

Performance Evaluation for Universal Grammar Engine

Silu Cao

Student ID: 0653808

**A dissertation submitted to Auckland University of Technology in partial
fulfillment of the requirements for the degree of Master of Computer and
Information Sciences (MCIS)**

2011

School of Computing and Mathematical Sciences

Supervisor: Albert Yeap

Table of Contents

Attention of Authorship	VI
Acknowledgements	VII
Abstract	VIII
1. Chapter 1: Introduction	1
1.1. Background and Motivation	1
1.2. Scope and Objectives	3
1.3. Dissertation Structure	4
2. Chapter 2: Literature Review	5
2.1. Basic Framework of UGE	5
2.2. Stanford parser and Link Grammar	7
2.3. Evaluating natural language parsers	11
2.4. Summary	13
3. Chapter 3: Research Methodology	14
3.1. Collecting the test data	14
4. Chapter 4: Implementation	17
4.1. Experiments and Results	19
4.2. Random Test Result	43
4.3. Summary	46
5. Chapter 5: Conclusion, Limitation and Future Direction	47
6. References	49
7. Appendix: random testing data set	52

List of Figures

Figure 2.1 Parse result generated by Stanford parser.....	8
Figure 2.2 Stanford parser graphical result.....	9
Figure 2.3 Link parser correct parse example.....	10
Figure 2.4 Link parser failed parse example.....	10
Figure 2.5 Parsing result generated by Link parser.....	10
Figure 4.1 Example of experiment environment setup.....	17
Figure 4.2 Adjective patterns	20
Figure 4.3 Verb Patterns	22
Figure 4.4 Be Verb Patterns	25
Figure 4.5 Auxiliary Verb Patterns	28
Figure 4.6 Infinitive to Patterns	30
Figure 4.7 Preposition Patterns	31
Figure 4.8 Gerund/Participle Patterns	34
Figure 4.9 Adverb Patterns	37
Figure 4.10 Article Patterns	39
Figure 4.11 Connective Patterns	41

List of Tables

Table 2.1 Stanford parser word NP subtree	8
Table 2.2 Stanford parser word label	8
Table 2.3 Link parser word formula examples	9
Table 4.1 Adjective patterns and examples.....	20
Table 4.2 Verb patterns and examples	23
Table 4.3 Be Verb patterns and examples.....	26
Table 4.4 Auxiliary Verb patterns and examples.....	28
Table 4.5 Infinitive patterns and examples	30
Table 4.6 Preposition patterns and examples	32
Table 4.7 Gerund/ Participle Patterns and examples	35
Table 4.8 Adverb patterns and examples	37
Table 4.9 Article patterns and examples.....	40
Table 4.10 Connective patterns and examples	42
Table 4.11 Results obtained from random testing	43
Table 4.12 Connective patterns and examples	45

List of Abbreviation

1. AUT: Auckland University of Technology
2. AI: Artificial Intelligence
3. CR: Clausal Representation
4. NLP: Natural language processing
5. UGE: Universal Grammar Engine
6. N: noun
7. Pron.: pronoun
8. Adj.: adjective
9. Num. : numeral
10. V.: verb
11. Adv.: adverb
12. Art.: article
13. Prep.: preposition
14. Conj.: conjunction

Attention of Authorship

“I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university of other institution of higher learning.”

Yours sincerely,

Silu Cao

2010

Acknowledgements

There are many people who supported, encouraged and accompanied me to the completion of my master's dissertation report. I express my most sincere appreciation to these individuals.

In particular, I am indebted to my supervisor, Albert Yeap, for his continued encouragement and invaluable suggestions during my research. He provided rigorous feedback and guided me to walk through the challenging steps from research proposal, conducting the research and finally to completing the report.

I would like to thank my colleagues and friends, Thamalini Arunachalam, Lasiew Chen and Waseem Ahmad, Anuj Bhowmik for their encouragement and assistance to help me solve various problems during my research.

Finally, I express my thanks to my parents in China for their support, trust, and to cheer me up during the tough time.

I would like to acknowledge the generous support of Auckland University of Technology for providing me with a good study environment and condition.

Abstract

The technology of natural language processing (NLP) in the field of Artificial Intelligent (AI) research has made much progress, enabling computers to parse human languages both in spoken and written form. Program that recovers the underlying syntactic structure of a sentence is known as a parser. This project focuses on the evaluation of one such parser developed by Professor Yeap. The parser is codenamed the Universal Grammar Engine (UGE) and it is developed to explain how children understand language. UGE is a newly developed parser. Its rules and performance are in need of evaluation.

Compared to traditional parser, UGE is a very different parser. The traditional parser implements a formal grammar of the language, whereas UGE classifies words in terms of how the information in each word is passed onto adjacent words for constructing the meanings of the phrase. Thus each word is encoded with a combination of ?L+ (accept words from its left), ?L- (pass this word to the left), ?R+ (accept words from its right), ?R- (pass this word to the right), and variations of these basic labels. It also has a set of rules for manipulating such words. While these rules are different from the formal rules used in traditional parsing, the labeling of words does correspond roughly to the syntactic categories of words. For example, a word with a ?R- functions as an adjective. Consequently, to test the performance of the rules developed in UGE, I take the syntactic category of each word label in UGE and find all possible sentences using its formal grammar. In particular, I consider many sentences that could be generated by using a combination of ten syntactic categories, namely: adjective, verb, be verb, auxiliary verb, infinitive to, preposition, gerund/participle, adverb, article and connective. This method tested the “completeness” of UGE as a parser. However, due to the creative use of language, sentences obtained from a formal grammar are not complete and there is no such set of complete test sentences for testing parsers. Consequently, a second test set is created for testing UGE and this set consists of almost four hundred real-world sentences randomly selected from newspaper articles.

My evaluation showed UGE performed well and it is a robust system. It parsed all the sentences in the first test and more than eighty percent of the second test

1. Chapter 1: Introduction

This chapter presents the background and motivation for this research that involves analyzing the performance of a language parser called UGE (Universal Grammar Engine). UGE is developed by Professor Yeap as part of his theory of how children acquire their first language (Yeap, 2005a). This chapter also presents the scope and objectives for this research and concludes with a synopsis of the thesis.

1.1. Background and Motivation

Natural language processing (NLP) is both an important and a popular field of research in Artificial Intelligence (AI) that uses computers to analyze, understand and represent human languages. Human languages used as an input for NLP system can be text, spoken or individual sentences. However, different input format brings with it extra dimensions to the problem of analysing sentences. For example, for text processing, one has to resolve anaphors and perform discourse analysis (Grosz & Sidner, 1986), both of which involve processing relations between sentences. For spoken sentences, one has to deal with incomplete sentences, repetitions, pauses, ellipses, and other issues which are rarely found in written sentences (Miller & Weiner, 2009). For simplicity, I will focus on testing UGE using only written sentences in my research.

In the early days of AI research into natural language processing, AI researchers (e.g. Ritchie & Thompson, 1983) have proposed the idea of developing “semantics” parsers, whereby the emphasis was on the use of semantics rather than syntax to parse a sentence. However, the implementation of the proposed idea had never been successful. Moreover, it is discarded due to the ambiguity problem of word representation and control. The ambiguity problem refers to successful meaning extraction of words in sentences (Allan, 2003). For example, the word “fly” has different meanings and grammatical classes in the dictionary. It is difficult to identify the most suitable meaning effortlessly by a computer system. The control problem refers to the order a grammar imposes upon the interpretation of a sentence. Without the grammar order control, a sentence like “The man bit the dog” would fail to be interpreted properly because our commonsense knowledge tells us that it is often a dog biting a man and not the reverse. Early efforts to produce a semantic parser

were not very successful. Subsequently, syntactic parsing became the default first step in any NLP program and this involves encoding the syntactic rules that govern the use of language. Many parsers were developed and implemented successfully based on syntactic parsing, for instance, Link parser (Naidu, Sharma & Bharati, 2009) and Stanford parser (Klein & Manning, 2003).

Yeap (2005a) rekindled the idea of semantic parsing. However, in his approach, he did not focus on deriving meanings of words in the sentences intentionally without the use of grammar rules. Rather, his effort was to develop a suitable framework to support semantic parsing. In other words, some rules are needed to recover the underlying structure of each sentence to support its interpretation. This is in contrast to the use of formal grammar rules to verify the correctness of the input as a legal sentence of the language. Yeap's research focused on what these rules might be. Moreover, he turned his attention to understanding how children acquire their first language. Studies on how children acquire their first language have raised the possibility that the grammar rules of language could not be learned, and yet all normal children have shown to have acquired them by the age of 5. This became, and still is, a paradox in language research (Pinker, 1989). A controversial solution was proposed by Chomsky (1965) who suggested that grammar knowledge is innately given. Till now, this idea – whether one's grammar is innately given or learned – is still being debated (for example, see Berwick & Chomsky, 2008; Clark & Eyraud, 2006; Kidd, 2004; Lidz & Gleitman, 2004). Yeap (2005b) proposed a solution to the paradox by arguing that what is needed is the learning of a different kind of language rules – one which focused on recovering the underlying structure of sentences for their interpretation rather than one which focused on verifying the correctness of the input according to some pre-determined rules (grammar rules). He also proposed a new algorithm to implement his solution.

The gist of Yeap's algorithm focuses on how the meanings of words appearing in a sentence are to be passed among themselves to produce an integrated meaning. Immediately, it is observed that there are four basic movements: passing a word meaning to its left, to its right, and receiving meanings from a word or a group of words on its left or on its right. These movements are encoded as part of the word lexicon as ?L-, ?L+, ?R-, and ?R+ respectively. However these basic movements cannot cope with the full

complexity of natural language, Yeap (2005b) showed how they could be refined to capture all the known syntactic categories identified by linguists. Therefore, it demonstrates that knowledge of grammar could be learned.

To combine the words, a set of rules are learned from observing how sentences are constructed. Unlike the formal rules, these rules are heuristics about language use. They grow in size as the rule may gain more experience of the way that a particular language is used. Thus, it is necessary ad hoc initially and there is no sense of completeness – like language, it is never complete or static; it keeps evolving. If a rule has not been learned, then a particular construct in the language would not be parsed correctly. Several years have been spent developing the existing rules for UGE for the English language.

UGE has now been tested for large variety of sentences and has shown to be quite powerful in dealing with real-world sentences. However, how good are the rules in UGE? How do we measure UGE's performance? Could we compare UGE with traditional parsers?

1.2. Scope and Objectives

The objective of this research is to evaluate UGE through using some useful criteria. However, what is the most useful criterion for evaluating a parser like UGE? In general, it is difficult to evaluate any natural language parser since it is impractical to set up a complete set of test cases for all known sentences (see discussion in the next chapter). With UGE, this is even more difficult since the rules are not fixed. Therefore, comparison of its performance with other rule-based parsers is difficult. Furthermore, the development of UGE is still at its infancy. Therefore, the comparison of performance might be pre-mature. Consequently, in discussion with my supervisor, I have chosen to evaluate UGE in terms of how well it handles ten known syntactic categories of the English language sentences: noun, adjectives, adverbs, verbs, be verb, auxiliary verb, connectives, preposition, gerund/participle, and article. These categories define the scope of my investigation.

A test set of more than three hundred sentences randomly drawn from English Grammar Handbooks and newspaper articles are used to test UGE. I used the English Grammar Handbooks to help me identify possible legal English sentences for each category so that I could test if UGE can also parse these sentences. Since we know that the real-world

sentences have more complicated structures, I used the newspaper sentences for random testing of UGE. The output of UGE is analyzed regarding to their handling of the above ten syntactic categories. In the analysis, particular attention is paid to the correct parsing of these cases and the nature of the output produced. By correct, it is meant that the output is judged appropriate by its design team. If it is deemed incorrect, then a bug is found in the UGE system. For the output, I analyzed to see if there is any discrepancy produced when the different categories are combined together in a given sentence.

In summary, the scope of this project is to evaluate how UGE handles ten syntactic categories of the English language sentences. More than three hundred sentences were used as test cases.

1.3. Dissertation Structure

This research report is organized into five chapters. Chapter one introduces background, motivations and objectives of the research. Chapter two presents a review of relevant research of NLP parsing system which includes the literature review of UGE, Stanford parser and Link parser. The techniques involved in NLP testing are also analyzed in the same chapter. Chapter three discusses the research methodology used and chapter four presents the design and implementation of the research objective which involved the evaluation of UGE. The evaluation presented is based on the testing result gained from sets of experiments associated with research objectives. Finally, Chapter five presents the conclusions for this research, aligns with the limitations of this project and provides recommendations for future work, such as the aspects where existing UGE needed to improve.

2. Chapter 2: Literature Review

As the objective of this research is to analyze the performance of a parser known as UGE, this chapter provides a review of UGE (section 2.1), a review of two existing parsers, Stanford Parser and Link Grammar (section 2.2), and a review of past work done on evaluating NLP systems (section 2.3). The first review will provide the necessary background for understanding UGE. The second review will briefly compare UGE with two publicly available parsers to show how different UGE is from traditional systems. The third review briefly looks at how NLP systems were evaluated in the past and in particular why it is so difficult to evaluate such systems and provide the rationale for the approach taken in this project.

2.1. Basic Framework of UGE

The basic idea of UGE is that the meaning of a sentence is obtained by combining meanings of individual words found in the sentence to the meanings obtained on the left or right of that word (Yeap, 2005a). For this purpose, each word must be tagged with information regarding how the word meanings could be moved (i.e. to the left or right or not at all). Furthermore, it must contain a set of rules for the language which is based upon how that language is used in a particular society. The parsing is done by using a stack. The result of parsing a simple sentence is shown below.

➤ (interpret '(Pretty girls laughed.))

```
[LAUGHED* (:ACTOR (GIRLS* (:NOUN
                                (:MODIFIER (PRETTY*)))))]
```

The tags encoded for each of the above words are as follows:

```
Pretty:      (?R- (:MODIFIER (PRETTY*)))
Girls:       (GIRLS* (:NOUN))
Laughed:     (LAUGHED* (:ACTOR ?L+))
```

There are four basic tags, namely ?R+, ?R-, ?L+, and ?L-. R/L indicates moving to the right/left of the word and -/+ indicate whether its content moved or contents from elsewhere are moved into it. These basic forms cover the following grammatical categories:

Adjectives:	(?R- (:modifier (beautiful*)))
Gerund:	(looking* (:what ?R+))
Infinitive-to:	(to* (:what ?R+))
Intransitive verb:	(blew* (:who ?L+))

Using a combination of these tags gives the following categories:

Be verb:	(be* (:actor ?L+) (:what ?R+))
Transitive verb:	(blew* (:who ?L+) (:what ?R+))

It turns out that other grammatical categories are special instance of the above. The variations are due to the way in which an extra constraint is put on the attachment. For example, a dative verb takes two arguments for its verb. Consequently, a new tag ?R++ is introduced. Five such variations have been suggested:

- ?R++: This implies it takes the second meaning from the right and is useful for datives
- ?R-*: This is for determiners – a kind of ?R- that must appear at the start of a ?R-series
- ?R-**: This is for pre-determiners – a kind of ?R- that can appear in front of determiners as well as any other ?R-
- ?L-*: Process the stack and attach the meaning of this word to the result. If the stack is empty, attach the meaning of this word to the word on its right
- ?L#: This is for adverbs. This implies that it processes the stack and attaches the meaning of this word to the result.

While all grammatical categories are covered by one of the above tags, it is not a one to one correspondence. For example, gerund is not distinguished from past participle verbs. That is, both have the tag ?R+. With these variations, those categories could now be defined:

Preposition: (?L* (:IN* ?R+=))
 Dative verbs: (give* (:actor ?L+) (:recipient ?R++) (:what ?R+))
 Adverb: (|?L#| (:MANNER (HEAVILY*)))

Routines for handling each tag usually begin with an elimination process and a combination process. An example of the former inside the routine for handling infinitive to includes the following rule: If the stack holds a verb with an entry “:recipient”, kill this stack. For example, consider parsing the sentence, “I gave the money to John”. When the parser reaches the word “money”, two stacks are produced as shown below:

- (interpret ‘(I gave the money to John))
 [GAVE* (:ACTOR (I* (:PNOUN)))
 (:WHAT (MONEY* (:MODIFIER (THE*)))))]

- [GAVE* (:ACTOR (I* (:PNOUN)))
 (:RECIPIENT (MONEY* (:MODIFIER (THE*))))
 (:WHAT ?R+)]

On encountering the infinitive to, the second stack above is eliminated.

2.2. Stanford parser and Link Grammar

Stanford parser

Stanford parser is a statistical parser which produces the most likely analysis of sentences by using the knowledge of language gained from hand-parsed sentences (Klein & Manning, 2003). The parser thus produces the correct grammatical structure of a sentence if the sentence is legal. It is implemented as a probabilistic natural language parser and outputs a tree tagged with part-of-speech and grammatical relations format. According to Rusu, Dali, Fortuna, Grobelnik and Mladeni (2007), through the process of Stanford parser, a sentence is presented as the root of the tree and tagged with S. The root(S) has three children which include a noun phrase (NP), a verbal phrase (VP) and the full stop. Each child has its own

subtree with its own descendent with specific attributes shown in the following NP subtree table as an example.

Table 2.1 Stanford parser word NP subtree

Subtree	The type of noun
NN	noun, common, singular or mass
NNS	noun, common, plural
NNP	noun, proper, singular
NNPS	Noun, proper, plural

According to De Marneffe, MacCartney and Manning (2006), there are two phases to parse a sentence: dependency extraction (parsing by phrase structure grammar parser) and dependency typing (label the extraction with a grammatical relation). The first stage intends to gather the words in a sentence into different group associated with grammar rules. The second stage is to label each word in the sentence with a grammatical attribute, for example:

Table 2.2 Stanford parser word label

Word	Attribute
I	PRP
love	VBP
apples	NNS

Each word is encoded with a specific label which must be the descendent of subtree NP and VP. Once the possessive relationship is found, the input sentence can be parsed, such as the example below:

<p><i>Sentence: I love apples.</i> Parsing result: (ROOT (S (NP (PRP I)) (VP (VBP love) (NP (NNS apples))))))</p>

Figure 2.1 Parse result generated by Stanford parser

Graphically, the output looks like:

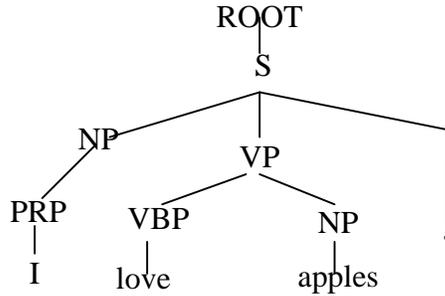


Figure 2.2 Stanford parser graphical result

Stanford parser is a typical example of how a parser using formal grammar rules is implemented. Its rules dictate how sentences are formed. UGE works with a different principle. Its rules dictate how meanings of words could be combined.

Link parser

The Link parser is another syntactic parser for English developed by Sleator and Temperley (1995). It is called a Link grammar because it builds relations (or links) between pairs of words rather than a constituent tree. The linking requirements of each word are contained in a dictionary and an example (from Sleator & Temperley, 1991, p.2) is shown below:

Table 2.3 Link parser word formula examples

Words	Formula
a, the	D+
cat snake	D- & (O- or S+)
chase, likes	S- & O+
Kate	O- or S+

Each word is thus encoded with labels that must be connected to a similar label on its left (-) or right (+). Thus the word “a” (which has a D+) could be attached to the word “snake” (which has a D-). Once all the words are connected, a correct parse of a sentence looks, in a diagram form, like:

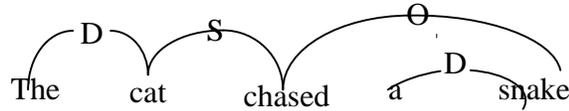


Figure 2.3 Link parser correct parse example

Thus a correct parse is one where the links are satisfied in such a way that they do not cross over as shown in the diagram above. In general, a successful linkage has to follow three conditions: Planarity (no cross link), connectivity (all words of the sequences have to be linked) and satisfaction (satisfy the link requirement). An example of a failed parse is as shown below:

Sentence: The Kate likes apples.

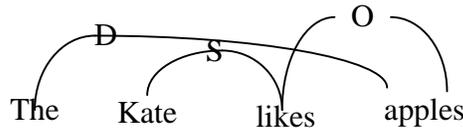


Figure 2.4 Link parser failed parse example

A link parser is publicly available and Figure 2.2 shows an output from running the program. Note that in the publicly available link parser, a constituent tree of the output is also produced.

```

Sentence: I love apples.
Parsing result:
++++Time                               0.00 seconds (577.41 total)
Found 1 linkage (1 with no P.P. violations)
Unique linkage. cost vector = (UNUSED=0 DIS=0 AND=0 LEN=3)
+-----Xp-----+
+-Wd--+Sp*i+---Op--+ |
| | | | |
LEFT-WALL I.p love.v apples.n .
Constituent tree:
(S (NP I)
  (VP love
    (NP apples))
.)
    
```

Figure 2.5 Parsing result generated by Link parser

A striking similarity of the Link parser with the UGE is its focus on left/right attachment of words. However, on closer inspection, the implementation of this basic idea is very different (Yeap, 2005b). Similar to Stanford parser, Link grammar is an implementation based upon the formal grammar rules of language, and hence it could be easily converted into a constituent tree. Whereas UGE is not based on the formal grammar rules and its output is different. This is also evidently clear as Sleator and Temperley (1991) stated explicitly in their abstract that what they have developed is a formal grammatical system. In contrast, Yeap (2005a) stated that UGE is not a formal grammatical system.

2.3. Evaluating natural language parsers

First and foremost, it is important to distinguish between a language parser and a language system. A language system consists other components in addition to having a language parser. For example, a Dialogue Management System is an NLP system that in addition to a language parser, it has other components such as dialogue management, dialogue planning, and dialogue generation. The testing of such systems poses a more serious problem since it might not be easy to isolate where a failure in the test lies. Examples of recent attempts to test such systems include the work of Barr (2003), Litman, Farrow and Dzikovska (2009), and Walker and Copperman (2010). This report is not about testing such systems. It is only concerned with the evaluation of a natural language parser.

There are many different criteria that a parser could be evaluated for. For example, most recently, Kakkonen (2007) listed five criteria: preciseness (concerning grammaticality or correctness of structural representations produced), coverage (ability to parse as many legal sentences as possible), robustness (ability to process noisy input), efficiency (measuring speed and memory usage), and subtlety (measuring how clever the system is). Some other criteria that researchers have used include:

King (1996): functionality (doing what it is supposed to do), and usability (easy to understand the information given);

Balkan, Arnold and Meijer (1993): diagnostic evaluation (detecting deficiencies), progress evaluation (comparing successive stages of development), and adequacy evaluation (meeting pre-specified requirements);

Guida and Mauri (1986): expressive power, adequacy and computational complexity.

At different stages of developing a parser and for different users' needs, one may focus on evaluating the parser based upon one or more of these criteria. For example, robust evaluation is important for a parser to be used in domains where one could expect significant errors in the text (e.g. processing text derived from speech). Robust evaluation of Link Grammar and the Stanford Parser described earlier were carried out in Kakkonen (2007) and an unsupervised method for evaluating parser robustness was described in Bigert et al. (2005). Balkan et al. (1993) noted that developers are chiefly interested in diagnostic and progress evaluation while users are mainly interested in adequacy evaluation.

Whatever the criteria used, one must evaluate the parser against a set of text data. One faces two major problems: the adequacy of the data collected and how to check if the output produced is correct. For the former, it is normal to use two kinds of test data: a test corpus which is a large amount of text collected from a given source (for instance, newspapers), and a test suite which is a list of specially constructed or selected sentences (Balkan et al., 1993). As in the development of UGE, a test suite was developed incrementally as the system was developed to test its handling of various syntactic constructs in a systematic way. Later, a large corpus of sentences was collected from daily newspaper to test its performance. The reason sentences from newspaper were used was because the developer wanted to test UGE with real-world sentences.

Checking the output of a parser to verify that a correct parse is obtained is time consuming. Consequently, treebanks or parsed corpuses containing sentences annotated with their correct parse results were set up. However, when it evaluates a new parser, re-using them is not a straightforward process unless the output of your system is the same as that used in the annotation (Hogenhout & Matsumoto, 1998). There is still one further complication when checking the correctness of a parse result. It is normal for a parser to come up with some alternative parses but most system will select what it thinks is the best parse. If it selects the incorrect one, the problem is not with the parsing of the sentence but rather with the selection of the best output.

2.4. Summary

The review of the UGE system and the more traditional parsers, Link Grammar and the Stanford parser, shows that there is a significant difference between them, both in terms of the way the sentences are parsed and the way in which the outputs are produced. Consequently, a different method of evaluating UGE is developed.

In view of the fact that the development of UGE is still at its infancy, it is suggested that one should evaluate it in terms of its functionality.

3. Chapter 3: Research Methodology

Since this project does not involve implementing a program or developing a prototype, traditional software development methodologies such as the RAD (rapid application development), waterfall, RUP (rational unified process), and others are not appropriate. Instead, a straightforward method of test and evaluate is used. This method would involve three key steps: (i) understanding UGE, (ii) collecting appropriate test data, and (iii) analyzing the results obtained. The first step is carried out in my review of UGE as reported in Chapter 2. This chapter discusses the second step. The next chapter discusses the third step.

3.1. Collecting the test data

One of the key problems for testing a natural language system based upon a formal approach is the adequacy of the test data. In particular, how does one know if the test data provides a sufficient sample for testing (Li, Dahl, Norton, Linebager & Chen, 1998)? This is a serious problem, because there is no clear way of knowing how the rules, formal or otherwise, are allowed to combine to form a sentence. In addition, it is also because the rules might not be complete. Human use of language is so creative that this poses a major problem.

For UGE, it is observed that the rules for handling each word category are primarily heuristics concerning how adjacent words are combined. For example, if the word has ?R-, one heuristics would be the object appearing on the right side of a verb. Note that the heuristics needed is more than “looking for an object on its right and pass this word to it”. The latter is a rule for constructing a result and the former is a rule for eliminating a choice. The reason an eliminating rule is needed is because a word can have more than one tag. For example, the word “run” can have no tag (i.e. a noun word) or a ?L+/?R+ pair (i.e. a verb). It is important UGE has sufficient heuristics to deal with the full complexity of language use. Unfortunately there is no way to know if this is the case. For these reasons, I created two different sets of data to test UGE’s performance.

The first set of data comes from considering a fixed set of 15 syntactic categories in the English language against each of the tag used in UGE. The set of 15 syntactic categories of

words used are: noun, preposition, adverb, conjunction, possessive case, verb, be-verb, aux-verb, article, infinitive-to, gerund/ participle, connective, adjective, double nouns, and clause. All legal sentences are then constructed and tested with UGE. Note that due to time constraint, not all tags have been tested and are created with the use of the formal grammar for English.

For example, consider creating a test set for evaluating the rules for ?R-. Since ?R- is identified as adjectives, I will look for all sentences whereby one of the 15 syntactic categories can appear on the right of an adjective. Figure 3.1.1 shows the attachment diagrammatically. In this example, 8 of the 15 categories produce meaningful sentences. Likewise, if there is a set of rules for handling words consisting of only a ?L+, then I will look for all possible sentences where the right side of those words is one of the fifteen categories.

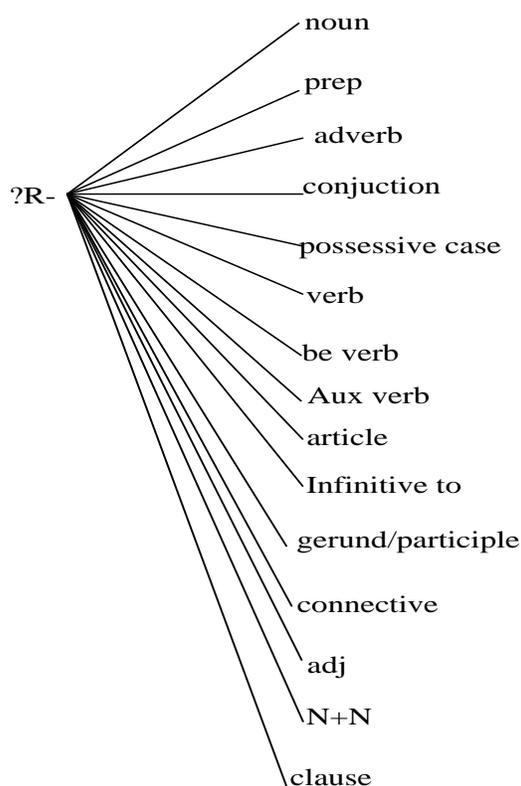


Figure 3.1.1 Evaluating Adjective attachment

To construct these sentences, various materials are used. One of the most important one is FLAX tool (FLAX, 2010), which is a website where I can find all the attachment possibilities for each category against the other categories. However, the tool itself does not necessarily generate an example sentence for a possible combination. Thus, I used the tool to find out what attachments are possible. Then, I use an English grammar book to find the necessary example sentences. One unique sentence for each construction is sufficient.

As mentioned, there are two sets of data used for testing the performance of UGE. The first test set is to find those legal sentences whereby one of the 15 categories can appear on the correct side of the test category. The second test set is meant to be a randomly collected set of real world sentences. For this, I use the New Zealand Herald website as a source. Newspaper articles are excellent examples of sentences constructed in the real world. To build up a reasonable test set, a minimum of twenty sentences are selected for each category.

There are two possible outcomes when evaluating UGE's performance: one is that UGE is complete and can parse all the test sentences. The other is that UGE is incomplete and there are still some errors or missing rules in UGE.

4. Chapter 4: Implementation

The UGE system was developed using Allegro Common LISP working under Microsoft Window XP operation system. To load and compile UGE, one has only to click “load.cl” under the SmartINFO root folder (see Figure 5.1).

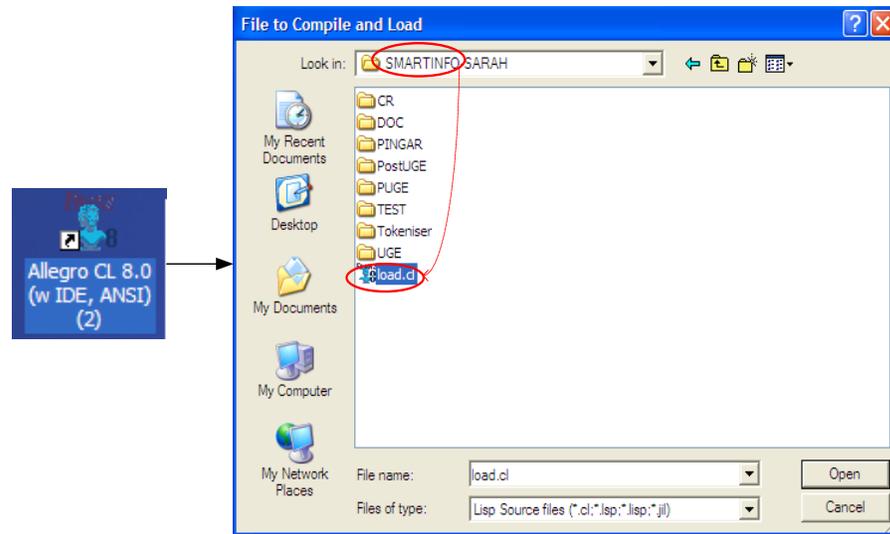


Figure 4.1 Example of experiment environment setup

The next step is to use the following commands to parse a sentence:

(pprint (uge” input sentence”)) – to get the best possible interpretation

(uge*”input sentence”) – to get all possible interpretations

(ugep”input sentence”) – to generate a trace of the parsing

For example : CG-USER(1): (pprint(uge”I eat an apple.”))

The output is:

```
[EAT* (:ACTOR (I* (:PNOUN)))
 (:WHAT (APPLE* (:NOUN) (:MODIFIER (AN*))))]
```

The above shows only the most preferred output generated. To get more details, we use another command which is `uge*`.

For example: `(uge*"I admit that the work is very hard.")`

Unknown Words: NIL

Number of MSs: 3

```
([ADMIT* (:ACTOR (I* (:PNOUN)))  
  (:WHAT (:THAT* (:MS1 (WORK* (:NOUN) (:MODIFIER (THE*)))  
    (:IS* (?R) (:MANNER (VERY*)) (:MANNER (HARD*)))))))]
```

```
[ADMIT* (:ACTOR (I* (:PNOUN)))  
  (:WHAT (:THAT* (:MS1 (WORK* (:NOUN) (:MODIFIER (THE*)))  
    (:IS* (HARD* (:NOUN)) (:MANNER (VERY*)))))))]
```

```
[ADMIT* (:ACTOR (I* (:PNOUN)))  
  (:WHAT (:THAT* (:MS1 (WORK* (:NOUN) (:MODIFIER (THE*)))  
    (:IS* (HARD* (:NOUN) (:MODIFIER (VERY*)))))))]
```

This command shows all outputs produced by UGE. The first output interprets both words, “very” and “hard”, as an adverb. The second and third output interprets the word “hard” as a noun but a different role for the word “very”. Using this command, one could check if UGE can parse the input text properly. The remainder of this chapter shows the evaluation of UGE on the following 10 different word categories: adjective, verb, be-verb, auxiliary verb, infinitive-to, prepositions, gerund/past-participle, adverb, article, and connective. Due to time constraints, some other categories (such as conjunction) are left out. For each case, I will show the meaningful sentences constructed for each possible combination and discuss the result of parsing these sentences using UGE. In the conclusion, UGE is tested against a set of random real-world sentences and its result is presented.

4.1. Experiments and Results

Adjective

An example of an adjective in UGE is: (?R- (:MODIFIER (BEAUTIFUL*))). Figure 4.2 shows that 7 out of 15 categories were never found to be present. Table 4.2 shows the test sentences constructed for testing UGE and the outputs produced by UGE. UGE parsed all the sentences correctly.

Thus, 8 combinations of adjective words with other syntactic categories were tested and these produced three unique patterns as output, namely:

1. (:modifier (a*) (new*)) – an object has been found on the right; a straightforward case.
2. (?R (:modifier (happy*)) (:manner (now*))) – no object is found on the right and additional information added to the ?R list.
3. (?R (:modifier (nice*) (and*) (clean*))) – no object is found on the right and additional information added to the :modifier list.

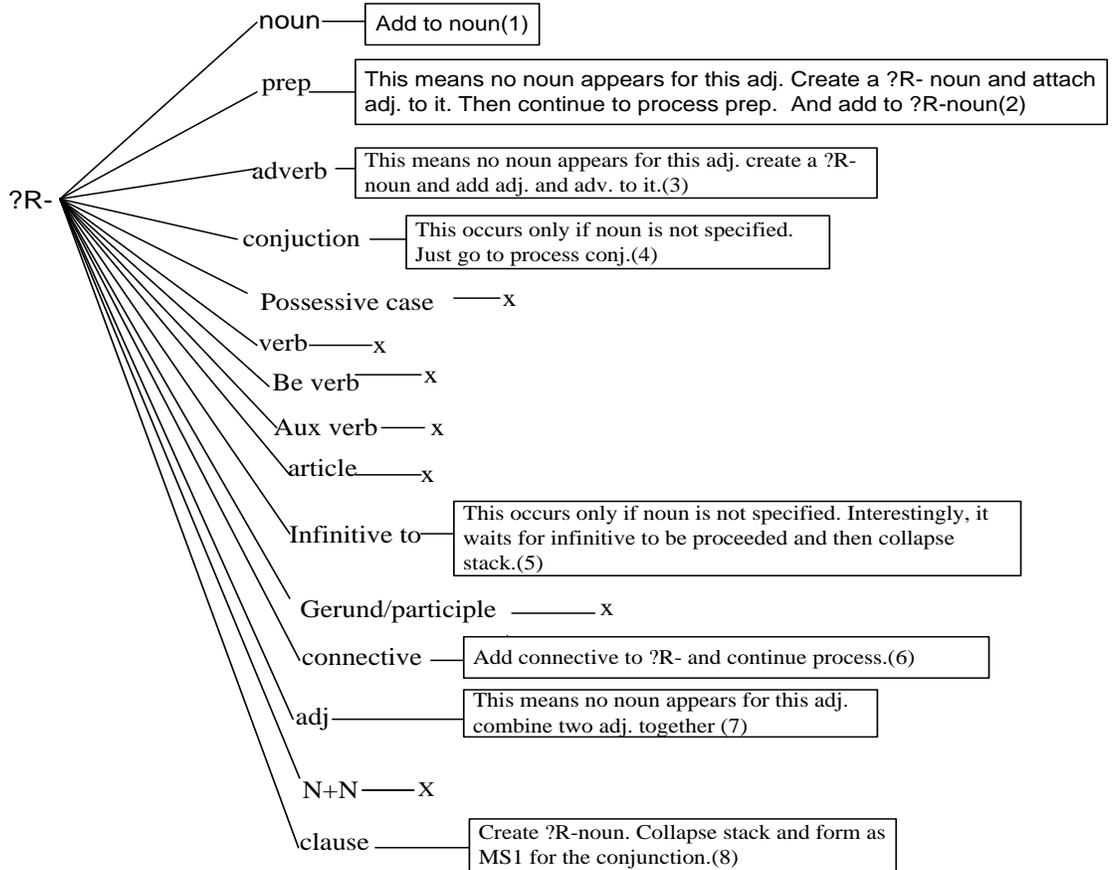


Figure 4.2 Adjective patterns

Table 4.1 Adjective patterns and examples

UGE Patterns	Grammatical Patterns	Sentence examples (input) and UGE outputs
(:modifier adj.)	Adj. +noun	(1) This is a <u>new house</u> . [THIS* (:NOUN) (:IS* (HOUSE* (:NOUN) (:MODIFIER (A*) (NEW*))))]
?R (:modifier adj. (:preposition))	adj. + preposition	(2) It is <u>nice for</u> you. [IT* (:PNOUN) (:IS* (?R (:MODIFIER (NICE*)) (:FOR* (YOU* (:PNOUN)))))]
(?R(:modifier adj.) (:manner)	Adj. +adv.	(3) I am <u>happy now</u> . [I* (:PNOUN) (:AM* (?R (:MODIFIER (HAPPY*)) (:MANNER (NOW*))))]

Implementation

?R (:modifier adj.)	Adj.+ conj. Adj.+ infinitive	(4) You are <u>happy and I am happy</u> . [AND* (:MS2 (I* (:PNOUN) (:AM* (?R (:MODIFIER (HAPPY*)))))) (:MS1 (YOU* (:PNOUN) (:ARE* (?R (:MODIFIER(HAPPY*))))))] (5) It is very <u>nice to see you</u> . [IT* (:PNOUN) (:IS* (?R (:MODIFIER (NICE*))) (:MANNER (VERY*))) (:TO** (SEE* (:WHAT (YOU* (:PNOUN)))))]
?R (:modifier Combined adj.)	Adj. + connective	(6) The room is <u>nice and clean</u> . [ROOM* (:NOUN) (:MODIFIER (THE*)) (:IS* (?R (:MODIFIER (NICE*) (AND*) (CLEAN*))))]
?R(:modifier adj. adj.)	adj. + adj.	(7) I wish <u>you happy new year</u> . [WISH* (:ACTOR (I* (:PNOUN))) (:RECIPIENT (YOU* (:PNOUN))) (:WHAT (YEAR* (:NOUN) (:MODIFIER (HAPPY*) (NEW*))))]
?R(:modifier MS1)	Adj. + clause	(8) It would be nice <u>that we found the answer</u> . [WOULD* (:ACTOR (IT* (:PNOUN))) (:MS1 (BE* (:ACTOR ?L) (:WHAT (?R (:MODIFIER (NICE*))) (:THAT* (:MS1 (FOUND* (:ACTOR (WE* (:PNOUN))) (:WHAT (ANSWER* (:NOUN) (:MODIFIER (THE*)))))))))]

Verb

There are three verb patterns in UGE, namely: (WALK* (:ACTOR ?L+)) , (FLY* (:ACTOR ?L+) (:WHAT ?R+*)) and (GIVE* (:ACTOR ?L+) (:WHAT ?R+) (:WHAT ?R++)). The first one is an example of an intransitive verb, the second is a normal verb and the third is a dative verb. Note that the left side is only a noun or a noun phrase and it is the right side that has many variations. This test will look only at all the possibilities on the right side. Figure 4.3 shows that 7 out of 15 categories were never found to be present. Table 4.2 shows the test sentences constructed for testing UGE and the outputs produced by UGE.

Thus, 8 combinations of verbs with other syntactic categories were tested and these produced four unique patterns as output, namely:

1. (verb* (:actor) (:what))—an object has been found on the right. Then it is a straightforward case, just create :what and simply attach to the result.
2. (verb* (:actor) (:through))—there is no object in the sentence. Then parse the rest part at the right side of verb, and attach to the result.
3. (AND* (:MS2) (MS1))—two verbs are connected by connective word and, parse each verb phrase respectively then collapse.
4. (verb* (:actor) (MS1))—when the object is a clause, then generate MS1.

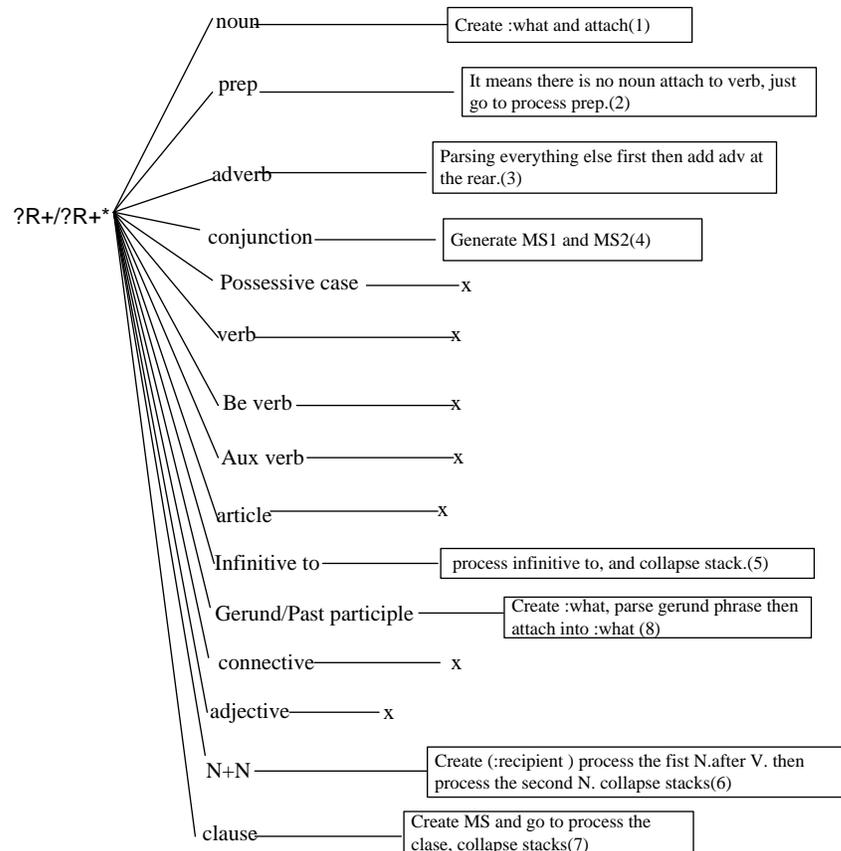


Figure 4.3 Verb Patterns

When UGE parse verb phrase which connected with infinitive to, it sometimes embed infinitive to into (:what) structure (see example 1.1 in table 4.2) and sometimes not (see example sentence (5)). This inconsistency might cause problem in the interpretation of the output and perhaps it is best that this is modified to display the attachment relationship clearly for all cases. More importantly, since there is no object attached with the verb word when the “infinitive to” appears, it is better to delete (:what) which means that the verb word is not having any object attached to it.

Table 4.2 Verb patterns and examples

UGE patterns	Grammatical patterns	Sentences examples and UGE output for examples
[verb :actor :what]	verb + object verb +infinitive to verb + gerund	(1)He <u>helps</u> the police. [HELPS* (:ACTOR (HE* (:PNOUN))) <u>(:WHAT (POLICE* (:NOUN) (:MODIFIER</u> <u>(THE*))))]</u> (1.1)Police <u>refused</u> to provide a copy of the statement. [REFUSED* (:ACTOR (POLICE* (:NOUN))) <u>(:WHAT (:TO** (PROVIDE*</u> <u>(:WHAT (COPY* (:NOUN) (:MODIFIER (A*)</u> <u>(:OF* (STATEMENT* (:NOUN) (:MODIFIER (THE*))))))))]</u> (8) We <u>enjoy spending</u> time with you. [ENJOY* (:ACTOR (WE* (:PNOUN))) <u>(:WHAT (SPENDING* (:WHAT (TIME* (:NOUN)</u> <u>(:WITH* (YOU* (:PNOUN)))))))]</u>
[verb :actor preposition]	verb + preposition + object	(2)She <u>went through</u> a lot of trouble. [WENT* (:ACTOR (SHE* (:PNOUN))) <u>(:THROUGH* (LOT* (:NOUN) (:MODIFIER (A*)</u> <u>(:OF* (TROUBLE* (:NOUN)))))]</u>
[verb :actor :manner]	verb + adv	(3)Amy <u>laughed</u> loud. [LAUGHED* (:ACTOR (AMY* (:UNKNOWN) (:NAME))) <u>(:MANNER (LOUD*))]</u>
[MS1 MS2]	verb +and+ verb	(4) She can <u>sing and</u> dance. [AND* (:MS2 (DANCE* (:ACTOR ?L))) <u>(:MS1 (CAN* (:ACTOR (SHE* (:PNOUN)))</u> <u>(:MS1 (SING* (:ACTOR ?L) (:WHAT ?R))))]</u>

Implementation

[verb :actor :to **]	verb +infinitive to	(5)I <u>promise to wait</u> . [PROMISE* (:ACTOR (I* (:PNOUN))) (:TO** (WAIT*))]
[verb :actor :recipient : what]	verb + object + object(recipient)	(6)She <u>told Ema</u> the news. [TOLD* (:ACTOR (SHE* (:PNOUN))) (:RECIPIENT (EMA* (:UNKNOWN) (:NAME))) (:WHAT (NEWS* (:NOUN) (:MODIFIER (THE*))))]
[verb :actor MS1]	verb + clause	(7)Citron said <u>he is not concerned about bringing a new technology to markets</u> . [SAID* (:ACTOR (CITRON* (:NOUN))) (:MS1 (HE* (:PNOUN) (:IS* (?R (:MODIFIER (CONCERNED*)) (:ABOUT* (BRINGING* (:WHAT (TECHNOLOGY* (:NOUN) (:MODIFIER (A*) (NEW*)) (:TO* (MARKETS* (:NOUN)))))))) (:MANNER (NOT*)))))]

Be Verb

In UGE dictionary, an example of be verb “are” is: (?L* (:ARE* ?R+)). Similar to verb structure, we only look at the right side of be verb. 5 possible sentences are found to be able to attach at the right side of “be verb” (shown in the Figure 4.4 below). There are 10 out of 15 categories found not to be present. Table 4.3 shows the test sentences constructed for testing UGE and the outputs produced by UGE. UGE parse most of the sentences correctly, but there are some errors in the sentence when the be verb connects with prepositions.

Thus, 5 combinations of be verbs with other syntactic categories were tested and these produced two unique patterns as output, namely:

1. (:is* (:noun))--When there is a noun or noun phrase as an object attached to the right side of “be verb”, UGE parse the noun or noun phrase and directly attach to “be verb”.
2. If it connected with a gerund, participle, adjective or clause, UGE parse the part of those categories firstly, and attach to be verb as an entity. For example, a pattern is formed as (:is* (suffering* (:what ?R)))

When be word is connected with preposition at its right side, the parsing result does not make sense (see sentence 1.1 in table 4.3), because preposition phrase treated as adjective. It is suggested to process preposition and rest parts of the sentence first and then attach to “be verb”.

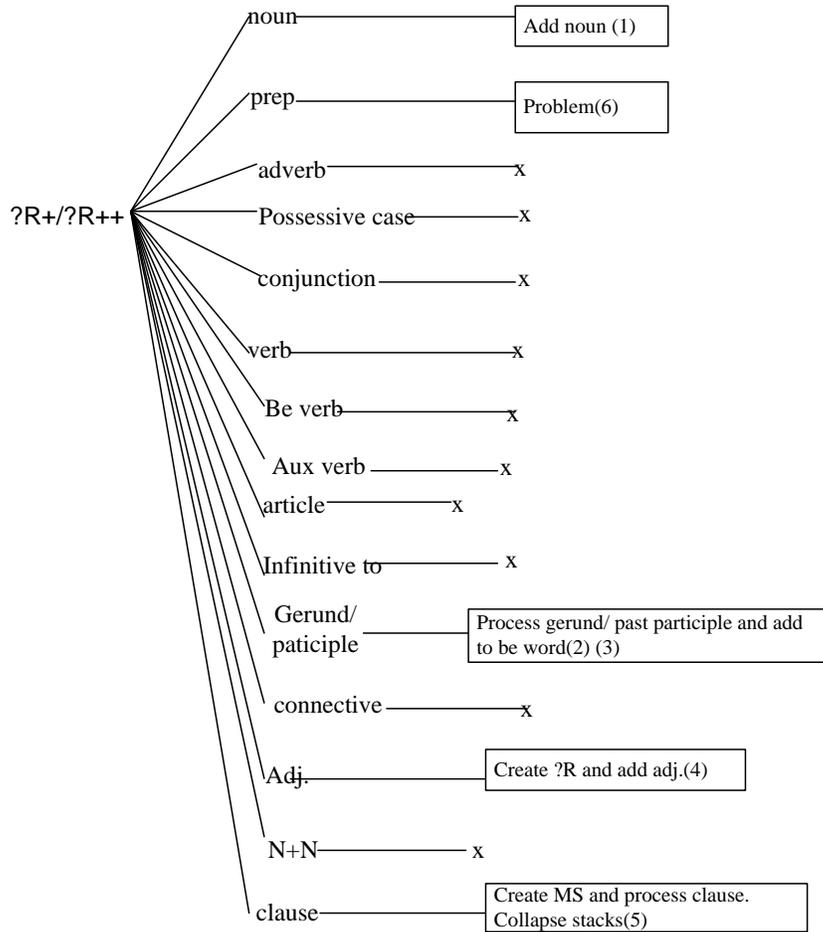


Figure 4.4 Be Verb Patterns

Table 4.3 Be Verb patterns and examples

UGE Patterns	Grammatical pattern	Sentences examples and UGE output for examples
(: be verb * (:noun))	Be +noun Be + preposition	<p>(1) Gagarin was the first to fly in a spaceship. [GAGARIN* (:UNKNOWN) (:NAME) (:WAS* (FIRST* (:NOUN) (:MODIFIER (THE*))) (:TO** (FLY* (:IN* (SPACESHIP* (:NOUN) (:MODIFIER (A*)))))))]</p> <p>(1.1) Pharmaceutical firms are about helping people. [FIRMS* (:NOUN) (:MODIFIER (PHARMACEUTICAL*)) (:ARE* (PEOPLE* (:NOUN) (:MODIFIER (ABOUT* (HELPING*)))))]</p>
(:be verb* (gerund *))	Be +gerund	<p>(2) French wine is suffering in the US [WIME* (:NOUN) (:MODIFIER (FRENCH*)) (:IS* (SUFFERING* (:WHAT ?R) (:IN* (*US** (:UNKNOWN) (:NAME) (:MODIFIER (THE*)))))))]</p> <p>They are taking the same exam. [THEY* (:PNOUN) (:ARE* (TAKING* (:WHAT (EXAM* (:NOUN) (:MODIFIER (THE*) (SAME*)))))))]</p>
(: be verb* (participle *))	Be+ participle	<p>(3) The spots were done by the Los Angeles office. [SPOTS* (:NOUN) (:MODIFIER (THE*)) (:WERE* (DONE* (:WHAT ?R) (:BY* (OFFICE* (:NOUN) (:X-WORDS (LOS-ANGELES*) (OFFICE*)) (:NAME) (:MODIFIER (THE*)))))))]</p>

Implementation

(: be verb * (?R (: modifier)))	Be +adj	(4) One thing is clear. [THING* (:NOUN) (:MODIFIER (ONE*)) (:IS* (?R (:MODIFIER (CLEAR*)))))]
(:be verb* (:that (:MS1)))	Be+ clause	(5) The initial reaction of many people is that Walter has won. [REACTION* (:NOUN) (:MODIFIER (THE*) (INITIAL*)) (:OF* (PEOPLE* (:NOUN) (:MODIFIER (MANY*)))) (:IS* (:THAT* (:MS1 (HAS* (:ACTOR (WALTER* (:UNKNOWN)(:NAME))) (:WHAT (WON* (:WHAT ?R)))))))]

Auxiliary Verb

In UGE dictionary, auxiliary verb is used similarly to normal verb except what is attached to its right is a complex clause. UGE uses the tag :MS to hold a clause and if they are more than one, it uses the tages MS1:, MS2:, etc. For example, the word “could” is represented in the dictionary as ((COULD* (:ACTOR ?L+) (:MS1 ?R+)). Such words could also start a question but this usage of auxiliary verbs is not considered here. Again, because it is always a noun or a noun phrase on the left side and it is the right side that varies, I will only look for the categories on the right side. Figure 4.5 shows that 4 out of 15 categories were found to be present. Table 4.4 shows the test sentences constructed for testing UGE and the output produced by UGE. UGE parsed all the sentences correctly.

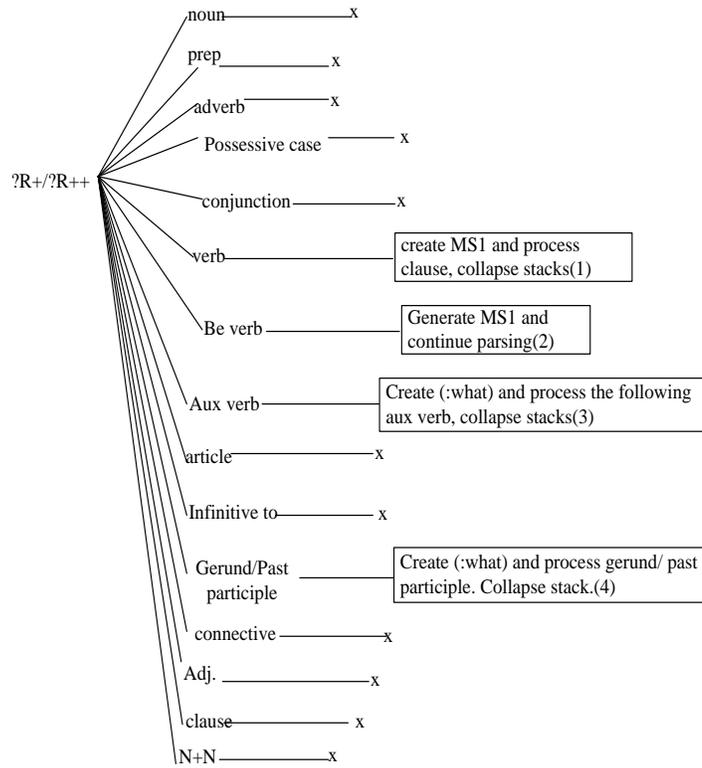


Figure 4.5 Auxiliary Verb Patterns

Table 4.4 Auxiliary Verb patterns and examples

UGE Patterns	Grammatical patterns	Sentences examples and UGE output for examples
[Aux verb* (:actor) (:MS1)]	Aux verb + verb Aux verb+ be verb	(1) I <u>would like</u> to have a chat with you. [WOULD* (:ACTOR (I* (:PNOUN))) (:MS1 (LIKE* (:ACTOR ?L) (:TO** (HAVE* (:WHAT (CHAT* (:NOUN) (:MODIFIER (A*)) (:WITH* (YOU* (:PNOUN)))))))] (2) I <u>will be</u> there soon. [WILL* (:ACTOR (I* (:PNOUN))) (:MS1 (BE* (:ACTOR ?L) (:WHAT ?R) (:MANNER (THERE*)) (:MANNER (SOON*)))))]

		<p>They <u>will enjoy</u> the game very much</p> <p>[WILL* (:ACTOR (THEY* (:PNOUN))) (:MS1 (ENJOY* (:ACTOR ?L) (:WHAT (GAME* (:NOUN) (:MODIFIER (THE*)))) (:MANNER (VERY*)) (:MANNER (MUCH*)))))]</p>
<p>[Aux verb* (:actor) (:what)]</p>	<p>Aux verb +Aux verb Aux verb+ participle</p>	<p>(3)He <u>had been speaking</u> for one day.</p> <p>[HAD* (:ACTOR (HE* (:PNOUN))) (:WHAT (BEEN* (:WHAT (SPEAKING* (:WHAT ?R) (:FOR* (DAY* (:NOUN) (:MODIFIER (ONE*)))))))))]</p> <p>(4)Someone <u>has taken</u> her daughter from the street.</p> <p>[HAS* (:ACTOR (SOMEONE* (:NOUN))) (:WHAT (TAKEN* (:RECIPIENT (HER* (:PNOUN))) (:WHAT (DAUGHTER* (:NOUN) (:FROM* (STREET* (:NOUN) (:MODIFIER (THE*)))))))))]</p>

Infinitive to

In UGE dictionary, an example of infinitive is (:TO** ?R+). Note that preposition “to” is represented as :TO*. It was found that 3 out of 15 possible sentences can be attached to the right side of “infinitive to” (shown in figure 4.6). Table 4.5 shows the test sentences constructed for testing UGE and the outputs produced by UGE. The patterns generated as output are as shown below:

1. (:TO** (WAIT*))-- When there is an action verb connected to “infinitive to”, it attaches “action verb” directly to “infinitive to”.
2. (:TO** (BE*))-- Similar to action verb, when parsing be verb sentences, UGE parse the rest part of sentence which after “infinitive to” and then attached the parsing result to “infinitive to” as an entity.

3. (:HOW-TO*)--In the third case, when there is wh/how clause connected at the right side of “infinitive to”, UGE combines wh/how words with “infinitive to” as an entity. UGE parses the rest of sentences after wh/how words and attaches the parsing result to “infinitive to”.

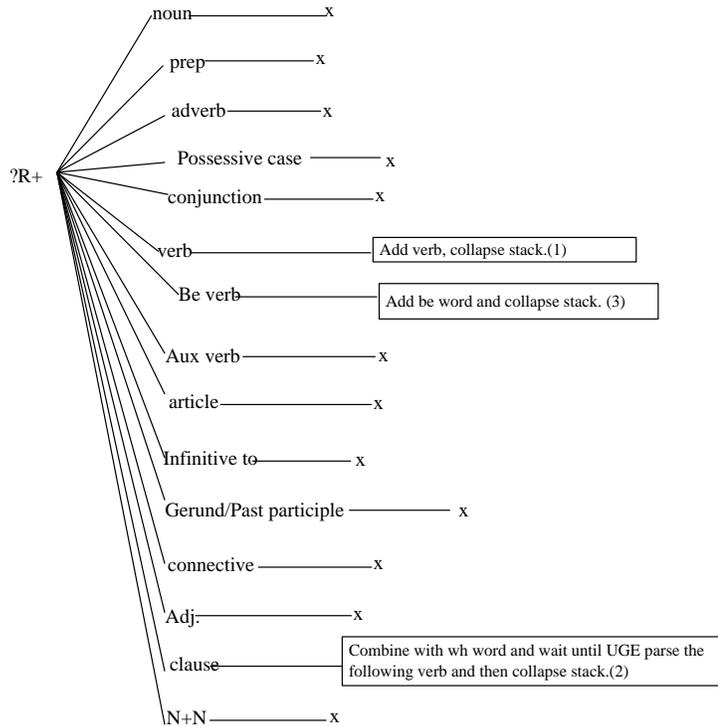


Figure 4.6 Infinitive to Patterns

Table 4.5 Infinitive patterns and examples

UGE Patterns	Grammatical Patterns	Sentences examples and UGE output for examples
(:to (verb))	Infinitive+ (Verb)	(1)I promise <u>to wait</u> . [PROMISE* (:ACTOR (I* (:PNOUN))) (:TO** (WAIT*))]
(How-to)	Infinitive Combined with How-to	(2)He discovered <u>how to</u> open the safe [DISCOVERED* (:ACTOR (HE* (:PNOUN))) (:WHAT (:HOW-TO* (:MS1 (OPEN* (:ACTOR ?L) (:WHAT (SAFE* (:NOUN) (:MODIFIER (THE*)))))))))]
(:to (be))	Infinitive+ (be Verb)	(3)He pretended <u>to be</u> angry. [PRETENDED* (:ACTOR (HE* (:PNOUN))) (:TO** (BE* (:WHAT (?R (:MODIFIER (ANGRY*)))))))]

Preposition

An example of a preposition word in UGE dictionary is: (?L* (:AT* ?R+=)). Again I look at the connections on the right and found that there are 5 out of 15 categories which can be attached. This is shown in figure 4.7. Table 4.6 shows the test sentences constructed for testing UGE and the outputs produced by UGE. Two unique patterns are generated as output:

1. (:preposition (:noun))—this pattern is not just for noun word but also suitable for connective, preposition and participle words. When these categories are attached with preposition, UGE parses the phrase for these categories first and then attach them to the preposition structure.
2. ((:preposition) (:manner))—when there is an adverb attached to the preposition, UGE simply attach add the adverb to it.

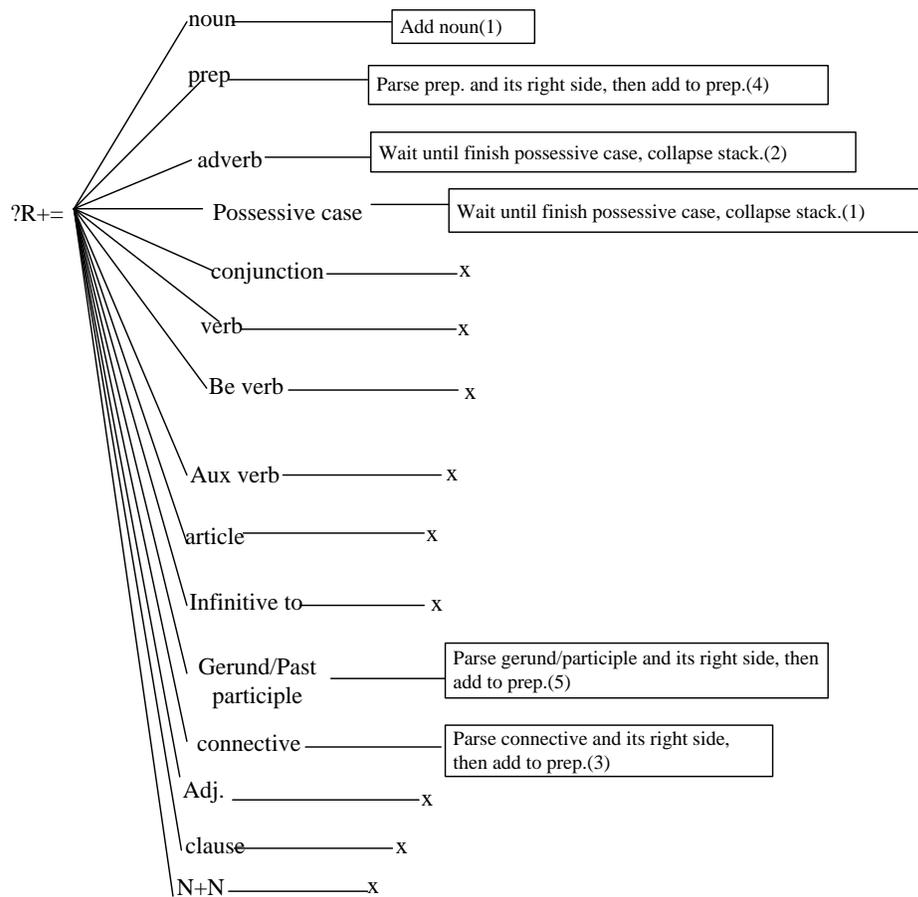


Figure 4.7 Preposition Patterns

Table 4.6 Preposition patterns and examples

UGE Patterns	Grammatical Patterns	Sentences examples and UGE examples for output
(:preposition :noun))	Preposition+ N	(1)This book is a cut <u>above his previous one</u> . [BOOK* (:NOUN) (:MODIFIER (THIS*)) (:IS* (CUT* (:NOUN) (:MODIFIER (A*)) (:ABOVE* (ONE* (:NOUN) (:MODIFIER (HIS*) (PREVIOUS*)))))))]
(:preposition (:manner))	Prep +adv.	(2)The new project was coming <u>along quite smoothly</u> . [PROJECT* (:NOUN) (:MODIFIER (THE*) (NEW*)) (:WAS* (COMING* (:WHAT ?R) (:ALONG* ?R) (:MANNER (QUITE*)) (:MANNER (SMOOTHLY*)))))]
(:preposition connective)	Prep+ connective	(3) He keeps walking <u>in and</u> out of the room. [KEEPS* (:ACTOR (HE* (:PNOUN))) (:WHAT (WALKING* (:WHAT ?R) (:IN* (AND* (:OUT* (:OF* (ROOM* (:NOUN) (:MODIFIER (THE*)))))))))]
(:preposition :preposition)	Prep+ prep	(4)News of the budget <u>got out before</u> it was officially announced. [GOT* (:ACTOR (NEWS*)))] (:OUT* (:BEFORE* (:MS1 (IT* (:PNOUN) (:WAS* (ANNOUNCED* (:WHAT ?R)) (:MANNER (OFFICIALLY*)))))))]
(:preposition :participle)	Prep +participle	(5)She finally comes around <u>after given</u> few minutes of CPR. [COMES* (:ACTOR (SHE* (:PNOUN) (:MANNER (FINALLY*))) (:MANNER (AROUND*)) (:AFTER* (GIVEN* (:WHAT (MINUTES* (:NOUN) (:MODIFIER (FEW*)) (:OF* (*CPR* (:UNKNOWN) (:NAME)))))))]

Gerund or Participle

In UGE dictionary, gerund and participle words have the same structure. For example, the word “speak” has its gerund form represented as ((SPEAKING* (:WHAT ?R+))) and its past participle form as ((SPOKEN* (:WHAT ?R+))). This means UGE treats them the same. Figure 4.8 shows that 6 out of 15 categories are able to attach at the right side of gerund or participle. Table 4.7 shows the test sentences constructed for testing UGE and the outputs produced by UGE. 2/3 of the parsing is correct and 1/3 of parsing has some errors.

For the processing of gerund/participle words, two output patterns are generated as shown below:

1. (:what)—an object or participle found on the right side, parse the noun phrase or participle phrase then add up straightforward.
2. (:what ?R)—no object nor participle found and additional information attach to(:what ?R) list

In UGE, a gerund always treated as adjective when a noun attached to the right side of gerund (see example sentence(1) in table 4.7). However, it is not a right parsing. It is suggested that, in the gerund-noun case, a (:what ?R) should be created to deal with the phrase of gerund and noun. In the case of gerund connected with conjunction clause, the clause is attached to the noun which is the object of the main clause. It is better to create a MS and start to parse it as a clause sentence.

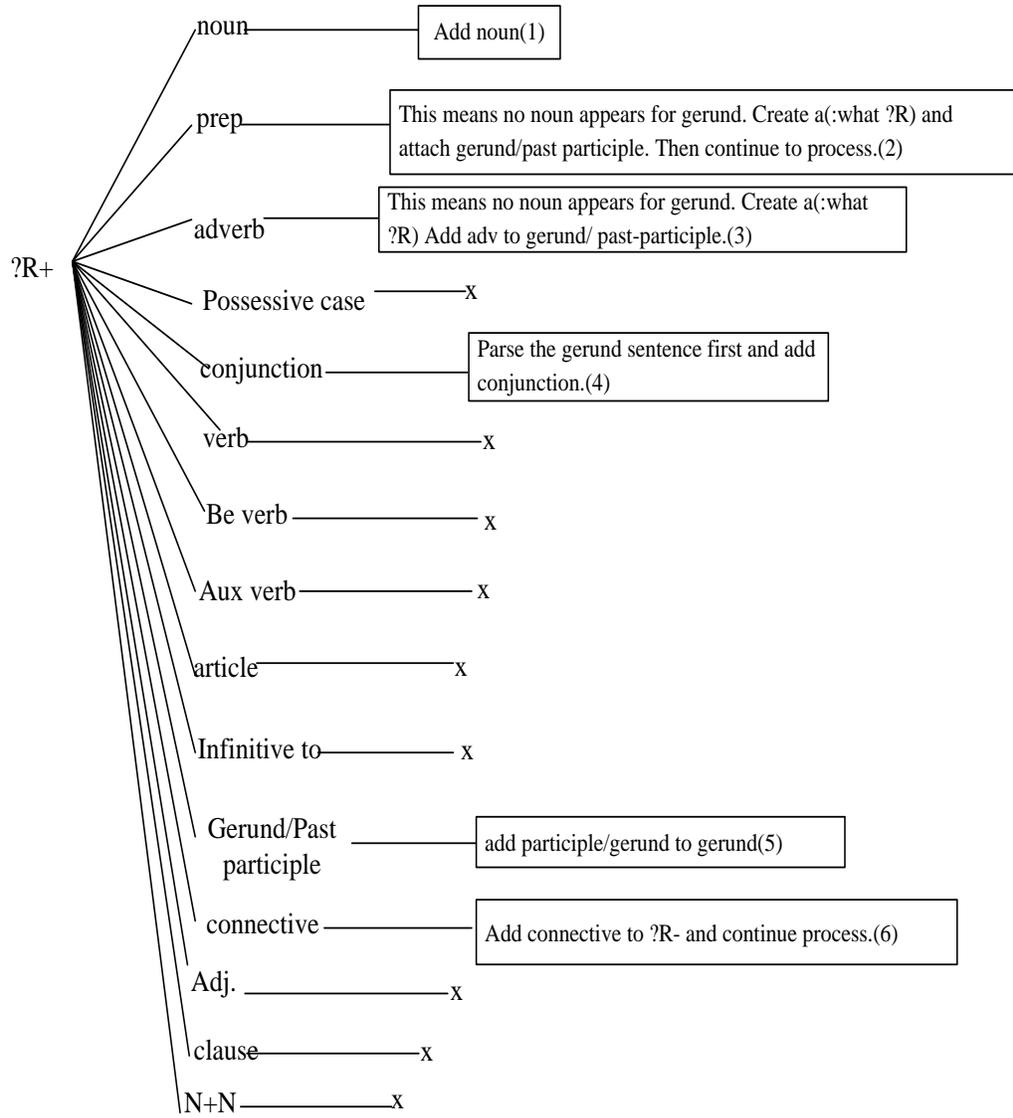


Figure 4.8 Gerund/Participle Patterns

Table 4.7 Gerund/ Participle Patterns and examples

UGE patterns	Grammatical patterns	Sentence examples and UGE output for example
Treat as adjective	Gerund + N	<p>(1) <u>Hunting elephants</u> is dangerous.</p> <p>[ELEPHANTS* (:NOUN) (:MODIFIER (HUNTING*)) (:IS* (?R (:MODIFIER (DANGEROUS*)))))]</p>
(:what ?R)	Gerund + preposition Gerund+ adverb	<p>(2) <u>Studying at school</u> is good for you.</p> <p>[STUDYING* (:WHAT ?R) <u>(:AT* (SCHOOL* (:NOUN)))</u> (:IS* (GOOD* (:NOUN) (:FOR* (YOU* (:PNOUN)))))]</p> <p>(3) He admitted <u>having driven too</u> fast.</p> <p>[ADMITTED* (:ACTOR (HE* (:PNOUN))) (:WHAT (HAVING* (:WHAT (<u>DRIVEN* (:WHAT ?R)</u> <u>(:MANNER (TOO*)) (:MANNER (FAST*))</u>)))))]</p>
(:what)	Gerund +participle Gerund +conj.	<p>(5) He admitted <u>having driven too</u> fast.</p> <p>[ADMITTED* (:ACTOR (HE* (:PNOUN))) (:WHAT (<u>HAVING* (:WHAT (<u>DRIVEN* (:WHAT ?R)</u> <u>(:MANNER (TOO*)) (:MANNER (FAST*))</u>)))))]</u></p> <p>(4) I delayed <u>telling Max the news and hiding the truth</u>.</p> <p>[DELAYED* (:ACTOR (I* (:PNOUN))) <u>(:WHAT (TELLING* (:WHAT (MAX* (:UNKNOWN) (:NAME)</u> <u>(NEWS* (:NOUN) (:MODIFIER (THE*))</u> <u>(AND* (HIDING* (:NOUN)</u> <u>(TRUTH* (:NOUN) (:MODIFIER (THE*))</u>)))))))]</p>
(:what?R) (and*)	Gerund + connective	<p>(6) They miss <u>playing and</u> singing with their friends.</p> <p>[MISS* (:ACTOR (THEY* (:PNOUN))) (:WHAT (SINGING* (:WHAT ?R) <u>(AND* (PLAYING* (:WHAT ?R)))</u> (:WITH* (FRIENDS* (:NOUN) (:MODIFIER (THEIR*)))))]</p>

Adverb

In UGE, adverb is a list of words tag with “:manner”. For example, the word “heavily” is encoded in the UGE dictionary as: (((!?L#| (:MANNER (HEAVILY*))))). Since there is only a left attachment, I look for the words appearing on the left side of adverbs. It is found that there are 8 possible attachments (shown in figure 4.9). Table 4.8 shows the test sentences constructed for testing UGE and the outputs produced by UGE. Four unique output patterns are generated as shown below:

1. When there is a noun appearing at the left side of adverb, the adverb attaches to the noun directly, and then UGE parse the rest of the sentence.
2. If there are two adverbs combine together, UGE parses the former one firstly and attach it to the left side object, then parses the second adverb and attach it to the former one.
3. In the third case, if there is an adverb appearing at the right side of an action verb, UGE processes the rest of the sentence firstly, then parses the adverb and attaches it at the end of the parsing result.
4. Similar to the attachment relationship of action verb, UGE parses the rest of the sentence but adverb firstly, and attach the parsing result of adverb at the rear of whole parsing result and collapse stack.

Interestingly, when a connective word (such as and) connect two adverbs, UGE parses the first adverb normally. Later on, it attaches the parsing result of second adverb to that connective word as an entity and then attaches it to the former adverb.

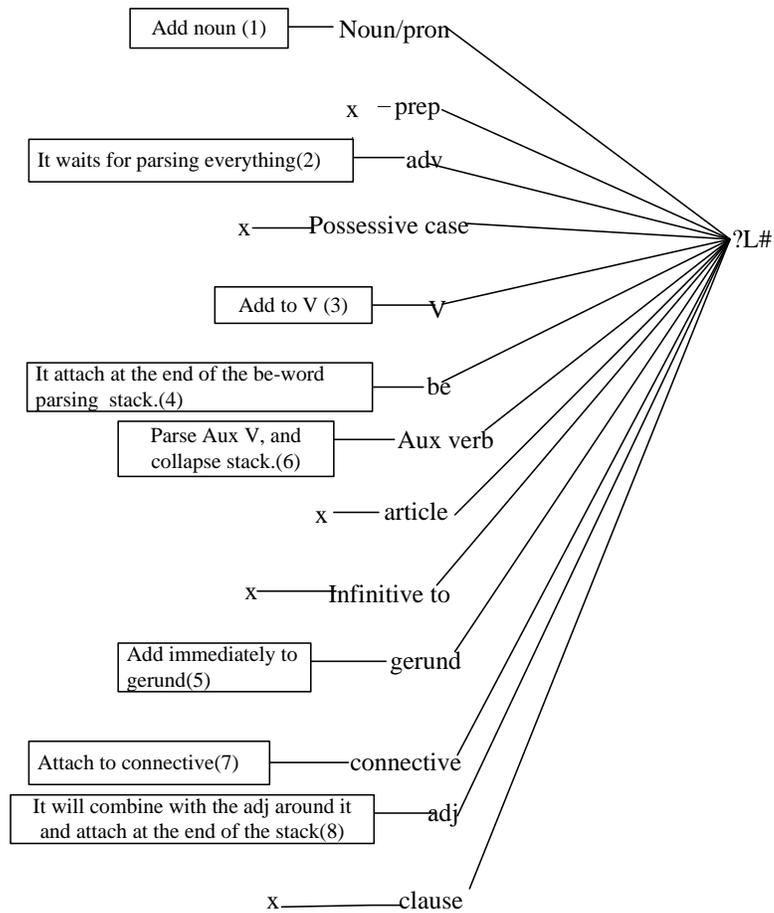


Figure 4.9 Adverb Patterns

Table 4.8 Adverb patterns and examples

UGE patterns	Grammatical patterns	Sentence examples and UGE output for example
(: pronoun) (: manner)	N+ adv	(1) <u>She finally</u> comes around after given few minutes of CPR. [COMES* (:ACTOR (SHE* (:PNOUN) (:MANNER (FINALLY*)))) (:MANNER (AROUND*)) (:AFTER*.....)))]
(:manner) (:manner)	Adv+ adv	(2) The new project was coming along <u>quite smoothly</u> . [PROJECT* (:NOUN) (:MODIFIER (THE*) (NEW*)) (:WAS* (COMING* (:WHAT ?R) (:ALONG* ?R) (:MANNER (QUITE*)) (:MANNER (SMOOTHLY*))))]

Implementation

(Verb: :n :manner)	Verb +adv	(3) He <u>runs fast</u> . [RUNS* (:ACTOR (HE* (:PNOUN))) (:MANNER (FAST*))]
(:Be (:manner))	Be verb + adv Gerund + adv	(4) News of the budget <u>was officially announced</u> . [NEWS* (:NOUN) (:OF* (BUDGET* (:NOUN) (:MODIFIER (THE*)))) (:WAS* (<u>ANNOUNCED*</u> (:WHAT ?R)) (:MANNER <u>OFFICIALLY*</u>))] (5) It was <u>raining heavily</u> . [IT* (:PNOUN) (:WAS* (<u>RAINING*</u> (:WHAT ?R) (:MANNER (HEAVILY*))))]
(Aux verb* (:manner))	Aux verb+ adv	(6) I <u>could hardly</u> hear you. [COULD* (:ACTOR (I* (:PNOUN))) (:MS1 (HEAR* (:ACTOR ?L) (:WHAT (YOU* (:PNOUN)))))) (:MANNER (HARDLY*))]
(:manner) (:connective)	Adv+ connective + adv	(7) You have to do <u>it quickly and accurately</u> . [HAVE* (:ACTOR (YOU* (:PNOUN))) (:WHAT (:TO** (DO* (:WHAT (IT* (:PNOUN)))))) (:MANNER (QUICKLY*)) <u>(AND* (:MANNER (ACCURATELY*)))]</u>
(?R(:modifier adj.) (:manner))	Adj. +adv.	(8) I am <u>happy now</u> . [I* (:PNOUN) (:AM* (?R (:MODIFIER (HAPPY*)) (:MANNER (NOW*))))]

Article

In UGE dictionary, the structure of article is similar to adjectives, for example, word “the” is explained in UGE dictionary as: ((?R-* (:MODIFIER (THE*))). However, articles can only appear at the start of a series of ?R-. 2 out of the 15 categories are found to be able to attach to an article (see figure 4.10). Table 4.9 shows the test sentences constructed for testing UGE and the outputs produced by UGE. UGE parsed all the sentences correctly. Two unique patterns are generated as output and as shown below:

1. (reason* (:noun) (:modifier (the*))-- When article modifies a noun, the parsing rule is same as adjective.
2. (:modifier (the*) (good*))-- When it modifies an adjective, the article word is always placed in front of adjective word. UGE combines them together as an entity modifier and parse the whole entity as normal adjective phrase.

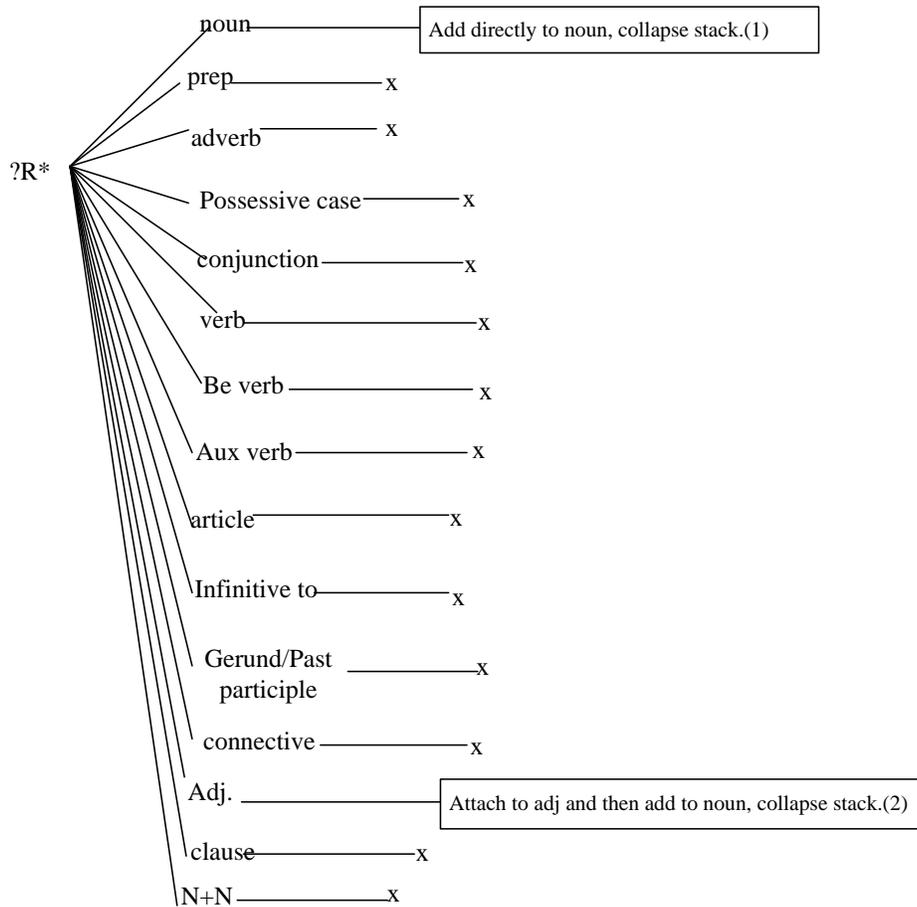


Figure 4.10 Article Patterns

Table 4.9 Article patterns and examples

UGE Patterns	Grammatical patterns	Sentence examples and UGE output for example
(:noun) (:modifier)	Article + Noun	(1) <u>The reason</u> of this case is not clear. [REASON* (:NOUN) (:MODIFIER (THE*)) (:OF* (CASE* (:NOUN) (:MODIFIER (THIS*)))) (:IS* (?R (:MODIFIER (CLEAR*))) (:MANNER (NOT*)))]
(:modifier (the*)) (good*))	Article +Adjective	(2) <u>The good grades</u> always come from hard work. [COME* (:ACTOR (GRADES* (:NOUN) (:MODIFIER (THE*) (GOOD*)) (:MANNER (ALWAYS*))) (:FROM* (WORK* (:NOUN) (:MODIFIER (HARD*))))]

Connective

In UGE dictionary, a connective word is represented as: (?L* (AND* ?R+)). Again, I only look at the attachments on the right side. There are 6 out of 15 categories appearing at the right side of connective (shown in figure 4.11). Table 4.10 shows the test sentences constructed for testing UGE and the outputs produced by UGE. UGE parse most of the sentences correctly. Two unique patterns are generated as output and as shown below:

1. (nice*) (and*) (sunny*)—when two adjective words are connected by connective word, UGE combines them together as an entity in one structure.
2. (add*)—when connective word meets other categories, additional information is attached to connective word to form a entire list.

It is interesting that most of the cases have the same parsing rules. In the case of connective with noun, adverb, preposition and “infinitive to”, the information of semantic objects always add on to connective word. In contrast, the parsing rule is slightly different when connective word attached to adjective word. UGE separated each word (connective and adjective) and treated them as isolated entity. Then it combined those entities together as integrity of adjective phrase.

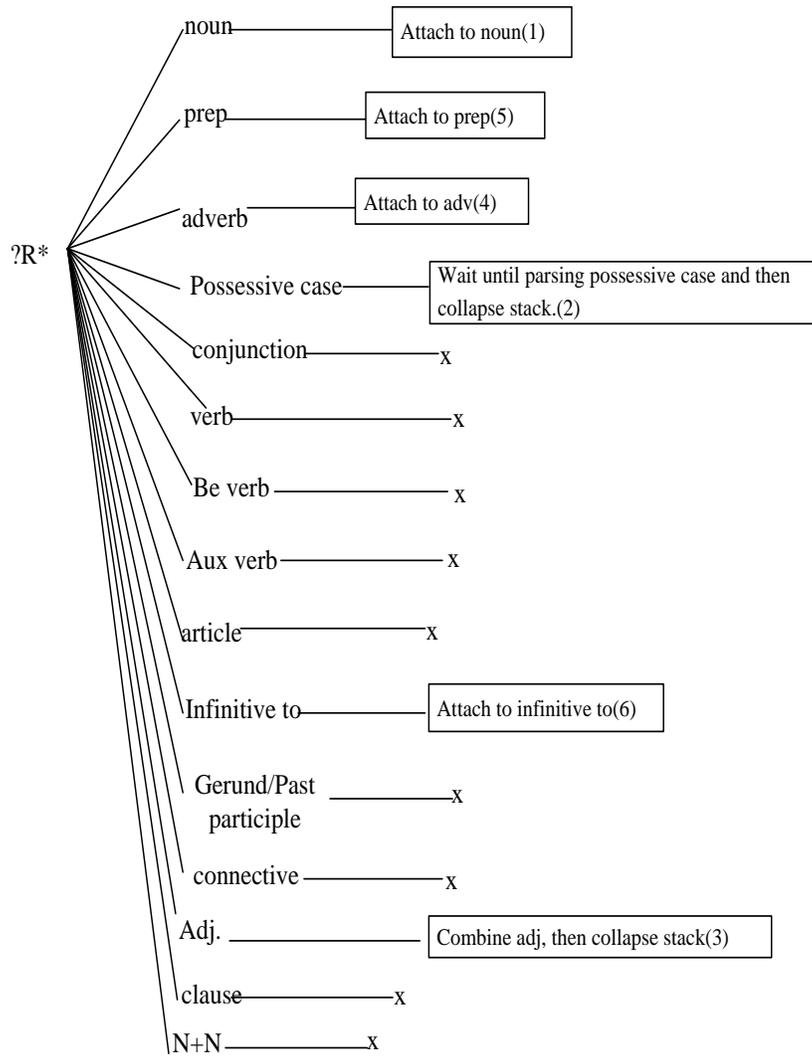


Figure 4.11 Connective Patterns

Table 4.10 Connective patterns and examples

UGE Patterns	Grammatical Patterns	Sentences examples and UGE output for examples
(and * noun)	And + N And + possessive case	<p>(1) Apple <u>and banana</u> are both fruits. [APPLE* (:NOUN) (AND* (BANANA* (:NOUN))) (:ARE* (FRUITS* (:NOUN) (:MODIFIER (BOTH*)))))]</p> <p>(2) Tom <u>and his wife</u> are leaving for the holiday. [TOM* (:NOUN) (AND* (WIFE* (:NOUN) (:MODIFIER (HIS*)))) (:ARE* (LEAVING* (:WHAT ?R) (:FOR* (HOLIDAY* (:NOUN) (:MODIFIER (THE*)))))))]</p>
(and)(adj. word)	Adj + and + adj.	<p>(3) This is a nice <u>and sunny</u> day. [THIS* (:NOUN) (:IS* (DAY* (:NOUN) (:MODIFIER (A*) (NICE*) (AND*) (SUNNY*)))))]</p>
(and* (:manner))	And + adv.	<p>(4) He drives fast <u>and safely</u>. [DRIVES* (:ACTOR (HE* (:PNOUN))) (:MANNER (FAST*)) (AND* (:MANNER (SAFELY*))))]</p>
(and* (prep. word))	And +prep.	<p>(5) He works for money <u>and for</u> fun. [WORKS* (:ACTOR (HE* (:PNOUN))) (:FOR* (MONEY* (:NOUN))) (AND* (:FOR* (FUN* (:NOUN)))))]</p>
(and* (infinitive to))	Add + infinitive to	<p>(6) He came back to check the result <u>and to</u> continue the work. [CAME* (:ACTOR (HE* (:PNOUN))) (:MANNER (BACK*)) (:TO** (CHECK* (:WHAT (RESULT* (:NOUN) (:MODIFIER (THE*)))))) (AND* (:TO** (CONTINUE* (:WHAT (WORK* (:NOUN) (:MODIFIER (THE*)))))))]</p>

4.2. Random Test Result

Almost 400 sentences were chosen randomly to perform this test. We found some of the outputs generated are consistent with those found from the first test but some are new. Table 4.11 shows the result of the random test. Some sentences failed to parse and table 4.12 show what these sentences are.

Table 4.11 Results obtained from random testing

Category	Total Sentence	Tested Pattern	Parsed sentences
Adjective	65	[modifier adj.]	31
		[?R (:modifier adj.)]	24
		[?R(:modifier adj. adj.)]	3
		[?R(:modifier adj.) (:manner)]	1
		[?R(:modifier MS1)]	2
		[?R (:modifier adj. (:preposition))]	3
		[?R (:modifier Combined adj.)]	1
		Verb	164
[verb :actor preposition]	27		
[verb :actor :manner]	11		

Implementation

		[verb :actor :to **]	8
		[verb :actor :recipient : what]	5
Be-verb	160	(: be verb * (:noun))	85
		(:be verb* (gerund *))	4
		(: be verb* (participle *))	16
		(: be verb * (?R (: modifier)))	51
		(:be verb* (:that (:MS1)))	4
Auxiliary verb	38	[Aux verb* (:actor) (:MS1)]	11
		[Aux verb* (:actor) (:what)]	27
Infinitive-to	31	[:to (verb)]	21
		[:to (be)]	10
Prepositions	19	[:preposition :noun]	18
		[(:preposition) (:manner)]	1
Gerund/ participle	7	[:what ?R]	4
		[:what]	3
Adverb	21	[(:manner) (:manner)]	2
		[(Verb: :n :manner)	11

Implementation

		(Aux verb* (:manner))	1
		(:manner) (:connective)	1
		[?R(:modifier adj.) (:manner)]	6
article	19	(:noun) (:modifier)	13
		(:modifier (the*) (good*))	6
connective	7	(and * noun)	5
		(and)(adj. word)	1
		(and* (:manner))	1

Table 4.12 Connective patterns and examples

Problem Sentences	Parsing Result	Diagnose
I am happy now as you were before.	NIL	Connective word “as “ has link problem
A separate report to be published this week will show the economy grew.	NIL	UGE cannot parse a noun clause
There is no point in waiting any longer.	[THERE* (:NOUN) (:IS* (POINT* (:NOUN) (:MODIFIER (NO*)) (:IN* (WAITING* (:WHAT (?R (:MODIFIER (ANY*) (LONGER*)))))))]	Any longer should be treated as an adverb phrase rather than adjective phrase.
The extra decorations make the house expensive.	[MAKE* (:ACTOR (DECORATIONS* (:UNKNOWN) (:MODIFIER (THE*) (EXTRA*))) (:RECIPIENT (HOUSE* (:NOUN) (:MODIFIER (THE*)))(:WHAT (?R (:MODIFIER(EXPENSIVE*)))))]	There is no recipient in this sentence, therefore, word”make” should not be treated as dative verb.

Implementation

It is nice running down the hill.	[IT* (:PNOUN) (:IS* (RUNNING* (:NOUN) (:MODIFIER (NICE*))) (:DOWN* (HILL* (:NOUN) (:MODIFIER (THE*))))))]]	Wrong sentence, when change is to sentence "it is running down the hill", UGE parse it correctly. Word "nice" make a confusion, rather than structure problem.
I lost my card and the security helped me	[LOST* (:ACTOR (I* (:PNOUN))) (:WHAT (CARD* (:NOUN) (:MODIFIER (MY*))) (AND* (SECURITY* (:NOUN) (:MODIFIER (THE*)))) (HELPED* (:WHAT (ME* (:PNOUN))))))]]	There should have an MS because the conjunction clause appear.

The random test shows two problems. One is related to a clause appearing within a sentence and the other with the use of connective words. Furthermore, a new pattern was found. The new phrase is "be verb + preposition and the example sentence is "The other two are for electric trains". The UGE pattern generated for (Be* (:preposition)) is as shown below:

```
[TWO* (:NOUN) (:MODIFIER (THE*) (OTHER*))  
(:ARE* (:FOR* (TRAINS* (:NOUN) (:MODIFIER (ELECTRIC*)))))]
```

4.3. Summary

In the experiments conducted, it was found that UGE performs well even though there are still word classes not being covered. The result has been displayed on Figures 4.1 to 4.11 and tables 4.1 to 4.12. From the analysis of the test results above, it was found that one UGE pattern can be used to handle one or more grammatical pattern.

5. Chapter 5: Conclusion, Limitations and Future Direction

This research implements a method for analyzing a new parser for natural language named UGE. UGE does not process sentences using a set of formal grammar rules. This makes it problematic to test its correctness and completeness.

A new method is used to test UGE. It consists of developing two special sets of sentences for evaluating UGE's performance. The first set is constructed by finding as many legal sentences as possible by using the English grammar book and FLAX tool (FLAX, 2010). UGE is tested to see if it manages to parse all of them. The second set is constructed from a random collection of sentences in newspapers. UGE was found to perform consistently, and its output does not produce inconsistent representations. UGE was also found to handle well with different sentences (producing a total of 47 unique patterns) involving 10 syntactic categories, namely: adjectives, adverbs, verb, be verb, auxiliary verb, infinitive to, prepositions, gerund/participle, articles and connective words.

In summary, we found 87% of the sentences were successfully parsed. Of the remaining 13%, it is important to notice that not all failed parse is the fault of UGE framework. 85% of those failed sentences are due to some words missing from the UGE dictionary. Once, they are added, these sentences were parsed successfully. The remaining 15% of the failed sentences is due to UGE's interpretation. For example, connective word can cause the attachment problems and verb has some structure problems.

Finally, the work done has three serious limitations. First, we need to increase the coverage of the test. For now, all of the selected sentences are declarative sentences. The evaluation is restrained with parsing only one sentence type. It is still undiscovered in the area of parsing other type of sentences, such as imperative sentence, exclamatory sentence, and interrogative sentence.

Secondly, most of the data is selected from newspapers, and a few are selected from English grammar book. Using the newspaper as a resource for testing NLP system might be too limited. One should include sources from elsewhere such as journal articles, emails, internet sites and others.

Thirdly, in the random test, the data set might be considered too small. We used less than 400 sentences to test all the 10 categories. To conduct a more sufficient testing, we need to increase the data size to a medium or large size.

6. References

- Allen, J., F. (2003). *Natural language processing*. San Francisco, CA, US: John Wiley and Sons Ltd.
- Barr, V. (2003). A proposed model for effective verifications of natural language generation systems. FLAIRS
- Balkan, L., Arnold, D. and Meijer, S. (1993). Test suites for natural language processing. *Machine Translation Today: Translating and the Computer*, 51-58.
- Berwick, R.C., & Chomsky, N. (2008). 'Poverty of the stimulus' revisited: Recent challenges reconsidered. *Proc. 30th Annual Conference of the Cognitive Science Society*, 383, Washinton D.C., USA.
- Bigert, J., Sjobergh, J., Knutsson, O., and Sahlgren, M. (2005). Unsupervised evaluation of parser robustness. *Proceedings of CICLing*, 142-154.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Massachusetts: MIT Press.
- Clark, A., & Eyraud, R. (2006). Learning auxiliary fronting with grammatical inference. *Proc. 28th Annual Conference of the Cognitive Science Society*, 1127-1132, Vancouver, Canada.
- De Marneffe, M. C., MacCartney, B., & Manning, C. D. (2006). Generating typed dependency parses from phrase structure parses *Citeseer*.
- Flax. (2010). Interactive Language Learning.
Retrieved from <http://flax.nzdl.org/greenstone3/flax>
- Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational linguistics*, 12(3), 175-204.
- Guida, G. and Mauri, G. (1986). Evaluation of natural language processing systems: Issues and approaches. *Proceedings of IEEE*, 74, 1026-1035.
- Hogenhout, W. R., & Matsumoto, Y. (1998). A fast method for statistical grammar induction. *Natural Language Engineering*, 4(3), 191-209.
- James, F. A. (2003). *Natural language processing*: John Wiley and Sons Ltd.
- Kakkonen, T. (2007). Robustness evaluation of two CCG, a PCFG and a Link Grammar Parsers. *Proceedings of the 3rd Language and Technology Conference: Human*

- Language Technologies as a Challenge for Computer Science and Linguistics, Poznan, Poland.
- Kidd, E. (2004). Grammars, parsers, and language acquisition. *Journal of Child Language*, 31(02), 480-483.
- King, M. (1996). Evaluating natural language processing systems. *CACM*, 39,73-79.
- Klein, D., & Manning, C., D. . (2003). *Accurate unlexicalized parsing*. Paper presented at the Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume1, Sapporo, Japan.
Retrieved from doi:<http://dx.doi.org/10.3115/1075096.1075150>
- Lidz, J., & Gleitman, L.R. (2004). Yes, we still need universal grammar, *Cognition* 94, 85-93.
- Liddy, E. D., Hovy, E., Lin, J., Prager, J., Radev, D., Vanderwende, L., & Weischedel, R. (2003). Natural language processing. *Encyclopedia of library and information science*, 2126–2136.
- Li, L., Dahl, D.A., Norton, L.M., Linebarger, M.C., & Chen, D. (1998). A Test Environment for Natural Language Understanding Systems. In *Proceedings of the 36th ACL and the 17th COLING (ACL-COLING '98)*. Montreal.
- Litman, D., Moore, J., Dzikovska, M. O., & Farrow, E. (2009). Using Natural Language Processing to Analyze Tutorial Dialogue Corpora Across Domains Modalities *IOS Press*.
- Miller, J., & Weinert, R. (2009). *Spontaneous spoken language: Syntax and discourse*: Oxford University Press, USA.
- Naidu, V., Singh, A. K., Sharma, D. M., & Bharati, A. (2009). *Improving the Performance of the Link Parser*. Paper presented at the Asian Language Processing, Singapore.
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*, Cambridge, Massachusetts: MIT Press.
- Rodney, D. N., Wayne, W., James, H. M., & Martha, P. (2008). *Extracting a representation from text for semantic analysis*. Paper presented at the Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers, Columbus, Ohio. from

- Ritchie, G., & Thompson, H.(1983). *Natural Language Processing*, In O'Shea, T., Eisentadt, M.(eds.): *Artificial Intelligence: Tools, Techniques and Applications*. Harper & Row, New York.
- Rusu, D., Dali, L., Fortuna, B., Grobelnik, M., & Mladeni , D. (2007). Triplet extraction from sentences. *Proceedings of the 10th International Multiconference" Information Society-IS*, 8–12.
- Savin, H. B., & Perchonock, E. (1965). Grammatical structure and the immediate recall of English sentences. *Journal of Verbal Learning and Verbal Behavior*, 4(5), 348-353.
- Sleator, D. D. K., & Temperley, D. (1995). Parsing English with a link grammar. *Arxiv preprint cmp-lg/9508004*.
- Walker, C.R. and Copperman, H.(2010). Evaluating complex semantic artifacts. LREC 2010: Seventh International Conference on Language Resources and Evaluation. Malta.
- Yeap W.K. (2005a). Semantics parsing revisited or how a tadpole could turn into a frog. *In Proceedings of the 2nd Language & Technology conference*, 468 – 472
- Yeap W.K. (2005b). A new GOFAI theory: How language works. *In Proceedings of the International LISP Conference*, 371 – 378
- Yen, D. C., & Davis, W. S. (1999). *Natural language processing*. CRC Press LLC.
- Zandvoort, R. W., & van Ek, J. A. (1969). *A handbook of English grammar*: Longmans.

7. Appendix: random testing data set

1. This is a new house.
2. The answer is clear.
3. I wish you happy New Year.
4. I am happy now as you were before.
5. It is good a task has been done.
6. It is nice for you.
7. The room is nice and clean.
8. It is very nice to see you.
9. It would be very nice that we found the answer.
10. A separate report to be published this week will show the economy grew.
11. The company had record global sales of 1.95 million cases.
12. The higher value of the NZ dollar had a negative effect.
13. The fair value movement on biological assets was included in the top line.
14. Commercial and rural businesses declined slightly.
15. This is similar to the comments in the parent's result.
16. McKay is switched onto the global trend.
17. McKay gears up to lead the new Auckland Council's executive team.
18. The Super City's effectiveness might be improved.
19. Greater focus on connectedness in its transport structures.
20. A considerable amount of work is being done.
21. The project will be quite a phenomenal bit of work.
22. It will be done at the right time.
23. Three stations will be located under public roads.
24. The city has seen an economic rejuvenation.
25. The same kind of petty disputes that occurred.
26. Having a friendly council is useful because you can shoot on the streets.
27. You are in close proximity to pretty much any location.
28. We are a significant employer.
29. The council looks after its key businesses.

30. Film facilities have also attracted new residents.
31. He remembers Henderson Square as an awful place.
32. He is also willing to admit his defeats.
33. I am wildly enthusiastic about the ARC.
34. It could have also created vital jobs for a city.
35. He is happy with Prime Minister's decision.
36. He finally hangs up his mayoral chain later this year.
37. The country enjoyed a natural advantage over international competitors.
38. Professor Porter named the marine industry as one with considerable potential.
39. It feeds a great number of suppliers of goods and services.
40. Major public and private investment is going into Wynyard Quarter.
41. Progress in the domestic market is somewhat dependent on population growth.
42. He would be happy to sit now with the new Auckland Council.
43. Here has been an enormous amount of work done in recent years.
44. Electrification is a major project.
45. Infrastructure development is economically sustainable.
46. Such improvements are unlikely while so many are misinformed about Auckland's environment today.
47. The subsequent intensification was enforced.
48. The genuine urban area is the continuous area.
49. Some cities are much denser than Auckland.
50. Auckland is too dense already.
51. It is more likely to happen in the current political environment.
52. The issues were political and the lead times were long.
53. It was unlikely there would be any reduction in rates.
54. There is a significant amount of work over the long haul.
55. The Government was getting excellent value for money from construction companies.
56. The other two are for electric trains.
57. Pedestrians and cyclists would be accommodated on the existing harbor bridge.
58. The bridge is a far more resilient structure to have in an earthquake.

59. She took a very good picture.
60. He drove too fast yesterday.
61. She finally comes around after given few minutes of CPR.
62. It was raining heavily the day before yesterday.
63. He knocked heavily against the wall.
64. You have to do it quickly and accurately.
65. They work continuously to build the house.
66. It was raining heavily when I was leaving.
67. Commerce Minister Simon Power announced today.
68. Article continues below.
69. Richard Simpson and Trevor Thornton were also statutory managers.
70. The Registrar of Companies case had been watched closely.
71. It subsequently filed a High Court statement of claim.
72. The talks ended very quickly.
73. Two very important announcements are looming.
74. The companies are also expected to announce whether funding has been found.
75. They would be would only be identified.
76. Companies should be more concerned with responding to the mountain of claims.
77. It was too early to estimate how big of a problem were likely to be.
78. Dishonest claims may sneak through simply.
79. Inland Revenue's systems are becoming more sophisticated.
80. International tax rules were continuously changing.
81. The deals on large purchase usually come with administration.
82. He speaks English quite well.
83. This is a fairly useful tool.
84. She often went there.
85. Some companies also charged a penalty for paying the debt early.
86. It's quite hard to work out what is in the contract.
87. I'm not caught out in the end.
88. I refer you to an article most recently on July 31.
89. We may be in such a phase currently.

90. The 10-year averages have been negative recently.
91. Those gains are added to relatively high interest.
92. The Japanese market is hardly typical.
93. This suggests shareholders will usually receive higher returns than bondholders.
94. It's too easy to get it wrong.
95. You sometimes buy cheaply and sometimes expensively.
96. Our solution is locally owned.
97. Generally we're seeing it pretty much in every market.
98. I used to really like staying there.
99. The Westin was being partially closed.
100. I was very disappointed.
101. They are just there to get their clients.
102. It is much quicker and easier to just shut something down.
103. Some businesses may have been tempted recently to load up their swaps hedging.
104. I firmly believe interest rates have hit the bottom.
105. Ultimately the best strategy is to base decisions.
106. You could still be shouting to an empty room.
107. The market has changed significantly since the end of the last recession in 2001.
108. New Zealand has just suffered its biggest corporate collapse.
109. It always results in improved performance and never fails to reduce the company's overhead.
110. They recently analysed decisions of the Employment Relations Authority.
111. Recently I read Angela Atkins' book.
112. It is internationally competitive.
113. Australia has benefited hugely from the emergence of China.
114. They usually cannot invest before they earn.
115. It is obviously correct.
116. The earthquake now forces us to unlock a large chunk of stored wealth.
117. The reason of this case is not clear.
118. The good grades always come from hard work.
119. The most important thing is the task.

120. It is well capitalised and has strong liquidity.
121. It had gone as far as it could in investigating Blue Chip.
122. Insurance companies were trying to expedite the process as quickly as they could.
123. Banks will have to significantly increase their capital reserves.
124. The world's major central banks are trying to prevent another financial collapse.
125. The new banking rules are designed to strengthen bank finances.
126. They may dampen the recovery by forcing them to reduce the lending.
127. Forcing banks to keep more capital on hand will restrict the amount of loans.
128. It will make them better able to withstand the blow if many of those loan go sour.
129. The rules also are intended to boost confidence.
130. Banks must hold back at least 4 per cent of their balance sheet to cover their risks.
131. Banks would be required to keep an emergency reserve.
132. It still has to be presented to leaders of the Group.
133. They will find new ways to appear more conservative.
134. Earlier Sunday plans to raise at least 9.8 billion euro.
135. There is likely to be a small upward impact on the rate of inflation.
136. Natural disasters tend to discourage tourists.
137. New Zealand is a good destination to visit.
138. Hotels were the only accommodation type to record an increase.
139. European banks would struggle to raise the additional cash.
140. The new rules may cause them to reduce the lending.
141. New Zealand Government had also acted quickly to provide a recovery package.
142. Everyone in New Zealand will have to be patient.
143. Extensive discounting is thought to have played a part.
144. He always has to worry about things.
145. A committee of Icelandic lawmakers is split over whether to recommend prosecution of former leaders.
146. The Independence Party refused to endorse charges against any of the politicians.
147. It is difficult to accurately predict a result for the period.
148. The company would need to continue the same level of aggression.
149. Customer confidence continues to be adversely affected by many factors.

150. Poor economic news from overseas also continues to erode consumer confidence in New Zealand.
151. Those deals on offer are too good to miss.
152. Warehouse is expecting to remain difficult for a period of time.
153. He passed his paper to me.
154. This book is a cut above his previous one.
155. He is drawing on new windows.
156. The new project was coming along quite smoothly.
157. We walked into the classroom.
158. News of the budget got out before it was officially announced.
159. He keeps walking in and out of the room.
160. He came in to tell me the result.
161. She finally comes around after given few minutes of CPR.
162. It will make them better able to withstand the blow if many of those loans go sour.
163. This mandatory reserve would rise to 4.5 per cent by 2013.
164. He expressed doubts on whether the new rules will avert another crisis.
165. Businesses and consumers decided to save instead of borrow.
166. Treasury made the statement in a note on the economic impact of the quake.
167. A range of scenarios put this figure between 0.2 percent and 0.8 percent.
168. The Bay of Plenty had the biggest increase in guest nights.
169. The most significant individual contributions came from higher prices.
170. Bank regulators were as tough on lenders with new requirements.
171. Ben Bernanke in a joint statement called the new standards.
172. Cindy Kiro has made for the sake of New Zealand's economy.
173. She told a conference on welfare reform.
174. These differences were driven by a wider increase in inequality.
175. The future NZ economy would depend on today's Maori and Pacific children.
176. It came down slightly with rising employment and Working for Families tax credits.
177. The non-Maori population combined with relatively high Maori rates influence NZ economy.

178. Maori children were growing up in families on welfare.
179. Welfare payments had to be high enough to support good lives in decent housing.
180. We have to think about what kinds of investment we are going to make.
181. 25 per cent of their smaller companies would most likely close permanently.
182. The data indicated it had time on its side.
183. Export commodity prices have fallen for the past three months.
184. Opinions differ on where that is.
185. Briscoe Group is increasing its interim dividend after reporting a 41.85 per cent rise in earnings.
186. Most retailers would be thrilled with that result.
187. The company was cautiously optimistic about the second half of the year.
188. This figure includes a tax adjustment for government tax changes.
189. The group also booked an impairment adjustment for under-performing specialty stores.
190. Aggression in promotions predictably resulted in some margin erosion.
191. The race is on to ensure Auckland has adequate support infrastructure.
192. Air New Zealand is shrugging off the gloom of the global financial crisis.
193. He helps the police.
194. I have a car.
195. My car has more family space.
196. The consortium includes private equity firm Texas Pacific Group.
197. Police refused to provide a copy of the statement.
198. They asked me to write the document.
199. I heard people talking about the news.
200. I admit that the work is very hard.
201. I believe it to be good news.
202. Everyone knows what I mean.
203. My dad bought this car from Tom.
204. That young girl plays the piano wonderfully.
205. I work fast when I am under pressure.
206. He takes a walk after dinner every day.

207. Rotorua had its coldest day since 1964 on Monday.
208. Amy laughed loud.
209. The rain fell heavily.
210. The theft drove very fast.
211. She went through a lot of trouble.
212. It debuted on the stock market in 2005.
213. It happened in June 1932.
214. She forgot to put on her coat this morning.
215. He came up to me.
216. Auckland finished seventh out of the 35 cities.
217. Sydney came in a distant 23rd.
218. They finished in the top 10 for all categories except for holding doors open.
219. Mr Mallard pointed to positive comments in the research.
220. Abu Musab died in a US air strike.
221. She told Ema the news.
222. Principal Martin Keast told The Dominion Post the prohibition on corporal punishment was a law.
223. I show them some pictures.
224. Miss Smith teaches us English.
225. She call me little John.
226. My mother keeps everything tidy.
227. We consider Collins a good friend.
228. The government appointed him ambassador.
229. He got everything ready before he left.
230. They named the baby Anne.
231. Citron said he is not concerned about bringing a new technology to markets.
232. Qantas said its board unanimously recommends the offer.
233. Police said the girl lived locally were continuing at her home.
234. Today it said the profit was expected to be 50 per cent down on last year.
235. Managing Director said the involvement with Contract Labour Services NZ Ltd has ceased.

236. I felt the house shake.
237. Civil defence officials hoped this snowfall would cause no more damage.
238. Mr McDavitt said the coldest days usually came about six weeks after the winter solstice.
239. Mrs Smith denied her client was refusing to help the police.
240. The woman said she was too upset to talk about the death or the fire.
241. My mum takes care of the baby in the day time.
242. Ms Bradford said she wrote to Mr Lee six months ago.
243. Previous ERO reports said the school administered corporal punishment with the approval of parents.
244. Mr Keast said that as a Christian he believed the law.
245. He said corporal punishment was a private matter.
246. Ms Bradford said the law banning corporal punishment in New Zealand schools.
247. The survey found 44 per cent had drunk so much they passed out on at least one occasion.
248. The survey also showed binge drinkers tended to cling together.
249. The AER study also revealed binge drinkers were more likely to have a one night stand when they were drunk.
250. Vodafone general manager says number portability is a key part of its attack on Telecom's voice market.
251. Police refused to provide a copy of the statement.
252. Japan wants to shift the focus to whale management rather.
253. I want to eat the apple.
254. I like to half close my eyes.
255. I want to go to eat an apple.
256. She used to like him a lot.
257. The disclosure threatens to embarrass the Government.
258. Mr Mallard pointed to positive comments in the research.
259. They want to hear from anyone who has seen her in the past few days.
260. The defence minister pledged to step up action against cross-border rocket salvos.
261. There were no casualties.

- 262. She is a university student.
- 263. It was a football game.
- 264. Normal weight for adolescents was 50kg to 80kg.
- 265. The global average was 75 per cent.
- 266. There were no casualties.
- 267. It is just a matter of time.
- 268. These groups are a big threat to the ceasefire and the peace process.
- 269. It is too late.
- 270. Smoking is harmful.
- 271. Nine other bystanders were wounded.
- 272. Most are poor.
- 273. The movie is instructive.
- 274. Her room is clean and tidy.
- 275. The latest version of the work is better.
- 276. He is rich but not happy.
- 277. The outlook is negative.
- 278. The increasing financial burden is slow.
- 279. I am pleased to meet you.
- 280. The raw Auckland duo is sure to be targeted by a Pumas side.
- 281. It was difficult to keep the pace up.
- 282. My goal is to make the biggest difference to the highest number.
- 283. The winner is to be announced tonight.
- 284. It is better to study at school.
- 285. I am about to go home.
- 286. It is dangerous to lean out of the window.
- 287. It is pleasure to talk with him.
- 288. It is necessary to make a plan for this course of study.
- 289. 800 homes are going into their 10th day without power.
- 290. The father of twin boys is now refusing to speak to police.
- 291. Paramilitary groups are colluding with the armed forces.
- 292. Leon MacDonald is talking up the new.

293. Tom was talking to Jessie then.
294. I am surprising that you are here as well.
295. They are counting their expense for this month.
296. The company is going to reduce half of the staff this year.
297. The economist is looking for the reason of the global financial crisis.
298. She is driving to school then.
299. The cold front was expected to hit the South Island overnight and Cook Strait tonight.
300. The man was arrested in South Auckland and returned to Northland.
301. The woman is understood to be originally from the Tauranga area.
302. She was convicted and remanded for a restorative justice meeting with the SPCA.
303. Another soldier was killed in the attack.
304. The SkySeer is powered by replaceable battery that lasts about 70 minutes.
305. It is equipped with an infrared sensor.
306. Refugees are usually required to apply for asylum in the first safe country they reach after fleeing.
307. England was rattled by the sight of Owen leaving on a stretcher.
308. He was troubled by an ankle injury after landing heavily.
309. He was someone who was totally committed to selling them all off.
310. She was making tea when I came to see her.
311. This is the room where Edison once worked.
312. His idea is that we clean the floor first.
313. My opinion is that he will not agree.
314. That is why she is so happy.
315. This is what we want.
316. The question is whether we can finish our work by tomorrow.
317. It is uncertain when she is coming this afternoon.
318. I am glad that you have come.
319. Irish Government should help the economy recover.
320. They will result in an increasing financial burden that will be slow to unwind.
321. That certain applies will get better value for money.

- 322. A standard private finance contract will usually include a service agreement.
- 323. The scope for profiteering can be staggering.
- 324. The banks will remain exposed to the vagaries of global markets.
- 325. International interest rates will push up only slowly from their current low levels.
- 326. The financial markets may be erring on the dovish side.
- 327. It will take a lot more than regulatory changes.
- 328. It will be a powerful influence on the fiscal position and the extent to which the Government is adding to.
- 329. The country has depended on high-growth countries such as China to fuel the economy.
- 330. Standard has downgraded Irish government bonds by one notch.
- 331. The Government has now added the private model to a policy platform.
- 332. The largest British private finance investor has shares in 269 schools and 28 hospitals.
- 333. The global financial crisis has highlighted how vulnerable we are.
- 334. We have gone from being up to our collective nostrils in debt to the rest of world to up to our chins.
- 335. Westpac's economists have been reflecting in a recent series of research notes.
- 336. Banks have voluntarily reduced their reliance on this potentially unstable market.
- 337. Banks have been willing to pay up for more stable sources of funding.
- 338. This has driven up the price of these sources of funds, relative to pre-crisis conditions.
- 339. The case study will be published in a conference.
- 340. Tom's sister is my friend.
- 341. That beautiful girl is very polite.
- 342. The students here are from China.
- 343. I eat apple.
- 344. I am from China.
- 345. I would like to stay here.
- 346. The book on the table is mine.
- 347. Jimmy and Jessie are twins.

- 348. That is the way to create a table.
- 349. This is the apartment that they built last year.
- 350. The global economy loses momentum.
- 351. A strong yen threatens to derail the country's recovery.
- 352. Export values climbed 23 percent from a year earlier to 5.98 trillion.
- 353. Japan's export growth slowed for the fifth straight month in July.