# SEGMENTATION AND INTEGRATION IN TEXT COMPREHENSION: A MODEL OF CONCEPT NETWORK GROWTH

A dissertation submitted to Kent State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

by

Manas Sudhakar Hardas May 2012 Dissertation written by Manas Sudhakar Hardas B.E., University of Mumbai, 2003 M.S., Kent State University, 2006 Ph.D., Kent State University, 2012

# Approved by

Dr. Javed I. Khan , Chair, Doctoral Dissertation CommitteeDr. Austin Melton , Members, Doctoral Dissertation CommitteeDr. Arvind BansalDr. Katherine Rawson

# Accepted by

Dr. Denise Bedford

Dr. Javed I. Khan, Chair, Department of Computer Science

Dr. John Stalvey , Dean, College of Arts and Sciences

# TABLE OF CONTENTS

LI	ST O	F FIGU	JRES	vii
LIST OF TABLES				
A	eknow	ledgem	ents	xii
De	edicat	ion		xiv
1	Intro	oduction	n: Text Comprehension	1
	1.1	Text c	comprehension in literature	9
		1.1.1	Knowledge representation in text comprehension	10
		1.1.2	Cognitive processes in text comprehension	12
		1.1.3	Other factors in text comprehension	17
	1.2	Thesis	contributions	18
	1.3	Thesis	organization	19
2	Com	putatio	onal segmentation and integration	20
	2.1	An exa	ample of concept network growth by SI	20
	2.2	Comp	utational model for SI	23
		2.2.1	Model parameters	24
		2.2.2	Notation	25
		2.2.3	Segmentation of concepts	27

		2.2.4	Integration of concepts	28
		2.2.5	Constraints formation in SI	28
	2.3	Comp	uting the model parameters	31
	2.4	Solvab	ility issues: mutual exclusivity	33
		2.4.1	Mutual exclusivity between two individual readers	34
		2.4.2	Mutual exclusivity due to delayed concept recognition	36
	2.5	Impler	nentation and computability	38
	2.6	Exerci	se: Computing a base concept network for a text	40
	2.7	Conclu	usion	40
2	Func	tion of	concept seguence	12
3	Func		concept sequence	40
	3.1	Motiva	ation	43
	3.2	Experi	iment	44
		3.2.1	Design	44
		3.2.2	Assumptions	46
		3.2.3	Statistics obtained from individual concept networks $\ldots$	47
		3.2.4	Statistics of base concept network for different sequences	50
	3.3	Conclu	ision	58
4	Stru	atura of	a grown concept network	50
4	561 0			09
	4.1	Algori	thm for the SI model of growth	60
	4.2	Impor	tant properties of the growth algorithm	61
	4.3	Struct	ural properties	63
		4.3.1	Size of the network	64

		4.3.2	Connectivity/sparsity	65
		4.3.3	Neighborhood clustering	66
		4.3.4	Reachability	69
		4.3.5	Degree distribution	71
	4.4	Implic	ations of structural properties	72
		4.4.1	Small worlds	72
		4.4.2	Proportional connectivity	73
		4.4.3	Other implications	74
	4.5	Conclu	asion	75
5	Con	nectedn	ess of a concept network	76
	5.1	Motiva	ation	77
	5.2	Comp	onent analysis of simulated concept networks	78
		5.2.1	Number of components	80
		5.2.2	Fraction of nodes in the giant component	83
		5.2.3	Average size of non giant component U	84
		5.2.4	Phase transition point versus average degree z	86
	5.3	Implic	ations	89
	5.4	Conclu	asion	90
6	Con	alucion		01
0	6.1	Some	thoughts about the mathematical model	91 01
	6.2	Short	term versus long term memory	03
	63	What	is a concopt?	90 05
	0.0	100100		30

	6.4	Constructivism and the SI model	96
	6.5	Concept network as an objective measure of quality of comprehension	97
	6.6	Conclusion	99
7	App	endix A: Data Collection	100
8	App	endix B: Derivation for degree distribution	103
BI	BLIO	OGRAPHY	107

# LIST OF FIGURES

2.1		21
2.1	Example of concept network growth.(a) A toy concept network constructed	
	for the example sentence 1 (b) Interaction with background concepts (c)	
	An example construction of concept network after comprehending second	
	example sentence 2. The concept network in (b) is now the background	
	concept knowledge for (c). (d) Textbase and Situation model for example	
	sentence 1 [Kintsch 1994]	22
2.2	Example of constraints formation in segmentation and integration	28
2.3	Transformation of a concept network in an episode	31
2.4	The issue of linear separability between individuals	34
2.5	Solution to mutual exclusivity between two individuals	36
2.6	Solution to mutual exclusivity due to delayed concept recognition	37
2.7	Example of a computed base concept network. The nodes named std1;	
	std2, etc. represent the background knowledge of each of the students. The	
	paragraph of text used and the ICNs from which this BCN is computed	
	are shown in Appendix A	42
3.1	Age of acquisition against concept weight for four different sequences	53
3.2	Concept id against concept weight in four different sequences. $\ldots$ .	53
3.3	Age of acquisition against variance in concept weight within a sequence,	
	plotted for four different sequences	55

3.4	Concept id against variance in the concept weight within a sequence, plot-	
	ted for four different sequences	56
3.5	Plot of average concept weight of a concept in between four different se-	
	quences against variance in concept weight in between four different se-	
	quences	58
4.1	Analysis of size of simulated concept networks (a) Number of concepts	
	recognized $(n_{rec})$ as a function of node recognition rare $(\vartheta)$ (b) number of	
	edges recognized $(l_{rec})$ as a function of node recognition rate $(\vartheta)$ . The size	
	of the graph is constant at $n = 100$	64
4.2	Analysis of connectivity/sparsity of simulated concept networks (a) Av-	
	erage degree (z) as a function of node recognition rate ( $\vartheta$ ) (b) average	
	weighted degree $(z_w)$ as a function of node recognition rate $(\vartheta)$ . The size	
	of the graph is constant at $n = 100$	65
4.3	Analysis of clustering coefficients of the simulated concept networks (a)	
	Clustering coefficient 1 $(CC_1)$ [Barrat, Weigt 2000] against node recog-	
	nition rate $(\vartheta)$ . (b) Clustering coefficient 2 $(CC_2)$ [Watts, Strogatz 1998]	
	against node recognition rate ( $\vartheta$ ). The size of the graph is constant at $n = 100$ .	67
4.4	Analysis of reachability of simulated concept networks (a) Average shortest	
	path length $(L)$ as a function of node recognition rate $(\vartheta)$ (b) Diameter	
	$(D)$ as a function of node recognition rate $(\vartheta)$ . The size of the graph is	
	constant at $n = 100$	69
4.5	Degree distribution and cumulative degree distribution. Values of $k$ are	
	generated over 10000 runs. The size of the graph is constant at $n = 200$ .	70

- 5.1 Evolution of a network structure and the phenomenon of phase transition. The lower figure shows the fraction of the nodes in the giant component against time. The upper three figures show the time step at which snapshots of the evolving networks are taken at t = 5, 20, 40.  $\beta = 0.5$  is the association threshold. Red nodes are a part of the giant component. Phase transition occurs at t = 20. At t = 40 most of the nodes in the network are in the giant component with very few disconnected components. . . .

79

5.5	Relationship between average degree and phase transition point. (a) Phase	
	transition occurs when 50% of the nodes in giant component, (b) 70% and	
	(c) 90% of the nodes. The initial graph contains $n_0=50$ nodes as discon-	
	nected components. In each figure, phase transition point $(p_T)$ versus t is	
	plotted for three values of association threshold $\beta = [0.15, 0.5, 0.85]$	86
7.0	Data collection: Drawn ICNs	02

# LIST OF TABLES

3.1	Correlation between presentation episode for concepts between sequences.	46
3.2	Average number of concepts recognized $(n)$ and average number of asso-	
	ciations made $(l)$ by each of the four groups of readers are their standard	
	deviations.	47
3.3	Moving averages of number of concepts and associations in each of the 8	
	episodes fitted to a linear regression model	48
3.4	Correlation between node weights of each concept in different sequences.	51
3.5	Correlation values $(r)$ between concept weights and the time $(age)$ of their	
	acquisition in different sequences.	54
4.1	Statistics for simulated concept networks	63

# Acknowledgements

At the outset I would like to thank Dr. Javed I. Khan, my graduate advisor, for teaching me everything I know. I have known Dr. Khan for almost 9 years now, first as my advisor for Master's degree and then the doctoral degree. I am still amazed by his enthusiasm for new ideas, his ability to think critically and unabashedly belief in audacious ideas. He has the tremendous ability to inspire people and dream with him. When I started with him I was fresh out of college who wanted to do work but did not know how to go about it. I thank him for guiding me from those initial stages right through to the end with assisting me in writing the dissertation document almost through 12 versions! I would specially like to thank Dr. Melton for agreeing to be a part of my defense committee and giving me invaluable inputs on the write up. Those comments and email exchanges really helped me improve the write-up and in fact helped me understand a whole lot more by getting the math right. I would also like to thank Dr. Katherine Rawson, Dr. Arvind Bansal and Dr. Denise Bedford for being a part of the defense committee and giving insightful feedback. I would also like to thank Dr. Kazim Khan of the Math department for helping me out with some statistical derivations. He is probably the coolest Math teacher I've known!

Marcy Curtiss and Sue Peti have been helping international students like me for almost forever and like most others I have been a beneficiary of their warmth and gratitude. I would like to thank them from the bottom of my heart for helping out not only me but many others like me. I would specially like to mention Dr. Yuri Brietbart, who unfortunately is no more amongst us. A chance meeting with him in the Spring of 2007 in fact gave me the initial idea which eventually evolved into the main problem. I regret I did not get to thank him personally.

I have spent a good part of my formative adult life in Kent. I have learnt a lot from this little college town and I have a lot of memories attached with places in Kent. Memories which are enriched by my very good friends - the two Amits', Sajid, Parag, Sid, Shandy, Ashish, Pradeep and many more whom if I begin to mention, the list would get prohibitively long! Everyone one of them have helped make my time in Kent probably the most enriching experience ever. A home away from home!

For the last 9 years I have live away from home in the pursuit of a something which my parents and family didn't understand why I needed! I cannot thank them enough for bearing with me for all these years. I would like to specially thank my grandfather, who passed on during the course of this degree, for his inquisitive nature and quite determination which I can only aspire to achieve. My parents are the reason for whatever I am today and I am thankful for their unwavering support. I would also like to thank my fiance and soon to be wife, Amruta Kulkarni, for bearing with me and being patient through these final stages of the completion of the degree as the last of my graduate student life comes to an end!

Finally, I apologize if I have forgotten to mention anyone. In spite of all the help I am sure to have made mistakes, all of which, of course, are my responsibility alone.

This dissertation is dedicated to my parents, Jidnyasa and Sudhakar Hardas, who have been constant pillars of support in all my endeavors.

# CHAPTER 1

## Introduction: Text Comprehension

So please, oh please, we beg, we pray, go throw your TV set away, and in its place you can install, a lovely bookshelf on the wall. -Roald Dahl, Charlie and the Chocolate Factory, 1964, pp. 145

Reading is one of the most fundamental ways in which a child will obtain knowledge about the external world. A child starts with learning the "a-b-c's", then moves on to the rhymes and fairy tales and then on to bigger books with more advanced subjects and complex concepts. Books will take the child to the depths of the oceans and the middle of the desserts, to distant planets and far off lands, become a fighter pilot in a war or a social worker in the jungles of South America, cure cancer or write another fascinating book. Throughout childhood and well into the adulthood, the child will learn new things and get a better understanding of the world around by reading and comprehending text. Human beings have the capacity to acquire knowledge and generate a representation of the knowledge which will help them in comprehending any situation. People continuously generate representations which help them in explaining the perceived data. These representations vary from person to person.

Reading comprehension is a very distinguished form of knowledge acquisition because of the property of text as a highly structured method of information delivery. A person can only read one sentence at a time and has to pause between sentences to build some intermediate understanding before proceeding with the next sentence [Just, Carpenter 1982]. Understanding by reading is therefore discretely episodic and state based. Picture comprehension on the other hand may require non-sequential processing [Schnotz 1993; Schnotz, Grzondziel 1996] and involves more complicated cognitive processes.

It is known that in text comprehension readers breakdown the presented information into concepts which are relevant or irrelevant from the standpoint of understanding the topic [Wittrock 1990, 1992; Kintsch, 1988, 1994]. The relevant concepts are recognized and then integrated into the existing body of knowledge which takes the reader to the next state. The concept knowledge can be represented as a complex network of nodes and links, often referred to as the concept network. The nodes represent the concepts (or words) and the links between the nodes signify a semantic relationship. In this document isolated references to the word network means concept network unless explicitly stated otherwise.<sup>1</sup>

Thus, text comprehension can be thought of as a process in which a readers current state of comprehension, which is represented by his concept network, undergoes continuous transformation by the addition of newly recognized concepts and associations in an episodic manner until the last episode is completed. The concept network of the individual passes through as many intermediate states as there are episodes to ultimately reach the final state at the end of the last episode. Episodes are periods in which concepts are presented for recognition. Usually in the context of text comprehension, episodes

<sup>&</sup>lt;sup>1</sup>In this document concept networks refer to networks constructed from connecting concepts with associations. The associations are allowed to have different semantics associated with them. However, in this dissertation the semantics are not of vital concern and therefore not used. Therefore concepts networks are much more flexible than other methods of representation of knowledge like semantic nets or associative nets or other schemas like scripts, frames or conceptual dependency diagrams.

are sentences. Thus, text comprehension is a process of segmentation and integration of concepts from text.

So, in the attempt to model a phenomenon (like text comprehension) two things are the most important: the process and the product. The process or processes are the various cognitive tasks which are carried out during text comprehension while the product is the concept network. The objective of any process is to create the desired product. In this context, the objective of the cognitive processes of text comprehension is to construct the concept network. This is done in a step-wise manner in which at every step the concept network moves closer to a final state. Every concept network is the direct manifestation of the current comprehension state and every comprehension process seeks to drive the concept network towards the final state. The final state is the state in which the concept network of the individual denotes his level of comprehension after having read the entire text. However, reaching a final state is different than having performed the "act of comprehension".

So, is the final state same for all individuals? And how is having performed the act of comprehension defined in terms of the concept network? As stated earlier, the final state is the state an individuals concept network is in after the last episode is complete. Obviously, it is desirable that the final state signify a condition of complete and deep understanding of the text<sup>2</sup>. This condition is precisely the condition by which the act of comprehension is identified. The final state is different for every individual. However, two individuals can form two distinct concept networks and still achieve deep comprehension

 $<sup>^{2}</sup>$ The assumption is that reaching the final state can serve as the necessary condition for evaluating the act of comprehension i.e. the quality of comprehension.

of the text. So how does the concept network in its final state signify the condition of having performed the act of comprehension? The only criterion on which the act of comprehension is contingent is that the concept networks of the individuals in their final states must contain at least a subset of concepts from the text which are required for the comprehension of the text. To explicitly state the above:

- 1. The final state is different for every individual and there is no external comprehension state that every individual must aspire to achieve. Individuals can reach their own final states and still successfully perform comprehension. The final states for different individuals do not necessarily have to be different either.
- 2. The necessary condition for having performed the "act of comprehension" is that the concept network of any individual in its final state must contain a subset of concepts from the text. This also means that final state for any individual reader cannot be reached before the end of the last episode. This is because every episode is assumed to contain at least one new concept. Therefore, until the last episode is completed, and the set of concepts in that episode are segmented and integrated comprehension cannot be completed.

Both these statements have profound implications. According to the first assumption, no two individuals have same concept network and also there is no perfect state of comprehension which the reader must achieve. This is loosely analogous to the theory of constructivism [Piaget 1957]. It states that all comprehension (acquisition of meaning) is an act of construction instead of search in an external reality (ontological reality)<sup>3</sup>.

<sup>&</sup>lt;sup>3</sup>Ontology is the philosophical study of the nature of being, existence, or reality as such, as well as the basic categories of being and their relations [ONT 2012].

Constructivists believe that comprehension is an act of knowledge internalization through processes of assimilation and accommodation. In this sense the model of segmentation and integration is an implementation of the constructivist epistemology in terms of a dynamic network model. This is discussed in further detail in Section 6.4

The second assumption is that although the final states for individuals may be different, their concept network in this state *must* contain a subset of concepts from the text to make the claim that comprehension has been performed. The explanation for this assumption is in the answer to the question "what is the purpose of a coherent and purposeful text?" The objective of a text is to deliver to the reader a set of concepts in an episodic fashion so that the reader can gradually build up a representation of the final state for that particular text. Therefore, the product of comprehension here is not a single concept, but a set of concepts. For example, consider a paragraph from a biology text which talks about the concept of lungs. Let this paragraph contain four other concepts namely breathing, carbon dioxide, oxygen, and blood which are directly related to gaining an understanding about lungs. Without exception, all of the concepts directly related to lungs must be acquired from the text by the reader for comprehension. However, it is also possible that one individual in his final state may contain other concepts like "carbon monoxide" and "pollution" and another individual may contain "mountains" and "clean air" in his final state. The final states of these two individuals are different, but both of them are said to have comprehended the text because they also have acquired the basic four concepts about lungs from the text. If one of the four concepts is absent, then it is said that deep comprehension is not performed. We now define deep comprehension as the acquisition of a subset of concepts from the text required to understand the main topic the text is about.

This does not however mean that the concept of lungs can be fully explained by only this set of concepts. This only means that those concepts in the text which are pertinent to understanding the concept of lungs must be acquired. This belief stems from a moderate atomistic understanding of concepts in which complex concepts are "assembled" [Laurence, Margolis 1999 pp.62-63] from primitive concepts<sup>4</sup>. This will be further discussed in Section 6.3.

Note here that there is a distinct difference between reaching a final state for the *concept* (say lungs) and reaching a final state for a *text*. The former involves the acquisition of many more primitive concepts which might not be present in the text. The later involves the acquisition of only the concepts presented in the text about lungs. Thus, reaching a final state of a text is not possible, i.e., comprehension is incomplete, without the acquisition of a subset of pertinent concepts from the text.

Since there is no ontological reality, i.e., there is no "perfect" representation of the concept of lungs, the concept of lungs is constructed individually by every reader with the help of the text. This is the reason why no two individual concept networks look the same. However, according to the second assumption, constructing the concept of lungs with respect to the text requires the acquisition of the primitive concepts given in the text. Thus, to achieve the state of comprehension of a concept, it is essential to achieve a state of comprehension of a text. The purpose of the text now correctly becomes to deliver the primitive set of concepts which are required for the construction

<sup>&</sup>lt;sup>4</sup>In conceptual atomism and representational theory of the mind, a primitive concept is a concept which cannot be broken down further into composite primitive concepts [Fodor 1975]. For example, the concept "green" can be thought of as a atomistic concept, where as "traffic" is probably a complex concept composed of primitive concepts like "cars", "highway" and many more.

(or assembly) of the complex concept. However, remember that although the concepts in the text might be complex themselves, they are primitive from the point of view of the higher level concept (i.e. lungs) that the text is trying to explain. Of course, the implicit assumption here is that to acquire any concept it is necessary to first acquire a basic set of primitive concepts. This, however, is in line with the innateness of knowledge proposed by rationalists for over two millennia.

The first part of this dissertation will concentrate on modeling the process of segmentation and integration of the recognized concepts (SI). The process of SI is directly responsible for the construction of the concept network of the primitive concepts given in the text. This process is repeatedly carried out in every episode until the final state is reached. The model is a framework with a representational system and an algorithm to compute the solution.

Since the final state of comprehension for each individual is different, measuring the result of comprehension is a challenge. One way to externally measure the final state is by measuring reader performance on recall and inference type tests specifically designed for that text. A good score would mean that the individual has comprehended the text. This method of measuring the success of comprehension has been studied in much detail [Kintsch 1994; Mannes, Kintsch 1987]. Another way to measure the final state is by analysis of the internal concept network. In the last part of this dissertation concept network growth in comprehension is simulated to form artificial networks of comprehension. The dynamical structural properties of these networks are analyzed to make comments about the quantity and the quality of comprehension from a theoretical viewpoint. Until now the principal focus of research in text comprehension in the psychological sciences has been to understand the cognitive processes of text comprehension and other factors like reader goals, motivations, attention, skills, etc. rather than study the concept network. As a result many theories about the cognitive processes have emerged but few theories related to the representation of knowledge as a concept network<sup>5</sup> [Kintsch 1988; van Dijk, Kintsch 1983; van der Broek et al. 1995]. Much research has been devoted to the understanding of the role of syntax, morphology, contextual disambiguation, etc. of the words and sentences which are heavily dependent upon context. The obviously algorithmic aspects of knowledge acquisition, like the SI process, seemed to have enjoyed little attention from a mathematical modeling perspective.

On the other hand, computer scientists studying machine natural language understanding have concentrated on the product but not paid much heed to the intricate processes described so well by psychologists. For example, the CYC common sense knowledge project [Randall, Lenat 1982] and strict rule based systems to create machine understandable data from natural language like MARGIE [Schank 1975], SHRDLU [Winograd 1972], etc. are good examples of efforts to create understanding systems. Most modern computer science techniques of text understanding are deeply inspired by statistical natural language processing. The main aim is to induce a semantic space using machine learning algorithms for semantic information retrieval from a large corpus of text. LSA [Landauer, Dumais 1998], LSI [Deerwester et al. 1988; Hoffman 1999], and topics models

<sup>&</sup>lt;sup>5</sup>The reason for this may be due to the advent of algorithmic information retrieval from large corpora of text to induce a high dimensional semantic space in which words are represented as high dimensional vectors and computations for similarity often represent associations between words. To their credit these techniques have fared reasonably well in representing human knowledge in comprehension [Landauer et al. 1998; Kintsch 2000, 2001].

[Griffiths, Steyvers 2003; Steyvers, Griffiths 2007] like LDA [Blei et al. 2003, 2004] are a few examples.

This does not mean that the computer scientists have ignored the process aspect of text understanding. Indeed there are algorithms for many of the mechanistic aspects underscoring language text like parsing, named entity recognition, parts-of-speech tagging, segmentation, word sense disambiguation, etc. Thus, although computer scientists have invented increasingly efficient algorithms for semantic space induction and syntactical, structural and even semantic processing, it has not been in the context of theories of comprehension. Interdisciplinary efforts to use machine induced semantic space representations in context of theories from psychology have been relatively scarce until recently [Kintsch 2000; Kintsch, Mangalath 2011].

The aim of this dissertation is to explore a natural phenomenon (SI) and present a mathematical model for it. The model has its own deficiencies and presents scope for improvement. In the process however, the exercise of mathematically modeling a natural phenomenon also ends up providing interesting insights into the cognitive process of text comprehension and the very nature of knowledge and its acquisition.

# 1.1 Text comprehension in literature

This section will look at the current understanding about the process and the product in text comprehension. The product is the concept network while the processes are the cognitive processes of text comprehension.

#### 1.1.1 Knowledge representation in text comprehension

There are three important questions concerning knowledge representation during comprehension:

- 1. How is the comprehended knowledge represented?
- 2. How is the knowledge represented in the intermediate stages of comprehension?
- 3. Where is the knowledge stored?

There is good evidence that knowledge is represented is some form of a cognitive concept network. Computer scientist and linguists have often assumed these concepts and relations to be discrete in nature and used artificial concept networks to represent cognitive knowledge. Concept networks are complex networks<sup>6</sup> of concepts and associations. In linguistics, concept networks are often networks of words (lexicon) referred to as the semantic networks. Thus, to linguists the acquisition of the lexicon is analogous to the acquisition of the semantic network<sup>7</sup>. The associations between two concepts may be loaded with semantics such as similarity, part-of, has-a, is-a, etc. The associations may also have weights which indicate to the strength of the relationship. The idea of concept networks as a representation of organized knowledge in memory is quite appealing. Knowledge organized into semantic or a concept network is not only an intuitive representation but also allows for different kinds of computations like search and retrieval to be performed relatively easily. Concept networks have been represented as semantic

<sup>&</sup>lt;sup>6</sup>In the context of network theory, a complex network is a graph with non-trivial topological featuresfeatures that do not occur in simple networks such as lattices or random graphs but often occur in real graphs. [CN 2012].

<sup>&</sup>lt;sup>7</sup>Lexicon acquisition is a process studied in much detail in cognitive linguistics. Some of the theories of lexical acquisition are [Carey 1978; Clark 1993, 2001; Slobin 1973]. However, acquisition of a lexicon is very different than the construction of a concept network.

networks with tree like hierarchies [Collins, Quillian 1969], unstructured graphs [Collins, Loftus 1975], bipartite graphs [Miller et al. 1995, Roget 1911] or generic associative networks [Anderson 2000; Nelson, McEvoy, Schreiber 1999].

Some researchers have also suggested that comprehended knowledge may be represented as a high level interconnected scheme of concepts called as cognitive schemata [van Dijk, Kintsch 1983; Johnson-Laird 1983; Carrell, 1984b; Schnotz, Grzondziel 1996]. Mayer (1993) suggests that the concept network is "a mental representation consisting of parts that interact with one another according to principle-based rules". For example, in the case of literary texts, a mental model can be models of settings, events, and characters in a novel. When reading a text, mental models are created based on propositional representations, using prior knowledge and experiences in the form of existing cognitive schemata. The concept network has also been thought of as an internalized object that describes the relationship between the subject matter and the individual during comprehension [Chun, Plass 1997].

The second concern is how is the knowledge represented in the intermediate stages of comprehension? Psychological experimentation suggests that multiple types of mental representations might be involved at multiple stages of comprehension. One of most influential theories on text comprehension is the construction-integration model by Kintsch (1988). It suggests that comprehension involves first building a mental representation of the text's propositional representation which is called as the "textbase" and then integrating the textbase with the background knowledge to form a mental model of the subject matter called as the "situation model". Several researchers [Schnotz 1993; Schnotz, Grzondziel 1996] suggest that text represents information in symbolic or sub symbolic structures of a language and is processed quite sequentially, that is, word by word or sentence by sentence. Schnotz (1993) suggests that the construction of a mental model from text also requires the construction of propositional representations, which then have to be integrated into the mental model.

Thus, it is usually believed that text comprehension results in a gradual building of some form of a high level concept network by connecting new concepts to existing concepts. Though the intermediate forms of representation might be different the final understanding is independent of language and method of learning. Given the general lack of our understanding a great many researchers have proceeded with some form of concept network to portray the relationships between concepts captured from the text. Also, a formal model necessitates the need for a standard representation of knowledge in all the stages of processing. A concept network representation is most appropriate for this exercise.

The answer to the question, "where is knowledge stored?" is in the semantic memory. Semantic memory can be long term or short term. It is usually thought that the existing knowledge is stored in the long term memory and during comprehension this knowledge is called upon to make interconnections with the concepts in the short term memory (obtained from reading the text) to generate a cohesive state of comprehension. Arguments about long and short term memory are further discussed in Section 6.2.

# 1.1.2 Cognitive processes in text comprehension

This section will discuss,

1. The different cognitive processes in text comprehension

- 2. How these process can be elementarily broken down to basic processes
- 3. The generative nature of the comprehension process

# Evidence for SI process

The central objective is to identify the exact processes which the reader's abilities will be employed to perform. Various processes employed in the construction of a representation have been identified, like word identification, sense disambiguation, syntactic and semantic parsing, etc. [Kintsch 1988, van den Broek et al. 1995 (landscape model); Gerrig, McKoon 1998; Myers, OBrien 1998]. It is generally recognized that the understanding of written text calls upon both bottom-up word recognition processes and top-down comprehension processes. The process of word/concept recognition (segmentation) and integration with prior knowledge seems to be at the heart of comprehension [Verhoeven, Perfetti 2008]. Studies in eye fixations during reading [Just, Carpenter 1980, 1982] have shown that readers tend to fixate on specific words longer than others. The process of word identification is a discrete process of recognition and sense making in which a group of words are collected and then integrated with background knowledge (immediacy assumption).

Reading is viewed as a set of ordered stages, consisting of a beginning state, an end state, and intervening transformations [Just, Carpenter 1987]. This is nicely summarized in the language of [Swaffar, Arens, Byres 1991] who define comprehension as a process that "fluent readers synthesize textual subsystems (e.g., content, context, intent, language) into a larger meta-system of meaning" and that "in short, readers comprehend a text when they construct a mental representation for incoming pieces of information". de Beaugrande (1982), among others, posits that what is in fact comprehended is not a sentence, but the conceptual content. The process steps range from recognizing letters, characters, and words, to analyzing the syntactic and semantic structure of clauses and sentences, and these must take place in an orchestrated manner [Rayner, Pollatsek 1989].

The three most common lines of thought are bottom-up processing, top-down processing, and interactive processing [Samuels, Kamil 1984; Silberstein 1987; Swaffar et al., 1991]. Some researchers believe that processes are hierarchically related to one another [Horiba 1996]. The dominant view is that that these processes, however, take place nonlinearly [Just, Carpenter 1980]. While readers are active in the selection of portions of the text for processing, former portions of the text may inform latter ones, just as latter portions of the text may inform former ones through feedback.

To summarize, there is evidence for the existence of a SI like process of text comprehension.

#### Basic semantic memory tasks

All high level cognitive processes (like comprehension) can be elementarily broken down into the three core semantic memory tasks: encoding, storage and recall. Encoding is the process in which the incoming knowledge is converted into a representation which is usable later for recall. Encoding of information implicitly involves the recognition of words and alphabet symbols in the context of text comprehension. It has been observed that proficient readers have the ability to recognize words more quickly and effortlessly [Adams, Marilyn and Jager 1994]. Sensory input obtained from the audio, visual or tactile faculties is not usable directly but has to be converted into constructs which can be used later for recall. Sometimes the act of locating relevant information and recognizing this information is also referred to as tracing and chunking.

Storage is the processes in which the encoded information is actually stored and retained in the semantic memory or the episodic memory using some representation for a short period or a prolonged period of time. Active information is stored in the short term memory while the passive in long term. Storage is a core semantic memory processes and efficiency in storage dictates the ability to perform other tasks efficiently like recall and encoding too.

Recall refers to the process of retrieving the encoded information which is stored in the memory. Psychologists suggest that there are three forms of recall namely, free recall, cued recall and serial recall. In free recall an individual is asked to read a list of words and then remember them in any order. In cued recall a person is asked to read a list of paired words and asked to reproduce one of the words when cued with the other word from the pair. In serial recall the individuals are asked to reproduce the words in the exact order as the given list. The ability of humans to store items in the memory and recall them when ever needed is an important in the use of language.

These semantic memory processes operate directly upon the concept network in the memory. For instance, the task of recall can be imagined as searching through the concept network to find the most relevant concept by comparison and then retrieving. By the two stage theory of the recall process, recall begins with a search process and then a recognition process in which the correct information is retrieved from the results [Watkins, Gardiner 1979]. Search can be thought of as a simple spread activation and retrieval of the activated nodes.

The process of SI can also be understood in terms of these elementary operations. Segmentation can be thought of as the task of encoding the recognized concepts and then associating them with the background conceptual knowledge. Through the task of encoding, the concept knowledge from the text is transferred from the text and encoded in the form of concept network. Therefore, the ability to perform well on recall and inference type tests is highly dependent on the ability to encode the concept knowledge well into the concept networks.

# The generative nature of comprehension

Comprehension is a fundamentally constructive process of building meaning [Kintsch 2009]. Meaning (concept) is coupled to the structure of the concept network of a reader. The meaning of a concept (or word) is not only derived from that concept but also from its neighboring concepts. In short, meaning is "contextual". In comprehension, meaning is an act of construction which is reflected in the structure of the concept network.

Human beings do not passively consume text but actively generate representations [Wittrock 1990, 1992]. The operative word here is generate because models of the perceived raw information are generated during comprehension. This generative nature of learning and comprehension has been shown in teaching science [Osborne, Wittrock 1983, 1985], mathematics [Peled, Wittrock 1990; Wittrock 1974a] and even economics [Kourilsky, Wittrock 1987]. In short there is evidence that reading comprehension is a generative process of "meaning" construction and model formation [Wittrock 1974b; Wittrock 1977; Wittrock, Marks, Doctorow 1975]. We model this process of meaning generation as a purely graphical process of concept network growth in which nodes and edges are added to the graph in the pursuit of generating meaning from perceived information.

#### 1.1.3 Other factors in text comprehension

There are other factors which affect text comprehension. These factors are outside of the SI model and not discussed in depth in this dissertation. Examples of other factors are; reader skills, reader goals, syntax, and writing quality of text. However, it is important to note that the model presented in the following pages does incorporate some of the other factors upon which text comprehension is contingent. Two of them are 1) the difference between readers skills and background knowledge and 2) quality of the text.

It is generally understood that the process of text comprehension involves both high level skills (such as understanding and maintenance of a concept space) and low level skills (such as text processing skills) [Goodman 1967; Smith 1971, 1979; Landauer et al. 1997, 1998; Grabe 1991]. Principally six types of skills are posited: automatic recognition skills, vocabulary and structural knowledge, formal discourse structure knowledge, content/world background knowledge, synthesis and evaluation skills/strategies, metacognitive knowledge and skills monitoring. Some researchers have also ranked these factors in terms of their criticality in the readers comprehension performance. For example, Laufer and Sim (1985) posit the following factors in order of decreasing importance: knowledge of vocabulary (concept network), subject matter, discourse markers and syntactic structure. Many posit that vocabulary is the most important and syntax the least.

Another important aspect which determines the process of comprehension is the goal of the reader. A deep understanding is only achieved if the reader consciously aims to achieve a deeper understanding by actively processing the text, making connections with the prior knowledge which are not explicit, inferring from the presented data, filling in the coherence gaps left by the text itself, generating a macro structure from the micro structure given in the text, etc. A passive reader who does not perform these tasks may only be able to build a superficial understanding of the subject. If the goal of the reader is mere fact retrieval, he will most likely focus of different aspects of the text while others who read to gain a deeper understanding and inference will focus on something else. A well written text certainly has cues for building up a coherent representation of the given text and the readers motivations also play an important role in using these cues.

There are many more important factors upon which text comprehension is contingent like attention, quality of the text, time required for reading, time between reading and testing, etc. but we will not discuss those.

### 1.2 Thesis contributions

The main contribution of this thesis is the mathematical framework for the concept segmentation and integration process (SI). The model can explain how and why different readers construct different concept networks on reading the same text. It can also describe why some readers may understand a text easily as compared to other readers, and also why some texts are difficult to understand than other texts for the same reader. The model is also used to explain the effect of the age of acquisition of a concept on comprehension. It is seen that earlier a concept is acquired the more important it is for comprehension of other concepts.

The model leads to an algorithm which is used to simulate concept network growth

during text comprehension. These networks are then analyzed to investigate their structural properties. It is seen that these networks are small worlds with high local clustering and a normal degree distribution. These properties are indicative of the high connectivity and reachability not observed in similar random networks. It is also seen that although concept networks may start off with multiple disconnected components, the process of comprehension leads to most of the nodes getting connected to form a single giant component.

- 1.3 Thesis organization
  - Chapter 2 describes in detail the computational model for SI. The model is used to learn the association strengths of fully connected concept network. Data is collected from experimentation and an example of the working of the model is given.
  - 2. In Chapter 3 the SI model is applied to inquire about the impact of the sequence of concept presentation and acquisition on the finally comprehended concept network. Data is collected through classroom experiments and the analysis is used to test the hypothesis.
  - 3. Chapter 4 presents the concept network growth algorithm based on the SI model. Network growth is simulated and the structural properties of these networks are computed. Applications of such a growth model are discussed.
  - 4. In chapter 5, the algorithm is used to simulate network growth and observe the evolution of connectedness of knowledge.
  - 5. Chapter 6 presents some of the interesting discussions that arise from this work.

# CHAPTER 2

#### Computational segmentation and integration

"... to completely analyze what we do when we read would almost be the acme of the psychologist's achievements, for it would be to describe very many of the most intricate workings of the human mind ..." - Edmond B. Huey, the Psychology of Reading, 1908, p. 6

In this Chapter we will present the mathematical framework for the SI model. The chapter starts with an example of the SI process and then proceeds to explain the model in detail. Some of the issues which can arise from the mathematics are discussed and solutions are presented. Finally, an example is discussed which is used for validating the model.

## 2.1 An example of concept network growth by SI

Consider the example shown Figure 2.1a. It shows a concept network generated by recognizing some concepts and discarding some concepts from an example sentence. There are many different ways in which the concept network might be constructed. A deeper understanding cannot be achieved unless the recognized concepts integrate with the background knowledge by associating with existing concepts. As shown in Figure 2.1b a deeper understanding is achieved by integrating the recognized concepts with the background concepts like lungs and other concepts like "release  $CO_2$ " and "inhale  $O_2$ ". The background concepts are shown by the dotted lines. This deep understanding is represented by a more elaborate concept graph which includes the background concept knowledge as shown in Figure 2.1b.



(a) Sentence 1: "When a baby has a septal defect, the blood cannot get rid of enough carbon dioxide through the lungs. Therefore, it looks purple."



(c) Sentence 2 - "This defect maybe caused by "Infant respiratory distress syndrome" (IRDS), a pulmonary alveolar disease caused by the lack of surfactant in the babys lungs."

Figure 2.1:



Figure 2.1: Example of concept network growth.(a) A toy concept network constructed for the example sentence 1 (b) Interaction with background concepts (c) An example construction of concept network after comprehending second example sentence 2. The concept network in (b) is now the background concept knowledge for (c). (d) Textbase and Situation model for example sentence 1 [Kintsch 1994].

Now consider another example sentence which is presented to the reader. In the next sentence the reader is presented with more concepts and the reader again recognizes a few concepts from this sentence and relates them to the background knowledge. The Figure 2.1c shows a more evolved concept network after completion of two episodes of comprehension. In this episode, all of the concepts acquired in the concept network, including the background concepts and those acquired in the previous sentence are treated as the background knowledge. In this manner, with the progression of every episode a deeper understanding of the subject matter is attained. The above example shows how the process of text comprehension can be conceptualized as selective recognition of concepts from texts and building up an incremental understanding of the concept knowledge. Kintsch (1994) describes this process as construction-integration in which
first a textbase is formed from the text surface semantic structure and then a situation model (state of comprehension) is formed by integrating it with the background concepts, Figure 2.1d.

# 2.2 Computational model for SI

At this point the following hypotheses are made about the SI process model;

1. Hypothesis 1 A concept x is recognized if the cumulative sum of the association strengths to this concept from existing concept is greater than or equal to than a threshold value <sup>8</sup>. This is called as segmentation of concepts (S).

$$S(x) = \begin{cases} 1 & f_1(x) \ge \alpha \\ 0 & f_1(x) < \alpha \end{cases}$$
 given  $f_1(x) = \sum_{i=1}^n w_i$  where  $(w_1, w_2, ..., w_n)$  are the weights of the associations from existing concepts  $(1, 2, ..., n)$  to the new concept  $x$ 

and  $\alpha$  is the threshold. S(x)=1 implies concept is recognized.

2. Hypothesis 2 A concept x is integrated with another existing concept y if the strength of association between the two concepts is greater than or equal to a threshold. This is called as integration of concepts (I).

 $I(x,y) = \begin{cases} 1 & f_2(x,y) \ge \beta \\ 0 & f_2(x,y) < \beta \end{cases}$  given  $f_2(x,y) = w_{x,y}$  where  $w_{x,y}$  is the strength of the association between x and y and  $\beta$  is the association threshold. I(x,y)=1 implies association is recognized.

The test for validity of these hypotheses if whether it can explain all the possible observed data. In the rest of the chapter we will go about testing the validity of the model.

<sup>&</sup>lt;sup>8</sup>The cumulative sum is analogous to the weighted average in artificial neural networks which is in turn inspired from the integrate-and-fire mechanism of the physical neuron. The threshold acts like a step function. It can also be modeled with a continuously differentiable sigmoid function where,  $f_1(x) = \frac{1}{1+e^{-x}}$ 

### 2.2.1 Model parameters

The model parameters in the SI model are discussed here.

# 1. Background knowledge

The most variable factor in the model is the background knowledge possessed by every reader. Therefore, in the model every readers background knowledge is represented by a single node in the concept network called the background concept.

# 2. Episode comprehension factor

The comprehensibility of an episode is also determined by the syntax, clarity of writing style, linguistic structure and other issues generally pertaining to the quality of the text. This variability is represented by another node in the concept network, one for each episode, and is called the episode comprehension factor.

3. Association strength

It is the scalar weight value of the association between two nodes in the network. The value of association weight varies depending upon what kind of an association it is. If it is an association between two concepts then it is a positive value between 0 and 1 and it does not vary from person to person or episode to episode. If it is an association between a background node and a concept node, or an episode comprehension node and a concept node, or between a background node and episode comprehension node then it can have negative values and may vary from person to person or episode to episode.

4. Concept recognition threshold

In the SI model the recognition of a concept is modeled as a threshold phenomenon. Every individual has a different threshold to recognize a concept. This is because every individual has different skills, motivation, attention spans, etc. More over these factors can change during the process of comprehension. Therefore, the concept recognition threshold variable varies from person to person and may vary from episode to episode too.

5. Concept association threshold

In the SI model, integration of a concept with the background knowledge is also a threshold phenomenon. Similar to the concept recognition threshold, the association threshold varies from person to person.

# 2.2.2 Notation

Let a piece of text  $\Omega$  be represented by a concept network of concepts and associations. This network is called as the *base concept network (BCN)*. The BCN contains set of concepts which are required to achieve a state of comprehension of the text. Any reader who acquires these concepts and associations is said to have reached the state of comprehension.

A BCN is represented as a graph,  $G_{\Omega} = \{V_{\Omega}, E_{\Omega}\}$  where  $V_{\Omega} = \{v_i | 1 \leq i \leq n\}$  is the set of vertices (or nodes or concepts) of the graph and  $E_{\Omega} = \{(v_x, v_y) | 1 \leq x, y \leq n \text{ and } v_x, v_y \in V_{\Omega}\}$  is the set of edges (or links or associations) between the all the concepts in the BCN.  $E_{\Omega}$  is the relation which represents all possible edges in the network. Every element in  $E_{\Omega}$  represents a single edge. If an edge is given by e then the relation can also be represented as  $E_{\Omega} = \{e_i | 1 \leq i \leq l \text{ where } l = \frac{n(n-1)}{2}\}$ . There is a function  $\varphi_{\Omega}$  which associates via a one-to-one mapping the two representations of the elements in  $E_{\Omega}$ .  $\varphi_{\Omega}$  is called the edge mapping, and it is given by  $\varphi_{\Omega}(e_i) = (v_x, v_y)$  where  $v_x$  and  $v_y$  are endpoints of  $e_i$ . The weight of an edge e is given by  $w_e$ .

Let there be a total of  $\mu$  sentences in the text. Therefore, there are also  $\mu$  episodes each corresponding to reading a sentence. The episodes are represented by the times they are said to occur  $(t_1, t_2, ..., t_{\mu})$ . Thus, the episode  $t_2$  is said to start at the end of  $t_1$  and end before  $t_3$ . The episode comprehension factor node is represented as  $v_t$ . Since every episode is different, a distinct node is needed for every episode. The set of episode comprehension nodes is  $(v_{t_1}, v_{t_2}, ..., v_{t_{\mu}})$ . The positive or negative comprehensibility of a sentence is signified by different association strengths from this node. This will be further discussed in Section 2.4.2.

The process of comprehension for an individual is defined as a process in which an individual reader  $\theta$  reads the text  $\Omega$  and constructs a concept network of his/her own. This is called as the *individual concept network (ICN)*. It is a subset of the base concept network. The ICN is represented by a graph  $G_{\theta} = \{V_{\theta}, E_{\theta}\}$  where  $\theta$  is the reader and  $G_{\theta} \subseteq G_{\Omega}$  i.e.  $V_{\theta} \subseteq V_{\Omega}$  and  $E_{\theta} \subseteq E_{\Omega}$ . For the ICN, the edge mapping function is represented by  $\varphi_{\theta}$ .

The reader starts with an initial individual concept network which only contains one concept denoted by  $v_{\theta}$  for a particular reader  $\theta$ . This represents the readers background knowledge. The importance of this background node is discussed further in Section 2.4.1. Comprehension occurs on a sentence by sentence basis. A reader reads a sentence, recognizes some concepts from that sentence and then connects the recognized concepts to the existing concepts in the concept network. In this process the ICN is transformed from  $G_{\theta}^{t} = \{V_{\theta}^{t}, E_{\theta}^{t}\}$  to  $G_{\theta}^{t+1} = \{V_{\theta}^{t+1}, E_{\theta}^{t+1}\}$  in the episode t. The set of concepts presented in episode t is given by  $D^{t}$ . This is called as the set of *latent concepts* available for recognition in this episode. It is important to note here that a concept may be repeatedly presented in many episodes. If the concept is not recognized in an episode then it may be recognized in a latter episode when it is presented again. If, however, the concept is recognized in an episode then it cannot be recognized again in a later episode (since it now exists in  $V_{\theta}^{t}$ ).

The set of possible associations which are generated when new concepts are presented in an episode t is represented by  $L^t$ . It is called as the set of *latent associations* available for recognition in an episode and is given by,

$$L^t = \{e \in E_\Omega \mid \exists i \in V^t_\theta \text{ and } j \in D^t \text{ and } \varphi(e) = (i, j)\}$$

. The set  $L^t$  does not contain the associations which might exist between the concepts which are presented in the same episode.

Out of the presented concepts a reader recognizes a subset of vertices,  $R_{\theta}^t \subseteq D^t$  and subset of edges,  $S_{\theta}^t \subseteq L^t$ . Then the set of vertices and edges for the resultant graph at time t + 1 is given by  $V_{\theta}^{t+1} = \{V_{\theta}^t \cup R_{\theta}^t\}$  and  $E_{\theta}^{t+1} = \{E_{\theta}^t \cup S_{\theta}^t\}$  respectively.

# 2.2.3 Segmentation of concepts

In this step the set of recognized concepts  $R_{\theta}^{t}$  is formed by evaluating the comprehension strength  $(\delta_{v})$  for each concept v in  $D^{t}$ . The concepts for which  $\delta_{v}$  greater than or equal to a certain threshold  $(\alpha_{\theta})$  are recognized and added to  $R_{\theta}^{t}$ . The comprehension



Figure 2.2: Example of constraints formation in segmentation and integration.

strength for a node v is sum of the association strengths of the links between concept vand the concepts in  $V_{\theta}^{t}$ . Thus,

 $R_{\theta}^{t} = \{v \in D^{t} | \delta_{v} \geq \alpha_{\theta}\}$  where  $\alpha_{\theta}$ =concept recognition threshold for reader  $\theta$ , and  $\delta_{v} = \sum w_{e}$  where  $\varphi(e) = (i, v)$  and  $i \in V_{\theta}^{t}$ .

The set of unrecognized concepts for episode t is,  $\overline{R_{\theta}^t} = D^t - R_{\theta}^t$ .

# 2.2.4 Integration of concepts

In the next step, the recognized association set  $S^t_{\theta}$  is formed by evaluating the association strengths of edges. If this value greater than or equal to a certain threshold  $\beta_{\theta}$  for the reader  $\theta$  then that edge is recognized.

 $S_{\theta}^{t} = \{e \in L^{t} | w_{e} \geq \beta_{\theta}\}$  where  $\beta_{\theta}$ =association recognition threshold for reader  $\theta$ .

The set of unrecognized concept is,  $\overline{S^t} = D^t - S_{\theta}^t$ .

### 2.2.5 Constraints formation in SI

Now let us look into an example of how constraints are generated for the segmentation and integration of a concept in the SI model.

Figure 2.2a shows the processes of segmentation. Let the set of concepts in the concept network of a reader  $\theta$  in episode t be  $V_{\theta}^t = \{v_1, v_2, v_3\}$ . At this time a set of concepts  $D^t = \{v_4, v_5\}$  are presented out of which only concept  $v_4$  is recognized (as shown by the solid line). Since  $v_4$  is recognized it follows that the summation of link strengths of the links to it exceeds the threshold  $\alpha_{\theta}$ . Therefore,

$$e_1 + e_2 + e_3 \ge \alpha_\theta$$

Also now the set of recognized nodes is given by,  $R^t = \{v_4\}$ .

The concept  $v_5$  is not recognized (as shown by dotted lines) which means that the summation of the link strengths to this concept does not exceed the threshold. Therefore,

$$e_4 + e_5 + e_6 < \alpha_\theta$$

The set of unrecognized nodes is given by,  $\overline{R^t} = \{v_5\}$ .

Thus, at the end of episode t+1 the set of vertices for the new graph is,

$$V_{\theta}^{t+1} = V_{\theta}^{t} \cup R_{\theta}^{t} = \{v_1, v_2, v_3\} \cup \{v_4\} = \{v_1, v_2, v_3, v_4\}$$

Figure 2.2b shows the process of integration. At episode t let there be no associations existing between the concepts in a concept network of a reader  $\theta$ ,  $E_{\theta}^{t} = \{\emptyset\}$ . The set of associations available for recognition is  $L^{t} = \{e_{1}, e_{2}, e_{3}\}$ . The associations  $\{e_{4}, e_{5}, e_{6}\}$  are not available for recognition because concept  $v_{5}$  is not recognized. Assume that out of the possible associations, the association  $e_{3}$  is not recognized. Since associations  $e_{1}$  and  $e_{2}$  are recognized their strengths exceed the association threshold  $\beta_{\theta}$  whereas that of  $e_{3}$  does not exceed. This is represented in equation form as,

$$e_1 \ge \beta_{\theta} > 0$$
$$e_2 \ge \beta_{\theta} > 0$$
$$e_3 < \beta_{\theta} > 0$$

(

Therefore the set of recognized and unrecognized associations are  $S_{\theta}^{t} = \{e_{1}, e_{2}\}$  and  $\overline{S_{\theta}^t} = \{e_3\}$  respectively.

Thus, at the end of episode t+1 the set of vertices for the new graph is,

$$S_{\theta}^{t+1} = E_{\theta}^t \cup S_{\theta}^t = \{\emptyset\} \cup \{e_1, e_2\} = \{e_1, e_2\}$$

Thus the process of segmentation and integration can be represented in the form of inequality constraints. In every episode the process of segmentation and integration transforms the graph from  $G_{\theta}^{t}$  to  $G_{\theta}^{t+1}$ . The process of comprehension is completed at the end of all  $\mu$  episodes in which the graph transforms from  $(G_{\theta}^{t_0}, G_{\theta}^{t_1}, ..., G_{\theta}^{t_{\mu}})$ . Thus if we have the individual concept networks generated by an individual reader at the end of every episode  $(G_{\theta}^{t_0}, G_{\theta}^{t_1}, ..., G_{\theta}^{t_{\mu}})$  then we can represent the growth in terms of a set of inequality constraints, and solving the constraints will result in the values of the associations and thresholds which satisfy the constraints.

Figure 2.3 shows the process of graph growth from  $G_{\theta}^{t}$  to  $G_{\theta}^{t+1}$  through segmentation and integration. Assume the initially learned graph  $G^t_{\theta}$  till episode t contains concepts



Figure 2.3: Transformation of a concept network in an episode.

 $V_{\theta}^{t} = \{v_{1}, v_{2}, v_{3}\}$ . At this point three new concepts are presented  $D^{t} = \{v_{4}, v_{5}, v_{6}\}$ . In the first step of concept recognition the comprehension strengths of each of the concepts  $v_{4}$ ,  $v_{5}$  and  $v_{6}$  are computed by adding up the association weights to each of the concepts from already existing concepts in  $V_{\theta}^{t}$ . Accordingly the comprehension strengths are  $\delta_{v_{4}}=13$ ,  $\delta_{v_{5}}=7$  and  $\delta_{v_{6}}=13$ . Assume that the recognition threshold for this particular example is  $\alpha_{\theta} = 10$ . The comprehension strengths for concepts  $v_{4}$  and  $v_{6}$  are greater than the threshold and therefore they are recognized (as indicated by the solid outline). In the next step the associations which have association strength less the association threshold are removed. Assuming association threshold  $\beta_{\theta}=5$ , the associations  $e_{5}$  and  $e_{11}$  are removed. The associations between any of the concepts in  $V_{\theta}^{t}$  and concept  $v_{5}$  are never considered because concept  $v_{5}$  is not recognized and is not a part of  $V_{\theta}^{t+1}$ . Thus concept network incrementally evolves from  $G_{\theta}^{t}$  to  $G_{\theta}^{t+1}$  during comprehension.

# 2.3 Computing the model parameters

On solving the constraints presented in the last section we can get the values of the associations and the thresholds. This can explain the recognition and non-recognition of every latent concept and association for that particular individual. However, any generalized model of text comprehension should be able to explain the process of comprehension for a group of readers and not just one reader. Different readers recognize different concepts and associations. In this section we will see how the model incorporates more than one reader.

Data is gathered by collecting the snapshots of the concept network of every reader after every episode. This records which new concepts and associations are recognized and/or not recognized in every episode. If the total number of readers are given by pand total episodes by  $\mu$  then the total number of snapshots of the concept networks will be  $p\mu$ . Let the individual concept networks (ICNs) for the set of readers ( $\theta_1, \theta_2, ..., \theta_p$ ) generated at the end of every episode ( $t_1, t_2, ..., t_\mu$ ) be represented by:

$$\{(G_{\theta_1}^{t_1}, G_{\theta_1}^{t_2}, ..., G_{\theta_1}^{t}), (G_{\theta_2}^{t_1}, G_{\theta_2}^{t_2}, ..., G_{\theta_2}^{t}), ..., (G_{\theta_p}^{t_1}, G_{\theta_p}^{t_2}, ..., G_{\theta_p}^{t})\}$$

Each of the  $p\mu$  snapshots will generate a set of inequality constraints. Let the inequality constraints for a reader  $\theta$  in episode t in which the graph transforms from  $G_{\theta}^{t}$  to  $G_{\theta}^{t+1}$  be represented by  $q_{\theta}^{t}$ .  $q_{\theta}^{t}$  can be represented in the form of a [0, 1] matrix from the coefficients of the variables in the generated constraints as follows,

$$q_{\theta}^{t} = \begin{cases} \sum w_{e} \ge \alpha_{\theta} & e \in L_{\theta}^{t}, \varphi(e) = (i, j), i \in V_{\theta}^{t}, j \in R_{\theta}^{t} \\ \sum w_{e} < \alpha_{\theta} & e \in L_{\theta}^{t}, \varphi(e) = (i, j), i \in V_{\theta}^{t}, j \in \overline{R_{\theta}^{t}} \\ w_{e} \ge \beta_{\theta} & e \in S_{\theta}^{t} \\ w_{e} < \beta_{\theta} & e \in \overline{S_{\theta}^{t}} \end{cases}$$

The computation of the association strengths of the base concept network can then be represented as a constraint satisfaction problem:  $C = \psi(Q, X)$  where  $\psi$  is an optimization function, Q is the set of constraints obtained from the p examples and X is the set of variables (association weights and thresholds), where

$$Q = [q_{\theta_1}^{t_1} \quad q_{\theta_1}^{t_2} \quad \dots \quad q_{\theta_1}^{t_{\mu}} \quad q_{\theta_2}^{t_1} \quad \dots \quad q_{\theta_p}^{t_1} \quad \dots \quad q_{\theta_p}^{t_{\mu}}]$$

$$X = [e, \alpha_{\theta_i}, \beta_{\theta_j} : e \in E_\Omega, 1 \le i, j \le p]$$

On solving these set of constraints we can get the values of the associations for the base concept network of the text. Using a linear optimization based algorithm<sup>9</sup> for solving the constraints the solution converges to a point in X dimensions. The constraints define the hyper planes which enclose the solution space in which the solution point is located. There can be multiple solution points in a solution space which satisfy all the constraints. The algorithm converges to one such solution point. The coordinates of the point are the values of the model parameters.

# 2.4 Solvability issues: mutual exclusivity

In this section we briefly discuss some of the issues which can limit the model from learning the association strengths.

<sup>&</sup>lt;sup>9</sup>Any linear optimization algorithm can be used such as simplex or an interior point method. Alternative approaches to compute the model parameters are learning by gradient descent using back propagation of errors in a simple artificial neural network [Rumelhart, Hinton, Williams 1986], or other nonlinear or convex optimization techniques. Back propagation of errors using ANNs is slow and does not guarantee convergence to an optimum solution. It has a tendency to settle into local minima or maxima. Boltzmann machines are often used to overcome these problems [Hinton, Sejnowski, Ackley, 1984].



Figure 2.4: The issue of linear separability between individuals.

# 2.4.1 Mutual exclusivity between two individual readers

The problem of mutual exclusivity may arise between two individuals when given the same background knowledge two or more individuals recognize a different set of concepts. This condition and its solution are explained in the following example.

Consider a case of graph transformation from  $G^t = \{V^t, E^t\}$  to  $G^{t+1} = \{V^{t+1}, E^{t+1}\}$ for two different readers  $\theta_1$  and  $\theta_2$  with different thresholds for concept recognition  $\alpha_{\theta_1}$ and  $\alpha_{\theta_2}$  respectively. The constraint representations of the transformations are shown in the Figure 2.4.

Assume that initially at end of episode t-1 both the graphs contain the concepts  $v_1$ and  $v_2$ . In episode t, reader  $\theta_1$  recognizes concept  $v_3$  and does not recognize concept  $v_4$ while the reverse is true for reader  $\theta_2$ . The constraints generated from the above two cases are also shown in the Figure 2.4. These cases offer a case of mutual exclusivity in which the constraints cannot be solved to obtain the values for  $e_1$ ,  $e_2$ ,  $e_3$  and  $e_4$ .

From the inequalities in Figure 2.4a we see that  $e_1 + e_2 > e_3 + e_4$  and from Figure 2.4b we see that  $e_1 + e_2 < e_3 + e_4$ . This contradiction cannot be resolved by varying the values of the thresholds  $\alpha_{\theta_1}$  and  $\alpha_{\theta_2}$ . The constraints are insolvable for any real values

of thresholds or association weights.

The model necessitates that if two individuals have the same background knowledge (as in the example) and also the same node recognition threshold, then they must recognize the same concepts. However this assumption may not conform to reality.

Therefore the solution to this problem is obtained by adding a layer of hidden nodes one corresponding to each reader. By introducing hidden nodes the dimensionality of the problem space is increased in the hope of finding a solution point within an increased dimensional space<sup>10</sup>. The hidden nodes also have semantic significance. A hidden node represents all the previous knowledge possessed by a reader  $\theta$  and is represented as  $v_{\theta}$ . The reasoning behind the inclusion of hidden nodes is this, if two readers possess the same background concept knowledge and if the same two readers are presented with an identical set of concepts and associations for recognition then they both have to recognize exactly the same subset of concepts and associations. If they do not then it conversely means that the background knowledge possessed by the two readers has to be different. This assumption stems from our belief in the innateness of primitive concepts. Since some innate concepts are native to every individual the background knowledge of every individual is different.

The difference in the background knowledge is instantiated by the associations which are made from the single node  $v_{\theta}$  which is unique for every individual  $\theta$ . The associations made from this node are allowed to have negative weights. Consider the above example but this time two hidden nodes  $v_{\theta_1}$  and  $v_{\theta_2}$  are added for the two individual cases (two

<sup>&</sup>lt;sup>10</sup>In machine learning this is called as the kernel trick [Aizerman et al. 1964]. The kernel trick is a way of mapping observations from a general set S into an inner product space V (usually in much higher polynomial dimensions) in the hope that the observations will be linearly separable in V.



Figure 2.5: Solution to mutual exclusivity between two individuals.

readers) respectively in Figure 2.5

From the inequalities presented in Figure 2.5 (a) and (b) it can be seen that by adjusting the values of  $e_5$ ,  $e_6$ ,  $e_7$  and  $e_8$  as desired and since all of these associations can take negative strengths, all the constraints can be satisfied to get the values of the association strengths.

### 2.4.2 Mutual exclusivity due to delayed concept recognition

The mutual exclusivity problem can also manifest itself in a different format in the case of a single individual. The problem occurs when there is a delay in the recognition of a concept and in the time elapsed there is no change in the background knowledge. This creates the discrepancy that although in both episodes the background knowledge is the same, the concept is recognized only in the later episode and not the former<sup>11</sup>.

An example of this condition is discussed as follows. Consider the case of graph transformation from  $G^t$  to  $G^{t+1}$  to  $G^{t+2}$  for a single reader  $\theta$  with thresholds for concept

<sup>&</sup>lt;sup>11</sup>A fundamental assumption is made here that if an individual has all the primitive concepts required to correctly recognize and integrate a new concept, and if that concept is presented to the individual, then the concept will be recognized without exception, otherwise the problem of delayed recognition is encountered.



Figure 2.6: Solution to mutual exclusivity due to delayed concept recognition.

recognition given by  $\alpha_{\theta}$ . The constraints representation of the transformation is shown in the Figure 2.6a. In episode t, the concept  $v_3$  is presented for recognition. Assume that it is not recognized. In the next episode however, at time t+1, the concept  $v_3$ is again presented for recognition and this time the concept is recognized and added to the comprehended graph. How is this possible? This presents a case of delayed recognition of concepts. From the definition of linear separability it is evident that the two equations shown in Figure 2.6 (a) and (b) cannot give rise to a solution space. Therefore we introduce another hidden node  $v_t$  where t is the episode. It represents the episode comprehension factor (or the sentence comprehension factor) and is the same for all readers (because the reader read the same sentence) but varies for every episode. Thus a total of  $\mu$  variables are added to the model each corresponding to an episode. From equation in Figure 2.6c it can be seen that by adjusting the value of variable  $e_4$  the mutual exclusion between the constraints can be resolved.

Mutual exclusivity between two readers and mutual exclusivity for single reader due to delayed recognition are the two known cases which can halt the model from computing the parameters. The SI process is modeled as a linear separability problem between recognized and unrecognized concepts and associations. So, even if a similar problem arises due to some other factor, it can be dealt with in a similar way by increasing the dimensionality of the solution space and thus relaxing the constraints on the solution point. Thus we show that the SI process can be modeled as a threshold phenomenon as given in hypothesis 1 and 2 as stated in Section 2.2.

### 2.5 Implementation and computability

The model is implemented in MATLAB. It implements an interior point algorithm to solve the linear optimization problem. Specifically it implements a large scale linear programming method called LIPSOL [Zhang 1995] which is a variant of Mehrotras predictor-corrector algorithm [Mehrotra 1992], a primal-dual interior point method. In general the computational complexity of current interior point methods is  $O(N^3L)$  where N is the number of variables and L is the size of data (i.e. number of inequalities).

Let the total number of concepts nodes in the concept network be n. Let total number of readers be p and total sentences in the text (episodes) be  $\mu$ . A total of pmore background nodes will be added to the concept network for every reader. Another  $\mu$  nodes will be added one each for every episode. Thus, the concept network will contain a total of  $n + p + \mu$  nodes. Therefore, the number of edges is,

$$\gamma = \frac{(n+p+\mu)(n+p+\mu-1)}{2}$$

Also for every reader the model will contain p node recognition thresholds and association recognition thresholds. Therefore the total number of variables is

$$N = \gamma + 2p$$

Every individual reader generates a set of inequality constraints. The constraints can be divided into those generated by the recognition or non-recognition of a concept and those generated for associations. Therefore, the total number of constraints generated for nconcept nodes by p readers is np and the total constraints generated for the associations are  $\gamma p$ . Therefore the size of the data, i.e., the total number of inequality constraints is,

$$L = p(n + \gamma)$$

The time complexity of the algorithm varies with the size of the data and total number of variables. In the SI model this is directly proportional to the number of readers, the number of sentences in the text and also the number of concepts in the text.

### 2.6 Exercise: Computing a base concept network for a text

An experiment is performed to learn the weights of a BCN of a text by fitting the reader generated ICNs to the model. The reader generated ICNs are shown in the Appendix A. We presented a group of 9 readers with a paragraph of text. The paragraph contained 8 sentences. In the each episode (sentence) the reader is asked to draw a graphical representation of the concepts and association which are recognized. In subsequent episodes the same procedure is repeated. A reader can make new associations between the new concepts recognized and any of the existing concepts in the graph. Each reader was asked to show the time stamped growth of the graph representing generative comprehension. The result of this exercise is a time stamped concept map consisting of concepts and associations acquired over each sentence indicating the dynamic nature of the concept network during comprehension. These graphs provide a window into the concept network construction process inside a text readers mind. Following the procedure detailed in Section 2.3 each episode for each reader is represented by a set of constraints and posed as a constraints satisfaction problem. The solution to the problem outputs the values of the variables, i.e., the association strengths between the concepts and the individual thresholds which explain the comprehension process. Figure 2.7 shows the computed base concept network with the learned weights for the given text.

# 2.7 Conclusion

This Chapter presented the SI model as an elementary process in comprehension and gave it a mathematical framework. The presented model has the following properties:

1. The model incorporates the differences between the skills of readers in the form of

the concept segmentation and integration thresholds

- 2. The model incorporates the difference in the background knowledge possessed by readers which is instantiated by the strengths of the associations from the hidden node for every reader
- 3. The model also incorporates the variability in the quality of the text in every sentence (episode) as instantiated by the strengths of associations from the episode comprehension node.

At this point we would like to explicitly state that this model is not intended to be a predictive model which can explain comprehension for any other individual reader not a part of the group under consideration. Although the model is capable of fitting any new example observed, its initial intent is not to predict the comprehension process of unseen examples but rather to explain the process of the seen examples. In short, it is not a model of concept learning (in the traditional sense of hypothesis induction) but a model of text comprehension.



Figure 2.7: Example of a computed base concept network. The nodes named std1; std2, etc. represent the background knowledge of each of the students. The paragraph of text used and the ICNs from which this BCN is computed are shown in Appendix A.

# CHAPTER 3

## Function of concept sequence

Time is not an empirical concept. For neither co-existence nor succession would be perceived by us, if the representation of time did not exist as a foundation a priori. - Immanuel Kant, Critique of pure reason, 1781-1991, p.29/45.

In this Chapter we will discuss the importance of the time of acquisition of a concept on the networks construction. First an experiment is described and then the SI model is used to analyze the data.

### 3.1 Motivation

It is nearly a universal human experience that the sequence in which people experience events matters in a fundamental way. This is because of the perceived linear structure of time [Elman 1990]. In this chapter we propose that the sequence in which concepts are acquired by a reader determine the way in which concept networks of comprehension are constructed. The following observation is made; the sequence in which concepts are presented in a text (and therefore acquired) is one of the reasons for the difference in the concept networks constructed by the individuals. In this chapter we propose that the said difference can be measured by,

- 1. Average number of concepts and associations recognized by individuals given concepts in different sequences.
- 2. The node weight of the concepts in the BCN obtained using the SI model from Section 2.3.

A text presented in different sequences can produce different concept network constructions in single reader. The same text can produce different concept network constructions in different readers. When in the former the difference is because of the recursively different background knowledge in the same reader, in the latter the difference is because of the difference in the background knowledge of different readers. In a study using the advanced organizer paradigm, [Mannes, Kintsch 1987] demonstrate the importance of organization of background knowledge in recall and inference experiments. It was found that students who are given background knowledge which is organized in the same manner as the text on which the questionnaire is based, fared well on recall questions. Whereas students which answered the same questionnaire based on a randomly organized background knowledge fared well on inference type questions. It is often found that different individuals prefer different text books depending upon the authors top-down, bottom-up, or any other approach of organizing knowledge.

The rest of this Chapter is devoted to validating the observation using the proposed methods by performing reading experiments in a classroom setting.

# 3.2 Experiment

### 3.2.1 Design

In the experiment, we present a set of readers with a paragraph of text as shown in Figure 2.7. It is taken from the standardized text book *Computer Communication Networks* by Peterson and Davies 5th edition. The paragraph contained 8 sentences (episodes). In the first episode the reader is asked to read the first sentence, comprehend and draw a graphical representation of the concepts and associations which are recognized. In subsequent episodes the same procedure is repeated and the reader is allowed to make new associations between the new concepts or some of the existing concepts in the network. Each reader was asked to show the time stepped growth of the network representing generative comprehension. The readers mark the drawn concepts and associations with the episode in which they were drawn. The result of this exercise is a time stamped concept map consisting of concepts and associations acquired over time indicating the dynamic nature of the concept map during comprehension. These graphs provide a window into the graph construction process inside a text readers mind. The experiment was conducted in spring 2009 in an undergraduate class of "Computer Networks". The subjects were divided into four groups of 16, 13, 8, and 8 readers in it respectively. The groups generated a total of 45x8=360 concept networks. To simulate the different sequences of concept acquisition the sentences were rearranged in different orders and were presented to each of the 4 groups.

- 1. Group 1 (G1) the control group were presented the paragraph as it is in the text book. This sequence is the sequence designed by the authors. It is called as the author sequence  $(Seq_{auth})$ .
- 2. Group 2 (G2) is presented with the same sentences rearranged in a random order  $(Seq_{rand})$ .
- 3. Group 3 (G3) is presented with a sequence in which the sentences with fewer concepts are presented first and the number of concepts in each sentence increasing progressively. It is called as the light start sequence  $(Seq_{light})$ .
- 4. Group 4 (G4) is presented with a sequence which is exactly the opposite to that of

	$Seq_{auth}$	$Seq_{rand}$	$Seq_{light}$	$Seq_{heavy}$
$Seq_{auth}$	1	-0.34	0.80	-0.88
$Seq_{rand}$	-0.34	1	-0.31	0.58
$Seq_{light}$	0.80	-0.31	1	-0.69
$Seq_{heavy}$	-0.88	0.58	-0.69	1

Table 3.1: Correlation between presentation episode for concepts between sequences.

G3, i.e. sentences with higher perceived concepts are presented first and number of concepts per sentence progressively decrease. It is called as the heavy start sequence  $(Seq_{heavy})$ .

All the arrangements of the text contained the same concepts. Table 3.1 shows the difference and similarities between the four sequences as measured by the correlation between the presentation times of the concepts. The value is calculated by listing the presentation times for the concepts in all sequences and then taking the correlation between these vectors. Higher correlations indicated that the same concepts were presented at more or less the same time in the two sequences. We observe that  $Seq_{auth}$  and  $Seq_{light}$ are the most similar with a correlation coefficient 0.8 as were  $Seq_{rand}$  and  $Seq_{heavy}$  with a correlation of 0.58. All others were negative.

Assuming that the sequence in which the concepts are presented in the text is responsible the difference in constructions, the following prediction is made; the students in the different groups will recognize different numbers of concepts and associations.

# 3.2.2 Assumptions

Before analyzing the data here are a few assumptions we have to make because of the errors which might arise due to the experimental methodology.

1. Readers strictly read the paragraph in the given sequence.

	n	n(s.d.)	l	l(s.d.)
$Seq_{auth}$	15.06	3.27	15	3.76
$Seq_{rand}$	14.92	2.75	15.08	5.45
$Seq_{light}$	16.83	3.54	14	2.60
$Seq_{heavy}$	14.75	2.36	16.75	4.11
variance	0.93	-	1.23	-

Table 3.2: Average number of concepts recognized (n) and average number of associations made (l) by each of the four groups of readers are their standard deviations.

2. The concept maps drawn truly reflect the concepts and associations recognized.

3. Readers draw all the concepts and associations they recognized.

# 3.2.3 Statistics obtained from individual concept networks

Table 3.2 compiles the average number of concepts (n) and average number of associations (l) recognized at the end of the comprehension by each group. The average value is the ratio of the total number of concepts identified and total number of associations made by subjects in each group to the total number of subjects in that group. It also shows the standard deviation observed within the members of each group and the variance between sequences.

It is observed that all the groups on average recognized almost the same number of concepts and associations as the control group [var(concepts)=0.93 and var(associations)=1.23]. However, there are some minor observations. The concept recognition was highest for light start (16.83) then for authors and randomized and lowest for heavy start (14.75) sequences. Interestingly the association recognition was somewhat inverse- higher for the heavy start and least for the light start sequence (16.75 vs. 14) - the authors sequence and randomized sequence performing somewhere in the middle (15). This indeed is a surprise. Irrespective of the sequence in which concepts are presented, it seems all

	$S_n$	$R^2$	$S_l$	$R^2$
$Seq_{auth}$	+0.044	0.04	+0.15	0.35
$Seq_{rand}$	-0.02	0.61	-0.10	0.70
$Seq_{light}$	-0.002	2E-2	+0.03	0.07
$Seq_{heavy}$	-0.32	0.86	-0.11	0.24

Table 3.3: Moving averages of number of concepts and associations in each of the 8 episodes fitted to a linear regression model.

the groups were able to acquire more or less the same number of average concepts and associations.

However, is it quite possible that the final values for the n and l variables provide a rather simplistic view of the effect of sequence in concept graph growth. Therefore we decided to observe the moving averages of n and l during the growth of the graph. Table 3.3 shows the values of the slope (S) and fitness coefficient  $R^2$  obtained by fitting the data to a linear regression model. n and l are calculated at the end of each episode. The trend is given by a slope (S) (the rate at which n and l increases/decreases over a period of eight episodes) and *fitness coefficient*  $R^2$  gives an estimate of how well the linear trend line fits the data. A negative value of slope implies decreasing values whereas positive slope implies increasing values. A value of  $R^2$  closer to 1 indicates a good fit and 0 indicates a bad fit to the data.

The following observations are made. The  $R^2$  values show that the growth in n and l resembles near linear change for  $Seq_{rand}$  and  $Seq_{heavy}$  compared to  $Seq_{auth}$  and  $Seq_{light}$ . Conversely in sequences  $Seq_{rand}$  and  $Seq_{heavy}$ , the l value is initially higher and steadily decreases over time steps. The trend is  $Seq_{auth}$  resembles more that of  $Seq_{light}$  whereas the trend during  $Seq_{rand}$  resembles closely that of  $Seq_{heavy}$ . This can be explained intuitively.  $Seq_{heavy}$  offers more concepts at the start, thus, with time the number of newly acquired concepts steadily decreases. On the other hand  $Seq_{light}$  offers fewer concepts at the start and the rate of new concept acquisition is much more flat. In a way the slope indicates some form of *learning curve* effect. The same effect is also evident in the number of associations made. The recognition behavior of the readers reading the authors sequence and the random sequence lie somewhere in between the above two groups. A closer inspection tends to indicate that the pattern for the authors sequence is closer to that of  $Seq_{light}$  than  $Seq_{heavy}$ . This result corresponds very well to the similarity correlation coefficients we obtained from the concept presentation sequence (Table 3.1). The random sequence performed closer to  $Seq_{heavy}$  than to  $Seq_{auth}$  and  $Seq_{light}$ . What is also evident is that the authors sequence resulted in a near positive slope (+0.0439) for concept recognition and highly positive slope (0.1518) for association recognition- implying easy learning at the start and gradually high cognitive load as the text progressed. This begs the question; is it possible that the authors sorted the sentences not only based on the number of concepts in them but also based on the number of associations presented in those to create a sequence with best learning experience? The random sequence also showed a decreasing trend in concept and association recognition. It means that if the concepts in a text are presented in a random order, the initial effort expended in terms of making contextual associations is more.

Overall we can conclude that although the final values for n and l remain more or less constant over different sequences (Table 3.1), the growth behavior of n and l varies quite distinctly based on the sequences. Also, we note that  $Seq_{auth}$  exhibits easier start and balanced learning load.

#### 3.2.4 Statistics of base concept network for different sequences

The above analysis summarizes some of the interesting observations on the 360 concept maps drawn by the readers. Although the above analysis gives an interesting perspective, it does not account for the differences in the internal structure. The analysis presented above only gives a partial empirical view of the observed phenomenon. A more detailed analysis can be performed by analyzing the computed base concept networks using the SI model. The values of the model parameters can be used to explain the comprehension process for all readers in each group.

# BCNs for different sequences

The BCN has the following properties. It is a union of all the concepts and associations learned by the readers in their ICNs. Each of the ICNs is a subset of the BCN. We collect ICNs for the readers in the four groups (total 360 ICNs). By fitting the ICNs to a model explained in Section 2.3 we can learn the association strengths of the BCNs for each of the four different sequences. Thus we get four different BCNs with the same set of concept nodes, but the associations between concepts have different strengths. The association strengths which are learnt are reflective of the growth processes of the ICNs for each of the different sequences. By performing an analysis on the four different BCNs each corresponding to a different sequence we can make inferences about the impact of the sequence in the formation of the ICNs.

A parameter called concept weight (or node weight)  $z_c$  of a concept c is used to interpret the results. Concept weight is defined as the ratio of individual node weight (i.e., summation of strengths of all the associations to a particular node) to the cumulative

Table 3.4: Correlation between node weights of each concept in different sequences.

		~		-
	$Seq_{auth}$	$Seq_{rand}$	$Seq_{light}$	$Seq_{heavy}$
$Seq_{auth}$	1.00	-0.61	0.78	-0.55
$Seq_{rand}$	-0.61	1.00	-0.35	0.82
$Seq_{light}$	0.78	-0.35	1.00	-0.23
$Seq_{heavy}$	-0.55	0.2	-0.23	1.00

node strengths (i.e., summation of concept weights of all nodes).

 $z_c = \frac{\text{individual node strength for concept } c}{\text{cumulative node strength for all nodes}}$ 

Individual node strength for concept c,

$$\widehat{z_c} = \sum w_e \text{ where } \varphi(e) = (i, c) \text{ and } i \in \overline{N}, \overline{N} = set \text{ of } c's \text{ neighbors.}$$

Cumulative node strengths for all concepts,

$$Z_{V_{\Omega}} = \sum \hat{z_i} \text{ where } i \in V_{\Omega}$$

Therefore the concept weight is given by,

$$z_c = \frac{\widehat{z_c}}{Z_{V_\Omega}}$$

It gives the importance of a concept in a particular sequence. Higher the concept weight more central is the concept in understanding the  $text^{12}$ .

The first question we ask is how does the concept weight vary in the four BCNs for each concept in a different sequence? Table 3.4 shows the correlations between the normalized concept weights of all concepts for the four different sequences. As expected concept weights are found to be highly correlated in sequence  $Seq_{auth}$  and  $Seq_{light}$  and among sequence  $Seq_{rand}$  and  $Seq_{heavy}$ . This means the concepts which are centrally important in  $Seq_{auth}$  are also centrally important in sequence  $Seq_{light}$ . A similar relationship exists between node weight of concepts in sequence  $Seq_{rand}$  and  $Seq_{heavy}$ . Similar sequence

<sup>&</sup>lt;sup>12</sup>A central concept is a concept which is important for comprehension as indicated by the high number of individuals who recognized it. Since a central concept is also recognized by individuals with high threshold, it follows that a central concept has high node weight.

resulted in similar concept weights.

### Impact of Sequence on the Centrality of Concepts

In many studies it has been observed that age of acquisition of a concept plays a role in higher connectivity of a node [Gilhooly, Loogie 1980; Morrison et al. 1997; Hills et al. 2008a, 2009a; Ellis, Ralph 2000]. Age of acquisition is a commonly used parameter to indicate the time (i.e., episode) at which a particular concept or association was presented and acquired in a concept network. The concept weight signifies the connectivity strength of a concept in the network. Figure 3.1 shows the concept weights against the episode (time) in which they were recognized or acquired and a linear trend line for each of the sequences. All four sequences show a distinct negative slope (reasonable linear fit with  $R^2=0.5$  to 0.8). It can be observed from the plot that earlier a concept is recognized, the higher is its concept weight. The centrality of the concept is dependent upon the age of its acquisition. Therefore, this study reinforces the finding that concept sequence plays an important role in forming centrality of a concept in the final concept network. Figure 3.2 shows the same set of data plotted against the concept ids.  $Seq_{auth}$  and  $Seq_{light}$  have negative slope while  $Seq_{rand}$  and  $Seq_{heavy}$  have a positive slope. This implies that the concept sequence in itself biases the centrality of the concepts. Concepts which tend to be central in a particular sequence may or may not be central in another given sequence. A concept which is presented early in one sequence may be central in that sequence, however, may not be central in another sequence in which it is not presented relatively early in time.

Table 3.5 shows the interpretation of Figure 3.1 in terms of correlation values (r)



Figure 3.1: Age of acquisition against concept weight for four different sequences.



Figure 3.2: Concept id against concept weight in four different sequences.

Table 3.5: Correlation values (r) between concept weights and the time (age) of their acquisition in different sequences.

Sequence	r
$Seq_{auth}$	-0.7
$Seq_{rand}$	-0.75
$Seq_{light}$	-0.9
$Seq_{heavy}$	-0.85

between concept weights and the episode (age) of their acquisition in different sequences. It can be seen that for all the sequences, node weight of a concept is highly negatively correlated (-0.7 to -0.9) with the episode in which it is acquired just as seen in Figure 3.1. This means the concepts which are acquired earlier have a higher node weight than those acquired later. Concepts which are acquired earlier thus seem to be assuming more central role compared to those which are acquired later. Hence we conclude that concepts which are presented earlier are therefore more likely to have stronger connections with newly acquired concepts and have a greater central role in the concept network defined in terms of the node weight.

# Stability of Node Weights within a sequence

Another interesting growth characteristic which can be tracked is the evolution of the concept weights as the comprehension progresses episode-by-episode. We calculate the concept weights for each concept at the end of every episode. We then estimate how it varies over the learning sequence. This is done by computing the BCN at the end of every episode. It may be possible that a particular concept may not be recognized by any of the readers in a particular episode. Indeed this is the case in the earlier episodes when not many concepts are presented. The computed BCN will contain the concept and the association strengths to that concept will reflect its apparent non recognition.



Figure 3.3: Age of acquisition against variance in concept weight within a sequence, plotted for four different sequences

Figure 3.3 plots the variance in the concept weights over 8 episodes for a concept against the time (age) at which that concept is acquired for each of the four sequences. The trend confidence levels are lower ( $R^2$ =0.06-0.25) as well as the slope of the trend line. Figure 3.4 plots the same data however against the concept id with confidence ( $R^2$ =0.03-0.5). Again it can be seen that the earlier the concept is presented in the sequence, the lower is the variance. If a concept has a high concept weight variance, it means that at each time step the concept weight is varying a lot, effectively changing the concepts centrality. Apparently, the earlier a concept is presented, the more central role it takes in the concept network.



Figure 3.4: Concept id against variance in the concept weight within a sequence, plotted for four different sequences

# Stability in between sequences (intrinsic centrality of concepts)

In the previous analysis we observed that the concepts which were presented earlier have node weights which are higher than those presented later. In this section we ask the question if some concepts inherently have higher node weights irrespective of the sequence they are presented in. A part of this question is answered in the node weight stability analysis which says that the concepts acquired earlier seem to be more stable. But is this relationship also reciprocal; meaning, are some concepts with intrinsic higher stability acquired earlier because in fact they are more stable?

Some researchers have [Hills et al. 2009a, 2009b] provided experimental evidence that some concepts may be acquired earlier because they have higher concept weights instead of the other way round. We also make a similar observation. The previous analysis seems to indicate that the centrality of a concept in a network is heavily dependent on the presentation sequence and the age of acquisition. However the varying slopes of Figure 3.3 and Figure 3.4 seem to indicate that the reason might be more involved. If a concept does have any intrinsic centrality (i.e., is more stable) then it means that the concept weight should remain proportionally the same in all the four different sequences.

To gain some insight into the issue, Figure 3.5 plots the 16 concepts (x-axis) and the observed variance of concept weight between the four sequences (y-axis). The x-axis sorts the concepts from lowest to highest variance. Interestingly, it can be observed that the concept weights of few of the concepts stay quite stable irrespective of the sequence in which these were presented. It can be seen from the plots that concepts like "host", "network class", "network", etc. have low node weight variance. Do their stable weights indicate that these are more intrinsically central than concepts like "file", "distributed system" or "domain name servers?" To further explore this issue in Figure 3.5 we plot the average node weights (averaged over all the sequences). The figure plots concepts in order of nodes with lowest to highest variance (x-axis). Interestingly one can see a substantial overlap of the concept sets with lowest variance and relatively high node weight across the sequences. The experiment seems to provide evidence that there is a possibility that some concepts are *intrinsically central* irrespective of the presentation sequence. Naturally there exists the other possibility that some concepts are *intrinsically peripheral* possibly signified by high variance and high average concept weight value. More importantly, the analysis done above provides a means to compute intrinsic central/ peripheral value of concepts in learning.



Figure 3.5: Plot of average concept weight of a concept in between four different sequences against variance in concept weight in between four different sequences.

# 3.3 Conclusion

The observation that different sequences of concept acquisition lead to divergent constructions of the concept network and as a result a different comprehension state in individuals has been proved to be valid. The SI model is used as a tool to validate the observation. It is seen that the concepts which are acquired earlier consistently display a higher centrality in terms of their concept weights. Also observed is that these early acquired concepts seem to be more *stable* in terms of variance of the concept weight during comprehension. This is another reason why comprehension is anchored around the earlier acquired concepts.
## CHAPTER 4

# Structure of a grown concept network

Knowing reality means constructing systems of transformations that correspond, more or less adequately, to reality. - Jean Piaget, Genetic epistemology lecture, 1968.

The SI model is in essence a growth model. It naturally inspires a growth algorithm which is used to simulate how concept networks would grow in certain conditions controlled by the model parameters. The structural properties of the simulated networks like size, clustering, reachability and connectedness, are useful in studying the nature of comprehension.

The thriving research in the complex network research community has inspired many cognitive scientists to study the structural properties of concept networks [Steyvers, Tenenbaum 2001, 2005; Hills et al. 2008a, 2008b, 2009a, 2009b]. For instance, it has been observed that these concept networks are small worlds characterized by short average path length and high clustering. These networks are sparse with low average degree of nodes. The degree distribution has been observed to follow a power law or even exponential or Gaussian distributions. However, all of these studies examine the concept network in the context of acquisition of the lexicon in an individual. In this context, we employ our model to present a completely novel method of concept network growth by segmentation and integration (SI) during text comprehension. This process of network growth is distinctly different than the process of lexicon acquisition.

#### 4.1 Algorithm for the SI model of growth

The SI growth algorithm is as follows;

- 1. Let the number of nodes and links in a network at the end of an episode t be n and l, respectively. Initially the network is assumed to contain a single node with no edges. Therefore, at time  $t_0$ , n=1 and l=0.
- 2. A new node is offered for recognition to be added to the network in each episode. This is the step of segmentation. If recognized this node can generate a maximum of n new edges, one corresponding to each existing node. The node is recognized if the summation of the link strengths of all the edges to this new node is greater than the node recognition threshold  $\alpha$ . If the node is recognized, then the node is added to the graph. If  $(w_1, w_2, ..., w_n)$  is the vector of weights for the possible edges. Therefore,

$$y = \begin{cases} 1 : \sum_{i=1}^{n} w_i \ge \alpha \\ 0 : \sum_{i=1}^{n} w_i < \alpha \end{cases}$$

An output of y=1 implies that the node is recognized and will be added to the graph in that time step. The node recognition threshold is calculated as follows,

$$\alpha = \frac{n(1-\beta)\omega}{\vartheta}$$

where  $\beta$ =association threshold,  $\omega$ =average weight of an association and  $\vartheta$ =node recognition rate. The weights for the edges are picked uniformly at random<sup>13</sup>. The

<sup>&</sup>lt;sup>13</sup>This may not be a realistic assumption. There is at least some evidence that association weight distribution in concept networks constructed from LSA [Landauer et al. 1988] may follow a normal distribution [Steyvers, Tenenbaum 2005].

value of  $\vartheta$  is in the interval [0, 1].  $\vartheta$  closer to 0 indicates a low recognition rate while closer to 1 indicates high recognition rate. Controlling  $\vartheta$  allows us to control the number of concepts recognized. If the node is added then the total number of nodes becomes n+1. If the node is not recognized then it is not added to the graph.

- 3. Variable  $\beta$  is the association threshold. It is the percentage of edges which are not recognized. By applying  $\beta$  the number of recognized edges is controlled. The recognized concept is associated with a subset of existing concepts whose association strength values are greater than or equal to  $\beta$ . This is the step of semantic integration. If  $\beta$  is set closer to 1 then few edges will be recognized else if  $\beta$  is closer to 0 more edges will be recognized. Number of recognized edges =  $n(1 - \beta)$ .
- 4. Repeat steps 2 and 3 till desired number of nodes are added to the graph or until limited number of iterations are reached.
- 4.2 Important properties of the growth algorithm
  - At each time step, the value of changes. is a function of the size of the graph in the previous time step and as the size increases, increases too. This implies that as the size of the network increases, so does the threshold for recognition of new concepts<sup>14</sup>.
  - 2. Node recognition rate allows us to simulate the final states of comprehension for different individuals. A low node recognition rate implies fewer concepts will be recognized. A final state which has fewer concepts is suggestive of a state which

<sup>&</sup>lt;sup>14</sup>This property of the algorithm, although not a part of the SI model, is in agreement with constructivist philosophy of meaning construction. It states that as more and more knowledge is assimilated, the resistance to accommodation of new knowledge keeps on increasing.

is far from performing deep comprehension. A high node recognition rate would mean more concepts are recognized and therefore would define the final state which is closer to satisfying the necessary condition for performing the deep comprehension. Thus, varying the node recognition rate simulates a range of final states for different readers. Consequently it is indicative of the level of comprehension. A node recognition rate closer to 0 implies low comprehension whereas once closer to 1 would indicate deeper comprehension. In the following sections we will discuss structural properties of the network over a range of node recognition rates. Thus, we can have a spectrum of structural properties which can indicate a wide range of levels of comprehension.

3. For the purpose of this simulation we consider the strengths of associations from the background node and the episode comprehension node to all other nodes equal to 0.

Firstly, a single run of the algorithm is meant to simulate the growth of the concept network for a single individual. Therefore, the possibility of mutual exclusivity as described in Section 2.4 in nonexistent. The most important function of the background node is to resolve this mutually exclusive condition. Since this condition will not arise, the background node is not considered in the algorithm. This is done by assuming that the value of the associations from the background node to every other node is 0.

Secondly, since the simulation is individual specific, the condition of mutual exclusivity due to delay recognition may arise. However, in the simulation it is assumed

n	$\beta$	$n_{rec}$	$l_{rec}$	z	$z_w$	L	$L_w$	D	$D_w$	$CC_1$	$CC_2$
100	0.15	89	3338	75.01	43.31	1.14	0.38	2	0.64	0.85	0.25
	0.5	94	2202	46.85	35.02	1.49	0.93	2	1.40	0.50	0.15
	0.85	75	439	11.70	10.79	1.97	1.77	3	2.74	0.15	0.03
200	0.15	187	14806	158.35	91.34	1.14	0.36	2	0.54	0.85	0.26
	0.5	199	9861	99.10	74.19	1.49	0.91	2	1.29	0.49	0.16
	0.85	165	2071	25.10	23.22	1.86	1.67	3	2.68	0.14	0.03
500	0.15	487	100644	413.32	237.47	1.14	0.33	2	0.49	0.85	0.27
	0.5	494	60841	246.31	184.59	1.50	0.90	2	1.22	0.49	0.16
	0.85	480	17287	72.029	66.63	1.84	1.63	3	2.59	0.14	0.04
The number of presented nodes varies from $n = [100; 200; 500]$ . The node recognition											

Table 4.1: Statistics for simulated concept networks

The number of presented nodes varies from n = [100; 200; 500]. The node recognition rate is kept constant at  $\vartheta = 1$ .  $\beta$  is the association threshold (or the percentage of edges not recognized),  $n_{rec}$  is the number of recognized nodes,  $l_{rec}$  is the number of recognized edges, z is mean degree,  $z_w$  is the mean weighted degree, L is the average shortest path length,  $L_w$  is the average shortest weighted path length, Dis the diameter,  $D_w$  is the mean weighted diameter,  $CC_1$  and  $CC_2$  are clustering coefficients, refer Section 4.3.3.

that a particular node is not offered repeatedly for recognition. Once the node is presented for recognition it will not be offered again either if it is recognized or not recognized. Due to this assumption, the possibility of mutual exclusive condition due to delay recognition is non-existent. Therefore, there is no need to add the episode comprehension factor node since the "quality" of presentation that the node represents never arises. This is done by assuming that the association weight value from this node to every other node is 0.

## 4.3 Structural properties

Table 4.1 shows the statistics obtained from simulated concept networks for n = [100, 200, 500],  $\beta = [0.15, 0.5, 0.85]$ ,  $\omega = 0.5$  and  $\vartheta = 1$ . A node recognition rate of 1 implies that node recognition is dependent upon only the number of nodes in the network at the time and the strength of association between the new node and the existing nodes.



Figure 4.1: Analysis of size of simulated concept networks (a) Number of concepts recognized  $(n_{rec})$  as a function of node recognition rare  $(\vartheta)$  (b) number of edges recognized  $(l_{rec})$  as a function of node recognition rate  $(\vartheta)$ . The size of the graph is constant at n = 100.

To perform a deeper analysis we vary  $\vartheta = [0.1 : 0.1 : 1]$  for n = 100 and observe the following properties.

#### 4.3.1 Size of the network

The size of the network is measure by the number of recognized nodes (n) and edges (l). Figure 4.1 (a) and (b) show the plot of  $n_{rec}$  and  $l_{rec}$  for different node recognition rates  $(\vartheta)$ . It is seen that as  $\vartheta$  increases so do  $n_{rec}$  and  $l_{rec}$ . This is very self-evident because as  $\vartheta$  increases, the node recognition threshold  $(\alpha)$  decreases and therefore more nodes are recognized. As this happens obviously more edges are also recognized. It is also seen that  $n_{rec}$  and  $l_{rec}$  start increasing the last for  $\beta=0.15$ . This is simply because  $l_{rec}$  depends upon the nodes existing in the network  $(n_{rec})$  and  $n_{rec}$  in turn is inversely proportional to  $\beta$ . For lower values of  $\beta$  the recognition threshold  $\alpha$  is high, which limits the number of recognized nodes  $n_{rec}$ . As slowly more and more nodes start getting recognized the values of  $n_{rec}$  slowly begin to peak. This peaking happens a little sooner for



Figure 4.2: Analysis of connectivity/sparsity of simulated concept networks (a) Average degree (z) as a function of node recognition rate ( $\vartheta$ ) (b) average weighted degree ( $z_w$ ) as a function of node recognition rate ( $\vartheta$ ). The size of the graph is constant at n = 100.

 $\beta=0.5$  and  $\beta=0.85$ . This implies that higher values of  $n_{rec}$  and  $l_{rec}$  are indicative of deeper comprehension.

# 4.3.2 Connectivity/sparsity

The connectivity is measured by average degree (z) and average weighted degree  $(z_w)$ . The average degree is the fraction of summation of degrees of all nodes and the number of nodes. Whereas the average weighted degree is the fraction of the summation of association strengths of the associations to a node, and the number of nodes. Average degree is indicative of the sparsity of the graph or conversely the average connectivity of a node in the graph. The greater the average degree greater is the connectivity of a node. Average weighted degree is indicative of the average degree is indicative of the nodes in the graph. Figure 4.2 shows the plot of average degree (z) and average weighted degree  $(z_w)$  against increasing node recognition rate  $(\vartheta)$  for three different values of  $\beta$ . Quite expectedly as the recognition rate increases, so do the number of recognized nodes and therefore so do z and  $z_w$ . Thus, high average node connectivity is indicative of deeper

comprehension.

It can be seen from the figures that for a low association threshold of  $\beta$ =0.15 there is a sudden increase in the average degree around  $\vartheta$ =0.6. For high values of  $\beta$ , the sudden rise in not observed. This is because for lower node recognition rates there are not many nodes in the network to connect to. However around 0.4 the nodes start acquiring degrees and around 0.6 and above there are enough nodes in the network to connect to. So for low values of  $\beta$  these nodes quickly get connected to each other suddenly increasing the average degree. From the Figure 4.2 it can be seen that for an association threshold of  $\beta$ =0.85 the average connectivity (z) comes to around 15% of the graph. This means that a node is connected to about 15% of the rest of the nodes. In other studies [Steyvers, Tenenbaum 2005] it is observed that the concept networks are very sparse. For example, in the word association network, a word is connected on an average to 22 nodes out of the possible 5018 nodes (0.44%). The percentage connectivity of a nodes for the other networks is even less. Thus, according to our model of growth, networks of comprehension are unlike other semantic networks of words.

## 4.3.3 Neighborhood clustering

Neighborhood clustering offers a way to measure the relatedness of nodes and their neighbors. One of the definitions of the clustering coefficient is as follows [Barrat, Weigt 2000].

$$CC_1 = \frac{3 * number of triangles in the graph}{number of connected triples in the graph}$$

This definition measures the fraction of triples (paths of length two) which form a triangle in the graph. A value closer to 1 indicates a highly clustered graph. Figure 4.3(a) shows



Figure 4.3: Analysis of clustering coefficients of the simulated concept networks (a) Clustering coefficient 1 ( $CC_1$ ) [Barrat, Weigt 2000] against node recognition rate ( $\vartheta$ ). (b) Clustering coefficient 2 ( $CC_2$ ) [Watts, Strogatz 1998] against node recognition rate ( $\vartheta$ ). The size of the graph is constant at n = 100.

the plot of  $CC_1$  against node recognition rate  $(\vartheta)$ . It can be seen that the lowest value of  $CC_1$  for  $\beta=0.15$  is around 0.8 which results in a highly clustered graph but with much higher average degree of nodes as compared to say Rogets thesaurus [Steyvers, Tenenbaum 2005].

Another definition of the clustering coefficient due to Watts and Strogatz (1998) measures the local clustering at a particular node and then averages these values over all the nodes to get the clustering coefficient of the whole graph. It is defined as,

$$CC_2 = \frac{1}{n} \sum_{i} \frac{number \ of \ triangles \ connected \ to \ node \ i}{number \ of \ triples \ centered \ at \ node \ i}$$

This definition measures the means of the ratios of the triangles to triples unlike the first definition which measures the ratio of the means of triangle to triples. This puts extra emphasis on individual nodes so that low degree nodes have considerable importance in the calculation. Figure 4.3(b) shows a plot of  $CC_2$  versus the node recognition rate.

Barrat and Weights definition of clustering  $(CC_1)$  tries to measure what percent of connected triples form triangles. This is a good way to measure the global clustering among the nodes in a network. From Figure 4.3(a) as the node recognition rate increases more and more nodes are added to the network, but the number of triangles with respect to the triples does not increase dramatically which is why the value of  $CC_1$  decreases. Initially when the node recognition rate is low, only very few nodes are even recognized and most of them form triangles with each other. Therefore,  $CC_1$  starts with a maximum value of 1 because there are as many triangles as there are nodes.

The behavior of  $CC_2$  is different than  $CC_1$  as seen in Figure 4.3(b). As the node recognition rate increases, the number of triples increases dramatically in proportion to the number of triangles. However by the definition of  $CC_2$  the formation of triangles is weighted more at each individual node and since  $CC_2$  considers only triples centered at a node, the effect is reduced as the probability of triangle existing at a centered triple is higher. That is the reason  $CC_2$  effectively calculates local clustering. Consequently  $CC_2$ increases with the formation of triangles even with the dramatic increase in the number of triples. As seen from Figure 4.3(b)  $CC_2$  starts from a value of 0 when no triangles exist and steadily rise to a maximum of 0.25 for  $\beta=0.15$ .

Both measures of clustering coefficients  $CC_1$  and  $CC_2$  indicate a high degree of clustering in the graph. The model results in networks which have higher local clustering than global clustering and is able to achieve clustering which is far better than random graphs. The parameters of the model can also be tuned to achieve the desired level of network clustering. As seen from the clustering analysis of real semantic networks, there is no single consistent value of clustering observed. It ranges from as low as 0.0265



Figure 4.4: Analysis of reachability of simulated concept networks (a) Average shortest path length (L) as a function of node recognition rate  $(\vartheta)$  (b) Diameter (D) as a function of node recognition rate  $(\vartheta)$ . The size of the graph is constant at n = 100.

for WordNet to as high as 0.875 for Roget's thesaurus and differing values in between [Steyvers, Tenenbaum, 2005]. Thus, low global clustering and high local clustering are indicative of deeper comprehension.

## 4.3.4 Reachability

Reachability is the property of a network which measures the efficiency of traversal from on node to another in a network. It is measure by the average shortest path length (L) or the diameter of the network (D). The reachability is poor if both these values are high.

It is seen that even with the increase in  $\beta$  the values of L and D do not change drastically. For example, the values of L for  $\beta=0.15$  and  $\beta=0.85$  are 1.149 and 1.84 respectively and the values of D are 2 and 3, respectively. It can be seen that even if the values of n increase, the values of L and D dont grow proportionally. As more nodes are recognized the values of L and D reach a plateau never going beyond 2 and 3, respectively.



Figure 4.5: Degree distribution and cumulative degree distribution. Values of k are generated over 10000 runs. The size of the graph is constant at n = 200.

Figure 4.4a shows the plot of L versus node recognition rate ( $\vartheta$ ) for different values of  $\beta$ . As  $\vartheta$  increases, more nodes are recognized thereby increasing the average shortest path length. As expected L is lowest for  $\beta=0.15$  never reaching beyond 1.2, whereas L is highest for  $\beta=0.85$  when it reaches to almost 2. This is because as  $\beta$  increases, fewer of edges are recognized thus increasing the values of L. Figure 4.4b shows the plot of the diameter versus the node recognition rate. It follows exactly as expected like the plot for average shortest path length.

Thus, even though it is seen that reachability of the network increases with comprehension, after a certain point it reaches a plateau and stops growing. Increasing comprehension (as indicated by node recognition rate) does not mean lower reachability (as indicated by increasing L and D). A good reachability is important for deeper comprehension.

#### 4.3.5 Degree distribution

The distribution function of the fraction of vertices which have degrees (k) ranging from minimum (k = 0) to maximum (k = n-1) is called the degree distribution function. Concept networks of comprehension generated by following our model show a modified Gaussian network as shown in Figure 4.5(a). There is no scale freeness in these networks, meaning a small number of nodes with high degrees and a large number of nodes with low degrees do not exist. Most of the nodes seem to be acquiring degrees around the values of (z), i.e., the average degree. From Table 4.1 it is seen that for  $\beta=0.15$ , 0.5 and 0.85 the average degree is 158.35, 99.1 and 25.1, respectively.

The degree distribution curve can be divided into 3 parts. The first part is for k = 0when there is an accumulation of zero degree nodes at the beginning for all the nonrecognized nodes. The second part is for  $1 \le k < \alpha$  when the edges could be recognized but are not because the nodes are not recognized. The third part of the curve is for  $\alpha \le k \le n$  when the nodes are recognized with accumulating degrees greater than the association threshold. The function can be generalized to a binomial function which for high values of n can be represented as a Gaussian. The detailed derivation for the degree distribution is given in the Appendix B. The degree distribution indicates the spread of connectivity among nodes. A normal distribution, like the one observed, indicates that a majority of the nodes have degree near the average value and few new in the tails. This implies that most of the nodes are well connected. As seen before, connectivity is indicative of good comprehension.

#### 4.4 Implications of structural properties

In this section we discuss some resulting implications of the structural properties as discussed above.

## 4.4.1 Small worlds

Small worlds are characterized by high reachability (low average short path lengths) and high neighborhood clustering (high clustering coefficients). In terms of graph theory, this phenomenon is observed when the average distance between two nodes grows logarithmically with the size of network. Small worlds have been observed in many real world networks like the power grid of United States, collaboration of film actors, and the neural network of the worm C. Elegans [Watts, Strogatz 1998]. Small world have also been shown to exist in common semantic networks like word association network, Rogets thesaurus and WordNet [Steyvers, Tenenbaum 2005]. The largest observed shortest path length for a semantic network is 10.56 for WordNet with  $n \approx 122,000$  and the most clustered is Rogets thesaurus with  $CC_1=0.87$ . The high local clustering and short average path length are indicative of a network with high reachability in low costs.

The concept networks of comprehension generated from the algorithm also shows low average shortest path lengths comparable to the other small world semantic networks. However, the clustering coefficient observed in these simulated networks varies with the network density. Common semantic networks are highly clustered and at the same time very sparse [Steyvers, Tenenbaum 2005]. This property is not found in simulated networks from our model. In our model if the density of the network is decreased, the clustering drops too. The simulated networks are small worlds but do not show the existence of hubs or cliques. Small world in concept networks can be backed by the logical intuition of low cost reachability which is required in information processing cognitive tasks of recall and retrieval. Small worlds are a characteristic structure in concept networks which is indicative of deep comprehension.

## 4.4.2 Proportional connectivity

The degree distribution gives the proportional connectivity of a network. For instance, in a scale free network (power law degree distribution) few nodes are well connected whereas a large number of nodes are not well connected. Therefore, the proportional connectivity is very low. In contrast, a network with normal degree distribution has high proportional connectivity. Power laws have been observed to naturally arise from a number of generative systems like word-frequency distributions, income distributions, cell growth network distributions, distributions of citation networks, WWW, etc. [Albert, Barabasi 2002]. When plotted on a log-log scale these networks show a linear decay in the probability of existence of hubs against normal nodes. These networks are called scale free because the behavior of degree distribution changes logarithmically with changing size of the networks. The generative nature of scale free networks has been a subject of much inquiry and debate in the recent past [Mitzenmacher 2004].

In common semantic networks like word association networks, Rogets thesaurus and WordNet and observed that degree distribution does follow the power law [Steyvers, Tenenbaum 2005]. In a comparative analysis of a concept network of words generated by Latent Semantic Analysis (LSA) [Landauer et al. 1998], the resultant network failed to show a scale free distribution of degrees. Hills et al. (2009) show that developmental networks of lexicon in 16 to 30 month old babies do not necessarily show power laws. Networks which grow under constraints do not exhibit scale free degree distribution but more of an exponential or Gaussian degree distribution [Amaral et al. 2000]. The utility of a scale free distribution is a matter of open debate. The simulated concept networks of comprehension from our models do not show a scale free distribution.

The property related the most to degree distribution is the resilience of the network to node removal. If vertices from the network are removed at a constant rate, then average shortest path length will increase until node pairs will get disconnected and communication will become impossible. The forgetting of a concept from semantic memory can be modeled as node removal. It has been observed that scale free networks are very resilient to random node removal [Albert et al. 1999] but they are extremely susceptible to attacks target towards the hubs [Callaway 2000]. Random graphs with differing degree distributions such as Poisson-ion, Gaussian and exponential, are resilient against random or targeted attacks [Callaway 2000; Dunne et al. 2002a, 2000b]. High connectivity demonstrated through the degree distribution of the concept networks is indicative of deep comprehension.

# 4.4.3 Other implications

One of the reasons to study the nature of the concept network during comprehension is to understand the structural properties of the networks in neuro-degenerative diseases like Alzheimers dementia (AD). It has been observed that in clinical populations many of the structural properties are lost which disables the patients from performing satisfactorily on comprehension tasks. The organization of semantic or concept networks plays an important role in predicting the performance of an individual in cognitive tasks of comprehension. There is evidence that the deterioration of the concept network structure leads to this degraded performance on cognitive tasks [Chan et al. 1993a, 1993b, 1997, 1998; Martin 1992; Nebes 1989, 1992; Hodges JR et al. 1992; Medler 1998]. Degradation can be in the form of weakening and loss of connections [Chan 1997; Au et al., 2003], reduced concepts and associations [Salmon et al., 1999; Binetti et.al. 1995], reduced connectivity with respect to average degree [Chan et al. 1995, 1997], fewer common links between concepts (sparsity) and atypical association strength values [Chan et al., 1995].

The structural properties presented in this Chapter will be helpful in giving a concrete quantitative measure to the degradation of concept networks in clinical patients like those with AD. By measuring these properties for a concept network of a patient, judgments can be made about how well it scores in the spectrum of comprehension from superficial to deep. Other applications include gauging the proficiency of individuals who require a highly reliable concept network for effective cognitive performance like pilots, sportsmen, etc. in terms of their network properties.

## 4.5 Conclusion

The SI model naturally motivates an algorithm for concept network growth. The model can be used to simulate a range of network growth which depict low to high comprehension. The simulated concept networks have peculiar structural properties such as they are small worlds which make them highly reachable and are highly connected with a varied scale of degree distributions which make then highly resilient. Both of these qualities are indicative of concept networks which represent deep comprehension.

## CHAPTER 5

## Connectedness of a concept network

Reality cannot be found except in one single source, because of the interconnection of all things with one another. - Gottfried Leibniz, Monadology, 1670.

In the previous Chapter we saw that reachability and connectivity of a network are indicative of the level of comprehension. Another way to look at reachability and connectivity is by analyzing the component<sup>15</sup> structure of the network. A network divided into many disconnected components has low reachability and connectivity unlike one with a single giant component. The component structure of the concept network is also indicative of the level of comprehension.

In this Chapter we present an interesting discussion on the evolution of the network structure in particular the phase transition in the component structure. The growth algorithm described in Section 4.1 is used to simulate network growth and the structural evolution is closely observed and analyzed. It is seen that concept networks start off with a few components which grow independently but soon converge into a single giant component. This result provides a glimpse into how concepts in a concept network get interconnected into a connected network at the end of successful comprehension.

<sup>&</sup>lt;sup>15</sup>A component in a graph is a group of nodes which are connected. A graph may contain many components which may be disconnected with each other or lightly connected to each other only through a few nodes. The component which contains most of the nodes in a graph is called as the giant component.

## 5.1 Motivation

The statement that all knowledge is connected is a truism. For example, consider physical knowledge networks of the WWW, Wikipedia page network, citation network, and word co-occurrence network and in a similar vein naturally occurring knowledge networks, i.e., the concept network of concepts (or things). All of these are almost fully connected. The connectedness of all knowledge has been professed by philosophies from all parts of the world and reiterated at all times in the history of mankind.

One of the interesting questions here is; how does the connectedness of these concept networks come into existence? Recently many researchers have collected data and analyzed the evolution of the connectedness of many man made and naturally occurring complex networks [Amaral 2000; Dorogovtsev et al. 2002; Leone et al. 2002; Aleksiejuk et al. 2002; Bianconi 2002]. The main results suggest that these networks are generally composed of many individual components which are densely connected within themselves but have sparse connections between the components. However, these sparse connections between the components allow for the network to be completely connected in which any node is reachable from any other node, meaning that there is also present a single giant component. In a structural analysis of some of the common semantic networks in literature such as the word association network, Rogets thesaurus and WordNet it was found that almost 96% of the nodes in the networks are reachable through a path [Steyvers, Tenenbaum 2005]. We perform a similar study for the evolving concept networks of comprehension based on the hypothesis that concept networks too show the emergence of a giant component like the other common semantic networks. Studies [Chan et. al. 1993a, 1993b; Just et al. 2004] have indicated that some form of disjointedness of semantic memory space might be related to diseases like Autism. This indicates a potential practical significance of component structure analysis for concept networks.

#### 5.2 Component analysis of simulated concept networks

A network may be composed of components which are strongly connected within them but may not have any connections in between the components rendering them totally isolated from each other. These can be thought of as pockets of knowledge with are totally disconnected from one another. However, it is the very character of knowledge that a concept may derive its meaning from the concepts it connects too. Therefore a concept network divided into components is not desirable.

The phase transition point is defined as the critical point at which most of the nodes in a network belong to a single connected component called the giant component. For the purpose of the simulation we start with a graph which initially contains one or more disconnected nodes. In the subsequent time steps nodes are offered for recognition and if they are recognized then they are added to the graph. Once the node is recognized, it can potentially make connections to all of the existing nodes in the network. However, not all of these connections are made. Only those connections which have strength greater than the association threshold are recognized. Thus at every time step either no new node is added to the network, or a new node with none or few connections to the existing nodes is added. If the node is recognized but none of its associations are recognized then the node is not a part of the giant component and can possibly start a separate component. Figure 5.1 gives a visualization of the formation of the giant component in a network.



Figure 5.1: Evolution of a network structure and the phenomenon of phase transition. The lower figure shows the fraction of the nodes in the giant component against time. The upper three figures show the time step at which snapshots of the evolving networks are taken at t = 5, 20, 40.  $\beta = 0.5$  is the association threshold. Red nodes are a part of the giant component. Phase transition occurs at t = 20. At t = 40 most of the nodes in the network are in the giant component with very few disconnected components.

The initial network at time  $t_0$  is assumed to have a certain number of nodes such that  $n_0 > 1$ , but no edges between the nodes  $l_0 = 0$ . Since these nodes are not connected to each other they are different components. Thus, initial graph has  $n_0$  disconnected components. As new nodes are recognized and added to the network, connections start forming with the new nodes and existing nodes forming components of sizes bigger than 1. Adding more nodes increases the size of these components proportionally, until a giant component emerges which include most of the nodes in the network. The critical point at which this happens is called the phase transition point. The node recognition rate is kept at a maximum of  $\vartheta = 1$ . A node recognition rate of 1 implies that node recognition is dependent upon only the number of nodes and the strength of associations. Also assuming  $\vartheta = 1$  will mean that the algorithm will generate a concept network which in its final state will indicate the highest level of comprehension possible.

In this section we will observe the occurrence of the phase transition in terms of these three measures.

#### 5.2.1 Number of components

Initially, the network consists of  $n_0 > 1$  number of disconnected nodes which are the disconnected components. As more nodes are added the number of disconnected components begins to decrease. This happens because as more nodes are added more edges are also added between the existing nodes and the new node. If an edge is added between two or more disconnected component then those disconnected components become a single component. Obviously as more edges are added more disconnected components start getting connected to each other.



Figure 5.2: Number of components  $(C_n)$  versus time for a graph of 200 nodes. In figure (a)  $n_0=15\%$  of the nodes are initially assumed to be included in the graph as disconnected components, in (b)  $n_0=50\%$  and in (c)  $n_0=85\%$ . In each figure,  $C_n$  versus t is plotted for three values of association threshold  $\beta=[0.15, 0.5, 0.85]$ .

Figure 5.2 shows the plot of the number of components in a network  $(C_n)$  against time t. It is observed that for (a) the number of components increases where as for the rest of the figures the number of components decreases. As more nodes are added during graph growth, most of the newly added nodes are added to the giant component which decreases the number of disconnected components in the network. From Figure 5.2 (b) and (c) the behavior of  $C_n$  follows expectations. However, as seen in Figure 5.2(a) the number of components increase over time. In this case the network initially consisted of only 15% of the total number of nodes as disconnected components. Thus the initial number of components in the network is  $n_0 = 30$ . Consider the cases for  $\beta = 0.15$  and  $\beta = 0.5$ . As seen from the Figure 5.2, initially as new nodes are added, the number of components in the network starts to decrease till about t = 40. However after that the new nodes being added to the network do not join the giant component. They join the network as disconnected individual components. Therefore, the number of components goes on increasing. For low association thresholds like  $\beta = 0.15$  and  $\beta = 0.5$ , 85% and 50% of the possible edges are recognized, respectively. Therefore, initially  $C_n$  falls sharply to around 15 components. However, as new nodes are added,  $C_n$  increases. For the case of  $\beta = 0.85$  the association threshold is very high and therefore not many edges are recognized and added to the network. As a result, the new nodes being added to the network are included in the network as disconnected components. Therefore the value of  $C_n$  increases consistently. However, the fact that in all the cases of  $\beta$  in Figure 5.2(a), the number of components increases is a surprising finding.

In Figure 5.2(b) the initial number of components is 50% of the total network size,  $n_0 = 100$ . When a new node is recognized and added to the network, it connects many of the existing nodes forming a component which is greater in size than all other single node components. Therefore the total number of components decreases sharply in this case. The behavior is consistent for  $\beta$  values of 0.15, 0.5 and 0.85 except that for higher values of association threshold  $\beta$ , the change is gradual. This is because the higher  $\beta$  value implies a high number of edges are not recognized. Therefore it takes a while before the appearance of a giant component among other components. For  $\beta = 0.5$  the total number of components almost reaches a low minimum of 10 at the end of the growth process and for  $\beta = 0.85$  it almost reaches to 40 components from 100 components initially.

In Figure 5.2(c) the initial number of components is 85% of the total network size,  $n_0 = 170$ . This is similar to the case of  $n_0 = 100$ . For  $\beta = 0.15$  and  $\beta = 0.5$  the final number of components almost reaches a low minimum of approximately 40 from an initial high of 170 components.



Figure 5.3: Ratio of size of giant component to the total nodes in a network against time  $\left(\frac{S}{n}\right)$  for a graph of n=200 nodes. In figure (a)  $n_0=15\%$  of the nodes are initially assumed to be included in the graph as disconnected components, in (b)  $n_0=50\%$  and in (c)  $n_0=85\%$ . In each figure,  $\frac{S}{n}$  versus t is plotted for three values of association threshold  $\beta=[0.15, 0.5, 0.85]$ .

#### 5.2.2 Fraction of nodes in the giant component

The next measure is the ratio of size of the giant component (S) to the total size of the network (n) against time. This ratio is the fraction of the nodes in the graph which belong to the giant component at that time. It can be thought of as a probability that a node picked at random from the network belongs to the giant component.

Figure 5.3 shows the way in which the number of nodes in the giant component goes through a sudden phase transition. It can be seen that after a certain critical point most of the nodes belong to the giant component. The behavior of this measure is similar across all of the sample initial number of components considered, i.e., for n=30, 100 and 170. Initially,  $\frac{S}{n} = 0$  since all the nodes in the network are disconnected components and there is no giant component. However, as new nodes are recognized and added, it starts picking up edges with existing nodes thereby starting the formation of a giant component. This continues until suddenly due to the addition of a node or a few nodes, most of the other disconnected nodes now become connected to the majority of the other nodes forming a giant component. In Figure 5.3 (a) and (b) it can be seen that this happens at around t < 40, when a single component appears which contains the majority of the nodes in the network. The newly added nodes either join the giant component formed from some of the other smaller components or they form a single disconnected component of their own. If the number of new nodes joining the non-giant component nodes increases, then the value of  $\left(\frac{S}{n}\right)$  decreases slightly as can be seen in some instances. After the critical point is reached and a prominent giant component comes into existence, the addition of new nodes does not alter the behavior of  $\left(\frac{S}{n}\right)$  except for minor fluctuations.

The result is more pronounced for  $\beta = 0.15$  and  $\beta = 0.5$ , since these lower thresholds allow for more edges to be recognized and therefore components to form at a faster rate. The value of  $(\frac{S}{n})$  never completely reaches its peak at 1 because at the end of the growth process, there are still few disconnected parts of the network which add up to the number of non-giant components. Therefore,  $(\frac{S}{n})$  never reaches 1 as seen from the Figure 5.3. In Figure 5.3(a) the value of  $(\frac{S}{n})$  almost reaches 0.8 which means that 80% of the nodes are in giant component. For Figure 5.3(b) around 95% of the nodes are in the giant component and for Figure 5.3(c) around 80%.

# 5.2.3 Average size of non giant component U

The nodes which do not join the giant component are a part of the non-giant components. There can be many non-giant components consisting of varying number of nodes or even single disconnected nodes. The average size of non-giant component (U) is calculated by taking the ratio of the total number of nodes in the non-giant components to



Figure 5.4: Average size of non-giant components (U) over time for a graph of n = 200. In figure (a)  $n_0=15\%$  of the nodes are initially assumed to be included in the graph as disconnected components, in (b)  $n_0=50\%$  and in (c)  $n_0=85\%$ . In each figure, U versus t is plotted for three values of association threshold  $\beta=[0.15,0.5,0.85]$ .

the total number of non-giant components,

$$U = \frac{(n-S)}{u}$$

where n=total size of the network, S=number of nodes in the giant component, u=total number of non-giant components.

This measure is also a pretty good indicator of the phase transition phenomenon. As new nodes are added to the network they may either join the giant component or join nongiant component. Before phase transition the new nodes added to the network are more or less uniformly added to the smaller components which exist before the existence of a giant component. As a result the average size of the non-giant components (U) gradually increases till the critical point is reached. At this point the smaller components which have been forming suddenly join together to form a giant component. There is a drop in the value of U as the smaller components disappear to give rise to a giant component. This drop in the U value can be seen from Figure 5.4 (a), (b) and (c). Initially the value



Figure 5.5: Relationship between average degree and phase transition point. (a) Phase transition occurs when 50% of the nodes in giant component, (b) 70% and (c) 90% of the nodes. The initial graph contains  $n_0=50$  nodes as disconnected components. In each figure, phase transition point  $(p_T)$  versus t is plotted for three values of association threshold  $\beta = [0.15, 0.5, 0.85]$ .

of U increases till the critical point and then suddenly the value drops till it reaches the low minimum of 1. It means that other than the giant component there are single disconnected nodes in the network. Irrespective of the initial number of components in the network like for Figure 5.4 (a), (b) and (c) the behavior of U versus time is consistent. It is seen that U never goes beyond a maximum of approximately 1.4. This means that on an average there are fewer than two nodes in each of the smaller component. The results are more pronounced for higher association threshold, e.g.  $\beta = 0.85$ . This is according to expectation because as the threshold increases the number of recognized edges decreases giving rise to more disconnected components consisting of one or two nodes until gradually the critical point is reached and a giant component emerges.

#### 5.2.4 Phase transition point versus average degree z

Here we measure at what point the phase transition occurs for networks of varying average degrees (z). The average degree is indicative of the connectivity of the network.

It is seen in Section 4.3.2 that connectivity is a direct indicator of the level of comprehension. This section seeks to establish the relationship between the phase transition point, i.e., the moment at which the giant component emerges, and the average degree, i.e., the connectivity of the concept network which is indicative of the level of comprehension. It is seen that an earlier emergence of the giant component is indicative of deeper comprehension.

Figure 5.5 (a), (b) and (c) show the plot of the critical point as a function of the average degree z. It is assumed that initially there are 50 disconnected nodes in the network which form the individual components of the network. The final size of the network is fixed to n = 100. A different network is grown for each value of average degree z = [1:20]. For instance, in a network of average degree 1, the number of edges in a network is equal to the number of nodes. Initially this graph will contain  $n_0 = 50$  nodes. At each time step a node is presented for recognition. If it passes the recognition threshold test then it is added to the network. Once it is added, those edges which pass the association threshold test are also added to the network. At this point a check is performed to see if there exists a giant component in the network. We test for three benchmark measures for the existence of a giant component. In the first one, if more than 50% of the nodes in the graph are in one component then it is assigned as a giant component. The second and third benchmark is for 70% of the nodes and 90% of the nodes in the giant component. Figure 5.5 (a), (b) and (c) show the plot for each of the three benchmarks respectively.

Figure 5.5(a) shows the plots of the time step at which 50% of the nodes are in the giant component. The point at which phase transition is assumed to have occurred is

defined as the critical point or the phase transition point  $(p_T)$ . This point is recorded for each of the grown graphs with different average degrees. Thus we get a total of 20 phase transition points  $(p_T)$  for each value of the average degree (z). It can be clearly seen from the figures that as the average degree increases, the phase transition occurs at an earlier time. For example, in Figure 5.5(a) for an association threshold value of  $\beta = 0.15$ , for a network with degree z = 1, the phase transition occurs at  $p_T = 80$ . This means that since the network initially contains 50 individual components, it takes 30 more time steps for 50% or more of the nodes to be included in the giant component. However, as the average degree z increases, the number of time steps at which this critical points occurs goes on decreasing. For instance, for an association threshold value of  $\beta = 0.5$ and for a network with average degree z = 10, the critical point at which more than 50% of the nodes are included in the giant component has at  $p_T \approx 55$ , which means only after 5 time steps.

The reason for this phenomenon is that as the average degree of the network increases, the number of edges in the network increases. With increase in the number of edges in the network the probability of a majority of the nodes being connected in a single giant component increases. This can be illustrated by constructing the networks for different values of association threshold  $\beta = [0.15, 0.5, 0.85]$ . As  $\beta$  increases the number of associations or edges in a network decreases. Therefore the phase transition occurs at a later time than it would have for a lower value of  $\beta$ . For instance, in Figure 5.5(a) for a network with z = 20, the phase transition point for  $\beta = 0.15$  is at  $p_T \approx 55$  and for  $\beta = 0.85$  is at  $p_T \approx 60$ .

Figure 5.5 (b) and (c) shows similar plots for benchmarks 2 and 3. The general

behavior of the plot is very similar to that of Figure 5.5(a) as expected. However it is surprisingly seen that although in benchmarks 2 and 3, the size of giant component increases, the critical points at which phase transition occurs do not show much of a difference. For instance consider Figure 5.5(b) in which critical point is the point at which 70% of the nodes are in the giant component. Now theoretically the phase transition of this benchmark must occur at a later point than for benchmark 1 in which phase transition occurs at a point when only 50% of the nodes are in the giant component. However as seen from Figure 5.5(b), say at z = 20, the  $p_T$  values for  $\beta = [0.15, 0.5, 0.85]$ are  $p_T \approx [53, 54, 60]$ , much like those seen in Figure 5.5(a). The same is true for Figure 5.5(c). This means that, when phase transition occurs in these concept networks, at least 90% of all the nodes in the network are connected in the giant component. In Figure 5.5(c) for a higher value of  $\beta = 0.85$ , the phase transition occurs at  $p_T = 70$ . It is expected that as  $\beta$  increases the phase transition will start occurring at greater time steps.

#### 5.3 Implications

Children with Autism Spectrum Disorder (ASD) have significant problems in performing semantic tasks related to comprehension like recall and inference [Chan et al. 1997, 1998] because of the disconnectedness of their semantic memory structures. ASD patients do not show a considerable loss of semantic memory content like Alzheimers patients but show a significant degradation in the network structure. For instance, there is performance degradation in recall tasks of retrieving neighborhood concepts pointing to a decrease in context utilization and superior performance on false memory tasks in ASD patients [Beversdorf et al. 1998, 2000, 2007]. There is also a low degree of information integration and synchronization suggesting deterioration in reachability and connectedness in terms of pathways and scale of degree distribution [Just et al. 2004]. There is also evidence of a weakening of connections between concepts measureable as connections strengths [Tivarus et al. 2006] or the number of hops and also a loss of associations [Chan 1997]. It is seen that not only are there a significantly fewer of clusters in the concept/semantic network [Binetti et al. 1995; Bousfield, 1953], but also the deterioration is not evenly distributed among the components.

In general, ASD patients display disorganized concept networks in terms of the absence of a giant component and also low individual node connectivity. The study of evolution of component structure and phase transition in comprehension networks will provide some quantitative insight to researchers studying the deterioration of concept networks in ASD patients.

## 5.4 Conclusion

The connectedness of in complex networks is supported by studies of networks like power grids, metabolic networks, protein interactions, neural networks etc. [Albert et al. 2000]. We ask the same question in the context of knowledge networks. It is seen that as the components grow in size independently, there comes a point often referred to as the phase transition point at which all components get connected and converge to a single giant component. After this point most of the concepts are accessible through a path. It can be concluded that the existence of a giant component is indicative of higher level of comprehension. Also, sooner the giant component appears, higher is the comprehension.

# CHAPTER 6

## Conclusion

In this Chapter we will discuss some of the issues and future considerations of our presented research which may help in putting the work into perspective.

6.1 Some thoughts about the mathematical model ...

- Many different connectionist theories of learning incorporate the possibility of node deletion or what is generally known as forgetting or de-learning learned knowledge. At this moment our model does not incorporate this facility. However, given the nature of our model it would not be difficult to represent the deletion of an existing node in terms of inequalities. The deletion could be produced by a drop in the association strengths of links to that node.
- 2. Our model does not incorporate the existence of associations between concepts which are offered in the same episode. This is because of the granularity of the episode rather than a deficiency is in the model itself. An episode consists of reading a whole sentence, therefore a number of concepts can be offered in that episode. However in reality an episode may be still finer grained, i.e., may consist of reading a word at a time, or even an alphabet at a time. This discrepancy can be remedied by using better data collection methodologies like eye trackers, magnetic resonance imaging or electroencephalography. In the simulation experiment we handle this by offering only one node for recognition at a time.

- 3. The association weights are computed by solving a linear optimization problem in which the objective function is to minimize the weights. The values of the association weights are therefore a product of the optimization algorithm. It would be interesting to experiment with different optimization algorithms.
- 4. An interesting debate is whether all knowledge is connected in itself? We believe that the conscious concept network (in the working memory) might have disconnected components. In the model it is completely possible that a concept may be recognized but not linked to any other concept in the concept network. This is actually the reason why islands of concepts (disconnected components) may form in a concept network. However, philosophically, the possibility of this happening is nonexistent [Fodor 1975]. If the atomistic theory of concepts is taken to be true, then any concept is the assembly or construction of further primitive concepts. In terms of the concept network this means that a node is basically the assembly of its primitive nodes to which it connects. Therefore it is impossible to imagine a node which can have meaning in absence of any connections to its primitives in this framework. As Wittgenstein puts it, the meaning of a word is in the company of other words. Our model is able to subsume this condition without sounding any alarms.
- 5. The simulation in itself is done for networks of small size up to 500 concepts. In reality concept networks may contain a number of concepts which is too large to even fathom. Comparable studies have simulated networks which are not more than a few thousand nodes (maximum 5018). In the future realistic studies of human

concept networks are going to demand advanced methods of data collection and network construction and simulations which can scale to millions of nodes.

- 6. The simulation done to study the phase transition phenomenon in Section 5.2 assumes that there are pockets of disconnected concepts even before the beginning of comprehension which may not be a realistic assumption (although various meanings of the word disconnected can be employed here to suit our case). For instance, components which have in between them associations below certain minimum association strengths can be considered disconnected in spite of there being a week connection.
- 7. The model does not incorporate the possibility of acquiring the concept before it is presented.
- 6.2 Short term versus long term memory

Where does the SI model stand with respect to semantic memory? In current psychology and cognitive science literature it is usually understood that there are two types of memories namely short term memory and long term memory. The long term memory is usually assumed to store decontextualized concepts knowledge like alphabets, words, propositions, grammar rules and elementary aspects of language. The short term memory stores the knowledge which is actively being processed. For example, according to [Kintsch, Mangalnath 2011], when reading a sentence the recognized concepts are usually stored in the short term memory and comprehension occurs when the recognized concepts react with the stored concepts in the long term memory to form a coherent representation. The SI model in its current form is a theoretically infinite memory model in terms of size and time. Also it does not distinguish between the short term and long term memories. In this model, both the concepts which are recognized and the concepts which are required apriori are assumed to be stored in the same memory. There is no distinction between contextual and decontextualized knowledge. All knowledge is assumed to be contextual knowledge. In fact, even the background node which represents all the knowledge possessed by the reader is available during the SI process. Other models consider this background knowledge as decontextualized knowledge stored in the long term memory.

Studies in eye fixations during reading [Just, Carpenter 1992] suggest that there is a limit to the number of concepts which can be stored while processing a sentence and factors like size of working memory does affect syntactic processing [Kemper, Crow, Kemtes 2004; Kintsch, Patel, Ericsson 1999]. So, there is some value to designing a model which keeps the differences between working and long term memory in mind. As of now the SI model is indifferent to this aspect but can be modeled to incorporate the difference in the mathematical workings. This difference can be thought of as a difference between definitions of what *long* and *short* in the context of semantic memory mean. At what point does knowledge in the active working memory ceases to be in the short term memory and graduate to long term memory. Another issue is the elementary nature of the knowledge which is supposedly stored in the long term memory. For instance, it is assumed that the long term memory does not store complex abstractions but stores elementary concepts which can be later combined and contextualized to generate complex abstractions. This raises the question, what are the criteria for judging the concepts in
the working memory if they are elementary enough to graduate to the long term memory? This again requires a very precise definition of what an elementary concept is and the discussion beings to get fuzzy.

### 6.3 What is a concept?

Almost every faculty of scientific thought which tries to inquire about learning and knowledge has its views about what a concept is. However, in the presentation of this dissertation we take caution to not make any declarative statements about the notion of a concept. The reason is twofold. Firstly, this dissertation does not seek to inquire about the nature of a "concept" but merely tries to use the well-established belief that leaning systems seek to understand the algorithms by which "concepts" are learned into useful representations. The SI algorithm in itself is independent of the nature of concepts although it deals with the structural manipulation of concepts and the relationships between them. Concepts in the SI model sense are considered as objects with properties of their own which the SI model does not seek to change. It is often said that meaning is derived from the context, which is this case are the nodes that a concept connects to and this is the aspect with which this dissertation concerns itself the most. The SI interpretation does not care if the concept is based on the atomistic, classical, or exemplar/classification based model but merely seeks to understand the mechanistic nature of structural changes in the representation due to concept acquisition during text comprehension.

This does not by any means imply that it is not worthwhile to get a more complete understanding about the notions of concepts. This in turn brings to the fore the second reason to abstain from discussing the issue of concepts in greater detail than might seems useful which is, there is a great debate as to what exactly is the philosophical notion of a concept. Indeed there are some who believe in the complete atomistic nature of concepts [Fodor 1975] and the notion that in fact the whole cognitive science faculty has gone "wrong" is assuming the anti-atomistic (propositional, categorical, etc.) nature of concepts instead. This belief is strongly held by psychologists and computer scientists alike. In line with this thought the SI model of comprehension subscribes to the atomistic view of concepts in which the content of concepts or the structure of concepts is inconsequential unless the concept can itself be broken down into even more fundamental concepts.

# 6.4 Constructivism and the SI model

Comprehension is often thought of as a paradigm for cognition [Kintsch 1999] in which meaning is actively constructed from the incoming knowledge and prior knowledge [Kintsch 2009]. Text comprehension, therefore, essentially is an act of construction according to the theory of constructivism. So how does the concept network growth using the SI model relate to the cognitive theory and the epistemology of constructivism? The answer to this question lies in the meaning of constructivism. Constructivism implies that the construction of a knowledge space is irrespective of any ontological reality because even if there were such an ontological reality, there is no way of validating the correctness of this ontological reality. Hence constructivists believe that although knowledge may be organized in some correct way, there is no way of validating this statement and therefore it does not exist. According to constructivism, knowledge is constructed by the person and not discovered from the external world. All knowledge is thus internally constructed. In the face of this realization, constructivism is then a theory of construction instead of discovery.

This idea fits extremely well with the model of concept network growth. Constructivist models derive more from the interactionist perspective of language development. These models try to incorporate both the nativist and behaviorist point of views. Firstly, like the empiricists, constructivists believe that information processing is connectionist and secondly like the nativist they believe that there is an innate desire for constructive growth from interaction. However, in the past constructivism has had some problems with the nativist idea of innate concepts, since all knowledge is believed to be constructed [Elman 1996; Marcus 1998]. The SI model steers clear of making any judgments about the nature of knowledge in itself although the idea that complex concepts to arise from primitive concepts necessitates innate concepts. In this dissertation we have studied a very specific problem, that of the segmentation and integration of knowledge in text comprehension. It is loosely inspired by the theory of constructivism. We in fact show that by representing assimilation as a constraint satisfaction problem the necessary conditions for knowledge acquisition can be computed.

6.5 Concept network as an objective measure of quality of comprehension

In Chapters 4 and 5 we saw how certain properties of concept networks of comprehension are indicative of the quality of comprehension. Here we list those properties.

- Size (number of concepts and associations)- Higher number of nodes and associations are usually indicative of a well formed concept network and deeper comprehension.
- Reachability (average shortest path length, diameter)- A reachable networks is one in which most of the nodes are reachable from each other through a path or multiple paths. A highly reachable network is indicative of good comprehension.
- 3. Connectivity/Sparsity (average degree, average weighted degree)- A higher average degree indicates greater connectivity of nodes and is indicative of deeper comprehension.
- 4. Connectivity/Sparsity (average degree, average weighted degree)- A higher average degree indicates greater connectivity of nodes and is indicative of deeper comprehension.
- 5. *Neighborhood clustering-* A high local clustering indicates good connectivity are therefore indicates deep comprehension.
- 6. Component structure (number of components, size of giant component, size of nongiant components)- A lot of disconnected components indicate low reachability and therefore single giant component is indicative of a well reachable network and deep comprehension.
- 7. Community structure (number of communities, density of connections within and in between communities)- A community is the organization of similar nodes in a network. A well-defined community structure indicates good organization indicative

of good comprehension. This is a very interesting research avenue for the future.

8. *Resilience* (degree distribution)- Resilience indicates the robustness of a network against random or targeted node or association removal attacks. A normal like distribution is able to main the reachability and connectivity fairly well in case of these attacks and is therefore indicative of good comprehension.

### 6.6 Conclusion

The main thesis of this dissertation is as follows one of the main cognitive processes in text comprehension is the segmentation and integration of concepts (SI) from the text. This process can be mathematically modeled as thresholded concept network growth process. The model parameters are used to explain how and why different readers construct different concept networks on reading the same text. It can also describe why some readers may understand a text easily as compared to other readers, and also why some texts are difficult to understand than other texts for the same reader.

The model can be used to study phenomenon like the effect of concept sequence on text comprehension. Also, the concept network which grows during comprehension shows some peculiar features like high reachability, high connectivity and existence of a single connected component which are indicative of the good comprehension. These dynamic structural properties can provide insights into the level of comprehension from the point of view of the concept networks constructed.

## CHAPTER 7

# Appendix A: Data Collection

1. The paragraph of text extracted from Computer Communication Networks, Peterson and Davies, 5th edition, pp. 150.

"The Ethernet is easily the most successful local area networking technology of the last 20 years. Developed in the mid-1970s by researchers at the Xerox Palo Alto Research Center (PARC), the Ethernet is a working example of the more general Carrier Sense, Multiple Accesses with Collision Detect (CSMA/CD) local area networking technology. As indicated by the CSMA name, the Ethernet is a multiple-access network, meaning that a set of nodes send and receive frames over a shared link. You can, therefore, think of an Ethernet as being like a bus that has multiple stations plugged into it. The carrier sense in CSMA/CD means that all the nodes can distinguish between idle and a bust link, and collision detect means that a node listens as it transmits and can therefore detect when a frame it is transmitting has interfered (collided) with a frame transmitted by another node. The Ethernet has its roots in an early packet radio network, called Aloha, developed at the University of Hawaii to support computer communication across the Hawaiian Islands. Like the Aloha network, the fundamental problem faced by the Ethernet is how to mediate access to a shared medium fairly and efficiently (in Aloha the medium was the atmosphere, while in Ethernet the medium is a coax cable). That is, the core idea in both Aloha and the Ethernet is an algorithm that controls when each node can transmit."



2. Reader generated individual concept networks (ICN) for the total 9 readers. Each node represents the concept and the timestamp on each concept represents the episode in which the concept was recognized. Concepts which were presented in an episode and not recognized are shows by dotted outlines. The edges between concepts are as drawn by the readers. The time at which an association is recognized is denoted by the higher time stamp between the two concepts to which is connects.



Figure 7.0: Data collection: Drawn ICNs

# CHAPTER 8

Appendix B: Derivation for degree distribution

Let the random variable  $u_{ij}$  represent the link between the  $i^{th}$  and the  $j^{th}$  node. If a link is present between the two nodes then this random variable will take the value 1 otherwise is 0. The recognition of a node is a two-step process. At the first step all the links which are greater than association threshold  $\beta$  are recognized. In the next step the summation of all the links greater than the threshold is computed and if this sum is greater than the recognition threshold  $\alpha$  then the concept too is recognized. If a concept is recognized, then all the links greater than  $\beta$ , are retained. If the concept is not recognized, all the links are discarded. Therefore at the first step of constraints the degree of a node i is given by,

$$D_i = \sum_{j=1}^n u_{ij}$$

where  $u_{ij}$  is a Bernoulli random variable in (1, p) where p is the probability of existence of a link. Therefore,  $P(u_{ij} = 1) = p$ .

Since the association threshold  $\beta$  is the chance that the link will not be recognized, the probability of link being recognized is  $p = 1 - \beta$ .

After the second threshold is applied the degree of a node changes. Some links are kept, some are dropped. If node is dropped then links are dropped too. Let  $v_{ij}$  be the random variable for the links of the node after the application of the recognition threshold

 $\alpha$ .

According to the model the degree of a node only changes when the random variable  $v_{ij}$  exceeds the threshold  $\alpha$ , otherwise all the links are dropped and  $v_{ij}$  assumes the value of 0. Thus,

$$v_{ij} = \begin{cases} u_{ij} & \sum_{j=1}^{n} u_{ij} \ge \alpha \quad where \quad j = 1, 2, ..., n \\ 0 & \sum_{j=1}^{n} u_{ij} < \alpha \quad where \quad j = 1, 2, ..., n \end{cases}$$

The final degree is,

$$D_i = \sum_{j=1}^n v_{ij}$$

Let the degrees of nodes be denoted by a vector  $(D_1, D_2, ..., D_n)$ . Here let us define another random variable K where,

Probability that a node has degree k = P(K = k)=  $\frac{number \ of \ nodes \ with \ D_i = k; \ i = 1, 2, ..., n}{n}$ 

The mean degree d is calculated as,

$$d = E[P(K = k)]$$
  
=  $\frac{P(D_1 = k) + P(D_2 = k) + \dots + P(D_n = k)}{n}$ 

But since each node has equal probability of having the degree k let,

$$P(D_1 = k) = P(D_2 = k) = \dots = P(D_n = k) = P(D = k)$$

Thus, d = E[P(K = k)] = P(D = k)

Now to find the degree distribution we divide the curve into 3 parts. The first part is for k = 0 when there is an accumulation of zero degree nodes at the beginning for all the non-recognized nodes.

1. For k=0,

$$E[P(K=0)] = P(D=0) = P(\sum_{k=0}^{n} v_{nk} = 0) = P(\sum_{k=0}^{n} u_{nk} < \alpha)$$

This is a binomial which can be represented as follows,

$$E[P(K=0)] = \sum_{k=0}^{\alpha} \binom{n}{k} (1-\beta)^k \beta^{(\alpha-k)}$$

The second part is for  $1 \le k < \alpha$  when the edges are recognized but are dropped because the nodes which they connect to are not recognized.

2. For  $1 \leq k < \alpha$ ,

$$E[P(K = \alpha)] = P(D = \alpha) = P(\sum_{k=1}^{n} u_{nk} = \alpha; \sum_{k=1}^{n} u_{nk} < \alpha) = 0$$

The third part of the curve is for  $\alpha \leq k \leq n$  when the nodes are recognized and start accumulating degrees greater than the association threshold.

3. For  $\alpha \leq k \leq n$ ,

$$E[P(K = k)] = P(D = k) = P(\sum_{k=1}^{n} u_{nk} = k)$$

Again this is a binomial which can be represented as follows,

$$E[P(K=k)] = \binom{n}{k} (1-\beta)^k \beta^{(n-k)}$$

Thus, the curve can be represented as a three part function,

$$f(n) = \begin{cases} \sum_{k=0}^{\alpha} {n \choose k} (1-\beta)^k \beta^{(\alpha-k)} & k = 0\\ 0 & 1 \le k < \alpha\\ {n \choose k} (1-\beta)^k \beta^{(n-k)} & \alpha \le k \le n \end{cases}$$

Generalizing the above binomial to a normal distribution for high values of n,

$$P(k) = \frac{1}{\sqrt{n(1-\beta)\beta}} \frac{1}{\sqrt{2\pi}} e^{\frac{-(k-n(1-\beta))^2}{2n(1-\beta)\beta}}$$

# BIBLIOGRAPHY

- Adams, Marilyn Jager (1994). Beginning to read: thinking and learning about print. Cambridge, Mass: MIT Press.
- [2] Albert, R. and Barabasi, A.-L., Statistical mechanics of complex networks, Rev. Mod. Phys. 74, 47-97 (2002).
- [3] Albert, R., Jeong, H., and Barabasi, A.-L., Attack and error tolerance of complex networks, Nature 406, 378-382 (2000).
- [4] Aleksiejuk, A., Hoalyst, J. A., and Staurer, D., Ferro-magnetic phase transition in Barabasi Albert networks, Physica A 310, 260 266 (2002).
- [5] Amaral, L.A.N., Scala, A., Barthelemy, M., & Stanley, H.E. (2000). Classes of smallworld networks. Proceedings of the National Academy of Sciences, 97, 11149-11152.
- [6] Anderson, J.R. (2000). Learning and memory: An integrated approach (2nd ed.). New York, NY: John Wiley & Sons, Inc.
- [7] Artificial intelligence, 2nd edition, Rich, E. and Knight, K., 1991, McGraw-Hill.
- [8] Au A. Chan A. S. & Chiu H. Conceptual organization in Alzheimer's dementia Journal of Clinical and Experimental Neuropsychology2003 Sep; 25(6):737-50.
- [9] Aizerman M., Braverman E., and Rozonoer, L. (1964). "Theoretical foundations of the potential function method in pattern recognition learning". Automation and Remote Control 25: 821837
- [10] Barrat, A. and Weigt, M., On the properties of small world networks, Eur. Phys. J. B 13, 547560 (2000).
- [11] Beversdorf DQ, Anderson JM, Manning SE, Anderson SL, Nordgren RE, Felopulos GJ, et al. The effect of semantic and emotional context on written recall for verbal language in high functioning adults with autism spectrum disorder. Journal of Neurology, Neurosurgery, and Psychiatry. 1998; 65(5):685692.
- [12] Beversdorf DQ, Narayanan A, Hillier A., Hughes JD. Network model of decreased context utilization in autism spectrum disorder. J Autism Dev Disord. 2007 Jul; 37(6):1040-8.
- [13] Bianconi, G., Mean field solution of the Ising model on a Barabasi-Albert network, Preprint cond-mat/0204455 (2002).

- Blei, David M.; Jordan, Michael I.; Griffiths, Thomas L.; Tenenbaum; Joshua B (2004). "Hierarchical Topic Models and the Nested Chinese Restaurant Process". Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference. MIT Press.
- [15] Blei, David M.; Ng, Andrew Y.; Jordan, Michael I (January 2003). Lafferty, John. ed. "Latent Dirichlet allocation". Journal of Machine Learning Research 3 (45): pp. 9931022.
- [16] Bousfield W.A. (1953). The occurrence of clustering in the recall of randomly arranged associates. Journal of General Psychology, 49, 229-240.
- [17] Callaway, D. S., Newman, M. E. J., Strogatz, S. H., and Watts, D. J., Network robustness and fragility: Percolation on random graphs, Phys. Rev. Letters. 85, 5468-5471 (2000)
- [18] Carey, S. (1978). The child as word learner. In J. Bresnan, G. Miller, and M. Halle (Eds.), Linguistic theory and psychological reality. Cambridge, MA: The MIT Press.
- [19] Carrell, P. L. (1984). Evidence of a formal schema in second language comprehension. Language Learning, 34, 87-113.
- [20] Chan A. S., Salmon D. P., Butters N., Johnson S. A. (1995) Semantic network abnormality predicts rate of cognitive decline in patients with probable Alzheimer's disease. Journal of the International Neuropsychological Society. Online publication date: 1-May-1995.
- [21] Chan A., Butters N., & Salmon D. (1997). The deterioration of semantic networks in patients with Alzheimer's disease: A cross-sectional study. Neuropsychologia, 241-248.
- [22] Chan A., Butters N., & Salmon D. (1998). Semantic networks abnormality of patients with Alzheimer's disease (p 381-393). In R. W. Parks D. S. Levine & D. L. Long (Eds). Fundamentals of neural network modeling: Neuropsychology and cognitive neuroscience Mass: MIT Press.
- [23] Chan A., Butters N., Paulson J., Salmon D., Swenson M. & Maloney L. (1993). An assessment of the semantic network in patients with Alzheimer's disease. Journal of Cognitive Neuroscience 254-261.
- [24] Chan A., Butters N., Salmon D., & McGuire K. (1993). Dimensionality and clustering in the semantic network of patients with Alzheimer's disease. Psychology and Aging 411-419.
- [25] Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. Modern Language Journal, 80, 183-198.
- [26] Chun, D. M., & Plass, J. L. (1996). Facilitating reading comprehension with multimedia. System, 24(4), 503-519.

- [27] Clark, E.V. (1993). The lexicon in acquisition. New York, NY: Cambridge University Press.
- [28] Clark, E.V. (2001). Making use of pragmatic inferences in the acquisition of meaning. In D. Beaver, S. Kaufman, B. Clark, and L. Cazillas (Eds.), Stanford Papers in Semantics. Stanford CA: CSLI Publications, 23-36.
- [29] Collins, A.M., & Quillian, M.R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-248.
- [30] Collins, A.M., Loftus, E.F. (1975). A spreading activation theory of semantic processing. Psychological Review, 82, 407-428.
- [31] CN 2012, "Complex Network": http://en.wikipedia.org/wiki/Complex\_ network. Retrieved Feb 2012.
- [32] Davis, Randall, and Douglas B. Lenat. Knowledge-Based Systems in Artificial Intelligence. New York: McGraw-Hill International Book Co, 1982.
- [33] de Beaugrande, R. (1982). General constraints on process models of language comprehension. In J.-F. L. Ny & W. Kintsch (Eds.), Language and comprehension (pp. 179-192). Amsterdam: North-Holland.
- [34] Deerwester, S., Dumais, S. T., Landauer, T. K., Furnas, G. W. and Harshman, R. A. (1990) "Indexing by latent semantic analysis." Journal of the Society for Information Science, 41(6), 391-407.
- [35] Dorogovtsev, S. N., Goltsev, A. V., and Mendes, J. F. F., Ising model on networks with an arbitrary distribution of connections, Phys. Rev. E 66, 016104 (2002).
- [36] Dunne, J. A., Williams, R. J., and Martinez, N. D., Food-web structure and network theory: The role of connectance and size, Proc. Natl. Acad. Sci. USA 99, 12917-12922 (2002).
- [37] Dunne, J. A., Williams, R. J., and Martinez, N. D., Network structure and biodiversity loss in food webs: Robustness increases with connectance, Ecology Letters 5, 558-567 (2002).
- [38] Eliasmith, C. (1996). The third contender: A critical examination of the dynamicist theory of cognition. Philosophical Psychology, 9, 441-463. Reprinted in P. Thagard (Ed.). (1998). Mind readings: Introductory selections in cognitive science. MIT Press.
- [39] Eliasmith, C. (1997). Computation and dynamical models of mind. Minds and Machines, 7, 531-541.
- [40] Ellis, A.W., & Lambon Ralph, M.A. (2000). Age of acquisition effects in adult lexical processing reflects loss of plasticity in maturing systems: insights from connectionist networks. Journal of Experimental Psychology: Learning, Memory, & Cognition.
- [41] Elman, J. L. (1990) Finding structure in time. Cognitive Science, 14(2):179–211.

- [42] Elman, J.L., Bates, E., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., Plunkett, K., 1996. Rethinking Innateness: A Connectionist Perspective on Development. Cambridge, MA: MIT Press
- [43] Fodor, Jerry (1975). The Language of Thought. Harvard University Press, Cambridge, Mass.
- [44] Fodor, Jerry (1998) Concepts: Where cognitive science went wrong?
- [45] G. A. Miller, R. Beckwith, C. D. Fellbaum, D. Gross, K. Miller. 1990. WordNet: An online lexical database. Int. J. Lexicograph. 3, 4, pp. 235-244.
- [46] Gerrig, R., & McKoon, G. (1998). The readiness is all: The functionality of memorybased text processing. Invited article, Discourse Processes, 26, 67-86.
- [47] Gilhooly, K.J. and Logie, R.H. (1980). Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. Behavior Research Methods and Instrumentation, 12, 395-427.
- [48] Giuliano Binetti, Eugenio Magni, Stefano Cappa, Alessandro Padovani, Angelo Bianchetti, Marco Trabucchi. (1995) Semantic memory in Alzheimers disease: An analysis of category fluency. Journal of Clinical and Experimental Neuropsychology 17:182-89 Online publication date: 1-Feb-1995.
- [49] Goodman, K. (1967). Reading: A psycholinguistic guessing game. Journal of the Reading Specialist, 6, 126-135.
- [50] Grabe, W. (1991). Current developments in second language reading research. TESOL Quarterly, 25(3), 375-406.
- [51] Griffiths, Thomas L.; Steyvers, Mark (April 6 2004). "Finding scientific topics". Proceedings of the National Academy of Sciences 101 (Suppl. 1): 52285235.
- [52] Hills, T., Maouene, M., Maouene, J., Sheya, A., and Smith, L. B. (2008). Categorical structure in early semantic networks of nouns. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society (pp. 161-166). Austin, TX: Cognitive Science Society.
- [53] Hills, T., Maouene, M., Maouene, J., Sheya, A., and Smith, L. B. (2008). Is there preferential attachment in the growth of early semantic noun networks? In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society (pp. 155-160). Austin, TX: Cognitive Science Society.
- [54] Hills, T., Maouene, M., Maouene, J., Sheya, A., and Smith, L. B., (2009) Longitudinal analysis of early semantic networks: preferential attachment or preferential acquisition. Psychological Science June 2009 vol. 20 no. 6 729-739
- [55] Hills, T., Maouene, M., Maouene, J., Sheya, A., and Smith, L. B., (2009) Categorical structure among shared features in networks of early-learned nouns, Cognition, Volume 112, Issue 3, September 2009, Pages 381-396.

- [56] Hinton, G.E., Sejnowski, T. J., and Ackley, D. H. Boltzmann Machines: Constraint satisfaction networks that learn. Technical Report CMU-CS-84-119, Carnegie-Mellon University.
- [57] Hodges JR, Salmon DP, Butters N. Semantic memory impairment in Alzheimer's disease: failure of access or degraded knowledge? Neuropsychologia. 1992 Apr; 30(4):301-14.
- [58] Hofmann, Thomas (1999). "Probabilistic Latent Semantic Indexing". Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval.
- [59] Johnson-Laird, P. N. (1983). Mental Models. Towards a cognitive science of language, inference, and consciousness. Cambridge, MA: Harvard University Press.
- [60] Just Marcel Adam; Cherkassky V L.; Keller T A.; and Minshew N J. "Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of under connectivity" (2004). Department of Psychology. Paper 323.
- [61] Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixation to comprehension. Psychological Review, 87, 329354.
- [62] Just, M., & Carpenter, P. (1987). The psychology of reading and language comprehension. Boston: Allyn & Bacon.
- [63] Just, Marcel A.; Carpenter, Patricia A. (1992). A capacity theory of comprehension: Individual differences in working memory. Psychological Review, Vol 99(1), Jan 1992, 122-149.
- [64] Just, Marcel A.; Carpenter, Patricia A.; Woolley, Jacqueline D. (1982). Paradigms and processes in reading comprehension. Journal of Experimental Psychology: General, Vol 111(2), Jun 1982, 228-238.
- [65] Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95(2), 163-182.
- [66] Kintsch, W. (1994). Text comprehension, memory, and learning. American Psychologist, 49(4), 294-303.
- [67] Kintsch, W. (1998). Comprehension: A paradigm for cognition. New York: Cambridge University Press.
- [68] Kintsch, W. (2000) Metaphor comprehension: A computational theory. Psychonomic Bulletin & Review, 7, 257-266.
- [69] Kintsch, W. (2001) Predication. Cognitive Science, 25, 173-202.
- [70] Kintsch, W. & Mangalath, P. (2011) The construction of meaning. Topics in Cognitive Science, 3, 346-370.

- [71] Kintsch, W., Patel, V., Ericsson, K. A. (1999) The role of Long-term working memory in text comprehension. Psychologia, 42, 186-198.
- [72] Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.
- [73] Landauer, T. K., Laham, D., Rehder, B., & Schreiner, M. E., How well can passage meaning be derived without using word order? A comparison of Latent Semantic Analysis and humans. In M. G. Shafto & P. Langley (Eds.), Proceedings of the 19th annual meeting of the Cognitive Science Society (pp. 412-417). 1996 Mahwah, NJ: Erlbaum.
- [74] Landauer, T., Laham, D., Foltz, P., Learning Human-like Knowledge by Singular Value Decomposition: A Progress Report, M. I. Jordan, M. J. Kearns & S. A. Solla (Eds.), Advances in Neural Information Processing Systems 10, Cambridge: MIT Press, 1998, pp. 4551
- [75] Laufer, B., & Sim, D. (1985). Measuring and explaining the reading threshold needed for English for academic purposes texts. Foreign Language Annals, 18, 405-411.
- [76] Laurence, Stephen and Margolis, Eric. 1999. Concepts and Cognitive Science. In Margolis and Laurence 1999, 3-81.
- [77] Leone, M., Vazquez, A., Vespignani, A., and Zecchina, R., Ferromagnetic ordering in graphs with arbitrary degree distribution, Eur. Phys. J. B 28, 191197 (2002).
- [78] Mitzenmacher, M.A Brief History of Generative Models for Power Law and Lognormal Distributions. Internet Mathematics, vol 1, No. 2, pp. 226-251, 2004.
- [79] Manas S. Hardas, Javed Khan, "Concept Learning in Text Comprehension", Y.Y. Yao et al. (Eds.): BI 2010, LNAI 6334, pp. 240-251, 2010.
- [80] Mannes, S. M., & Kintsch, W. (1987). Knowledge organization and text organization. Cognition and Instruction, 4, 91115.
- [81] Marcus G. F., Can connectionism save constructivism? Cognition.1998 May;66(2):153-82.
- [82] Marilyn Kourilsky, M.C. Wittrock, Verbal and graphical strategies in the teaching of economics, Teaching and Teacher Education, Volume 3, Issue 1, 1987, Pages 1-12.
- [83] Marilyn Kourilsky, Merlin C. Wittrock (1992) Generative Teaching: An Enhancement Strategy for the Learning of Economics in Cooperative Groups. American Educational Research Journal, Vol. 29, No. 4 (Winter, 1992), pp. 861-876.
- [84] Mark Steyvers; Tom Griffiths (2007) "Probabilistic Topic Models" In: T. Landauer, D McNamara, S. Dennis, and W. Kintsch (eds), Handbook of Latent Semantic Analysis, Psychology Press.

- [85] Martin A. (1992). Degraded knowledge representations in patients with Alzheimers disease: Implications for models of semantic and repetition priming. In L.R. Squire & N. Butters (Eds.) Neuropsychology of memory (2nd ed.) (pp. 220232). New York: Guilford Press
- [86] Marvin Minsky and Oliver G. Selfridge. "Learning in random nets" Proc. Fourth London Symposium on Information Theory, Butterworth, Ltd., London, 1961.
- [87] Mayer, R. E. (1993). Comprehension of Graphics in texts: An overview. Learning and Instruction, 3, 239-245.
- [88] McClelland, J.L., D.E. Rumelhart and the PDP Research Group (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models, Cambridge, MA: MIT Press
- [89] Medler D. A. (1998). A connectionist model of semantic network disturbances evident in Alzheimers patients [Abstract]. Journal of Cognitive Neuroscience (Suppl.) 31.
- [90] Mehrotra, S., "On the Implementation of a Primal-Dual Interior Point Method," SIAM Journal on Optimization, Vol. 2, pp. 575601, 1992.
- [91] Morrison, C.M., Chappell, T.D., & Ellis, A.W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. Quarterly Journal of Experimental Psychology, 50A, 528-559.
- [92] Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. Discourse Processes, 26, 131-157
- [93] Nebes R. D. (1992). Semantic memory dysfunction in Alzheimers disease: Disruption of semantic knowledge or information processing limitation? In L.R. Squire & N. Butters (Eds.) Neuropsychology of memory (2nd ed. pp. 233240). New York: Guilford Press.
- [94] Nebes R. D. Semantic memory in Alzheimer's disease. Psychological Bulletin Vol 106(3) Nov 1989 377-394. doi: 10.1037/0033-2909.106.3.377
- [95] Nelson, D.L., McEvoy, C.L., & Schreiber, T.A. (1999). The University of South Florida word association norms. http://www.usf.edu/FreeAssociation.
- [96] Neural Nets and the Brain Model Problem, Ph.D. dissertation in Mathematics, Princeton, 1954.
- [97] Frank Rosenblatt (1962): Principles of neurodynamics; perceptrons and the theory of brain mechanisms. Washington: Spartan Books.
- [98] Newman, M. E. J., The structure and function of complex networks, SIAM Review 45, 167-256 (2003).
- [99] ONT 2012, Ontology: http://en.wikipedia.org/wiki/Ontology. Retrieved Feb 2012

- [100] Osborne, R. J. & Wittrock, M. C. (1983) Learning science: A generative process. Science Education, vol. 67, Issue 4, pp.489-508.
- [101] Osborne, R.J., & Wittrock, M.C. (1985). The generative learning model and its applications for science learning. Studies in Science Education, 12, 59-87.
- [102] Peled, Z., & Wittrock, M. C. (1990) Generated meanings in the comprehension of words problems in mathematics. Instructional Science, 19, 171-205.
- [103] Piaget, J. Logic and psychology. New York: Basic Books, 1957.
- [104] Port, Robert (2002) 'The dynamical systems hypothesis in cognitive science.' Entry for the MacMillan "Encyclopedia of Cognitive Science". L. Nadel, Associate Editor. Volume 1, pp. 1027-1032
- [105] Rayner, K., & Pollatsek, A. (1989). The psychology of reading. Englewood Cliffs, NJ: Prentice Hall.
- [106] Roger Schank, 1969, A conceptual dependency parser for natural language Proceedings of the 1969 conference on Computational linguistics, Sang-Saby, Sweden, pages 1-3
- [107] Roget, P.M. (1911). Rogets Thesaurus of English Words and Phrases (1911 edition). Available from Project Gutenberg, Illinois Benedictine College, Lisle, IL.
- [108] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. (1986) Nature, 323, 533–536.
- [109] Rumelhart, D.E., J.L. McClelland and the PDP Research Group (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations, Cambridge, MA: MIT Press
- [110] Salmon David P.; Butters Nelson; Chan Agnes S. The deterioration of semantic memory in Alzheimer's disease. Canadian Journal of Experimental Psychology Vol 53(1) Mar 1999 108-117.
- [111] Samuels, J., & Kamil, M. (1984). Models of the reading process. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), The handbook of reading research (pp. 185-224). New York: Longman.
- [112] Schnotz, W. (1993). On the relation between dual coding and mental models in graphics comprehension. Learning and Instruction, 3, 247-249.
- [113] Schnotz, W., & Grzondziel, H. (1996). Knowledge acquisition with static and animated pictures in computer-based learning. Paper presented at the Annual Meeting of the American Educational Research Association (AERA) in New York, April 8-12, 1996.
- [114] Silberstein, S. 1987. Let's take another look at reading: Twenty-five years of reading instruction. English Teaching Forum, 25.

- [115] Slobin, D.I. (1973). Cognitive prerequisites for the acquisition of grammar. In C.A. Ferguson and D.I. Slobin (Eds.), Studies of child language development. New York: Holt, Rinehart and Winston, pp. 173-208.
- [116] Smith, F. (1971). Understanding reading. New York: Holt, Rinehart & Winston.
- [117] Steyvers, M., & Tenenbaum, J. (2001). Small worlds in semantic networks. Unpublished manuscript
- [118] Steyvers, M.; Tenenbaum, J.B. (2005). "The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth". Cognitive Science 29 (1): 41-78.
- [119] Swaffar, J., Arens, K., & Byrnes, H. (1991). Reading for meaning: An integrated approach to language learning. Englewood Cliffs, NJ: Prentice Hall.
- [120] Tivarus ME, Ibinson JW, Hillier A, Schmalbrock P, Beversdorf DQ. An fMRI study of semantic priming: modulation of brain activity by varying semantic distances. Cognitive Behavioral Neurol. 2006 Dec; 19(4):194-201.
- [121] Winograd, T. Understanding Natural Language, Academic Press, 1972
- [122] van den Broek, P., Risden, K., Fletcher, R., & Thurlow, R. (1995) A 'landscape' view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In B. Britton & A. Graesser, Eds. Models of Understanding Text. Hillsdale, NJ: Erlbaum.
- [123] van Dijk, T., & Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic Press.
- [124] van Gelder, T. J. (1995) What might cognition be, if not computation? Journal of Philosophy, 91, 345-381.
- [125] van Gelder, Timothy, and Robert Port. (1995) It's about time: Overview of the dynamical approach to cognition. In Robert Port and Timothy van Gelder, editors, Mind as motion: Explorations in the dynamics of cognition. MITP, Cambridge, Mass. pp. 1-43.
- [126] Verhoeven, L., & Perfetti, C. (2008). Advances in text comprehension: Model, process and development. Applied Cognitive Psychology, 22, 293301.
- [127] Vygotsky, L. S. (1986). Thought and language Cambridge, MA: MIT Press. (Original work published 1934)
- [128] Watkins, M., & Gardiner, J. M. (1979). An appreciation of the generate recognize theory of recall. Journal of Verbal Learning and Verbal Behavior, 18, 687704.
- [129] Watts, D.J., & Strogatz, S.H. (1998). Collective dynamics of small-world networks. Nature, 393, 440-442.
- [130] Wittrock, M. C. (1992) Generative learning processes of the brain. Educational Psychologist, 27, 531-541.

- [131] Wittrock, M.C. (1974). A generative model for mathematics learning. Journal for Research in Mathematics Education, 5, 181-197.
- [132] Wittrock, M.C. (1974). Learning as a generative process. Educational Psychologist, 11, 87-95.
- [133] Wittrock, M.C. (1977) Generative processes of memory. In M. C. Wittrock, J. Beatty, J. E. Bogen, M. S. Gazzaniga, H. J. Krashen, R. D. Nebes, & T. J. Teyler (Eds.), The human brain (pp. 153-183). Englewood Cliffs, NJ: Prentice-Hall.
- [134] Wittrock, M.C. (1990). Generative Processes of Comprehension. Educational Psychologist, 24, 345-376, 1990. The Handbook of Research on Teaching, 3rd edition (Editor) Macmillan, 1986.
- [135] Wittrock, M.C., Marks, C. B., & Doctorow, M.J. (1975). Reading as a generative process. Journal of Educational Psychology, 67, 484-489.
- [136] Zhang, Y., "Solving Large-Scale Linear Programs by Interior-Point Methods Under the MATLAB Environment," Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD, Technical Report TR96-01, July, 1995.