A COMPARISON OF COMPREHENSIBILITY BETWEEN CONTROLLED TURKISH AND GRAPHICAL REPRESENTATIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

MÜGE CANPOLAT ŞAHİN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF
MASTER OF SCIENCE
IN
COGNITIVE SCIENCE PROGRAM

September 2013

A COMPARISON OF COMPREHENSIBILITY BETWEEN CONTROLLED TURKISH AND GRAPHICAL REPRESENTATIONS

Submitted by MÜGE CANPOLAT ŞAHİN in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science, Middle East Technical University** by,

Prof. Dr. Nazife Baykal ____________________
Director, **Informatics Institute, METU**

Prof. Dr. Cem Bozşahin ____________________
Head of Department, **Cognitive Science, METU**

Dr. Ceyhan Temürcü ____________________
Supervisor**, Cognitive Science, METU**

Examining Committee Members:

Prof. Dr. Cem Bozşahin ____________________
Cognitive Science, METU

Dr. Ceyhan Temürcü ____________________
Cognitive Science, METU

Assist. Prof. Dr. Murat Perit Çakır ____________________
Cognitive Science, METU

Assist. Prof. Dr. Cengiz Acartürk ____________________
Cognitive Science, METU

Assist. Prof. Dr. Bilal Kırkıcı ____________________
Foreign Language Education, METU

Date: 03.09.2013

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Müge CANPOLAT ŞAHİN

Signature:_____________________

ABSTRACT

A COMPARISON OF COMPREHENSIBILITY BETWEEN CONTROLLED TURKISH AND GRAPHICAL REPRESENTATIONS

Canpolat Şahin, Müge
M. Sc., Department of Cognitive Science
Supervisor: Dr. Ceyhan Temürcü
Co-supervisor: Assist. Prof. Cengiz Acartürk

September 2013, 112 pages

There are studies which show that visual representations, given in appropriate formats consistent with the context, are more informative and easy to understand as compared to textual representations (e.g., Simon & Larkin, 1980; Schnotz & Bannert, 2003; Carney & Levin, 2002).

However, the factors affecting the comprehensibility of visual and textual representations should be taken into account before coming up with this conclusion. The quality of a text is closely related to the linguistic abilities of its writer. This subjective factor makes it difficult to compare textual representations with graphical ones, which may yield unreliable results. To be more precise, depending on the linguistic abilities of the writer, a plain text may contain various redundant or ambiguous words or sentences which may confuse readers and lower the comprehensibility of the text. Although graphical representations generally code only core information, depending on the quality of the used notation they may also be ambiguous for its interpreters.

In order to make a comparison of comprehensibility between texts and graphical representations in a right manner, it must be ensured that the text to be compared does not contain ambiguous elements or unnecessarily complex sentence structures. Similarly, the graphical representation should also be eliminated from ambiguous structures.

This study aims to provide a comparison of comprehensibility between a text representing a task hierarchy and its graphical representation. To ensure the quality of the textual representation, the text was generated in Controlled Turkish. Similarly the quality of the graphical representation was ensured by conducting a preliminary experiment.

The results of this study have shown that although texts are given in Controlled Turkish, graphical representations are still better for ease of comprehension.

Keywords: Controlled Turkish, Controlled natural language, textual representation, graphical representation, comprehensibility.

ÖZ

KONTROLLÜ TÜRKÇENİN VE GRAFİKSEL GÖSTERİMLERİN ANLAŞILABİLİRLİK YÖNÜNDEN KARŞILAŞTIRILMASI

Canpolat Şahin, Müge

Yüksek Lisans, Bilişsel Bilimler Bölümü

Tez Danışmanı: Dr. Ceyhan Temürcü

İkinci Tez Danışmanı: Yrd. Dr. Cengiz Acartürk

Eylül 2013, 112 sayfa

Metinsel betimlemeler ile görsel betimlemeler kapsamında yapılan çalışmalar, kullanılan bağlama uygun olan görsel notasyonlar seçildiğinde, görsel betimlemelerin daha öğretici ve kolay anlaşılabilir olduğunu göstermiştir (Simon & Larkin, 1980; Schnotz & Bannert, 2003; Carney & Levin, 2002).

Metinsel betimlemelerin kalitesi, yazarın sözel kabiliyeti ile yakından ilişkilidir. Buna göre yazarın kabiliyeti ölçüsünde, çok sade bir metin bile belirsizlik barındıran kelime ve/veya cümleler içerebilir. Bu durum okuyucunun kafasının karışmasına ve okunan metnin anlaşılabilirliğinin azalmasına neden olur. Diğer taraftan, görsel bir betimlemenin kalitesi,ise betimlemede kullanılan notasyonlar ile ilişkilidir. Metinsel betimlemede olduğu gibi, görsel betimlemeler de belirsizlik barındırarak yorumlayan kişiyi ikileme düşürebilir.
Metinsel ve görsel betimlemelerin anlaşılabilirliğinin karşılaştırılması durumunda betimlemelerin kalitesi önemli rol oynamaktadır. Bu sebeple, doğru bir karşılaştırma yapabilmek için hem görsel hem de metinsel betimlemelerin kalitelerinin yeterli seviyede olması gerekmektedir. Bu

nedenle metinsel betimlemelerde belirsizlik yaratan kelime ve/veya cümlelerin; görsel betimlemelerde ise bağlam ile ilgisi olmayan ve/veya yanlış kullanılan notasyonların elenmesi gerekmektedir. Bu eleme işlemi yapılmadığı taktirde karşılaştırma sonucu ilgili betimlemeyi yapan kişinin yeteneğine göre değişeceğinden betimlemelerin karakteristiklerinin karşılaştırılması yönünde sağlıklı bir sonuç alınamaz. Bu yüzden, iki betimleme arasında yapılacak anlaşılabilirlik karşılastırması betimlemelerde bulunan uygunsuz ve belirsiz yapıların temizlenmesinden sonra gerçeklestirilmelidir.

Bu çalışma kapsamında, bir görev sırasını içeren görsel betimleme ile aynı görev sırasının verilmiş olduğu bir metinsel betimlemenin anlaşılabilirlik yönünden karşılaştırılması yapılacaktır.Metinsel betimlemenin kalitesini artırmak ve belirsiz yapıları elemek amacıyla metin Kontrollü Türkçe kullanılarak üretilmiştir. Aynı şekilde görsel betimlemenin kalitesinin artırılması için görsel betimleme deneye tabi tutulmuştur.

Çalısma sonucunda, metinsel betimlemelerin kalitesi Kontrollü Türkçe kullanılarak artırılsa da görsel betimlemelerin daha kolay anlaşılabildiği gözlemlenmistir.

Anahtar kelimeler: Kontrollü doğal diller, Kontrollü Türkçe, metinsel betimleme, görsel betimleme, anlaşılabilirlik.

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor, Dr. Ceyhan TEMÜRCÜ and my co-advisor Dr. Cengiz ACARTÜRK for their excellent guidance, patience, and valuable feedbacks for doing this research.

Finally, I would like to thank my husband and my parents for their patience and support.

TABLE OF CONTENTS

LIST of TABLES

LIST of FIGURES

CHAPTER 1

INTRODUCTION

Various studies show that if visual representations are given in an appropriate manner consistent with the context, they are more informative and easy to understand compared to textual representations (Simon & Larkin, 1980; Schnotz & Bannert, 2003; Carney & Levin, 2002). This is because visual representations can be less prone to redundancies, ambiguities and complexities as compared to linguistic ones.

However, there are factors that affect the comprehensibility of textual and visual representations. Comprehensibility of texts is dependent on the linguistic abilities of the writer and the comprehensibility of the visual representations is dependent on the notations used. Those dependencies may lead to unreliable results for the comparison. Thus, in order to make a fair comparison of comprehensibility between textual and visual representations, one option is to eliminate, as much as possible, ambiguous elements and unnecessarily complex structures from both and visual representations.

On the textual or linguistic side, this can be achieved by using a language defined by a small subset of vocabulary and grammatical rules instead of the full grammar of a natural language. At this point, controlled natural languages provide an appropriate solution. A controlled natural language is a subset of a natural language with a constrained set of grammatical rules and restricted lexicon. They are designed to generate sentences which are unambiguous and structurally simpler as compared to plain languages. Similarly, visual representations may involve notational conventions which may give rise to redundancies or ambiguities and may in turn reduce their comprehensibility. Thus, in a study that aims to compare comprehensibility aspects of textual and visual modes of communication, it is important to adopt a clear and unambiguous notation as much as possible.

In this study a comparison of comprehensibility between a text given in Controlled Turkish and its mapping graphical representation is made. In order to eliminate the ambiguities from the textual representation Controlled Turkish is chosen as textual representation. Similarly, in order to eliminate redundancies and ambiguities from the visual representation, the visual representation is subjected to an experiment and finalized according to experiment results.

To my knowledge, there has been no study which attempted to compare visual representations with the texts given in a controlled natural language. The primary motivation behind this study is to contribute to filling this gap in the literature.

1.1 Aim and Scope of the Study

The aim of this study is to compare comprehensibility aspects of texts presented in a Controlled Turkish with a semantically equivalent graphical representation. Textual representations and their graphical counterparts involve messages that express task hierarchies in a military planning domain.

Thus, the main research question is the following:

- Is the visual (graphical) representation of a text easier to understand than the text itself written in a simple controlled natural language?

According to the conducted literature survey, the initial prediction is as follows:

- The visual (graphical) representation is more comprehensible than the controlled natural language representation.

In the context of this study, a computer application was developed. This application was based on the “Collaborative Planning Model” (CPM), which is an ontology developed to support military planning. Visual notations in this study were the same notations used in CPM. The detail of CPM is explained in the literature review section.

The application developed in this study has two modules: One module is responsible for generating texts in Controlled Turkish. As part of this module a Controlled Turkish generator was developed, which yields sentences that express a military task hierarchy. The second module is a visualization tool which maps, or translates between Controlled Turkish statements and visual representations. This visualization tool was also used to compose texts, since writing in Controlled natural language in conformity with the defined rule set is a tedious task.

In the analytic part of the study, two different representations with same semantic content, texts in Controlled Turkish and their visual counterparts, were compared for comprehensibility. This comparison was made by means of an experiment. The stages of the study are summarized in the following diagram.

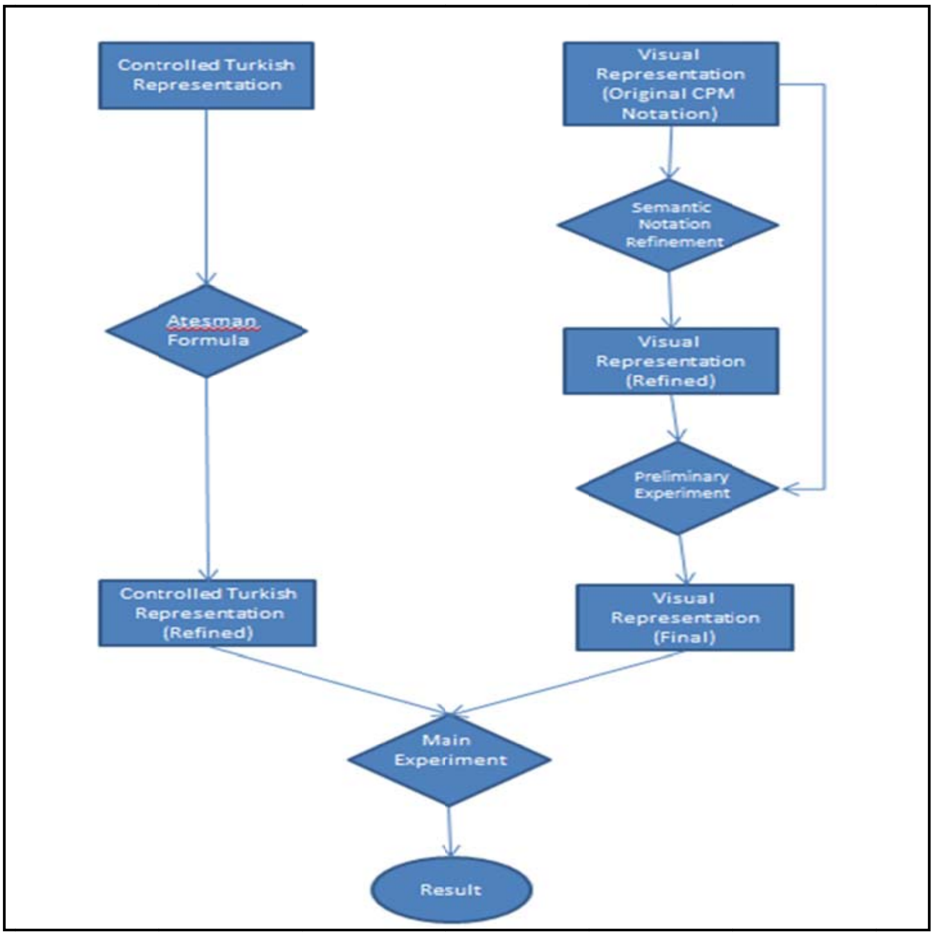


Figure 1 The Flow for the Study

In the preparatory stage of the experiment, the textual representation generated by the computer application was subjected to Ateşman's readability formula, in order to get clues about the ease of reading of the text. The rule set for Controlled Turkish was then refined until the statements are simple enough. The preparatory stage also involved refinements on the graphical notation. The graphical representation generated by the visualization tool was evaluated according to its semantic notations. The original graphical representation, which was in accordance with the original Collaborative Planning Model (CPM), was found to have ambiguities. Then, a modified visual notation was adopted, in which the problematic semantic notations were tried to be eliminated. In order to ensure that this modified notation was clearer and easier to understand, a preliminary experiment was carried out. In the experiment two different visual representations were compared for comprehension. One of the visual representations was designed according to original CPM notations and the other representation was a modified one where the semantic notations of the first visual representation that could lead to ambiguities were identified and eliminated. After the Controlled Turkish text and graphical representation were ensured to have no deficiencies, the main experiment was conducted to compare the two representations, namely the text in Controlled Turkish and the mapping visual representation, in terms of comprehensibility.

The specific question that is examined to find answer to the main research question is as follows:

- Does representing a task hierarchy in Controlled Turkish contribute to comprehensibility as compared to representing the same task hierarchy using visual elements?

1.2 Outline of the Thesis

In Chapter 2, the methods used in the study are explained. The chosen comprehensibility evaluation techniques and the rationale behind these choices are discussed. Chapter 3 is a literature review. It introduces controlled natural languages, mentions their different types, and discusses their advantages and disadvantages in various applications. This chapter also provides concepts related to comprehensibility in visual and textual notations, as well as methods used for the evaluation of comprehensibility. The last part of the literature review is a cognitive background for this study from different perspectives. In Chapter 4, the computer application developed for generating graphical representations and their Controlled Turkish counterparts is explained. The Collaborative Planning Model used in this application is explained in detail. The principles and rules used in developing Controlled Turkish are also given in this chapter. In Chapter 5, the preliminary studies and the experiment conducted for this thesis are explained in detail, and the results are exposed In Chapter 6, Discussion and

Conclusion, the results of the experiment are analyzed and examined in the light of information provided in the “Cognitive Background” subsection of literature review. Throughout Chapter 6, some prospects for future research are given and finally the findings of the study are summarized.

CHAPTER 2

METHODOLOGY

2.1 Overview

In this chapter the approach to find an answer to the research question is introduced. The experimental methods chosen for evaluating the comprehensibility measures of visual and textual notations are provided.

2.2 Approach

As already mentioned, the purpose of the study is to compare the comprehensibility of Controlled Turkish and a graphical notation, both of which are expressing the same task hierarchies. The semantic content involves a sample scenario which includes the tasks to be performed and their temporal relations with each other.

Before starting comparison, a Controlled Turkish text and a visual representation were produced for the task hierarchy contained in the scenario. The visual representation was generated using the visual editor of the developed software, details of which are given in Chapter 4. Controlled Turkish statements were also generated by this tool, by the application of mapping rules and heuristics as explained in Section 4.3.

One concern in developing a visual notation was the fact that notational and syntagmatic choices would have a direct effect on the comprehensibility. The primary principle for high quality visual representations is to use visual elements appropriate to the context (Schnotz & Bannert, 2003). In this study, CPM (Collaborative Planning Model), which is used in military task planning, was chosen as a starting point.

First, a simple scenario was developed which expressed a sample task hierarchy. A Controlled Turkish representation and a visual representation for this scenario, in conformity with the original CPM notation, are given below:

Table 1. Controlled Turkish Representation for a simple scenario (sample task hierarchy)

Adı "Görev_A" olan bir görev tanımlıdır. Adı "Görev_B" olan bir görev tanımlıdır. "Görev_B" görevi, "Görev_A" görevinin bitişinden sonra başlar. Adı "GÖREV_C" olan bir görev tanımlıdır. "Görev_C" görevi, "Görev_B" görevinin bitişinden sonra başlar. "Görev_C" görevi, "Görev_E" görevi ile aynı anda başlar. Adı "Görev_E" olan bir görev tanımlıdır. "Görev_E" görevi "Görev_C" görevi ile aynı anda başlar. Adı "Görev_D" olan bir görev tanımlıdır. "Görev_D" görevi Görev_E görevinin bitişinden sonra başlar. Adı "Görev_F" olan bir görev tanımlıdır. "Görev_F" görevi "Görev_D" görevinin bitişinden sonra başlar.

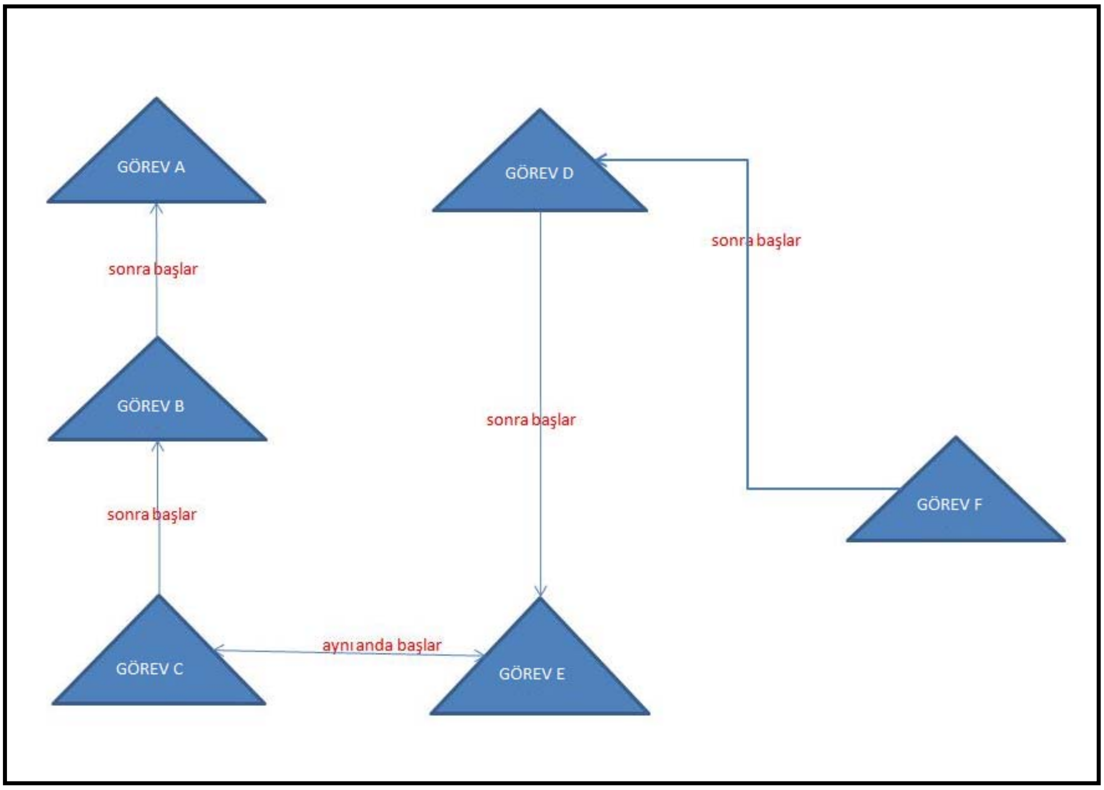


Figure 2. A CPM Representation for the Sample Task Hierarchy in Table 1

When examined, the visual representation shown above can be argued to involve ambiguities as to the temporal relationships between tasks: The arrowheads showing the direction of relations can lead to misinterpretations. Thus, there can be situations where information provided by the arrowheads can be interpreted as conflicting with information provided by the tags such as "sonra başlar" (starts after). Such situations may result in comprehension difficulties. In order to increase the quality of the notation used in this graphical representation, an undirected but fully tagged version was developed. This modified notation is shown in the figure below.

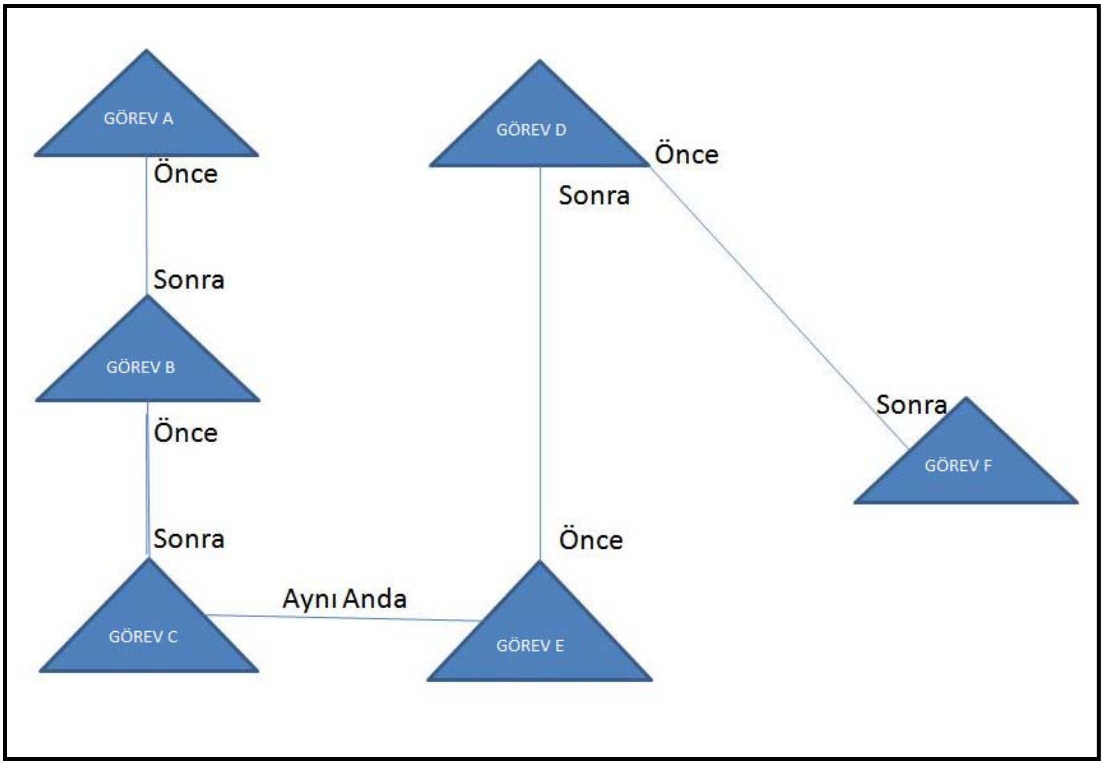


Figure 3. Improved CPM Representation for Sample Task Hierarchy

A preliminary experiment was conducted to evaluate comprehensibility measures of these two graphical notations. The aim was to determine which one notation would be more comprehensible. This preliminary experiment is explained in Chapter 5.

After the preliminary experiment, the visual representation that was performing the best was selected and this version became the final graphical notation to be used in the main experiment, which is then compared with its mapping Controlled Turkish representation.

The Controlled Turkish to be used in the main experiment was also pre-evaluated for ease of readability and comprehensibility. Ateşman's readability formula, the unique readability formula for Turkish, was applied to sentences in the first version of developed Controlled Turkish. Structures that lead to ambiguities were identified and the rule set was refined until the sentences could be categorized as “simple” in Ateşman’s readability scale. The details of this formula are provided in the Section 3.3.

After the simplicity of visual representation and Controlled Turkish sentences were ensured, a performance experiment was conducted for comparing these two modes of representation. In order to make a fair comparison, measurement techniques applicable to both visual and textual representations were chosen.
Among some alternative techniques, which are examined in the literature review, performance testing was chosen among reader focused methods, and comprehension testing was chosen among retrospective protocols.

For the comprehension test, a task was given to participants about a sample scenario expressed in a both graphical representation and a Controlled Turkish representation. The subjects were asked to sort the tasks in the scenario as to their temporal order. The success of the subjects in ordering of the tasks correctly was taken as the decisive criteria for comprehensibility.

The other technique used in the comparison experiment was performance testing: The time elapsed to finish the sorting task was used as the performance criteria, which in turn was taken as a measure for comprehensibility.

The details and the results of the experiments are presented in Chapter 5.

CHAPTER 3

LITERATURE REVIEW

3.1 Overview

The literature review in this study will be given in four sections. In Section 3.2, a brief explanation of Controlled Natural Languages (CNLs), types of CNLs, constructed CNLs in the literature, controlled natural language rule sets and the advantages and disadvantages of CNLs are provided. In Section 3.3, evaluation techniques for textual representations and visual representations are introduced separately. In the last section, the cognitive background for visual and textual comprehension is provided in some detail. Throughout this section, Kintsch and Van Dijk's theory for textual comprehension is briefly mentioned. Then, the cognitive perspectives of text comprehension and picture comprehension are explained in the light of the Schnotz's explanations.

3.2 Controlled Natural Languages

A Controlled Natural Language (CNL) is a subset of natural language, which is artificially formed by using a constrained set of syntactic and semantic rules of natural language. They have become more salient since the necessity for precise, unambiguous, fast readability and understandability arose. The machine readability asset of CNLs make them valuable in many fields such as aerospace, military, law and finance, especially in areas with intensive and ambiguous documentation. Controlled Natural Languages have recently become a popular area of study in Natural Language Processing and Computational Linguistics.

There are controlled natural languages constructed for English, German, Spanish, Swedish, Greek and Chinese (Namahn, 2001).

3.2.1 Types of Controlled Natural Languages

The controlled natural languages can be classified into two categories when their intent of use is considered: human-processable controlled natural languages and machine-processable controlled natural languages. Human-processable controlled natural languages are the ones that facilitate readability for humans. Machine-processable controlled natural languages are those that enable statements and texts to be interpreted by computers. Both of them have specialized lexicons and rule set checkers.

3.2.2 Controlled Natural Languages in the Literature

Several controlled natural languages are constructed for English for different purposes. Below are listed those for English:

- *AECMA Simplified English* is released in aerospace industry for maintenance manuals. It is categorized as "Human-Oriented Controlled Natural Language" since it is intended to form a common understanding for aerospace maintenance documentation (O'Brien, 2003).

- *Attempto Controlled English (ACE)* is developed by the University of Zurich as a research project. It contains a wide set of English rules and words and it is one of the most mature controlled natural languages. It is a general purpose controlled natural language and thus is not constrained to a specific domain.
  One of the most advantageous sides of ACE is its machine-processability. It provides facilities such as the translation of ACE texts into discourse-representation structures. This facility is intensively used in many realms

of computational linguistics. Another advantage of this controlled natural language is that it simplifies the job of writing statements which conform to its specifications, by providing user a writing editor. It has been referred to in the works of computational linguists such as John. F. Sowa (Attempto Project, 2011).

In ACE, there are two word classes. The first word class is the content words, which are user defined words. Lexical words, which can be entries into a lexicon, such as adjectives, adverbs, nouns and verbs, can be given as examples to content words. The second word class is the function words, which are predefined words. Determiners, prepositions and conjunctions are examples of function words.

- *Alcatel's COGRAM* is aimed to form a standard for technical documentation of course materials and systems in the realm of telecommunication. Its development started by investigating the inadequacies of present controlled natural languages such as AECMA and IBM's Easy Language. It continued by the formation of its own controlled grammar. It is currently used in grammar and style checking tools (Adriaens, 1992).

- *IBM's Easy English*, like AECMA, is designed for technical documentation. Like most of controlled natural languages it is machine-processable, and it is used in the automatic translation of documentation (O'Brien, 2003).

- *Oce's Controlled English* is designed for the automation of technical documentation and translation. It provided a significant progress in the reduction of technical documentation and translation duration (O'Brien, 2003).

- *Sun Microsystem's Controlled English* is the subset of English formed by Sun Software Company for the automation of technical documentation (O'Brien, 2003).

3.2.3 Controlled Natural Language Rule Sets

O'Brien (2003) concludes that the rules in a controlled natural language can be categorized into three groups: Lexical rules cover the situations that are related to word selection. They contain rules such as vocabulary usage, abbreviation/acronym usage, prefix/suffix usage, spelling, comparatives/superlatives, synonymy, polysemy, pronoun usage, anaphoric reference, quantifier usage, conjunction usage, negation and dictionary usage. The second category is the syntactic category which covers the rules of subject-verb agreement, modifier usage, adverb and adjective functionality, article usage, pronoun usage, preposition usage, particle, tense, person and number, voice, mood, modals and punctuation. The third category is the textual rules. Its subcategories, according to O'Brien, are related to sentence length, information load, sentence structure, paragraph structure and length, and word count. The last category is the pragmatic rules which contain subcategories such as textual devices such as idioms, metaphors, etc.

Many other rules can be specified for the construction of a controlled natural language but it is not possible to think of a controlled natural language supporting the whole set. Each controlled natural language is constructed according to a specific application domain.

3.2.4 Advantages and Disadvantages of Controlled Natural Languages

CNLs have significant advantages in the process of representing complex documents for both humans and machines (Namahn, 2011). Although CNLs exploit these advantages, they also bring some challenges. One of them is the difficulty in writing CNL statements by end users. End users may have problems writing controlled natural language since they may miss syntactic compliance. This problem of controlled natural languages is solved by the use of ontology editing tools or other visualizing tools.

3.3 Comprehensibility of Textual Representations and Visual Representations

In this section, the concept of comprehensibility for texts and visual representations is explained, evaluation techniques for measuring the readability and comprehensibility of texts are provided, comprehensibility concerns for visual representations are presented and methods for measuring comprehensibility for visual representations are explained.

3.3.1 Comprehensibility and Readability for Textual Representations

There are many definitions of readability and comprehensibility for texts in the literature. Starting from first decades of the 20^{th} century, there have been numerous studies on readability, comprehensibility and evaluation of text quality. There are various, sometimes conflicting opinions on the concepts of readability and comprehensibility and on how to measure them.

The concepts of readability and comprehensibility are intertwined, and many studies use them interchangeably (Adelberg and Razek, 1984). Because of the variety of studies and opinions about these concepts, it is important to come up with clear definitions for both concepts.

3.3.1.1 Readability

In general, readability is the property of a text that reflects how easily it can be read and understood. In fact, there are many different definitions for readability. For example, Klare defines readability as *"the ease of understanding or comprehension due to the style of writing" (DuBay, 2004,* *p.3).* This definition relies on the physical aspects of a text, such as the color, font size, etc.

However, McLaughlin defines readability *as "the degree to which a given class of people find certain reading matter compelling and comprehensible" (DuBay, 2004, p.3)*, and this definition refers to the performance of the readers**.** Chall and Dale come up with a more generic definition of readability, as covering many aspects such as **"***The sum total (including all the interactions) of all those elements within a given piece of printed material that affect the success that a group of readers have with it***"** *(DuBay, 2004, p.3).*

According to all these definitions, it can be proposed that the term readability has many aspects and the definitions may differ accordingly. So, while examining the readability concept one can handle it as a super-concept which includes various hidden aspects. Klare (1984) identifies three different aspects of readability. These aspects are legibility, ease of reading and the ease of understanding of the text.

Legibility is about the material properties of a text, such as font size, font style, color, etc. The second aspect, the ease of reading, is about the style of wording used to attract readers' attention and the topic of the text. The third aspect, the ease of understanding, is about the understandability of the text with respect to the sentence structure and the words it includes, which is closely related with the term comprehensibility.

3.3.1.2 Comprehensibility

Comprehension, in connection with reading, is a cognitive phenomenon which involves the perception, interpretation (assigning meanings) and understanding of a text. Thus, the term comprehensibility is related both with the properties of the text and the cognitive capabilities of readers.

Recalling from the previous section, comprehensibility can be seen as that aspects of readability which reflects properties of a text which affect readers' understanding. This idea is supported by Sutaria's (1965) study. According to Sutaria (1965), the characteristic which makes a text interesting and easy to read is the readability; on the other hand the characteristic that makes a text easy to understand for the target audience is the comprehensibility.

Even with these distinctions between the terms comprehensibility and readability, comprehensibility is still closely related with other aspects of readability defined in the previous section. In other words, a text cannot be comprehensible without being legible, or well styled. So, readability according to stylistic and linguistic perspective is a prerequisite for comprehensibility. In order to measure the comprehensibility of a text one should also measure its readability.

3.3.1.3 Evaluating Readability

In parallel with the readability definition of Klare (1984), readability should be evaluated both from a linguistic and a reader-focused perspective.

In her study, "Evaluating Text Quality: The Continuum from Text-Focused to Reader-Focused Methods", Schriver (1990) classifies evaluation techniques under three main groups: text-, expert- and reader-focused types of evaluation.

3.3.1.3.1 Text Focused Evaluation

In this evaluation technique, the reader is ignored. The primary subject of investigation is the text and its linguistic properties. In this technique, a text is examined according to predefined linguistic properties (such as the average word count per sentence, average sentence count per paragraph, line count per paragraph, etc.). The quality of the text is calculated, in terms of its level of readability, by .using readability formulas.

Readability Formulas

The most widely used readability formulas for English are SMOG, Gunning Fog, ARI, FRES (Flesh Reading Ease Score) and Flesch-Kincaid.

SMOG formula was defined by McLaughlin in 1969. The technique used to calculate the readability by SMOG formula is as follows:

Firstly, 10 sentence sequences are taken from the beginning, middle and last part of the text. Then, from each sequence, the numbers of polysyllabic words (words that have at least three syllables) are found. The formula, which is given below, is used to find the readability level of the text according to American Education System:

$$SMOG = 1.043 X \sqrt{[30x((\text{Polysyllabic Word Count}) \div (\text{Sentence Count}))]} + 3,1291$$

Another formula is the **Gunning-Fog readability index** (1952). This formula depends on word and sentence length. Results gained from this formula indicate the age group that the text is suitable for. This also gives a clue on the ease or the difficulty of the text. This formula has been used in many newspapers and journals, as it is easy to calculate.

$$\text{Gunning – Fog} = 0,4 \text{ X } [(\text{Word Count} \div \text{Sentence Count}) + 100 \text{ x } (\text{Polysyllabic Word Count} \div \text{Word Count})]$$

ARI-automatic readability index (1967) is used by the American Army in order to standardize technical documents. This formula, although different from other formulas, also considers the number of letters in an average-length word in the text.

ARI = 4,71 X [(Letter Count ÷ Word Count) +
(0,6 X (Word Count ÷ Sentence Count)) – 21,43]

Fresch Reading Ease Score (FRES) is the most valid readability score (Flesch, 1948). The formula is as follows.

FRES=206,835-[1,015 X (Word Count ÷ Sentence Count) –
[84,6 X (Syllable Count ÷ Word Count)]]

The value calculated from this formula is between 0 and 100. This score is an indication of the ease of readability of the text.

Table 2. Fresch score scale for readability (Bezirci & Yılmaz, 2010)

Fresch Score	Readability of text
0-30	Very difficult
30-40	Difficult
40-60	Somewhat difficult
60-70	Normal
70-80	Somewhat easy
80-90	Easy
90-100	Very easy

Flesch-Kincaid index (1948) makes use of Fresh reading ease score and combines it with the Kincaid's NRI (Navy Readability Index) index.

Flesch-Kincaid = [0,39 x (Word Count ÷ Sentence Count)]+
[1,18 x (Syllable Count ÷ Word Count) -15,59]

Ateşman's Readability Index is the only readability index developed for Turkish. It is published by Ateşman (1997). According to Ateşman, the average sentence length in Turkish is 9-10 word, and the average word length is 2,6 spell. The formula of Ateşman is as follows.

$$\text{Atesman} = 198{,}825 - [40{,}175 \times (\text{Sylabble Count} \div \text{Word Count})] - [2{,}610 \times (\text{Word Count} \div \text{Sentence Count})]$$

This formula, similar to Fresch's readability formula, again gives an indication on the ease or difficulty of reading a text, with a score between 0 and 100. The scale of ease or difficulty of a text according to this formula is as follows.

Table 3. Readability scale according to Ateşman (1997)

Ateşman Score	Text's readability
1-29	Very difficult
30-49	Difficult
50-69	Somewhat difficult
70-89	Easy
89-100	Very easy

Readability for Turkish

Compliance of English Readability Formulas to Turkish

Ateşman (1997) has carried out a study to investigate the compliance of English readability formulas to Turkish texts. His findings were as follows: SMOG value is not appropriate for Turkish since the formula is only dependent to multi-syllable words in the text. Turkish is an agglutinative language, in which words can be derived by the addition of suffixes to roots (Göksel & Kerslake, 2005). Hence, a message that is expressed by a sentence in some language can be expressed by a word in Turkish. Ateşman explains that widely used formulas, namely the Gunning Fog, ARI, and

Flesch Kincaid formulas, are almost same in the sense that they deal with the same features (average word and sentence length). The results obtained from these formulas reflect grade levels of US education system. When these formulas are adapted for Turkish, results were found to be very high. Then, some of the coefficients in the formulas were manipulated in order to obtain a formula appropriate for Turkish. However, there were still deficiencies in the formula since the results obtained from the formula were condensed in a specific range, thus not providing a differentiation as in English. In the case of Fresh Reading Ease Score formula some of the results were negative for Turkish and it was concluded that this formula is not suitable for Turkish either.

Ateşman's formula for Turkish was derived from the FRES formula. Books from different categories, such as Sigmund Freud's Totem and Taboo, Essays by Montaigne, Charles Dickens' Oliver Twist and Reşat Nuri Güntekin's Yaprak Dökümü were analyzed, and results were found to be plausible for Turkish.

Factors Effecting Readability in Turkish

According to Ateşman (1997), the main factors that play important role in readability in Turkish is the average sentence length and the frequency of syllables. In his study he explains these factors as follows:

- Higher word count decreases the readability and the comprehensibility of a sentence. This situation is not just valid for Turkish, it also valid for other languages. The average word count in Turkish sentence is 9-10 words.
- The average syllable count of a sentence in Turkish is about 2,6. When we analyze sentences in Turkish, they are mostly composed of one, two and three syllables. This structure makes readability easier (Bezirci & Yılmaz, 2011).

In a more broad perspective, readability in Turkish can be captured in three levels of analysis. In the first level, the features, sentence count, word count, syllable count, letter count, are analyzed. In the second level of analysis, number of the words with one syllable, two syllables, and three syllables, four and five syllables are identified. The third level of analysis concerns the average sentence length, word length and word frequencies according to number of syllables, average number of words according to syllables (Bezirci & Yılmaz, 2011).

Critique on Readability Formulas

Göpferich (2009) points out that although these formulas are easily applicable to computer programs, which makes them popular; they only use lexical and syntactical properties of the text which is only one of the aspects of readability. Therefore, these methods should not be the only evaluation strategies for readability since they do not tell much about the text comprehensibility from reader's perspective.

3.3.1.3.2 Expert Focused Evaluation

This evaluation technique relies on readers who have experience about the domain of a text. Schriver (1989) examines expert focused evaluation techniques in three classes: peer review, editorial reviews, and external reviews.

Peer review

In peer reviews experts sharing the same background assess text according to different properties such as consistency, style, etc. According to Schriver (1989), peer reviews are very useful especially determining text problems and stylistic issues. Furthermore, in peer reviews peers may suggest solutions for the detected problems about the text which may be useful.

However, there are some disadvantages of this methodology. In her study, Schriver (1989) emphasizes that the detected problems by the peers may not always be consistent with each other. Another disadvantage that Schriver (1989) points out is that, since experts are familiar with the topic, they may be insensitive to potential reader problems and may skip or underestimate issues that may be problematic for readers.

Editorial reviews

In editorial reviews, a text is reviewed by an editorial expert according to its style, grammar and consistency. Schriver (1989) believes that editors having enough experience and background about the domain are more successful in thinking from reader's perspective. Thus, different from writers, editors may be successful in detecting target audiences potential problems about the text. However, similar to domain experts, editors may also get insensitive to some potential reader problems and may miss some of them.

External reviews

This technique is used when insider experts are not successful in determining problems in a text or it is impractical to use insider experts. In such situations external experts are needed to review the text from a different and fresh perspective (Schriver, 1989).

3.3.1.3.3 Reader Focused Evaluation

Schriver (1989) defines reader focused evaluation as the procedure that relies on incoming feedback from the target audience and examines these procedures under two branches which are Concurrent and Retrospective Testing.

Concurrent Testing

As the word concurrent also describes, concurrent testing is about examining a reader's attitudes when s/he is reading a text. Schriver (1989, p. 247) defines concurrent testing as "evaluating problem-solving behaviors of readers while they are actively engaged in comprehending and using the text for its intended purpose"*.* According to her study, concurrent testing methodology includes Cloze Testing, Behaviour Protocols, Eye Movement Protocols, Performance Testing and Thinking Aloud Protocols as testing procedures.

Cloze Testing

The idea behind a cloze test is to collect data directly from the reader in order to determine how easy it is for the reader to understand the subject text. In order to achieve this, some of the words in the subject texts are deleted according to a procedure referred to as deletion procedure. Then, readers are asked to fill against the deleted words and their answers are collected to determine the difficulty of the text.

According to Bormuth (1968), the main idea in constructing a cloze test is to delete every fifth word and replace it with a blank character having the same size. The subjects that will attend cloze tests should have never seen the text before. The subjects then requested to read the text and fill the blanks. When subjects fill the blank with the exact deleted word then it is counted as a score. Then, the percentages of the participants' correct responses are calculated. The difficulty of the text.is than calculated as the mean of these percentages.

There are many different adaptations of cloze tests which vary according to different deletion techniques. Although in the original test every fifth word is deleted, this may be changed to deletion of every *n*th word or deletion of some type of words (nouns, verbs etc…) in order to eliminate extreme cases.

Steinman (2002) distinguishes cloze tests as Random and Rational according to the deletion rates as described above. A random cloze test deletes every *n*th word where a rational cloze test deletes words according to a linguistic principle, such as nouns, verbs, adjectives, etc.

Schriver (1989) also argues that the validity of cloze tests varies depending on the nature of the subject text. According to her, cloze tests give best results on narrative and expository texts, while they perform poorly on reference and procedural texts.

Behaviour Protocols

In behavior protocols readers are asked to complete a task. The main aim of behavior protocols is to observe readers while they are trying to accomplish a task. This observation may or may not be computer aided. One key rule in behavior protocols is that readers are not allowed to talk aloud (Schriver, 1989).

Eye Movement Protocols

An eye movement protocol is a subset of behavior protocols where readers' eye movements are examined while reading the subject text. Readers' eye movements are interpreted to come up with a conclusion about the text readability and comprehensibility. By using eye tracking systems one can detect eye fixations and gaze times, both of which may give important clues about the readability and comprehensibility of a text.

Performance Testing

According to Schriver (1989, p.248) performance testing includes the evaluation of "factors such as readers' task performance, retrieval and access behaviors, error recovery strategies, cognitive load, and general ability to use a text". In general, criteria for performance testing are speed and accuracy. Schriver (1989) suggests preventing talking or thinking aloud which may affect these performance criteria measures.

Thinking Aloud Protocols

Thinking aloud protocols are somewhat similar to behavioral protocols. Both protocols are designed such that first a task is assigned to the reader and then observers collect data while readers perform their tasks. According to Schriver (1989), different from behavioral protocols, participants in thinking aloud protocols are asked to say anything that comes to their minds while performing their tasks. Schriver (1989) also proposes that, thinking aloud (especially while having difficulty) usually provides useful comments and feedbacks about the type and the location of problems that cause difficulties for the reader.

According to Schriver (1989), one of the main advantages of this protocol is that they yield feedbacks which are more useful than those taken after readers complete their tasks.

Retrospective Testing

Schriver (1989) classifies retrospective testing under reader focused protocols. The main aim in retrospective testing is to collect data about the comprehensibility of the text after readers finish reading. Feedback from the user can be collected in various ways containing interviews, focus groups, critical incidents and reader feedback cards.

Although they are easy to conduct, there are some disadvantages of retrospective testing protocols. Those are explained by Ericson and Simon (1993). According to them, while conducting retrospective memory tests readers generally retrieve information from their memories. According to the reader skills, they may use their long term memory (LTM) or short term memory (STM). The problem arises when readers refer to their LTM that lead to errors and incompleteness proved to be probable in experiments and studies on memory.

Schriver (1989) further classifies retrospective testing strategies under Comprehension Testing, Surveys and Interviews and Critical Incident Methods.

Comprehension Testing

Comprehensibility is closely related with the reader’s understanding of the text. So, the main aim of comprehension testing is to measure the degree of understanding of the readers of a text. It is a retrospective protocol, which means that it is conducted after readers finish their tasks. In order to measure the reader's understanding of the text different techniques are used. Schriver (1989) reports that these techniques include asking users to paraphrase or summarize the text, soliciting inferences about the text, and using multiple choice, space filling or true/false tests.

At first glance, the easiest method seems to conduct multiple choice or true/false tests to measure understanding. However, the nature of the tasks and questions may seriously affect the quality of the test. Thus, in order to ensure that the prepared questions are feasible to measure comprehensibility, they should be validated by applying time consuming validation tests. The other alternative is to conduct summarizing of paraphrasing sessions with the reader but that also requires time and can be subjective (Guillemette, 1987).

Because of unreliable affects, such comprehension testing techniques are criticized by many studies, including Schriver (1989).

Surveys and Interviews

Similar to other comprehension testing procedures, surveys and reviews are also used to measure readers' understanding of a text.

In surveys and interviews readers are subjected to open- or closed-ended questions in order to find out readers' opinions about the stylistic and wording features of a text, which affect the comprehensibility level of the text. Like other techniques, surveys and interviews also come with their advantages and disadvantages.

Schriver (1989) describes the advantages and disadvantages of surveys and interviews. These advantages and disadvantages are summarized as in the table below.

Table .4.Pros and Cons of Surveys and Interviews (Schriver, 1989)

Technique Type	Advantages	Disadvantages
Surveys	• inexpensive • not time consuming	• Participants generally ignore open ended questions. • Self-selected participants may lead to biased results. • Low rates require further testing to determine the source of problem in the text.
Interviews	• Rich data source to determine the problem in the text. • Participants feel more comfortable while answering	• time consuming • difficult to analyze the results

Critical Incidents Method

The aim of this technique is to collect the reader’s positive and negative opinions about the text. In this technique readers are asked to remember the critical aspects of the text and afterwards these positive or negative memories of readers are collected.

According to Schriver (1989), the major disadvantage of this technique is the memory load brought on reader, which may usually lead to unreliable results.

3.3.2 Comprehensibility of Visual Representations

Visual representations have to be designed according to comprehensibility and communicative purposes. That is, the semantic notations should be appropriate for the context and should not give rise to comprehensibility conflicts and ambiguities. If visual representations are given in a right manner they facilitate communication and comprehension.

The comprehensibility of a graphical representation is measured in terms of the degree of ease and rapidity that a user can make inferences from it. Below are provided main methods used to evaluate the comprehensibility of visual representations.

3.3.2.1 Verbal Protocol Method

One of the methods for collecting information about visual representations is the verbal protocol method. Simon and Ericsson (1980) identify two different classifications of verbal reports. The first classification depends on the form of stored information, while the other depends on the probing technique:

Classification according to form of stored information:

- In Level 1 protocol analysis, participants talk while they are processing the given information in their short term memory.
- In Level 2 protocol analysis, the information provided to participants is not in the verbal form. Information can be an image or a diagram. Participants are asked to verbalize the given information.
- In Level 3 protocol analysis, participants do not just tell their thoughts; they are also expected to process the information according to a given

criterion. In this type of verbal protocols, participants verbalize the information according to given criteria or make some inferences out of the given information.

Classification according to probing technique:

1. In think-aloud protocol, subjects are asked to talk about their thoughts while processing the information.

2. In the second class of probing technique, subjects are asked to complete their tasks without talking, but at the same time a controller probes for specific information.

3. In the third class of probing technique, a controller probes for specific information after participants complete their tasks.

In verbal protocol methodology, participants' voice records are transcribed into verbal reports for analysis. Ericsson and Simon (1980) propose to analyze the data in a sequence of steps. These steps begin with tape recording. In the next step, this raw data has to be processed and a written transcript has to be prepared. Then, a segmentation process is applied. That is, the participant's verbalization is divided into segments according to a criterion. Ericsson and Simon (1980) refer to this step as preprocessing step.

At the next step, generated segments are classified according to a criterion. The assessment for the classification of the segments is up to the controller's judgment. Ericsson and Simon (1980, p.5) propose that "*If each of the segments is to be treated as an independent datum, then the encoding of that segment must be made on the basis of the information contained in it, independently of the surrounding segments*."

After the preprocessing step, the categorization of the segments takes place. Since the segmentation is made so that each segment contains single information, the categorization idea of Ericsson and Simon's seems very suitable in this case. The idea is to encode segments according to the information contained within them and to encode them separately from the neighboring segments where each segment contains single information (Ericsson & Simon, 1980, p.5).

3.3.2.2 Performance Criteria

Another method for the evaluation of graphical representations is the use of performance criteria. Efficiency of a graphical representation can be evaluated according to how easily and quickly a participant can infer the expected information. In order to measure this, time to complete the task is used as a performance criterion.

3.3.2.3 Eye Tracking

Eye tracking method is used to compare the convenience and comprehensibility aspects of graphical representations. In this method, fixation counts and fixation durations are measured. These metrics are briefly explained below:

- Fixation duration is the duration of each fixation in an area of interest.
- Fixation count is number of times a participant fixates in an area of interest.

These two metrics give important feedback on the comprehension of graphical representations.

3.4 Cognitive Background

3.4.1 Models for Text and Picture Comprehension (Schnotz, 2009)

According to Schnotz (2009), people construct mental representations both while understanding a textual sample and a pictorial sample. There are three levels of mental representations involved in understanding a text. The levels are as follows:

- At the first level, the reader reads the sentence and forms a mental representation which is in the form of text-surface structure. This text-surface structure is still not related to the comprehension of the text.
- After the formation of this text-surface structure, the reader forms a propositional structure of the text. This level is related to the conceptualization of the text and it is distinct from the syntax of the sentence.
- At the third level, the reader forms the mental representation of the sentence content.

While understanding a depictive description a learner may create many mental representations for the depictive description other than the original one. S/he first constructs a perceptual representation of what s/he has seen in the depictive description and by using that perceptual model, creates different mental models for the same depictive description. As an example, when a learner understands a pie chart representation, s/he can construct different representations (bar chart etc.) for the same information.

As it is the case in representations, the same depictive and descriptive distinction also applies to mental representations. According to this, text surface and propositional representations fall into descriptive representation category while perceptual models and visual images fall into depictive representation category (Kosslyn, 1994).

Physical representations (textual and visual) and mental representations differ according to whether they are sensory-specific or not. Physical representations are sensory-specific since they are closely related with the visual model while mental representations are not, since they can be constructed by using information coming from various sources such as touch, hearing etc. Another area of distinction is the information content. According to this, mental models may discard unnecessary information contained in the visual image or they may contain some other information coming from the prior knowledge that is not involved in the visual picture.

According to the distinctions between descriptive and descriptive representations, Schotz and Bannert (2003) constructed a framework for textual and visual comprehension. The framework is depicted in the figure below:

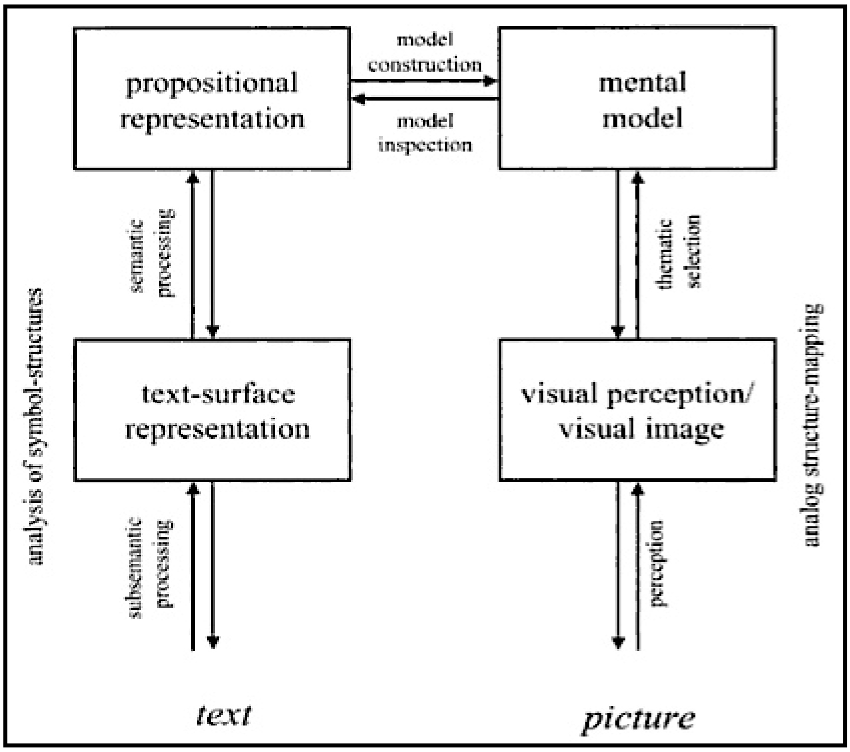


Figure 4. Two representational channels in text and picture comprehension

Schnotz and Bannert (2003), by considering the differences between depictive and descriptive representations, construct a theoretical framework which is provided in the figure above. Using this framework they analysed text and picture comprehension. As it can be seen from the figure above there are two main channels: One of the channels is based on textual comprehension and it is called verbal channel. The flow of process in this channel starts with the external text. After the semantic processing on this text, a text surface representation is constructed. Further process on this representation results in propositional representation. Finally, from this propositional representation, a mental model is constructed. In the verbal channel, information is processed by symbols. The other channel is the pictorial channel. The flow of processing in the pictorial channel, like in the verbal channel, starts with the external picture representation. In contrast to verbal channel, the external picture representation is processed by structure mapping. As a result of this process visual perception of the external image is realized, then the mental model for the image is constructed.

3.4.1.1 Multiple Memory Systems in Picture Comprehension

For textual and picture comprehension, people construct different mental representations with different cognitive architecture. Atkinson and Shiffrin (1971) propose that these cognitive architectures include different memory systems, including working memory, sensory registers and long-term memory.

People perceive external information via various sensory means and store these information in sensory registers.. There are many different channels (through eyes, touching etc.) for information gathering. Visual channel is one of the most important channel in order to gather information and is also the main sensory register. Visual registers can store the transferred information less than one second. Information that is stored in the visual register can be transmitted to the visual working memory if and only if the necessary attention is given by the subject.

Working memory is the place that is used to form a mental model for the transferred information. Baddeley (1986) proposes that working memory contains an execution center as well as subsystems for information storage. Limitations for descriptive and depictive channels are actually storage limitations of these subsystems.

One of these subsystems is the visual working memory. It deals with spatial information perceived through visual modalities and has a limit of five units at a time.

According to (Freidman and Miyake, 2000), there is another type of working memory that is responsible for the construction of mental models, which is heavily influenced by the capacity of visual working memory. So one can conclude that the visual comprehensibility is closely related with the capacity of the working memory (Daneman and Carpenter, 1983).
Long term memory is the place that subject stores the comprehended information. This information is then used as prior knowledge in order to ease the comprehension of other information. So, the textual and visual comprehension is closely related with prior knowledge.

The syntactic competence of readers enables the construction of a mental text-surface representation. Similarly, in picture comprehension, prior knowledge of conventions facilitates the perception of the external picture. Prior knowledge of the use of symbols is a precondition for constructing propositional and mental models. Prior knowledge in other domains can also compensate for deficiencies or lack of information in the external source. Carney and Lavin (2002) propose that pictures are analysed more intensively when the subject's prior knowledge is poor.

3.4.1.2 Integrated comprehension of text and pictures

Many different processes have been shown to be at work in textual and visual comprehension. In order to explain how these processes interact, an integrative model was proposed (Schnotz, 2009). This model both defines both the textual and depictive comprehension. The figure below describes this integrated model.

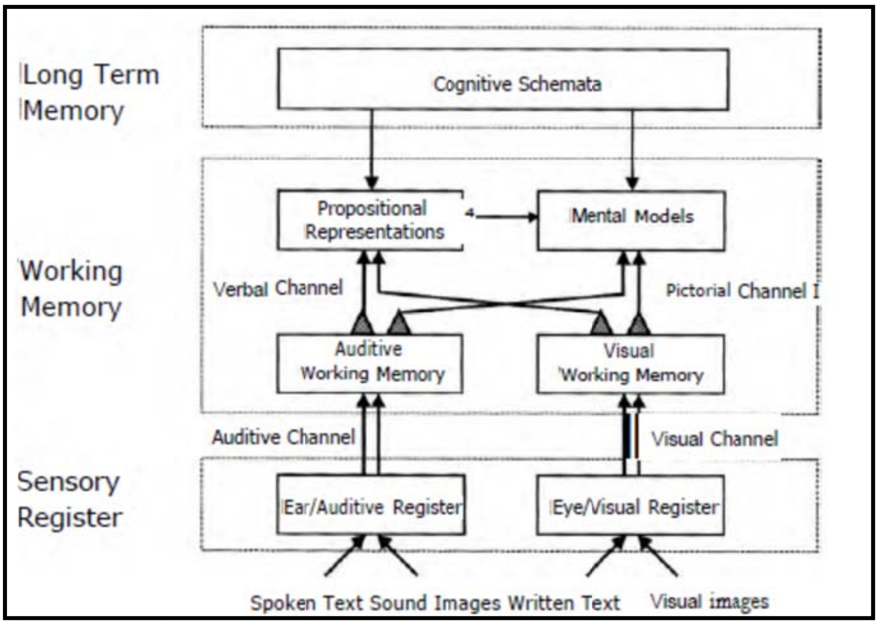


Figure 5 Integrative model of text and picture comprehension (Schnotz, 2009, p. 57)

The cognitive architecture defined by the figure above consists of three main components, namely the sensory registers, limited working memories and long term memory. According to the figure, textual and picture comprehension is achieved by these interacting components.

There appear two main channels that are responsible to transfer external information to the working memory. These channels are the auditive and visual channels. The external verbal information perceived by auditive registers is transferred through the auditive channel whereas the external visual pictures perceived by eye/visual sensory registers are transferred through visual channels.

The verbal information (both spoken and written) transferred into working memory is processed in the verbal channel and the pictorial information in the working memory is processed in the pictorial channel. As previously mentioned, these channels have limited processing and storing capacities.

Lastly, as mentioned before, prior information plays an important role in text and picture comprehension. The long term memory component takes place in the model where the already acquired information is stored and this information influences text and picture comprehension positively.

The model can be divided into two sections according to the type of processing. These two sections are the perception section and the cognitive section. In the perceptual section the external information is transferred into the working memory through sensory registers and sensory channels. The cognitive section is the section where long term memory and the working memory interact and the information in the working memory is processed to construct a mental model.

In the current study only reading and picture comprehension is being studied.

3.4.1.2.1 Reading comprehension

According to
Figure 5, reading comprehension starts with the perception of the external visual verbal information through eyes. Then this perceived external information is transferred to the visual working memory through visual channel. Next, information is further processed and a text-surface representation is constructed. By the help of verbal filter, the verbal information in the text-surface representation is picked and sent to the propositional working memory through verbal channel. By this way, a propositional representation is formed. The construction of the propositional representation then results in construction of mental model.

3.4.1.2.2 Visual Picture Comprehension

Similar to reading comprehension, picture comprehension also starts through eyes where the external pictorial information is perceived. This perceived information is then transferred to the visual working memory through visual channel. Then, a pictorial filter picks up the pictorial information and sends it to the pictorial channel in order to form a mental model.

For comprehension to take place, a set of cognitive processes should occur. These processes interact with each other. To be more precise, information selection, organizing selected information, using prior knowledge and integration of information from different resources are the most important processes that lead to comprehension.

3.4.1.3 Cognitive Theory of Multimedia Learning (CTML)

Another theory on text and picture comprehension is the Mayer's CTML. The figure below depicts the CTML model.

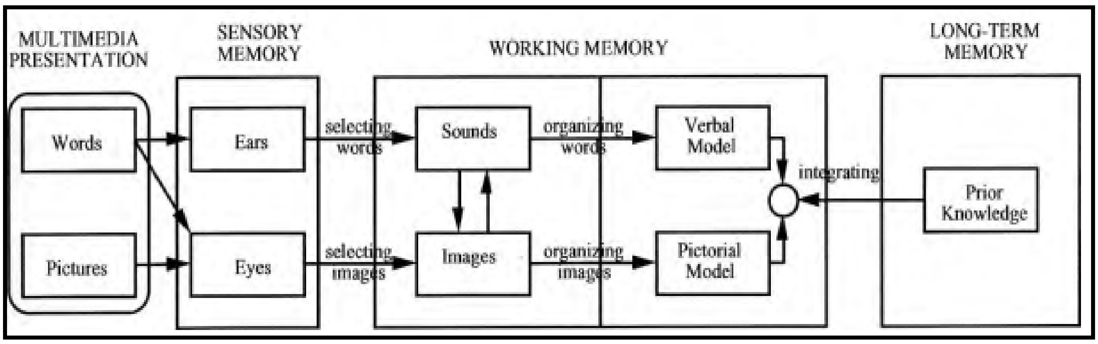


Figure 6 Cognitive Theory of Multimedia Learning Model (Mayer, 2009, p.37)

According to the picture there are three types of memories used in multimedia learning. These are the sensory memory, working memory and the long-term memory. The sensory memory is the memory that stores the text and picture coming from the outside (Represented as Multimedia Presentation Box). The pictures and texts are registered to the sensory memory through the ears and eyes. The sensory memory can keep these transferred information as exact visual images for a very short amount of time.

The actual work takes place in the working memory. Working memory is used to store the information and process this information in order to construct verbal or pictorial models. According to Mayer (2009), the processes taking place in the working memory involves conscious awareness. That is, when reading texts one may give more importance to some words or when looking to an image one may give more attention to some parts of the image. The arrows for “selecting words”, “selecting images”, “organizing words” and “organizing images” depicts the processes takes place in the working memory. The “selecting words” arrow means that the person focuses on significant words in order to construct the sound in his/her working memory. The “selecting images” arrow similarly means that the person focuses on significant pictures or words in order to construct the corresponding image in his/her working memory. The “organizing words” arrow means that, the person relates the selected words in his working memory to create the verbal model. The “organizing images” arrow means that the person relates the selected images and from these relations s/he creates the pictorial model.

The right side of the figure above depicts the long-term memory. Different from working memory, long-term memory can store more information for long time. One can think long term memory as a storage for the previously learnt knowledge. To use the knowledge stored in the long term memory, the knowledge has to be transferred into the working memory. This is indicated by the “integrating” arrow in the figure above.

3.4.1.4 Comparison between ITCP and CTML Models

There are many similarities between the CTML and ITPC models. The architectures proposed by these two models are similar in the sense that they both rely on the multiple memory types. They propose different channels to transfer external information and processing according to the type of sensory registers. One of the different aspects among these two models is that, CTML proposes two different sensory modalities to perceive external information namely the visual and verbal channels. On the other hand, ITPC model proposes more sensory channels in order to perceive external information such as visual, auditory, touch and others. Another difference is that, CTML proposes the production of verbal mental model and the pictorial mental model and these two models have to be integrated. On the contrary, the ITPC model claims that only one mental model is constructed and all the necessary information has to be integrated before to construct the mental model.

3.4.1.5 Empirical Evidence

In this part the efficiency of the ITPC model is discussed and positive and negative effects of using (i) only pictures, (ii) only text or (ii) both are explained.

3.4.1.6 Positive Effects of Combining Texts and Picture

3.4.1.6.1 Coherence and Contiguity

The ITPC model proposes that learning gets easier when the text and the picture are related to each other. This phenomenon is known as the coherence condition. The coherence condition is established when the text and the picture belong to the same context. Since the decay time of the information in working memory is too short, ITPC model further proposes that learning occurs easily if both the pictorial and textual mental models are simultaneously in the working memory, which is known as the contiguity condition. In other words, contiguity condition is established when the text and the picture are physically close to each other.

According to the ITPC model, if a picture is given with a spoken text, the learning occurs easier than the picture is given with a written text. The reason for this behaviour is that, in the picture-spoken text couple, pictorial information is transferred through the visual channel and the spoken text information is transferred via auditory channel simultaneously. However, in the picture-written text couple, both the pictorial and the textual information are transferred through the visual channel, which triggers the eye to switch focus between the picture and the text. This is called split of attention, which requires a visual search through picture and text. This behaviour also emphasizes the importance of the contiguity condition.

3.4.1.6.2 Modality

As already mentioned, comprehension occurs more easily when the picture is given with a spoken text instead of written text. This is somewhat expected since the split of attention does not occur when the picture is given with a spoken text as described in the previous section.

However, according to ITPC model, modality effects cover more than the elimination of split of attention. In order to prove this, Moreno and Mayer (1999) conducted an experiment. In order to remove the split of attention, volunteers were first subjected to a picture and a written text consecutively, and then to a picture and a spoken text consecutively. The result was that even when the split of attention effect was eliminated, the picture-spoken text pair resulted in easier learning. ITPC model explains this result with reference to working memory capacity. Even when the written text and the picture are supplied consecutively, some part of the visual working memory is used to store the external information and the rest is used to process the external information which results in weak learning behaviour.

3.4.1.6.3 Sequencing

Sometimes there is no chance to give picture and the text close together and contiguity condition cannot be established. In this case, giving the picture before the related text is more effective. The reason to this is that with only textual information, it is hard to establish the mental model. However, the mental model constructed from pictorial information is closer to the mental model. So, if the picture information is presented after the text, the mental model established by reading the text and the mental model constructed after viewing the picture may differ and conflicts may occur. One can easily eliminate these conflicts by giving the pictorial representation beforehand.

3.4.1.6.4 Readability and Prior Knowledge

According to ITPC model, the learning process uses different sources of information including text, picture and prior knowledge. When somehow one of these sources is weak or cannot be used, the other strong or usable sources become more important. To give an example, it is relatively hard to construct a mental model only from textual information when subjects have no prior knowledge about the topic. So, including a related picture will introduce another information channel which will highly increase the success of learning. However, subjects having prior knowledge can still construct a mental model only from textual information and may not need a picture. According to this, ITPC model proposes that subjects with no prior knowledge exploit pictorial information more than subjects with prior knowledge.

3.4.1.7 Negative Effects of Combining Texts and Pictures

3.4.1.7.1 Redundancy (Specific)

Sometimes texts are given with both written and spoken form. The idea behind this is to satisfy subject needs and let s/he choose which one to focus. But ITPC model proposes that the comprehensibility of a picture given with both spoken and written text is worse than a picture given only with spoken text. There are two main reasons for this: One of them is that people cannot fully focus on the spoken text when it is presented simultaneously with the written text. This will cause a split of attention which decreases comprehensibility. Another reason is due to the problem of synchronization between the agent's processing speed and the speed of the auditory text. This situation especially occurs if the subject is a good reader and reads faster than the spoken text, which causes conflicts. This conflict then results in a decrease in comprehension, which is also called the specific redundancy effect.

3.4.1.7.2 Redundancy (General)

The dual-coding theory of Paivio (1986) proposes that texts supplied with pictures are always better than texts alone, since in the former two different external information sources are transferred into the working memory. However, the ITPC model proposes that textual information supplied with picture leads to better comprehension only in some circumstances. For instance, when subjects are experienced readers they can easily construct a mental model from the text alone and do not need further pictorial information. When such subjects are given texts that are combined with pictures, additional pictorial information does not lead to better understanding nor a more detailed mental model. Furthermore, this type of redundancy causes a split of attention between the text and the picture, which in turn causes loss in time and working memory capacity.

3.4.1.7.3 Structure Mapping

ITPC model proposes that a text supplied with a picture leads to better comprehensibility if the picture is given in an appropriate structure for the related concept. The rationale behind this is that in the mental model construction process, pictorial channel works by structure mapping. So, the structure of the supplied picture affects understanding and thus comprehensibility. Schnotz and Bannert (2003) also support this proposal of ITPC. They provided subjects two different sets of text and picture information. Textual information and information contained in the pictures were the same in both sets, but pictures differed only in their structure. According to the results, ease of learning increases when the form of the pictorial information is appropriate to the concept. To give an example, diamond shapes are used to indicate decision points in flow diagrams. If the context of the pictorial information is different than a flow, using a diamond shape to denote another type of visual element may conflict with the subjects prior knowledge about flow diagrams. Then subject has to resolve this confliction which latens the visual comprehension.

3.4.1.7.4 Deep versus Superficial Processing

Another prediction of the ITPC model is that, if a text and a picture are simultaneously given, the addition of a picture makes textual information less important. This results in superficial processing of the text, which implies less working memory capacity usage than in the case of only textual information.

3.4.1.7.5 Cognitive Economy

The ITPC model also provides a method for cognitive cost analysis. As mentioned before, the existence of multiple external information sources generally increase comprehensibility. But each different source bring a cognitive cost for transfering and analyzing information. If the cognitive cost of the additional information source is more than the gained comprehensibility benefit, then the subject will not focus on this additional information source and opt to skip it.

3.4.1.8 Instructional Implications

While preparing visual and textual representations, ITPC and CTML frameworks' deductions on textual and picture comprehension should be taken into account. To sum up, both frameworks make the following suggestions as rules of thumb.

- It is worth to use textual information combined with picture when subjects have no or less prior knowledge about the subject and they have no perception deficiencies to process text and picture simultaneously.
- When a picture is given with a written text, the picture and the text should be close to each other (spatial contiguity).
- When a picture is given with a spoken text, picture and the text should be presented simultaneously (temporal contiguity).
- If animated pictures are used, they should be given simultaneously with spoken texts (Modality).
- Do not supply pictures with both written and spoken texts (Specific Redundancy).
- Pictures should be appropriate to the concept and irrelevant structures should not be used (Coherence).

ITPC model goes further and makes the following additional suggestions:

- Pictures should be presented before the related texts if it is not possible to provide them simultaneously (Sequencing principle).
- The structure of the pictorial information should conform to the context (Structure Mapping).

- If subjects have prior knowledge and experience about the context then it is not appropriate to use both text and the picture simultaneously (General redundancy).

The modality effect implies the advantage of spoken text over written text. However, the ITPC model proposes that this is not the case if the learning time is not limited. This is assumed to be due to the effect of subject's control in written and spoken text. Subjects have no control over the pace of spoken texts but have much more control over static written texts. This control is especially important if the text is difficult. So, a subject may choose to reread the text or move forward or backwards in the text.

Thus, the last additional rule proposed by the ITPC is:

- If the text is difficult to understand and there is no time limitation for learning, then pictures should be given with written texts instead of spoken texts (Control-of-processing principle).

3.4.2 Kintsch and Van Dijk’s Model of Text Comprehension

An influential theory of textual comprehension is the Kintsch and Van Dijk’s (1983) theory. This theory is an integrated theory which involve stages starting from recognition of words to construction of a representation for the meaning of the text. In the core of its assumptions lies the idea that three different mental representations are constructed during text comprehension. One of the mental representation is the surface representation of the text where each word in the text is parsed seperately. The second representation is called the textbase where a network of propositions is constructed (Kintsch, 1998). In this network, the nodes are connected to each other if they share a structural feature such as two propositions sharing an argument (Frank et. al, 2007) or there exists a connective such as “and”, “but”, “because”, etc. The last level of the mental representation is the combination of the elements in the text with the elements in reader's knowledge. This is called situational model. In this situational model, the relations does not have dependency on structural features. The relations in the situational model which Kintsch and Van Dijk call as "facts", depend on the the effect the items have on another probability of occuring. This model has been taken as a starting point for other models of text comprehension (Frank et. al, 2007).

CHAPTER 4

A TOOL FOR GENERATING GRAPHICAL REPRESENTATIONS AND CORRESPONDING CONTROLLED TURKISH STATEMENTS

4.1 The Collaborative Planning Model (CPM)

Military is one of the areas where CNLs proved to be useful. One of the basic tasks of commanders is to plan operations. To achieve this goal, commanders prepare documents describing the plan. At this point, a shared understanding of those plans gains importance since they are exchanged with subordinates and vice versa. The success of operations is tightly coupled with the quality of plans. Plans should be easily understandable, should not contain ambiguities and should have a precise and well-defined structure. All of these requirements make military planning process an appropriate area for the use of CNLs.

CPM is an ontology developed to support military planning in order to provide a shared understanding. CPM can be regarded as a representation strategy for operation plans. In order to form a common understanding, it presents its end users a set of constructs that are used to show plan information, such as tasks, goals, resources, activities, constraints (Allen et al, 2008).

The CPM ontology, can be represented by first-order languages like OWL (Web Ontology Language), or by CNLs. CPM was first developed using OWL, but in order to ensure shared understanding, more expressive means to represent a plan were needed. As a result of this necessity, Controlled English (CE) has been used to represent CPM. The CE used in CPM model is based on John Sowa's Common Logic Controlled English (Sowa, 2007), which supports a constrained set of lexical items and grammatical rules for interpreting and generating English statements (Mott & Giammanco, 2008). CPM uses a subset of Controlled English that supports the military planning terminology in order to generate easily understandable and unambiguous plans that can be understood among commanders from many different nations especially in coalition operations (Mott & Giammanco, 2008).

The table below shows is provided to give a summary of basic plan elements and relationships between those plan elements and their CE statements in CPM (Ibbotson, 2012).

Table 5.Basic plan elements and their corresponding CE statements (Mott, D. 2011b)

Element	Explanation/Actual statement	CE statement
Task	Tasks are low level military activities executed to achieve a goal in a time period (start and end time) which has an assignee and assigner.	*"the task TASK1 has the agent INFUNIT1 as executor and has the agent UNIT1 as assigning agent and has 600 as earliest start time and has 700 as latest start and time has 650 as earliest completion time and has 800 as latest completion time and has 50 as minimum duration and has 100 as maximum duration." (Mott, D. 2011b)*
Goal	A goal is a targeted situation or advantage gained by executing a task.	*"there is a goal named GOAL7 that has the division 'US DIV' as owner." (Mott, D. 2011b)*
Mission	Missions can be regarded as high level tasks assigned to a high level military unit whose assigner is usually a political agent such as government.	*"there is a mission MISSION1 that has the division 'US DIV' as setter and has the brigade 'UK BDE' as author." (Mott, D. 2011b)*

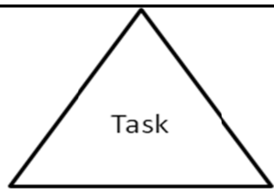
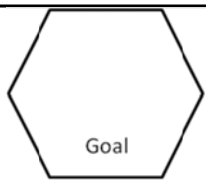
Intent	Intent is a high level plan that is produced by the assignee of a mission representing the interpretation of that mission.	*“there is an intent INTENT1 that has the agent ‘UK BDE’ as executor“ (Mott, D. 2011b)* *“the mission MISSION1 collaborates with the intent INTENT1.” (Mott, D. 2011b)*

Although CNLs are appropriate means for representing plans, they also bring some difficulties. One of them is that writing CNL statements is a tedious task and requires training. Another one is that, although they are both machine and human readable, they can still be ambiguous to human end-users. “For example, do human end-users fully understand the meaning (i.e. the semantic implications) of statements such as: ‘If x is a dog then it has 4 legs’ and ‘If x has 4 legs then it is a dog’ ” (Allen et.al, 2008).

The first difficulty can be solved by supplying plan visualization and editing tools and auto generate CE statements from plan graphs, thus eliminating the process of writing in CE. CPM uses several visual editing tools developed by IBM and Boeing. The second difficulty is not investigated in the context of this study.

Basic planning elements used in CPM and their visual representations are given in the table below (Mott, 2011a). The notations used in the current study are the same used in CPM framework.

Table 6.Plan Elements and Their Visual Mappings

Element	Visual Mapping
Task	
Goal	
Mission	
Intent	

4.2 Controlled Turkish and Collaborative Planning Model

To my knowledge, there is no work in Turkish in the field of controlled natural language creation. In this study a simple Controlled Turkish is developed for constructing a sample task hierarchy.

The controlled language constructed in this study is confined to the main rules given, as given by O'Brien (2003). The rules are classified into four categories: lexical rules, syntactic rules, text structure rules and pragmatic rules. These are:

Lexical rules:

- A word has only one meaning.
- Abbreviations and acronyms are not allowed.
- Pronouns are not allowed.

Syntactic Rules:

- The rule of agreement between subject and verb is applied.
- No conjunctive words are allowed.
- Simple present tense is used.
- Active voice is used.
- Passive voice is not allowed.

Semantic rules:

- Synonyms and homonyms are not allowed.
- Metaphors are not allowed.

Text structure rules:

- Sentence length is limited to a maximum number of 15 words.

4.3 A Visual Modeling Tool for Controlled Turkish

In this section the validity of the graphical representation, namely its syntactic conformance is inspected. Then, the design issues and the rules used in this study are given in detail.

4.3.1 Design Issues - Syntactic Conformance of Graphical Representation

In order to generate a proper graphical representation, Engelhardt's study (2007) is taken as reference. According to Engelhardt (2007), there are three main building blocks of a graphic. These are graphic objects, graphic spaces and the graphic properties. In this study the triangles representing the tasks, the lines connecting them which represent the temporal relationships between tasks and the labels both for the tasks and the relationships fall into graphical objects category. The space that is used to arrange tasks locations according to their relationships falls into the graphic space category. The color and font used in labels and the other graphical objects fall into the graphic properties.

According to Engelhardt (2007), graphical objects can also be classified according to the syntactic categories as in table below.

Table 7.Syntactic Categories of Graphic Objects and Rules for Their Combination (Engelhardt,2007,p.29)

Syntactic categories of graphic objects:	Type of attachment:	Example(s):
node	is attached to: either a point in a meaningful graphic space, or to nothing	a dot marking a city on a map, or a text box in a flow chart
label	is attached to: a graphic object that is labelled by it	a name labelling an object on a map
connector	is attached to: two graphic objects that are connected by it	a line connecting two names in a family tree
line locator	is attached to: a specific line in a meaningful graphic space	a river on a map, or the curve of an electrocardiogram
surface locator	is attached to: a specific surface in a meaningful graphic space	a colored surface on a map, representing a lake or a country
grid marker	is attached to: points and lines of orientation in a meaningful graphic space	latitude/longitude lines on a map, or axes and tick marks in a chart
proportional segment	is attached to: a segment of the surface of a graphic object	a pie segment in a pie chart
frame	is attached to: the graphic object that is framed by it	the line around the panel in a comic book
etc…	etc....	etc....

According to table given above, in this study the graphical objects used can be categorised as the nodes which correspond to triangles representing he tasks, the connectors correspond to the lines relating the tasks and the labels correspond to the text representing the relationship or the name of the task.

Engelhardt also classifies the graphical space as shown in the table below.

Table 8. A Typology of Meaningful Graphic Shape (Engelhardt,2007,p.33)

A typology of meaningful graphic space: Alternative terminology and explanations:		*representation of* ***physical space***	*representation of* ***conceptual space***
"proportion" (in French, Bertin 1967) "interval" (Tversky 1995) "quantitative" (Engelhardt *et al.* 1996) "ratios of spatial distances [...] are perceived as meaningful" (Engelhardt 2002) "quantitative grid" (Card 2003)	**metric space** (shows proportions)	e.g., a topographic map, most pictures	e.g., a time axis, any other quantitative axis
"ordre" (in French, Bertin 1967) "ordinal" (Tversky 1995, Engelhardt *et al.* 1996) "a metric space that was printed on a 'rubber sheet' and then stretched non-homogenously" (Engelhardt 2002) "ordinal grid" (Card 2003)	**topological space** (shows order)	e.g., the London Underground map, an "exploded view" of a machine	e.g., chronological ordering of panels in a comic, any other meaningful spatial ordering
"association" (in French, Bertin 1967) "categorical" (Tversky 1995, Engelhardt *et al.* 1996) "segmentation" (Engelhardt 1998, 1999) "spatial clustering" (Engelhardt 2002) "nominal grid" (Card 2003)	**grouping space** (shows association)	e.g., columns and rows in a table, any other meaningful spatial grouping	
a) "recursion is the repeated subdivision of space" (Card *et al.* 1999) "nesting", "embedding" b) (Engelhardt 2002) b) "orthogonal placement of axes" (Card *et al.* 1999) "simultaneous combination" (Engelhardt 2002)	**composite space** (constructed from combinations of the spaces above)	a) e.g., the (metric, physical) space of a picture within the (topological, conceptual) space of a chronological sequence, b) e.g., a chart that combines a (metric, conceptual) horizontal time axis with a (metric, conceptual) vertical quantitative axis	

According to table above, the graphical space used in this study falls into the topological space category since the ordering of the tasks are arranged so that they are meaningful according to the type of the relationship.

According to the Engelhardt, the syntactics of the graphics is closely related to the main building blocks of the graphics discussed above. Syntactics of the graphics investigates the syntactic categories of the graphical objects, the rules and relationships among the graphial objects and the graphic spaces. According to him, "*the syntactic structure of a graphic representation is determined by the rules of attachment for each of the involved syntactic categories and by the structure of the meaningful graphic space that is involved".* Since we are able to categorize the graphical objects and the graphics space used in this study the notion of syntactics applies to the visual representation used in our case.

4.3.2 High Level System Design

The figure below represents the high level system design of the visual modeling tool. The boxes represent the components of the systems and the arrows represent the data flow between components.

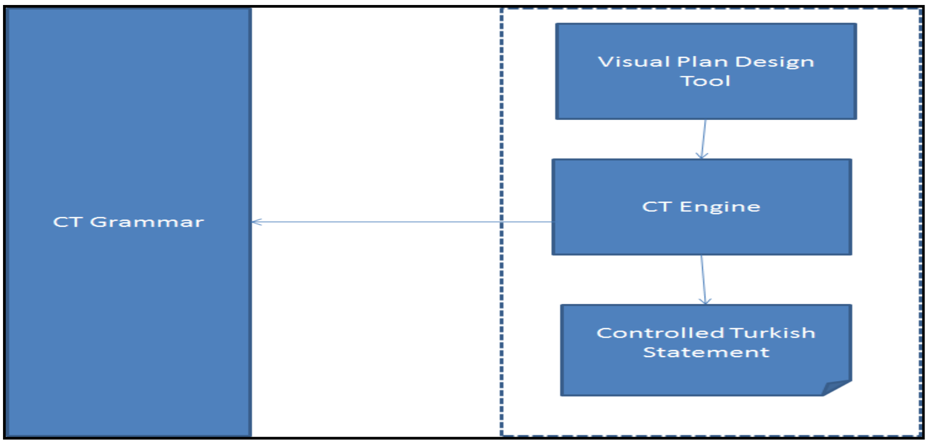


Figure 7. High Level System View for Visual Modeling Tool

4.3.3 Visual Plan Design Tool (VPDT)

Visual Plan Design Tool (VPDT) is the component that is responsible of visualizing plan elements. There is a visual element mapped to each plan element. Relations between plan elements are also represented with tagged lines. Users use these visual plan elements in order to construct the actual task hierarchy.

4.3.4 CT Engine

The CT Engine is the component responsible of processing visual plan elements and converting them into CT statements. It takes a user-designed visual task hierarchy as input. In order to process this visual plan, it uses the CT grammar and predefined visual mappings to generate equivalent Controlled Turkish texts.

4.3.5 Controlled Turkish Grammar

The visual representation of a task hierarchy constructed by the user is transformed to a Controlled Turkish representation by the use of a grammatical rule set.

Since users are not expected to write Controlled Turkish statements, the grammar is completely invisible to the user. When a user chooses to transform the task hierarchy into the Controlled Turkish statement, the CT Engine picks the appropriate grammar rules and transforms them into a Controlled Turkish representation.

The grammar is written in xml file format containing tags and the value pairs. Each tag represents a visual design element or intermediate grammar elements that are used while processing other tags. For example, <TASK> tag is a direct representation for the Task visual element while

<OBJECT_GEN> tag represents an intermediate state that could occur during the transformation process.

The rules are organized so that each represents a state in the transformation process. The initial state is when a tag maps to a visual representation element. Examples for this case are the TASK tag and the REL tags. The TASK tag corresponds to a Task in the visual representation where the REL tag corresponds to the relationships among tasks. The TASK tag is represented with the rule given below.

```
<entry key="TASK">"ad"-DET-Task.taskIdentifier-",&#10;"-"başlangıç zaman"-DET-Task.startTime-",&#10;"-"bitiş zaman"-DET-Task.endTime-",&#10;"-OBJECTIVE-RESOURCE-SUB_TASK-"olan"-" bir"-Task.taskType-" görev"-DET-" tanımlıdır."</entry>
```

This TASK tag represents an initial state in order to transform a task visual element into a Controlled Turkish statement. It represents the structure of the equivalent Controlled Turkish statement.

The REL tags are represented with the rule set given below.

```
<entry key="REL,AYNI_ANDA_BASLAR">SUBJECT-OBJECT-" ile "-TaskAssoc.type</entry>

<entry key="REL,AYNI_ANDA_BITER">SUBJECT-OBJECT-" ile "-TaskAssoc.type</entry>

<entry key="REL,ONCE_BASLAR">SUBJECT-OBJECT_ABL-TaskAssoc.type</entry>

<entry key="REL,BITISINDEN_SONRA_BASLAR">SUBJECT-OBJECT_GEN-TaskAssoc.type</entry>

<entry key="REL,SONRA_BASLAR">SUBJECT-OBJECT_ABL-TaskAssoc.type</entry>

<entry key="REL,ESNASINDA_BASLAR">SUBJECT-OBJECT-TaskAssoc.type</entry
```

The REL tag in this rule set is used in conjunction with the relationship type between the tasks where different relationship types need different grammatical structures.

A quick look to the rules shows that they contain structures in quotes such as ***"başlangıç zaman"***, and structures represented in capitals such as ***OBJECTIVE-RESOURCE-SUB_TASK.*** In addition to these, one can notice dashes splitting these structures. The details of these structures are explained below.

The words in quotes are the reserved words for the Controlled Turkish representations. They are the leaf nodes for the grammar and they do not need further processing. They are directly projected into the output.

The dash symbol separates grammar structures in order to facilitate the parsing process. They do not have any effect on the output.

The capitalized words represent the other intermediate grammar rules. Each of these rules has a mapping tag which fully defines the rule. These intermediate rules can also be grouped according to their functions. These groups are explained below.

4.3.5.1 Subject & Object

In the current study, the main visual elements are tasks and their relationships. Both of these elements are defined with respect to a subject and an object. The related rules are given as follows.

```
<entry key="SUBJECT">Task.taskIdentifier-" görev"-DET-";"</entry>
<entry key="OBJECT">Task.taskIdentifier-" görev"-DET-" "</entry>
<entry key="OBJECT_GEN">Task.taskIdentifier-" görev"-DET-GEN-" "</entry>
<entry key="OBJECT_ABL">Task.taskIdentifier-" görev"-DET-ABL-" "</entry>
```

As seen from the rules above, there is more than one rule for subjects and objects. The reason for this is due to transitivity and case-assigning features of verbs used for the relations. According to the type of the verb, the rules may differ so that the subject or the object task can contain genitive or ablative affixes. For example, when the relationship type is ***“AYNI_ANDA_BASLAR”*** the rule ***<entry key="REL,AYNI_ANDA_BASLAR">SUBJECT-OBJECT-" ile "-TaskAssoc.type</entry>*** is applied. According to this rule, there is no need to use genitive, ablative or dative affixes. But, when the relationship type is chosen as ***“ONCE_BASLAR”*** then the rule ***<entry key="REL,ONCE_BASLAR">SUBJECT-OBJECT_ABL-TaskAssoc.type</entry>*** is mapped to the relationship requiring an ablative affix.

4.3.5.2 Case Suffixes

In the current study, case suffixes are used in conjunction with the last vowel and the flag indicating if the word using the case suffix ends with a vowel or not in order to determine the right suffix for the output. An example rule is given below.

<entry key="CASE_SUFFIX,e,true"> yi </entry>

According to the example rule given above, the last vowel is the “e” and the word ends with a vowel so the resulting suffix is “yi” where the “y” is a buffer phoneme

4.3.5.3 Genitives

The logic used in the genitive rule set is also similar to the logic used in case suffixes. Genitives are required to represent possession for the used noun. In this study, the use of genitives are the cases where they are attached to subjects and objects of the sentences. The genitives are used in conjunction with a last vowel and a flag indicating if the word using the genitive ends with a vowel or not in order to determine the right affix for the output. An example rule is given as follows.

```
<entry key="GEN,e,true">nin</entry>
```

According to the rule, if the last vowel is “e” and the word ends with a vowel the resulting suffix is “nin”, where “n” stands for the buffer phoneme.

4.3.5.4 Ablatives

Ablatives are used to indicate separation, or direction away from something. In the grammar used for the current study ablatives are used with objects in order to represent time relations between tasks. Suffixes coming from this rule set also differs according to the last vowel of the root; a flag is used to determine if the word ends with has a vowel or not. An example from ablative rule set is given below:

```
<entry key="ABL,ü,false">den</entry>
```

According to the rule, if the last vowel is “ü” and the word does not end with a vowel, the resulting suffix is “den”.

4.3.5.5 Visual Mapping

In addition to the grammar, a mapping for the visual elements and the tags used in the grammar is needed. The task visual element represented with a  is mapped to the TASK tag which is represented with the rule:

```
<entry key="TASK">"ad"-DET-Task.taskIdentifier-",&#10;"-"başlangıç zaman"-DET-Task.startTime-",&#10;"-"bitiş zaman"-DET-Task.endTime-",&#10;"-OBJECTIVE-RESOURCE-SUB_TASK-"olan"-" bir"-Task.taskType-" görev"-DET-" tanımlıdır."</entry>
```

In addition to the tasks, the relationships between visual elements represented with a line connecting the tasks are mapped to the REL tag which is represented with the REL rule set. An example for the REL rule set is:

<entry key="REL,SONRA_BASLAR">SUBJECT-OBJECT_ABL-TaskAssoc.type</entry>

4.3.5.6 *Modeling the Plain Task Hierarchy*

In order to construct a task hierarchy, users can add task elements and associate them with each other according to timing and functional relations. In order to add a task, a drawing toolbar at the right side of the application is used. When clicking on the task element a dialog shows up in order to enter element properties. Then, when user clicks on the modeling area the task item is dropped on the modeling area. The figure below shows a sample task hierarchy.

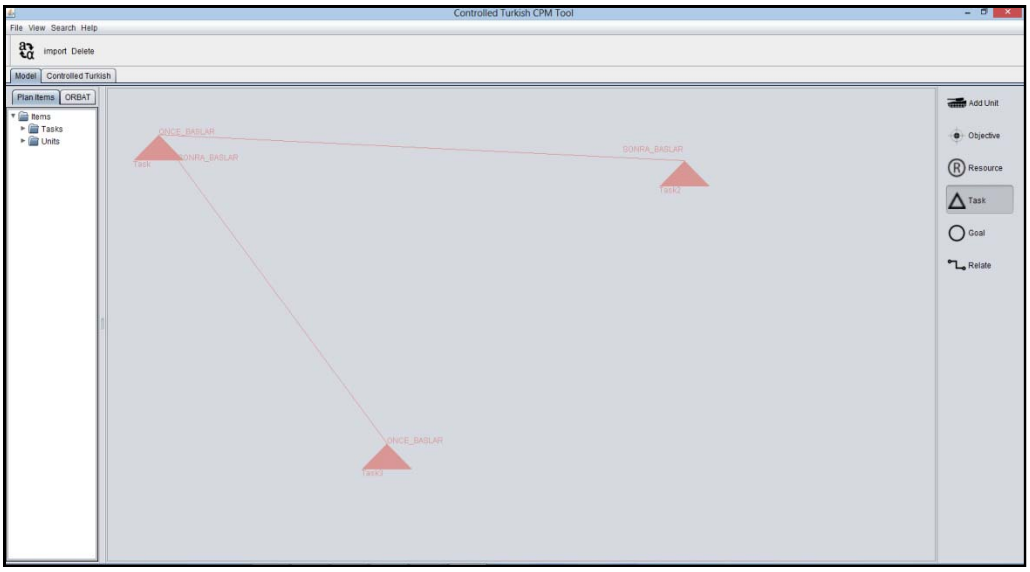


Figure 8. Sample Task Hierarchy

After the user finishes modeling the task hierarchy, the Controlled Turkish representation can be viewed from the Controlled Turkish tab. The Controlled Turkish representation of the modeled task hierarchy can be exported to the file system by clicking on the Export button. A sample Controlled Turkish version of a sample order model is provided in Annex B.

CHAPTER 5

EXPERIMENTS

5.1 Preliminary Experiment

Although graphical representations are valuable for comprehension, one has to consider the factors that affect their quality. As explained in Section 3.3.2, the design of a graphical representation is very crucial for comprehensibility. Semantic notations used in graphical designs should be chosen carefully so that users do not experience ambiguity and difficulty while interpreting them.

The purpose of this preliminary study is to compare the quality of two different graphical representations which represent similar information in two different ways. The data represented in these graphical representations, is a group of tasks with temporal relations. One graph is a directed graph with one tag information on a relation (see Figure 9) and the other is a fully-tagged but not directed graph (see Figure 10).

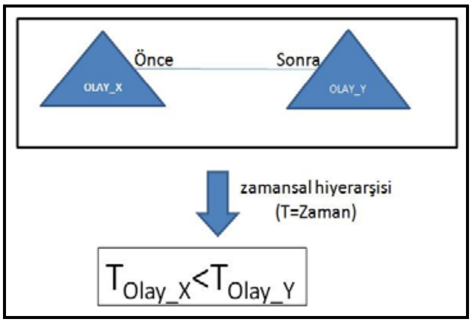


Figure 9. Graphical Representation for Task-1 Example

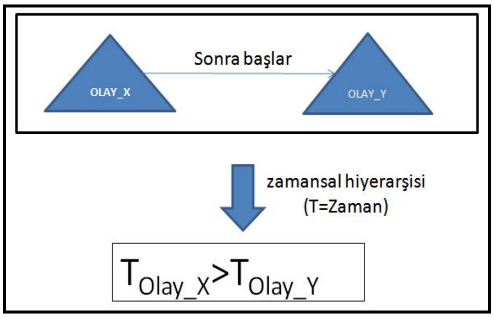


Figure 10. Graphical Representation for Task-2 Example

The questions in this preliminary study are given as follows:

- Which graphical representation is more comprehensible and easier for readers, directed graphs or fully-tagged graphs?
- Do directed graphs lead to misunderstandings for its readers?

In order to find answers to these questions, the comprehensibility aspects of these two graphical representations have to be evaluated. Participants were asked to sort the actions placed in the graphs according to their temporal precedence. During the task, participants' eye tracking metrics and verbal recordings were collected. In addition, the elapsed time for the task was gathered for performance evaluation.

In order to evaluate which representation is better, the prerequisite condition is that the representations are informationally similar. Larkin and Simon (1987, p.67), explain that the two representations are informationally equivalent *"if all of the information in the one is also inferable from the other and vice versa".* After the informational equivalence is ensured, the comprehensibility, ease and convenience aspects of these two representations can be investigated. According to Simon and Larkin (1987) these aspects can be measured in terms of their computational efficiencies. According to them, two representations are computationally equivalent *"if they are informationally equivalent and, in addition, any inference that can be drawn easily and quickly from the information given explicitly in the one can also be drawn easily and quickly from the information given explicitly in the other, and vice versa"* (Simon and Larkin, 1987, p.67).

However, the term "informationally equivalent" used by Larkin and Simon has different aspects. The inference process of human mind is complex and has many dependencies such as prior knowledge which can make one to assert information presence even if the information is not presented there. So, instead of using the term "informationally equivalence", it seems more appropriate to use the term “informationally similar”. Thus, in this study two informationally similar visual representations are compared for comprehension. The representation which is more difficult to comprehend requires more cognitive processing and thus requires more cognitive effort.

In this preliminary study, in order to measure the efficiency of these two representations, eye tracking metrics and verbal reports of the participants were analyzed. Moreover, the duration of the task was measured for each participant. At the end, the correctness and the rapidity of the participants were interpreted together to come up with a conclusion about the comprehensibility of the graphical representations.

In eye tracking protocol, participants’ eye movements were recorded while viewing the graphs. The aim of this protocol is to get clues about graphical representations from participants’ eye movements. By using eye tracking method, one can detect eye fixations, gaze counts and gaze durations which provide significant feedback for the representations.

Another type of protocol used in this study is the verbal protocols. Ericsson and Simon (1980) state that *“…verbal reports, elicited with care and interpreted with full understanding of the circumstances under which they were obtained, are a valuable and thoroughly reliable source of information about cognitive processes”* (p. 215). Although verbal protocols require interpretation and analysis effort, they are valuable source of information.

The rationale behind using verbal protocol is that, the data obtained from this protocol can be valuable since it represents the raw thoughts of the participants while performing the task. They can give important feedback on problem detection about the representations since they are obtained at the same time the participant faces with a problem.

5.1.1 Preliminary Experiment Details

The experiment conducted was to evaluate the comprehensibility of two different graphical representations of an action hierarchy. Below is given the task details, the participants that take place in the experiment, experiment procedure and data gathering methods used in this study.

5.1.1.1 Preliminary Experiment Task

Participants were asked to sort the tasks according to their temporal precedence; they are expected to verbalize the tasks' temporal hierarchy for two different graphical representations. Two variants of graphical representations that are informationally similar were provided to the participants. The graphical representations used in this study are shown in Figure 11 and Figure 12. As it can be seen in figures, the triangular shapes stand for "tasks" and the lines between them stand for the temporal relations between them. The Task 1 in this experiment is to verbalize the temporal relations for the first graphical representation stimulus given in Figure 11. As it can be seen from the figure, this graph is an undirected and a fully tagged graph. Both ends of relations between tasks are tagged as "önce" (before) and "sonra" (after). After participants finished verbalizing the first graphical schema, they were presented Task 2 which asked them to verbalize the graphical schema given in Figure 12. As it can be seen in Figure 12, this version of graphical representation is a directed graph, where the direction information is specified by an arrowhead on the relation. The relations between actions in this representation contain only one tag information.

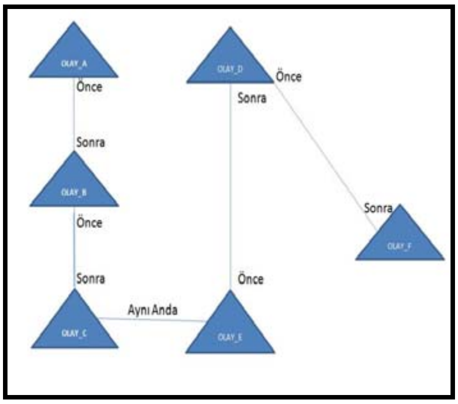


Figure 11. Undirected Graph Representation

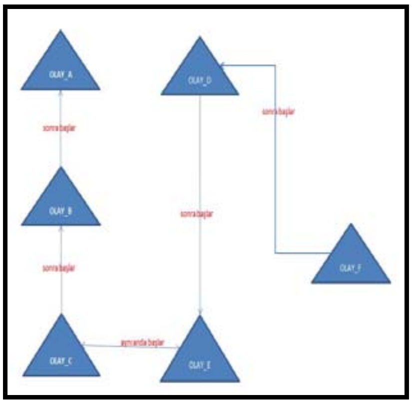


Figure 12. Directed Graph Representation

5.1.1.2 Participants

The experiment was conducted with six people. All of them were students of METU Cognitive Science Department. 5 participants' data were collected and analyzed in this experiment.

5.1.1.3 Preliminary Experiment Procedure

The experiment was conducted at Eye Tracker and Test Computer in Human Computer Interaction Research and Application Laboratory of Middle East Technical University (METU). The computer in this laboratory is used to collect useful data such as where the user looks at on the screen at a specific time and metrics such as gaze intervals, the fixation counts and durations. Moreover, voice recordings can be performed. The experiment was carried out on the software called "TOBII Studio v.3.1.0" which is installed on the computer in the eye tracking laboratory. This software was also used in the analysis stage. It mainly converts the recorded eye movements and voice records into visual and digital data and it enables experimenter to analyze the recorded data.

At the beginning of the experiment, a short calibration session was conducted for each participant. The eye tracker's position was arranged relative to the participant, and then, participants were asked follow a red spot appearing on the screen. By this way, the data quality of eye tracking records was ensured.

Before starting the experiment, participants were informed about the performance criteria and they were asked to perform their tasks as soon as possible. In addition, participants were asked to verbalize the actions' temporal precedence while performing their tasks. A preliminary slide was shown to the participants in order to inform them about the notations used in the experiment. Another preliminary slide was given before the stimulus of the first graphical representation for training purpose. In this slide, a simple example action hierarchy was provided for graphical representation (see Figure 9). After, the representation in Figure 11 was shown to the participant and his/her eye tracking metrics; voice records and time to complete the task were collected. Next, another example slide was provided (See Figure 10) for graphical representation shown in Figure 12. Again, the eye movements, voice recordings and time to complete the task were gathered.

5.1.2 Data Gathering Methods

5.1.2.1 Verbal Protocol Methodology

According to the classification given in the literature review, the verbal protocol used in this study falls into Level 3 class. The reason is that in this study participants were asked to achieve a sorting task according to temporal precedence. At first sight, this seems to fall into Level 2 class since the stimulus provided to participants is in graphical form. However, in this task, the participants were asked not only to verbalize the given graphical representation but they were also asked to perform a sorting according to temporal relations. So, the participants filtered or processed the information according to a criterion which is the temporal relations in this case. Thus, it can be concluded that the verbal protocol used in this study falls into Level 3 class according to the information stored.

According to the probing technique classification explained in the literature review, the verbal protocol used in this study is a think aloud protocol. Data collected from participants represent their thoughts while they were processing the information. Data also contains the participant thoughts while s/he faces with a problem while performing the task. Data obtained from this protocol is very useful to detect problems and in the diagram as well as important clues for their solutions.

After the experiment, participants' voice records were transcribed into verbal reports in order to analyze. Then a segmentation process was applied.
At the next step, generated segments were classified according to a criterion. In this study, the segmentation criteria were the information included in that segment. The segmentation was made so that each segment contains only one piece of data about the sequencing task which generally corresponds to a sentence.

After the preprocessing step, the categorization of the segments took place. The categorization idea of Ericsson and Simon's is very suitable in this case: The idea was to encode segments according to the information contained within them and to encode them separately from the neighboring segments where each segment contains single information (Ericsson & Simon, 1980).

According to this idea, the classification was carried out according to two different criteria. One of them is the "Conceptual Category" which was used in Tenbrink's study (2008). In this category, segments were classified according to information type such as "temporal structure", "causality and reasoning". In addition to that, "hesitations" was also used as a classifying category since they give valuable information about the comprehension of tasks. The other criterion was the problem type which classifies segments according to the problem areas they indicate. The problem areas that were used in this type of category were "reverse sequencing" which indicates that participant ordering is in reverse order, and “missing/wrong sequence” which indicates a wrong or missing order has occurred in task sequence.

5.1.2.2 Performance Criteria

In the current preliminary study two informationally similar graphical diagrams were compared for their computational efficiencies. Computational efficiencies of the diagrams were evaluated according to how easily and quickly a participant can infer the expected information. In order to measure this, time was used as a performance criterion. One of the motivations in using time criterion was to increase the participant’s motivation and performance.

5.1.2.3 Eye Tracking Method

In eye tracking method, fixation counts and fixation durations were measured for two different graphical representations. These metrics are briefly explained below:

- The fixation duration is the duration of each fixation in an area of interest.
- The fixation count is number of times a participant fixates in an area of interest.

These two metrics can give important feedback on tasks’ comprehension.

5.1.3 Results

5.1.3.1 Verbal Protocol Results

As mentioned in the methodology section 3.2.2.1, two different classifications for the information segments were given for analysis. The results according to these classifications and the classification criteria are listed in

Table 10. Conceptual Category Results for Task 2 to

below.

Table 10. Conceptual Category Results for Task 2 and **Error! Reference source not found.** below show the classification done according to the conceptual category for two different representations. First column in the table indicates the task identifier; the second column indicates the participant and the number of segments that are given by that participant during the experiment. The third column shows the number of segments containing temporal structures like "önce" (before), "sonra" (after), "şimdi" (now); the fourth column shows the segments containing causality and reasoning statements such as "çünkü" (because), "bundan dolayı" (therefore) etc…". The last column shows the number of hesitations and the hesitation time in seconds. The last row of the table shows the total results.

Table 9. Conceptual Category Results for Task 1

Task Identifier	Participant/# of segments	temporal structure (önce,sonra,şimdi …)	causality /reasons (çünkü, bundan dolayı…)	Hesitations
T-1	**P1-6 segment**	**3**	**5**	**4 times (4 sec)**
	P2-5 segment	**5**	**-**	**2 times (2 sec)**
	P3-4 segment	**7**	**-**	**1 times (2 sec)**
	P4-5 segment	**5**	**-**	**-**
	P5-3 segment	**3**	**1**	**1 times (1 sec)**
Total	**23 segment**	**23**	**6**	**8 times (9 sec)**

Table 10. Conceptual Category Results for Task 2

Task Identifier	Participant/# of segments	temporal structure (önce,sonra,şimdi …)	causality /reasons (çünkü, bundan dolayı…)	Hesitations
T-2	**P1-5 segment**	**5**	**-**	**2 times (3 sec)**
	P2-6 segment	**5**	**1**	**5 times (17 sec)**
	P3-6 segment	**5**	**-**	**2 times (2 sec)**

	P4-4 segment	**4**	**-**	**2 times (2 sec)**
	P5-7 segment	**5**	**1**	**1 time (8 sec)**
Total	**29 segment**	**24**	**2**	**12 times (32 sec)**

Table 12. Problem Type Results for Task 2 and below represent results for the second classification. This type of classification is done according to the problem types that are determined through the experiment. The first column is the task identifier, the second column represents the participant and other columns show the number of occurrences of reverse sequencing and missing sequence errors. The last row shows the total number of errors.

Reverse sequencing is the error type in which participant gives the overall sequence of the tasks in the reverse direction. For instance, if the correct sequencing is T1, T2, T3 participant gives the sequence as T3, T2 and T1. Missing/Wrong sequencing is the error type in which participant forgets to mention the sequence of a Task or gives the sequence of a Task in a wrong place. For instance, if the correct sequence is T1, T2, T3 participant gives the sequencing as T1, T3 or T1, T3, T2.

Table 11. Problem Type Results for Task 1

Task Identifier	Participant	Reverse sequencing	Missing sequencing
T-1	**P1**	**-**	**-**
	P2	**-**	**-**
	P3	**-**	**-**
	P4	**-**	**-**
	P5	**-**	**1**
Total		**-**	**1**

Table 12. Problem Type Results for Task 2

Task Identifier	Participant	Reverse sequencing	Missing

			sequencing
T-2	P1	1	6
	P2	-	-
	P3	1	6
	P4	-	1
	P5	1	7
Total		3	20

5.1.3.2 Performance Criteria Results

Table 14. Time Performance Results for Task 2 and

below shows the performance of participant's while achieving the given task. The participant's performance is measured according to the time required to complete the task.

The first and the second column are to identify the task and the participant. The third column shows the total time needed to complete the task, and the last column shows the participant's examination time on diagram before starting to talk about the diagram. The last row shows the average of the total time needed to complete the task and the average examination time.

Table 13. Time Performance Results for Task 1

		Total Time	Examination time
T-1	P1	30 sec	8 sec
	P2	26 sec	5 sec
	P3	21 sec	5 sec
	P4	13 sec	2 sec
	P5	21 sec	7 sec
Average		20.2 sec	5.4 sec

Table 14. Time Performance Results for Task 2

		Total Time	Examination time
T-2	P1	26 sec	11 sec

	P2	**37 sec**	**5 sec**
	P3	**22 sec**	**4 sec**
	P4	**13 sec**	**3 sec**
	P5	**31 sec**	**6 sec**
Average		**25.8**	**5.8 sec**

5.1.3.3 Eye Tracking Results

Table 13 represents the fixation duration results and **Error! Reference source not found.** represents the results for fixation counts that were collected in eye tracking session.

Table 15 Results for Fixation Duration

	Fixation Duration					
	Directed Graph			Undirected Graph		
	Directed Graph AOI			Undirected Graph AOI		
Recordings	N (Count)	Mean (Seconds)	Sum (Seconds)	N (Count)	Mean (Seconds)	Sum (Seconds)
Rec 01	95	0,23	22,28	82	0,25	20,12
Rec 02	95	0,21	20,22	119	0,25	29,51
Rec 03	41	0,28	11,63	48	0,21	10,09
Rec 04	65	0,24	15,62	57	0,27	15,24
Rec 06-08	76	0,22	16,93	105	0,23	24,28
All Recordings	372	0,23	86,67	411	0,24	99,25

Table 16 Results for Fixation Count

Fixation Count					
Directed Graph			Undirected Graph		
Directed Graph AOI			Undirected Graph AOI		
N (Count)	Mean (Count)	Sum (Count)	N (Count)	Mean (Count)	Sum (Count)
1	95,00	95	1	82,00	82
1	95,00	95	1	119,00	119
1	41,00	41	1	48,00	48
1	65,00	65	1	57,00	57
1	76,00	76	1	105,00	105
5	74,40	372	5	82,20	411

5.1.4 Analysis of Results

Preliminary study results and the analysis showed that the undirected (fully tagged) graphs were easier and required less time to understand. In accordance with this, it showed that many of the participants misinterpret the directed graph representation by sorting given tasks in reverse order although there were stereotypes on each edge for the directed graph. So, one can conclude that directed graphs leads to misinterpretations by making readers skip the stereotype information on the directed edges and just focus on the direction.

5.1.4.1 Eye Tracking Analysis

The results of eye tracking that were collected from the five participants, that took place in the preliminary experiment, can be interpreted as follows:

- There was more fixation count on Task 2. This result is an indicator of having more difficulty in comprehending the Task 2.

- The fixation durations were longer in Task 2. This result can be thought that participants have more difficulty in comprehending Task 2.

The heat-maps for Task-1 and Task-2 are provided below:

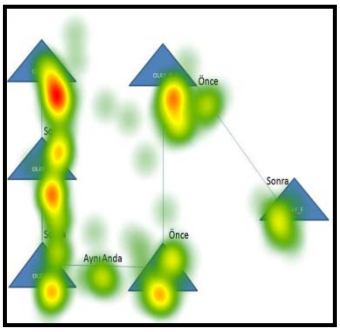


Figure 13. Heat map for Task 1

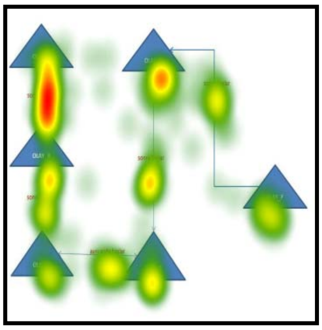


Figure 14 Heat map for Task 2

As it can be seen from the heat maps, the red area on the relationship between "Olay-A" and "Olay-B" is bigger for Task-2. By looking at the heat maps, we cannot know whether the participant looked at the name of task or the name of the relation. Thus, we cannot make inference on task comprehension by just looking at the heat maps. When this result is evaluated with the overall gaze duration difference between these two graphical representations, it may be inferred that task comprehension for Task-2 is lower than Task-1.

5.1.4.2 Verbal protocol analysis

The findings of this method and their interpretations are listed below:

- Hesitation counts and durations in Task 2 were more than Task 1.

When hesitation counts are examined per participant, although the hesitation counts in Task-2 were more than in Task-1, this difference was not significant to infer something. However, these two tasks differ when one takes hesitation durations into account. In Task-2 every participant hesitated longer than

Task-1. This difference in hesitation time was significant. Hesitations, in this case, indicated the participant's need to reanalyze the information or show that the participant's information processing period was still continuing. Both cases were indications of an understanding difficulty. According to these, one can conclude that Task-2 was harder to understand for each participant.

When hesitation counts and durations are examined together, the situation is the same as described above.

- The temporal structures used in Task-1 and Task-2 were almost same in count.
- The total segment count in Task-1 was lower than the total segment count in Task-2

Two findings above are discussed together since they give valuable information when the ratio of them is examined. That is, the number of temporal structures used in Task-1 was 23 and the number of temporal structures used in Task-2 was 24. The difference is negligible. This is not surprising since the number of temporal relations between tasks was the same in both representations.

The difference among the total number of segments used in Task-1 and Task-2 is considerable. According to this, one can conclude that participants tended to use more sentences in order to verbalize the representation in Task-2. This indicates that participants had hard times to understand the representation given in Task-2 and they tended to use unnecessary sentences in order to give supporting information such as ***"Hepsi de ardışık olarak ilerliyor."(All of them are progressing consecutively)***; ***"Burada ise tersi mevcut" (Here the opposite situation is present).*** Such segments are regarded as unnecessary since one can verbalize the whole task sequence by just using the temporal structures. This situation can be supported when the number of temporal structures used in both tasks is examined with respect to the number of segments used. According to this, the ratio (# of temporal structure / # of segments) is greater in Task-1 which shows the existence of the unnecessary supporting information used in Task-2. One can again conclude that representation used in Task-2 was harder to understand when it is compared to the representation used in Task-1.

- The number of causality/reasoning statements used in Task-1 was higher than Task-2.

This finding seems to be valuable when one looks at the total numbers. However, when the number of statements is examined per participant, it shows that only one participant used such causality/reasoning statements in Task-1, which is the reason for the big difference between two representations. Since the general approach or style while performing the task differs from participant to participant, this category segment is not taken into account and it is not discussed.

- Task-2 contains more reverse ordering mistake than Task-1.

Reverse sequencing is an incorrect pattern encountered in the experiments. None of the participants gave this incorrect pattern for the representation used in Task-1. But, 3 participants out of 5 made incorrect sequencing while verbalizing the representation in Task-2. According to this finding, participants tended to use only the direction given in the representation used for Task-2 and skipped the stereotypes used in the directed edges. As a result, they sorted the tasks in reverse order. Although these mistakes can also be related with the participant’s attention, they are definitely an indicator of a problem in the representation. The problem was that when the directions and the stereotypes were chosen in reverse direction, participants tended to just follow the edge directions and ignore the edge stereotype.

- The number of wrong/missing sequence given in Task-2 was much more than Task-1.

Although the number of wrong/missing sequence given in Task-2 was 20, this value for Task-1 was 1. In addition to the explanation given for the previous finding, this big difference between the wrong/missing sequences also indicates that Task-2 was harder and trickier to understand and interpret when compared with Task-1.

5.1.4.3 Performance Analysis

- The average time spent to complete Task-1 was less than the average time needed to complete Task-2.

This finding is may be the clearest indication of the difference between participant’s efforts to complete the tasks. It can be concluded that representation used in Task-2 was harder to understand and needed more attention to complete than the representation used in Task-1.

- The examination durations for both representations were almost same.

This finding is very valuable since it indicates that participants tended to examine both representations in equal time spans. Participants examined two representations approximately in same durations but the total time to complete Task-2 was longer thanTask-1. Since the representation used in Task-2 was more difficult to understand and needed more attention, it became obvious that the same examination duration for Task-2 obviously was not enough for the participants. This situation led to sequencing errors and long hesitations (increasing the total time needed to complete the task).

5.1.4.4 Statistical Analysis of Results

In order to determine the analysis method for the collected data, the appropriate tests have to be applied. The decision for the tests is closely dependent on the nature of the collected data (e.g. the normality and the homogeneity of the data). To decide on which test is more appropriate, the nature of the data has to be explored. To determine the normality of the data, Kolmogorov Smirnov test is applied to the data. The results of the Kolmogorov Smirnov test for the time performance and the correctness are given below for p = .05.

Table 17. Kolmogorov Smirnov Test Results for Normality

CRITERIA	CONTROL_GROUP	Kolmogorov-Smirnov[a]		
		Statistic	Df	Sig.
CORRECTNESS	Undirected	.473	5	.001
	Directed	.331	5	.076
TIME_PERFORMANCE	Undirected	.225	5	.200*
	Directed	.138	5	.200*

The results in the table above can be reported as follows.

- The correctness for the undirected group, D (5) = 0.473, p<.05
- The correctness foe the directed group, D (5) = 0.331, p<.05
- The time performance for the undirected group, D (5) = 0.225, p<.05
- The time performance for the directed group, D (5) = 0.138, p<.05.

According to the results the sig value for the TIME_PERFORMANCE for the Directed and Undirected groups are greater than the .05 indicating that the data is normally distributed within undirected group. The situation is similar for the CORRECTNESS data in directed group where .076 is greater than the .05. However the data for the CORRECTNESS in undirected group is not normally distributed since the sig value .001 is smaller than the .05.

Another property of the data that need to be explored is the homogeneity. In order to determine if the data is homogeneous with in the groups, Levene test has to be applied. The results for the Levene test are given in the table below.

Table 18. Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
TIME_PERFORMANCE	Based on Mean	.483	1	8	.507
	Based on Median	.519	1	8	.492
	Based on Median and with adjusted df	.519	1	7.648	.493
	Based on trimmed mean	.466	1	8	.514

The results for the Levene test can be reported as:

- Since the CORRECTNESS data is not normally distributed among directed and undirected groups. The levene test is skipped for correctness data.

- For the time performance, the variance for the directed and undirected groups are not significantly different $F (1, 8) = 0.483$, $p < .05$

According to the results the sig values for the TIME_PERFORMANCE data are greater than the .05 value indicating that the variances are not significant. Thus the data is homogeneous within the groups. However the situation is different for the CORRECTNESS data since the values in the Sig column are not greater than the .05 (.000 values for the based on Mean and the Based on trimmed mean).

According to these results, the t-test for independent groups can be applied for the TIME_PERFORMANCE data. The results of the t-test are as follows.

Table 19. Group Statistics for t-test

	CONTROL_GROUP	N	Mean	Std. Deviation	Std. Error Mean
TIME_PERFORMANCE	Undirected	5	22.20	6.380	2.853
	Directed	5	25.80	9.094	4.067

Table 20. Independent Samples Test for t-test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
TIME_PERFORMANCE	Equal variances assumed	.483	.507	-.725	8	.489	-3.600	4.968	-15.056	7.856
	Equal variances not assumed			-.725	7.169	.492	-3.600	4.968	-15.291	8.091

According to the results the sig (2-tailed) value is .489 which is greater than the .05 value indicating that the there is no significant difference between the means of the two samples. From this one can conclude that the using directed or undirected graphs make no difference on the time performance.

Since the CORRECTNESS data is not normally distributed among directed and undirected groups. The t-test cannot be applied. Instead of t-test Mann Whitney U test will be applied to analyse the data for CORRECTNESS. The results of the test are given below.

Table 21. Ranks for Mann Whitney U Test

	CONTROL_GROUP	N	Mean Rank	Sum of Ranks
CORRECTNESS	**Undirected**	5	3.70	18.50
	Directed	5	7.30	36.50
	Total	10		

Table 22. Test Statistics for Mann Whitney U Test

	CORRECTNESS
Mann-Whitney U	3.500
Wilcoxon W	18.500
Z	-2.019
Asymp. Sig. (2-tailed)	.043
Exact Sig. [2*(1-tailed Sig.)]	.056

According to the results the asymp sig value is .043 which is less than .05 indicating that the results among directed and undirected groups are significantly different. To calculate the exact effect size, the formula of r = Z /

√N is used where r is the effect size; Z is the z score in given in the test results and the N is the number of the total observations. According to this the r = -2.02/ √10 = -0.63 which is above the .5 threshold for a large effect.

In the light of these results, looking at the Ranks table Since the mean rank of the directed group is 7.3 which is greater than 3.7 which is the mean rank of the undirected group, one also conclude that the directed group makes more mistakes than the undirected group.

5.2 Conformity of Controlled Turkish Statements to Ateşman's Formula

In order to see if the statements in Controlled Turkish are simple enough, Ateşman's readability formula was applied on these statements.

Since the text in Controlled Turkish was not a discourse or narrative text, each sentence was evaluated independently. Thus, Ateşman's formula was applied on each sentence. The text contained two types of sentences. One type of sentence was in the form: "Adı "Görev X" olan bir görev tanımlıdır." (There is a task that has name "Task X") and the other type of sentence was in the form "Görev X, Görev Z'den önce başlar." (Task X starts before Task Z).

Ateşman's index for sentence "Adı "Görev X" olan bir görev tanımlıdır." (There is a task that has name "Task X") is given below:

198,825 – (40,175 * (13 /6)) – [2,610 * (6/1)] = 198,825 – 87,045 – 15,66 = 96,12

According to Ateşman's readability category, this index is in the "very simple" category.

Ateşman's index for sentence "Görev X, Görev Z'den önce başlar" is given below:

198,825 – (40,175 * (11 /4)) – [2,610 * (4/1)] = 198,825 – 110,48 – 10,44 =77,905

According to Ateşman's readability category, this index is in the "simple" category.

Thus, we can say that the sentences provided in Controlled Turkish are simple enough according to Ateşman's readability formula.

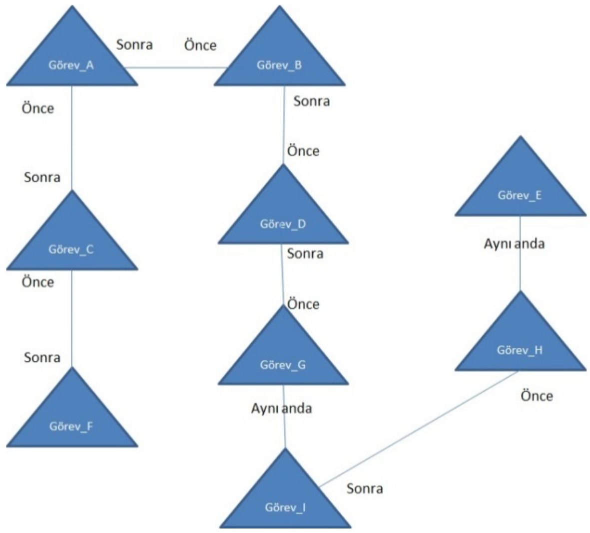
Since Ateşman's formula is only dependent on the syllable count, word count and sentence count, we cannot make any inference about possible ambiguities of the sentences. Such circumstances are eliminated by rules provided in section 4.3.

5.3 Main Experiment

In the preliminary experiment (See Section 5.1), the visual representation used in this study was examined. The original CPM visual representation and the modified visual representation were compared for comprehension and the results showed that the modified visual representation performed better for comprehension. Similarly, a preliminary study was carried out for the textual representations given in Controlled Turkish. The text was subjected to Ateşman's readability formula in order to ensure that the text was simple enough. The details of this study are given in Section 5.2.

After the visual representation and textual representation was evaluated for comprehensibility and ease of reading, the main experiment, which was the comparison of graphical representation and a textual representation in Controlled Turkish both representing a same task hierarchy, was conducted. The textual representation and the graphical representation in this experiment can be seen in the table below.

Table 23. Textual and Graphical Representation in Main Experiment

<table>
<tr>
<td>Adı “Görev X” olan bir görev tanımlıdır.
Adı “Görev Z” olan bir görev tanımlıdır.
Görev X, Görev Z’den önce başlar.
Adı “Görev A” olan bir görev tanımlıdır.
Görev Z, Görev A’dan önce başlar.
Adı “Görev Y” olan bir görev tanımlıdır.
Görev X, Görev Y’den sonra başlar.
Adı “Görev T” olan bir görev tanımlıdır.
Görev Y, Görev T’den sonra başlar.
Adı “Görev B” olan bir görev tanımlıdır.
Görev T, Görev B’den sonra başlar.
Adı “Görev C” olan bir görev tanımlıdır.
Görev B, Görev C ile aynı anda başlar.
Adı “Görev E” olan bir görev tanımlıdır.
Görev C, Görev E’den sonra başlar.
Adı “Görev D” olan bir görev tanımlıdır.
Görev D, Görev E ile aynı anda başlar.</td>
<td>
</td>
</tr>
</table>

In the remaining of this section, the details of the main experiment are provided.

5.3.1 Experiment Details

The purpose of the main experiment is to compare the comprehensibility aspects of a graphical representation and the controlled natural language representation of a task hierarchy, both of which are generated by the Visual Planning Tool explained in Chapter 4.

5.3.1.1 Participants

The experiment was conducted with 18 people. 9 of them had a bachelor degree on various departments. 5 among them were female participants with ages in the interval of 22-30. The remaining 4 of them were male participants again in the age intervals of 22-30. The other 9 people were working for a company on defence sector. All of them had a bachelor degree from different universities. 3 of them were male participants with ages above 24. The other six participants were female and all of them had ages above 24. The mean age for the all participants was 26. All of the 18 participants' data were collected and analysed in this experiment.

5.3.1.2 Experiment Procedure

The experiment was conducted in two sessions. One of the sessions was held in a company which is in the defence sector in Turkey. The other session was conducted in the Informatics Institute of Middle East Technical University.

At the beginning of each session, a short introduction was given on the scope and the purpose of the study. Then the participants were informed about the stages of the experiment. They were then asked to write a temporal hierarchy for the tasks presented to them as texts and graphical representations. They were also informed about the time criteria so that they would try to finish the experiment as fast as possible.

The experiment involved two different subtasks. In one, a textual document was provided to participants. In this document a task hierarchy in the form of Controlled Turkish was presented to participants. Then, after each participant finished and submitted the task hierarchy, the elapsed time was measured and noted. In the second subtask a graph which was informationally equivalent to the text in the fist subtask were presented to the participantsand they were again asked to generate a temporal-sequence for the task hierarchy. Again, they were informed about the time criteria. Like in the former session, each participant's answer was collected in papers and the elapsed time was noted on the papers they submitted.

5.3.2 Data Gathering Methods

The representations' computational efficiencies were evaluated according to how easily and quickly a participant can infer the relevant information. In order to measure this, correct responses and time were used as performance criteria. Correctness was also taken as a sign of comprehensibility.

5.3.3 Results

The results obtained from first groups of participants are provided below.

Table 24. First Session Results

Participant ID	CNL - Correctness Sequencing / Time Elapsed	Diagram - Correctness Sequencing / Time Elapsed
Participant_1	Incorrect / 110 sec.	Correct / 103 sec.
Participant_2	Correct / 153 sec.	Correct / 108 sec.
Participant_3	Correct / 147 sec.	Correct / 72 sec.
Participant_4	Correct / 97 sec.	Correct / 63 sec.
Participant_5	Incorrect / 174 sec.	Correct / 122 sec.
Participant_6	Correct / 117 sec.	Correct / 108 sec.
Participant_7	Correct / 66 sec.	Correct / 72 sec.
Participant_8	Correct / 112 sec.	Correct / 91 sec.
Participant_9	Correct / 108 sec.	Correct / 101 sec.

Results obtained from the second group of participants are provided below.

Table 25. Second Session Results

Participant ID	CNL - Correctness Sequencing / Time Elapsed	Diagram – Correctness Sequencing / Time Elapsed
Participant_1	Correct / 197 sec.	Correct / 65 sec.
Participant_2	Correct / 62 sec.	Correct / 53 sec.
Participant_3	Correct / 65 sec.	Correct / 47 sec.
Participant_4	Correct / 57 sec.	Correct / 49 sec.
Participant_5	Correct / 125 sec.	Correct / 72 sec.
Participant_6	Incorrect / 147 sec.	Correct / 76 sec.

Participant_7	Correct / 117 sec.	Correct / 73 sec.
Participant_8	Correct / 121 sec.	Correct / 66 sec.
Participant_9	Correct / 92 sec.	Correct / 79 sec.

5.3.4 Analysis of Results

As it can be seen from the tables provided in the section 5.3.3, all participants gave correct sequencing when the action sequencing was provided as graphical representations. However, some of the participants gave incorrect results when the sequencing was provided in Controlled natural language form.

In both sessions, the experiment took at least 60 seconds for each participant. The participants in second session were faster than the participants that took place in the first session on average. It took an average of 90 seconds for graphical representations and it took an average of 120 seconds for controlled natural language representations in the first session. In the second session, it took an average of 100 seconds for Controlled Turkish representation and an average of 65 seconds for graphical representation.

The results above showed that participants were more successful in sequencing in graphical representation. Moreover, when we look at the time for performance criteria, we can easily see that sequencing in graphical representation almost in all the cases took less time compared to sequencing in Controlled Turkish representation.

5.3.4.1 Statistical Analysis of Results

Similar to the data collected for the directed and undirected groups, in order to determine the normality of the data, Kolmogorov Smirnov test is applied. The results of the Kolmogorov Smirnov test for the time performance and the correctness are given below for $p = .05$.

Table 26. Kolmogorov-Smirnov for Normality.

	CONTROL_GROUP	Kolmogorov-Smirnov		
		Statistic	df	Sig.
TIME_PERFORMANCE	CNL_GROUP	.119	18	.200*
	DIAGRAM_GROUP	.165	18	.200*
CORRECTNESS	CNL_GROUP	.501	18	.000

CORRECTNESS is constant when CONTROL_GROUP = DIAGRAM_GROUP. It has been omitted.

The results in the table above can be reported as follows.

- The correctness for the CNL group, D (18) = 0.501, p<.05
- CORRECTNESS is constant when CONTROL_GROUP = DIAGRAM_GROUP. It has been omitted,
- The time performance for the CNL group, D(18) = 0.119, p<.05 and
- The time performance for the DIAGRAM group, D (18) = 0.501, p<.05.

According to the results the sig value for the TIME_PERFORMANCE for the CNL_GROUP and DIAGRAM_GROUP are greater than the .05 indicating that the data is normally distributed within these group. However the situation is different for the CORRECTNESS data having the value .000 which is smaller than .05 indicating that the data is not normally distributed.

Another property of the data that need to be explored is the homogeneity. In order to determine if the data is homogeneous with in the groups, Levene test has to be applied. The results for the Levene test are given in the table below.

Table 27. Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
TIME_PERFORMANCE	Based on Mean	3.242	1	34	.081
	Based on Median	3.446	1	34	.072
	Based on Median and with adjusted df	3.446	1	28.071	.074
	Based on trimmed mean	3.287	1	34	.079

CORRECTNESS is constant when CONTROL_GROUP = DIAGRAM_GROUP. It has been omitted.

The results for the Levene test can be reported as:

- For the time performance, the variance for the directed and undirected groups are not significantly different F (1, 34) = 3.24, p < .05
- Since the correctness data is not normally distributed among directed and undirected groups the Levene test is skipped for the correctness value.

According to the results the sig values for the TIME_PERFORMANCE data are greater than the .05 value indicating that the variances are not significant. Thus the data is homogeneous within the groups.

According to these results, the t-test for independent groups can be applied for the TIME_PERFORMANCE data. The results of the t-test are as follows.

Table 28 Group Statistics for t-test

	CONTROL_GROUP	N	Mean	Std. Deviation	Std. Error Mean
TIME_PERFORMANCE	CNL_GROUP	18	114.83	38.995	9.191
	DIAGRAM_GROUP	18	78.89	21.892	5.160

Table 29. Independent Samples Test for t-test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
TIME_PERFORMANCE	**Equal variances assumed**	3.242	.081	3.410	34	.002	35.944	10.541	14.523	57.366
	Equal variances not assumed			3.410	26.748	.002	35.944	10.541	14.307	57.582

According to the results the sig (2-tailed) value is .002 which is less than the .05 value indicating that the there is a significant difference between the means of the two samples.

By looking at the Mean values in the Group Statistics Table, the mean of the CNL_GROUP is 114,83 and the mean for the DIAGRAM_GROUP is 78.89 indicates that time to complete the task in CNL_GROUP takes more time than the DIAGRAM_GROUP. From this one can conclude that diagrams perform better than Controlled Turkish in terms of time.

Since the CORRECTNESS data is not normally distributed among directed and undirected groups. The t-test cannot be applied. Instead of t-test Mann Whitney U test will be applied to analyse the data for CORRECTNESS. The results of the test are given below.

Table 30. Ranks for Mann Whitney U Test

	CONTROL_GROUP	N	Mean Rank	Sum of Ranks
CORRECTNESS	CNL_GROUP	18	20.00	360.00
	DIAGRAM_GROUP	18	17.00	306.00
	Total	36		

Table 31. Test Statistics for Mann Whitney U Test

	CORRECTNESS
Mann-Whitney U	135.000
Wilcoxon W	306.000
Z	-1.784
Asymp. Sig. (2-tailed)	.074
Exact Sig. [2*(1-tailed Sig.)]	.406[b]
Exact Sig. (2-tailed)	.229
Exact Sig. (1-tailed)	.114
Point Probability	.114

According to the results the asymp sig value is .074 which is greater than .05 indicating that the results among directed and undirected groups are not significantly differ. To calculate the exact effect size, the formula of $r = Z / \sqrt{N}$

is used where r is the effect size; Z is the z score in given in the test results and the N is the number of the total observations. According to this the r = -1.78/ √36 = -0.29 which is less than the .3 threshold for a medium effect size.

In the light of these results, one can conclude that the CNL and its visual representation do not have significant effect according to the correctness.

5.4 Summary of Results

To sum up, the interpretations of the results for the preliminary study and the main experiment are respectively given as follows:

- The undirected diagrams performs better than the directed diagrams according to the correcteness performance criteria. However the directed and undirected diagrams have no difference in terms of time perfroamce criteria.

- Diagrams are more comprehensible than the Controlled Turkish according to the chosen time performance criteria. However according to the correctness performance criteria there is no significant difference among diagrams and Controlled Turkish.

CHAPTER 6

DISCUSSION AND CONCLUSION

The aim of this study was to compare a text for a task hierarchy given in Controlled Turkish with a graphical representation of the same task hierarchy for comprehension. The results of this study showed that participants were successful on graphical representations and they were faster and efficient with the graphical representation. Thus, although the text was presented in controlled language form, the participants were more comfortable with graphical representation.

In this study, the specific graphical representation adopted for a task hierarchy showed to be more successful than its Controlled Language representation. One of the reasons might be that we tried to transform plain Turkish into Controlled Turkish. This specific Controlled Turkish design might be poor since there is no Controlled Turkish study in the literature to be compared. If a more sophisticated Controlled Turkish was used, then the results could change.

Other factors should also be regarded while inspecting the performance of Controlled Turkish and graphical representation. The success of Controlled Language may be close to the performance of graphical representation when the factors such as the extent and complexity of the messages were taken into consideration. For instance, longer messages might be more readable in its textual representation than its mapping graphical representation since the graphical representation needs more visual elements to express the same stuff and thus be more crowded and difficult to comprehend.

One issue related to the Controlled Turkish generation is that the CPM rule set was taken as reference. In the original CPM's Controlled Language, discourse elements are not used and sentences are generated explicitly in order to eliminate ambiguity. Also, in this study one of the rules was the disallow of pronoun usage and the sentences were generated explicitly. For instance, in the textual representation given in Section 5.3, the name of the tasks are provided explicitly for the sake of clarity (e.g. "Adı "Görev_A" olan bir görev tanımlıdır"). Although such rules reduce ambiguity, they may result in superfluous structures. If such rules were eliminated complexity could be reduced and this may result in higher task performance. Thus, a further experiment can be carried out on the alternative Controlled Turkish texts.

One of the toughest issue in this study was to decide how to visually represent the Controlled Turkish statements. Controlled Turkish statements could be visually represented in many ways. Since the notations used in visual representations is crucial for comprehensibility, the notations could not be chosen arbitrary. For this reason, a visual notation that was proven in the military planning domain for representing task hierarchies should be used. Literature survey showed that visual representation proposed in CPM was the most appropriate and proven notation. Thus, the visual representation proposed in the original CPM model was taken as reference. To ensure that the original CPM graphical representation was easy to understand, the CPM visual representation was examined and the problematic semantic notations used in the visual display were identified; the visual representation was modified and enhanced according to rules of thumbs given in the "Cognitive Background" chapter. The resulting visual representation was finalized by an eye tracking experiment. The results of this experiment showed that the original

CPM visual display resulted in comprehension difficulties and led to interpretation complications, thus the modified representation was used. The reason of comprehension difficulties in the original graphical representation can be interpreted according to “Structure Mapping” subsection of Section 3.5 which states that if the structure of the picture is not well suited to the expressed concept, then the mental model constructed conflicts with the pictorial structure and this latens the comprehension process. This can be the case in the original CPM representation where the arrowheads on the edges led to misinterpretations and conflicts with the stereotypes given on the edges. Another inference on this reason can be made according to “Redundancy” subsection of Section 3.5 which states that when learners can construct mental models from textual or visual representation alone, further information does not lead to better understanding. Moreover, such a case leads split of attention between the text and picture. In the original CPM representation, one could claim that redundancy was caused because of the existence of both the directed edge and its stereotype. According to Tversky et.al (2011), arrows can map too many concepts such as order, direction, movement, causality etc. Thus, visual elements such as lines and arrows used in CPM could have multiple meanings and contain ambiguity. Since CPM uses directed lines with relationship tags, this situation can lead to ambiguity. Conflicts may occur while creating conceptual mappings for the visual elements. Another reason can be interpreted according to “Cognitive Cost” subsection of Section 3.5 which states that multiple external information source increases the comprehensibility. But, each different source of information also means cost to transfer this information and analyse them. In the original CPM representation the same information was given by the directed edges and the stereotypes on those edges led to further perception time, split of attention and more time to process the perceived information to construct a mental model. Thus, in order to eliminate the potential ambiguity, an experiment is conducted to measure the efficiency of the directed lines used in the original CPM model. As a result, the undirected lines performed better than the directed lines when used with relationship tags.

An important concern in the main experiment was related to the issue that in which order the tasks of the experiment, namely the task in textual representation and the task in graphical representation, should be given. In main experiment, participants were first given the Controlled Turkish representation and then they were given the visual representation. According to the "Contiguity" subsection in the Section 3.5, providing the textual representation before visual representation is not appropriate. Since one of the main aims of this study was to measure the comprehensibility of the proposed Controlled Turkish, the visual representation was used as a control element and in order to eliminate such unreliable increases in comprehensibility of the Controlled Turkish, the textual representation was given first. If the stimuli was presented in the reverse order, that is if the graphical representation was given before textual representation, then the experiment could end up with unreliable results since providing the graphical representation before the textual one might ease the comprehension process for textual representation according to "Contiguity" principle.

This result of this study, which is the graphical representation is better than Controlled Turkish for comprehension, can also be explained by referring to the ITPC model given in Section 3.5. According to ITPC model, visual representations seem to have an advantage over textual representations. As it can be seen from

Figure 5, the textual and visual representations are both transferred to the short-term memory from the same channel, namely the visual channel by using the same registers, the eye/visual registers. The situation begins to differ from this point. In pictorial representation the transferred information in the short term memory is directly used to construct a mental model. But, in the verbal representation case, the information transferred is subjected to further processing to construct a propositional model and then the mental model is formed from this propositional model. These extra processing steps in textual comprehension process can be thought to be responsible for the extra time needed to comprehend a textual representation. Another possible explanation for this situation could be as follows: Since the experiment content was simple, its corresponding visual representation was also simple. It contained a small set of visual elements so that the reader could easily transfer all of the visual elements into the working memory. But, the textual representation contained many words, which are also visual elements that need to be transferred into working memory. Thus, it took longer to transfer and interpret all these textual information.

As mentioned above, the diagrams used in this study were trivial and did not require much prior knowledge. The reason for keeping experiments simple was to decrease the experiment duration and increase the number of subjects.

According to what is mentioned so far, this study can be extended in a way to test how the behaviours of the subjects change when the tests get more complicated. That is, the experiment may be conducted with a diagram and its mapping Controlled Turkish text containing 20-50, 50-100 and over 100 visual elements. In such cases, diagrams will require more effort to be transferred to the working memory and as a result, subjects would probably need more time to construct mental models. By this way, an optimum number of visual elements for comprehensibility can be obtained for a diagram compared to text based representation.

In addition to increasing the number of the visual elements in the sample case, increasing the number of visual element types is also an important factor that needs to be evaluated. According to this, the diagram used in the experiment may be changed to include more types of visual elements. In that case subjects will have more item types to interpret. Thus, constructing mental model from the visual mappings would probably get more complex.

Another factor that needs to be further investigated is the effect of adding another information channel to the experiment according to the ITPC Model. In order to investigate this, the experiment can be expanded by adding another comparison branch such as a diagram supported with spoken text. By this way, Controlled Turkish, its corresponding diagram with and without spoken texts can be compared. This situation may increase the comprehensibility of the diagram especially if it contains many visual elements.

This study can be further extended by asking participants to sketch patterns for the textual representation. By this way, participants' comprehension process can be analysed in depth. Sketches have advantage over written language since they have the advantage of transferring the visuospatial ideas directly (Tversky, 2002). By sketching, participants can express abstract entities by using visual elements. This process also promotes memory and makes comprehension and inference easier. Moreover, the quality of textual representation can be investigated.

As a conclusion, the simple graphical representations adopted in this study proved to be more promising in expressing stuff at short notice than its mapping Controlled Turkish representations. However, the output of this study should be inspected elaborately in different cognitive dimensions.

REFERENCES

Adelberg, A. H. & Razek, J. R. (1984). The Cloze procedure: a methodology for determining the understandability of accounting textbooks. The Accounting Review, 59 (1), 109–122.

Adriaens, G. (1992). Proceedings of the 14th International Conference on Computational Linguistics ‘92: From COGRAM to ALCOGRAM: Toward a Controlled English Grammar Checker. Nantes, France.

Allen, J. A., Mott, D., Bahrami, A., Yuan, J., Giammanco, C., & Patel, J. (2008). Proceedings of the Second Annual Conference of the International Technology Alliance ‘2008: A Framework for Supporting Human Military Planning. London, UK.

E. Fuchs, Norbert. 2013. Attempto Project. Retrieved from http://attempto.ifi.uzh.ch/site/.

Ateşman, E. (1997). Türkçe’de Okunabilirliğin Ölçülmesi. Dil Dergisi, 58, 71-74.

Atkinson, C, & Shiffrin, R. M. (1971). The control of short-term memory. Scientific American, 225, 82-90.

Baddeley, A. D. (1986). Working memory. Oxford, England: Clarendon Press.

Bezirci, B & Yılmaz, A. E. (2010). A Software Library for Measurement of Readability of Text and a New Readability Metric for Turkish. DEU Engineering Faculty Scientific Journal, 12, 49-62.

Bormuth, J. R. (1968). Cloze Test Readability: Criterion Scores. Journal of Educational Measurement, 5(3), 189-196.

Bozşahin, C., Zeyrek D. & Pembeci I. (1998). Proceedings of Joint Conference ACH ALLC '98: Computer-aided Learning of Turkish Morphology. Debrecen, Hungary.

Brookings, J. B. Wilson, G. F. & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. Biological Psychology. 42, 361-377.

Bruneau, D., Sasse, M. A. & McCarthy, J. D. (2002). Proceedings of the CHI'02 Workshop on Physiological Computing '2002: The eyes never lie: The use of eye tracking data in HCI research. NY.

Byrne, M. D., Anderson, J. R., Douglas, S. & Matessa, M. (1999). Proceedings of CHI'99: Eye tracking the visual search of click down menus. NY.

Carney, R. N. & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. Educational Psychology Review. 14, 5-26.

Cowen, L., Ball, L. J., & Delin, J. (2002). Proceedings of HCI '2002: An eye movement analysis of web-page usability. London: UK.

Dale, E. & J. S. Chall. (1949). The concept of readability. Elementary English 26 (23).

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. Journal of Experimental Psychology: Learning, Memory, and Cognition. 9, 561-583.

Doherty, S. (2012). Investigating the Effects of Controlled natural language on the Reading and Comprehension of Machine Translated Texts: A Mixed-Methods Approach. (Doctoral dissertation). Retrieved from Machine Translation Archive.

Dubay, W.H. (2004). The Principles of Readability. Impact Information. Costa-Mesa, CA

Engelhardt, Y. (2007). Syntactic Structures in Graphics. Image Journal of Interdisciplinary Image Science (5), Special issue on Computational Visualistics and Picture Morphology, 23-35.

Ericsson, K. A. & Simon, H. (1993). Protocol analysis: Verbal reports as data. Cambridge, UK: MIT Press.

Ericsson, K. & Simon, H. (1987). Verbal reports on thinking. Introspection in Second Language Research. (24–54). Clevedon, Avon.

Ericsson, K. & Simon, H. (1980). Verbal reports as data. Psychological Review. 87 (3), 215–251.

Flesch R. (1948). A New Readability Yardstick. Journal of Applied Psychology. 32, 221-233.

Frank, S. Koppen M., Noordman, L. & Vonk, W. (2007) Modeling Multiple Levels of Text Representation. Discourse Studies. (2-29). Tilburg University, Netherlands.

Friedman, N. P. & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. Journal of Experimental Psychology, 129, 61-83.

Goldberg, H. J. & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. International Journal of Industrial Ergonomics, 24, 631-645.

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002). Proceedings of the Eye Tracking Research and Applications Symposium ‘2002. Eye tracking in web search tasks: Design implications. NY: US.

Goldberg, H. J. & Wichansky, A. M. (2003) Eye tracking in usability evaluation: A practitioner’s guide. The mind's eye: Cognitive and applied aspects of eye movement research. Amsterdam: Elsevier Science B. V.

Göpferich, S. (2009). Comprehensibility assessment using the Karlsruhe Comprehensibility Concept. The Journal of Specialised Translation, 11, 31-52.

Graesser, A. C., Millis, K. K. & Zwaan, R. A. (1997). Discourse comprehension. Annual Re-view of Psychology, 48, 163-189.

Gunning R. (1952). The Technique of Clear Writing, New York, US: McGraw-Hill Press.

Haak van den, Maaike J., Jong de, Menno D.T. & Schellens, Peter Jan. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. Behaviour & Information Technology, 22 (5), 339-351.

Ibbotson, John. (2012). The ITA Collaborative Planning Model (CPM).

Kintsch, W. (1998). Comprehension: A paradigm for cognition. Cambridge, UK: Cambridge University Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95, 163-182.

Kintsch, W. & Van Dijk, T.A. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Klare, G.R. (1984). Readability Handbook of Research. New York and London: Longman.

Klarę, G.R. (1963). The Measurement of readability. Ames, IA: Iowa State University Press.

Kosslyn, S. M. (1994). Image and brain. Cambridge, MA: MIT Press.

Kuihavy, R. W, Stock, W A. & Caterino, L. C. (1994). Reference maps as a frame-work for remembering text. Comprehension of graphics, 153-162. Amsterdam: Elsevier Science B. V.

Larkin, J. & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. Cognitive Science, 11, 65-99.

Leo Lentz en Menno de, J. (2005). Proceedings of the International Professional Communication Conference ‘2005. Anticipating reader problems: cognitive shortcuts that prevent experts from focusing on readers. Limerick: Ireland.

Mayer, R. E. (2009). Multimedia learning (2nd ed). New York: Cambridge University Press.

McLaughlin G. H. (1969). SMOG Grading-A New Readability Formula. Journal of Reading, 12(8), 639–646.

McLaughlin, G. H. (1968). The Third International Reading Symposium ’68. Proposals for British readability measures. London: Cassell.

Mott, D. & Giammanco, C. (2008). Proceedings of the 2nd Annual Conference of the International Technology Alliance ‘2008. The Use of Rationale in Collaborative Planning. London: UK

Mott, D. (2011a). CPM: Visual Guide to the CPM v3 [PowerPoint Slides]. Retrieved from https://www.usukitacs.com/node/1712.

Mott. D, (2011b). CPM Guide to Objects and Relations (CPM 3.3h) v9 [PowerPoint Slides]. Retrieved from. https://www.usukita.org/sites/default/files/GuideToCPM_9.ppt, 2011b.

Paivio, A. (1986). Mental representations: A dual coding approach. Oxford, UK: Oxford University Press.

Poole, A. & Ball, L. (2005). Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future. In Encyclopedia of Human-Computer Interaction. Pennsylvania: Idea Group, Inc.

Namahn. (2001). Controlled natural languages: A Research Note by Namahn. Retrieved from. http://www.namahn.com/resources/documents/note-CL.pdf.

Handbook of Statistical Methods. (2003). Retrieved from. http://www.itl.nist.gov/div898/handbook/. NIST/SEMATECH’s

O’Brien, S. (2003). Controlling Controlled English: An Analysis of Several Controlled natural language Rule Sets. Dublin City University, Ireland.

Ronald A. G. (1987). Predicting Readability of Data Processing Written Materials. Database, 18(4), 40-47.
Schnotz, W. (2009). An Integrated Model of Text and Picture Comprehension. In R.E. Mayer (Ed.), The Cambridge Handbook of Multimedia Learning. (pp. 1-16). New York: Cambridge.

Schnotz, W & Bannert, M. (2003). Construction and interference in learning from multiple representations. Learning and Instruction, 13, 141- 156.

Schriver, K. A. (1989). Evaluating Text Quality: The Continuum from Text-Focused to Reader-Focused Methods. IEEE Transactions on Professional Communication, 32 (4).

Senter R. J., Smith E. A. (1967). Automated Readability Index. Technical Report, Cincinnati University, Ohio.

Sowa, J. (2007). Common Logic Controlled English, retrieved from http://www.jfsowa.com/clce/clce07.htm.

Sutaria, M.C. (1965). The Basic Readers for EnglishTeaching. Quezon City: Phoemix Press.
Steinman, L. (2002). Considering the cloze. Canadian Modern Language Review, 59(2), 291-301.

Tenbrink, T. (2008). The verbalization of cognitive processes: Thinking-aloud data and retrospective reports. University of Bremen, Germany.

Tversky, B (2002). What do Sketch say about Thinking. AAAI Spring Symposium, Sketch Understanding Workshop, Stanford University, AAAI Technical Report SS-02-08.

Tversky, B., Corter, J.E., Y, L., Mason, D. and Nickerson, J.V. (2011). Visualizing Thought: Mapping Category and Continuum. In Proceedings of the 33rd Annual Conference of the Cognitive Science Society (Cogsci'11).

Van Dijk, T. A. & Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic Press.

ANNEX –A

WRITTEN TRANSCRIPT

	Task 1	Task 2
Participant 1	Önce olay A olmuş. Bu olay B'nin olmasına neden olmuş. --- (1 sec) Olay B'de olay C'ye neden olmuş. --- (1 sec) Aynı anda olay B, olay C ile olay E'ye --- (1 sec) neden olmuş ve olay E olay D'ye --- (1 sec) olay D'de olay F'ye neden olmuş.	Bence burada olay F'den sonra olay D başlar. Olay D'den sonra olay E başlar. Olay E ile aynı anda olay C başlar.(--- 2 sec) Olay C (--- 1 sec) 'den sonra olay B. Olay B'den sonra ise olay A başlar.
Participant 2	Olay A, olay B den önce (--- 1 sec); olay B, olay C'den önce oluyor. Olay C ve olay E aynı anda oluyor. (--- 1 sec) Olay E, olay D'den önce oluyor ve olay D de olay F'den önce gerçekleşiyor.	Olay A, olay B'den (--- 6 sec) Olay A, Olay b'den önce oluyor. Burda (--- 1sec) Olay B, olay C'den önce oluyor. C ve E aynı anda başlamışlar. (---4 sec) D, E'den sonra başlamış yani E önce oluyor (---2 sec) Burada da yine D (---2 sec) F'den önce (--- 2 sec) olmuş.
Participant 3	A olayından sonrasında B olayına geçilmiş Daha sonrasında C (---2 sec) B olayından sonra C olayıyla E olayı aynı anda olmaktadır. Daha sonra bu iki olaydan sonra tekrar D olayına geçilmektedir. D olayından sonra da F olayına geçilmektedir.	İlk başta olay F ile başlıyor daha sonra (---1sec) olay F'den sonra olay D olmakta (---1sec) Olay D den sonra olay E başlıyor. Olay E ile birlikte olay C aynı anda başlıyor Olay C'den sonra olay B gerçekleşiyor Olay B'den sonra da en son olay A gerçekleşiyor.
Participant 4	Olay A, olay B'den önce. Olay B, olay C'den önce Olay C, olay E ile aynı zamanda, olay	Olay F, olay D den sonra başlar. Olay D, olay E'den (---1sec) ve olay C'den sonra başlar.

	D’den önce Olay D, olay F’den önce	Olay C, olay B den sonra başlar, Olay B (---1 sec) olay A’dan sonra başlar.
Participant 5	Önce Olay a. Hepsinden önce Olay A gerçekleşmiş. Daha sonra olay B, olay C, olay E, olay d ve son olarak da olay F (---1 sec) gerçekleşmiş. Hepsi de ardışık olarak ilerliyor.	Burada ise tersi mevcut. Önce olay F olmuş, olay F’den sonra olay D, olay D‘den sonra olay E, olay E’den sonra olay C, olay C’den sonra olay B, ve son olarak da olay A gerçekleşmiş. (--- 8 sec) olay C ve olay E aynı anda başlamış.

Table 32. Written Transcript

ANNEX-B

CONTROLLED TURKISH REPRESENTATION FOR SAMPLE TASK HIERARCHY

A sample is in the attachment.

Task Hierarcy.docx

ANNEX-C

EXPERIMENT

Experiment.rar

ANNEX-D

TOOL FOR GENERATING GRAPHICAL REPRESENTATIONS AND CORRESPONDING CONTROLLED TURKISH STATEMENTS

Visualization Tool.rar