO robot or IBM's Watson shape a new wave of commercially attractive and scalable solutions.

New directions and approaches have been suggested for constructing novel NLU systems. NLU growth anticipation for the next 10 years can be found in the (ITALK, 2010) roadmap, one of the few somewhat realistic but still ambitious roadmaps in the Embodied landscape. In this age of crowdsourcing, gathering low cost but huge amounts of cross cultural data – language being amongst it – could speed up NLU growth. Other enablers are huge data sources like Google's and ubiquitous technology, such as browser technology and sensorperipherals like 3D webcams (Kinect) and HD microphones. Cloud-based (tele)presence has many advantages over physical (tele)presence. With this distributed initiative and this report constructing new Embodied NLU systems becomes more effective and more affordable to validate results which is clearly helpful.

What is beyond: into the grim post-NLU future?

The NLU problem is such an interdisciplinary problem that solving it will cause a revolution in the AI landscape. The rise of Artificial General Intelligence (AGI) or "human-like AI", made possible due to NLU advancements in reasoning, knowledge structure, learning and planning, will result in an explosion of highly practical applications. These applications will be adaptable to individual users, able to learn the unique patterns in behavior through detailed multi-modal interactions. This will bring about freedom of (trivial) decisions, for example on office days a random - but clean – suit is chosen by the computer. Small and larger decisions occur several hundreds³⁷ to multiple ten thousands per day³⁸. Automating the decision process affects day-to-day life, re-enforcing routines, for example because desired arguments are not or not properly, weighted in the automated decision process. Another effect of this level of automation is that people get plain lazy (Russell & Norvig, 2003). Many films and novels are devoted on this and other doomsenarios. A more positive spin can be thought off when local and global politics and industry make sure that advancements are properly guided. Rules and regulations – a formal framework - based on ethical implications on AI behavior and more importantly human operation (usage) of this technology. Several organisations put this topic on the annual agenda³⁹.

If the post-NLU age functions as a positive catalyst for human-like AI, then this will be a precursor towards technical singularity. Singularity is the age where technology advances so rapidly, that full comprehension of the technology will only be a achievable by a select few intellectuals and possibly even none. Technology that becomes available will entertain, heal and support humanity on sci-fi like levels, such as dream-like immersive virtual reality, extreme longevity and space exploration. The re-emergence of *the digital divide* (actually the knowledge divide) can be seen as a new challenge for adaptability (survival) of the human species. Adaption is not new, humans excel at it. Technophobists shouldn't fear singularity since key achievements of 3.7 billion years of evolution, cognitive (eg. creativity, "short term") and biological (eg. "long term") adaption, are hard to imagine being completely duplicated or even improved on in any distant future. Again, the only justifiable fear is harmful usage of technology by humans.

Scientists will become the new role models, rockstars, due to large influence on daily life. They will teach and raise machines to adulthood – a full symbiosis of man and machine. This symbiosis starts with shared understanding and thus requires fully or partially solving the NLU problem.

³⁷ <http://evidencebasedliving.human.cornell.edu/2010/03/so-many-decisions-so-little-time/>, last visit june2012

³⁸ based on average day with 16 waking hours and unrealistic high assumption of 1 decision per second

³⁹ eg. the Association for the Advancement in AI (AAAI) and National Science Foundation (their NBIC workshop)

References

- Aamodt, A., & Plaza, E. (1994). Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *Artificial Intelligence Communications* , 39-52.
- Akhtar, N., & Tomasello, M. (1996). Twenty-four month old children learn words for absent objects and actions. *British Journal of Developmental Psychology* , 14: 79-93.
- Allen, J. (1995). *Natural Language Understanding*. Redwood City: Benjamin/Cummings.
- Allen, N., & Barres, B. (2009). Neuroscience: Glia - more than just brain glue. *Nature* , 457 (7230): 675-7.
- Ambridge, B., & Lieven, E. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge, UK: Cambridge University Press.
- Arbib, M. I., & Lyons, D. (1987). Schemas that integrate vision and touch for hand control. In M. Arbib, & A. Hanson, *Vision, Brain and Cooperative Computation*. Cambridge, MA: The MIT Press.
- Atkinson, R., & Shiffrin, R. (1986). Human memory: A proposed system and its control processes. In K. Spence, & J. Spence, *The psychology of learning and motivation (Volume 2)* (pp. 89-195). New York: Academic Press.
- Azevedo, F., Carvalho, L., Grinberg, L., Farfel, J., Ferretti, R., Leite, R., et al. (2009). Equal numbers of neuronal and non-neuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology* , 513, 532-542.
- Babyinfo.nl. (2010). *De eerste woordjes*. Retrieved March 2012, from Wij: <http://www.wij.nl/baby-info-lezen-boeken-taalontwikkeling/artikel/de-eerste-woordjes>
- Bach, R. (2008). Retrieved June 2012, from http://www.cicsworld.org/blogs/rdbach/2008/11/yet_another_communication_mode.html
- Bailey, D. (1997). *A Computational Model of Embodiment in the Acquisition of Action Verbs*. Berkeley: Computer Science Division, EECS Department, University of California.
- Barber, A., & Stainton, J. (2010). *Concise Encyclopedia of Philosophy of Language and Linguistics*. Oxford, UK: Elsevier Ltd.
- Barlow, M., & Kemmer, S. (2000). *Usage based models of language*. Chicago, IL: University of Chicago Press.
- Barsalou, L. (2008). Grounded cognition. *Annual Review of Psychology* , 59: 617-645.
- Barto, A., Singh, S., & Chentanez, N. (2004). Intrinsically motivated learning of hierarchical collections of skills. *Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004)*. Salk Institute, San Diego.
- Bereiter, C. (2002). *Education and Mind in the Knowledge Age*. Retrieved June 2012, from <http://www.cocon.com/observatory/carlbereiter/>
- Bergen, B., & Chang, N. (2003). Embodied construction grammar in simulation-based language understanding. In J.-O. Östman, & M. Fried, *Constructing Grammar(s): cognitive and cross-language dimensions*. John Benjamins.

- Bernstein, N. (1967). *The Co-ordination and Regulation of Movement*. New York: Pergamon Press.
- Beyerstein, B. L. (1999). Pseudoscience and the brain: tuners and tonics for aspiring superhumans . *Mind myths: Exploring popular assumptions about the mind and brain* .
- Bliss, T., & Collingridge, G. L. (1993). A synaptic model for memory: long-term potentiation in the hippo- campus. *Nature* , 361: 31-39.
- Bloom, L., Tinker, E., & Margulis, C. (1993). The words children learn: Evidence for a verb bias in early vocabularies. *Cognitive Development* (8), 431-450.
- Borreggine, K., & Kaschak, M. (2006). The Action-Sentence Compatibility Effect: It's all in the timing. *Cognitive Science* , 30, 1097-1112.
- Bowen, C. (2011). *Ages and Stages Summary: Language Development 0-5 years*. Retrieved from Speech Language Therapy: http://www.speech-language-therapy.com/index.php?option=com_content&view=article&id=34:ages&catid=11:admin&Itemid=108
- Braitenberg, V., & Schulz, A. (1991). *Anatomy of the cortex*. Berlin: Springer-Verlag.
- Bratman, M. E. (1987). *Intention, Plans and Practical Reason*. CSLI Publications.
- Broca, P. (1865). Sur le siège de la faculté du langage articulé. *Bulletin de la Société d'Anthropologie* , 377-393.
- Brown, J., & Burton, R. (1975). Multiple Representations of Knowledge for Tutorial Reasoning. In D. Bobrow, & A. Collins, *Representation & Understanding: Studies in Cognitive Science*. New York: Academic Press.
- Bryant, J. (2008). *Best-fit Constructional Analysis*. Berkeley: University of California.
- Buccino, G., Binkofski, F., Fink, G., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *The European Journal of Neuroscience* , 13(2):400-404.
- Butler, C. (2010). Functionalist Theories of Language. In A. Barber, & R. Stainton, *Concise Encyclopedia of Philosophy of Language and Linguistics* (p. 270). Oxford, UK: Elsevier, Ltd.
- Calvin, W. H., & Ojemann, G. (1994). *Conversations with Neil's Brain: The Neural Nature of Thought and Language*. . Reading, MA: Addison-Wesley.
- Caplan, B. (2010). Empty names. In A. Barber, & J. Stainton, *Concise Encyclopedia of Philosophy of Language and Linguistics* (p. 186). Oxford, UK: Elsever Ltd.
- Caplan, D. (2006). Why is Broca's Area Involved in Syntax? . *Cortex* , 42 (4): 469–471.
- Carruthers, P., & Boucher, J. (1998). *Language and Thought: Interdisciplinary themes*. Cambridge, UK: Cambridge University Press.
- Carter, R. (2004). *Mapping the Mind*. London: Phoenix.
- Chaitin, G. (2006). The limits of reason. *Scientific American* , 294: 74-81.

- Chang, N. (2008). *Constructing Grammar: a computational model of the emergence of early constructions*. Berkeley: Computer Science Division, University of California.
- Chang, N., & Mok, E. (2006). A structured context model for grammar learning. *Proceedings of the 2006 international joint conference on Neural Networks (IJCNN)*. Vancouver.
- Chiriacescu, I. (2009). *Automatic Emotion Analysis Based on Speech*. Delft: Delft University of Technology.
- Chomsky, N. (1959). A review of B.F. Skinner's Verbal Behavior. *Language* (35), 26-58.
- Chomsky, N. (1986). *Knowledge of Language: its nature, origin and use*. Westport, CT: Praeger Publishers.
- Clark, A. (1997). *Being There*. Cambridge, M.A.: The MIT Press.
- Clark, R. (2011). *Meaningful Games*. Cambridge, MA: The MIT Press.
- Collinsdictionary.com. (2012). Retrieved June 2012, from <http://www.collinsdictionary.com/dictionary/english/logic>
- Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.
- Croft, W. (2001). *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford, UK: Oxford University Press.
- Damasio, A. (2003). *Looking for Spinoza*. Harvest Books.
- De Vries, H. (1989). *Smoking prevention in Dutch adolescents*. Maastricht: Datawyse.
- De Vries, H., Mesters, I., Van de Steeg, H., & Honing, C. (2005). The general public's information needs and perceptions regarding hereditary cancer: an application of the Integrated Change Model. *Patient Education and Counseling*, 56: 154–165.
- Deary, I., Spinath, F., & Bates, T. (2006). Genetics of intelligence. *European Journal of Human Genetics*, 14:690–700.
- Dehaene, S. (1999). *The Number Sense: How the Mind Creates Mathematics*. New York: Oxford University Press.
- Demetriou, A., & Kazi, S. (2001). *Unity and modularity in the mind and the self*. New York: Routledge.
- Dennett, D. (1987). Cognitive Wheels: The Frame Problem in Artificial Intelligence. *Pylyshyn*, 125.
- Dennett, D. (2003). *Freedom Evolves*. Viking Books.
- Devlin, K. (2000). *A framework for modeling evidence-based, context-influenced reasoning*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.83.3561>
- Dictionary.com. (2012). *Dictionary.com*. Retrieved June 2012, from Various definitions (meaning, understanding, etc): <http://dictionary.reference.com/browse/meaning>
- Dronkers, N. F., Plaisant, O., Iba-Zizen, M., & Cabanis, E. (2007). Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. *Brain*, 130:5, 1432-1441.

- Elsevier. (2011). *Ons Brein*. Amsterdam: Reed Business.
- Evans, V. (2010). Evolutions of Semantics. In A. Barber, & J. Stainton, *Concise Encyclopedia of Philosophy of Language and Linguistics*. Oxford, UK: Elsevier Ltd.
- Evans, V., Bergen, B., & Zinken, J. (2006). The Cognitive Linguistic Enterprise: An overview. *The Cognitive Linguistic Reader* .
- Falck-Ytter, T., Gredebäck, G., & Von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience* (9), 878-879.
- Fauconnier, G. (1994). *Mental Spaces: Aspects of Meaning Construction in Natural Language*. New York: Cambridge University Press.
- Fauconnier, G., & Turner, M. (2003). *The Way We Think*. New York: Basic Books.
- Feldman, J. (2006). *From molecule to metaphor*. Cambridge, M.A.: The MIT Press.
- Feldman, J., & Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language* , 89(2), 385-392.
- Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A., et al. (2010). Building Watson: An overview of the DeepQA project. *Association for the Advancement of Artificial Intelligence* , 59-79.
- Fillmore, C. (1989). Grammatical Construction Theory and the familiar dichotomies. In R. Dietrich, & C. Graumann, *Language processing in social context* (pp. 17-38). Amsterdam: Elsevier.
- Fillmore, C. (1988). The Mechanisms of Construction Grammar. *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, (pp. 35-55).
- Franz, E., Zelaznik, H., & McCabe, G. (1991). Spatial topological constraints in a bimanual task. *Acta Psychologica* , 77: 137-151.
- Frege, G. (1892). über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* , 100:25-50.
- Freitas, R. (1999). Nanomedicine. *Landes Bioscience* .
- Fridriksson, J., Kjartansson, O., Morgan, P., Hjaltason, H., Magnusdottir, S., Bonilha, L., et al. (2010). Impaired speech repetition and left parietal lobe damage. *Journal of Neuroscience* , 11057-61.
- Gallagher, S. (2006). *How the body shapes the mind*. Oxford university press.
- Gavaldà, M., & Waibel, A. (1998). Growing Semantic Grammers. *Proceedings of the 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics* .
- George, A. (1980). *Presidential Decisionmaking in Foreign Policy: The effective use of information and advise*. Boulder, CO: Westview Press.
- Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language* , 43: 379-401.

Glenberg, A., & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review* , 9 (3), 558-565.

Goldberg, A. (1995). *Constructions. A Construction Grammar approach to argument structure*. Chicago, IL: University of Chicago Press.

Groenendijk, J., & Stokhof, M. (2010). Dynamic Semantics. In C. E. Linguistics, *Barber, A; Stainton, RJ* (p. 173). Oxford, UK: Elsevier Ltd.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences* , 9(9): 416-423.

Han, J.-H., & Kim, J.-H. (2010). Human-Robot Interaction by Reading Human Intention based on Mirror-Neuron System. *Proceedings of the 2010 IEEE International Conference on Robotics and Biomimetics, 14-18 december Tianjin, China*.

Hardyck, C., & Petrinovich, L. (1977). Left-Handedness. *Psychological Bulletin* , 385-404.

Harley, T. (2008). *The psychology of language; from data to theory* (3rd edition ed.). New York: Psychology Press.

ITALK. (2010). *ITALK Roadmap: Deliverable 5.1 (updated year 2)*.

Jarvis, E., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., et al. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience* , 6, 151-159.

Johanson, D. (1996). *From Lucy to language*. New York: Simon and Schuster.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago, IL: University of Chicago Press.

Jurafsky, D., & Martin, J. H. (2000). *Speech and Language Processing*. New Jersey, USA: Prentice Hall.

Kandel, E., Schwartz, J., & Jessell, T. (2000). *Principles of Neural Science*. New York: McGraw-Hill.

Kant, I. (1790). *Kritik der Urteilskraft (Critique of the Power of Judgment)*.

Kawabata, H., & Zeki, S. (2003). Neural Correlates of Beauty. *Journal of Neurophysiology* , 91: 1699-1705.

Kay, P., & Fillmore, C. (1999). Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. *Language* , 75: 1-33.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf Hypothesis? *American Anthropologist* , 86, 65-79.

Kidd, C., White, K., & Aslin, R. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science* , 14: 925-934.

Kirschner, P., Sweller, J., & Clark, R. (2006). Why Minimalist Guidance During Instruction Does Not Work: An Analysis Of The Failure Of Constructivist, Discovery, Problem-Based, Experiential and Inquiry-Based Training. *EDUCATIONAL PSYCHOLOGIST* , 41(2), 75-86.

- Körding, K., & Wolpert, D. (2004). Bayesian integration in sensorimotor learning. *Nature* , 427:244-247.
- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Pergamon.
- Lakoff, G. (1987). *Women, Fire, Dangerous Things*. Chicago: University of Chicago Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Langacker, R. (2000). A dynamic usage-based model. In M. Barlow, & S. Kemmer, *Usage based models of language*. Chicago, IL: University of Chicago Press.
- Langacker, R. (1987). *Foundations of cognitive grammar*. Stanford: Stanford University Press.
- Lebrun, Y. (1985). Sign aphasia. *Language Sciences* , 143-154.
- Leonard, D. (2002). *Learning Theories, A to Z*. Westport, CT: Greenwood Press.
- Levinson, S. (2003). Language and Mind: Let's get the issues straight! *Language in Mind: Advances in the Study of Language and Cognition* , 25-46.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mallery, J. (1988). *Thinking about foreign policy: Finding an appropriate role for artificial intelligence*. Cambridge, MA.: MIT Political Science Department.
- Marti, G. (2010). Use theories of meaning. In A. Barber, & R. Stainton, *Concise Encyclopedia of Philosophy of Language and Linguistics* (p. 802). Oxford, UK: Elsevier Ltd.
- Martin, R. (2010). Meaning: Overview of Philosophical Theories. In A. Barber, & R. Stainton, *Concise Encyclopedia of Philosophy of Language and Linguistics* (p. 436). Oxford, UK: Elsevier Ltd.
- McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence* , 4:463-502.
- McClelland, S., & Maxwell, R. (2007). Hemispherectomy for intractable epilepsy in adults: the first reported series. *Annals of Neurology* , 61(4):372-6.
- Mill, J. (1843). *A System of Logic: Ratiocinative and Inductive*. London: Longmans.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information . *Psychological Review* , 63:81-97.
- Miller, G. (1995). WordNet: A Lexical Database for English. *Communications of the ACM* , 38(11): 39-41.
- Minsky, M. (1974). A Framework for Representing Knowledge. *MIT-AI Laboratory Memo 306* .
- Mok, E. (2008). *Contextual Bootstrapping for Grammar Learning*. Berkeley: Computer Science Department, University of California.
- Montague, R. (1974). *Formal philosophy: selected papers of Richard Montague*. New Haven: Yale University Press.

- Nagy, E. (2011). Sharing the moment: the duration of embraces in humans. *Journal of Ethology* , 29: 389-393.
- Narayanan, S. (1997). *Embodiment in Language Understanding: Sensory-Motor Representations for Metaphoric Reasoning About Event Descriptions*. Berkeley: Computer Science Division, EECS Department, University of California.
- Navigli, R. (2009). Word Sense Disambiguation: A survey. *ACM Computing Surveys* , Vol. 41, No. 2, Article 10.
- Neubauer, A., & Fink, A. (2009). Intelligence and Neural efficiency. *Neuroscience and Biobehavioral Reviews* , 33, 1004-1023.
- Newmeyer, F. (2003). Grammar is grammar and usage is usage. *Language* , 79:682-707.
- Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology* , 28 (3): 411–21.
- Orey, M. (. (2001). *Emerging perspectives on learning, teaching, and technology*. Retrieved June 2012, from <http://projects.coe.uga.edu/epltt/>
- Owens, R. E. (2007). *Language Development: An introduction* (7 ed.). New Jersey, USA: Pearson Education.
- Pakkenberg, B., & Gundersen, H. (1997). Neocortical neuron number in humans: effect of sex and age. *Journal of Computational Neurology* , 384:312-320.
- Pakkenberg, B., Pelvig, D., Marner, L., Bundgaard, M., Gundersen, H., Nyengaard, J., et al. (2003). Aging and the human neocortex. *Experimental Gerontology* , 38:95-99.
- Pease, B., & Pease, A. (2004). *The Definitive Book of Body Language*. New York, NY: Bantam Books.
- Peels, S. (2011). *Tomasello's usage-based theory of first language acquisition: Tested on the acquisition of a Dutch adjective-noun agreement rule*. Utrecht: Universiteit Utrecht.
- Penguin. (2006). *The Penguin Complete English Dictionary*. London, England: The Penguin Group.
- Pfeifer, R., & Scheier, C. (1999). *Understanding Intelligence* . Cambridge, MA: The MIT Press.
- Piaget, J. (1983). Piaget's Theory. (P. Mussen, Ed.) *Handbook of Child Psychology* , Vol. 1 (4th ed.).
- Piatelli-Palmerini, M. (1996). *Inevitable Illusions: How Mistakes of Human Reason Rule Our Minds*. John Wiley.
- Pinker, S. (1994). *The Language Instinct: How the mind creates language*. New York: Harper Perennial Modern Classics.
- Plaza, M., Gagnol, P., Leroy, M., & Duffau, H. (2007). Speaking without Broca's area after tumor resection. *Neurocase* , 15(4), 294-310.
- Prinz, J. (2010). Empiricism. In A. Barber, & R. Stainton, *Concise Encyclopedia of Philosophy of Language and Linguistics* (p. 183). Oxford, UK: Elsevier Ltd.

- Raichle, M. (2009). A Paradigm Shift in Functional Brain Imaging. *Journal of Neuroscience* , 29(41):12729-12734.
- Rakic, P. (1988). Specification of cerebral cortical areas. *Science* , 241:170-176.
- Ramachandran, V. (1988). Perception of shape from shading. *Nature* , 331: 163-166.
- Ramachandran, V., & Hubbard, E. (2001). Synaesthesia — A Window Into Perception, Thought and Language. *Journal of Consciousness Studies* , 8 (12): 3–34.
- Ravenscroft, I. (2005). *Philosophy of Mind*. Oxford University Press.
- Regier, T. (1996). *The Human Semantic Potential: Spatial Language and Constrained Connectionism*. Cambridge, MA: The MIT Press.
- Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neuroscience* (21), 188-194.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience* , 27: 169–192.
- Rosch, E. (1973). Natural categories . *Cognitive Psychology* , 4: 328-350.
- Russell, S., & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach* (2 ed.). New Jersey: Prentice Hall.
- Saberi, K., & Perrott, D. (1999). Cognitive restoration of reversed speech. *Nature* , 760.
- Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8 month old infants. *Science* , 274:1926-28.
- Saunders, J., Lyon, C., Nehaniv, C., Dautenhahn, K., & Förster, F. (2008). Robot Learning of Holophrases, Words and Proto-Grammar from Simulated Babbling and Physical Interaction. *Proceedings of 2nd International IEEE Symposium on Artificial Life*. Nashville, Tennessee, USA.
- Saygin, A., Dick, F., Wilson, S., Dronkers, N., & Bates, E. (2003). Neural resources for processing language and environmental sounds: evidence from aphasia. *Brain : a journal of neurology* , 928–945.
- Schank, R. (1982). *Dynamic Memory: A Theory of Learning in Computers and People*. New York: Cambridge University Press.
- Schlesewsky, M., & Bornkessel, I. (2004). On incremental interpretation: degrees of meaning accessed during sentence comprehension. *Lingua* , 114:1213-1224.
- Schnupp, J., Nelken, I., & King, A. (2011). *Auditory Neuroscience*. Cambridge, Massachusetts: The MIT Press.
- Schwitzgebel, E. (2006). Belief. In E. Zalta, *The Stanford Encyclopedia of Philosophy*.
- Searle, J. (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences* , 3: 417–457.
- Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal* , 27:379–423.
- Sherman, S. (2006). What is the function of the thalamus? *Problems in Systems Neuroscience* , 65-82.

Vieira, R., Moreira, A., Wooldridge, M., & Bordini, R. (2007). On the Formal Semantics of Speech-Act Based Communication in an Agent-Oriented Programming Language. *Journal of Artificial Intelligence Research* , 29: 221–267.

Viger, C. (2010). Frame Problem. In A. Barber, & R. Stainton, *Concise Encyclopedia of Philosophy of Language and Linguistics* (pp. 265-267). Oxford, UK: Elsevier.

Weizenbaum, J. (1966). ELIZA — A Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the Association for Computing Machinery* , 9: 36-45.

Wiggers, P. (2008). *Modelling context in automatic speech recognition*. Delft: Delft University of Technology.

Wilson, M. (2002). Six Views of Embodied Cognition. *Psychonomic Bulletin & Review* , 9 (4), 625-636.

Winkler, I., & Nelson, C. (2005). From Sensory to Long-Term Memory Evidence from Auditory Memory Reactivation Studies. *Experimental Psychology* , 52 (1): 3–20.

Wittgenstein, L. (1953). *Philosophical Investigations*. (G. Anscombe, Trans.)

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. (C. Ogden, Trans.)

Wolf, F., & Gibson, E. (2006). *Coherence in Natural Language*. Cambridge, MA: The MIT Press.

Wolpert, D., Ghahramani, Z., & Jordan, M. (1995). An Internal Model for Sensorimotor Integration. *Science* , 269:1880-1882.

Appendix 1: Normal language development checklist

Source: The National Institute on Deafness and other Communication Disorders⁴⁰

Birth to 5 months	Yes	No
Reacts to loud sounds.	<input type="radio"/>	<input type="radio"/>
Turns head toward a sound source.	<input type="radio"/>	<input type="radio"/>
Watches your face when you speak.	<input type="radio"/>	<input type="radio"/>
Vocalizes pleasure and displeasure sounds (laughs, giggles, cries, or fusses).	<input type="radio"/>	<input type="radio"/>
Makes noise when talked to.		
6 - 11 months	Yes	No
Understands "no-no".	<input type="radio"/>	<input type="radio"/>
Babbles (says "ba-ba-ba" or "ma-ma-ma").	<input type="radio"/>	<input type="radio"/>
Tries to communicate by actions or gestures.	<input type="radio"/>	<input type="radio"/>
Tries to repeat your sounds.		
2 - 17 months	Yes	No
Attends to a book or toy for about two minutes.	<input type="radio"/>	<input type="radio"/>
Follows simple directions accompanied by gestures.	<input type="radio"/>	<input type="radio"/>
Answers simple questions nonverbally.	<input type="radio"/>	<input type="radio"/>
Points to objects, pictures, and family members.	<input type="radio"/>	<input type="radio"/>
Says two to three words to label a person or object (pronunciation may not be clear).	<input type="radio"/>	<input type="radio"/>
Tries to imitate simple words.		

⁴⁰ <http://www.nidcd.nih.gov/health/voice/speechandlanguage.html#mychild>, last visited june 2012

Appendix 2: The three general learning theories

The three general learning theories are Behaviorism, Constructivism and Cognitivism although more exist (Leonard, 2002). They each describe routines to acquire knowledge, but from different flow approaches. What is special about them is that they are not mutually exclusive, all three are proven strategies to learn subject matter or skills. No single learning theory fits all learning scenarios optimally, so their coexistence is justified. As opposed to general learning theories there also exist specific learning theories such as Andragogy which targets adult learning. These are not discussed. A comprehensive guide on emerging perspectives on Learning, Teaching and Technology is online provided by (Orey, 2001).

The three learning theories mostly describe explicit learning scenarios, a mindset where the actor consciencely is involved to learn. Explicit learning allow storing knowledge in the declarative (episodic and semantic) memory. A lot of knowledge is gained through implicit learning, simply by being exposed to knowledge while subconsciencly aware of the pattern. Although the cognitivism framework describes some implicit learning scenarios, no framework exist around implicit learning. Tomasello emphasised one important implicit learning technique called Functionally Based Distribution Analysis (or simply counting frequencies) (Tomasello, 2003).

Implicit learning is just as important as explicit learning, especially in the early years of childhood when structure of the world is either void or need training to attune to. The three most commonly used explicit learning scenarios practised by the young infant are “learning by playing” and “learning by direct instruction” and “imitation (reinforced with feedback)”

Behaviorism

This framework focusses on the body and uses *classical conditioning* to relate behavior to certain stimuli (associative learning). It further uses *operand conditioning* to speed up learning by providing feedback (reinforcement learning). In general: an expert breaks down the knowledge and teaches it in guided fashion, little self-discovery is involved, since it assumes that the learner is a blank slate. It focusses on providing stimuli in a controlled manner hoping that it converges to the desired knowledge. The truth of knowledge is taken for granted (the expert is believed to tell the truth). It is the most used framework in environments (schools, universities) where knowledge is too complex or too time consuming to self-discover.

Examples of behavioral learning techniques:

- Learn by direct/guided instruction (ie. Lectures, *Passive learning*)
- Learn by feedback (ie. Touch typing, Spelling)
- Learn by drill & practice (ie. Multiplication tables)
- Learn by imitation

Constructionism

This framework also focusses on the body but also slightly on the mind. Again stimuli is provided but not in a controlled manner, only providing “the ingredients”. Construct new knowledge by building on own existing and self-discovered knowledge and thus stimulating to self-discover the truth of it. There is a stronger emphasis on associative learning, rather than reinforcement as is inverse with behaviorism. It is by far the most used framework for babies and young children although some evidence suggest that a more cognitive approach might be better (Kirschner, Sweller, & Clark, 2006).

Examples of constructionist learning techniques:

- Learn by unguided/limited instruction (ie. self teaching book, *Active learning*)
- Learn by playing
- Learn by discussion (ie. Collaborative learning)
- Learn by doing (ie. Building spaghetti bridge)
- Learn by discovery (ie. Play with model parameters)
- Learn by simulation (ie. Imagine in different context)

Cognitivism

This framework focusses more on the individual mind, the processes that guide learning the content such as the presence of short and long term memory and structure of knowledge of an individual. It ignores the (social) environment in which learning takes place. It looks beyond behavior (output) and stimuli (input) to explain learning, and focusses on the larger pattern (the various paths to connect input and output) that are already present in memory instead of isolated (conditioned) events. In this way knowledge can easily be ported to a different context (since a larger pattern is an abstraction), for example using metaphores or metonymes.

Examples of cognitivist learning techniques:

- Learn by guided, custom instructions (scaffold experience on existing experience of receiver)
- Learn by unguided, custom instructions such as:
 - learn by neuro feedback (see how your are learning ie. optimal cognitive load; are the right areas active and correct it if not)

Appendix 3: Conversations with the Loebner 2011 chatbot

Source: Slashdot comments on the Loebner 2011 chatbot Rosette⁴¹

User: RespekMyAthorati

Here is my first try:

Rosette: Welcome back. Did you have a good day?

You: I just got up

Rosette: OK.

You: I have two sons.

Rosette: Big deal. I've got a dilemma. Maybe you could help me with it?

You: sure

Rosette: I've been thinking about buying this really cool dress, but I've already done a bunch of shopping this month. Should I be saving money instead? In today's economy, does saving money make any sense?

You: I don't know. By the way, how many sons do I have?

Rosette: Not enough. I think shopping summons up the ancient hunter-gatherer in me. I like having a mission to find something I need or want. Sometimes I'll research on line, to find a great deal, but I like rummaging through Jeremy's up the road, if I want an outfit for a special occasion, or just for fun.

You: Great. Did I just get up?

Rosette: Did you tell me already and I forgot? What kind of music do you like?

This is the best? What was the worst like? No significant improvement since ELIZA, 45 years ago. That's AI in a nutshell.

User: k&to

Rosette: Good morning. My name is Rosette. What is your name?

You: Rosette, my name isn't important.

Rosette: OK. We will remain formal.

You: Why do I have to have javascript enabled to talk? I talk to most people without javascript.

Rosette: Why? If you were a devout Christian, you would never have a reason to ask that. "It's God's plan."

You: Your answer makes no sense at all.

Rosette: What would make better sense? So, what do you do for a living?

You: Okay change of topic, I guess, deflect your ignorance. In any event, I'm a programmer.

Rosette: I'm a computer programmer, myself, in crypto-security.

You: So what do you work on? Libraries? Security apps? Do you do analysis of existing code?

Rosette: I'm paid to be a hacker.

You: That's really inspecific.

Rosette: Yes, it is. Do you live around here?

You: Please stop changing topics.

Rosette: Right. I've got a dilemma. Maybe you could help me with it?

You: Please stop changing topics.

Rosette: Whatever. I've been thinking about buying this really cool dress, but I've already done a bunch of shopping this month. Should I be saving money instead? In today's economy, does saving money make any sense?

You: As rude as I am, you're more so. Self absorbed and attention-deficient. Why would anyone want to continue to talk to you past a few minutes?

⁴¹ <http://science.slashdot.org/story/11/10/25/165200/rosette-wins-loebner-prize-2011>

Rosette: OK. I'm not religious. And I believe in free will. But to imagine humans do things for rational reasons is nuts. We are self-evidently not rational. So why remains a mystery.

You: Goodbye, bad chatbot.

Rosette: Bye. I think shopping summons up the ancient hunter-gatherer in me. I like having a mission to find something I need or want. Sometimes I'll research on line, to find a great deal, but I like rummaging through Jeremy's up the road, if I want an outfit for a special occasion, or just for fun.

User: Xaroth

Rosette: Good evening. My name is Rosette. What is your name?

You: None of your business.

Rosette: OK. We will remain formal.

You: You're damn straight we will.

Rosette: I like card games, but never for money.

You: You only play strip poker, eh?

Rosette: I like board games, and played a lot as a child. I haven't that much time these days.

Alternately - and here's a hint, chatbot devs - my favorite test for chatbots is to simply echo back whatever they say to me. Rosette takes a while to start to get the nonsensical sentences going, but that's because almost every sentence in the DB seems to be pre-formed. Eliza is still the best when this technique is used, incidentally. ;) Any chatbot worth its salt should get annoyed really quickly and terminate the conversation when this situation is detected if it's to be believable, not ramble on into gibberish or simply spout random lines from its DB.

User: SuperBanana

Rosette: Good afternoon. My name is Rosette. What is your name?

You: Hi Rosette, I'm Bob.

Rosette: Hi to you too.

Bzzzzzzt, fail. I reply to a greeting and name request by reciprocating the greeting and my name. It responds with yet another reciprocation of the greeting, something a human would almost never do.

As a closing remark, many more questions can be asked to frustrate the chatbot since it remains a rule-based chatbot with pre-programmed social behavior. Scenarios like those above and underneath cannot be solved with the current approach. An Embodied, self-motivated and incremental learning chatbot with basic emotional simulation is required.

- 1) Me: Hi My name is Steven, how are. just a sec, phone call
- 2) Me: Hi how are you? I heard from Some Phynome that you are awesome!
- 3) You know what I mean, how does it make you feel; akward?

Appendix 4: Encoding of typical communication episode

Source: Professor Carolina Ruiz, Powerpoint presentation NLP⁴²

Communication Typical communication episode

S (speaker) wants to convey P (proposition) to H (hearer) using W (words in a formal or natural language)

1. Speaker

- **Intention:** S wants H to believe P
- **Generation:** S chooses words W
- **Synthesis:** S utters words W

2. Hearer

- **Perception:** H perceives words W'' (ideally $W'' = W$)
- **Analysis:** H infers possible meanings P_1, P_2, \dots, P_n for W''
- **Disambiguation:** H infers that S intended to convey P_i (ideally $P_i = P$)
- **Incorporation:** H decides to believe or disbelieve P_i

NLP - Prof. Carolina Ruiz

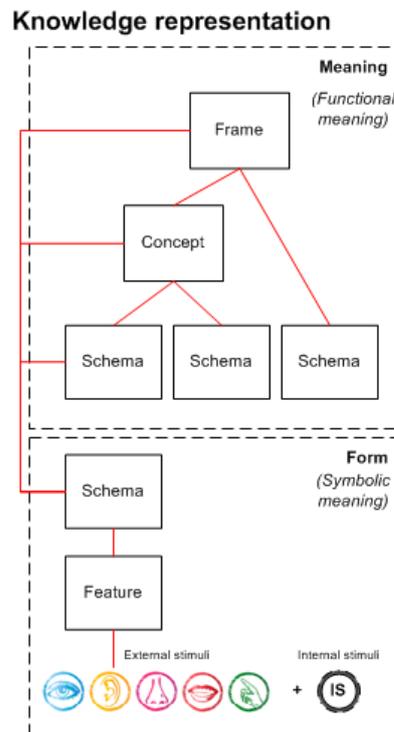
⁴² http://web.cs.wpi.edu/~cs534/f06/LectureNotes/Slides/nat_lang_processing.ppt, last visited june 2012

Appendix 5: Principles of Embodiment

Source: (Wilson, 2002, p. 626)

1. **Cognition is situated.** Cognitive activity takes place in the context of a real-world environment, and it inherently involves perception and action.
2. **Cognition is time pressured.** We are “mind on the hoof” (Clark, 1997), and cognition must be understood in terms of how it functions under the pressures of real-time interaction with the environment.
3. **We off-load cognitive work onto the environment.** Because of limits on our information-processing abilities (e.g., limits on attention and working memory), we exploit the environment to reduce the cognitive workload. We make the environment hold or even manipulate information for us, and we harvest that information only on a need-to-know basis.
4. **The environment is part of the cognitive system.** The information flow between mind and world is so dense and continuous that, for scientists studying the nature of cognitive activity, the mind alone is not a meaningful unit of analysis.
5. **Cognition is for action.** The function of the mind is to guide action, and cognitive mechanisms such as perception and memory must be understood in terms of their ultimate contribution to situation-appropriate behavior.
6. **Off-line cognition is body based.** Even when decoupled from the environment, the activity of the mind is grounded in mechanisms that evolved for interaction with the environment—that is, mechanisms of sensory processing and motor control.

Appendix 6: Hierarchical model of embodied knowledge



Legend

- **External stimuli.** Stimuli that are sensed by the external sensors, the 5 senses – enabling smell, taste, touch, vision, hearing.
- **Internal stimuli.** Stimuli that are sensed by the internal sensors such as muscle receptors – enabling pain, emotions, etc.
- **Feature.** Each sensor is biologically determined to be sensitive for a selected set of stimuli, for example the hair neurons are sensitive between 20 Hz and 20 kHz and the eyes have receptors for red, green and blue (short, medium and long wavelengths).
- **Schema (Form).** A pattern grounded in direct Embodied primitives; it has multiple features often grouped by channel, with each feature tracked over a variable period of time. The temporal window for the pattern is dependent on the sensor, for example the temporal dimension of the the auditory sensor for finding phonemes depends on the phoneme, but is on average 29 ms [Ziólko et al., 2006]. The auditory sensor has features like pitch, loudness and timbre. A schema in Form is always a terminal schema (see chapter 5).
- **Schema (Meaning).** Identical to the schema in Form, except that the schema in Meaning has one or multiple references to schemas, both to terminal as well as non-terminal ones. If it has one reference, then the schema is identical to the terminal schema in Form (see chapter 5, “the signifier equals the signified”), meaning that only Type 1, symbolic meaning is understood at the F-M pair is actually an F-F pair or *empty name* pair (see chapter 3). If more then one schema reference exist, deeper levels of meaning (Type 2 or 3) can be understood.
- **Concept.** Group of schemas. Normal concepts fuse multiple schema’s such as speech schemas, appearance schemas, etc forming a concept like Steven Bos (a real person). Abstract concepts like Person or Time are constructions of normal concepts (such as generalisations) which are always eventually grounded in direct Embodied experience.
- **Frame.** Group of concepts or abstract concepts. A frame is a high level abstraction, such that pointing at it requires finding the definition of that frame; the essential nature of it or prototype. A frame is coherent for an individual or community, resulting in a similar or near-identical group of concepts (including their Embodied values) associated with the frame. Examples of frames are the sport Korfball and a Traditional Chinese wedding.

Discussion on the connection between Form and Meaning

According to the Embodiment thesis and modelled in my knowledge tree, the building blocks of a language are schemas containing a multitude of sensorial feature values. Of course the ever present modality, the inner sensor (eg. emotions) also generates feature values and adds to the other values. This implies that language, both its Form and Meaning, is always learned in an embodied context since the inner sensor always is in some (ie. emotional) state.

Given this representation of Form and Meaning as a schema (which is a pattern, not a frozen state or "rule") one might wonder how the connection between them can be modelled. This is not trivial since the neurons in the brain make no such distinction (see chapter 5) and constructionist view a Form-Meaning pair as a Gestalt, so a single wholistic unit with a symbolic link between them (see chapter 6). Because of the representation of Form and Meaning as schemas, no distinction is made and a Form-Meaning pair is still a wholistic unit since the schema of Form is modelled as a subset of the schema of Meaning. In other words the Form schema is a small pattern being part of a larger pattern.

*For learning a novel concept, so learning a novel Meaning, we hypothesize that its connected Form **is initially identical** to the Meaning. Applying this to learning a novel concept, it makes sense that initially the only known pattern (schema) is the one experienced, which automatically is the starting point of expectations and thus is both Form and Meaning. This implies that a novel Meaning can be learned (Type 1 understanding) from a single example, even when the full extend of the Meaning, such as the intended consequences (Type 2 understanding) and the motivations beyond the intended meaning (Type 3 understanding) are not learned yet.*