An extension and implementation of a computational theory of language

Thamilini Arunachalam

School of Computing and Mathematical Science

Auckland University of Technology

A thesis submitted for the degree of

Doctor of Philosophy

2011

Contents

List of Figures

List of Tables

Attestation 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 or other institution of higher learning.

- Thamilini Arunachalam

Acknowledgements

> At times our own light goes out and is rekindled by a spark from another person. Each of us has cause to think with deep gratitude of those who have lighted the flame within us.
>
> - Albert Schweitzer

It is my pleasure to thank those who have made this thesis possible. Without their support and help, this thesis would have just been a dream. First of all, I would like to thank god for giving me the strength and for answering my prayers. Thank you so much, my lord.

It is with immense gratitude that I acknowledge the help, support and guidance of my supervisor, Prof. Wai Yeap (Albert). He was available to me for thesis related discussions, at any time. I consider it an honour to have worked with him. Albert, thank you very much.

I am indebted to many of my colleagues inside and outside AUT who have supported me throughout this tough journey. I owe a lot to all the members in the Centre for Artificial Intelligent Research for sharing their knowledge and helping me to develop the parser's dictionary.

I am grateful for my husband, Ramesh Arunachalam and my two lovely kids, Theepika and Sahaana. Thank you, Ramesh, for being a proof reader and for giving me strength and courage whenever I needed it. I always feel that god has gifted me two wonderful kids. They are very patient with me for stealing their mum's time.

How can I finish my acknowledgment without mentioning my parents (Mr. A. Vamathevan & Mrs. B. Vamathevan), my brothers and my in-laws for their invisible support and encouragement? To tell the truth, it is only because of my dad I am here finishing my thesis. Dad, I love you a lot.

Last but not the least, I take this opportunity to thank all my school teachers and well-wishers who have lifted me up to this stage and have inspired me over the years: Mr. G. Manoranjan (secondary teacher), Mr.S. Sivasambu (secondary teacher), Mrs. T. Shanmuhanathan (secondary teacher), Mrs. C. Ratnam (secondary teacher), Mr. K. Gnanasekaram (secondary teacher), Mr. S. Ganeshalingam (secondary teacher), Mr. S. Sathiyamoorthy (high school teacher), the late Mr. A. Mahadevan (high school teacher), Mr. T. T. Murugaiya (high school teacher), Mr. A. Partheepan (cousin) and Jim Niven (ex boss).

To Saima

Abstract

In the past 40 years, there has been much debate between nativists and empiricists about how children acquire the grammar of their first language. Nativists believe that children are born with a "Language Acquisition Device" containing innate knowledge of grammar to bootstrap the learning of grammar while empiricists insist that the input and feedback, which children receive from their parents or caretakers, are sufficient.

More recently, Yeap offered an alternative explanation: Children acquire their initial knowledge of grammar from understanding multi-word utterances. It is argued that what is learned, as part of one's syntactic knowledge, is how meanings of individual words are passed between adjacent words. Some words' meaning are passed to the words on their left or on their right while others accept meanings of the words from their left, right or both. It was claimed that the encoding of these basic movements in a word is similar to learning the word syntactic category as identified by linguists. Then, rather than learning formal grammar rules, children learn rules/heuristics to process words in terms of these allowable movements.

Yeap tested his theory with a program, code-named UGE, written in LISP. However, the test was done using a small sample of either individually constructed sentences or well-constructed sentences drawn from published English text. The grammar covered is only a fraction of those one would encounter in the daily use of the English language. This thesis extended UGE to enable it to parse real world sentences. These sentences were drawn from newspaper articles written for different domains (such as business, sports, etc.) in the New Zealand Herald. Many of these sentences are thus non-trivial sentences, highly ambiguous, and lengthy. If UGE could be extended to parse many of these sentences, then UGE would have "learned" a sophisticated English grammar and one is then more confident that the theory is feasible.

The work done involves the testing of UGE using more than 1900 sentences drawn from more than 100 newspaper articles. The result is that UGE was re-organized into four main modules, namely: a pre-processor, a more powerful UGE, an intelligent dictionary, and a post-UGE. Furthermore, a fixed test set of data (with 843 sentences) with the expected parsing result from UGE is collected and used for automatic re-test of UGE after any major modification have been done to it. This ensures that the new modifications do not undo what UGE was doing right before it was being modified. New labels were found missing and were introduced without the need to deviate from the theory. New rules were introduced to process the new labels. Sev-

eral bugs were fixed and the dictionary size has been extended from the initial 5000 words to 145699 words. In short, a more powerful UGE has been developed and the theory was found to be feasible.

The performance of UGE has also been evaluated, both in terms of how well it parses sentences and in terms of how it performs compared to some existing parsers. The former is done via an evaluation of its speed, accuracy, and number of parse options generated. The latter is done via comparing it with the Link Grammar and the Stanford Parser. The overall result shows that UGE performed better than both the Link Grammar and the Stanford Parser and that UGE can now be put to use in practical applications.

Chapter 1

Introduction

This thesis describes an extended implementation of Yeap's (2005a, 2005b) computational theory of language which was developed based upon investigating how children acquire their first language. Yeap's theory was developed as a solution to Baker's paradox. Baker's paradox is a term coined by Pinker to describe an apparently unexplainable situation in which children learn their first language (Pinker, 1984, 1987, 1989).

The paradox is that the input and feedback that children get when learning their first language are not rich enough to support what they appear to have learned which is, a sophisticated grammar of the first language. Many researchers (Hirsh-Pasek, Treiman, & Schneiderman, 1984; Demetras, Post, & Snow, 1986; Penner, 1987; Bohannon & Stanowicz, 1988; Hirsh-Pasek & Golinkoff, 1996) have attempted to show that one or more of the conditions which lead to the paradox are not true. However, the arguments put forward in Pinker's studies appear to be most convincing and the paradox is generally accepted to be true (Pinker,

1984, 1987, 1989). If this is the case, how do children learn their first language?

Many theories for child language acquisition have been developed over the years. These theories can be divided into three main categories, namely:

- **Imitation Theory:** For a long time, it was believed that children learn to speak by imitating adults (Holt, 1931). Children learn the first language by repeating what the adults are telling them or by listening to adults talking to others. The main problem with this theory is that very young children do not repeat or imitate what they hear, exactly. They may be able to repeat one or two words but they are unable to repeat the whole sentence.

- **Reinforcement Theory:** Next, it was believed that children learn by reinforcement. Children learn the language by being corrected and taught by their parents or caregivers. The main idea for this theory came from B.F.Skinner's reinforcement theory (Skinner, 1957). To believe this theory is feasible, it must be assumed that children are positively reinforced for correctness and are negatively reinforced for error. But, in reality, children are not corrected for every mistake they make and are not always encouraged for their correctness (Braine, 1971). Even, adults do not use proper grammar when they communicate. Brown and Hanlon (1970) tested B.F.Skinner's behaviorist claim that language learning is based on parents' reinforcement of children's grammatical behaviors. In their research, they found that parents did not differentially express their approval or disapproval of their children's speech. They also found that parents did not understand their children's well-formed questions any better than their badly-formed ques-

tions.

- **Innateness Theory:** This theory is based on the fact that humans have a genetic predisposition to learn language. This is similar to the childrens' ability to learn how to walk at a specific stage in their development. Therefore, naturally children have the ability to learn the language with little effort. Chomsky (1965) first proposed this idea. In his claim, children are born with a "Language Acquisition Device" (LAD) for the acquisition of language.

Debate continues as to which is the better theory. In recent times, the innateness theory has been hotly debated (for recent discussions, see Foraker & Tenenbaum, 2009; Lasnik & Uriagereka, 2002; Lawrence & Margolis, 2001; Legate & Yang, 2002; Lidz & Gleitman, 2004; Pullum & Scholz, 2002; Reali & Christiansen, 2005; Sampson, 2002). The field of child language research is often said to be split along two views, a native view (an innateness theory) versus an empirical view (Hirsh-Pasek & Golinkoff, 1996).

Nativists follow the innateness theory providing a strong argument that the input to any child learning his/her grammar of the first language is simply not rich enough. This argument is referred to as the Poverty of the Stimulus (PoS) argument (Baker, 1978; Chomsky, 1965, 1968; Pinker, 1989). However, they have never produced an algorithm detailed enough to show how an innate solution could work.

Empiricists promote the reinforcement theory believing that the input to the

child is rich enough to learn the language. Empiricists continue to develop many algorithms, mainly using connectionist and probabilities approaches, to show how some form of grammar could be learned directly from the input. Some examples of the latter work developed recently include Clark and Eyraud (2006), Foraker and Tenenbaum (2009), Redington and Chater (1998), Reali and Christiansen (2005), and Regier and Gahl (2004).

As we shall soon see in Chapter 2, both approaches have not produced a convincing model to show how language is learnt. A new computational theory was developed by Yeap (2005a, 2005b) who took a radical view of the child language acquisition problem. Assuming that the PoS argument is correct, Yeap then argued that Chomsky's conclusion is not the only logical conclusion that can be drawn from the argument. Another equally valid conclusion is that children do not learn the formal grammar rules of language first but some other forms of grammar rules. Yeap's theory identified what these rules are and developed a model to show how such rules could help one to comprehend sentences. The implementation model is called Universal Grammar Engine (UGE).

Briefly, Yeap argued that one's grammar rules emerge from a need to combine adjacent words to produce the required meanings. This argument is based upon the common observation that all children learning their first language go through a stage of multiword utterances. Yeap argued that what is learned are rules for combining individual meanings of words into their combined interpretations. Yeap identified four simple ways in which meanings of adjacent words could be combined initially, namely: (i-ii) that the meanings of the current word be added

to the meanings of the words coming from its left or from its right, and (iii-iv) that the meanings of the current word could be enriched or made more complete by adding the meanings derived from the words on its left and/or right. Such simple mechanisms, left/right attachments, form the basis of Yeap's approach for a child to develop a grammar for interpreting the language encountered. Yeap then extended these four basic mechanisms and showed that they are equivalent to many of the syntactic structures of language as studied by linguists. To process these left/right attachments, Yeap proposed a stack and the development of special routines to handle their appearance in a sentence.

Readers who are familiar with the parsing literature would realise immediately that the left/right attachment mechanism is nothing new and has been exploited in many earlier works on parser development. Examples of the latter are Categorical Grammar (Wood, 1993) and Link Grammar (Sleator & Temperley, 1993). However, the motivation behind Yeap's work is different: the mechanism is not used as an alternative means to implement a formal grammar. Categorical Grammar and Link Grammar do just that (see Chapter 3). Rather, the mechanism is identified as the first piece of knowledge that children learn, about how words are combined. Their complex use in a sentence comes from developing rules to manipulate them on a stack. Note that psycholinguistics' observations indicating that children do pay attention to local cues such as word ordering and case markings in their input, provide support that such mechanisms are learned (Abney, 1989; Hirsh-Pasek & Golinkoff, 1996). This is important since one of the major controversial issues surrounding how children acquire their first language is about learnability (Pinker, 1989).

Here is a simple example demonstrating how the basic UGE works. Consider the processing of sentence (1) below:

1. John ate an apple.

The lexical entries of these words in UGE are shown below:

Ate: (ate* (:actor ?L+) (:what ?R+))
John: John*
An: (?R- (:modifier an*))
Apple: apple*

The word with an asterisk denotes a pointer pointing towards the meaning of the word (which is expected to be a complex representation that a child would have learned in a separate process). If the entry is a list, it means that the child has learned some syntactic knowledge about that word. For example, in the definition for the word, *"ate"*, the child learns the need to attach two objects to its definition, one from the left (indicated by ?L+) and one from the right (?R+). These two objects are added to the meaning of the word with some defined roles, :actor and :what, respectively. The rules governing what kind of objects are appropriate coming from the left and right are the new grammar rules to be learned. These are not formal rules but rather rules learned from observing how language is used and which, in turn, will be used to help guide one's interpretation of language itself. Consequently, the output produced is a representation that guides one's

interpretation of the sentence. In contrast, the representation produced in traditional parsers is a verification that the sentence is legal. The "+/-" sign describes how the variable is to be replaced. If it is a "+" sign, it means it will be replaced by some information coming from the appropriate direction. If it is a "-" sign, its content will be passed on to the word coming from the appropriate direction. A stack is used to process the incoming sentence and a frame-like output is produced.

Figure 1.1 shows how a stack is used to parse a sentence in UGE. Given an input sentence, each word is read one at a time. Its lexical entry is first retrieved and for each entry, one could potentially create a new stack. A stack has several entries, each denoted by a []. For example, when the word *"John"* was read, the stack was empty and hence, [John*] is created as its first entry. Recalled that "John*" indicates a pointer to the meaning of John that one has learned (which is distinguished from the word, *"John"*). When the word *"ate"* was read, its lexical entry is combined with what is on the stack to form a combined new entry. When the word *"an"* is read, its lexical entry expects something from its right and hence, it is put on the top of the stack, waiting for something to arrive. When the word *"apple"* is read, it is combined with what is on the stack to form the interpretation for *"John ate an apple"*. However, for all nouns, one also creates a new stack whereby the noun is put as its first entry. This stack is thus waiting for a noun to appear. If it does, it means one has a compound noun. For example, a possible next word could be *"pie"*, thus creating *"apple pie"*. Hence, two stacks are created when the word *"apple"* is read. When we encounter the end of the sentence, the stack with a single entry is selected as the interpretation for the sentence. The other is simply discarded.

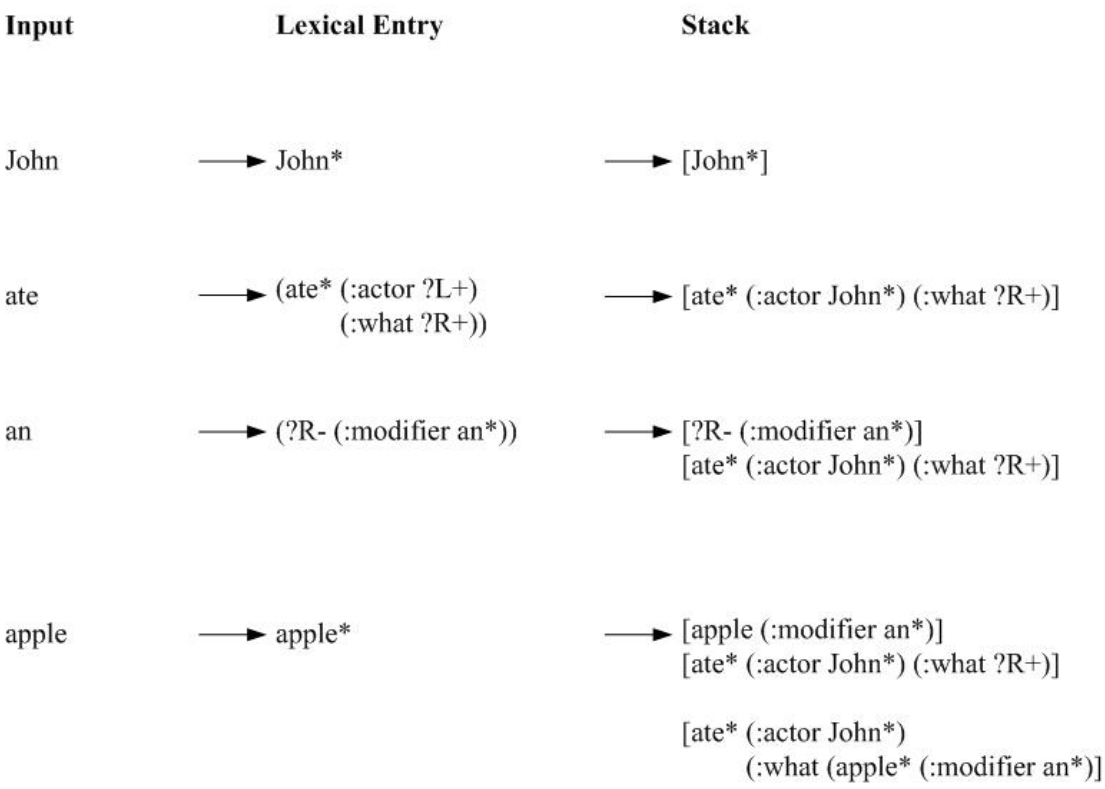

Figure 1.1: The parsing of sentence (1) - using Yeap's theory of language

Figure 1.2 shows a simplified fragment of the rules used in the routine to handle ?L+. As can be seen from Figure 1.2, there are no formal grammar rules to drive the operations on the stack. Instead, it is driven solely by inspecting what is in the lexical entry of each word and rules gleaned from observing how a word with that particular lexical entry has been used in the past. However, note that we often encode some of these rules using related formal terms but that is just for our own ease of reading the code. For example, the function "is-noun?" is used to detect words with no ?L or ?R labels. These words are formally referred to by linguists as the noun terms and hence the name of the function. Similarly, the function "inf-to?" is for detecting the word *"to"* with a ?R+ label (i.e. an infinitive to); Note that a prepositional *"to"* has a ?L* and a ?R+= labels; for a discussion of the latter labels, see Chapter 2. Linguists formally refer to such a word as an infinitive-to and hence the function is so named. Yeap implemented UGE using LISP to show how the four basic mechanisms, focusing on the left/right attachments of words, could turn into a full-blown parser (Yeap, 2005a, 2005b).

```
;; MS is the stack
;; LM is the current word which has a ?L+
(defun fill?L+MS (MS LM)

        (cond
         ; first a series of tests to decide if this word should really function as a verb

         ; Nothing on the stack
         ; a verb that starts the sentence is picked up here.
         ; e.g. "give me that"
         ((null MS) (new-MS nil LM))

         ; Noun on the stack
         ((is-noun? MS)
                (cond
                 ; if the noun on the stack is the word "her" and
                 ; the current word has a noun definition then fail
                 ; e.g. "her book" - here verb definition of "book" fails.
                 ((and (is-pnoun? MS)
                        (member (LMWord (MS-result MS)) '(his* her*))
                        (is-dict? LM :type 'noun)) nil)
                 .
                 .
         )

         ; Second, a series of tests to see which verb case is this

         ; Stack is expecting ?R+
         ((expect?R+ MS)
                (cond
                 ; infinitive-to on the stack
                 ; e.g. "I want to eat an apple"
                 ((inf-to? MS) ....)

                 ; Connective on the stack
                 ; e.g. "I eat an apple and drink an orange juice daily"
                 ((expect?conn MS) ....)
                 .
                 .
         )
         .
         .
)
```

Figure 1.2: Simplified fragment - the rules used to handle ?L+

The goal of this research is to scale up Yeap's implementation of the theory, paying particular attention to parsing real world complex sentences which are often composed of nested clauses, quotation marks and special symbols. An example of a typical real world sentence is shown in (2) below:

2. But if Berry was to be released with no more than a "bad luck, my boy - hope you learned something by all this", then the wrong message would be going out, said Mr Neels.

Sentence (2) has special symbols, quotation marks and nested clauses. UGE could not parse such a sentence before, but now it can. In order to do that, significant extension was made to the theory by extending Yeap's initial labeling scheme, and its implementation. The output (It is a complex representation and will be explained in more details in Chapter 4) is shown below:

```
(SAID*
 (:ACTOR (NEELS*  (:NAME) (:MODIFIER (MR*))))
 (:MS1
  (BUT*
   (:MS1
    (IF* (:MS1
          (BERRY* (:NAME)
           (:WAS*
            (:TO**
             (BE*
              (:WHAT
               (RELEASED*
                (:WITH*
                 (MORE*
                  (:MODIFIER (NO*))
                  (:THAN*
                   (:SS-BLK
                    (LUCK* (:NOUN) (:MODIFIER (BAD*))
                     (COMMA*
                      (BOY* (:NOUN) (:MODIFIER (MY*))
                       (:B-BLOCK
                        (HOPE* (:ACTOR ?L)
                               (:MS1
                                (LEARNED*
                                 (:ACTOR (YOU* (:PNOUN)))
                                 (:WHAT (SOMETHING* (:NOUN)
                                                    (:BY*
                                                     (THIS* (:NOUN)
                                                            (:MODIFIER (ALL*))))))))))))))))))))
                  (:MODIFIER (A*))))))))))

          (:MS2
           (WOULD* (:ACTOR (MESSAGE* (:NOUN) (:MODIFIER (THE*) (WRONG*))
                                     (:MANNER (THEN*))))
                   (:MS1 (BE* (:ACTOR ?L)
                              (:WHAT (GOING* (:WHAT ?R) (:OUT* ?R)))))))))))
```

By doing this research, we aimed to answering two important questions related to Yeap's theory:

1. Can the theory be extended to parse real world complex sentences in English?, and
2. How well would such a model of parsing perform in the real world?

Given humans' creative use of language, it is not clear whether Yeap's model can be extended to handle the creative use of language found in real world sentences. Just like it is difficult to construct a complete formal grammar for a given language, it is equally difficult to prove that the current model can handle *all* creative use of language. Furthermore, the latter problem is made more difficult since each rule is "learned" (constructed by hand in Yeap's model) via exposure to their use (see Figure 1.2). For example, to parse sentence (1) (Figure 1.1), one learns how to process a compound noun (i.e. when a noun appears, one must create a new stack to wait for another noun term to appear).

This research thus carries out the task of inspecting numerous new constructs and "learning" the new rules as required. The result is a more powerful program that can handle a variety of constructs. The successful extension of the program shows that Yeap's theory is feasible in the sense that we have not uncovered a construct that could not be learned using the same basic principles advocated in the theory. The resulting program is also found to be able to process a large amount of real world text effectively and it is thus argued that the program is also suitable for real world applications. The only major drawback is that, for now, each new case has to be discovered manually and the new rule has to be

handcrafted and added to the system. Future work will need to discover ways in which new rules can be learned automatically.

This thesis is organised in the following manner:

- Chapter 2, **Child Language Acquisition**, describes the child language acquisition problem and Yeap's solution to this problem along with other similar solutions. Firstly, the current debate arising from the Poverty of the Stimulus (PoS) argument is explained in detail with examples. Then, other theories which are related to the nativist and empiricist camps are analysed for their strengths and weaknesses. Next, Yeap's solution to child language acquisition is explained in detail. Finally, a brief review of other works on child language acquisition theories is given.

- Chapter 3, **UGE and Related Parsers**, describes the implementation of UGE and similar parsers that utilize the left/right attachment of words. The latter parsers are Dependency Grammar, Categorial Grammar and Link Grammar. We showed that such parsers are not the same as UGE and we also noted the limitations of Yeap's implementation of UGE.

- Chapter 4, **Implementation**, describes the extensions made to UGE to enable it to parse complex real world sentences. Four modules were introduced, namely: a pre-processor module, an intelligent dictionary module, an expanded UGE module, and a PostUGE module. In addition, a separate set of test data is collected and used to ensure that any modification to UGE does not have hidden side-effects. This chapter provides an example of parsing a complex sentence using the newly expanded UGE.

- Chapter 5, **Experimental Evaluation**, evaluates the new improved UGE using two different experiments. The first experiment evaluates UGE's performance in terms of its accuracy, speed and number of parse options returned. The second experiment evaluates UGE's performance against two publicly available parsers, namely Link Grammar and Stanford parser.

- Chapter 6, **Conclusion and Future work**, concludes the thesis with a summary of its findings and recommendations for future research.

Chapter 2

Child Language Acquisition

This chapter extends the discussion on the child language acquisition problem which was mentioned briefly in the introduction. In particular, it focuses on the current debate arising from the Poverty of the Stimulus (PoS) argument. As noted earlier, this debate splits the research on child language acquisition into two, a nativist and an empiricist camp. The discussion here will first focus on this debate and review some of the arguments put forward. It will then describe some of the research from both camps and Yeap's solution to the problem. Finally, a brief review of other work on child language acquisition theories is given before this chapter concludes.

A synopsis of this chapter is as follows. Section 2.1 describes the Poverty of the Stimulus argument (PoS). To date, proponents of the argument have developed few detailed models, if any, in support of this argument. Section 2.2 reveals two such work, Pinker's (1984, 1989)bootstrapping algorithm and Siskind's (1996) cross-situational algorithm. In contrast, the empiricists' approach developed sev-

eral algorithms to show how language is learned. One important work, distributional approach, is described in Section 2.3 together with the nativists' arguments as to why these solutions are inadequate. Section 2.4 describes Yeap's model in sufficient detail to enable the reader to understand the work done in subsequent chapters. Section 2.5 briefly sketches other works in child language acquisition research. Section 2.6 concludes this chapter.

2.1 PoS argument

Every speaker of a language knows the grammar of that language. It helps one to construct sentences so that others could interpret them with ease. If not, even a simple sentence such as (1) below could become highly ambiguous; one does not know who is kicking who.

1. Jane kicked John

However, how we learn the grammar of our first language remains a mystery. Since normal children acquire their grammar by the age of three, such knowledge must be learned during the early years (Ingram, 1989). Furthermore, since children cannot be told what to learn at that age, it must be learned via exposures to its use. The mystery arises when psycholinguist began to demonstrate that the input is not rich enough to enable children to learn the grammar of the language. It is puzzling then why every normal child learns it, almost effortlessly. For a start, the learning situation is already a difficult one. Children were not given a large corpus of example sentences so that they could be trained for what is right

or wrong and the input words do not come with their syntactic labels attached. However, the problem becomes a mystery and paradoxical when it was discovered that children's use of language show that their grammar knowledge exceeds what they could have learned directly from the input data (see Pinker, 1989 for a detail and excellent discussion of the problem). This argument is referred to as the Poverty of the Stimulus (PoS) argument (Baker, 1978; Chomsky, 1965, 1968; Pinker, 1989).

Two classic examples were often used in the literature to support the PoS argument that the input is not rich enough. The first is the phenomenon of auxiliary fronting in interrogative sentences (Kimball, 1973) and the second is learning antecedents for anaphoric one (Baker, 1978; Hornstein & Lightfoot, 1981). Consider the first example, the phenomenon of auxiliary fronting in interrogative sentences. The question is, how does a child learn to form an interrogative from a declarative sentence such as (2-3) below?

2. This is a dog
3. The man who was in the park is at the door

If a child were to learn from first observing how simple sentences such as (2) above are formed, it is possible that they will learn a simple rule such as "move the leftmost (first) occurrence of the auxiliary in the sentence to the front" (Chomsky, 1968). Unfortunately, such a rule will produce an ungrammatical sentence (4) for more complex sentences such as (3):

4. *Was the man who in the park is at the door?

The correct rule to learn ought to be "move the auxiliary from the main clause to the front" but it was claimed that examples needed to learn such a rule are rarely found in children's conversations (Crain & Pietroski, 2001; Legate & Yang, 2002). Furthermore, and unlike other child language learning tasks such as regularisation of plural forms or generalisation of past tense formed of verbs where there are clear signs of mistakes, children show no sign of having learned the simple rule first and then changing it to the correct one (Crain & Nakayama, 1987). Both observations strongly support the idea that children could not have learned such a rule from experience. However, if they are born with the knowledge that one's grammar is structure dependent then such a rule could easily be learned (Berwick & Chomsky, 2008).

In the second example, the anaphoric uses of *"one"* concern what children need to learn in order to resolve the meaning of *"one"* in sentences such as (5) below (from Lidz, Waxman, & Freedman, 2003):

5. I'll play with this red ball and you can play with that one.

Again, here the child could easily learn a flat structure representation although it is the nested structure that provides the most correct usage (see Figure 2.1):

Lidz et al. (2003) pointed out that the right kind of evidence might be a situation in which (6) is uttered and *"Max has a blue ball"*.

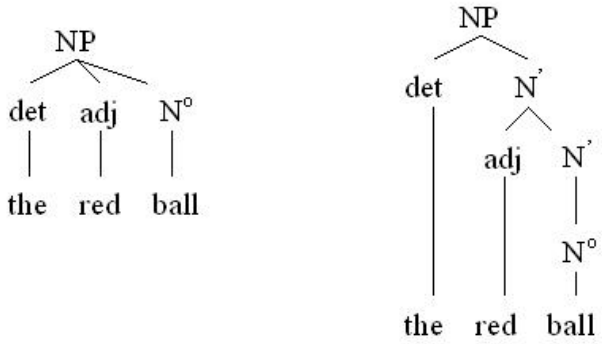

Figure 2.1: A flat structure representation versus a nested structure representation - "the red ball"

6. Chris has a red ball but Max doesn't have one.

However, they found little empirical evidence that such evidence is available to children and yet they found evidence that the child knows that *"one"* is anaphoric to the phrasal category N'.

Learning argument structures for verbs also pose a major problem and Pinker (1989) provided an in-depth discussion of these problems. Examples used include learning dative verbs (7), passive verbs (8), lexical causative alternation (9) and locative alternation (10) (from p.7 and 8 of Pinker, 1989):

7. John gave a dish to Sam/ John gave Sam a dish
John donated a painting to the museum/*John donated the museum a painting

8. John touched Fred/Fred touched by John

John resembled Fred/*Fred resembled by John

9. The ball rolled/John rolled the ball
The baby cried/*John cried the baby

10. Irv loaded eggs into the basket/*Irv loaded the basket with eggs
Irv poured water into the glass/*Irv poured the glass with water

Together, these examples above provide a strong argument that the input to any child learning his/her grammar of the first language is simply not rich enough. It poses a serious problem for understanding how children learn their first language - if children could not have learned the grammar rules from experience, where do these rules come from? Chomsky (1965, 1968) argued that if it could not be learned, it must be innately given and introduced the idea that a child is born with a universal grammar. To date, one popular version of the universal grammar idea is that a child is born with some principles and parameters of language. The former (such as the locality principle) are universal properties and the latter (such as pro-drop parameter and head directionality parameter) are set to turn the universal grammar into a unique grammar (Cook & Newson, 2007). Thus, according to this version, a child learns a language by being guided by its innate principles and by setting the parameters in it to turn it into the grammar of the language being experienced. From then on, the child learns to become a fully competent speaker of that language.

Both the PoS argument and the universal grammar idea have been fiercely de-

bated for the past 40 years and are still being hotly debated in recent times (for recent discussions, see Foraker & Tenenbaum, 2009; Lasnik & Uriagereka, 2002; Lawrence & Margolis, 2001; Legate & Yang, 2002; Lidz & Gleitman, 2004; Pullum & Scholz, 2002; Reali & Christiansen, 2005; Sampson, 2002). As discussed in Chapter 1, the field of child language research is often said to be split along these two views, a native view versus an empirical view (Hirsh-Pasek & Golinkoff, 1996).

Opponents of these ideas feverously seek evidence to show that the input is not poor and that learning is possible (notably, see Sampson, 2002). They continue to develop many algorithms, mainly using connectionist and probabilities approaches, to show how some form of grammar could be learned directly from the input. Some examples of the latter work developed recently include Clark and Eyraud (2006), Foraker and Tenenbaum (2009), Redington and Chater (1998), Regier and Gahl (2004), Reali and Christiansen (2005), and Seidenberg and MacDonald (1999).

Proponents of these ideas remain unconvinced, pointing out that the wide range of problems (some of which as described above) mean that the tiny evidence the opponents produced is simply not adequate. For example, Pinker (1984), responding to MacWhinney's (2004) article noted (on p. 951):

> "With MacWhiney, I accept Pullum & Sholzś (2002) challenge to linguists to DOCUMENT the putative rarity of sentence constructions, but the combinatorial explosion of interlocking constructions in any language would surely yield many examples of sentence types that speakers accept without prior exposure".

The algorithms the opponents developed so far have also been criticised as inadequate or unrealistic (see for example Berwick & Chomsky, 2008; Pinker & Prince, 1988)).

However, the proponents' own effort to develop a working algorithm is also lacking. For example, Hohle (2009, p. 359), while commenting on bootstrapping mechanisms, noted: "But even after decades of intensive research in language acquisition, our understanding of how these essential components interact to produce a full competence of the native language is still fragmentary". While some studies have been conducted to show how parameters in a Universal Grammar could be set (Fodor, 1998; Gibson & Wexler, 1994; Niyogi & Berwick, 1997; Sakas & Fodor, 2001, 2003), the problem of how a formal grammar is linked to meanings remains, at best, sketchy. Recall that the input to a child does not come with the syntactic categories of words attached. So, how are the linkages formed? The prominent ideas proposed to date to solve the latter is Pinker's (1987, 1989) idea of "semantic bootstrapping" and Siskind's (1996) interleaved strategy for language acquisition.

2.2 Nativist approach

The focus of this section is primarily on the nativist approach (innateness theories) which explains how children bootstrap the linguistic entities of language (e.g: noun, verb, subject, object, auxiliaries and tense) to develop a grammar for the language. This is now referred to as, the "bootstrapping problem" (Pinker, 1984, 1989). As mentioned in the previous section, to date, two works have pro-

duced detailed ideas about how bootstrapping works, which are Pinker's (1984) semantic bootstrapping algorithm and Siskind's (1996) interleaved strategy for language acquisition. These two works are reviewed in this section.

2.2.1 Pinker's semantic bootstrapping

The main idea of Pinker's semantic bootstrapping is that some innate linking rules exist to inform a child of the basic syntactic categories of words via their semantic properties (Pinker, 1984). When the child learns a word which is referring to the agent in the current scene, then that word would become the subject of the input sentence. The existing innate rule says that the agent maps onto the subject. When the child learns the word which is referring to the action in the current scene, then that word would become the verb of the input sentence. Given these rules, a child can then begin to shape their internal grammar to the one which adults use.

Even though Pinker did not explain clearly how these innate linking rules work to relate agent and action, he attempted to show how different syntactic argument structure could be derived from the semantic structure for verb.

For example, consider the different argument structure for verbs such as *"give"*:

- I give a book to John
- I give John a book (dative form)

Pinker argued that taking into consideration the semantics, one could derive a new argument structure for the verb *"give"*. This is because the first sentence

would presumably have the meaning, *"Cause X to go to Y"*, and the second sentence, *"Cause Y to have X"*. Note that, there are verbs (example: drive) that do not have a dative form and presumably these words do not have the alternative semantics.

Although one could see how semantics could assist in developing the syntactic structure of words in a language, Pinker never explained the necessary rules required to generate the different syntactic structure given the different meanings of the different sentences. Furthermore, if the meaning of a sentence can be understood without knowing the syntactic structure, it is puzzling why there is still the need to learn the sentence structure at all. Is not knowledge of syntactic structure needed to help us to derive the meaning of a sentence? There appears to be a chicken and egg problem in Pinker's solution.

Pinker also pointed out two major problems in implementing his idea (Pinker, 1989). First, since a child is not born with an innate ability to learn a specific language, the linking rules must be specified in such a way that they are valid for all languages. In other words, it is easy to decide that agents are mapped onto subject and actions to verbs but what about other categories in a language that are not easily defined. Second, there is a problem in handling the noncorrelated structures (e.g. passive, deverbal nouns, etc.). The noncorrelated structures need to be filtered by the parents out of their own speech to the child or by the child using some independent criterion.

Basically, Pinker suggested that some innate linking rules exist to "inform" a

child the basic syntactic categories of words via their semantic properties. Thus, on learning that a word is referring to an agent in the current scene, the word would become the subject of the input sentence. This is because an innate linking rule exists which says that agent maps onto subject. However, how many of these rules are sufficient and necessary for the process to work, has never been detailed and there are doubts if these rules are possible (Tomasello, 2000). To conclude, Pinker's solution is at best incomplete and at worst, inappropriate.

2.2.2 Siskind's interleaved strategy for language acquisition

Siskind uses a cross-situational strategy to account for language bootstrapping (Siskind, 1996). Again, the idea is that the combined effect of syntactic and semantic properties of word is used to derive the syntactic category of words. However, Siskind showed how each process should use the partial information provided by the other.

This process is done in two phases (Brent, 1997). In the first phase, the exact set of symbols for the meaning of the word is determined. In the second phase, the arrangement of the symbols for the meaning of the word is determined. This process is done using the set of inference rules which either narrow the set of possible meaning of the words in the utterance or narrow the set of possible interpretations of the utterance.

First phase:

To determine the meaning of the word, Siskind (1996) uses the table with the following information:

- P(w) - a possible conceptual-symbol table that contains the possible symbols for the meaning of the word w. This holds all the possibilities for the word w.
- N(w) - a necessary conceptual-symbol table that contains the necessary symbols for the meaning of the word w. This set can be determined based on one's knowledge on particular word w. If the word w is unknown then this set is empty.

For an unknown or newly used word, the algorithm knows nothing about that word w. Therefore, the N(w) has an empty set of symbols and P(w) has all the possible set of the symbols. N(w) is always a subset of P(w). By applying some inference rules, the algorithm either removes the symbols from P(w) or add the symbols to N(w). Once both sets are equal, it finishes the first phase and moves to the next phase.

Example (from, Siskind, 1996, p. 57):

Consider the sentence *"John took the ball"*. Suppose that the algorithm is part-way through its lexical-acquisition task and has the following lexicon:

w	**N(w)**	**P(w)**
John	{John}	{John, ball}
took	{}	{CAUSE, WANT, GO, TO, arm}
the	{}	{WANT, arm}
ball	{ball}	{ball}

Now consider the algorithm receives the following hypothesised meanings for this utterance *"John took the ball"*:

1. CAUSE (John, GO(ball, TO(John)))
2. WANT(John, ball)
3. CAUSE(John, GO(PART-OF(LEFT(arm), John), TO(ball)))

By applying inference rules along with the above hypothesised meanings (1-3) for this utterance, the algorithm will converge into the following lexicon:

w	**N(w)**	**P(w)**
John	{John}	{John}
took	{CAUSE, GO, TO}	{CAUSE, GO, TO}
the	{}	{}
ball	{ball}	{ball}

Since N(w) and P(w) are identical, the algorithm passes the first phase and moves to the second phase.

Second phase:

In the second phase, the possible expression which represents the arrangement of the symbols is determined. By using the inference rules and necessary conceptual-symbol table N(w), the algorithm determines the CAUSE(x, GO(y, TO(z))) as a possible meaning of the word *"took"*.

Some examples of utterances and meaning of the utterance are shown below (from Siskind, 1996):

Utterances	**Meaning**
John took the ball	CAUSE(John, GO(ball, TO(John))
Mary took the ball	CAUSE(Mary, GO(ball, TO(Mary))
John had a ball	POSSESS(John, spherical-toy)
	CONDUCT(John, -
	-formal-dance-party)
John saw Mary arrive at the ball	SEE(John, GO(Mary,-
	-TO (BE(Mary, -
	-AT(formal-dance-party)))))
John danced with Susan at the party	DANCE(John, WITH(Susan),-
	-AT(celebration))
John kissed Susan at the party	KISS(John, Susan, AT(celebration))

Siskind (1996) claimed that this algorithm solves the mapping problem (how do the children map new words to meaning). Also, he believed that this algorithm addresses the five central problems in lexical acquisition which were considered

difficult earlier:

1. Learning from multi-word input
2. Disambiguating referential uncertainty
3. Bootstrapping without prior knowledge that is specific to the language being learned
4. Noisy input
5. Homonymy

However, this algorithm is based on five strong assumptions which are not the worst-case scenario. They are (from, Siskind, 1996, p. 84-85):

- The model assumes that children can extract correct utterance meaning from the set of uncertain meaning hypotheses with sufficient regularity.
- The model assumes that homonymy can be modeled by pairing words with small sets of distinct unrelated semantic representations.
- The model cannot learn idioms and metaphors by themselves, rather it treats them as noise.
- The model assumes that a simple linking rule that treats compositional semantics purely as argument substitution.
- Inference rules assume a strict correspondence between the semantic content of an utterance and the meaning hypothesized for that utterance.

2.3 Empiricists approach

The focus of this section is to review the empirical approaches put forward against the nativist argument of PoS problem which states some form of grammar rules cannot be learned directly from input data available. The idea behind empirical approach is to use the distributional information among sequence of words in language acquisition. Distributional approach (Redington & Chater, 1998), is reviewed along with the nativist argument against empirical approaches.

2.3.1 Connectionist's distributional approach

According to Redington and Chater (1998), distributional information extracted from relationship between words, phonemes and morphemes in a sequence of words plays a major role in language acquisition. This is called distributional learning approach. Such information can be extracted using statistical models (such as contingency table) or connectionist networks.

Consider finding this phrase *"to be or not to be"* in a corpus, the co-occurrence statistic of adjacent words *"to be"* occurs twice, while *"not to"* occurs once. This co-occurrence statistic is then used as a cue to extracting syntactic category of words. This statistical information is represented in Figure 2.2. Here, word_n represents current word and word_{n+1} represents next word. The weight for the co-occurrence between *"to"* and *"be"* is 2. When the weight of the co-occurrence statistic increases the connection between two words strengthen. Using contingency table, many other distributional properties can be captured, such as presence/absence of different combinations of phonetic features or syntactic cat-

egories. These properties also can be captured using connectionist networks (see Redington and Chater (1998)).

		$Word_{n+1}$			
		to	be	or	not
$Word_n$	to		2		
	be			1	
	or				1
	not	1			

Figure 2.2: A contingency table for corpus *"to be not to be"* - As in Redington and Chater (1998).

Redington and Chater (1998) argued that the significance of statistical approaches in language acquisition are: 1) They provide principled conceptions of learning and learnability, 2) They provide potential learning mechanisms for particular aspects of language, and 3) They allow inferences in nature which extends the innate knowledge. However, Redington and Chater (1998) also cautioned that the distributional approach is not appropriate for every aspect of language acquisition.

2.3.2 Arguments against distributional approach

The distributional information is useful in language acquisition, however it is not the way how a child learns their first language. One limitation in distributional approaches is that it rarely uses the actual child input data (e.g. mothers - child-

directed speech) or have not been linked to the linguistic phenomenon of human data (Christiansen & Chater, 2001; Gobet, Freudenthal, & Pine, 2004).

Pinker (1984) placed four main points against the usefulness of distributional information in language acquisition (for more detail, see Pinker, 1984). Redington, Chater, and Finch (1998) carefully analysed Pinker's four points and argued these are inadequate. For example, Redington et al. (1998), responding to Pinker (1984) points noted (on p. 431):

> "None of these arguments are persuasive. Pinker's first point, the danger of a combinatorial explosion assumes that distributional learning mechanisms will blindly search for relationships between a vast range of properties. While this may be a fair criticism of early, unimplemented distributional proposals (e.g., Maratsos & Chalkley, 1980), the kinds of learning mechanisms that contemporary researchers have considered and implemented ten to focus on highly specific properties of the input...".

However, Redington et al. (1998) also accepted that the current system does not take into account the syntactically ambiguous words. To conclude, the distributional approach is useful but it does not completely address all the issues of child language acquisition.

2.4 Yeap's computational theory of language

Yeap's solution to the problem is based upon identifying what information is available initially to the child and from that, how grammar rules might emerge

and why (Yeap, 2005a, 2005b). The grammar rules that a child develops should not be made up of abstract categories of words such as prepositions, adverbs and others. This is because it is uncertain how a child could detect such categories from their input and this is the cause of the learning problem in the first instance.

To understand what information is available to the child, Yeap made the following observations:

1. Children first communicate via the use of a single word. This implies that they have used a word to refer to some meanings about the world.

2. Children later communicate using small phrases. This implies that they have learned to combine two or more words to express themselves.

When combining multiple words to form meanings for a phrase, Yeap hypothesized that what children have learned is how the meaning of each word is passed among them to construct the meaning of the phrase. Furthermore, given that the input is a linear sequence of words, what is learned initially is how two adjacent words are combined. Note that, there are four basic ways in which two adjacent words could be combined, namely:

- **?R+** to add information to the meaning formed by the words on its right.

- **?R-** to remove information from itself and add it to the meaning formed by the words on its right.

- **?L+** to add information to itself from the meaning formed by the words on its left.

- **?L-** to remove information from itself and add it to the meaning formed by the words on its left.

Children have been shown to be sensitive to the position information of incoming words (Hirsh-Pasek & Golinkoff, 1996) and human language is incremental which builds the grammatical structure word by word (Abney, 1989; Demberg & Keller, 2009). This provides evidence that the above is learnable. However, the above does not capture the wide range of structures that one evidently finds in languages. This posed a major problem for this idea to become a theory of how language works. For example, one of the early problem noted is that of handling of distance attachments. For example, consider the sentence (11) below:

11. The man the police wanted is here.

The object of the verb, *"wanted"*, is the subject of the sentence, which is found on the left side of the verb. How could one attach such objects if the lexical definition of the verb looks for an object on its right? Yeap solved this problem via the use of a stack to form complex objects (such as those denoted by a phrase like *"running up the hill is a good form of exercise"*) and to enable more creative ways to form attachments out of the basic forms.

Figure 2.3 shows how the sentence (11) is parsed. Notice how *"the police"* concept is stacked on top of *"the man"* and the action word *"wanted"*, is then attached to *"the police"* phrase. When the word *"is"* appears, UGE is forced to collapse the stack to form a noun.

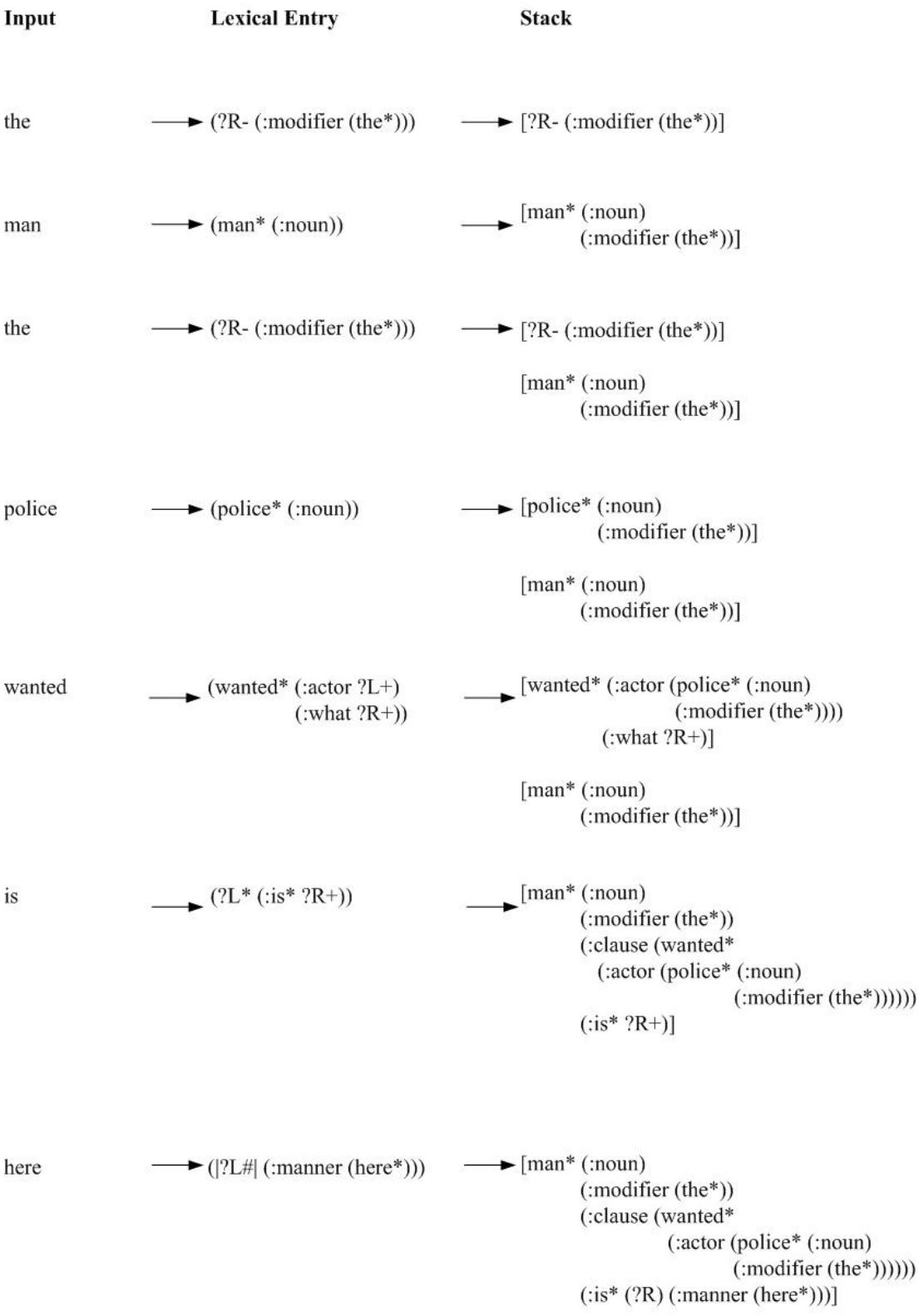

Figure 2.3: The parsing of sentence (11) - *"The man the police wanted is here."*

Note that in Figure 2.3, two new labels , ?L* and ?L#, have been introduced. Having analysed many different English sentences, Yeap introduced five more labels to the four basic labels. The rationale for them are as follows:

1. **?R++:** This is used together with a ?R+ and is for representing dative verbs i.e verbs with two arguments on their right. This label takes in the extra arguments.

2. **?R+=:** Many words contain a ?R+. Examples are: (eat* (:actor ?L+) (:what ?R+)), (eaten* (:what ?R+)), (:to** ?R+), (?L* (:with ?R+)), (?L* (:and ?R+)). The patterns for the first three are unique but the pattern for a preposition and a connective are not. However, since these patterns of words are learned together with the meanings of the words, it is possible that words with similar patterns but with different functions are then grouped differently. To make it easy to distinguish them in the program, the label ?R+= is used and represent prepositions. Thus, the word, *"with"*, is represented as (?L* (:with ?R+=)). Such an extension is introduced to simplify the program.

3. **?R-*:** Similarly there are many different kind of words that are labeled as ?R-. Some examples include *"her"*, *"a"*, *"nice"*, and others. Again the initial labeling scheme failed to provide a finer distinction between articles, adjectives and pronouns for the English language. Both articles and pronouns normally can only appear at the start of a sequence of ?R- words. Hence, a ?R-* is introduced for such words to prevent them appearing in the middle of a sequence of ?R- words. An example sentence would be: *"I gave her a book"*. Without defining *"a"* as ?R-*, one would parse *"her a*

book" as a phrase and that is wrong.

4. **?L*:** This is similar to ?L- except that such a word is immediately attached to the word on its left. If the word on the stack has a ?R+ then it attaches itself after changing ?R+ to ?R. For example, consider the phrase *"eating in"*:

   The dictionary entries for the words *"eating"* and *"in"*:

   ```
   eating - (EATING* (:WHAT ?R+))
   in - (?L* (:IN* ?R+=))
   ```

   Stack created for the word *"eating"* is:

   ```
   [EATING* (:WHAT ?R+)]
   ```

   Stack Created after the word *"in"* comes:

   ```
   [EATING* (:WHAT ?R)
            (:IN* ?R+=)]
   ```

   Here, the ?R+ of *"eating"* is changed to ?R before attaching the word *"in"*.

5. **?L#:** This is similar to ?L- except that such a word is immediately attached to the left regardless of what is on the stack. Here the characteristic of the stack remains the same. For example if the stack has ?R+, it still has ?R+ after attaching ?L#. Consider the phrase *"eating happily"*:

The dictionary entries for the words *"eating"* and *"happily"*:

```
eating - (EATING* (:WHAT ?R+))
happily - (|?L#| (:MANNER (HAPPILY*)))
```

Stack created for the word *"eating"* is:

```
[EATING* (:WHAT ?R+)]
```

Stack Created after the word *"in"* comes:

```
[EATING* (:WHAT ?R+)
         (:MANNER (HAPPILY*))]
```

Here, the ?R+ of *"eating"* is not changed. In short, ?L* and ?L# indicate immediate attachment of this word to the word on the stack. However, ?L# allows more to be added to the word on the stack whereas ?L* would allow something on the right to be added only to itself.

With these additional labels, one could represent all the formal syntactic categories of words as identified by linguists. Their correspondence is shown in Table 2.1. Note that higher-order sentences in a language such as the use of conjunction and clauses to combine sentences to form a new sentence can also be represented by allowing a sentence to be attached to a word. Such attachment is done using two tags :MS1 and :MS2. An example of this use is seen in the parsing of sentence

(12) below:

12. I am happy when I pass my thesis.

The dictionary entry for the word *"when"*:

```
(WHEN* (:MS2 ?R+) (:MS1 ?L+))
```

UGE output for the sentence (12):

```
(WHEN* (:MS2 (PASS* (:ACTOR (I* (:PNOUN)))
                    (:WHAT (THESIS* (:NOUN) (:MODIFIER (MY*))))))
       (:MS1 (I* (:PNOUN)
                 (:AM* (?R (:MODIFIER (HAPPY*)))))))
```

Syntactic categories	Yeap's labeling scheme
Adjective	(?R- (:MODIFIER (BEAUTIFUL*)))
Article or Determiner	(?R-* (:MODIFIER (A*)))
Transitive verb	(EAT* (:ACTOR ?L+) (:WHAT ?R+))
Intransitive verb	(EAT* (:ACTOR ?L+))
Dative verb	(GIVE* (:ACTOR ?L+) (:RECIPIENT ?R++) (:WHAT ?R+))
Gerund	(EATING* (:WHAT ?R+))
Past participle	(EATEN* (:WHAT ?R+))
Infinitive to	(:TO** ?R+)
Adverb	(?L# (:MANNER (HAPPILY*)))
Connective	(?L* (AND* ?R+))
Preposition	(?L* (:TO* ?R+=))
Be verb	(?L* (:AM* ?R+))
conjunction	(WHEN* (:MS1 ?R++) (:MS2 ?R+)) or (WHEN* (:MS2 ?R+) (:MS1 ?L+))

Table 2.1: Yeap's labeling scheme and its equivalence to the syntactic categories of words

2.5 Other child language acquisition theories or approaches

This section gives an overview of other child language acquisition theories. Since it is an overview, the readers are directed to the relevant articles for more details.

2.5.1 Cognitive theory

Cognitive view of language acquisition is based on the child's intellectual development. The child only uses the linguistic structure after they have developed the conceptual ability to understand it. This theory is based on Piaget's four stage cognitive development theory (Piaget, 1969). Piaget argues that a child goes through four separate stages of cognitive development as they are growing up. These stages are universal to all children, however the amount of information acquired differ from child to child. According to Piaget (1969), the language acquisition does not take place until the child is psychologically matured. The child moves from one stage to another when they reach the maturation in one stage through experience. Piaget's four stages are:

- **Sensorimotor stage -** from birth to approximately two years of age. A child relatively has little competence in expressing the environment using images, language, or symbols during this stage. A child has no consciousness of objects or people that are not present immediately at a given moment. According to Piaget, this is a lack of object permanence. Object permanence is the consciousness that objects and people exist continuously even if they are out of sight.

- **Preoperational stage -** from age two to seven. This is the most important stage for language acquisition. Children go through an internal representation of the world that allows them to describe people, events, and feelings using symbols. The thinking of the child is more advanced than in the sensorimotor stage. However, it is still inferior qualitatively to that of an adult. Piaget called this stage as egocentric thought stage. In this stage, the world is viewed entirely from the child's own perspective. Therefore, a child's explanation to an adult can be uninformative.

- **Concrete operational stage -** Children in the concrete operational stage are aged from seven years to twelve years old. In this stage, they have a better understanding of time and space. Piaget called this stage as abstract thinking stage, in which children have limits to their abstract thinking.

- **Formal operational stage -** This stage begins in most children at age twelve and extends to adulthood. Thinking is not connected to events that are observed. A new kind of thinking, that is abstract, formal, and logical, are produce in this stage. A child can think hypothetically and logically to solve problems.

Following Piaget, Drescher (1991) developed a Schema Mechanism, a general learning and concept-building mechanism, based on Piaget's theory. The Schema Mechanism is one of the few implementations of constructivist learning. The Schema Mechanism is intended to replicate key aspects of cognitive development during infancy (refer Drescher (1991) for more detail).

Congnitive theory explains the order in which certain aspects of language are acquired. However, this theory does not explain why language emerges in the first place. Apes' cognitive development is similar to that of young children in the first few years of life, but language acquisition does not follow naturally from their development.

2.5.2 Input / Interactionist theory

This theory views language acquisition based on the interaction between child and the care taker, since language is for communication purpose. Interactionists Bruner (1983) suggests that the adult behavior of talking to a child (example turn taking conversation structure) supports the language acquisition process. Bruner (1983) agrees Chomsky's idea of LAD, however he suggests on top of LAD, there is a Language Acquisition Support System (LASS) which makes the language acquisition task possible. Using LASS, parents or care takers use books and images to develop the child's naming abilities and the child's ability to get involved in conversation.

It is not proven that a child learns more quickly with frequent interactions. However, children in all cultures pass through the same stages in acquiring their first language. In some cultures, the care taker does not adapt a special way to talk to their child. Therefore LASS might be useful, but not essential to language acquisition.

2.5.3 Crystal's theory

Crystal's theory of language acquisition states that the child learns language in five stages (Crystal, 1970). These five stages are:

- **Stage 1 -** Child who is in this stage says something to get what they want, to get someone's attention or to draw attention to something. In this stage, the child does not have a large vocabulary.

- **Stage 2 -** In this stage the child asks questions using *"where" or "what"* followed by noun term like *"where mummy"*. Also the child learns to talk in the characteristic pairs like *big/small* and opposite pairs like *up/down*.

- **Stage 3 -** The child asks a lot of questions which are questions with intonation alone (i.e. making a sentence into a question using the tone of voice), for example: *"John eat in park mummy?"*. In this stage, the basic sentence structure is expanded by inserting adverbs.

- **Stage 4 -** The child, in this stage, uses complex sentence structures to explain things, to ask for explanations using *"why?"* and to make requests like *"shall I do it?"*. The child also starts to use the auxiliary verb to form a sentence and question.

- **Stage 5 -** In this stage the child uses language for all things. The child starts to use nested clauses in the day to day communication like *"If I want this I can have this"*.

The problem with this theory is that the five stages are not clearly defined and overlap with each other.

2.5.4 Selectionist theory

The selectionist theory is developed based on Darwinian evolution thinking. Here, the language acquisition is viewed as a population of grammars with associated weights. Then, given a linguistic environment, the weight is a function of linguistic environment itself and the time since the onset of language acquisition. Learning algorithm accepts the sentence for a particular grammar if the the grammar can analyse the sentence(for more details refer Yang (1999)).

2.5.5 The modular approach - INFANT system

The modular approach views language acquisition as an integration of independent procedural modules (such as Lexical analysis, syntactic analysis module, anaphoric analysis module, etc). Each module is responsible for a specific, predictable and mundane task. The integration of these modules is called INFANT system, which makes up an understanding system. The output of the INFANT system varies and is unpredictable (for mor details refer Buchheit (1993))

2.6 Conclusion

This chapter describes the child language acquisition problem in detail, specifically the current PoS argument in language acquisition. It also discusses language acquisition theories to date along with Yeap's computational theory of language.

From this review, other language acquisition theories are useful in language acquisition, but it is not the way the child acquires language. Since Yeap's theory only

uses the information available to the child, UGE is argued as a viable solution to the child language acquisition problem.

Chapter 3

UGE and Related Parsers

Works on theories of syntax have made significant progress since Chomsky's (1965, 1968) seminal work and numerous parsers have been developed based upon these theories. Some examples of works are Dependency Grammar (Nivre, 2005), Word Grammar (Hudson, 2007), Dependency Unification Grammar (Hellwig, 1986), Relational Grammar (Vodenski, 2009), Categorial Grammar (Wood, 1993) and Link Grammar (Sleator & Temperley, 1993).

While the basic approach underlying these parsers are very different from UGE, some of these parsers were implemented using some form of left/right attachments of words which is very similar to how UGE works. This chapter briefly reviews three of these parsers, namely Dependency Grammar (Section 3.1), Categorical Grammar (Section 3.2) and Link Grammar (Section 3.3), to highlight their differences with UGE. It is important to establish that UGE is not the same as these parsers because humans cannot learn these parsers, according to the PoS argument. UGE is claimed to be learnable.

Before discussing these works, it is worth noting here that there are also many parsers which were developed based upon observing, or reasoning about, how humans (adults) process languages. For example, earlier work debated on what parsing mechanism best modeled the way humans process a sentence. Various models were proposed (such as top-down/bottom-up parsing versus left-corner parsing (Resnik, 1992), and licensing-structure parsing (Abney, 1989)). Then the popular Augmented Transition Networks (ATN) model in the early 70's was also analysed for its psychological plausibility (Kaplan, 1971). Observations regarding how humans resolve ambiguous sentences have led to the development of many models embodying various syntactic parsing strategies or preferences (for a discussion of these early models, see Altmann, 1988). Observations regarding word sense disambiguation led to distributed, connectionist parsing (e.g. Small & Shastri, 1982) and statistical parsing (e.g. Bikel, 2000). More recent work includes the development of a human-like parser constrained by the use of a short-term memory (Schuler, AbdelRahman, Miller, & Schwartz, 2010) and another using recursive neural networks to implement incremental parsing (Costa, Frasconi, Lombardo, & Soda, 2003).

Like UGE, these works attempt to model human parsing but unlike UGE, these works are based upon observing how adults process language and not on how a child learns its first language. Furthermore, these works assume that adults have some form of formal grammar rules as part of their knowledge of language and most of them are aimed at developing more powerful and efficient parsers. The development of UGE is not based upon a desire to develop a new and efficient

parser but rather to develop one for explaining how children acquire their first language. Given these different motivations, these works are viewed as peripheral to our interests and therefore will not be reviewed in this thesis.

Prior to concluding, this chapter briefly reviews, in Section 3.4, the existing implementation of UGE (i.e. Yeap's implementation) and in particular its limitations as a practical parser. Section 3.5 draws the chapter to a close.

3.1 Dependency Grammar

A Dependency Grammar makes explicit the syntactic connection between two words or constituents using only asymmetric relations. The binary relation is asymmetric since one word or constituent is viewed as dependent on the other. The latter is referred to as the head or governor and the former, as dependents. The relation is thus referred to as a dependency and the grammar, a Dependency Grammar. Figure 3.1 shows an example of a dependency tree generated for the sentence *"I eat an apple"*. In contrast, a phrase structure grammar would produce a constituent tree and an example is shown in Figure 3.2.

According to Nivre (2005), the development of Dependency Grammar has a long history and not surprisingly, many grammatical theories based on the notion of a Dependency Grammar have been developed. Some well-known ones are Word Grammar (Hudson, 2007), Dependency Unification Gramar (Hellwig, 1986), Functional Dependency Grammar (Jarvinen & Tapanainen, 1997), and Meaning Text Theory (Melcuk, 1988). These theoretical works focus on the cri-

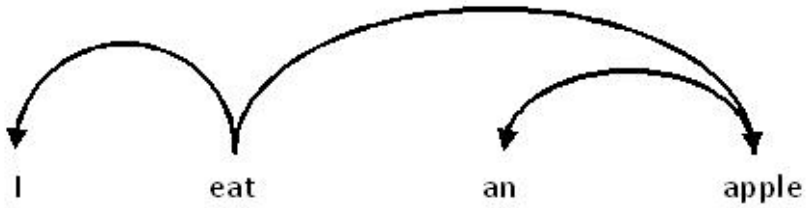

Figure 3.1: A dependency tree - *"I eat an apple"*

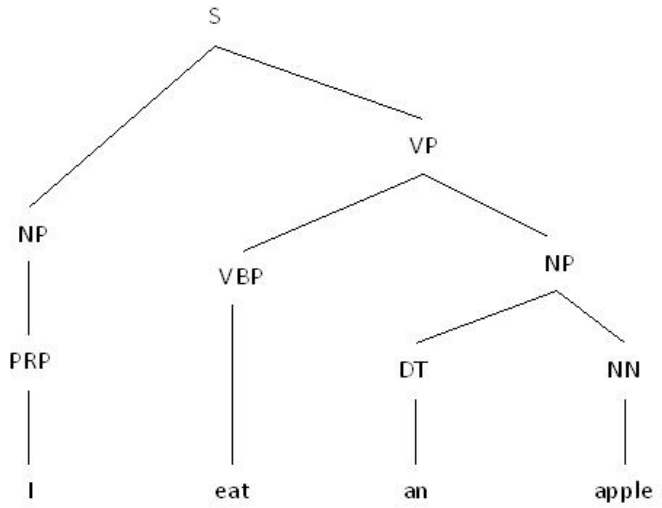

Figure 3.2: A constituency tree - *"I eat an apple"*

teria for establishing dependency relations and the rules for processing them. From our perspective, what is interesting is how the dependency relations are parsed or, in other words, how the rules are formulated (and hence learned). It turns out that the formal grammar rules, expressed in terms of dependency relations, could be captured using the following axioms first defined by Hays (1964) and Gaifman (1965) (see Debusmann, 2000):

1. x(w_1,...,*,...,w_k): w_1...w_k are dependent on x
2. x(*): x is a leaf node
3. *(x): x is a sentence root node

The star * indicates the position of governor x in the linear order of words w_1...w_k. A Dependency Grammar for parsing simple sentences like *"I eat an apple"* is as shown below:

*(V)
V(N,*,N)
N(Det, *)
N(*)
Det(*)

The above looks much like a context-free grammar and this has apparently stopped further interests in the study of this formalism (Nivre, 2005). However, in recent years, the field has re-ignited and many richer dependency grammars have been developed and more powerful parsers have been built. For example,

Eisner (1998) developed several probabilistic models for dependency parsing and evaluated them using supervised learning with data from the Wall Street Journal section of the Penn Treebank. More recently, Buch-Kromann (2006) developed a dependency-based model that emphasized parsing as an optimization problem and McDonald (2006) developed discriminative learning and spanning tree algorithms for dependency parsing.

Dependency Grammar is thus one of the earliest forms of grammar that emphasizes directly on computing the binary connections between words/constituents. In this respect, it is similar to UGE. However, at the heart of its formalism, a set of formal grammar rules exists for producing the desired outputs. This implies that Dependency Grammar, from the UGE perspective, is another kind of formal grammar and, according to the PoS argument, it cannot be learned. This strongly distinguishes Dependency Grammar from UGE.

3.2 Categorial Grammar

Categorial Grammar (CG) is one of the oldest "lexicalised" theories of grammar. Syntax and semantics properties of the lexicon are used to define the syntactic form of lexicon. Variations of Categorial Grammar include Head-driven Phrase Structure Grammar (HPSG), Lexical Functional Grammar (LFG), Tree-Adjoining Grammar (TAG), Montague Grammar, Relational Grammar and certain version of the Chomskean theory (Steedman, 1999).

The formal property of grammar is represented as follows in Categorical Gram-

mar (from Steedman, 1999):

S – NP VP

VP – TV NP

TV – {likes, sees, }

Here, "result leftmost" notation is used in which a/b and a\b represents functions from b to a. The "/" represents that the argument is to the right and the "\" represents that the argument is to the left. Each word in Categorial Grammar is associated with one or more symbolic expressions. This can be a single symbol or combination of symbols using "/" and "\".

Consider the sentence *"John likes Mary"*. The grammatical categories of the words *"John"*, *"likes"* and *"Mary"* are represented as follows in Categorial Grammar:

John – NP

Likes – (S \ NP) / NP

Mary – NP

Categorial Grammar combined the categories of words using the following two rules.

Rule 1 - X / Y Y = X

Rule 2 - Y X \ Y = X

So that the string *"John likes Mary"* is derived to sentence **S** as follows (reproduced from Steedman (1999); Wood (1993)):

John	likes	Mary
NP	(S \ NP) / NP	NP
	S \ NP	(applying Rule 1)
S	(applying Rule 2)	

Categorial Grammar is thus similar to UGE in that it combines words appearing to its left and right to form higher constituents, and that parsing is driven with information encoded at the lexical level. Thus compared with Dependency Grammar, Categorial Grammar has taken a step further in utilizing binary relations between words as a basis for parsing. However, unlike UGE, what is encoded at the lexical level are the formal grammar rules themselves. UGE encoded the equivalent of "syntactic knowledge" at the word level and has to learn a set of heuristics to parse the sentences. Again, we conclude that Categorial Grammar is yet another implementation of the formal grammar of languages.

3.3 Link Grammar

Link grammar is a formal grammatical system defined as follows (Sleator & Temperley, 1993):

> "A sequence of words is in the language of a Link grammar, if there is a way to draw links between words in such a way that (1) the local requirements of each word are satisfied, (2) the links do not cross

> when drawn above the words, and (3) the words form a connected graph, i.e. The links suffice to connect all the words of the sequence together."

Link Grammar combines adjacent words based upon some attachment rules and pattern matching mechanism. The Link Grammar dictionary contains the linking requirement of each word. In the linking requirement, the **"+"** sign or **"-"** sign indicate the direction. For example, **"+"** sign is used to point the direction of right and **"-"** sign is used to point the direction of left. The operators **&**, and **OR** are used to express multiple linking requirement for one word.

For example, consider the sentence *"The cat chased a snake"*. The dictionary entry of those words (*"the", "cat", "chased", "a", and "snake"*) are as follows (Sleator & Temperley, 1993; Grinberg & Sleator, 1995):

Words	**Formula**
a	D+
the	D+
cat	D- & (O- or S+)
snake	D- & (O- or S+)
chased	S- & O+

Here, D is denoting "Determiner", O is denoting "Object" and S is denoting "Subject". In order to make a valid link, + sign looks for - sign of the same value. For example, D+ looks for D- on its right to make the valid link and D- looks for D+ on its left to make a valid link. When all the words in the sentence

perform links (i.e. no words are left alone), then that particular sentence is a valid sentence.

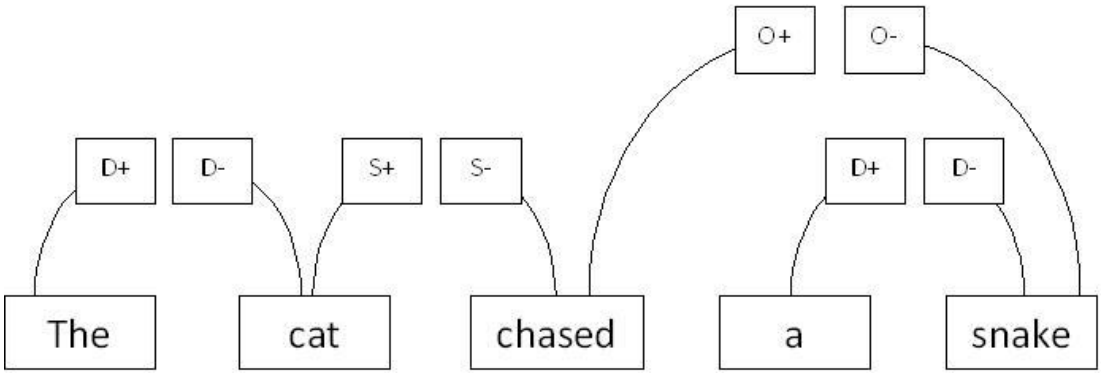

Figure 3.3: Parsing the sentence *"The cat chased a snake"* - before linkage performed (reproduced from Sleator and Temperley (1993))

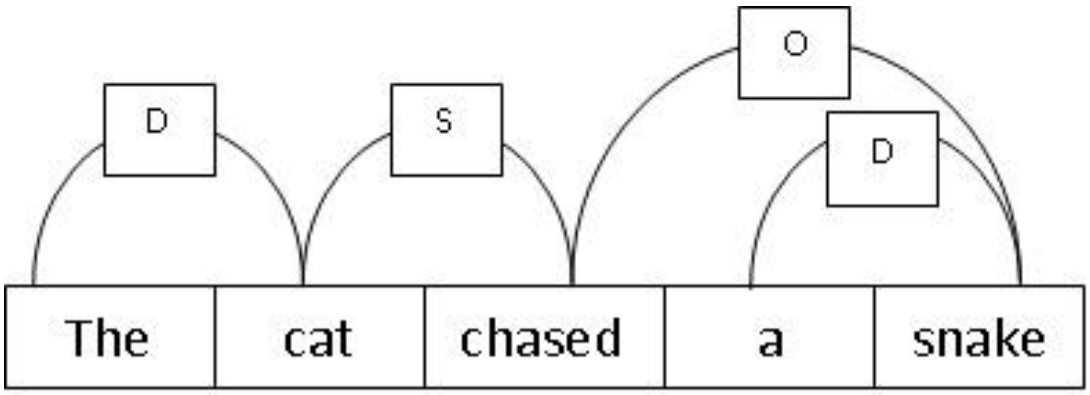

Figure 3.4: Parsing the sentence *"The cat chased a snake"* - after linkage performed (reproduced from Sleator and Temperley (1993))

In Link Grammar, D+ word looks for D- word for a linkage (a set of links that prove a sequence of words). If the D+ word could not find the D+ word on its right then there is no link for that word. The above diagram shows how the linkage is performed.

Link Grammar is thus very similar to Categorial Grammar. Its creator showed how a Categorial Grammar can be transformed into a Link Grammar and argued that the latter is an efficient way to parse the former. Two other distinctions between Link Grammar and UGE could also be highlighted:

1. Link Grammar uses a pattern matching mechanism to link words together. For example if the word on its left has D+ label, then it is looking for D- label on its right to make a successful link. If there is no D- on its right then no linkage is formed. In UGE, a "link" is formed by the functional role of a word. If the word is encoded as having a ?R+ slot, then it expects an object from its right to be "linked" to it to form a semantic entity. Thus UGE does not have pre-defined links (in the sense of finding the appropriate piece of the puzzle). Rather, it accepts that the pieces are there already and the question is how to make sense of it. The later must be learned by examples. UGE is arguably more interesting in that it views language as "understanding" what is said rather than "restricting" how one is going to say something.

2. In Link Grammar, each word must have a pre-defined linkage and multiple linkages are possible. However, in UGE, noun words do not need to have a label. This is because these words carry whatever meaning one assigns to it; they do not play a functional role in composing a sentence. However, they function as "subject" because a word appearing on its right behaves as, say, a verb role. Similarly, they function as "object" because a word appearing on its left behaves as, again, a verb role.

3.4 UGE

Yeap's implementation of UGE consists of only one module (see Figure 3.5). The input to the system is an English sentence or sequence of words and the output is a parse tree for the sentence which Yeap refers to as a Mental sketch [1] of the sentence. Lexical entries for each word is stored in a dictionary file which is also integrated into the main module.

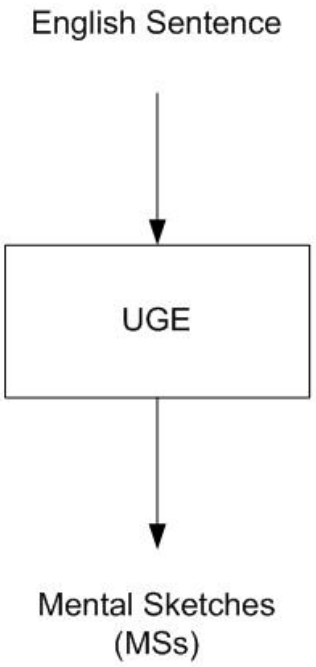

Figure 3.5: Yeap's Implementation - UGE

As already noted in the previous chapters, Yeap uses stacks to process words in the sentence and Table 3.1 shows its main routine (Yeap, 2005a).

[1] It is called a sketch because the output is not a full interpretation of the sentence. The latter is referred to as a Mental Picture and its construction is beyond the scope of this thesis

Interpret():

1. Perceive the next word and generate its object.

2. If it has a ?L, do:
{Pop a semantics object from the stack and use it to replace the ?L.
Push the result back onto the stack.
Go to step 6}.

3. If it has a ?R, push it onto the stack and go to step 6.

4. If it has no ?L or ?R,
check if the semantics object on the stack has a ?R.
If it has, do:
{Use the incoming semantics object to replace the ?R.
Go to step 6}.

5. Push the current semantics object onto the stack.

6. Repeat the process.

Table 3.1: Yeap's main algorithm for UGE (Yeap, 2005a)

Yeap's implementation works fine for testing simple sentences consisting of words only. Figure 3.6 provides another example of parsing a sentence using UGE. The sentence and the lexical entries are shown below:

1. I saw her in the park with a telescope.

I - (I* (:PNOUN))
Saw - (SAW* (:ACTOR ?L+) (:WHAT ?R+))
her - (HER* (:PNOUN))
in - (?L* (:IN* ?R+=))
the - (?R-* (:MODIFIER (THE*)))
park - (PARK* (:NOUN))
with - (?L* (:WITH* ?R+=))
a - (?R-* (:MODIFIER (A*)))
telescope - (TELESCOPE* (:NOUN))

Again, note that when the noun words are read (i.e. *"her"*, *"park"*, and *"telescope"*), two stacks are created (the second stack for *"telescope"* is not shown). However, of interest in this example is the way in which UGE handles prepositions. As noted in Chapter 1, a significant difference between traditional parsing and the approach taken in UGE is that the output produced by the former is a legal parse of the sentence whereas the latter is a representation to assist interpretation. This difference is best illustrated in the handling of prepositions. As Figure 3.6 shows, prepositions in UGE are always attached to whatever is on the stack as long as it is not a collapsed stack or a complete sentence. Traditional

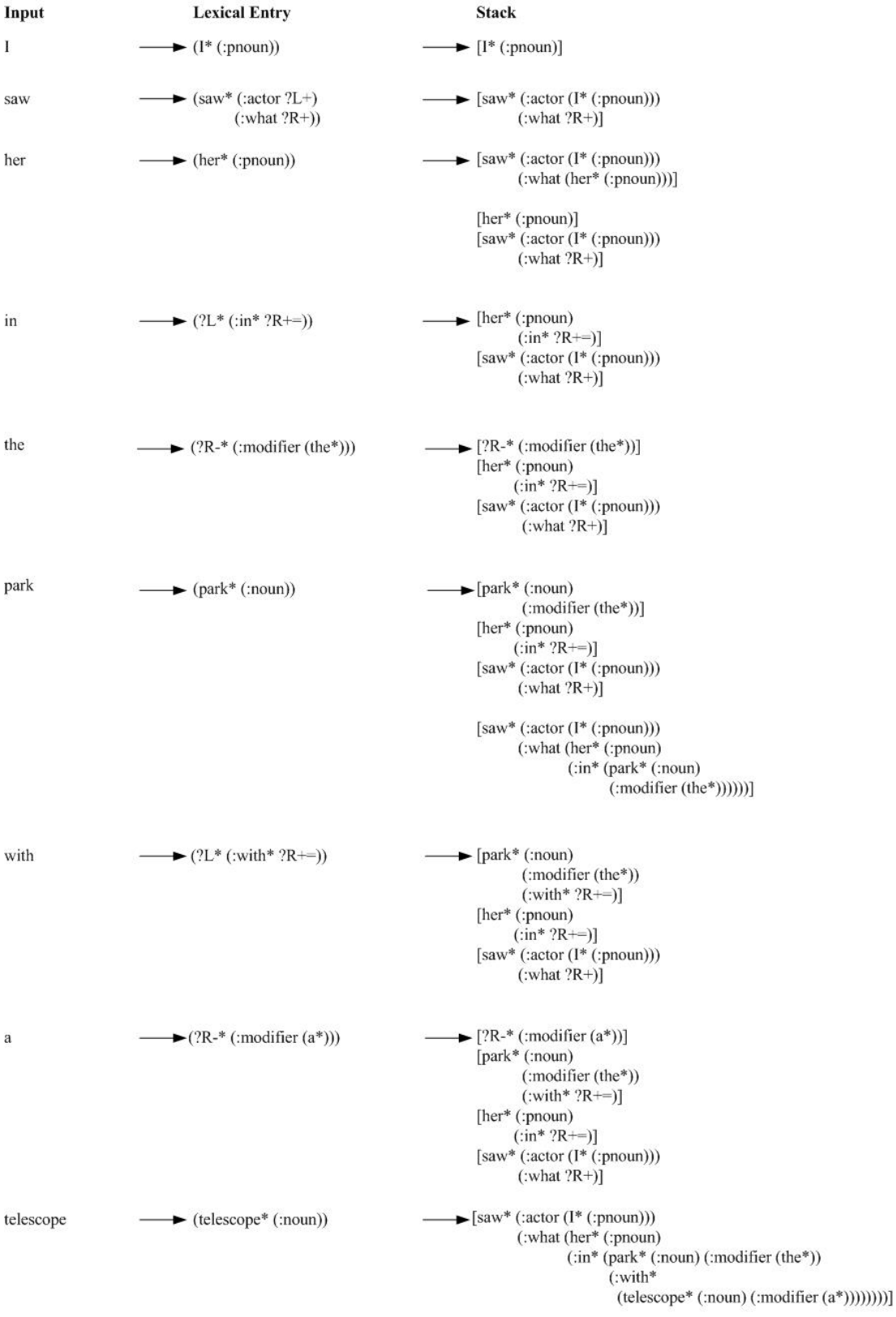

Figure 3.6: The parsing of sentence (1) - *"I saw her in the park with a telescope."*

parsing will produce different legal parses of such sentences. Figure 3.7 shows two outputs produced using the Link Grammar. Much has been discussed in the literature regarding how one might resolve this problem (Hindle & Rooth, 1993; Zhao & Lin, 2004; Asch & Daelemans, 2009).

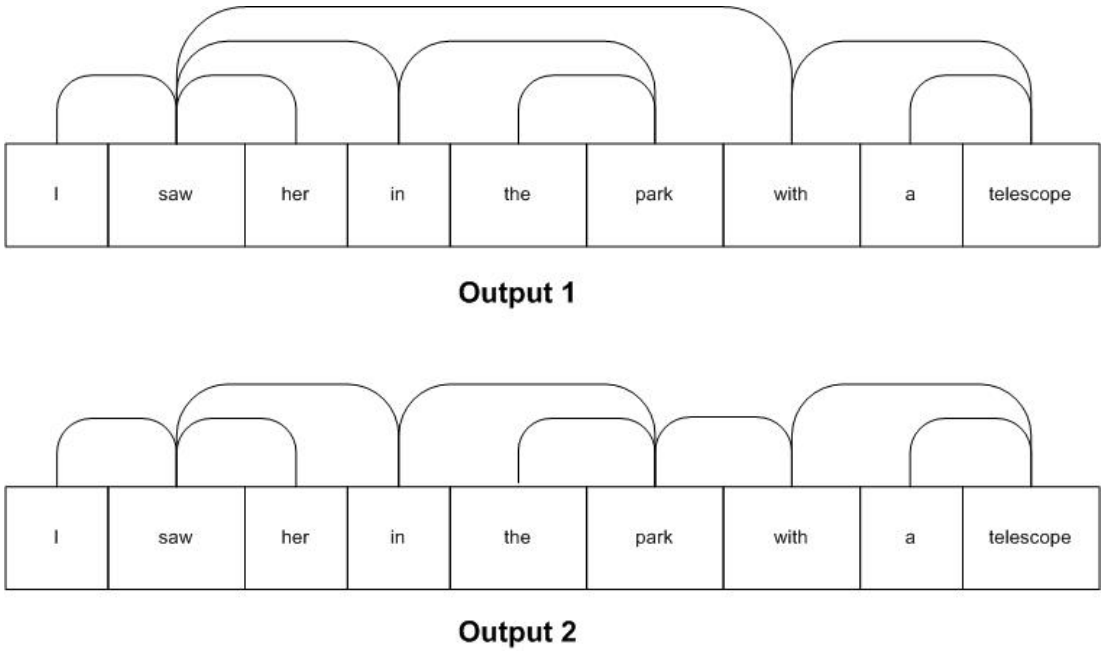

Figure 3.7: Link grammar outputs - for sentence (1)

UGE does not produce different "correct" attachments. Rather, it always attaches them to the noun that is on the stack and leaves the final interpretation to the interpreter. Thus, in the final output produced for sentence (1), the interpreter will have to decide, using whatever context available, whether *"in"* is attached to *"her"* or promoted to attach to *"saw"*. If the former, then similarly, it has to decide if *"with"* is attached to *"park"* or *"her"*.

Below shows the three different sentences that we have analysed so far using

UGE:

- John ate an apple (Figure 1.1)
- The man the police wanted is here (Figure 2.3)
- I saw her in the park with a telescope (Figure 3.6)

Each highlights a different aspect of the UGE method and demonstrates the difference between UGE and the formal parsers described earlier.

However, the parsing of real world sentences poses a more complex issue. Consider another example of a real world sentence, such as sentence (2) below:

2. The Government has refused to consider the applications until Mr Zaoui's case is resolved (he is on bail awaiting a review of his SIS security risk certificate, which could mean his deportation).

These sentences (like (2) above) often have special symbols, quotation marks, and nested clauses; all of which need extra processing. Furthermore, to extend UGE to process a text as opposed to individual sentences, UGE would also need to have some abilities to handle errors, to detect automatically the start of a new sentence, to select the best outputs and so on. From using UGE to process numerous individual sentences, the following limitations of UGE were identified:

1. The current implementation only handles words. However, the real world complex sentences have special symbols along with words which need to be handled differently.

2. The current dictionary makes no attempt to guess the lexical meaning of the word based on its context or characteristic. For example, if the word in a sentence is not in the dictionary, this implementation assumes it as a NOUN term. This leads to a wrong interpretation or, at worst, fails to parse the sentence.

3. The current dictionary is extremely limited, both in the number of words it has (about 5000 words) and the roles of words. Thus, many words are found with missing roles. For example, the transitive verb role of the word, *"eat"*, may not be found in the dictionary even though the word is in the dictionary.

4. The current labeling scheme is not adequate. For example, the relative clause usage of the word *"who"* has not been defined. Consequently, UGE could not handle sentences such as *"John who is a doctor likes to eat an apple"*. A new label must be introduced with care.

5. The current rules for processing each label are also not adequate. For example, consider the two sentences: *"I eat an apple"* and *"I normally eat an apple"*. Since the transitive verb *"eat"* can follow a pronoun or an adverb, two rules are needed to process the verb. Currently, these rules are discovered manually via encountering examples of their use and are thus done in an ad hoc manner. No learning algorithm is proposed but a more systematic way of discovering these rules is introduced.

6. The current implementation returns all possible MSs produced. There is no decision making mechanism to eliminate the non-sensible combination

based on context. This leads the process to return multiple MSs as a result. For example, consider the sentences *"He has been driving from Auckland to Hamilton"* and *"His driving is good"*. In the first sentence GERUND definition of the word *"driving"* is sensible rather than the NOUN definition. In the second sentence NOUN definition is more sensible. By adding the decision making rules to the system, we could prevent multiple parses. Again, how do we know which decision making rules are appropriate?

7. The current implementation does not validate the returned result for its completeness. In other words, if it fails in the middle of parsing the sentence, then the next fetched word is assumed to be the start of the sentence for parsing. This leads to having an incomplete parse.

8. In the current implementation, there is no algorithm to select the best parse among the multiple parses returned. It always returns the first parse as its final result which may not be always right.

UGE is extended to overcome the above limitations and the new version is discussed in the next chapter.

3.5 Conclusion

This chapter critically reviews three language parsers which are built based upon the idea of parsing using binary relations between words/constituents. This idea is also a key idea in UGE. However, the review shows UGE differs from them on one very important aspect:

- All these parsers provide a different implementation of the formal grammar of language as specified by linguists whereas UGE does not.

Theoretically, UGE first learns the different labeling of words from observing their presence in a sentence and the meaning of the sentence itself. It then learns a set of rules to deal with the complex handling of these labeled words to enable the creative use of language. However, its current implementation is limited, both in terms of its coverage of the language and its ability to handle efficiently other issues surrounding the practical parsing of text.

Chapter 4

Implementation

This chapter discusses the extension of the Universal Grammar Engine (UGE) for parsing real world English sentences. The new implementation has four main modules (see Figure 4.1), which are: 1) Pre- processor - Prepares the input for parsing. 2) UGE - Produces the Mental Sketches (MSs) from the pre-processed sentence. 3) Dictionary - Returns the lexical entries for a given word. 4) PostUGE - Selects the best Mental Sketch (MS) for the sentence from the MSs return by the UGE model. With these modifications, UGE becomes a much more powerful parser and most of the problems identified in Chapter 3 have been dealt with.

Sections 4.1 to 4.4 detail each of the components of UGE along with supporting examples. Section 4.5 describes a test procedure developed in this research to ensure the consistent development of UGE as new rules are found. Section 4.6 gives an example of parsing a real world sentence taken from the New Zealand Herald newspaper. Finally, Section 4.7 concludes this chapter.

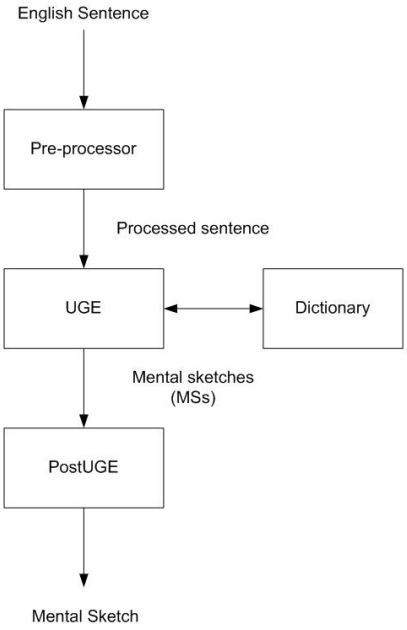

Figure 4.1: The UGE Model - Main models of the program

4.1 Pre-processor

Pre-processor module takes the sentence or sequence of words as input and performs the following tasks:

- **Identify sentence boundary** - The algorithm used to identify the sentence boundary is simplified. This is done by identifying the word sequence ending with period or question mark or exclamation mark or ellipses followed by space or new-line character with capitalised word. However, there are few abbreviations which leads to the incorrect sentence boundary using this method (e.g. *"Mr. "*). The program detects nearly forty two abbreviations separately before detecting the sentence boundary (see appendix A.1 for the full list).

- **Split sentence into tokens** - This process splits the sentence into meaningful chunks or tokens, then further categorises them as word, number, punctuation or special symbol. For example, the tokenised output for the sentence *"I eat 10 apples."* is given below:

```
((#S(LEXICAL-TOKEN :STRING "i"
                   :TYPE :WORD
                   :FLAVOR :STANDARD
                   :CASE :CAPITALIZED
                   :PROPERTIES NIL
                   :POINTER NIL)
 #S(LEXICAL-TOKEN :STRING "eat"
                  :TYPE :WORD
                  :FLAVOR :STANDARD
                  :CASE :LOWERCASE
                  :PROPERTIES NIL
                  :POINTER NIL)
 #S(LEXICAL-TOKEN :STRING "10"
                  :TYPE :NUMBER
                  :FLAVOR :INTEGER
                  :CASE NIL
                  :PROPERTIES (:VALUE 10)
                  :POINTER NIL)
 #S(LEXICAL-TOKEN :STRING "apples"
                  :TYPE :WORD
```

```
:FLAVOR :STANDARD
:CASE :LOWERCASE
:PROPERTIES NIL
:POINTER NIL)))
```

- **Group/Ungroup tokens into meaningful chunks** - Some words or symbols can be grouped together to produce meaningful chunks for parsing purposes. Then in the parsing process, these grouped tokens are considered as a single token. They are:
  - Compound Nouns - In English, proper names are capitalised and for the parsing purpose the process combines these words together to make one meaningful name component. Here, the process identifies these types of nouns by identifying the capitalised words and combines them together as a hyphenated one word (e.g. *"New-Zealand*) after eliminating list of known capitalised words like starting word of the sentence, titles and some types of pronouns (e.g. *"I", "Mr.", "Tomorrow" etc.*).
  - Compound words - This process is the same as above except that it combines any two words to make a meaningful component. Some examples of such words are *"ahead-of", "because-of"* and etc. The full list of compound words is given in appendix A.2.
  - Contractions - This process splits the contractions into two separate words. For example YOU'RE becomes *"YOU"* and *"ARE"*. For some contractions like MARY'S or MARY'LL', the process separates them into two word like *"MARY"* and *"'S"* or *"MARY"* and *"'LL"*. Then

the parser identifies the contraction whether it is *"IS or HAS"* or *"WILL or SHALL"* in the parsing process based on its context.

- Speech block - This process identifies the speech part within a sentence by finding the starting and ending of the single or double quotation marks, and specifies the content within the quotation as speech block. For example, consider the sentence *I said "Mary is ill"*. Here the phrase *"Mary is ill"* is a speech block. If the sentence is *I said Mary is ill*, then there is no speech part.
- Other blocks - This process identifies the grouped words segment together by finding the words within the brackets/dashes/double dashes and tags them as a block.
- Special Symbols - This process identifies the special symbols and/or word/number segments which can be grouped together and groups them as one component. Here we group the arithmetic equations (e.g. 3 * 4 = 12), numerals (e.g. 25 thousand) and money (e.g. 12 sen).

The sample Pre-processor output for the sentence *"But if Berry was to be released with no more than a "bad luck, my boy - hope you learned something by all this", then the wrong message would be going out, said Mr Neels."* is given in appendix A.3. Here each word or symbol in the sentence is considered as token which enables to capture the characteristic of the word or symbol.

4.2 Dictionary

According to Yeap's theory, each word has a lexical entry which indicates how the word can be used to form a sentence. There could be more than one entry, denoting multiple usage. For example, consider the dictionary entry for the word *"play"*:

```
((PLAY* (:ACTOR ?L+) (:WHAT ?R+))
 (PLAY* (:NOUN))
 (PLAY* (:ACTOR ?L+))
 (PLAY* (:ACTOR ?L+) (:RECIPIENT ?R++) (:WHAT ?R+)))
```

In the new implementation, the dictionary is separated into a new module and provided with smart functionality. The current dictionary module has 145699 entries. This module is equipped with the following smart functions:

- Dictionary module has a cache file with frequently accessed words for easy access.
- If the lexical entry of the word can be decided without accessing the dictionary, then generate the required entry. For example, if the word has first letter capitalised and it is not the beginning of the sentence or the upper-case word or mixed case word then the smart detector returns its lexical entry as NAME entity (NOUN term) without accessing the dictionary data set.
- If the word is not found in the dictionary, then the smart detector tries to guess the lexical entries of the word from its context and characteristic.

  - Characteristic: It uses the suffix of the word to detect the lexical entry or entries for particular word. For example, *"ing"* for GERUND definition, *"ed"* for verb and PAST PARTICIPLE definition and *"ly"* for ADVERB definition.
  - Context: This detects the lexical entry from its previous word in the sentence. For example if the previous word is AUX-VERB or INFINITIVE-TO, then smart detector returns verb definition. Similarly, if the previous word is BE-VERB, then it returns PAST PARTICIPLE.

- Interface function which enables us to add and remove entries to and from the data set.

4.3 UGE

UGE module is an extended version of Yeap's module. First of all Yeap's module is modified to handle the Pre-processor output as an input to the system and the new algorithms, final-interpret() and final-check(), are added to validate the returned results of the UGE module. Final-interpret() checks each stack created by the interpret() function to make sure the stack has all the words in the sentence along with compressing the stack for final output. This process prevents impartial parses of the sentence. Final-check() validates the return results of UGE module by checking whether the stack has the sentence constructor such as verb, be-verb or conjunction. This process makes sure the MSs returns by UGE module are actual parse trees for a sentence and not for a phrase or the result of a partial parse. The Table 4.1 shows the new top-level algorithms for UGE.

Apart from these two new algorithms, the original UGE module has been significally extended to handle the limitations of Yeap's implementation as identified in the previous chapter. These works form a major part of this research and they include:

- Identifying missing lexical entries
- Identifying new labeling schemes
- Developing new rules for new cases
- Developing decision making rules to eliminate the non-sensible combination

Note that the above is not a list of disjoint problems but rather a set of problems that need to be solved together, once a new word is entered into the dictionary (i.e. starting with the first problem on the list). In other words, when a missing entry is found, we have to identify its appropriate labels or introduce new ones for it. We then have to ensure that the existing rules for handling that word are adequate and if not, we need to add new rules. The new rules should not produce any side effects that will affect the correct workings of existing rules. Then we need to generate the necessary decision making rules for its use.

Sections 4.3.1 to 4.3.4 describes work done for each of the problems identified above.

Interpret():

```
Step1: FOR all the pre-processed tokens in the sentence
       Retrieve the lexical entries from the dictionary
       Step2: FOR all the lexical entries for the token
              IF the stack is empty
                     Create a new stack
                     Create MS from the lexical entry
                     Push the result into the stack
                     Go to Step2
              ELSE
                     Pop the MS from the stack
                     Call the Process for MS and the lexical entry
                     Push the result MS into the stack
                     Go to Step2
       Go to Step1
```

final-interpret():

```
Step1: FOR all the stacks of MS created for the sentence
       IF the stack does not have all the words in the sentence
              Delete the stack
              Go to Step1
       ELSE
              Back process the MS
              Clear the content in the stack
              Push the processed MS into the stack
              Go to Step1
```

final-check():

```
Step1: FOR all the stacks of MS created for the sentence
       If the MS is not a valid parse for the sentence
              Delete the stack
              Go to Step1
       Return the list of MSs for the sentence
```

Table 4.1: Top-level algorithms for UGE - Interpret(), Final-interpret() and Final-check()

4.3.1 Missing lexical entries

The missing lexical entry can be found by checking the word in the data set using get-word function. If this function returns :UNKNOWN key, then it indicates that that particular word is not in the data set. The lexical entries are added using the insert-corex function and are deleted using the delete-core function. The insert-corex function takes key word and list of lexical entries as an argument and inserts them into the dictionary data set. For example, the following entry inserts the transitive and intransitive verb definition for the key word *"eat"*. Multiple entries can be inserted using list of multiple key words.

```
(INSERT-COREX '(("eat" ((EAT* (:ACTOR ?L+))
                        (EAT* (:ACTOR ?L+) (:WHAT ?R+))))))
```

The delete-core function removes the key word from dictionary data set. This function takes list of key words as an argument. The following entry removes the key word *eat* and *apple* from the dictionary data set.

(delete-core '("eat" "apple"))

Initial dictionary data set is constructed by importing Yeap's dictionary data into the data set. There are many researchers and students who have contributed to develop the current dictionary data set by inserting new entries. Each word is identified with the English grammatical categories by examining the relevant example sentences, then the labeling scheme is identified based on Yeap's theory and then inserted using insert-corex function. Since manual verification is needed, no automatic import is performed from online dictionary. For example, dictio-

nary definition for the word *"play"* is described as NOUN, TRANSITIVE VERB and INTRANSITIVE VERB. However, consider the sentence *"John plays me a piano"*. Here *"plays"* is used as DATIVE VERB. Therefore, manual verification and example sentences are very important to decide in adding lexical entries to the dictionary. Current dictionary data set has 145699 entries. Roughly, 140000 new entries have been entered since Yeap's earlier model.

4.3.2 Missing labeling schemes

UGE has the following labels which are equivalent to the following grammatical categories of the English language (see Chapter 2):

- **Adjective** - (?R- (:MODIFIER (BEAUTIFUL*)))
- **Article or Determiner** - (?R-* (:MODIFIER (A*)))
- **Transitive verb** - (EAT* (:ACTOR ?L+) (:WHAT ?R+))
- **Intransitive verb** - (EAT* (:ACTOR ?L+))
- **Dative verb** - (GIVE* (:ACTOR ?L+) (:RECIPIENT ?R++) (:WHAT ?R+))
- **Gerund** - (EATING* (:WHAT ?R+))
- **Past participle** - (EATEN* (:WHAT ?R+))
- **Infinitive to** - (:TO** ?R+)
- **Adverb** - (?L# (:MANNER (HAPPILY*)))
- **Connective** - (?L* (AND* ?R+))

- **Preposition** - (?L* (:TO* ?R+=))
- **Be verb** - (?L* (:AM* ?R+))
- **Conjunction** - (WHEN* (:MS1 ?R++) (:MS2 ?R+)) or
  (WHEN* (:MS2 ?R+) (:MS1 ?L+))

Some new labels were introduced when parsing real world sentences. These labels were identified when testing more sentences. Once the new category is identified and before introducing a new label for it, many more similar sentences are collected from the newspaper articles and other online resources to examine how it is used. Then, the program is modified to handle words with the new labels. These new labels are:

- **Wh-word** - (?L* (:WHO* (:MS1 ?R+)))
  wh-words are used to connect relative clauses in English grammar. For example, in the sentence *"John **who** is a doctor eats an apple"*, *"who"* is a relative clause which is named as wh-word in the UGE module. Yeap's uses ?L* to attach itself to the left side and uses either (:MS1 ?R+) or (:MS2 ?R+) to attach clause from its right side. Thus, these words have the labels as shown above.
- **Wh-noun** - (WHAT* (:NOUN) (:WH-NOUN* (:MS1 ?R+)))
  This category functions as a noun term that also contains a relative clause in itself. Consider the sentence, *"**What** Alice did annoyed me"*, here, *"what"* functions as a noun term and *"Alice did"* clause needs to be attached to *"what"* to fulfill the meaning of the subject of the verb *"annoyed"*. Consider

the relative clause attached to the noun term *"apple"* in the phrase *"Apple which"*:

```
(APPLE* (:NOUN)
        (:WHICH* (:MS1 ?R+)))
```

Similarly, the labeling scheme is selected as (WHAT* (:NOUN) (:WH-NOUN* (:MS1 ?R+))) for wh-noun words. Here, the key *:WH-NOUN** added to indicate it is a special case of noun which has relative clause attached to itself.

- **Preposition with MS1** - (?L* (:AFTER* (:MS1 ?R+)))

  This category is same as preposition, however its object is a clause rather than a noun term. The argument is similar to wh-word case. This case is only used if it is preceded with a noun term. The example sentence for this category is *"The calls for the family to be allowed in before Mr Zaoui's case was settled were ludicrous."*. MS created for this sentence is:

```
(CALLS* (:NOUN)
        (:MODIFIER (THE*))
        (:FOR* (FAMILY* (:NOUN) (:MODIFIER (THE*))
                        (:TO**
                         (BE*
                          (:WHAT (ALLOWED*
                                  (:WHAT ?R)
                                   (:MANNER (IN*)))))))))
        (:BEFORE* (:MS1 (CASE* (:NOUN)
```

```
                                    (:MODIFIER (MR*)
                                               (|ZAOUI'S*|))
                                    (:WAS* (SETTLED*
                                            (:WHAT ?R))))))
          (:WERE* (?R (:MODIFIER (LUDICROUS*)))))
```

- **Verb with MS1** - (SAID* (:ACTOR ?L+) (:MS1 ?R+))

  This verb takes a clause as an object instead of noun term. In the sentence *"She said she started screaming after seeing that"*, the verb *"said"* take a clause *"she started screaming after seeing that"*, which is a complete sentence by itself, as its object. Here is the MS created for this sentence:

```
(SAID* (:ACTOR (SHE* (:PNOUN)))
       (:MS1
        (STARTED*
         (:ACTOR (SHE* (:PNOUN)))
         (:WHAT
          (SCREAMING* (:NOUN)
                      (:AFTER* (SEEING*
                                (:WHAT
                                 (THAT* (:NOUN)))))))))))
```

- **Time object** - (?L- (THIS* ?R+))

  These words are used to indicate the time of an event. It is similar to adverbs *"yesterday, morning, today, evening, etc."* and the phrases *"in the morning, at noon, etc."*. There is a need for the time object because in the

sentence *"I met John this morning"*, the phrase *"this morning"* denotes the time of the event. This is different from preposition, because it only precedes special kind of words called *time-words*, which denotes time of the event like *"MORNING, EVENING, etc."* or another time-object like *"late this morning"*. Since it attaches to the clause on its left, the labeling scheme is developed as above. Here, we do not use ?L*, since ?L* attaches itself to its left immediately. However, time-object only attaches itself to a noun term or complete sentence, which carries a complete meaning. The MSs created for the sentences *"I met John this morning"* is:

```
(MET* (:ACTOR (I* (:PNOUN)))
      (:WHAT (JOHN* (:UNKNOWN) (:NAME)))
      (:TIME (THIS* (MORNING* (:NOUN)))))
```

So far, there are eight time object and forty five time-words identified. They are:

Time-objects:

```
(THIS* LATE* NEXT* YESTERDAY* TOMORROW* LAST* ALL* EACH*)
```

Time-words:

```
(NIGHT* NIGHTS* AFTERNOON* AFTERNOONS* TIME* MINUTE*
        MINUTES* DAY* DAYS* WEEK-END* WEEK-ENDS* WEEKEND*
        WEEKENDS* MONTH* MONTHS* MORNING* WEEK* WEEKS*
        YEAR* YEARS* SUMMER* WINTER* SPRING* AUTUMN*
```

```
JANUARY* FEBRUARY* MARCH* APRIL* MAY*
JUNE* JULY* AUGUST* SEPTEMBER* OCTOBER*
NOVEMBER* DECEMBER* SUNDAY* MONDAY* TUESDAY*
WEDNESDAY* THURSDAY* FRIDAY* SATURDAY* DECADE* DECADES*)
```

- **Questions** - Questions are not handled in Yeap's implementation. Three different kinds of questions are identified. They are:
  - Be Questions - (IS? (:OBJ1 ?R++) (:OBJ2 ?R+))

    Example sentences: *"Are they coming?"*, *"Is he a doctor?"*, *"Is John with manager?"*, *"Is there a doctor in the house?"*, etc. An example parse for the sentence *"Are they coming?"*:

    ```
    (ARE? (:OBJ1 (THEY* (:PNOUN)))
          (:OBJ2 (COMING* (:WHAT ?R))))
    ```

  - Auxiliary Questions - (CAN? (:ACTOR ?R++) (:MS1 ?R+)) or (HAVE? (:ACTOR ?R++) (:WHAT ?R+))

    Example sentences: *"Do you smoke?"*, *"Did Peter enjoy the party?"*, *"Can you help me?"*, *"Have you finished?"*, etc. An example parse for the sentence *"Do you smoke?"*:

    ```
    (DO? (:ACTOR (YOU* (:PNOUN)))
         (:MS1 (SMOKE* (:ACTOR ?L))))
    ```

  - Wh Questions - (WHAT? (:MS1 ?R+)) or (WHAT? (:WHAT ?R++) (:MS1 ?R+))

Example sentences: *"Where does she live?"*, *"Where are you?"*, *"What time does the training start?"*, *"How far is it to York?"*, etc. An example parse for the sentence *"Where does she live?"*:

```
(WHERE? (:MS1
          (DOES?
           (:ACTOR (SHE* (:PNOUN)))
           (:MS1 (LIVE* (:ACTOR ?L))))))
```

4.3.3 New rules for new cases

Many different routines were implemented to handle the different labels in UGE. At the moment, 16 such routines have been implemented. These are: fill?aux-qs, fill?be-qs, fill?wh-qs, fill?conn-MS, fill?wh-noun, fill?wh-MS , fill?be-ms, fill?L*MS, fill?conj-MS, fill?L-adv-MS, fill?L+MS, fill?R-MS, fill?infto-MS, fill?L-MS, fill?R+only-MS, process-noun.

In the initial development, the focus was to discover these rules. However, such a process led to duplications and/or rules which later became obsolete or inappropriate. To overcome this problem, we identified 10 top-level labels. These labels are labels that could, in theory, appear adjacent to each other. Then we carried out a systematic investigation into whether they could or could not appear adjacent to each other. These 10 labels are:

- **Empty Stack** - Stack is empty and it is the case of starting point of the sentence

- **Null-set** - Stack is not expecting ?L or ?R and it is the case of noun term on the stack
- **?R+** - Stack is expecting ?R+ to fill its object from right side
- **MS?R+** - Stack is expecting another sentence from right side
- **?L+** - Stack is expecting ?L+ to fill its object or agent from left side
- **MS?L+** - Stack is expecting another sentence from left side
- **?R-** - Stack is waiting to pass its information to its right
- **?L-** - Stack is waiting to pass its information to its left
- **?L*** - Stacks is waiting to attach its information to its left.
- **?L#** - Stacks is waiting to attach its information to its left.

Figure 4.2 shows an example of considering all the above labels against the null set (i.e noun terms, those words without any ?L+ or ?R+). This means when such a word appears, we consider all the possible labels that could appear on the stack. Then, we consider whether that combination is possible and if so how should they be combined. For the former, we search for an appropriate example (note the reverse of the above approach). To find examples, we use online resources, journal articles, grammar books, and newspaper articles. If no sentence is found, we put in a piece of code to trap their occurrence in the future. If a sentence is found, code is put in place to combine them. Note that in the latter case, one may have to deal with various sub-categories. For example, within *?R+* there are *?R++* and *?R+=* to handle the different way of filling objects from its right side.

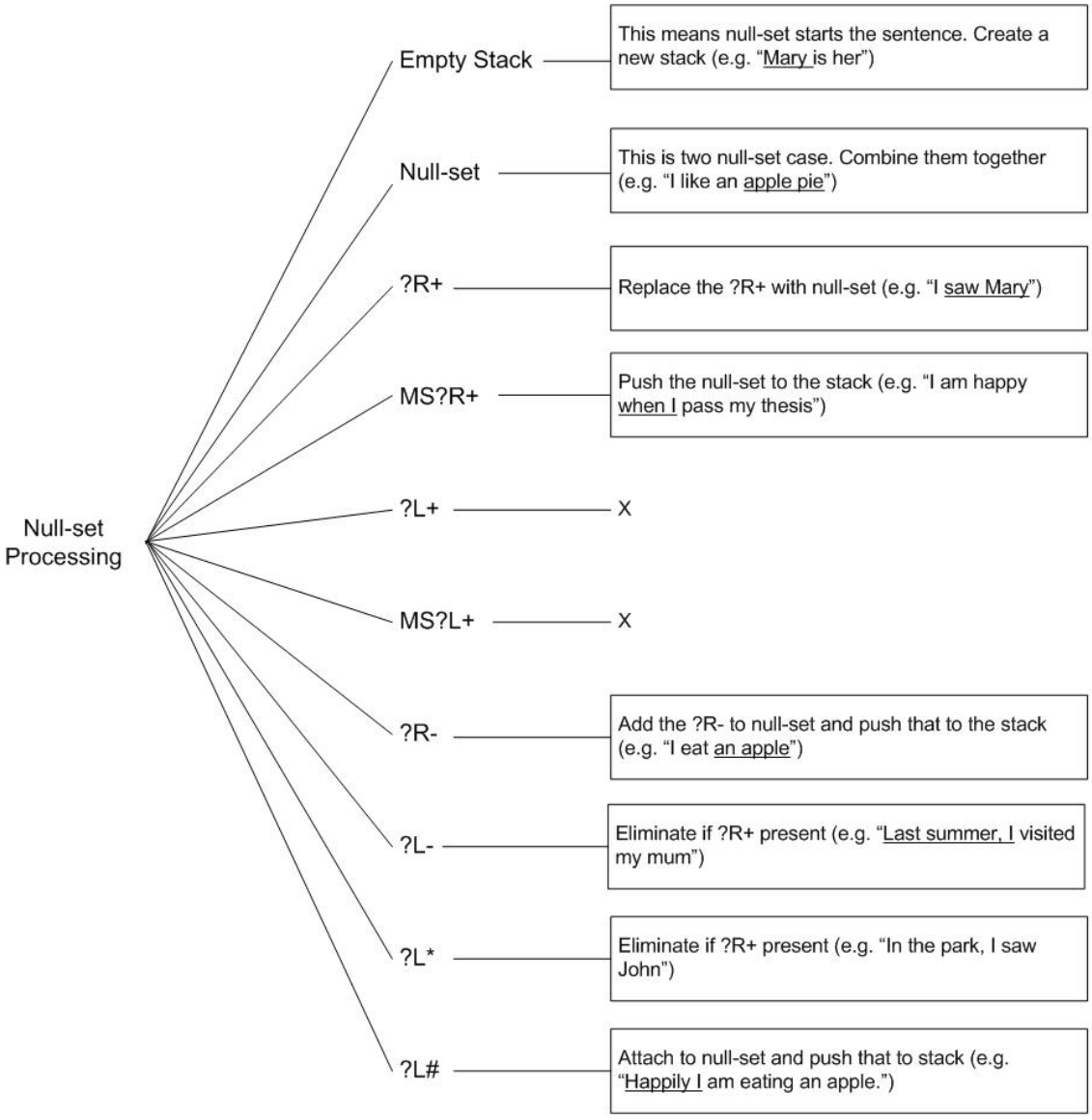

Figure 4.2: Null-set processing - with examples

Figure 4.2 shows that it is not possible to find on the stack an entry with a ?L+ or a MS?L+ when the input is a noun term (these cases are marked with "X"). This process is repeated for each of the routines handling each of the ten top-level labels. Table 4.2 shows the abstract code for the noun-process function (i.e. null-set processing). It takes a stack (stacks) and the lexical entry of the word (noun-LM) as input.

Using this systematic approach, the rules in UGE have been tidied up and several missing cases were found. Furthermore, cases that we have missed could now be trapped by the code if they appear in subsequent use.

4.3.4 Decision making rules

As mentioned earlier, some words function as more than one category. Based on its context, what grammatical category of the word is more sensible can sometimes be determined. Consider these two sentences, *"He has been driving from Auckland to Hamilton"* and *"His driving is good"*. Here, gerund definition of *"driving"* is more sensible in the first case, while noun definition is more sensible in the second case. In the new implementation, there are rules which decide on what is more sensible based on its context or the nature of the sentence. This is called decision making rules, which normally eliminate the stack for non sensible combination.

To eliminate the noun definition of the word *"driving"* in the sentence *"He has been driving from Auckland to Hamilton"*,the following condition (refer Table 4.3)

noun-process(stacks noun-LM):

FOR each stack check
- Case: **Empty Stack**
  - Create a new stack; Create MS from noun-LM
  - Push this MS into the stack;
- Case: **Null-set**
  - Pop the MS from the stack; Combine MS with noun-LM
  - Push the result into the stack;
- Case: **?R+**
  - Pop the MS from the stack; Replace its ?R+ with noun-LM
  - Push the result into the stack;
- Case: **MS?R+**
  - Push the noun-LM into the stack;
- Case: **?L+**
  - Eliminate the stack;
- Case: **MS?L+**
  - Eliminate the stack;
- Case: **?R-**
  - Pop the MS from the stack; Combine ?R- MS with noun-LM
  - Push the result into the stack;
- Case: **?L-**
  - IF stack has no ?R+ or filled ?R+
    - Pop the MS from the stack; Attach the MS to noun-LM
    - Push the result into the stack;
  - ELSE
    - Eliminate the stack;
- Case: **?L***
  - IF stack has no ?R+ or filled ?R+
    - Pop the MS from the stack; Attach the MS to noun-LM
    - Push the result into the stack;
  - ELSE
    - Eliminate the stack;
- Case: **?L#**
  - Pop the MS from the stack; Attach the MS to noun-LM
  - Push the result into the stack;
- Case: **Else**
  - Eliminate the stack;

Table 4.2: New Implementation Model - Noun process

is added to the noun-process. Here, the process checks the dictionary for gerund definition of the particular word before eliminating the stack, which only enables the system to eliminate the words which has both noun and gerund definition. The stack created for the phrase *"He has been"* is (here IN-PEEK-MS denotes the stack):

```
IN-PEEK-MS
[BEEN* (:WHAT ?R+)]

[HAS* (:ACTOR (HE* (:PNOUN)))
      (:WHAT ?R+)]
```

Similarly, the decision making rules are added to each process to eliminate the stacks. However, before adding these rules to eliminate the stack, lots of sample sentences need to be examined. This prevents accidently eliminating sensible combinations.

4.4 PostUGE

There are multiple MSs generated by the UGE module, due to the multiple meaning of words in the dictionary. This is one of the major problems in natural language parsing. To use UGE in natural language application, one must select the best parse tree for the sentence from the source of information available. Here, the only information available is the sentence and the MSs returned for the sentence. The PostUGE module computes the best MS from the MSs returned

noun-process(stacks noun-LM):

Step1: For each stack
.
.
Case: **?R+**
.
.
Case: the MS word is BEEN and
the dictionary has gerund definition for LM
Eliminate the stack; Go to Step1
.
.
.
.

Table 4.3: Noun process - with sample decision making rule

by UGE module. This is achieved by adding rules to the system.

The abstract algorithm is shown in Table 4.4. Select-best-MS function takes MSs (list of MS) returned by UGE module as an argument and returns the MS as a best parse tree for the sentence. Here, each process (Process-Rule1 or Process-Rule2, etc) is a complex process and works based on priority rules. For example, in the Process-Rule1 (refer Table 4.5), there is a rule which gives more priority to the verbs which are speech words, such as SAID* SAY* SAYS* TELL* TELLS* TOLD* etc. Then, there is a rule which gives more priority to the verbs which has MS1 on them. When a new rule is added to the system, it verifies its correctness by running the test module mentioned in Section 4.5. Here, each rule based process is not explained in detail, since it is far too detail to mention. The complete program of Process-Rule1 is given in appendix A.5.

Select-best-MS(MSs):

Case: MSs is empty
    Return NIL
Case: Length of MSs is 1 (only one MS is returned by UGE)
    Returned first member of MSs
Case: Else
        Case: Rule1 - All of the MSs are verb
            Call Process-Rule1
        Case: Rule2 - All of the MSs are be-verb
            Call Process-Rule2
        Case: Rule3 - All of the MSs are conjunction
            Call Process-Rule3
        Case: Rule4 - All of the MSs are conjunction with one clause
            Call Process-Rule4
        Case: Rule1&2 - All of the MSs are either verb or be-verb
            Call Process-Rule1&2
        .
        .
        Case: Else
            Return first member of MSs

Table 4.4: PostUGE - Select best MS

Process-Rule1(MSs):

Initialise list tmpresult to NIL
Remove all the MSs which do not have *speech-words
and Set this list tmpresult list
    When length tmpresult is 1
        Return first member of tmpresult
        as output of Process-Rule1
        Exit Process-Rule1
    When length tmpresult is less than length of MSs
        Set MSs to tmpresult list
        Continue..
Remove all the MSs which do not have MS1 on them
and Set this list tmpresult list
    When length tmpresult is 1
        Return first member of tmpresult
        as output of Process-Rule1
        Exit Process-Rule1
    When length tmpresult is less than length of MSs
        Set MSs to tmpresult list
        Continue..
.
.
.
.

Table 4.5: PostUGE - Process-Rule1

4.5 Testing mechanism

As mentioned in Chapter 3 (Section 3.4), the systematic testing mechanism is important to verify the changes to UGE are correct. This testing module must validate that newly established labeling scheme and the newly added rules are correct. This module has test data set and three different tests. The current test data has 843 entries. The sample entry in the test data set has three components. They are: 1) Sentence 2) Number of MSs returned by UGE at the time of adding the test entry 3) The best MS for the sentence which is selected manually. The sentences in the test data set is appended in appendix A.4. The sample test entry for the sentence *"I want to eat the apple quickly"* is shown below:

```
("I want to eat the apple quickly."
 2
 (WANT* (:ACTOR (I* (:PNOUN)))
        (:TO** (EAT*
                 (:WHAT (APPLE* (:NOUN)
                                (:MODIFIER (THE*)))))
                (:MANNER (QUICKLY*))))
 )
```

Test module is invoked by the function test-ugetest(). This function goes through the test data set and parses every sentence in the test entry, which is the first component of the test entry, using the current UGE (after the new rule is added to UGE). It returns **Pass** or **Fail** for each three tests based on MSs returned by current UGE. The three tests are:

Test1 This test checks that the current UGE is parsing all the sentences in the data set correctly. That means the added rule does not affect any of the correct parses in the test data set. This is considered as a very important test. If it fails, then the added rules need to be modified, or at worst, need to be removed.

Test1 - Pass: If the manually selected MS, which is the third component in test entry, is presented among the MSs returned by the current UGE, then the particular data entry is passed for **Test1**.

Test1 - Fail: If the above is not true, then the particular data entry is failed for **Test1**.

Test2 This test deals with number of MSs returned by the current UGE. It checks the current MSs with the previously returned MSs, which is the second component of the test entry, and reports the fail or pass based on it. Even though, this test is not as significant as the first one, it reveals how tuned UGE is. For example, if this test passes, then it means that the added rule either reduces or equates the MSs returned by UGE.

Test2 - Pass: If the number of MSs returned by current UGE is less than or equal to the number of MSs previously returned.

Test2 - Fail: If the number of MSs returned by current UGE is greater than the number of MSs previously returned.

Test3 This test is for the **PostUGE** module. This makes sure the selection rule added to **PostUGE** module is correct. Every time a new rule is added to **PostUGE**, the test is run to make sure the added rule does not spoil

the existing data set. This test examines the result returned by current **PostUGE** module with the manually selected MS for the sentence, which is the third component of the test entry. If both of these do not match then it reports **Test3 - Fail**.

Test module helps to identify incorrect rules in the system. For example, the new rules or decision making rules are added based on examining sample sentences. If that is the case, what happens if some specific case is missed out in the sample sentences? To avoid this, whenever the rule is added to handle the new case, UGE is tested with the test data set and the new case is also added to the test data set. If any of the tests fail, the specific case has to be carefully examined by collecting more examples.

4.6 An Example: Parsing a complex sentence using UGE

Parsing a complex sentence using UGE is shown in this section. Here, a sentence from New Zealand Herald newspaper is used as a sample sentence. The sentence is:

> "If Ahmed Zaoui's family succeed in their application to join him in New Zealand as refugees, the decision will annoy NZ First leader Winston Peters but delight a boy who has not seen his father for nearly three years."

This sentence has 39 words and nested clauses. Each word has either one or more lexical entries in the dictionary to denote the different definition of grammatical

usage of that word (for example the word *"driving"* has gerund and noun definition in the dictionary). In order to get the MS for the sentence, first UGE preprocess the sentence and then produce MSs by attaching ?L/?R attachment rules in which multiple stacks are created to handle multiple definition of words. Recall that the Pre-processor group or ungroup the words/word into meaningful chunks. The pre-processor output for the above sentence is:

> "If Ahmed-Zaoui 's family succeed in their application to join him in New-Zealand as refugees, the decision will annoy NZ-First leader Winston-Peters but delight a boy who has not seen his father for nearly three years."

The parsing of the above sentence is shown below. For each word parsed, the lexical entries in the dictionary for that word is first shown and then the number of stacks created is shown (normally only the first stack is shown in detail). An explanation of the output is then given.

```
Input: "If"
Dictionary entry: ((IF* (:MS1 ?R++) (:MS2 ?R+))
                   (IF* (:MS2 ?R+) (:MS1 ?L+)))

Number of stacks: 1

IN-PEEK-MS
[IF* (:MS1 ?R++)
(:MS2 ?R+)]
```

A stack is created using the entry (IF* (:MS1 ?R++) (:MS2 ?R+)).
This entry looks for parts of the conjunction if on its right
hand side. The other entry expects a sentence already on the stack.
No such sentence is found and hence this entry is eliminated.

```
Input: "Ahmed Zaoui"
Dictionary entry: (AHMED-ZAOUI* (:UNKNOWN)
                                (:NAME))

Number of stacks: 2

IN-PEEK-MS
[AHMED-ZAOUI* (:UNKNOWN)
              (:NAME)]

[IF* (:MS1 ?R++)
     (:MS2 ?R+)]

.
.
```

The name "AHMED-ZAOUI" is stacked on the stack, since (:MS1 ?R++)
or (:MS2 ?R+) only allow complete sentences to be attached to its
?R++ or ?R+.

```
Input: "'s"
Dictionary entry: ((?R- (:MODIFIER (|AHMED-ZAOUI'S*|)))
                   (IS-HAS? (:OBJ1 ?R++) (:OBJ2 ?R+))
                   (?L* (:IS-HAS* ?R+))
                   (HAS-IS* (:ACTOR ?L+) (:WHAT ?R+)))

Number of stacks: 3

IN-PEEK-MS
[?R- (:MODIFIER (|AHMED-ZAOUI'S*|))]

[IF* (:MS1 ?R++)
     (:MS2 ?R+)]

.
.
The first stack shows the use of the possessive definition
of the noun.


Input: "family"
Dictionary entry: ((FAMILY* (:NOUN)))
```

```
Number of stacks: 8


IN-PEEK-MS
[FAMILY* (:NOUN)
         (:MODIFIER (|AHMED-ZAOUI'S*|))]


[IF* (:MS1 ?R++)
     (:MS2 ?R+)]


.
.
Here, the "?R-" in "AHMED-ZAOUI'S" is attached to the
noun term "FAMILY". The result is pushed back on to the stack.



Input: "succeed"
Dictionary entry: ((SUCCEED* (:ACTOR ?L+)))


Number of stacks: 1


IN-PEEK-MS
[SUCCEED* (:ACTOR (FAMILY* (:NOUN)
                           (:MODIFIER
                            (|AHMED-ZAOUI'S*|))))]
```

```
[IF* (:MS1 ?R++)
     (:MS2 ?R+)]


The "?L+" of "SUCCEED" is filled by the noun entity
on the stack. Note, the number of stacks created is reduced
to 1.


Input: "in"
Dictionary entry: ((?L* (:IN* ?R+=))
                   (|?L#| (:MANNER (IN*))))

Number of stacks: 2

IN-PEEK-MS
[SUCCEED* (:ACTOR (FAMILY* (:NOUN)
                           (:MODIFIER
                            (|AHMED-ZAOUI'S*|))))
          (:IN* ?R+=)]

[IF* (:MS1 ?R++)
     (:MS2 ?R+)]

.
```

.

Here, "IN" is attached to the non-collapsed stack.

Input: "their"

Dictionary entry: ((?R-* (:MODIFIER (THEIR*))))

Number of stacks: 1

IN-PEEK-MS

[?R-* (:MODIFIER (THEIR*))]

[SUCCEED* (:ACTOR (FAMILY* (:NOUN)
(:MODIFIER
(|AHMED-ZAOUI'S*|))))
(:IN* ?R+=)]

[IF* (:MS1 ?R++)
(:MS2 ?R+)]

The "?R-" in "THEIR" is put on top of the stack, waiting for something from its right to attach itself to complete

```
the meaning.



Input: "application"
Dictionary entry: ((APPLICATION* (:NOUN)))

Number of stacks: 3

IN-PEEK-MS
[APPLICATION* (:NOUN)
              (:MODIFIER (THEIR*))]

[SUCCEED* (:ACTOR (FAMILY* (:NOUN)
                           (:MODIFIER
                            (|AHMED-ZAOUI'S*|))))
          (:IN* ?R+=)]

[IF* (:MS1 ?R++)
     (:MS2 ?R+)]

.
.


When the word "APPLICATION" appears, the "?R-" term on
```

```
the stack attaches itself to the null-set "APPLICATION"
(i.e. noun term).



Input: "to"
Dictionary entry: ((:TO** ?R+)
                   (?L* (:TO* ?R+=)))


Number of stacks: 3


IN-PEEK-MS
[:TO** ?R+]


[SUCCEED* (:ACTOR (FAMILY* (:NOUN)
                           (:MODIFIER
                            (|AHMED-ZAOUI'S*|))))
          (:IN* (APPLICATION* (:NOUN)
                    (:MODIFIER (THEIR*))))]


[IF* (:MS1 ?R++)
     (:MS2 ?R+)]


.
.
```

```
Infinitive-to is put on the stack and waiting for
something from its right.


Input: "join"
Dictionary entry: ((JOIN* (:ACTOR ?L+) (:WHAT ?R+))
                   (JOIN* (:ACTOR ?L+))
                   (JOIN* (:NOUN)))

Number of stacks: 4

IN-PEEK-MS
[:TO** (JOIN* (:WHAT ?R+))]

[SUCCEED* (:ACTOR (FAMILY* (:NOUN)
                           (:MODIFIER
                            (|AHMED-ZAOUI'S*|))))
          (:IN* (APPLICATION* (:NOUN)
                              (:MODIFIER (THEIR*))))]

[IF* (:MS1 ?R++)
     (:MS2 ?R+)]

.
```

.

```
Now, "JOIN" from right side attaches to "?R+" of infinitive-to,
now the stack is waiting to fill the ?R+ of "JOIN"


Input: "him"
Dictionary entry: ((HIM* (:PNOUN)))

Number of stacks: 5

IN-PEEK-MS
[IF* (:MS1
      (SUCCEED* (:ACTOR (FAMILY* (:NOUN)
                                 (:MODIFIER
                                  (|AHMED-ZAOUI'S*|))))
                (:IN* (APPLICATION* (:NOUN)
                                    (:MODIFIER
                                     (THEIR*))))
                (:TO** (JOIN*
                        (:WHAT
                         (HIM* (:PNOUN)))))))
     (:MS2 ?R+)]
```

```
.
.
Once "HIM" appears from rights, it fills the "?R+" of "JOIN".
Then, the process forces the stack to be collapsed.


.....
Here, we are not showing the processing of other words.
Below, the processing is shown for the last word
in the sentence which is "YEARS".


Input: "years"
Dictionary entry: ((YEARS* (:NOUN)))

Number of stacks: 1777

IN-PEEK-MS
[YEARS* (:NOUN)
        (:MODIFIER (THREE*))]

[FATHER* (:NOUN)
         (:MODIFIER (HIS*))
         (:FOR* ?R+=)
         (:MANNER (NEARLY*))]
```

```
[SEEN* (:WHAT ?R+)]

[HAS* (:ACTOR ?L)
      (:WHAT ?R+)
      (:MANNER (NOT*))]

[BOY* (:NOUN)
      (:MODIFIER (A*))
      (:WHO* (:MS1 ?R+))]

[DELIGHT* (:NOUN)]

[?L* (BUT* ?R+)]

[WINSTON-PETERS* (:UNKNOWN)
                 (:NAME)]

[LEADER* (:NOUN)
         (:X-WORDS (*NZ*-FIRST*)
                   (LEADER*))
         (:NAME)]

[ANNOY* (:ACTOR ?L)
        (:WHAT ?R+)]
```

```
[WILL* (:ACTOR (DECISION*
                (:NOUN)
                (:MODIFIER (THE*))))
       (:MS1 ?R+)]

[IF* (:MS1
      (SUCCEED* (:ACTOR
                 (FAMILY*
                  (:NOUN)
                  (:MODIFIER
                   (|AHMED-ZAOUI'S*|))))
                (:IN* (APPLICATION*
                       (:NOUN)
                       (:MODIFIER
                        (THEIR*))))
                (:TO**
                 (JOIN*
                  (:WHAT
                   (HIM*
                    (:PNOUN)
                    (:IN* (NEW-ZEALAND*
                           (:UNKNOWN)
                           (:NAME)
                           (:AS*
```

```
                                (REFUGEES*
                                 (:NOUN)))))))))))
      (:MS2 ?R+)
      (:COMMA-PRESENT)]


UGE final interpretations:


Number of MSs returned: 2
The output selected by PostUGE module:


(IF* (:MS1
       (SUCCEED* (:ACTOR (FAMILY*
                          (:NOUN)
                          (:MODIFIER
                           (|AHMED-ZAOUI'S*|)
                           )))
                 (:IN* (APPLICATION*
                        (:NOUN)
                        (:MODIFIER
                         (THEIR*))))
                 (:TO** (JOIN*
                         (:WHAT
                          (HIM*
                           (:PNOUN)
```

```
                          (:IN*
                           (NEW-ZEALAND*
                            (:UNKNOWN)
                            (:NAME)
                            (:AS*
                             (REFUGEES*
                              (:NOUN))
                            )))))))))
(:MS2
 (BUT*
  (:MS2
   (DELIGHT* (:ACTOR ?L)
    (:WHAT
     (BOY* (:NOUN)
           (:MODIFIER
            (A*))
      (:WHO*
       (:MS1
        (HAS* (:ACTOR ?L)
         (:WHAT
          (SEEN*
           (:WHAT
            (FATHER*
             (:NOUN)
             (:MODIFIER
```

```
                (HIS*))
               (:FOR*
                (YEARS*
                 (:NOUN)
                 (:MODIFIER
                  (THREE*))))
               (:MANNER
                (NEARLY*))
               ))))
                (:MANNER
                 (NOT*)
                 )
                )
      ))))))
(:MS1
 (WILL* (:ACTOR
         (DECISION*
          (:NOUN)
          (:MODIFIER
           (THE*))))
  (:MS1
   (ANNOY* (:ACTOR ?L)
           (:WHAT
            (WINSTON-PETERS*
             (:X-WORDS
```

```
    (*NZ*-FIRST*)
    (LEADER*)
    (WINSTON-PETERS*)
    )
   (:NAME)
   )
  )
)))))))
```

```
Even though the process creates 1777 stacks,
the final-interpret function only returns 2 outputs
as a complete MSs. This is due to the fact that other outputs
or stacks created are not complete MSs (i.e. it might be a
partial interpretation). The output shown above is the output
return by PostUGE module.
```

4.7 Conclusion

The new implementation is developed based on Yeap's implementation. The aim of this research is to show that Yeap's theory can be implemented to parse real world complex sentences. To do that, there are three new modules added to the system, which are 1) Pre-processor 2) PostUGE 3) Dictionary. Apart from that, Yeap's implementation, which is UGE, is extended to handle more complex sentences. This is achieved by 1) adding missing lexical entries to the dictionary 2)

identifying missing labeling scheme for new cases and establishing labeling scheme using Yeap's theory 3) Adding new rules to handle new cases 4) Adding decision making rules to the system to eliminate non sensible combinations. Moreover, the system is equipped with a test module, which provides a systematic mechanism for testing the correctness of any modification to the program.

Next section provides the experiments which evaluate the performance of the new UGE. These experiments are designed to test how well UGE can parse complex real world sentences found in newspaper articles. Here the New Zealand Herald newspaper is used as a source.

Chapter 5

Experimental Evaluation

This chapter describes the experiments done using UGE and analyses the results obtained. Two different evaluation methods were used. They are:

1. **Self evaluation** - This test directly evaluates the performance of UGE itself. The goal is to measure the percentage of the real world sentences UGE can parse, the accuracy of the results and the cause of failures.

2. **Comparative evaluation** - This test evaluates UGE against other parsers. The goal of this evaluation is to test how well UGE performs comparatively.

Section 5.1 presents the evaluation matrix and scheme used to measure UGE's performance. Section 5.2 describes the experiments conducted for the self evaluation method and its results. Section 5.3 outlines the experiments done for comparative evaluation methods and the results obtained. Section 5.4 concludes this chapter with experimental findings.

5.1 Evaluation matrix and scheme

Parser performance should be measured in terms of accuracy, speed and number of parse options it returns (Daniel Cer & Manning, 2010). This is because, the purpose of a parser is to solve natural language processing tasks such as information extraction and machine translation, in which speed and accuracy are the key elements. Even though the purpose of this research, first and the most, is to implement the theory of language to parse complex real world sentence, the accuracy of the parse is still important. Parsing is a complicated task due to the multiple meanings of words. The number of parse options increase exponentially with the length of the sentence. When the number of parse options are high, it is difficult to select the correct parse option as intended by the writer.

Therefore, the parser evaluation matrix should include speed, accuracy and the number of parse options returned for a particular sentence. Speed and parse options are however, easy to obtain. The former is determined by computing the time taken per parse while the later is calculated by counting the number of parse options returned for each sentence. However, determining the accuracy of a parser is much more difficult since it requires deep linguistic knowledge of the parser output representation and the grammatical information it possesses.

Many researchers have used Treebanks as a gold standard to measure the parser accuracy for many decades (Collins, 1999; Marcus, Marcinkiewicz, & Santorini, 1993). Treebank is a collection of sentences which are annotated manually with the correct parse tree (Collins, 1999; Marcus et al., 1993). These are built either

using constituency based annotation scheme or dependency based annotation scheme (Marcus et al., 1993; Sampson, 1993). Parser output representation must be similar enough to Treebank's annotation scheme in order to use the specific Treebank for evaluation (Kbler, 2005). Since UGE's output is different, it is not possible to use Treebank for its evaluation. Therefore, an alternate evaluation scheme, which is independent of parser output format, must be used here.

Grammatical relations (GRs) evaluation scheme proposed by Carroll, Briscoe, and Sanfilippo (1998) is used as a base for this research to measure the parser accuracy. GRs scheme is independent of the parser output format which is well suited for this research and its two evaluation methods. Carroll et al. (1998) proposed twenty different hierarchical parser independent GRs. The arguments for each GRs are head, dependent, type and initial grammatical relations. Figure 5.1 shows graphically the grammatical relation hierarchy and twelve GRs from this hierarchy with its description are outlined in Table 5.1.

Grammatical Relation	Description
ncmod(type,head,dependent) ncmod(hand,eat,with)	Non-clausal modifier sentence: "John ate an apple with hand"
xmod(type,head,dependent) xmod(without,eat,ask)	Clausal modifier controlled from without sentence: "John ate without asking"
cmod(type,head,dependent) cmod(because,eat,is)	Clausal modifier controlled from within sentence: "John ate because he was hungry"
arg_mod(type,head,dependent,initial_gr) arg_mod(by,kill,Mary,subj)	Modifier analysed as a bound adjunct sentence: "John is killed by Mary"
ncsubj(head,dependent,initial_gr) ncsubj(eat,John,_)	Non-clausal subject sentence: "John ate an apple"
xsubj(head,dependent,initial_gr) xsubj(win,require,_)	Clausal subject controlled from without sentence: "to win world cup requires heaps of practice"
csubj(head,dependent,initial_gr) csubj(leave,mean_)	Clausal subject controlled from within sentence: "John left without saying good-bye meant he was angry"
dobj(head,dependent,initial_gr) dobj(read,book,_)	Direct object sentence: "John read books"
iobj(type,head,dependent) iobj(to,give,poor)	Object introduced by preposition sentence: "John gives to the poor"
obj2(head,dependent) obj2(give,book)	Second ditransitive object sentence: "John give Mary a book"
xcomp(type,head,dependent) xcomp(to,intend,leave)	Clausal complement without overt subject sentence: "John intends to leave the job"
ccomp(type,head,dependent) ccomp(that,say,leave)	Clausal complement with overt subject sentence: "John said that Mary left"

Table 5.1: GRs used in this experiment

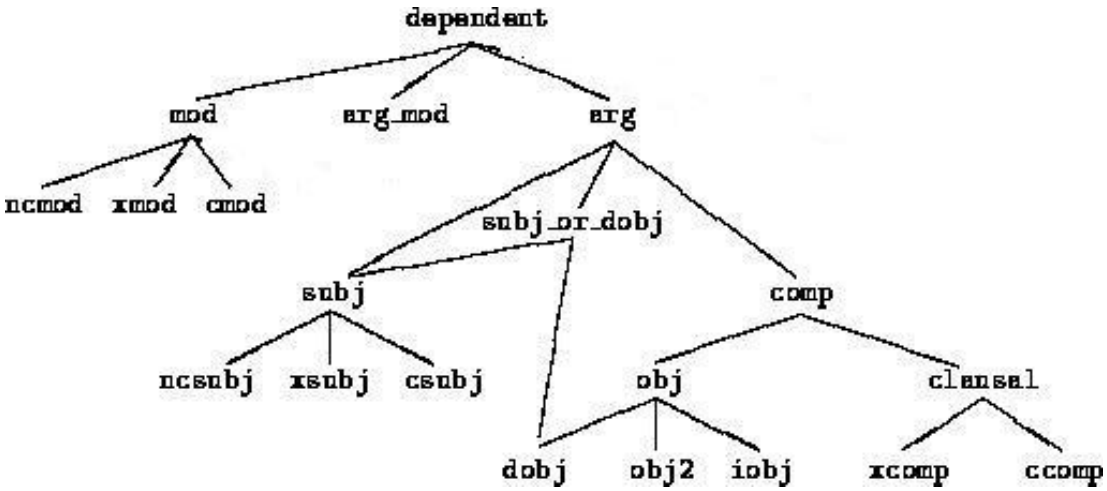

Figure 5.1: The grammatical relation hierarchy - As proposed by Carroll et al. (1998).

For example, the sentence *"John said that he ate apples and oranges with hand without asking"* has the following GRs:

ncsubj(say,john,_),
ncsubj(eat,he,_),
dobj(eat,apples,_),
dobj(eat,oranges,_),
ncmod(hand,eat,with),
xmod(without,eat,ask),
ccomp(that,say,eat)

To measure the parser accuracy, the parser output needs to be transformed into GRs (i.e. GRs needs to be extracted for the parser output). Then the extracted GRs from the parser output are compared with manually annotated GRs for a particular sentence. The accuracy is calculated as a percentage of the number of matching extracted GRs from parser output with the annotated GRs divided by the total annotated GRs for a particular sentence.

However, one problem with this approach is that it is not straightforward to compare GRs of parser output with annotated GRs, since different parsers attach arguments in different ways. For example, UGE attaches the preposition to the nearest noun and never to the verb. The following example illustrates this difference. Consider the sentence *I eat an apple in the park*:

The manually annotated GRS for this sentence:

ncsubj(eat,I,_)
dobj(eat,apple,_)
ncmod(in,eat,park)

UGE output and extracted GRs from UGE output are:

```
(EAT* (:ACTOR (I* (:PNOUN)))
      (:WHAT (APPLE* (:NOUN) (:MODIFIER (AN*))
                     (:IN* (PARK* (:NOUN) (:MODIFIER (THE*)))))))
```

ncsubj(eat,I,_)
dobj(eat,apple,_)
ncmod(in,apple,park)

Here, the type and dependent of ncmod is same with manually annotated GRs, however head is different. To handle these differences in annotated GRs and parser output, the accuracy is measured in two ways with and without **MOD/IOBJ** annotation, which are *ncmod, xmod, cmod and iobj*. For each way, the accuracy is measured. For example, for the sentence *I eat an apple in the park* the overall accuracy (which is with **MOD/IOBJ** annotation) is calculated at 66% (2/3 * 100) and accuracy without **MOD/IOBJ** is calculated at 100% (2/2 * 100).

5.2 Self evaluation

Self evaluation is done using three experiments. The intention here is to measure the UGE performance for the evaluation matrix such as speed, accuracy and number of parser options returned.

First experiment is a coverage test which measures the percentage the sentences that UGE assigns one or more parses, average time taken per parse and the number of parses UGE produces per sentence. Second experiment determines the accuracy of parses using the evaluation scheme proposed in Section 5.1. Third experiment examines the cause of failed parses.

The data set used for this evaluation comes from the New Zealand Herald newspaper. Appendix A.6 shows one of the articles used in this evaluation. To cover a wide range of vocabulary, articles from different domains such as business, sports and technology are selected.

5.2.1 Experiment 1

For this experiment, one hundred articles from the data set are parsed automatically using UGE. Total number of sentences in the articles (Total), the number of sentences which UGE returns parse options for (Parsed), total time taken for parsing (Time) and average number of parses (Avrg-MSs - total number of parses divided by parsed sentences) are recorded.

To strengthen this coverage test to avoid false positives, UGE only returns output

if the parse option has sentence constructor such as verb, conjunction or be-verb (Please refer Chapter 4 for more detail). Also UGE returns nothing if the parsing is broken in the middle; in other words, if it is a partial parse, then UGE returns nothing.

Table 5.2 shows the results. Here the success rate (success rate%) is calculated as the percentage of the number of parsed sentences divided by the total number of sentences. This result reveals that UGE returns parse option for 81% of the sentences and fails for 19% of sentences. Also from this result the average time taken per sentence can be calculated to be 0.456 seconds (Total time taken divided by Total sentences). Average parse options returned per sentence is 3, which is calculated as total number of parse options returned divided by number of parsed sentences. Average word per sentence is also calculated as total words in the parsed sentences divided by the parsed sentences.

Total sentences	1991
Parsed sentences	1612
Success rate%	81%
Total time taken	926 sec
Average parse options per sentence	3
Average words per sentence	26

Table 5.2: Self evaluation - Experiment 1 results

Here, the number of parse options returned is extremely low. The sentences parsed are quite complex because they are quite lengthy. The performance, both in terms of successful parses and time taken are acceptable. Note that, no optimisation of the code is done.

5.2.2 Experiment 2

In experiment 1, there is no guarantee that what the parser returns is correct, unless one performs manual checking. The goal of this experiment is to examine the accuracy of UGE parses. One hundred parsed sentences from experiment 1 are used as the data set. Each of these sentences is manually annotated for GRs and GRs from UGE results are extracted too. As mentioned in Section 5.1, UGE accuracy is measured with and without **MOD/IOBJ** annotation. Table 5.3 shows the results of this experiment.

Overall Accuracy	89%
Accuracy without **MOD/IOBJ**	98%

Table 5.3: Self evaluation - Experiment 2 results

The results in Table 5.3 reveal that accuracy of UGE parses are 89% with **MOD/IOBJ** annotation and 98% without **MOD/IOBJ** annotation. The difference between the two lies in the way in which UGE handles prepositional attachment and this difference causes a drop of 11% in accuracy in the **MOD/IOBJ** annotation. UGE looses 2% accuracy even in the without **MOD/IOBJ** annotation, which is caused by the incorrect parse selection by UGE's select-best algorithm. For example, if UGE returns more than one analysis for a particular sentence, the select-best algorithm selects the best parse (see Chapter 4 for selection algorithm) which is used for this evaluation method. This selected parse is not always the best parse and this cause the 2% accuracy loss in the without **MOD/IOBJ** annotation.

5.2.3 Experiment 3

The goal of this experiment is to examine the cause of the failed parses. From experiment 1, UGE fails to produce output for 19% of the sentences. The cause of the failure has been found to be due to four reasons, which are:

NEW CASE - this specific case is not handled in UGE

UNKNOWN WORDS - the dictionary does not have all the definitions of the word

NON-SENTENCE - the specific sentence is not a valid sentence for a given grammar

OTHER CASE - Special symbols or special characters or quotation marks within the sentence causes the failure OR other unknown causes trigger the abnormal termination of the program.

Here, one hundred failed sentences from experiment 1 are analysed to determine the cause of failure. Table 5.4 shows the results of this experiment.

Type of failure	**Percentage %**
NEW CASE	28%
UNKNOWN WORDS	38%
NON-SENTENCE	16%
OTHER CASE	18%

Table 5.4: Self evaluation - Experiment 3 results

Experiment 1 shows 19% of the sentences fail to parse by UGE. Within this 19%, only 28% of the sentences are new cases which need to be implemented.

For example, consider the sentence, *"Another relative, who was allowed to visit the scene to feed hungry stock, said family members knew little about what had happened, and what they did know they were struggling to comprehend."*. This is an example of a new case which is not handled by current UGE. Two different clauses are connected with connective *"and"*. First clause is *"what had happened"*, which is a question and second clause *"what they did know they were struggling to comprehend"*, which is a wh-noun clause. The current UGE only allows the connective to connect the clauses if both are of the same type. For example, two nouns can be connected using connective, but not noun and a verb term.

38% of the failed sentences failed due to missing word meaning. Here the missing word meaning can be fixed by adding the correct definition of the word into the dictionary. For example, the sentence *"In one of the most remarkable signs yet of the advance of global warming, Britain's first olive grove has just been planted in Devon."* failed due to the missing past participle definition of the word ***"planted"*** in the dictionary. The following analysis shows the UGE output before and after adding the correct definition for the word *"planted"*:

The dictionary entry for the word *planted*:

```
(get-word "planted")

((PLANTED* (:ACTOR ?L+) (:WHAT ?R+)))
```

UGE output:

```
(UGE* "In one of the most remarkable signs yet of the
```

```
advance of global warming, Britain's first olive grove
has just been planted in Devon")

Number of MSs: 0
NIL
```

After adding past participle of *"planted"* to the dictionary:

```
(GET-WORD "planted")

((PLANTED* (:ACTOR ?L+) (:WHAT ?R+))
 (PLANTED* (:WHAT ?R+)))
```

UGE output:

```
(UGE* "In one of the most remarkable signs yet of the
advance of global warming, Britain's first olive grove
has just been planted in Devon")

Number of MSs: 14
(.
 .
 (HAS* (:ACTOR
        (GROVE* (:NOUN)
                (:MODIFIER (|BRITAIN'S*|)
                           (FIRST*)
                           (OLIVE*))
```

```
   (:IN*
    (ONE* (:NOUN)
          (:OF* (SIGNS* (:NOUN)
                        (:MODIFIER
                         (THE*)
                         (MOST*)
                         (REMARKABLE*))))
          (:MANNER (YET*))
          (:OF* (ADVANCE* (:NOUN)
                          (:MODIFIER
                           (THE*))
                          (:OF*
                           (WARMING*
                            (:NOUN)
                            (:MODIFIER
                             (GLOBAL*)))))))))
(:WHAT (BEEN*
        (:WHAT (PLANTED* (:WHAT ?R)
                         (:IN*
                          (DEVON*
                           (:UNKNOWN)
                           (:NAME)))))))
(:MANNER (JUST*)))
```

.

.

```
)
```

16% of the 19% failed parses are due to grammatically incorrect sentences in which failure is expected. One of the non-sentence examples is *"* Alpine roads likely to be closed."* This sentence is more like a phrase rather than a complete sentence. If the sentence is rephrased by adding be-verb, *"* Alpine roads **are** likely to be closed"*, then the sentence can be parsed by UGE.

UGE output:

```
(uge* "* Alpine roads are likely to be closed.")

Number of MSs: 1

((ROADS* (:NOUN)
         (:MODIFIER (ALPINE*))
         (:ARE*
          (:TO** (BE*
                  (:WHAT (CLOSED* (:WHAT ?R)))))
          (:MANNER (LIKELY*))))

)
```

Other cases make up the remaining 18% of failed parses. These are due to special symbols in an unusual place which cause the sentence to fail. For example, in the sentence *"The statement said 36 "suspected anti-Iraqi forces" had been detained*

after the operation, and that two of the detainees had admitted to being al Qaeda members" the phrase *"suspected anti-Iraqi forces"* is within the double quotation mark. UGE tries to parse this phrase separately and because of that the above sentence fails. If we remove the double quotation marks from the sentence, UGE can then parse this sentence. To avoid this problem, a new algorithm needs to be added to the pre-processor to identify unusual special symbols within the sentence and remove them. The UGE output after removing the special symbol is:

```
(uge* "The statement said 36 suspected anti-Iraqi forces
had been detained after the operation,
and that two of the detainees had admitted
to being al Qaeda members.")

Number of MSs: 8

(
 .
 (SAID* (:ACTOR (STATEMENT* (:NOUN) (:MODIFIER (THE*))))
        (:MS1
         (AND*
          (:MS2
           (HAD* (:ACTOR (TWO* (:NOUN) (:MODIFIER (THAT*))
                               (:OF*
                                (DETAINEES*
                                 (:NOUN)
```

```
                                (:MODIFIER (THE*))))))
          (:WHAT
           (ADMITTED* (:WHAT ?R)
                      (:TO* (BEING*
                             (:WHAT (MEMBERS* (:NOUN)
                                              (:X-WORDS
                                               (AL*)
                                               (QAEDA*)
                                               (MEMBERS*)))))))))
        (:MS1
         (HAD* (:ACTOR (FORCES* (:NOUN)
                                (:X-WORDS
                                 (36*)
                                 (SUSPECTED*)
                                 (ANTI-IRAQI*)
                                 (FORCES*))))
               (:WHAT (BEEN* (:WHAT
                              (DETAINED* (:WHAT ?R)
                                         (:AFTER*
                                          (OPERATION*
                                           (:NOUN)
                                           (:MODIFIER
                                            (THE*)))))))))))
.
)
```

5.2.4 Summary

Self evaluation experiments reveal that 81% of the sentences (from experiment 1) can be parsed by UGE with 98% accuracy (from experiment 2). The average parses per sentence is 3 and the average parse time per sentence is 0.456 seconds. The sentences used for this evaluation had 26 words on average. This reveals the sentences used for this evaluation are complex ones.

Self evaluation results showed 19% failure rate. Within this only 28% of them are new cases and 38% were due to missing words' definition. The rest of the failures are due to ungrammatical sentences or other special characters or symbols in the sentences.

5.3 Comparative evaluation

The purpose of this evaluation is to compare how well UGE performs against the current parsers available. Two parsers were selected, *Stanford parser* and *Link Grammar (LG)* for comparative evaluation. The former is selected due its good performance and availability and the later is selected due to its availability and similarity to UGE (refer Chapter 3). This evaluation is done using two sets of experiments.

First experiment is a coverage test which tests what percentage of the sentences are parsed using all three parsers (Stanford parser, LG and UGE). For this exper-

iment, one hundred sentences from New Zealand Herald newspaper articles were used as a data set. These sentences are different from those that were used in the self evaluation tests. These hundred sentences are grammatically correct sentences according to English grammar and they do not have any special symbols inside them. To cover a wide range of vocabulary, articles from different domains such as business, sports and technology were selected. Here, if any words in the sentence are not found in the UGE dictionary, they are added to the UGE dictionary before testing against UGE.

Second experiment is an accuracy test which tests how meaningful the parser outputs are. Here, the two best performed parsers from experiment 1 are compared for meaningfulness of their output.

5.3.1 Experiment 1

For this experiment, each sentence is parsed using three parsers. The results are summed up in Table 5.5. Table 5.5 shows the percentage of sentences parsed (Parsed%) with each parser and average parse options (Average parses) each parser returns. For example, if the parser returns one or more parse analysis for a particular sentence, it is considered that the sentence is parsed by that parser. In the case of LG, if it returns one or more complete linkages for a particular sentence, it is considered as a parse. This is because, LG also returns incomplete linkage which does not link all the words in the sentence.

Stanford parser only returns one output per sentence, unless you use the option

to return multiple outputs. In the latter case, you have to specify how many. Thus, this option is not used. Therefore, the average number of parse options are only relevant to UGE and LG.

Parser	Parsed%	Average parses
UGE	100%	4
LG	68%	21199
Stanford parser	100%	Unknown

Table 5.5: Comparative evaluation - Percentage of parsed sentences and average parse options returned

From this experiment, UGE and Stanford parser performed the best. LG only produces complete linkages for 68% of the sentences. LG also produces many parse options with an average of 21199 per sentence while UGE only produces on average 4 parse options per sentence.

5.3.2 Experiment 2

This experiment analyses the meaningfulness of the parser results. Since UGE and Stanford parser performed the best in the previous experiment, the outputs of these two parsers are analysed more carefully. They are checked manually to see if the output captures the intended meaning. Consequently, sentences from two single articles are used (these two articles are about the same story published in different times). In this way, the intended meaning is available. These articles have 21 and 17 sentences respectively with an average word length of 23.

UGE returned 92% of meaningful results while Stanford parser only returned 71% (see Table 5.6). Appendix A.7 shows the parses obtained from UGE and

Parser	Accuracy %
UGE	92%
Stanford parser	71%

Table 5.6: Comparative evaluation - Accuracy of parsed sentences

Stanford parser for one article. Consider parsing the sentence *"Ahmed phones regularly but I think the satellite link was so powerful because they hadn't actually seen each other in the flesh for nearly three"*. The results from both parsers are shown below:

UGE result:

```
(BUT* (:MS2
       (THINK* (:ACTOR (I* (:PNOUN)))
        (:MS1
         (BECAUSE*
          (:MS2
           (HAD* (:ACTOR (THEY* (:PNOUN)))
            (:WHAT
             (SEEN*
              (:WHAT
               (EACH-OTHER* (:PNOUN)
                (:IN*
                 (FLESH* (:NOUN) (:MODIFIER (THE*))
                         (:FOR*
                          (YEARS* (:NOUN)
                                  (:MODIFIER (THREE*))))
```

```
                        (:MANNER (NEARLY*)))))))))
                  (:MANNER (NOT*)) (:MANNER (ACTUALLY*))))
          (:MS1
           (LINK* (:NOUN) (:X-WORDS (SATELLITE*) (LINK*))
                  (:MODIFIER (THE*))
                  (:WAS*
                   (?R (:MODIFIER (SO*) (POWERFUL*)))))))))))
      (:MS1 (PHONES* (:ACTOR (AHMED* (:UNKNOWN) (:NAME)))
                     (:MANNER (REGULARLY*)))))
```

Stanford parser result:

```
(ROOT
  (S
   (NP (JJ Ahmed) (NNS phones))
   (ADVP (RB regularly))
   (ADVP (CC but))
   (NP (PRP I))
   (VP (VBP think)
       (SBAR
        (S
         (NP (DT the)
             (NN satellite)
             (NN link))
```

```
        (VP (VBD was)
            (ADJP (RB so)
                  (JJ powerful))
            (SBAR (IN because)
                  (S
                   (NP (PRP they))
                   (VP (VBD had)
                       (RB n't)
                       (ADVP
                        (RB actually))
                       (VP (VBN seen)
                           (NP (DT each)
                               (JJ other))
                           (PP (IN in)
                               (NP
                                (NP (DT the)
                                    (NN flesh))
                                (PP (IN for)
                                    (NP
                                     (QP
                                      (RB nearly)
                                      (CD three))
                                     (NNS years))
                                    ))))))))))
    (. .)))
```

From the above results, the output of UGE is correct, but that for the Stanford parser is incorrect (i.e. the word *"phones"* is interpreted as a noun term in the Stanford parser, which is wrong for this sentence).

5.3.3 Summary

The comparative evaluation shows that UGE and Stanford parsers produce parse analysis for 100% of the sentences while LG returns complete linkage for 68% of the sentences. It is also noted that LG returns, on average, 21199 parse options per sentence while UGE returns only 4 parses on average. When considering the correctness of the output produced for two single articles, UGE outperforms the Stanford parser.

5.4 Conclusions

Self evaluation methods reveal that UGE does parse 81% of real, complex sentences with 98% accuracy. The average words per sentence were 26 which shows that the test sentences were quite complex.

Also from self evaluation tests, it was shown that 19% of the sentences failed to parse using UGE. Within this, only 28% are new cases which need to be implemented. 16% are grammatically incorrect sentences and 18% are other cases. 38% of failed parses are due to missing dictionary entries. This shows almost 2/3 of the unsuccessful parses is not due to the failure of UGE.

In the comparative evaluation, UGE and Stanford parsers performed better com-

pared to LG with 100% to 68% in the parses respectively. However, when analysing the meaningfulness of the parses, UGE was shown to perform better than the Stanford parser.

In conclusion, UGE satisfies the thesis claim that Yeap's theory can be implemented to parse real world complex sentences. It not only parses real world complex sentences but also performs well when compared to other parsers.

Chapter 6

Conclusion

This thesis is concerned with extending and implementing Yeap's theory of language to parse real world complex sentences. We asked:

1. Can the theory be extended to parse real world complex sentences in English?, and

2. How well would such a model of parsing perform in the real world?

The implementation has highlighted several shortcomings of the program. These include missing words, missing rules, missing labels, and its inability to handle special symbols appearing in a sentence. The extended program can now be used to parse any file downloaded from the New Zealand Herald newspaper site with a high parsing rate. From a theoretical standpoint, what is important though is that the extension is done in accordance with the theory. This then shows that the theory itself is adequate for handling English sentences. The answer to the first question is thus a confirmed yes.

With respect to the second question, we have measured the performance of UGE in terms of its speed, parsing rate, and accuracy. We have also compared its performance against two traditional parsers, the Link Grammar and the Stanford Parser. The test data set came from the New Zealand Herald newspaper site. To ensure that a good range of articles were used, the test data was selected from different domains. Given that UGE is implemented using a different design philosophy, it is not a straightforward test to compare its performance with other parsers. Still, our results show that UGE performs well and could be scaled up for real world applications. UGE is so good that it is currently being trialled as part of an application program known as SmartInfo, a program that automatically extracts relevant results from all the text files returned as a result of a search using a publicly available search engine (such as Google).

It is interesting to note that the extended implementation of the theory did not reveal further significant insights about the theory itself. Throughout the implementation, problem sentences were studied to see how it could be parsed by UGE and usually it required one to add a new rule or a new label to UGE. This process is equivalent to a child learning new rules for some new aspects of language that the child has not learned before or for some creative use of language that the child has never encountered before. The way UGE is expanded could definitely shed light on how children learn their first language. However, such psychological testing is beyond the scope of this thesis. Little modification to the theory is needed and this demonstrates that the theory is sound and works well for a computer.

6.1 Future directions

The successful implementation of UGE to parse real world sentences has opened up interesting future directions for work in this area. Some of these include:

1. **Psychological reality -** The strong performance of UGE implies that it is now ready to be evaluated against human data. Could the way UGE works tell us anything about how humans learn their first language? It is important that future work investigates how children could learn the kind of grammar rules suggested in UGE. Does the learning of the labels follow a universal sequence? How does the brain learn the complex rules needed and in particular, could the different abstract groups be formed? Will the stimulus for a child be rich enough to learn the rules in UGE?

2. **Learning algorithm -** One of the biggest problems in developing UGE is that the rules and labels for words are learned from examples. At the moment, this is learned manually. Could a learning algorithm be developed? To do this automatically poses a serious problem since learning the rules requires understanding of the input. Most learning algorithms are based upon detecting common patterns in the input, which are not interpreted.

3. **UGE for other languages -** UGE is intended to be a universal grammar engine. In other words, one expects that the approach could be used for parsing other languages. It would be interesting to develop UGE for a different language, such as Chinese.

4. **Benchmarks testing -** UGE performance is currently measured manually using Grammatical relations (GRs). It would be nice if one could compare

UGE with other parsers using one of the standardised testing method such as Treebanks. However, it needs an extensive research to see how UGE output can be converted into Treebanks parse tree.

5. **Constructing a testable model -** This thesis uses the simple percentage approach in its experiments. However, using the quatitative data gathered in the experiments, it is possible to construct a statistical testable model. In future, it would be interesting to see how such a testable model could be built with quatitative data gathered.

Appendix A

Appendix

A.1 Abbreviations

- "Adm.", "Apr.", "Aug.", "Ave."
- "Blvd."
- "Capt.", "Cf.", "Col.", "Corp.", "Co.", "Corp."
- "Dec.", "Dr."
- "etc."
- "Feb."
- "Gen."
- "Inc.", "Inc."
- "Jan.", "Jul.", "Jun."
- "Ltd.", "Lt."
- "Mar.", "Mr.", "Mrs.", "Ms.", "Mt."
- "Nov.", "No."
- "Oct."
- "Pres.", "Pvt.", "Prof.", "PLC."
- "Rep."
- "Sen.", "Sept.", "Sep.", "Sgt.", "St."
- "yrs."

A.2 Compound words

Compound words

- "ahead-of", "all-of", "all-over", "and-that", "as-though", "as-if", "as-soon-as", "as-much-as", "at-least"
- "because-of", "both-of"
- "each-other", "either-of", "every-time", "everyone-else", "everyone-else's", "even-if", "even-though", "even-as", "even-so", "even-now", "even-then"
- "face-to-face", "for-example", "for-sure", "for-about", "free-and-clear"
- "how-much", "how-to", "how-many"
- "in-connection-with", "in-connection-to", "in-connection-for", "instead-of"
- "job-search"
- "less-of", "little-of"
- "many-of", "more-of", "more-and-more", "more-or-less", "more-or", "most-of", "mr-and-mrs"
- "neither-of", "new-born", "no-wonder", "no-one", "not-only"
- "only-if", "other-than"
- "post-mortem"
- "rather-than"
- "so-that", "so-far", "so-much", "so-fast", "someone-else", "someone-else's", "such-as"
- "the-moment"
- "up-to"
- "what-to"

A.3 Pre-processor output

For the sentence *"But if Berry was to be released with no more than a "bad luck, my boy - hope you learned something by all this", then the wrong message would be going out, said Mr Neels."*

```
((#S(LEXICAL-TOKEN :STRING "but"
                   :TYPE :WORD
                   :FLAVOR :STANDARD
                   :CASE :CAPITALIZED
                   :PROPERTIES (:NE-STATUS NIL)
                   :POINTER NIL)
  #S(LEXICAL-TOKEN :STRING "if"
                   :TYPE :WORD
                   :FLAVOR :STANDARD
                   :CASE :LOWERCASE
                   :PROPERTIES (:NE-STATUS NIL)
                   :POINTER NIL)
  #S(LEXICAL-TOKEN :STRING "berry"
                   :TYPE :WORD
                   :FLAVOR :STANDARD
                   :CASE :CAPITALIZED
                   :PROPERTIES (:NE-STATUS NIL)
                   :POINTER NIL)
  #S(LEXICAL-TOKEN :STRING "was"
                   :TYPE :WORD
                   :FLAVOR :STANDARD
                   :CASE :LOWERCASE
                   :PROPERTIES (:NE-STATUS NIL)
                   :POINTER NIL)
  #S(LEXICAL-TOKEN :STRING "to"
                   :TYPE :WORD
                   :FLAVOR :STANDARD
                   :CASE :LOWERCASE
                   :PROPERTIES (:NE-STATUS NIL)
                   :POINTER NIL)
  #S(LEXICAL-TOKEN :STRING "be"
                   :TYPE :WORD
                   :FLAVOR :STANDARD
                   :CASE :LOWERCASE
                   :PROPERTIES (:NE-STATUS NIL)
                   :POINTER NIL)
  #S(LEXICAL-TOKEN :STRING "released"
                   :TYPE :WORD
                   :FLAVOR :STANDARD
                   :CASE :LOWERCASE
                   :PROPERTIES (:NE-STATUS NIL)
                   :POINTER NIL)
  #S(LEXICAL-TOKEN :STRING "with"
                   :TYPE :WORD
                   :FLAVOR :STANDARD
                   :CASE :LOWERCASE
                   :PROPERTIES (:NE-STATUS NIL)
                   :POINTER NIL)
```

```
#S(LEXICAL-TOKEN :STRING "no"
                 :TYPE :WORD
                 :FLAVOR :STANDARD
                 :CASE :LOWERCASE
                 :PROPERTIES (:NE-STATUS NIL)
                 :POINTER NIL)
#S(LEXICAL-TOKEN :STRING "more"
                 :TYPE :WORD
                 :FLAVOR :STANDARD
                 :CASE :LOWERCASE
                 :PROPERTIES (:NE-STATUS NIL)
                 :POINTER NIL)
#S(LEXICAL-TOKEN :STRING "than"
                 :TYPE :WORD
                 :FLAVOR :STANDARD
                 :CASE :LOWERCASE
                 :PROPERTIES (:NE-STATUS NIL)
                 :POINTER NIL)
#S(LEXICAL-TOKEN :STRING "a"
                 :TYPE :WORD
                 :FLAVOR :STANDARD
                 :CASE :LOWERCASE
                 :PROPERTIES (:NE-STATUS NIL)
                 :POINTER NIL)
   #S(LEXICAL-TOKEN :STRING
                    ((#S(LEXICAL-TOKEN
                         :STRING "bad"
                         :TYPE :WORD
                         :FLAVOR :STANDARD
                         :CASE :LOWERCASE
                         :PROPERTIES (:NE-STATUS NIL)
                         :POINTER NIL)
                         #S(LEXICAL-TOKEN
                            :STRING "luck"
                            :TYPE :WORD
                            :FLAVOR :STANDARD
                            :CASE :LOWERCASE
                            :PROPERTIES (:NE-STATUS NIL)
                            :POINTER NIL)
                         #S(LEXICAL-TOKEN
                            :STRING NIL
                            :TYPE :COMMA
                            :FLAVOR NIL
                            :CASE NIL
                            :PROPERTIES (:NE-STATUS NIL)
                            :POINTER NIL)
                         #S(LEXICAL-TOKEN
                            :STRING "my"
                            :TYPE :WORD
                            :FLAVOR :STANDARD
                            :CASE :LOWERCASE
                            :PROPERTIES (:NE-STATUS NIL)
                            :POINTER NIL)
                         #S(LEXICAL-TOKEN
                            :STRING "boy"
                            :TYPE :WORD
                            :FLAVOR :STANDARD
                            :CASE :LOWERCASE
                            :PROPERTIES (:NE-STATUS NIL)
                            :POINTER NIL)
                         #S(LEXICAL-TOKEN
                            :STRING
```

```
                        ((#S(LEXICAL-TOKEN
                             :STRING "hope"
                             :TYPE :WORD
                             :FLAVOR :STANDARD
                             :CASE :LOWERCASE
                             :PROPERTIES (:NE-STATUS NIL)
                             :POINTER NIL)
                             #S(LEXICAL-TOKEN
                                :STRING "you"
                                :TYPE :WORD
                                :FLAVOR :STANDARD
                                :CASE :LOWERCASE
                                :PROPERTIES (:NE-STATUS NIL)
                                :POINTER NIL)
                             #S(LEXICAL-TOKEN
                                :STRING "learned"
                                :TYPE :WORD
                                :FLAVOR :STANDARD
                                :CASE :LOWERCASE
                                :PROPERTIES (:NE-STATUS NIL)
                                :POINTER NIL)
                             #S(LEXICAL-TOKEN
                                :STRING "something"
                                :TYPE :WORD
                                :FLAVOR :STANDARD
                                :CASE :LOWERCASE
                                :PROPERTIES (:NE-STATUS NIL)
                                :POINTER NIL)
                             #S(LEXICAL-TOKEN
                                :STRING "by"
                                :TYPE :WORD
                                :FLAVOR :STANDARD
                                :CASE :LOWERCASE
                                :PROPERTIES (:NE-STATUS NIL)
                                :POINTER NIL)
                             #S(LEXICAL-TOKEN
                                :STRING "all"
                                :TYPE :WORD
                                :FLAVOR :STANDARD
                                :CASE :LOWERCASE
                                :PROPERTIES (:NE-STATUS NIL)
                                :POINTER NIL)
                             #S(LEXICAL-TOKEN
                                :STRING "this"
                                :TYPE :WORD
                                :FLAVOR :STANDARD
                                :CASE :LOWERCASE
                                :PROPERTIES (:NE-STATUS NIL)
                                :POINTER NIL)))
                        :TYPE :B-BLOCK
                        :FLAVOR NIL
                        :CASE NIL
                        :PROPERTIES (:NE-STATUS NIL)
                        :POINTER NIL)))
                  :TYPE :S-BLOCK
                  :FLAVOR NIL
                  :CASE NIL
                  :PROPERTIES NIL
                  :POINTER NIL)
#S(LEXICAL-TOKEN :STRING NIL
                 :TYPE :COMMA
                 :FLAVOR NIL
```

```
                :CASE NIL
                :PROPERTIES (:NE-STATUS NIL)
                :POINTER NIL)
#S(LEXICAL-TOKEN :STRING "then"
                :TYPE :WORD
                :FLAVOR :STANDARD
                :CASE :LOWERCASE
                :PROPERTIES (:NE-STATUS NIL)
                :POINTER NIL)
#S(LEXICAL-TOKEN :STRING "the"
                :TYPE :WORD
                :FLAVOR :STANDARD
                :CASE :LOWERCASE
                :PROPERTIES (:NE-STATUS NIL)
                :POINTER NIL)
#S(LEXICAL-TOKEN :STRING "wrong"
                :TYPE :WORD
                :FLAVOR :STANDARD
                :CASE :LOWERCASE
                :PROPERTIES (:NE-STATUS NIL)
                :POINTER NIL)
#S(LEXICAL-TOKEN :STRING "message"
                :TYPE :WORD
                :FLAVOR :STANDARD
                :CASE :LOWERCASE
                :PROPERTIES (:NE-STATUS NIL)
                :POINTER NIL)
#S(LEXICAL-TOKEN :STRING "would"
                :TYPE :WORD
                :FLAVOR :STANDARD
                :CASE :LOWERCASE
                :PROPERTIES (:NE-STATUS NIL)
                :POINTER NIL)
#S(LEXICAL-TOKEN :STRING "be"
                :TYPE :WORD
                :FLAVOR :STANDARD
                :CASE :LOWERCASE
                :PROPERTIES (:NE-STATUS NIL)
                :POINTER NIL)
#S(LEXICAL-TOKEN :STRING "going"
                :TYPE :WORD
                :FLAVOR :STANDARD
                :CASE :LOWERCASE
                :PROPERTIES (:NE-STATUS NIL)
                :POINTER NIL)
#S(LEXICAL-TOKEN :STRING "out"
                :TYPE :WORD
                :FLAVOR :STANDARD
                :CASE :LOWERCASE
                :PROPERTIES (:NE-STATUS NIL)
                :POINTER NIL)
#S(LEXICAL-TOKEN :STRING NIL :TYPE :COMMA
                :FLAVOR NIL :CASE NIL
                :PROPERTIES (:NE-STATUS NIL)
                :POINTER NIL)
#S(LEXICAL-TOKEN :STRING "said"
                :TYPE :WORD
                :FLAVOR :STANDARD
                :CASE :LOWERCASE
                :PROPERTIES (:NE-STATUS NIL)
                :POINTER NIL)
#S(LEXICAL-TOKEN :STRING "mr"
```

```
                    :TYPE :WORD
                    :FLAVOR :STANDARD
                    :CASE :CAPITALIZED
                    :PROPERTIES (:NE-STATUS NIL)
                    :POINTER NIL)
    #S(LEXICAL-TOKEN :STRING "neels"
                    :TYPE :WORD
                    :FLAVOR :STANDARD
                    :CASE :CAPITALIZED
                    :PROPERTIES
                    (:NE-STATUS
                     :COMPOUND-NOUN :COMPOUNDED-TOKENS
                     (#S(LEXICAL-TOKEN :STRING "neels"
                                       :TYPE :WORD
                                       :FLAVOR :STANDARD
                                       :CASE :CAPITALIZED
                                       :PROPERTIES
                                       (:NE-STATUS :NE-START)
                                       :POINTER NIL)))
                    :POINTER NIL)))
```

A.4 Test data

(“I eat an apple.” “he is so powerful.” “he never considered the situation an emergency.” “He will do two live interviews a day.” “Relational Technology went public in May 1988 at $14 a share.” “He charges 100 dollars an hour.” “He had a gun pointed at him.” “All I have is here.” “every firm saw costs grow more than expected, even after adjusting for inflation.” “We are at the office now.” “it was possible he did not initially realise he or she had hit someone.” “Then when he was around 14 or 15 I suddenly became aware he was becoming a man.” “I want to know right from wrong.” “The jokes put me in mind of a funny thing.” “But Metrick suggests investors keep in mind the generally negative effects of dual-class ownership when they consider buying stocks like Google.” “I will be there tomorrow.” “I might be late.” “He wants to be a computer scientist.” “Be nice now.” “Please be quiet.” “I will be back tomorrow.” “Ms. Mariam will be going to United Kingdom next month.” “John has been a doctor since 1998.” “That family have been living there since this April.” “She is just being nice to everyone in the office.” “Rachel has been there before.” “I have never been to Maine.” “We have been living here since June.” “You should have been paying attention.” “Being a good sport is important.” “Being on time is essential.” “I was being a jerk.” “we don’t have anything that competes with that.” “I promise to wait.” “He pretended to be angry.” “I happened to be looking out of the window.” “I must go to help my mother.” “He forgot to leave the key on the table.” “I want to eat the apple quickly.” “I like to half close my eyes.” “To save money now seems practically impossible.” “To lean out of the window is dangerous.” “He wanted to go but he wasn’t able to.” “I arranged for Jane to meet John.” “I arranged with Jane to meet John.” “He discovered how to open the safe.” “I want my children to learn Chinese well.” “He hurried to the house only to find that it was empty.” “No one is to leave this building“ “She is to be married in January.” “He is always the first to come.” “This is the best play to be performed this year.” “The windows ought to be opened.” “The money were to be kept here.” “It is stupid of him to smoke.” “It is careless of me to lose my umbrella at school.” “I go to school to play basketball.” “That is a stupid place to park a car.” “I was delighted to see him.” “It is easy to talk.” “She works too slowly to be much use to me.” “She is old enough to travel by herself.” “He drives fast to work.” “He made me move my car“ “I felt the house shake“ “She let us use her phone.” “Your job is to really understand these children.” “You are to give her an apple“ “I want to go to eat an apple.” “John is the man to ask the question.” “The person to arrive first was John.” “The first person to arrive was John.” “The first to arrive was John.” “I

wanted to go home to eat an apple." "I wanted to go home yesterday to eat an apple." "you ought to finish your work before going out." "she used to like him a lot but she does not like him anymore." "John is taller than Mary." "John is taller than Mary is." "A man taller than John ate all apples." "John has one leg shorter than the other one." "What Alice did annoyed me." "What Alice did really annoyed me." "I enjoyed what Alice did in the party." "I know what he was thinking." "He told me what he was doing." "I know what happened." "Her last thoughts was what was going to happen to granddad." "It is easy to see why he left home." "He asked John to go to the market." "He asked John not to go to the market." "he left home four months before the crash." "I met John who left home four months before the crash." "I mix them together in the park." "he had been asked not to talk to the media." "She believes someone has taken her daughter from the street." "The end markets for software products are intensely and increasingly competitive." "As of June 30, 2006, Microsoft employed approximately 71,000 people on a full-time basis, 44,000 in the United States and 27,000 internationally." "The Vanguard REIT Index Fund, to name one, saw returns of 36% last year, after average gains of 15% annually in the prior three years." "The distrust won't go away until the economy revs up again." "Microsoft and Ashton-Tate announce Microsoft-SQL-Server, relational-database server software for Local-Area-Networks (LANs) based on a relational-database management system licensed from Sybase." "Microsoft-Research has more than 700 employees, including some of the world's finest computer-scientists, sociologists, psychologists, mathematicians, physicists, and engineers." "We investigated whether experience of the siege did in fact lead to an increased risk of mortality, particularly from cardiovascular-disease." "For short-term prospects, close monitoring is required for the future movement of the U.S.-economy." "you have to do it quickly and accurately." "you will do it quickly and accurately." "I am doing it quickly and accurately." "The police advised against our entering to the building." "I cannot bear being pushed around in crowds" "He went on giggling, not having noticed the teacher enter" "I am considering sleeping over if you do not mind" "The cost of RFID remains difficult to justify." "16% of couples got married in clothes they already owned." "Weddings no longer signify the major life change." "But it is something they should know." "We can and should do better than this." "They sacrifice some performance to maintain control." "As voting power grew from zero to 45%, Tobins Q fell by 25%." "I was here to witness the accident." "Many other symptoms can occur including suicidal-thoughts." "I was here to witness the accident." "Total assets declined 1.3 percent, amounting to $ 66,371 million." "Some are visually-impaired or deaf." "While imports grew strongly, the trade-balance showed only marginal improvements." "It was also recognised as the first agricultural, industrial and service power." "But at dual-class

companies, ownership stake and voting control are not necessarily in the same hands." "The new advice raises several questions, two of which are addressed here." "Corporations sense of excess-capacity, which indicates future business-investment, has been on an improving trend." "Private consumption is almost flat, reflecting decelerating income, with consumer-confidence moving horizontally." "According to the BOJ tankan, fiscal 2006 sales are expected to post an increase for the fourth-consecutive-year." "Employment will increase moderately, and the unemployment-rate will continue to fall, albeit at a considerably slower pace than in 2006." "The economy is recovering, despite some weakness in consumption." "Microsoft's brand was ranked No 1 in terms of value, according to the 2006 Millward-Brown-Optimor Survey." "Online Services Microsoft distributes online content and services through MSN and other online channels." "Microsoft-products may include some components that are available from only one or limited sources." "Or he could have many super shares and few ordinary ones, enjoying lots of control but having a small financial stake." "Individual consumers obtain the products primarily through retail-outlets, including Best-Buy, Target, and Wal-Mart." "investors may be reluctant to purchase the inferior voting-stock of these firms, and they may therefore have to rely more heavily on debt-financing" "A potential explanation for firms heavier reliance on debt-financing is that investors may be reluctant to purchase the inferior voting-stock of these firms, and they may therefore have to rely more heavily on debt-financing," the authors write." "A potential explanation for dual-class firms heavier reliance on debt-financing is that investors may be reluctant to purchase the inferior voting-stock of these firms, and they may therefore have to rely more heavily on debt-financing," the authors write." "The same patterns were found with capital-expenditures and the expenses for research-and-development and advertising, with the pattern most pronounced in sales growth and advertising expenditures, Metrick said." "The authors looked at each company value, measured by Tobins Q, calculated by dividing the market-value of its assets by the assets replacement-value." "An executive might have a large ownership stake composed of ordinary single-vote shares, but at the same time possess little voting clout because the multi-vote shares are in the hands of others." "he is running fast" "he is doing well" "they consider buying stocks like Google." "Housing-starts are expected to move steadily as long as the income environments in households continue to recover along with the improvement in the employment situation." "The dissenters of the Warren Court were often defending a legal legacy that they inherited, says Prof. A.E Dick Howard." "I ate an apple" said John." "I ate an apple, said John of Dunedin." "I ate an apple, said John Nathan of Dunedin." "She said she started screaming after seeing some words written on a sack he had produced." "every firm saw costs grow more than expected"

“He said this morning the weather is going to be bad.” “Should you program to it or not¿‘ “Should you program it or not¿‘ “The boy who we met yesterday is very nice.” “He paid $500 in fines.” “he drives fast car“ “If Ahmed Zaoui’s family succeed in their application to join him in New Zealand as refugees, the decision will annoy NZ First leader Winston Peters but delight a boy who has not seen his father for nearly three years.” “Mr Zaoui’s wife, Leila, and four sons - Youssef, 7, Adbel, 14, Soheib, 17, and Hicham, 19 - have applied through Mr Zaoui’s lawyer, Deborah Manning, to move to New Zealand as refugees.” “Ms Manning told the Sunday Star-Times that supporters had offered the family financial assistance, but she could not preclude the need for the family to apply for a benefit, to which refugees are entitled.” “The Government has refused to consider the applications until Mr Zaoui’s case is resolved (he is on bail awaiting a review of his SIS security risk certificate, which could mean his deportation).” “The applications have been criticised by the National and New Zealand First parties.” “While the politicians bicker over issues of cost, security risk and human rights, for Mr Zaoui’s youngest son all it means is growing up without a father.” “The family spoke of the separation to TV3’s Campbell Live producer Carol Hirschfeld, who travelled to interview them at a Southeast Asian location for a segment to screen tonight.” “They also got a taste of what they were missing - a satellite link let Mr Zaoui and the family talk face to face for the first time since he left for New Zealand in December 2002.” “Hirschfeld said separation had a strong effect on son Youssef, who was 4 1/2 years old when he last saw his father.” “Ahmed phones regularly but I think the satellite link was so powerful because they hadn’t actually seen each other in the flesh for nearly three years.” “Seeing how much Youssef had grown was something Ahmed really responded to.” “Hirschfeld said the family had perceived NZ as “a peaceful land which offered a place of refuge with a good record for democracy“.” “National’s immigration spokesman, Tony Ryall, said calls for the family to be allowed in before Mr Zaoui’s case was settled were ludicrous.” “We shouldn’t have a system in New Zealand where you get one refugee and end up with several others.” “It puts a lot of pressure on the system.” “If it was decided Mr Zaoui should go but his family had already been let in, then Mr Zaoui would have another argument for coming back, said Mr Ryall.” “Mr Peters said at his campaign launch in Takapuna yesterday that the Zaoui case had already cost the country $2 million.” “Mr Zaoui, you can see your family tomorrow if you would just get on a plane and go and see them.” “But the Green Party said Mr Zaoui should be shown some “Kiwi compassion“ because his separation from his family was the fault of deficiencies in New Zealand’s own systems.” “He had also spent two years in jail “unnecessarily“.” “The party said allowing the family to get together could not represent a security danger to New Zealand.” “The owner of a 4WD vehicle seized in connection with

the Birgit Brauer murder case says it was stolen by his employee six weeks ago." "The Toyota Hilux is registered to Palmerston North man Brent Cleverley, who confirmed to the Herald yesterday that his vehicle had been taken by a man who worked for him." "The Herald understands the worker, who used the vehicle in his job cutting firewood, failed to turn up to work one day last month." "He has not been seen since." "Mr Cleverley, whose 4WD was taken for forensic tests, said he had been asked by police not to talk to the media." "His worker is understood to have grown up in the Himatangi/Foxton area but had recently spent time in the South Island before moving back to Levin." "He is known by various names." "He is in his mid-30s and lived in Levin, about 4km from the Ohau River where the vehicle was found yesterday." "After it was discovered dozens of police officers, including forensic experts and a dive squad, arrived at the river bed to secure and examine the scene." "Last night Detective Senior Sergeant Grant Coward would not call the man a suspect in the killing of the German backpacker, but has confirmed police know his name and want to find him." "Last Friday police made inquiries about the ownership of the specific 4WD vehicle as part of routine inquiries to account for all dark grey or black Toyota Hiluxes, like the one Ms Brauer travelled in before her death last week." "Her body was found at Lucys Gully, near New Plymouth, on September 20." "The 28-year-old had been stabbed and had head injuries." "Mr Coward, the head of the investigation, has repeatedly said the key to catching the killer lay in finding the mid-1980s Hilux Ms Brauer was seen getting into." "The one airlifted out of the Ohau River yesterday clearly matched the description." "Retired schoolteacher Jenny Burnell, who lives on the farm which backs on to the riverbed, said the vehicle was dumped there on Monday night." "Her nephews saw its lights along a track on Miss Burnell's farm about 9.30pm." "They investigated but did not find anything." "The vehicle was seen on the edge of the riverbed yesterday morning and reported to police, who used a helicopter to remove it from the water." "Miss Burnell said it was an unusual place for the vehicle to be dumped as few people would know how to get there." "Taranaki helicopter company owner Alan Beck said police called for one of his heavy-lift Super Huey Iroquois helicopters." "Our job was to lift it out because there were fears if the water suddenly rose they might lose valuable evidence," Mr Beck said." "Mr Coward said there was no evidence that the person who dumped the vehicle in the river was the murder suspect, but police would still like to talk to him." "Obviously we want the person who stole that vehicle to come forward." "That would be most helpful." "Someone who's committed this crime would, in my view, be acting differently than they would beforehand." "They have done something horrendous and they know it." "I can assure you if someone knows who the killer is and they want to be treated confidentially, they will be treated with confidence."

“Mr Coward said police had received many reports of vehicles matching the description of the one Ms Brauer was last seen in.” “Several had been recovered, including three stolen ones.” “Land Transport NZ figures show that almost 23,000 Toyota Hiluxes were registered in New Zealand between 1985 and 1990.” “Of those, 673 were in the Taranaki area.” “Mr Coward said police had tracked sightings of the vehicle from Waitotara, where Ms Brauer was seen getting into a Toyota Hilux, to Lucys Gully, where her body was found.” “The vehicle had then been seen at Cardiff, near Stratford.” “The Welsh rugby fan who killed a young Waikato woman when his campervan smashed into her oncoming car has been ordered to pay almost $10,000 in fines and reparation.” “James Berry, 23, had previously pleaded guilty to careless driving causing injury and careless driving causing the death of 18-year-old Cambridge woman Liz Neels.” “This morning in the Hamilton District Court he stood looking scared and vulnerable as he was handed down a $500 fine, plus costs, for the injury charge.” “Judge Anne McAloon said she took into account that Berry had been held in custody for three days after the June 23 crash and the volunteer work he had been doing since his initial court appearance.” “The accounting graduate from Swansea had come to New Zealand to follow the Lions tour but had not seen any of the games.” “The judge ordered him to pay $9000 in reparation to the Neels family on the causing death charge.” “Berry’s mother was in the court for the sentencing, and cried throughout the proceedings.” “Liz Neels’ father Michael told the court he thought more statistics needed to be kept on the number of accidents caused by tired drivers.” “Mr Neels and his wife, Lori, said before the sentencing that they did not want to see Berry jailed or “financially crippled with some huge fine“.” “But if Berry was to be released with no more than a “bad luck, my boy - hope you learned something by all this“, then the wrong message would be going out, said Mr Neels.” “None of our hopes and concerns for her future have any further meaning,“ Mr Neels wrote in his victim impact report to go before the judge.” “The clear message was that it was not acceptable to drive when tired, he said.” “Berry had arrived from Britain with two friends mid-morning on June 23.” “After a break at his sister’s house in Auckland, they left about 4pm, intending to stop for the night at 10pm and catch the interisland ferry from Wellington the next day to get to Christchurch for the first Lions versus All Blacks test.” “About 7.40pm, the campervan crossed the centre line on State Highway 1, near Karapiro, slamming into a car driven by Liz Neels, a student chef.” “After initial reluctance to meet Berry at a restorative justice conference, a deeply grieving Mr and Mrs Neels came to think of him as a victim too, “just in a completely different tragedy, connected only by Liz“.” “They thought they should take the opportunity before Berry returned to Wales.” “We’re glad we did.” “Mr Neels doubted they could have faced him had he been a “boy-racer or a drunk driver“

instead of an intelligent, sensitive young man with a promising future." "A nephew and Liz's two best friends accompanied the couple to the voluntary meeting, run by facilitators." "Berry came alone" "I have never seen a person so contrite as James when he walked into the conference room to meet us." "Words were not needed to express how sorry he was for what had happened," said Mr Neels." "Berry expressed a wish to meet the Neels, appeared remorseful and even sent a card." "It seemed that, in falling asleep [at the wheel], he had made a very human mistake with devastating consequences," Mr Neels said." "The Neels said they wanted Berry to live a good life, both for himself and for Liz, and not carry her in his memory as a burden, rather "letting our Lizzie rest lightly on his soul"." "No further communication was planned, said Mr Neels." "But if he wants to contact us in five years' time and let us know how he is doing, what his wife is like, I daresay we will show interest." "A woman has been jailed in Italy for four months for pretending to be a ghost." "She spent weeks terrorising her husband's employer at the 15th century Castel Coldrano, on the border between Italy and Austria, slamming doors, haunting hallways and making things go bump in the night." "The Austrian owner of the castle called the police, who decided to video the estate." "And instead of an apparition, they caught a 42-year-old Polish woman." "She had spent her nights masquerading as a phantom to scare the estate owner." "An unexplained grievance had provoked her campaign." "A shot was fired at a motorist after he stopped to find out why a man was crouching in the middle of State Highway 5 at 11.30pm on Friday night, police said today." "The incident was being treated as a case of attempted aggravated robbery as Napier and Taupo police continued their investigations into the incident today." "Sergeant Mal Lochrie said the motorist told police he had been driving from Napier to Taupo when he spotted someone crouching in the middle of the road about 1km from the Tarawera Tavern." "He told police he slowed and stopped to determine whether the person needed assistance and as a result, he had a gun pointed at him." "He was told to get out of his vehicle," Mr Lochrie said." "Instead, he "put the pedal to the metal" and drove on, his car glancing the gunman and knocking him to the ground." "The man told police that as he accelerated away he heard a shot fired." "When he arrived in Taupo he went to the police station and reported the incident." "He described the gunman as a clean-shaven Caucasian, in his 30s and wearing a checked bush-style shirt." "Police described the incident as "a strange one" and appealed for anyone who may have seen the offender or heard a shot in that area at the time, to contact police." "Police were also seeking sightings of the victim's silver Mazda MS6 car on the Napier-Taupo highway as part of piecing the investigation together." "A busy North Shore road was brought to a standstill last night when a Stagecoach bus burst into flames - the third one to do so in the past six months."

“The Herald understands the bus was being driven back to the depot by a maintenance man when flames appeared from the back of it.” “Albany resident Rex Auty was heading home from the beach along Oteha Valley Rd when he saw people waving at the intersection ahead of him.” “The next minute he heard a loud bang and saw the bus, which was on the side of the road, on fire.” “It was quite an inferno,“ said Mr Auty.” “The back of the bus was well ablaze.” “There were a series of loud bangs as more and more parts of the bus exploded.” “ “He said members of the public blocked off the road to prevent motorists driving into the path of the bus.” “Oteha Valley Rd resident David McGrath was looking out his window when he saw the bus pull up near his house just before 7pm.” “It just pulled up at the bus stop on fire.” “Mr McGrath said the fire started around the engine bay, causing insulation, plastic and rubber to burn.” “Firefighters arrived about 15 minutes after the bus pulled up and the fire was extinguished with foam.” “No passengers were on board the bus, which was handed over to the maintenance man after the driver experienced problems with the acceleration.” “Stagecoach spokesman Russell Turnbull said a full investigation would be held into what caused the fire.” “In July Stagecoach management moved to reassure passengers its vehicles were safe after two buses caught fire within a week.” “The first happened on a bus carrying about 50 passengers and was caused by an electrical fault at the rear of the bus.” “The second happened on an empty bus travelling along KRd and was caused by a fuse at the front of the bus.” “Mr Turnbull said it was too early to say what caused last night’s blaze.” “Any bus fire we have is of concern and the main thing is to find out what we can do to make sure there is no connection between them that we need to be looking out for.” “An investigation into the latest fire was expected to start first thing today.” “Ralph Timoteo was told to prepare for the worst as doctors wheeled his 16-year-old son, Lincoln Hapeta, into surgery following a hit-and-run.” “The impact of the accident had torn the teenager’s liver and caused massive bleeding, which an initial operation failed to fix.” “As Lincoln underwent his second operation on Saturday his family waited anxiously at Middlemore Hospital, hoping for the best.” “They were also hoping the person who hit Lincoln planned to hand him or herself in.” “The accident happened late on Friday night as Lincoln and his cousin were waiting at the end of their Raglan St, Mangere, address for relatives to pick them up.” “It is not known exactly what happened, but Lincoln, who was in Auckland visiting family for the school holidays, was hit by a van.” “The driver did not stop.” “The Rotorua Boys High School student was thrown across the footpath and onto the grass verge.” “He came to rest up against a white picket fence.” “Bleeding and in pain, he told his cousin he was struggling to breathe.” “It was roughly at that time that Mr Timoteo noticed some commotion at the end of his relative’s long driveway and went to investigate.” “I walked around

the corner, and he [Lincoln] was on the ground," he said." "It was gut wrenching." "I just held him and tried to control my emotions." " "Mr Timoteo travelled in the ambulance with his son, who needed four times his usual blood levels pumped into him to replace lost blood." "At the hospital Mr Timoteo called Lincoln's mother, Kirsten Hapeta, who lives in Otaki." "She boarded a plane and arrived soon after the second operation had begun." "They had said to Ralph, 'You should prepare for the worst'." "When I got there 10 minutes later, that's what he said to me so we were freaking out." "Fortunately doctors were able to stop the bleeding and repair Lincoln's torn liver." "Last night he remained in intensive care, but his parents are confident that he will recover." "With that in mind they now have time to think about the person who left Lincoln lying injured on the ground." "Mr Timoteo said he was initially really angry but that has subsided somewhat." "Our elders are keeping us grounded." "When one talks about retribution, it solves nothing." "Mr Timoteo and Ms Hapeta urged the driver to hand himself of herself in to police." "If they don't get caught ... there's a possibility it might happen again," said Ms Hapeta." "It would be good for them to come and see Lincoln, if he wanted to." "I would like [the driver] to see what it's done to all the family," she added." "Detective Sergeant Richard O'Connor said it was possible the driver did not initially realise he or she had hit someone." "However, there would probably be evidence of the accident on the front left-hand side of the van and the person would know by now what had happened." "The vehicle is described as a white van, possibly with bull bars." "The windscreen or lights may be cracked or shattered, and there is possibly panel damage." "Anyone with information can contact Mr O'Connor or the Counties Manukau crime squad on (09) 261-1327 or 027 498-9010." "The anguished parents of a Cambridge teenager who died after a campervan driven by a Lions rugby supporter crashed into her car have met and forgiven the young man." "It is another comfort to us." "We have no animosity towards him," Michael Neels said yesterday, on the eve of the sentencing of James Berry." "Mr Neels and his wife, Lori, who lost their only child, Elizabeth, 18, do not want to see Berry jailed or "financially crippled with some huge fine"." "But if Berry was to be released with no more than a "bad luck, my boy - hope you learned something by all this", then the wrong message would be going out, said Mr Neels." "The clear message was that it was not acceptable to drive when tired, he said." "Berry had arrived from Britain with two friends mid-morning on June 23." "After a break at his sister's house in Auckland, they left about 4pm, intending to stop for the night at 10pm and catch the interisland ferry from Wellington the next day to get to Christchurch for the first Lions versus All Blacks test." "About 7.40pm, the campervan crossed the centre line on State Highway 1, near Karapiro, slamming into a car driven by Liz Neels, a student chef." "Despite the heart-wrenching

memories, her parents' biggest comfort has been the donation of their daughter's organs to seven people waiting for transplants." "But it is still hard for them to believe that the treasured child they took 12 years to conceive, who had left home only four months before the crash and was enjoying life, has gone." "None of our hopes and concerns for her future have any further meaning," Mr Neels wrote in his victim impact report to go before the judge." "Alongside physical injuries, he put: "Two broken hearts." " "After initial reluctance to meet Berry at a restorative justice conference, a deeply grieving Mr and Mrs Neels came to think of him as a victim too, "just in a completely different tragedy, connected only by Liz"." "They thought they should take the opportunity before Berry returned to Wales." "We're glad we did." "Mr Neels doubted they could have faced him had he been a "boy-racer or a drunk driver" instead of an intelligent, sensitive young man with a promising future." "A nephew and Liz's two best friends accompanied the couple to the voluntary meeting, run by facilitators." "Berry came alone." "I have never seen a person so contrite as James when he walked into the conference room to meet us." "Words were not needed to express how sorry he was for what had happened," said Mr Neels." "Berry expressed a wish to meet the Neels, appeared remorseful and even sent a card." "It seemed that, in falling asleep [at the wheel], he had made a very human mistake with devastating consequences," Mr Neels said." "Mrs Neels showed Berry photographs of Liz and they gave him an [unsigned] card bearing her picture, copies of which have gone to all those who sent messages of sympathy." "The Neels are not sure how they will feel after the sentencing." "They want Berry to live a good life, both for himself and for Liz, and not carry her in his memory as a burden, rather "letting our Lizzie rest lightly on his soul"." "No further communication was planned, said Mr Neels." "But if he wants to contact us in five years' time and let us know how he is doing, what his wife is like, I daresay we will show interest." "Berry has admitted charges of careless driving causing death and injury." "The 23-year-old accounting graduate from Swansea, in Wales, will learn his fate in the Hamilton District Court today." "New Zealand First is opposing efforts to bring Ahmed Zaoui's family to New Zealand." "Mr Zaoui's lawyers lodged an appeal with Immigration Minister David Cunliffe before Christmas to allow his wife and four children to come here." "Yesterday the Greens backed that call, saying delays in the case were unreasonable and should not be allowed to keep the family apart." "Mr Zaoui has been struggling to stay in New Zealand since he arrived in December 2002 seeking refugee status." "He is awaiting a review of his case." "The Department of Labour is preparing advice for Mr Cunliffe before he decides." "NZ First associate immigration spokesman Peter Brown said the family should not be allowed to come here from Southeast Asia." "Put simply, if Mr Zaoui wants to be with his family so badly, then there is nothing preventing him from

getting on a plane and going to be with them today." "Green MP Keith Locke said the family were not a security risk and met criteria for New Zealand's United Nations refugee intake." "But Mr Brown said Mr Zaoui had terrorism-related convictions from European courts and his case had cost taxpayers millions." "Mr Zaoui was convicted in Belgian and French courts on charges of association with terrorists, but his lawyers say that those charges were groundless." "He was also expelled from Switzerland and left Malaysia after reports the Algerian regime was seeking his extradition." "Mr Brown did agree with Mr Locke that the case had dragged on too long." "When Mr Zaoui, once elected an MP in Algeria, came to New Zealand, he sought refugee status on the grounds he would be tortured or killed if he was sent back to his homeland." "He spent almost two years in prison waiting for his case to be decided as he fought an SIS security risk certificate and moves to expel him from the country." "Mr Zaoui was released on bail in December 2004 after a Supreme Court hearing, and has since lived with the Catholic community in the Dominican Priory in Auckland awaiting a hearing." "The hearing to review the security risk certificate was due to be held last August but will not now go ahead until between June and August this year." "A female prison guard who was allegedly held hostage at knifepoint in a jail storeroom before being rescued by the armed offenders squad has told a court how she feared for her life." "Jeremy William Mataira, 46, is accused of holding the woman hostage at Paremoremo Prison for almost three hours last September." "He has been ordered to stand trial." "He appeared in North Shore District Court yesterday for a depositions hearing on charges including kidnapping, assault, indecent assault, assault with a weapon and assault with intent to commit sexual violation." "The guard was on duty in Paremoremo Prison's medium security wing when the 46-year-old inmate allegedly pulled her into a 2.7sq m cleaning cupboard." "When the armed offenders squad freed her, nearly three hours later, the inmate was stripped to his underwear." "The woman yesterday revealed in a statement how she thought she would be killed as she was held captive inside the storeroom on September 8." "The 53-year-old told police Mataira's duty was to look after the storeroom." "She said he went into the store room and returned with a plastic container of chocolate biscuits." "He offered her one but she declined." "He went back into the storeroom saying, "I've got something else for you here from the boys." " "She said he came out carrying a small towel folded in half with a knife inside it." "He took the knife from the towel ... and held it to my throat." "Then he grabbed me and pushed me into the storeroom." " "She described how she fought back and yelled at him to stop." "He tied her up and barricaded the door shut with shelving units, draws, boxes and paper towels." "She said she tried to calm him down as he was extremely hyped up and agitated." "She said he placed a blanket on the floor and created a canopy with another

blanket." "He stripped to his underwear." "She said she refused his offer to sit on the blanket with him." "He made several lewd comments and told her about sexual abuse he said he had suffered as a child." "She said she continually tried to reason with him, including telling him she had medication at home she needed to take." "He blindfolded her using material from a sheet, which he cut with scissors." "She described how she started to panic when he began kissing her feet." "She said she started screaming after seeing some words written on a sack he had produced." "As she tried to climb some shelving units, a window smashed and she could see people in black clothing." "The door was smashed in and Mataira grabbed her in a headlock." "She managed to break free." "I went towards the door where the rescuers were." "She said she passed out and woke up in an ambulance." "She suffered a black eye and abrasions to her face, wrists and legs." "Mataira will stand trial in the High Court at Auckland in March." "An A$8.7 ($10.01) billion takeover bid for Australia's Qantas Airways Ltd by a consortium led by Macquarie Bank Ltd is fair and reasonable, an independent expert said today." "In a statement issued by the airline, the independent expert said the offer price of A$5.60 a share is within its valuation range of A$5.18 to A$5.98 and any offer above the bottom of that range would be fair value." "Qantas said its board unanimously recommends the offer." "The consortium, Airline Partners Australia (APA), also includes private equity firm Texas Pacific Group , Allco Equity Partners, Allco Finance Group and Canadian investment firm Onex Corp." "In a letter to shareholders, Qantas chairman Margaret Jackson said no superior offer had been forthcoming and the APA price was a substantial premium on recent trading." "Qantas has delivered year-on-year profits, growth and diversification," Ms Jackson said." "But while the business had prospered, the Qantas share price has not." "The offer is the best available option to enable Qantas shareholders to realise significant value for their investment." "Police are investigating the suspicious death of a 16-year-old girl who was pronounced dead when she arrived at Whakatane Hospital last night." "Police said the girl lived locally and inquiries were continuing at her home." "The cause of death was not yet known and a post mortem would dictate the direction further inquiries would take." "Japan and its pro-whaling allies are hoping a special meeting of the International Whaling Commission will build momentum to resume commercial hunting of the giant creatures." "Japan wants to shift the commission's focus to whale management rather than a moratorium, but with 26 anti-whaling nations - including Australia, New Zealand and the United States - boycotting the meeting, prospects for dialogue in the organisation appear slim." "Thirty-four of the commission's 72 members are attending." "One of our goals is to improve the atmosphere of the IWC, which has become one of confrontation, and to improve dialogue," Minoru Morimoto, the commissioner for Japan, told the meeting." "It's a shame that

most anti-whaling nations chose confrontation," he said." "He hoped the commission would seriously consider "normalisation", Japanese code for resuming commercial hunting." "Ahead of the meeting, a Japanese whaling ship and protest vessels collided in the Southern Ocean." "In Tokyo, three anti-whaling protesters, including a man wearing a mask of Prime Minister Shinzo Abe's face, carried a signboard which read Welcome to the commercialisation meeting." "One activist was dressed as a weeping whale." "Pasted to the sign were 10,000 yen ($120) notes and names of several countries, an allusion to charges by anti-whalers that Japan has bought pro-whaling votes at the IWC with foreign aid." "Japan has repeatedly denied the allegations." "The IWC instituted a commercial whaling ban in 1986." "But the group is now bitterly divided between countries that assert all whales need protection and others, like Japan, that say some species are now abundant enough for limited hunting." "Japan, which says whaling is a cherished cultural tradition, began scientific research whaling in 1987." "The meat, which under commission rules must be sold for consumption, ends up in supermarkets and pricey restaurants but is far from a daily menu choice." "Some experts say Japan fears that limits on whaling will lead to limits on all Japanese fishing, a crucial food source in a nation that has limited agricultural land." "Allied Workforce Group (AWF) is signalling a 50 per cent decline in annual profit, blaming an investment in the rural sector that went sour." "The Blue collar labour hire company reported a maiden $3.02 million net profit for the year to March 31 last year, which was slightly below its prospectus forecast." "Today it said the profit in the year to March 31, 2007 was expected to be 50 per cent down on last year after the writeoff of investments." "The final dividend was also likely to lower than last year." "Managing Director Simon Hull said the company's involvement with Contract Labour Services NZ Ltd (CLS) in the rural sector has ceased and the investment made in it was unlikely to be recovered." "Although the company remains convinced that the rural/seasonal sector offers opportunities for the future, given the CLS experience the Group would handle immediate opportunities through its existing AWF branch network," he said." "AWF had lost confidence in CLS's commitment to company policies and procedures and was forced to sever the relationship." "AWF is New Zealand's largest specialist blue collar labour hire company." "It debuted on the stock market in 2005." "Lower electricity generation and supply costs were the main driver behind falls for output and input prices in the Producers Price Index (PPI) for the December quarter." "Drops of 0.5 per cent in output prices and 0.3 per cent in input prices were the first quarterly falls since the March 2004 quarter, Statistics New Zealand (SNZ) said today." "The PPI measures the average level of industrial input prices (excluding labour) and output prices at the farm and factory gate." "The electricity generation and supply outputs index fell 5.8 per cent due to lower

spot market prices and lower commercial electricity retail prices in the December 2006 quarter, the third consecutive fall for that index." "Another significant downward contributor to the PPI output index was the meat and meat product manufacturing index which fell 6.1 per cent due to lower export prices for lamb and beef." "The dairy product manufacturing outputs index was also down, dropping 4.6 per cent due to lower prices for cheese, SNZ said." "Rises in outputs prices included increases of 1 per cent in the construction index, 6.2 per cent in the horticulture and fruit growing index, and 3.5 per cent in the sheet and fabricated metal product manufacturing index." "For the PPI inputs index the main feature was an 11.4 per cent fall in the electricity generation and supply index due to higher lake levels and lower fuel costs." "Also significantly down was the gas supply inputs index, with a 6.9 per cent fall, and the air transportation index, down 3.9 per cent." "A 0.4 per cent rise in the construction index was the main upward contributor to input prices in the December quarter." "On an annual basis, the PPI outputs index rose 3.6 per cent in the year to December, while the inputs index rose 5.3 per cent." "The Green Party today called on Education Minister Steve Maharey to act immediately to protect children who attend schools that continue to impose corporal punishment, in defiance of the law." "MP Sue Bradford was commenting on reports that Wainuiomata Christian College continued to impose corporal punishment, with the approval of parents." "Ms Bradford said she wrote to Mr Maharey six months ago about corporal punishment policy at Auckland's Tyndale Park Christian School and other private schools." "She had been assured the ministry would take the matter up with the NZ Association of Christian Schools, followed by steps to clarify the law on this subject." "Nothing has happened, judging by the Wainuiomata Christian College case," Ms Bradford said." "This school – and others like it around the country – is apparently continuing to defy a law that was passed 20 years ago in New Zealand." "Just because there is parental approval does not make it alright." "The Education Review Office noted that the principal and the board of Wainuiomata Christian College refused to confirm or deny the use of corporal punishment." "Principal Martin Keast told The Dominion Post the prohibition on corporal punishment was a "rotten" law because it infringed parental rights." "The private school of about 70 pupils has fallen foul of the ERO over the issue." "Previous ERO reports said the school administered corporal punishment with the approval of parents." "Mr Keast said that as a Christian he believed the law was "contrary to scripture"." "He said corporal punishment was a private matter and the ERO had no authority to interfere." "Ms Bradford said the law banning corporal punishment in New Zealand schools existed to protect children." "A school is supposed to be a safe place, where children are able to learn, free from the threat of violence." "She urged the minister to act immediately to enforce the law." "A

third of Australians aged 18 to 24 are classified as binge drinkers, with nearly one-in-four drinking to the point of passing out on at least five occasions, a new study reveals." "The research, commissioned by the Alcohol Education & Rehabilitation Foundation (AER), also found one-in-three people in that age bracket who typically drank 10 or more drinks did not see themselves as binge drinkers." "The survey, of 500 men and women across Australia, found 44 per cent had drunk so much they passed out on at least one occasion and 22 per cent on five or more occasions." "Four per cent admitted to having passed out in excess of 20 times." "The survey also showed binge drinkers tended to cling together - finding they are more likely to befriend, date and consider marrying other excessive drinkers." "According to National Health and Medical Research Council guidelines, men who drink 11 or more drinks in one sitting are considered to be binge drinking, while seven drinks puts women in the danger zone." "The AER study also revealed binge drinkers were more likely to have a one night stand when they were drunk." "But at the same time, 59 per cent of respondents said they felt alcohol affected their sex life in a negative way." "An overwhelming 96 per cent said it also had a negative effect on their weight." "However, 85 per cent of 18- to 24-year-olds said they would seek help if they had a drinking problem." "AER now plans to work with the music, fashion and media industries, to kick-off a national campaign called Fresh Party." "To begin on Saturday, April 14, in Sydney, it will be a daytime event aimed at turning around the image of binge drinking from being considered a form of entertainment." "Binge drinking is a wide-spread community problem," AER director Cheryl Bart said." "We're targeting 18- to 24-year-olds because that's where we feel the greatest challenge and opportunity lies to begin shifting this cultural problem." "Following on from the Sydney event, will be an on-going calendar of Fresh events." "We'd take the Valley Ranch free and clear as a booby prize." "He reiterated his opposition to such funding, but expressed hope of a compromise." "The White House said minors haven't any right to abortion without the consent of their parents." "Ten of the nation's governors, meanwhile, called on the justices to reject efforts to limit abortions." "The Justice Department announced that the FBI has been given the authority to seize U.S. fugitives overseas without the permission of foreign governments." "The device was replaced." "Details of the talks, described by a Zairean official as "very delicate," weren't disclosed." "Hurricane Jerry threatened to combine with the highest tides of the year to swamp the Texas-Louisiana coast." "Thousands of residents of low-lying areas were ordered to evacuate as the storm headed north in the Gulf of Mexico with 80 mph winds." "We support all efforts to remove Victor Posner from control of Arby's Inc. and the Arby's system." " "I would say this is not bad news; this is a blip," he said." "All year, energy prices have skewed the producer price index, which

measures changes in the prices producers receive for goods." "Energy prices then plummeted through the summer, causing the index to decline for three consecutive months." "Overall, the index has climbed at a 5.1% compound annual rate since the start of the year, the Labor Department said." "Prices for capital equipment rose a hefty 1.1% in September, while prices for home electronic equipment fell 1.1%." "Food prices declined 0.6%, after climbing 0.3% in August." "In the year-earlier period, CityFed had net income of $485,000, but no per-share earnings." "CityFed's president and chief executive officer, John Atherton, said the loss stems from several factors." "Regulators also ordered CenTrust to stop buying back the preferred stock." "He said the thrift will try to get regulators to reverse the decision." "These aren't mature assets, but they have the potential to be so," said Mr. Shidler." "In the same month, the Office of Thrift Supervision ordered the institution to stop paying common stock dividends until its operations were on track." "CenTrust, which is Florida's largest thrift, holds one of the largest junk-bond portfolios of any thrift in the nation." "The legislation requires thrifts to divest themselves of junk bonds in the new, somber regulatory climate." "In the third quarter, for instance, CenTrust added $22.5 million to its general reserves." "Most analysts and thrift executives had expected a decision on the proposed transaction, which was announced in July, long before now." "CenTrust said the branch sale would also reduce the company's large amount of good will by about $180 million." "European programs usually target only their own local audience, and often only a small portion of that." "It was front-page news in Italy earlier this year when the fictional inspector was gunned down in the series." "They will be available in minimum denominations of $10,000." "When the dollar is in a free-fall, even central banks can't stop it." "Now, however, commercial channels are coming to most European countries, and at the same time, satellite and cable technology is spreading rapidly." "Bids must be received by 1 p.m. EDT Thursday at the Treasury or at Federal Reserve banks or branches." "In recent months, a string of cross-border mergers and joint ventures have reshaped the once-balkanized world of European arms manufacture." "Thomson feels the future of its defense business depends on building cooperation with other Europeans." "Mr. Watkins said the main reason for the production decline is shrinking output of light crude from mature, conventional fields in western Canada." "That ranks Canada as the fourth-largest source of imported crude, behind Saudi Arabia, Nigeria and Mexico." "But when something is inevitable, you learn to live with it," he said." "The system currently has a capacity of 1.55 million barrels a day." "Predicting the financial results of computer firms has been a tough job lately." "Take Microsoft Corp., the largest maker of personal computer software and generally considered an industry bellwether." "Microsoft's surprising strength is one example of the difficulty facing investors

looking for reassurances about the financial health of the computer firms." "This month, however, Businessland warned investors that results for its first quarter ended Sept. 30 hadn't met expectations." "While the earnings picture confuses, observers say the major forces expected to shape the industry in the coming year are clearer." "Meanwhile, competition between various operating systems, which control the basic functions of a computer, spells trouble for software firms generally." "On the other hand, the battle of the bus is expected to grow increasingly irrelevant." "A bus is the data highway within a computer." "Users don't care about the bus," said Daniel Benton, an analyst at Goldman, Sachs & Co." "The gap between winners and laggards will grow." "Personal-computer makers will continue to eat away at the business of more traditional computer firms." "The guys that make traditional hardware are really being obsoleted by microprocessor-based machines," said Mr. Benton." "But they will have to act quickly." "So are Indiana, Ohio and Michigan." "The gains, to be sure, are rather small." "At the same time, several states in the South and West have had their own population turnaround." "That share has remained at about 24% since 1970." "What has changed is that more of the young elderly are living with spouses rather than with other relatives, such as children." "The likelihood of living alone beyond the age of 75 has increased to 40% from 32%." "The goal was to learn about one of today's fastest-growing income groups, the upper-middle class." "Across the board, these consumers value quality, buy what they like rather than just what they need, and appreciate products that are distinctive." "Thirty-five percent attend religious services regularly; at the same time, 60% feel that in life one sometimes has to compromise one's principles." "The American Bankers Association says that women make up 47% of officials and managers in the top 50 banks, up from 33% in 1978." "The share of minorities in those positions has risen to 16% from 12% ..." "Mario Gabelli, for instance, holds cash positions well above 20% in several of his funds." "This small Dallas suburb's got trouble." "And the mayor, in an admonition that bears a rhythmic resemblance to Prof. Hill's, warned that "alcohol leads to betting, which leads to fights." " "The council's action is yet another blow to a sport that its fans claim has been maligned unjustly for years." "At the lounge, manager Elizabeth Dyer won't admit patrons in jeans, T-shirts or tennis shoes." "I thought this was all taken care of in 'The Music Man." "For instance: "Haole" (white) is not the ultimate insult; "Mainland haole" is." "And the local expression for brother is "brah," not "bruddah." " "Of all the ethnic tensions in America, which is the most troublesome right now¿' "In addition, some of Mr. Mason's critics have implied that his type of ethnic humor is itself a form of racism." "For example, the New York state counsel for the NAACP said that Mr. Mason is "like a dinosaur." "People are fast leaving the place where he is stuck." " "But wielded by a pro like Jackie

Mason, it is a constructive form of mischief." "Charles J. Lawson Jr., 68, who had been acting chief executive since June 14, will continue as chairman." "Lawmakers drastically streamlined the bill to blunt criticism that it was bloated with special-interest tax breaks and spending increases." "In addition, the companion deficit-reduction bill already passed by the House includes a capital-gains provision." "The Senate bill was pared back in an attempt to speed deficit-reduction through Congress." "A key is whether House Republicans are willing to acquiesce to their Senate colleagues' decision to drop many pet provisions." "These include a child-care initiative and extensions of soon-to-expire tax breaks for low-income housing and research-and-development expenditures." "Sen. Dole said that the move required sacrifice by every senator." "Many fund managers argue that now's the time to buy." "It also drops a provision that would have permitted corporations to use excess pension funds to pay health benefits for current retirees." "The approval of the Senate bill was especially sweet for Sen. Mitchell, who had proposed the streamlining." "The deficit reduction bill contains $5.3 billion in tax increases in fiscal 1990, and $26 billion over five years." "Vincent Bajakian, manager of the $1.8 billion Wellington Fund, added to his positions in Bristol-Myers Squibb, Woolworth and Dun & Bradstreet Friday." "– Curb junk bonds by ending tax benefits for certain securities, such as zero-coupon bonds, that postpone cash interest payments." "Withhold income taxes from the paychecks of certain farm workers currently exempt from withholding." "Change the collection of gasoline excise taxes to weekly from semimonthly, effective next year." "Reduction of Medicare spending in fiscal 1990 by some $2.8 billion, in part by curbing increases in reimbursements to physicians." "Are you kidding¿' "Columbia Savings is a major holder of so-called junk bonds." "New federal legislation requires that all thrifts divest themselves of such speculative securities over a period of years." "They perhaps had concern that we were getting out of all these," said Franklin President Duane H. Hall." "I think it was a little premature on their part." "Well, in some ways it is different, but technically it is just the same." "If you're a technician, you obey the signals." "I see a possibility of going to 2200 this month." " "This was an October massacre" like those that occurred in 1978 and 1979." "Now, as in those two years, her stock market indicators are positive." "She says that ratio could climb to 14.5, given current interest rates, and still be within the range of "fair value." " "Trading volume was only modestly higher than normal." "But Mr. Davis, whose views are widely respected by money managers, says he expects no 1987-style crash." "I think the market will be heading down into November," he says." "Mr. Cooperman sees this as a good time to pick up bargains, but he doesn't think there's any need to rush." "Unlike 1987, interest rates have been falling this year." "So it's a very mixed bag." " "We're going to look for some of the better-known companies that got clocked" Friday." "I see this

as a reaction to the whole junk bond explosion," he says." "There is more resiliency in the economy at large than we commonly suppose," he says." "Mr. Rogers won't decide what to do today until he sees how the London and Tokyo markets go." "T-bills probably are the right place to be," he says." "I thought your editorial was factually accurate and deliberately elucidative." "The Baltimore-based group noted that some investors moved money from stock funds to money-market funds." "He used about 56 words defending the witnesses' constitutional rights." "Unfortunately, by my rough guess, he used better than 5,000 words heaping scorn on the witnesses for exercising the Fifth." "He sandwiched his praise of constitutional meat between large loaves of bilious commentary." "That certainly is not the supposed "distorted reading" indicated by Mr. Lantos." "But most investors seemed to be "in an information mode rather than in a transaction mode," said Steven Norwitz, a vice president." "Right now they're pursuing evidence." "He's right about his subcommittee's responsibilities when it comes to obtaining information from prior HUD officials." "it defended appropriate constitutional safeguards and practical common sense." "In an unusual move, several funds moved to calm investors with recordings on their toll-free phone lines." "Absent the risk of such prosecution, a court may order the defendant to testify." "At the end of the day, 251.2 million shares were traded." "In academia, a so-called Friday the 13th effect has been set up and shot down by different professors." "Robert Kolb and Ricardo Rodriguez, professors of finance at the University of Miami, found evidence that the market is spooked by Friday the 13th." "In the '70s, the market took falls nine times in a row on Friday the you-know-what." "It was like the Friday before Black Monday" two years ago." "Some early selling is likely to stem from investors and portfolio managers who want to lock in this year's fat profits." "Ten points of the drop occurred during the last 45 minutes of trading." "Most of the complaints about unanswered phone calls came from regional brokers rather than individual investors." "Stock funds have averaged a staggering gain of 25% through September, according to Lipper Analytical Services Inc." "It wasn't intentional, we were all busy." " "On days like Friday, that means they must buy shares from sellers when no one else is willing to." "On Friday, some market makers were selling again, traders said." "Everyone was hitting everyone else's bid," he said." "I don't know of anyone carrying big inventories now," said Mr. King of Robinson-Humphrey." "But they are worried." "This is not a major crash," she said." "It won't take much more to "scare the hell out of retail investors," he says." "Institutional investors, which had been selling stock throughout last week to lock in handsome gains made through the third quarter, were calmer." "Nevertheless, Ms. Garzarelli said she was swamped with phone calls over the weekend from nervous shareholders." "The turnover in both issues was roughly normal." "The key U.S. and

foreign annual interest rates below are a guide to general levels but don't always represent actual transactions." "COMMERCIAL PAPER placed directly by General Motors Acceptance Corp.:" "The backdrop to Friday's slide was markedly different from that of the October 1987 crash, fund managers argue." "I waited to make sure all the program trades had kicked through," he said." "Stocks, as measured by the Standard & Poor's 500-stock index, have been stellar performers this year, rising 27.97% before Friday's plunge, excluding dividends." "It can happen before you can turn around." " "Even without portfolio insurance, market conditions were grim Friday, money managers said." "Mr. Weisman predicts stocks will appear to stabilize in the next few days before declining again, trapping more investors." "The shorts sell borrowed shares, hoping to profit by replacing them later at a lower price." "We're not making a killing, but we had a good day." "Neither Vitarine nor any of the Springfield Gardens, N.Y., company's officials or employees have been charged with any crimes." "But so far the company hasn't complied with that request, the spokesman said." "The West German retailer ASKO Deutsche Kaufhaus AG plans to challenge the legality of a widely employed anti-takeover defense of companies in the Netherlands." "It was previously thought ASKO held a 13.6% stake that was accumulated since July." "A spokesman for Ahold said his company is confident of its own position and the propriety of the preferred-share issue." "He termed ASKO's legal actions as "unproductive" to international cooperation among European retailers." "The cuts are necessary because Congress and the administration have failed to reach agreement on a deficit-cutting bill." "– For the last two weeks, the Bush administration and the Federal Reserve have been engaged in a semi-public battle over international economic policy." "Last month, Transportation Secretary Sam Skinner forced Northwest Airlines to reduce a stake held by KLM Royal Dutch Airlines." "But he has since run into opposition from the Treasury and the White House over that decision." "Nevertheless, the company's reaction underscores the domino effect that a huge manufacturer such as Boeing can have on other parts of the economy." "No one in Washington was willing to take the blame for provoking Friday's drop in the stock market." "Three million shares of $25 preferred, via competitive bidding." "I don't think their customers would like it very much." " "But he's not so sure about everyone else." "Indeed, a random check Friday didn't seem to indicate that the strike was having much of an effect on other airline operations." "We hope to take advantage of it," said John Snyder, a member of a Los Angeles investors' club." "There's no question that there's a general distaste for leverage among lenders." "Sara Albert, a 34-year-old Dallas law student, says she's generally skittish about the stock market and the takeover activity that seems to fuel it." "I have this feeling that it's built on sand," she says, that the market rises "but there's no foundation to it." " "She

and her husband pulled most of their investments out of the market after the 1987 crash, although she still owns some Texaco stock." "It's so close to completion, Boeing's told us there won't be a problem," said a Southwest spokesman." "Others wonder how many more of these shocks the small investor can stand." "We all assumed October '87 was a one-time shot," said San Francisco attorney David Greenberg." "Still, he adds: "We can't have this kind of thing happen very often." "Merrill Lynch can't survive without the little guy." " "Small investors have tiptoed back into the market following Black Monday, but mostly through mutual funds." "Individual investors are still angry about program trading, Mr. Quackenbush says." "But it's not only the stock market that has some small investors worried." "I got out in 1987." "Would Mr. Antori ever get back in¿' "The crowded field for notebook-sized computers is about to become a lot more crowded." "Compaq's series of notebooks extends a trend toward downsizing in the personal computer market." "One manufacturer already has produced a clipboard-sized computer called a notepad, and two others have introduced even smaller "palmtops." " "At 4 1/2 pounds, it may be too ambitiously named, but it nevertheless opens up the kind of marketing possibilities that make analysts froth." "Laptops – generally anything under 15 pounds – have become the fastest-growing personal computer segment, with sales doubling this year." "Responding to that demand, however, has led to a variety of compromises." "It also has precluded use of the faster, more powerful microprocessors found in increasing numbers of desktop machines." "The competitive sniping can get pretty petty at times." "Toward that end, experts say the real battle will take place between center-stage players like Toshiba, Zenith and now Compaq." "No date has yet been set to get back to the bargaining table." "The problem Compaq is going to have is that they won't be able to make enough of them." " "Grumman Corp. received an $18.1 million Navy contract to upgrade aircraft electronics." "Of course, many more issues – 93 – hit new lows." "In October 1987, these margin calls were thought to have contributed to the downward spiral of the stock market." "Margin calls since Friday "have been higher than usual, but reasonable," a spokesman for Shearson Lehman Hutton Inc. said." "He said Schwab had increased margin requirements "so customers have more of a cushion." " "Industry estimates put Avis's annual cost of all five programs at between $8 million and 14 million." "Analysts and competitors, however, doubt the numbers were that high." "They've been looking to get their costs down, and this is a fairly sensible way to do it," he said." "CBS Inc. is cutting "The Pat Sajak Show" down to one hour from its current 90 minutes." "I wouldn't expect an immediate resolution to anything." " "Tandem said it expects to report revenue of about $450 million and earnings of 35 cents to 40 cents a share." "Obviously IBM can give bigger discounts to users immediately," said Mr. Weiss." "Don't jump yet."

“When it went down, by all tradition, the economy followed.” “Of course, the health of the economy will be threatened if the market continues to dive this week.” “Growth is slower.” “Profits are softer.” “The union is continuing to work through its expired contract, however.” “In the third quarter of 1987, the economy spurted at an inflation-adjusted annual rate of 5.3%.” “The effects were much less severe and less prolonged than some had feared or expected.” “A Dow spokeswoman declined to comment on the estimates.” “Still, some industry giants are expected to report continuing gains, largely because so much of their business is outside commodity chemicals.” “Du Pont Co. is thought to have had steady profit growth in white pigments, fibers and polymers.” “Most estimates for Monsanto run between $1.70 and $2 a share.” “By some accounts on Wall Street and in the industry, the inventory reductions are near an end, which may presage firmer demand.” “In the 1988 third quarter, Quantum earned $99.8 million, or $3.92 a share, on sales of $724.4 million.” “Rated Baa-1 by Moody’s and triple-B-plus by S&P, the issue will be sold through underwriters led by Morgan Stanley & Co.” “Bridget O’Brian contributed to this article.” “$500 million of Remic mortgage securities offered in 13 classes by Prudential-Bache Securities Inc.” “The principal-only securities will be repackaged by BT Securities into a Freddie Mac Remic, Series 103, that will have six classes.” “Our long suit is our proven ability to operate” power plants, he said.” “The company said the improvement is related to additional cogeneration facilities that have been put into operation.” “On the exchange floor, “as soon as UAL stopped trading, we braced for a panic,” said one top floor trader.” “When the price of plastics took off in 1987, Quantum Chemical Corp. went along for the ride.” “For weeks, the market had been nervous about takeovers, after Campeau Corp.’s cash crunch spurred concern about the prospects for future highly leveraged takeovers.” “And the financial decline of some looks steep only in comparison with the heady period that is just behind them.” “We were all wonderful heroes last year,” says an executive at one of Quantum’s competitors.” “At Quantum, which is based in New York, the trouble is magnified by the company’s heavy dependence on plastics.” “Quantum’s lot is mostly tied to polyethylene resin, used to make garbage bags, milk jugs, housewares, toys and meat packaging, among other items.” “Benchmark grades, which still sold for as much as 50 cents a pound last spring, have skidded to between 35 cents and 40 cents.” “Meanwhile, the price of ethylene, the chemical building block of polyethylene, hasn’t dropped nearly so fast.” “By many accounts, an early hint of a price rout in the making came at the start of this year.” “People were even hoarding bags,” he says.” “One doubter is George Krug, a chemical-industry analyst at Oppenheimer & Co. and a bear on plastics stocks.” “Some say November.” “Some analysts saw the payment as an effort also to dispel takeover speculation.” “Some viewed his response

– that company directors review the dividend regularly – as nothing more than the standard line from executives." "Until this year, the company had been steadily lowering its accident rate and picking up trade-group safety awards." "The plant usually accounts for 20% to 25% of Quantum's polyethylene production and 50% of its ethylene production." "Not everything looks grim for Quantum." "These stocks eventually reopened." "Petrolane is the second-largest propane distributor in the U.S." "The spokesman said the broadcast unit will be disbanded Dec. 1, and the move won't affect RJR's print, radio and spot-television buying practices." "Other countries that don't have formal steel quotas with the U.S., such as Taiwan, Sweden and Argentina, also have supplied steel." "That increase rises to slightly more than 2% of the U.S. market if a joint Korean-U.S. steel project is included." "Meanwhile, Brazil is expected to increase its allowance from the 1.43% share it has had in recent years." "Japan has been shipping steel to total about 4.5% of the U.S. market compared with a quota of 5.9%." "Most of the stock selling pressure came from Wall Street professionals, including computer-guided program traders." "The balance is supplied by a host of smaller exporters, such as Australia and Venezuela." "Traders said most of their major institutional investors, on the other hand, sat tight." "Mr. Pierce said Elcotel should realize a minimum of $10 of recurring net earnings for each machine each month." "A P&G spokeswoman confirmed that shipments to Phoenix started late last month." "She said the company will study results from this market before expanding to others." "And retailers are expected to embrace the product, in part because it will take up less shelf space." "If the new Cheer sells well, the trend toward smaller packaging is likely to accelerate as competitors follow with their own superconcentrates." "Others said the Bush administration may feel the rhetoric on both sides is getting out of hand." "And some said it reflected the growing debate in Washington over pursuing free trade with Japan versus some kind of managed trade." "I am painted sometimes as ferocious, perhaps because I have a ferocious list of statutes to implement." "I don't feel either hard or soft." "She said the trade imbalance was mainly due to macroeconomic factors and shouldn't be tackled by setting quantitative targets." "At her news conference for Japanese reporters, one economics journalist summed up the Japanese sense of relief." "But she stressed, "I am against managed trade." "In the year ended June 30, 1988, Traditional reported net income of $4.9 million, or $1.21 a share." "In the latest nine months net income was $4.7 million, or $1.31 a share, on revenue of $44.3 million." "Separately, the company said it would file a delayed fiscal-year report with the Securities and Exchange Commission "within approximately 45 days." " "Information International said it believes that the complaints, filed in federal court in Georgia, are without merit." "Closely held Morris Communications is based in Augusta, Ga." "The units that filed

the suit are Southeastern Newspapers Corp. and Florida Publishing Co." "A spokeswoman said Sulka operates a total of seven stores in the U.S. and overseas." "Syms operates 25 off-price apparel stores in the U.S." "Typical is what happened to the price of ethylene, a major commodity chemical produced in vast amounts by many oil companies." "But stocks kept falling." "The main reason remains weather." "This summer, on the other hand, had milder weather than usual." "The seller of photographic products and services said it is considering a number of financing alternatives, including seeking increases in its credit lines." "Sanford Sigoloff, chief executive of L.J. Hooker, said yesterday in a statement that he has not yet seen the bid but that he would review it and bring it to the attention of the creditors committee." "Products like Flash and Dreamweaver - and other brands - played a leading role in enabling people to make better experiences on the Internet." "So Macromedia played a leading role in CD-ROMs and the Internet." "Everybody is very excited about it." "Video's a great business for them, as is PostScript." "Photoshop's a great business for Adobe - we don't have anything that competes with that." "And that distribution now is not as important as it was." "How did you form that vision¿' "So we think everybody's going to get involved." "The other thing I would say is that our long-term competitive advantage is cross-platform." "And that happened for us with Flash." "There's no cost associated with [distributing the Flash Player] in the PC world." "The content ecosystem that is developing in Japan around Flash is absolutely phenomenal." "These are assets that make us very excited about the future." "When I look at where the companies are going, there's very little overlap." "Adobe's been a document-centric company, and it's done very, very well." "These are applications that benefit by operating on multiple platforms." "And then over the last several years we have been investing very heavily to broaden the "platformness" of it." "We've added Flex to that, which is now for programmers to bring Flash to classic enterprise applications." "So now, on the trains, it used to be that people would be banging away on e-mail." "We have 90-plus market share in the case of Dreamweaver and so it was really broadening our agenda." "So, the major issue for us is not competition as much as it is effectively executing on the market opportunity." "Because we own the Flash Player and that is so important to the ultimate end-user experience - that's a long-term competitive advantage." "There are several thousand web sites now supplying Flash content, and it's simply because, as human beings, we're multi-sensory creatures." "We've actually had 60 or 70 people working with Macromedia for a while now in India - although they weren't employees, they were [contracted] through an outsourcing company." "To begin with, we have the full expectation that many, many companies will get involved in much richer experiences." "You're seeing all kinds of examples now of incredibly better experiences - and a lot of them are

enabled by our technology." "You're going to see communications appliances that are entirely different, and capabilities that are entirely different." "Although, interestingly, there's a lot of similarity in business-model." "You also have FlashPaper now, which produces both a [Flash] SWF-based document format and a PDF-format." "All of a sudden you could do pictures and sounds and words on computers, and mix them together." "All Betsey and I did was to create an environment where they could actually bet on the vision that they had - and that was about the web." "Looking at it from where we are now, it seems clear that the web was the way to go, but it was not that obvious to everybody back then." "But I would think that if we ended up in a few years with equal businesses there - getting there the right way, not the wrong way - [laughs] that would be good." "Everybody is alarmed about the banking-system, but the Chinese have spent a tremendous amount of energy in the last two to three years resolving this problem and I think they are going to succeed," said Shafer, pointing out that the government has put-money into the banks to reduce the level of non-performing-loans." "Jeff Shafer, vice-chairman of Citigroup? public-sector business, discussed the Chinese banking-industry, which traditionally has been burdened by bad loans." "All of that rolls into what I mean when I say interest-rates." "If that happens, we'll have a long way to go to get out of this." "I only have this, for the future." "Some areas will grow faster, but by and large we expect economic-expansion across all businesses." "If you go hostile," you may not be able to conduct much due-diligence, "depending on how the target reacts." " "The Sarbanes-Oxley legislation, which makes corporate officials personally liable for fiduciary-duties, has also made independent board members more cautious about dismissing an unwanted bid." "I made it clear when I was appointed that I took this to be a particularly important area of MRC work to be strengthened, he told the BMJ." "It was acceptable to drive when not tired." "This double payment makes scientific publishing a highly lucrative business, worth $7bn a year." "It is hard for REITs to cook their books when they pay out 90% or more of their cash flow as dividends." "But, in fact, his days as head of Pepsi-Cola were numbered." "she assumed that the phenomenon she eats is limited to." "It is not even uniquely Western." "This is the main category of medication used in the treatment of depression in children and adolescents, and the announcement will have taken many young-people who take these drugs, their parents, and doctors by surprise." "Although the advice only applies to the United-Kingdom, it mirrors concerns that are also being considered by the US-Food-and-Drug-Administration." "It costs approximately 9bn ($16bn; 13bn) in England each year, and worldwide is the fourth most important cause of disability." "About half of the estimated 40 000 young-people under the age of 18 years using antidepressants in the United-Kingdom are currently taking one of the

newly “contraindicated“ antidepressant-medications.” “A more recent Cochrane systematic-review showed that they may offer some benefit for adolescents with depression but not for pre-pubertal children.” “The dramatic issuing of the guidance by the Committee-on-Safety-of-Medicines is likely to lead to considerable uncertainty and some difficulty for many patients and doctors.” “The image of the U.S. would improve if Obama wins because his rhetoric about people fighting for global causes extends beyond the U.S.” “Wages rose more in 2006 than during the two preceding years.” “Similarly, sales growth improved as insiders financial stakes grew, and worsened as they gained voting clout.” “They have offices in more than 90 countries, which are grouped into six corporate regions.” “Microsoft’s Xbox-360 console includes certain key components that are supplied by a single-source.” “High investment in R&D paid off in the form of new ground-breaking products and technologies that are helping redefine the next generation of information-technology.” “Nixon-era commerce secretary Peter-Peterson has lost count of the corporate-scandals he has seen during his long career in public and private organizations.” “When there are no more buyers, there are sellers,“ Zell reasoned.” “He gave three reasons why publicly-held REITs have escaped the recent wave of scandals.” “Peterson, who is now chairman and co-founder of The-Blackstone-Group, an investment and advisory services firm in New-York-City, finds that little has changed.” “Unlike the boom of the 1980s, whose aftermath revealed a host of shady deals between real-estate-developers and unscrupulous S&L executives, publicly-traded REITs have largely been untouched by the most recent scandals.” “(REITs are exempt from federal tax so long as they distribute at least 90% of taxable-income to investors each year.)“ “REIT stocks were going up 25-cents, 50-cents and 10 cents a day (during their climb), but when they came down, they were coming down $2 and $3 a day.” “But if and when private capital flows begin to ebb, “the longer term trend is going to be very positive,“ he said.” “Saltzman finds “the most prolific deal-flow“ in Asian-countries such as China, and the least amount of competition.” “If you look at the balance-sheets of the banks in China today, 40% of their loans have issues of one sort or another.”“ “Some real-estate leaders have passed the baton to the next generation to good effect“ “To be frank, the development side of the real-estate business is not where it used to be, because the demand isn’t there,“ he said.” “To be frank, the development side of the real-estate business eats where it used to be, because the demand isn’t there,“ he said.” “To be frank, the development side is not where it used to be, because the demand isn’t there.” “Compared with entrepreneurial firms, where leadership passes on to a family member, you are going to see a challenge in the transfer of institutional leadership to newer, younger talent.” “all that was not enough to divert attention from the one issue that kept coming back to the discussion table.” “There was so much hype in the

late nineties, and clearly with the bursting of the tech-bubble, the general population thinks it got conned," he said." "While most of its researchers are based at Microsoft's Redmond headquarters, Microsoft-Research has expanded globally to ensure that it can attract the richest pool of talent." "A U.S. federal jury found that Microsoft-Corp infringed on audio patents held by Alcatel-Lucent and should pay $1.52 billion in damages (Feb 2007)." "Among the most recent product areas benefiting from Microsoft Research's technology-transfer process in action is Microsoft's SmartScreen Technology, which is at the core of powerful anti-Spam filters within such products as Microsoft-Office-Outlook 2003, Microsoft-Exchange-Server 2003, MSN 8 and Hotmail." "please contact PARS International: reprints@parsintl.com P. (212) 221-9595 x07." "Zell acknowledged he does not always get it right." "Everybody else said, 'Sam, you don't understand.'" "It has a different risk-premium on it, but the actual amount of liquidity has not changed." "I would argue the excess liquidity that existed eight weeks ago still exists today." "They bought anything they wanted and were proud that they didn't do due-diligence." "The idea was built it and somebody will buy it, and that somebody was the Japanese," Zell recalled." "REIT stocks are back these days to the 87% range after the recent correction, Saltzman noted." "The flow of funds into REITs in the first quarter of 2004 was greater than that in all of 2003," he said.")

A.5 Post UGE - Process Rule1

Process-Rule1 (MSs):

```
((rule1p MSs)
    (when *debug-interp* (print 'rule1))
    (let ((tmpresult nil))


      ;; Action with *speech-words* and MS1
      ;; has more value than others
      (setf tmpresult (remove-duplicate-mss
                       (remove-if-not #'(lambda (x)
                       (and (member
                       (lmword (MS-result x))
                       *speech-words*)
                      (get-predicate x :pred ':ms1))
                      )
                      MSs)))

      (when (and tmpresult
             (equal (length tmpresult) 1))
        (return-from uge-select-bestMS-aux
        (car tmpresult))
        )

      ;; Set MSs to tmpresult if length is < MSs
      (when (and tmpresult (< (length tmpresult)
                              (length MSs)))
        (setf MSs tmpresult)
       )


      ;; Action with MS1 get high priority
      ;; ex: "every firm saw costs grow more than expected,
      ;; even after adjusting for inflation."
      (setf tmpresult (remove-duplicate-mss
                       (remove-if-not #'(lambda (x)
                       (get-predicate x :pred ':MS1))
                       MSs)))

      (when (and tmpresult
             (equal (length tmpresult) 1))
        (return-from uge-select-bestMS-aux
        (car tmpresult))
        )

      ;; Set MSs to tmpresult if length is < MSs
      (when (and tmpresult (< (length tmpresult)
                              (length MSs)))
        (setf MSs tmpresult)
       )
```

```
;; Action with *speech-words* has more value than others
(setf tmpresult (remove-duplicate-mss
                 (remove-if-not #'(lambda (x)
                  (member (lmword (MS-result x))
                  *speech-words*)
                    )
                    MSs)))

(when (and tmpresult
       (equal (length tmpresult) 1))
  (return-from uge-select-bestMS-aux
  (car tmpresult))
  )

;; Set MSs to tmpresult if length is < MSs
(when (and tmpresult (< (length tmpresult)
                        (length MSs)))
  (setf MSs tmpresult)
  )


;; Action with *say-that* has more priority
(setf tmpresult (remove-duplicate-mss
                 (remove-if-not #'(lambda (x)
                 (and
                 (member (lmword
                 MS-reslt x))
                 '(SAYS-THAT* SAY-THAT*
                 CONFIRMED-THAT*
                 CONFIRM-THAT*
                 CONFIRMS-THAT*
                 SHOW-THAT*))
                 (get-predicate x :pred ':ms1))
                 )
                 MSs)))

(when (and tmpresult
           (equal (length tmpresult) 1))
  (return-from uge-select-bestMS-aux
      (car tmpresult))
  )

;; Set MSs to tmpresult if length is < MSs
(when (and tmpresult (< (length tmpresult)
                        (length MSs)))
  (setf MSs tmpresult)
  )


;; Action with time attachement got more value
(setf tmpresult (remove-duplicate-mss
                 (remove-if-not #'(lambda (x)
                 (member
                 (car
                 (get-predicate x
                 :last-entry t)) '(:TIME))
                 )
                 MSs)))

(when (and tmpresult
       (equal (length tmpresult) 1))
  (return-from uge-select-bestMS-aux
```

```
  (car tmpresult))
  )

;; Set MSs to tmpresult if length is < MSs
(when (and tmpresult (< (length tmpresult)
                        (length MSs)))
  (setf MSs tmpresult)
  )


;; Action with :recipient attachment
;; has more priority
;; ex: "Mrs Neels showed Berry photographs of Liz."
(setf tmpresult (remove-duplicate-mss
                 (remove-if-not #'
                 (lambda (x)
(get-predicate x :pred ':RECIPIENT)
)
MSs)))

(when (and tmpresult
           (equal (length tmpresult) 1))
  (return-from uge-select-bestMS-aux
  (car tmpresult))
  )

;; Set MSs to tmpresult if length is < MSs
(when (and tmpresult (< (length tmpresult)
                        (length MSs)))
  (setf MSs tmpresult)
  )

;; Action with :recipient within inf-to
;; attachment has more priority
;; ex: "I want to tell you the truth."
(setf tmpresult (remove-duplicate-mss
(remove-if-not #'(lambda (x)
(and (get-predicate x :pred ':to**)
(listp (get-predicate x :pred ':to**))
(listp (second (get-predicate x :pred ':to**)))
(get-predicate (second
(get-predicate x :pred ':to**)) :pred ':recipient)
)
)
MSs)))

(when (and tmpresult
           (equal (length tmpresult) 1))
  (return-from uge-select-bestMS-aux
  (car tmpresult))
  )

;; Set MSs to tmpresult if length is < MSs
(when (and tmpresult (< (length tmpresult)
                        (length MSs)))
  (setf MSs tmpresult)
  )


;; Action with :ms1 attachement is comma then
;; it has less priority.
(setf tmpresult (remove-duplicate-mss
```

```
                (remove-if #'(lambda (x)
                (and (get-predicate x :pred ':MS1)
(member
(car (second (get-predicate x :pred ':MS1)))
'(comma*))
)
)
MSs)))

(when (and tmpresult
           (equal (length tmpresult) 1))
  (return-from uge-select-bestMS-aux
  (car tmpresult))
  )

;; Set MSs to tmpresult if length is < MSs
(when (and tmpresult (< (length tmpresult)
                      (length MSs)))
  (setf MSs tmpresult)
  )

;; Action with :ms1 attachement is comma or
;; for then it has less priority.
(setf tmpresult (remove-duplicate-mss
(remove-if #'(lambda (x)
(and (get-predicate x :pred ':MS1)
(member
(car (second (get-predicate x :pred ':MS1)))
'(comma* for*))
)
)
MSs)))

(when (and tmpresult
           (equal (length tmpresult) 1))
  (return-from uge-select-bestMS-aux
  (car tmpresult))
  )

;; Set MSs to tmpresult if length is < MSs
(when (and tmpresult (< (length tmpresult)
                          (length MSs)))
  (setf MSs tmpresult)
  )

;; Action with :what as last attachment over
;; :manner attachment got high priority
(setf tmpresult (remove-duplicate-mss
(remove-if-not #'(lambda (x)
(get-predicate x :pred ':what :last-entry t)
)
MSs)))

(when (and tmpresult
           (equal (length tmpresult) 1))
  (return-from uge-select-bestMS-aux
  (car tmpresult))
  )

;; Set MSs to tmpresult if length is < MSs
(when (and tmpresult (< (length tmpresult)
                          (length MSs)))
```

```
  (setf MSs tmpresult)
  )

;; Some verbs has higher priority than others.
(setf tmpresult (remove-duplicate-mss
(remove-if #'(lambda (x)
(member (lmword (ms-result x)) '(held*))
)
MSs)))

(when (and tmpresult
           (equal (length tmpresult) 1))
  (return-from uge-select-bestMS-aux
  (car tmpresult))
  )

;; Action with :PHRASE attachement has less priority.
(setf tmpresult (remove-duplicate-mss
(remove-if #'(lambda (x)
(get-predicate x :pred ':PHRASE)
)
MSs)))

(when (and tmpresult
           (equal (length tmpresult) 1))
  (return-from uge-select-bestMS-aux
  (car tmpresult))
  )

;; :MODIFIER attached to MS as last-entry
;; got less priority
(setf tmpresult (remove-duplicate-mss
(remove-if #'(lambda (x)
(get-predicate x :pred ':MODIFIER :last-entry t)
)
MSs)))

(when (and tmpresult
           (equal (length tmpresult) 1))
  (return-from uge-select-bestMS-aux
  (car tmpresult))
  )

;; prep attached to MS as last attachment
;; got less priority
;; If ms has :what attachment.
(setf tmpresult (remove-duplicate-mss
(remove-if #'(lambda (x)
(and (get-predicate x :pred ':what)
(get-predicate x :last-entry t)
(member (car
(get-predicate x :last-entry t)) *prep*)
)
)
MSs)))

(when (and tmpresult
           (equal (length tmpresult) 1))
  (return-from uge-select-bestMS-aux
  (car tmpresult))
  )
```

```
  ;; Finally
  (return-from uge-select-bestMS-aux (car tmpresult))

  )
)
```

A.6 The news article

(source: New Zealand Herald)

Investors are lining up behind the Accident Compensation Corporation's call for a rethink of the $8 billion merger of Contact Energy and Australian rival Origin Energy.

And at least one fund manager has pledged financial support for the ACC, which will ask the High Court to raise the number of shareholder votes required to approve the controversial plan.

BT Funds Management's Paul Richardson said the ACC's question was one among a number of matters that needed to be clarified.

"I would certainly be happy [to contribute] if necessary," Richardson said. "When these issues come up we can contribute, they are part of operating a funds business."

The merger of the two energy giants is to be brought about by two shareholder votes on the proposed scheme. The first vote will be on a special resolution, requiring a 75 per cent majority of all shareholders.

The second, an ordinary resolution, requires the support of at least half of the shareholders other than Origin, which already owns 51 per cent of the New Zealand company.

However, the ACC, New Zealand's mandatory workplace insurer, believes the proposal alters the rights attached to Origin shares in quite a different way to the way it affects minority shareholders.

As a result, the second vote should be a special resolution, requiring 75 per cent of minority shareholders to vote in favour.

The ACC will if necessary ask the High Court to support its view and has engaged barrister Bill Wilson, QC, and Gibson Sheat lawyer Nigel Moody.

Under the terms of the plan, Contact investors effectively swap their holdings for a stake in ContactOrigin, formed by a contract between the two companies. Contact and Origin shares will continue trading on the New Zealand and Australian exchanges.

Contact will be seeking orders from the High Court, proposing two shareholder resolutions to approve the merger proposal. Details are due to be delivered to shareholders early next month before a vote in August.

The ACC has told other fund managers that the broad principle of its objection was that support of more than half Contact's minority investors should be needed to force all minorities to accept a fundamental change in its investment.

Tyndall Investment Management's Rickey Ward said he had already indicated its support for the ACC, but was not considering financial support.

"I would expect there would be a lot of support for the move," Ward said.

Contact's independent directors Phil Pryke, John Milne and Tim Saunders have already said the merger is in the interests of shareholders.

Contact has told the Business Herald the proposed arrangements were consistent with the NZX Listing Rules for related-party transactions and the Takeovers Code.

"Contact Energy considers these thresholds meet statutory and regulatory requirements, and are consistent with those required for previous schemes of arrangement," said the company's general counsel, Ross O'Neill.

The merger would create the largest integrated energy company in Australia and New Zealand, with 2.6 million customers and sales of A$5.5 billion ($6.6 billion).

Sydney-based Origin says the plan would help alleviate Contact's dwindling gas supplies and rising fuel costs. But sceptical shareholders say there's nothing to stop Origin lifting exploration under the current structure.

Contact's shares have more than doubled since the company went public in May 1999. The shares closed up 7c yesterday at $7.42.

A.7 UGE vs Stanford parser

```
ARTICLE 1

Sentence 1:

"If Ahmed Zaoui's family succeed in their application to join him in New Zealand as refugees,
the decision will annoy NZ First leader Winston Peters but delight a boy who has not seen
his father for nearly three years."


UGE output:
(IF* (:MS1
      (SUCCEED* (:ACTOR (FAMILY* (:NOUN) (:MODIFIER (|AHMED-ZAOUI'S*|))))
       (:IN* (APPLICATION* (:NOUN) (:MODIFIER (THEIR*))))
                (:TO** (JOIN* (:WHAT (HIM* (:PNOUN) (:IN* (NEW-ZEALAND* (:UNKNOWN) (:NAME)
                                                                          (:AS* (REFUGEES* (:NOUN)))))))))))
     (:MS2
      (BUT*
       (:MS2
        (DELIGHT* (:ACTOR ?L)
         (:WHAT
          (BOY* (:NOUN) (:MODIFIER (A*))
           (:WHO*
            (:MS1
             (HAS* (:ACTOR ?L)
              (:WHAT
               (SEEN*
                (:WHAT
                 (FATHER* (:NOUN) (:MODIFIER (HIS*)) (:FOR* (YEARS* (:NOUN) (:MODIFIER (THREE*))))
                  (:MANNER (NEARLY*))))))
              (:MANNER (NOT*)))))))))
       (:MS1
        (WILL* (:ACTOR (DECISION* (:NOUN) (:MODIFIER (THE*))))
         (:MS1
          (ANNOY* (:ACTOR ?L)
                  (:WHAT (WINSTON-PETERS* (:X-WORDS (*NZ*-FIRST*) (LEADER*) (WINSTON-PETERS*)) (:NAME)))))))))


Stanford output:
(ROOT
  (S
    (SBAR (IN If)
      (S
        (NP
          (NP (NNP Ahmed) (NNP Zaoui) (POS 's))
          (NN family))
        (VP (VBP succeed)
          (PP (IN in)
            (NP (PRP$ their) (NN application)
              (S
                (VP (TO to)
                  (VP (VB join)
                    (NP (PRP him))
                    (PP (IN in)
                      (NP
                        (NP (NNP New) (NNP Zealand))
                        (PP (IN as)
                          (NP (NNS refugees)))))))))))))
    (, ,)
    (NP (DT the) (NN decision))
    (VP (MD will)
      (VP (VB annoy)
        (UCP
          (NP (NNP NZ) (NNP First) (NN leader) (NNP Winston) (NNP Peters))
          (CC but)
          (S
            (VP (VB delight)
              (NP
                (NP (DT a) (NN boy))
                (SBAR
                  (WHNP (WP who))
                  (S
                    (VP (VBZ has) (RB not)
                      (VP (VBN seen)
                        (NP (PRP$ his) (NN father))
                        (PP (IN for)
                          (NP
                            (QP (RB nearly) (CD three))
                            (NNS years)))))))))))))
```

```
    (. .)))



Sentence 2:

"Mr Zaoui's wife, Leila, and four sons - Youssef, 7, Adbel, 14, Soheib, 17, and Hicham, 19 -
have applied through Mr Zaoui's lawyer, Deborah Manning, to move to New Zealand as refugees."

UGE output:
(HAVE* (:ACTOR
        (WIFE* (:NOUN) (:MODIFIER (MR*) (|ZAOUI'S*|))
               (COMMA* (LEILA* (:UNKNOWN) (:NAME)))
         (AND*
          (SONS* (:NOUN) (:MODIFIER (FOUR*))
           (:B-BLK
            (YOUSSEF* (:UNKNOWN) (:NAME) (COMMA* (7* (:NOUN) (:NUMBER)))
                      (COMMA* (ADBEL* (:UNKNOWN) (:NAME)))
                      (COMMA* (14* (:NOUN) (:NUMBER))) (COMMA* (SOHEIB* (:UNKNOWN) (:NAME)))
                      (COMMA* (17* (:NOUN) (:NUMBER)))
             (AND* (HICHAM* (:UNKNOWN) (:NAME))) (COMMA* (19* (:NOUN) (:NUMBER))))))))
       (:WHAT
        (APPLIED* (:WHAT ?R)
                  (:THROUGH* (LAWYER* (:NOUN) (:MODIFIER (MR*) (|ZAOUI'S*|))
                                      (COMMA* (DEBORAH-MANNING* (:UNKNOWN) (:NAME)))))
                  (:TO** (MOVE*) (:TO* (NEW-ZEALAND* (:UNKNOWN) (:NAME)
                                                     (:AS* (REFUGEES* (:NOUN)))))))))


Stanford output:
(ROOT
  (S
    (NP
      (NP
        (NP
          (NP (NNP Mr) (NNP Zaoui) (POS 's))
          (NN wife))
        (, ,)
        (NP (NNP Leila))
        (, ,))
      (CC and)
      (NP
        (NP (CD four) (NNS sons))
        (PRN (: -)
          (NP
            (NP (NNP Youssef))
            (, ,)
            (NP (CD 7))
            (, ,)
            (NP (NNP Adbel))
            (, ,)
            (NP (CD 14))
            (, ,)
            (NP (NNP Soheib))
            (, ,)
            (NP (CD 17))
            (, ,)
            (CC and)
            (NP
              (NP (NNP Hicham))
              (, ,)
              (NP (CD 19))))
          (: -))))
    (VP (VBP have)
      (VP (VBN applied)
        (PP (IN through)
          (NP
            (NP
              (NP (NNP Mr) (NNP Zaoui) (POS 's))
              (NN lawyer))
            (, ,)
            (NP (NNP Deborah) (NNP Manning))
            (, ,)))
        (S
          (VP (TO to)
            (VP (VB move)
              (PP (TO to)
                (NP
                  (NP (NNP New) (NNP Zealand))
                  (PP (IN as)
                    (NP (NNS refugees))))))))))
  (. .)))
```

```
Sentence 3:

"Ms Manning told the Sunday Star-Times that supporters had offered the family financial assistance,
but she could not preclude the need for the family to apply for a benefit, to which refugees are entitled."


UGE output:
(TOLD-THAT* (:ACTOR (MANNING* (:UNKNOWN) (:NAME) (:MODIFIER (MS*))))
            (:RECIPIENT (SUNDAY-STAR-TIMES* (:UNKNOWN) (:NAME) (:MODIFIER (THE*))))
            (:MS1
             (BUT*
              (:MS2
               (COULD* (:ACTOR (SHE* (:PNOUN)))
                (:MS1
                 (PRECLUDE* (:ACTOR ?L)
                  (:WHAT
                   (NEED* (:NOUN) (:MODIFIER (THE*))
                    (:FOR*
                     (FAMILY* (:NOUN) (:MODIFIER (THE*))
                      (:TO** (APPLY*)
                       (:FOR*
                        (BENEFIT* (:NOUN) (:MODIFIER (A*))
                                  (:TO* (:WHICH*
                                         (:MS1 (REFUGEES* (:NOUN)
                                                          (:ARE* (ENTITLED*
                                                                  (:WHAT ?R))))))))))))))))
               (:MANNER (NOT*))))
              (:MS1
               (HAD* (:ACTOR (SUPPORTERS* (:NOUN)))
                (:WHAT
                 (OFFERED* (:RECIPIENT (FAMILY* (:NOUN) (:MODIFIER (THE*))))
                           (:WHAT (ASSISTANCE* (:NOUN) (:MODIFIER (FINANCIAL*)))))))))))


Stanford output:
(ROOT
  (S
    (S
      (NP (NNP Ms) (NNP Manning))
      (VP (VBD told)
        (NP (DT the) (NNP Sunday) (NNP Star-Times))
        (SBAR (IN that)
          (S
            (NP (NNS supporters))
            (VP (VBD had)
              (VP (VBN offered)
                (NP (DT the) (NN family))
                (NP (JJ financial) (NN assistance))))))))
    (, ,)
    (CC but)
    (S
      (NP (PRP she))
      (VP (MD could) (RB not)
        (VP (VB preclude)
          (NP (DT the) (NN need))
          (PP (IN for)
            (NP (DT the) (NN family)
              (S
                (VP (TO to)
                  (VP (VB apply)
                    (PP (IN for)
                      (NP
                        (NP (DT a) (NN benefit))
                        (, ,)
                        (SBAR
                          (WHPP (TO to)
                            (WHNP (WDT which)))
                          (S
                            (NP (NNS refugees))
                            (VP (VBP are)
                              (ADJP (VBN entitled))))))))))))))
    (. .)))



Sentence 4:

"The Government has refused to consider the applications until Mr Zaoui's case is resolved
```

```
(he is on bail awaiting a review of his SIS security risk certificate, which could mean his deportation)."


UGE output:
(HAS* (:ACTOR (GOVERNMENT* (:UNKNOWN) (:NAME) (:MODIFIER (THE*))))
      (:WHAT
       (REFUSED*
        (:WHAT
         (:TO**
          (CONSIDER*
           (:WHAT
            (APPLICATIONS* (:NOUN) (:MODIFIER (THE*))
             (:UNTIL*
              (:MS1
               (CASE* (:NOUN) (:MODIFIER (MR*) (|ZAOUI'S*|))
                (:IS*
                 (RESOLVED* (:WHAT ?R)
                  (:B-BLOCK
                   (HE* (:PNOUN)
                    (:IS*
                     (:ON*
                      (BAIL* (:NOUN)
                       (AWAITING*
                        (:WHAT
                         (REVIEW* (:NOUN) (:MODIFIER (A*))
                          (:OF*
                           (CERTIFICATE* (:NOUN)
                                         (:X-WORDS (*SIS**)
                                                   (SECURITY*)
                                                   (RISK*)
                                                   (CERTIFICATE*))
                                         (:NAME)
                            (:MODIFIER (HIS*))))
                          (:WHICH*
                           (:MS1
                            (COULD* (:ACTOR ?L)
                             (:MS1
                              (MEAN* (:ACTOR ?L)
                                     (:WHAT (DEPORTATION* (:NOUN)
                                                          (:MODIFIER (HIS*))))
                                     ))))))))))))))))))))))))


Stanford output:
(ROOT
  (S
    (NP (DT The) (NN Government))
    (VP (VBZ has)
      (VP (VBN refused)
        (S
         (VP (TO to)
             (VP (VB consider)
                 (NP (DT the) (NNS applications))
                 (SBAR (IN until)
                       (S
                        (NP
                         (NP (NNP Mr) (NNP Zaoui) (POS 's))
                         (NN case))
                        (VP (VBZ is)
                            (ADJP (VBN resolved)
                                  (PRN (-LRB- -LRB-)
                                       (S
                                        (NP (PRP he))
                                        (VP (VBZ is)
                                            (PP (IN on)
                                                (NP
                                                 (NP (NN bail))
                                                 (VP (VBG awaiting)
                                                     (NP
                                                      (NP (DT a)
                                                          (NN review))
                                                      (PP (IN of)
                                                          (NP (PRP$ his)
                                                              (NNP SIS)
                                                              (NN security)
                                                              (NN risk)
                                                              (NN certificate)))
                                                      (, ,)
                                                      (SBAR
                                                       (WHNP (WDT which))
                                                       (S
                                                        (VP (MD could)
                                                            (VP (VB mean)
```

```
                                                                (NP (PRP$ his)
                                                                    (NN deportation))
                                                                ))))))))))
                                        (-RRB- -RRB-)))))))))))
   (. .)))
```

```
Sentence 5:

"The applications have been criticised by the National and New Zealand First parties."


UGE output:
(HAVE* (:ACTOR (APPLICATIONS* (:NOUN) (:MODIFIER (THE*))))
       (:WHAT
        (BEEN*
         (:WHAT
          (CRITICISED* (:WHAT ?R)
           (:BY*
            (NATIONAL* (:UNKNOWN) (:NAME)
                       (:MODIFIER (THE*))
                       (AND* (PARTIES* (:NOUN)
                                       (:X-WORDS (NEW-ZEALAND-FIRST*)
                                                 (PARTIES*)) (:NAME)))))))))


Stanford output:
(ROOT
  (S
    (NP (DT The) (NNS applications))
    (VP (VBP have)
      (VP (VBN been)
        (VP (VBN criticised)
          (PP (IN by)
            (NP (DT the) (NNP National)
              (CC and)
              (NNP New) (NNP Zealand) (NNP First) (NNS parties))))))
   (. .)))



Sentence 6:

"While the politicians bicker over issues of cost, security risk and human rights, for
Mr Zaoui's youngest son all it means is growing up without a father."


UGE output:
(WHILE* (:MS1
         (BICKER* (:ACTOR (POLITICIANS* (:NOUN) (:MODIFIER (THE*))))
          (:OVER*
           (ISSUES* (:NOUN) (:OF* (COST* (:NOUN)))
                    (COMMA* (RISK* (:NOUN) (:X-WORDS (SECURITY*) (RISK*))))
             (AND* (RIGHTS* (:NOUN) (:MODIFIER (HUMAN*))))))))
        (:MS2
         (ALL* (:NOUN) (:FOR* (SON* (:NOUN) (:MODIFIER (MR*) (|ZAOUI'S*|) (YOUNGEST*))))
          (:CLAUSE (MEANS* (:ACTOR (IT* (:PNOUN))))
               (:IS* (GROWING* (:WHAT ?R) (:MANNER (UP*))
                               (:WITHOUT* (FATHER* (:NOUN) (:MODIFIER (A*)))))))))


Stanford output:
(ROOT
  (S
    (SBAR (IN While)
      (S
        (NP (DT the) (NNS politicians))
        (VP (VBP bicker)
          (PP (IN over)
            (NP
              (NP (NNS issues))
              (PP (IN of)
                (NP (NN cost))))))))
    (, ,)
    (NP
      (NP (NN security) (NN risk))
      (CC and)
      (NP (JJ human) (NNS rights))
      (, ,)
      (PP (IN for)
        (NP
```

```
          (NP
            (NP (NNP Mr) (NNP Zaoui) (POS 's))
            (JJS youngest) (NN son))
          (SBAR
            (WHNP (DT all))
            (S
              (NP (PRP it))
              (VP (VBZ means)))))))
    (VP (VBZ is)
      (VP (VBG growing)
        (PRT (RP up))
        (PP (IN without)
          (NP (DT a) (NN father)))))
   (. .)))
```

Sentence 7:

"The family spoke of the separation to TV3's Campbell Live producer Carol Hirschfeld,
who travelled to interview them at a Southeast Asian location for a segment to screen tonight."

UGE output:

```
(SPOKE* (:ACTOR (FAMILY* (:NOUN) (:MODIFIER (THE*))))
        (:OF* (SEPARATION* (:NOUN) (:MODIFIER (THE*))))
        (:TO*
         (CAROL-HIRSCHFELD* (:X-WORDS (CAMPBELL-LIVE*)
                                      (PRODUCER*)
                                      (CAROL-HIRSCHFELD*))
                            (:NAME) (:MODIFIER (|*TV3*'S*|))
          (:WHO*
           (:MS1
            (TRAVELLED* (:ACTOR ?L)
             (:TO**
              (INTERVIEW*
               (:WHAT
                (THEM* (:PNOUN)
                 (:AT*
                  (LOCATION* (:NOUN)
                             (:X-WORDS (SOUTHEAST-ASIAN*)
                                       (LOCATION*)) (:NAME) (:MODIFIER (A*))
                   (:FOR*
                    (SEGMENT* (:NOUN) (:MODIFIER (A*))
                              (:TO** (SCREEN* (:WHAT ?R))
                                     (:MANNER (TONIGHT*)))))))))))))))))
```

Stanford output:

```
(ROOT
  (S
    (NP (DT The) (NN family))
    (VP (VBD spoke)
      (PP (IN of)
        (NP (DT the) (NN separation)))
      (PP (TO to)
        (NP
          (NP (NNP TV3) (POS 's))
          (NNP Campbell) (NNP Live) (NN producer)))
      (NP
        (NP (NNP Carol) (NNP Hirschfeld))
        (, ,)
        (SBAR
          (WHNP (WP who))
          (S
            (VP (VBD traveled)
              (S
                (VP (TO to)
                  (VP (VB interview)
                    (NP (PRP them))
                    (PP (IN at)
                      (NP
                        (NP (DT a)
                          (ADJP (JJ Southeast) (JJ Asian))
                          (NN location))
                        (PP (IN for)
                          (NP (DT a) (NN segment)
                            (S
                              (VP (TO to)
                                (VP (VB screen)
                                  (NP (RB tonight)))))))))))))))
   (. .)))
```

```
Sentence 8:

"They also got a taste of what they were missing - a satellite link let Mr Zaoui and
the family talk face to face for the first time since he left for New Zealand in December 2002."


UGE output:
(GOT* (:ACTOR (THEY* (:PNOUN) (:MANNER (ALSO*))))
      (:WHAT
       (TASTE* (:NOUN) (:MODIFIER (A*))
        (:OF*
         (WHAT* (:NOUN)
          (:WH-NOUN*
           (:MS1
            (THEY* (:PNOUN)
             (:WERE*
              (MISSING* (:WHAT ?R)
               (:B-BLOCK
                (LET* (:ACTOR (LINK* (:NOUN)
                                     (:X-WORDS
                                      (SATELLITE*)
                                      (LINK*))
                                     (:MODIFIER (A*))))
                  (:MS1
                   (SINCE*
                    (:MS2
                     (LEFT* (:ACTOR (HE* (:PNOUN)))
                            (:FOR* (NEW-ZEALAND* (:UNKNOWN) (:NAME)
                                                 (:IN* (DECEMBER-2002*
                                                        (:UNKNOWN) (:NAME)))))))
                    (:MS1
                     (TALK*
                      (:ACTOR (ZAOUI* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))
                                      (AND* (FAMILY* (:NOUN) (:MODIFIER (THE*))))))
                      (:MANNER (FACE-TO-FACE*))
                      (:FOR* (TIME* (:NOUN) (:MODIFIER (THE*) (FIRST*))))))))))))))


Stanford output:
(ROOT
  (S
    (S
      (NP (PRP They))
      (ADVP (RB also))
      (VP (VBD got)
        (NP
          (NP (DT a) (NN taste))
          (PP (IN of)
            (SBAR
              (WHNP (WP what))
              (S
                (NP (PRP they))
                (VP (VBD were)
                  (ADJP (VBG missing)))))))))
    (: -)
    (S
      (NP (DT a) (NN satellite) (NN link))
      (VP (VBD let)
        (SBAR
          (S
            (NP
              (NP (NNP Mr) (NNP Zaoui))
              (CC and)
              (NP (DT the) (NN family) (NN talk)))
            (VP (VBP face)
              (S
                (VP (TO to)
                  (VP (VB face)
                    (PP (IN for)
                      (NP (DT the) (JJ first) (NN time)))
                    (SBAR (IN since)
                      (S
                        (NP (PRP he))
                        (VP (VBD left)
                          (PP (IN for)
                            (NP (NNP New) (NNP Zealand)))
                          (PP (IN in)
                            (NP (NNP December) (CD 2002)))))))))))))
    (. .)))
```

```
Sentence 9:

"Hirschfeld said separation had a strong effect on son Youssef, who was 4 1/2 years old
when he last saw his father."


UGE output:
(SAID* (:ACTOR (HIRSCHFELD* (:UNKNOWN) (:NAME)))
       (:MS1
        (HAD* (:ACTOR (SEPARATION* (:NOUN)))
         (:WHAT
          (EFFECT* (:NOUN) (:MODIFIER (A*) (STRONG*))
           (:ON*
            (YOUSSEF* (:NAME) (:X-WORDS (SON*) (YOUSSEF*))
             (:WHO*
              (:MS1
               (:WAS*
                (OLD* (:NOUN) (:X-WORDS (4*) (1/2*) (YEARS*) (OLD*)) (:NUMBER) (:NUMBER)
                 (:WHEN*
                  (:MS1
                   (SAW* (:ACTOR (HE* (:PNOUN) (:MANNER (LAST*))))
                         (:WHAT (FATHER* (:NOUN) (:MODIFIER (HIS*))))))))))))))))


Stanford output:
(ROOT
  (S
    (NP (NNP Hirschfeld))
    (VP (VBD said)
      (SBAR
        (S
          (NP (NN separation))
          (VP (VBD had)
            (NP
              (NP (DT a) (JJ strong) (NN effect))
              (PP (IN on)
                (NP
                  (NP (NNP son) (NNP Youssef))
                  (, ,)
                  (SBAR
                    (WHNP (WP who))
                    (S
                      (VP (VBD was)
                        (ADJP
                          (NP (CD 4 12) (NNS years))
                          (JJ old))
                        (SBAR
                          (WHADVP (WRB when))
                          (S
                            (NP (PRP he))
                            (ADVP (RB last))
                            (VP (VBD saw)
                              (NP (PRP$ his) (NN father))))))))))))))
    (. .)))



Sentence 10:

"Ähmed phones regularly but I think the satellite link was so powerful because
they hadn't actually seen each other in the flesh for nearly three years."


UGE output:
(BUT* (:MS2
       (THINK* (:ACTOR (I* (:PNOUN)))
        (:MS1
         (BECAUSE*
          (:MS2
           (HAD* (:ACTOR (THEY* (:PNOUN)))
            (:WHAT
             (SEEN*
              (:WHAT
               (EACH-OTHER* (:PNOUN)
                (:IN*
                 (FLESH* (:NOUN) (:MODIFIER (THE*))
                         (:FOR* (YEARS* (:NOUN) (:MODIFIER (THREE*))))
                  (:MANNER (NEARLY*)))))))))))))
```

```
           (:MANNER (NOT*)) (:MANNER (ACTUALLY*))))
         (:MS1
          (LINK* (:NOUN) (:X-WORDS (SATELLITE*) (LINK*)) (:MODIFIER (THE*))
           (:WAS* (?R (:MODIFIER (SO*) (POWERFUL*)))))))))
      (:MS1 (PHONES* (:ACTOR (AHMED* (:UNKNOWN) (:NAME)))
                     (:MANNER (REGULARLY*)))))


Stanford output:
(ROOT
  (S
    (NP (JJ Ahmed) (NNS phones))
    (ADVP (RB regularly))
    (ADVP (CC but))
    (NP (PRP I))
    (VP (VBP think)
      (SBAR
        (S
          (NP (DT the) (NN satellite) (NN link))
          (VP (VBD was)
            (ADJP (RB so) (JJ powerful))
            (SBAR (IN because)
              (S
                (NP (PRP they))
                (VP (VBD had) (RB n't)
                  (ADVP (RB actually))
                  (VP (VBN seen)
                    (NP (DT each) (JJ other))
                    (PP (IN in)
                      (NP
                        (NP (DT the) (NN flesh))
                        (PP (IN for)
                          (NP
                            (QP (RB nearly) (CD three))
                            (NNS years)))))))))))
    (. .)))



Sentence 11:

"Seeing how much Youssef had grown was something Ahmed really responded to."


UGE output:
(SEEING* (:WHAT (:HOW-MUCH*
                 (:MS1 (HAD* (:ACTOR (YOUSSEF* (:UNKNOWN) (:NAME)))
                             (:WHAT (GROWN* (:WHAT ?R)))))))
         (:WAS*
          (SOMETHING* (:NOUN)
                      (:CLAUSE
                       (RESPONDED* (:ACTOR (AHMED* (:UNKNOWN) (:NAME)
                                                   (:MANNER (REALLY*))))
                                   (:TO** ?R))))))


Stanford output:
(ROOT
  (S
    (S
      (VP (VBG Seeing)
        (SBAR
          (WHADJP (WRB how) (JJ much))
          (S
            (NP (NNP Youssef))
            (VP (VBD had)
              (VP (VBN grown)))))))
    (VP (VBD was)
      (NP
        (NP (NN something))
        (SBAR
          (S
            (NP (NNP Ahmed))
            (ADVP (RB really))
            (VP (VBD responded)
              (PP (TO to)))))))
    (. .)))



Sentence 12:
```

```
"Hirschfeld said the family had perceived NZ as ä peaceful land which offered
a place of refuge with a good record for democracy."


UGE output:
(SAID* (:ACTOR (HIRSCHFELD* (:UNKNOWN) (:NAME)))
       (:MS1
        (HAD* (:ACTOR (FAMILY* (:NOUN) (:MODIFIER (THE*))))
         (:WHAT
          (PERCEIVED*
           (:WHAT
            (*NZ** (:UNKNOWN) (:NAME)
             (:AS*
              (LAND* (:NOUN) (:MODIFIER (A*) (PEACEFUL*))
               (:WHICH*
                (:MS1
                 (OFFERED* (:ACTOR ?L)
                  (:WHAT
                   (PLACE* (:NOUN) (:MODIFIER (A*)) (:OF* (REFUGE* (:NOUN)))
                           (:WITH* (RECORD* (:NOUN) (:MODIFIER (A*) (GOOD*))
                                            (:FOR* (DEMOCRACY* (:NOUN))))))))))))))))


Stanford output:
(ROOT
  (S
    (NP (NNP Hirschfeld))
    (VP (VBD said)
      (SBAR
        (S
          (NP (DT the) (NN family))
          (VP (VBD had)
            (VP (VBN perceived)
              (NP (NNP NZ))
              (PP (IN as) (`` ``)
                (NP
                  (NP (DT a) (JJ peaceful) (NN land))
                  (SBAR
                    (WHNP (WDT which))
                    (S
                      (VP (VBD offered)
                        (NP
                          (NP (DT a) (NN place))
                          (PP (IN of)
                            (NP (NN refuge))))
                        (PP (IN with)
                          (NP
                            (NP (DT a) (JJ good) (NN record))
                            (PP (IN for)
                              (NP (NN democracy))))))))))
                (`` ``)))))))
   (. .)))



Sentence 13:

"National's immigration spokesman, Tony Ryall, said calls for the family to be
allowed in before Mr Zaoui's case was settled were ludicrous."


UGE output:
(SAID* (:ACTOR
        (SPOKESMAN* (:NOUN) (:X-WORDS (IMMIGRATION*) (SPOKESMAN*))
                    (:MODIFIER (|NATIONAL'S*|))
         (COMMA* (TONY-RYALL* (:UNKNOWN) (:NAME)))))
       (:MS1
        (CALLS* (:NOUN)
                (:FOR* (FAMILY* (:NOUN) (:MODIFIER (THE*))
                                (:TO** (BE* (:WHAT (ALLOWED* (:WHAT ?R)
                                                             (:MANNER (IN*))))))))
                (:BEFORE* (:MS1 (CASE* (:NOUN) (:MODIFIER (MR*) (|ZAOUI'S*|))
                                       (:WAS* (SETTLED* (:WHAT ?R))))))
                (:WERE* (?R (:MODIFIER (LUDICROUS*)))))


Stanford output:
(ROOT
  (S
    (NP
      (NP
```

```
        (NP (NNP National) (POS 's))
        (NN immigration) (NN spokesman))
      (, ,)
      (NP (NNP Tony) (NNP Ryall))
      (, ,))
    (VP (VBD said)
      (SBAR
        (S
          (NP
            (NP (NNS calls))
            (PP (IN for)
              (NP (DT the) (NN family)
                (S
                  (VP (TO to)
                    (VP (VB be)
                      (VP (VBN allowed)
                        (PP (IN in)
                          (SBAR (IN before)
                            (S
                              (NP
                                (NP (NNP Mr) (NNP Zaoui) (POS 's))
                                (NN case))
                              (VP (VBD was)
                                (VP (VBN settled)))))))))))))
          (VP (VBD were)
            (ADJP (JJ ludicrous))))))
    (. .)))
```

Sentence 14:

"We shouldn't have a system in New Zealand where you get one refugee and end up with several others."

UGE output:

```
(SHOULD* (:ACTOR (WE* (:PNOUN)))
         (:MS1
          (HAVE* (:ACTOR ?L)
           (:WHAT
            (SYSTEM* (:NOUN) (:MODIFIER (A*))
             (:IN*
              (NEW-ZEALAND* (:UNKNOWN) (:NAME)
               (:WHERE*
                (:MS1
                 (AND* (:MS2 (END* (:ACTOR ?L) (:MANNER (UP*))
                                   (:WITH* (OTHERS* (:NOUN)
                                                    (:MODIFIER (SEVERAL*))))))
                       (:MS1 (GET* (:ACTOR (YOU* (:PNOUN)))
                                   (:WHAT (REFUGEE* (:NOUN)
                                                    (:MODIFIER (ONE*)))))))))))))
         (:MANNER (NOT*)))
```

Stanford output:

```
(ROOT
  (S
    (NP (PRP We))
    (VP (MD should) (RB n't)
      (VP
        (VP (VB have)
          (NP
            (NP (DT a) (NN system))
            (PP (IN in)
              (NP (NNP New) (NNP Zealand))))
          (SBAR
            (WHADVP (WRB where))
            (S
              (NP (PRP you))
              (VP (VBP get)
                (NP (CD one) (NN refugee))))))
        (CC and)
        (VP (VB end)
          (PRT (RP up))
          (PP (IN with)
            (NP (JJ several) (NNS others))))))
   (. .)))
```

Sentence 15:

```
"It puts a lot of pressure on the system."


UGE output:
(PUTS* (:ACTOR (IT* (:PNOUN)))
       (:WHAT (LOT* (:NOUN) (:MODIFIER (A*)) (:OF* (PRESSURE* (:NOUN)))
                    (:ON* (SYSTEM* (:NOUN) (:MODIFIER (THE*)))))))


Stanford output:
(ROOT
  (S
    (NP (PRP It))
    (VP (VBZ puts)
      (NP
        (NP (DT a) (NN lot))
        (PP (IN of)
          (NP
            (NP (NN pressure))
            (PP (IN on)
              (NP (DT the) (NN system)))))))
   (. .)))



Sentence 16:

"If it was decided Mr Zaoui should go but his family had already been let in,
then Mr Zaoui would have another argument for coming back, said Mr Ryall."


UGE output:
(SAID* (:ACTOR (RYALL* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))))
       (:MS1
        (IF* (:MS1
              (IT* (:PNOUN)
               (:WAS*
                (DECIDED*
                 (:MS1
                  (BUT*
                   (:MS2
                    (HAD* (:ACTOR (FAMILY* (:NOUN) (:MODIFIER (HIS*))))
                          (:WHAT (BEEN*
                                  (:WHAT (LET* (:WHAT ?R) (:IN* ?R)))))
                          (:MANNER (ALREADY*))))
                   (:MS1
                    (SHOULD* (:ACTOR (ZAOUI* (:UNKNOWN) (:NAME)
                                             (:MODIFIER (MR*))))
                             (:MS1 (GO* (:ACTOR ?L))))))))))
             (:MS2
              (WOULD* (:ACTOR (ZAOUI* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))))
               (:MS1
                (HAVE* (:ACTOR ?L)
                       (:WHAT (ARGUMENT* (:NOUN) (:MODIFIER (ANOTHER*))
                                         (:FOR* (COMING* (:WHAT ?R)
                                                         (:MANNER (BACK*))))))))))))


Stanford output:
(ROOT
  (SINV
    (S
      (SBAR (IN If)
        (S
          (NP (PRP it))
          (VP (VBD was)
            (VP (VBN decided)
              (SBAR
                (S
                  (NP (NNP Mr) (NNP Zaoui))
                  (VP (MD should)
                    (VP (VB go)
                      (SBAR (CC but)
                        (S
                          (NP (PRP$ his) (NN family))
                          (VP (VBD had)
                            (ADVP (RB already))
                            (VP (VBN been)
                              (VP (VBN let)
                                (PP (IN in)))))))))))))))
      (, ,) (RB then)
```

```
      (NP (NNP Mr) (NNP Zaoui))
      (VP (MD would)
        (VP (VB have)
          (NP
            (NP (DT another) (NN argument))
            (PP (IN for)
              (S
                (VP (VBG coming)
                  (ADVP (RB back)))))))))
    (, ,)
    (VP (VBD said))
    (NP (NNP Mr) (NNP Ryall))
   (. .)))
```

```
Sentence 17:

"Mr Peters said at his campaign launch in Takapuna yesterday that the Zaoui case had
already cost the country $2 million."


UGE output:
(SAID-THAT* (:ACTOR (PETERS* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))))
            (:MS1
             (HAD* (:ACTOR (CASE* (:NOUN)
                                  (:X-WORDS (ZAOUI*) (CASE*)) (:NAME)
                                  (:MODIFIER (THE*))))
              (:MS1
               (COST* (:ACTOR ?L) (:RECIPIENT (COUNTRY* (:NOUN)
                                                        (:MODIFIER (THE*))))
                (:WHAT ($2-MILLION* (:NOUN) (:NUMBER)))))
              (:MANNER (ALREADY*))))
            (:AT*
             (LAUNCH* (:NOUN) (:X-WORDS (CAMPAIGN*) (LAUNCH*))
                      (:MODIFIER (HIS*))
                      (:IN* (TAKAPUNA* (:UNKNOWN) (:NAME)))))
            (:MANNER (YESTERDAY*)))


Stanford output:
(ROOT
  (S
    (NP (NNP Mr) (NNP Peters))
    (VP (VBD said)
      (PP (IN at)
        (NP
          (NP (PRP$ his) (NN campaign) (NN launch))
          (PP (IN in)
            (NP (NNP Takapuna)))))
      (NP (NN yesterday))
      (SBAR (IN that)
        (S
          (NP (DT the) (NNP Zaoui) (NN case))
          (VP (VBD had)
            (ADVP (RB already))
            (VP (VB cost)
              (NP (DT the) (NN country))
              (NP
                (QP ($ $) (CD 2) (CD million))))))))
    (. .)))
```

```
Sentence 18:

"Mr Zaoui, you can see your family tomorrow if you would just get on a plane and go and see them."


UGE output:
(IF* (:MS2
      (AND* (:MS2 (SEE* (:ACTOR ?L) (:WHAT (THEM* (:PNOUN)))))
       (:MS1
        (AND* (:MS2 (GO* (:ACTOR ?L)))
         (:MS1
          (WOULD* (:ACTOR (YOU* (:PNOUN)))
                  (:MS1 (GET* (:ACTOR ?L)
                              (:ON* (PLANE* (:NOUN) (:MODIFIER (A*))))))
           (:MANNER (JUST*))))))))
     (:MS1
      (CAN* (:ACTOR (ZAOUI* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))
```

```
                          (COMMA* (YOU* (:PNOUN)))))
            (:MS1 (SEE* (:ACTOR ?L) (:WHAT (FAMILY* (:NOUN) (:MODIFIER (YOUR*))))
                        (:MANNER (TOMORROW*)))))))
```

```
Stanford output:
(ROOT
  (S
    (NP (NNP Mr) (NNP Zaoui))
    (, ,)
    (NP (PRP you))
    (VP (MD can)
      (VP (VB see)
        (NP (PRP$ your) (NN family) (NN tomorrow))
        (SBAR (IN if)
          (S
            (NP (PRP you))
            (VP (MD would)
              (ADVP (RB just))
              (VP
                (VP (VB get)
                  (PP (IN on)
                    (NP (DT a) (NN plane))))
                (CC and)
                (VP (VB go)
                  (CC and)
                  (VB see)
                  (NP (PRP them)))))))))
   (. .)))
```

```
Sentence 19:

"But the Green Party said Mr Zaoui should be shown some "Kiwi compassion"
because his separation from his family was the fault of deficiencies in New Zealand's own systems."
```

```
UGE output:
(BUT* (:MS1
       (SAID* (:ACTOR (GREEN-PARTY* (:UNKNOWN) (:NAME) (:MODIFIER (THE*))))
        (:MS1
         (BECAUSE*
          (:MS2
           (SEPARATION* (:NOUN) (:MODIFIER (HIS*))
                        (:FROM* (FAMILY* (:NOUN)
                                         (:MODIFIER (HIS*))))
            (:WAS*
             (FAULT* (:NOUN) (:MODIFIER (THE*)) (:OF* (DEFICIENCY* (:NOUN)))
              (:IN* (SYSTEMS* (:NOUN) (:MODIFIER (|NEW-ZEALAND'S*|) (OWN*)))))))
          (:MS1
           (SHOULD* (:ACTOR (ZAOUI* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))))
            (:MS1
             (BE* (:ACTOR ?L)
                  (:WHAT (SHOWN*
                          (:WHAT (COMPASSION* (:NOUN) (:MODIFIER (KIWI*))
                                              (:MODIFIER (SOME*))))))))))))))
```

```
Stanford output:
(ROOT
  (S (CC But)
    (NP (DT the) (NNP Green) (NNP Party))
    (VP (VBD said)
      (SBAR
        (S
          (NP (NNP Mr) (NNP Zaoui))
          (VP (MD should)
            (VP (VB be)
              (VP (VBN shown)
                (NP (DT some) (`` ``) (JJ Kiwi) (NN compassion) ('' ''))
                (SBAR (IN because)
                  (S
                    (NP
                      (NP (PRP$ his) (NN separation))
                      (PP (IN from)
                        (NP (PRP$ his) (NN family))))
                    (VP (VBD was)
                      (NP
                        (NP (DT the) (NN fault))
                        (PP (IN of)
                          (NP
```

```
                          (NP (NNS deficiencies))
                          (PP (IN in)
                            (NP
                              (NP (NNP New) (NNP Zealand) (POS 's))
                              (JJ own) (NNS systems)))))))))))))))
    (. .)))
```

Sentence 20:

"He had also spent two years in jail ünnecessarily∴"

```
UGE output:
(HAD* (:ACTOR (HE* (:PNOUN)))
      (:WHAT (SPENT* (:WHAT (YEARS* (:NOUN) (:MODIFIER (TWO*))
                                    (:IN* (JAIL* (:NOUN)))))))
      (:MANNER (ALSO*))
      (:MANNER (UNNECESSARILY*)))
```

```
Stanford output:
(ROOT
  (S
    (NP (PRP He))
    (VP (VBD had)
      (VP
        (ADVP (RB also))
        (VBN spent)
        (PP
          (NP (CD two) (NNS years))
          (IN in)
          (NP (NN jail)))
        (`` ``)
        (ADVP (RB unnecessarily))
        ('' '')))
   (. .)))
```

Sentence 21:

"The party said allowing the family to get together could not
represent a security danger to New Zealand."

```
UGE output:
(SAID* (:ACTOR (PARTY* (:NOUN) (:MODIFIER (THE*))))
       (:MS1
        (COULD* (:ACTOR (ALLOWING*
                         (:WHAT (FAMILY* (:NOUN) (:MODIFIER (THE*))))
                         (:TO** (GET*) (:MANNER (TOGETHER*)))))
         (:MS1
          (REPRESENT* (:ACTOR ?L)
           (:WHAT
            (DANGER* (:NOUN) (:X-WORDS (SECURITY*) (DANGER*)) (:MODIFIER (A*))
             (:TO* (NEW-ZEALAND* (:UNKNOWN) (:NAME)))))))
              (:MANNER (NOT*)))))
```

```
Stanford output:
(ROOT
  (S
    (NP (DT The) (NN party))
    (VP (VBD said)
      (SBAR
        (S
          (S
            (VP (VBG allowing)
              (NP (DT the) (NN family)
                (S
                  (VP (TO to)
                    (VP (VB get)
                      (ADVP (RB together))))))))
          (VP (MD could) (RB not)
            (VP (VB represent)
              (NP (DT a) (NN security) (NN danger))
              (PP (TO to)
                (NP (NNP New) (NNP Zealand))))))))
   (. .)))
```

```
ARTICLE 2

Sentence 1:

"New Zealand First is opposing efforts to bring Ahmed Zaoui's family to New Zealand."


UGE output:
(NEW-ZEALAND-FIRST* (:UNKNOWN)
                    (:NAME)
                    (:IS*
                     (OPPOSING*
                      (:WHAT
                       (EFFORTS* (:NOUN)
                        (:TO**
                         (BRING*
                          (:WHAT
                           (FAMILY* (:NOUN) (:MODIFIER (|AHMED-ZAOUI'S*|))
                                    (:TO* (NEW-ZEALAND* (:UNKNOWN) (:NAME)))))))))))


Stanford output:
(ROOT
  (S
    (NP (NNP New) (NNP Zealand) (NNP First))
    (VP (VBZ is)
      (VP (VBG opposing)
        (NP (NNS efforts))
        (S
          (VP (TO to)
            (VP (VB bring)
              (NP
                (NP (NNP Ahmed) (NNP Zaoui) (POS 's))
                (NN family))
              (PP (TO to)
                (NP (NNP New) (NNP Zealand))))))))
   (. .)))



Sentence 2:

"Mr Zaoui's lawyers lodged an appeal with Immigration Minister David Cunliffe
before Christmas to allow his wife and four children to come here."


UGE output:
(LODGED* (:ACTOR (LAWYERS* (:NOUN) (:MODIFIER (MR*) (|ZAOUI'S*|))))
         (:WHAT
          (APPEAL* (:NOUN) (:MODIFIER (AN*))
           (:WITH*
            (IMMIGRATION-MINISTER-DAVID-CUNLIFFE* (:UNKNOWN) (:NAME)
                                                  (:BEFORE* (CHRISTMAS* (:UNKNOWN) (:NAME)))
             (:TO**
              (ALLOW*
               (:WHAT
                (WIFE* (:NOUN) (:MODIFIER (HIS*))
                       (AND* (CHILDREN* (:NOUN) (:MODIFIER (FOUR*))))
                       (:TO** (COME*) (:MANNER (HERE*)))))))))))


Stanford output:
(ROOT
  (S
    (NP
      (NP (NNP Mr) (NNP Zaoui) (POS 's))
      (NNS lawyers))
    (VP (VBN lodged)
      (NP (DT an) (NN appeal))
      (PP (IN with)
        (NP (NNP Immigration) (NNP Minister) (NNP David) (NNP Cunliffe)))
      (PP (IN before)
        (NP (NNP Christmas)))
      (S
        (VP (TO to)
          (VP (VB allow)
```

```
           (S
             (NP
               (NP (PRP$ his) (NN wife))
               (CC and)
               (NP (CD four) (NNS children)))
             (VP (TO to)
               (VP (VB come)
                 (ADVP (RB here)))))))))
   (. .)))
```

Sentence 3:

"Yesterday the Greens backed that call, saying delays in the case were unreasonable
and should not be allowed to keep the family apart."

UGE output:

```
(BACKED* (:ACTOR (GREENS* (:UNKNOWN) (:NAME) (:MODIFIER (THE*))
                          (:MANNER (YESTERDAY*))))
         (:WHAT
          (CALL* (:NOUN) (:MODIFIER (THAT*))
           (COMMA*
            (AND*
             (:MS2
              (SHOULD* (:ACTOR ?L)
               (:MS1
                (BE* (:ACTOR ?L)
                     (:WHAT (ALLOWED* (:WHAT (:TO** (KEEP*
                                                      (:WHAT (FAMILY* (:NOUN)
                                                                      (:MODIFIER (THE*)))))
                                                     (:MANNER (APART*))))))))
              (:MANNER (NOT*))))
             (:MS1
              (SAYING* (:WHAT (DELAYS* (:NOUN) (:IN* (CASE* (:NOUN) (:MODIFIER (THE*))))))
                       (:WERE* (?R (:MODIFIER (UNREASONABLE*)))))))))))
```

Stanford output:

```
(ROOT
  (S
    (NP (NN Yesterday))
    (NP (DT the) (NNS Greens))
    (VP
      (VP (VBD backed)
        (SBAR (IN that)
          (S
            (NP
              (NP (NN call))
              (, ,)
              (VP (VBG saying)
                (NP
                  (NP (NNS delays))
                  (PP (IN in)
                    (NP (DT the) (NN case))))))
            (VP (VBD were)
              (ADJP (JJ unreasonable))))))
      (CC and)
      (VP (MD should) (RB not)
        (VP (VB be)
          (VP (VBN allowed)
            (S
              (VP (TO to)
                (VP (VB keep)
                  (NP (DT the) (NN family))
                  (ADVP (RB apart)))))))))
   (. .)))
```

Sentence 4:

"Mr Zaoui has been struggling to stay in New Zealand since he arrived
in December 2002 seeking refugee status."

UGE output:

```
(SINCE* (:MS2
         (ARRIVED* (:ACTOR (HE* (:PNOUN)))
                   (:IN* (DECEMBER-2002*
```

```
                        (:UNKNOWN)
                        (:NAME)
                        (SEEKING*
                         (:WHAT (STATUS* (:NOUN) (:X-WORDS (REFUGEE*) (STATUS*)))))))))
        (:MS1
         (HAS* (:ACTOR (ZAOUI* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))))
               (:WHAT (BEEN* (:WHAT (STRUGGLING*
                                     (:WHAT (:TO** (STAY*)
                                                   (:IN* (NEW-ZEALAND*
                                                          (:UNKNOWN)
                                                          (:NAME)))))))))))


Stanford output:
(ROOT
  (S
    (NP (NNP Mr) (NNP Zaoui))
    (VP (VBZ has)
      (VP (VBN been)
        (VP (VBG struggling)
          (S
            (VP (TO to)
              (VP (VB stay)
                (PP (IN in)
                  (NP (NNP New) (NNP Zealand)))
                (SBAR (IN since)
                  (S
                    (NP (PRP he))
                    (VP (VBD arrived)
                      (PP (IN in)
                        (NP (NNP December) (CD 2002)))
                      (S
                        (VP (VBG seeking)
                          (NP (NN refugee) (NN status)))))))))))
   (. .)))



Setence 5:

"He is awaiting a review of his case."


UGE output:
(HE* (:PNOUN)
     (:IS* (AWAITING* (:WHAT (REVIEW* (:NOUN) (:MODIFIER (A*))
                                      (:OF* (CASE* (:NOUN)
                                                   (:MODIFIER (HIS*)))))))))


Stanford output:
(ROOT
  (S
    (NP (PRP He))
    (VP (VBZ is)
      (VP (VBG awaiting)
        (NP
          (NP (DT a) (NN review))
          (PP (IN of)
            (NP (PRP$ his) (NN case))))))
   (. .)))



Sentence 6:

"The Department of Labour is preparing advice for Mr Cunliffe before he decides."


UGE output:
(DEPARTMENT* (:UNKNOWN)
             (:NAME)
             (:MODIFIER (THE*))
             (:OF* (LABOUR* (:UNKNOWN) (:NAME)))
             (:IS*
              (PREPARING*
               (:WHAT
                (ADVICE* (:NOUN)
                 (:FOR*
                  (CUNLIFFE* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))
                             (:BEFORE* (:MS1 (DECIDES* (:ACTOR (HE* (:PNOUN))))))))))))
```

```
Stanford output:
(ROOT
  (S
    (NP
      (NP (DT The) (NNP Department))
      (PP (IN of)
        (NP (NNP Labor))))
    (VP (VBZ is)
      (VP (VBG preparing)
        (NP
          (NP (NN advice))
          (PP (IN for)
            (NP (NNP Mr) (NNP Cunliffe))))
        (SBAR (IN before)
          (S
            (NP (PRP he))
            (VP (VBZ decides))))))
   (. .)))
```

```
Sentence 7:

"NZ First associate immigration spokesman Peter Brown said the family
should not be allowed to come here from Southeast Asia."
```

```
UGE output:
(SAID* (:ACTOR (PETER-BROWN* (:X-WORDS (*NZ*-FIRST*)
                                       (ASSOCIATE*)
                                       (IMMIGRATION*)
                                       (SPOKESMAN*)
                                       (PETER-BROWN*)) (:NAME)))
       (:MS1
        (SHOULD* (:ACTOR (FAMILY* (:NOUN) (:MODIFIER (THE*))))
         (:MS1
          (BE* (:ACTOR ?L)
               (:WHAT (ALLOWED* (:WHAT (:TO** (COME*)
                                              (:MANNER (HERE*))
                                              (:FROM* (SOUTHEAST-ASIA*
                                                       (:UNKNOWN)
                                                       (:NAME)))))))))
                 (:MANNER (NOT*)))))
```

```
Stanford output:
(ROOT
  (S
   (NP (NNP NZ) (NNP First) (JJ associate)
       (NN immigration) (NN spokesman)
       (NNP Peter) (NNP Brown))
    (VP (VBD said)
      (SBAR
        (S
          (NP (DT the) (NN family))
          (VP (MD should) (RB not)
            (VP (VB be)
              (VP (VBN allowed)
                (S
                  (VP (TO to)
                    (VP (VB come)
                      (ADVP (RB here))
                      (PP (IN from)
                          (NP (NNP Southeast)
                              (NNP Asia)))))))))))
   (. .)))
```

```
Setence 8:

"Put simply, if Mr Zaoui wants to be with his family so badly,
then there is nothing preventing him from getting on a plane and
going to be with them today."
```

```
UGE output:
(COMMA* (:MS2
         (IF* (:MS1
```

```
              (WANTS* (:ACTOR (ZAOUI* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))))
               (:WHAT
                (:TO** (BE* (:WHAT ?R))
                       (:WITH* (FAMILY* (:NOUN) (:MODIFIER (HIS*))))
                       (:MANNER (SO*)) (:MANNER (BADLY*))))))
             (:MS2
              (THERE* (:NOUN)
               (:IS*
                (PREVENTING*
                 (:WHAT
                  (HIM* (:PNOUN)
                        (:FROM*
                         (GETTING*
                          (:WHAT ?R)
                          (:ON* (?R (:MODIFIER (A*) (PLANE*) (AND*) (GOING*)))))))))
                 (:MANNER (NOTHING*)))))))
       (:MS1 (PUT* (:ACTOR ?L) (:MANNER (SIMPLY*)))))


Stanford output:
(ROOT
  (S
    (S
      (VP (VB Put)
        (ADVP (RB simply))
        (, ,)
        (SBAR (IN if)
          (S
            (NP (NNP Mr) (NNP Zaoui))
            (VP (VBZ wants)
              (S
                (VP (TO to)
                  (VP (VB be)
                    (PP (IN with)
                      (NP (PRP$ his) (NN family)))
                    (ADVP (RB so) (RB badly)))))))))
    (, ,)
    (ADVP (RB then))
    (NP (EX there))
    (VP (VBZ is)
      (ADVP (NN nothing))
      (VP (VBG preventing)
        (NP (PRP him))
        (PP (IN from)
          (S
            (VP
              (VP (VBG getting)
                (PP (IN on)
                  (NP (DT a) (NN plane))))
              (CC and)
              (VP (VBG going)
                (S
                  (VP (TO to)
                    (VP (VB be)
                      (PP (IN with)
                        (NP (PRP them)))
                      (NP (NN today)))))))))))
   (. .)))



Setence 9:

"Green MP Keith Locke said the family were not a security risk and met
criteria for New Zealand's United Nations refugee intake."


UGE output:
(SAID* (:ACTOR (GREEN-*MP*-KEITH-LOCKE* (:UNKNOWN) (:NAME)))
       (:MS1
        (AND*
         (:MS2
          (MET* (:ACTOR ?L)
           (:WHAT
            (CRITERIA* (:NOUN)
             (:FOR*
              (INTAKE* (:NOUN) (:X-WORDS (UNITED-NATIONS*) (REFUGEE*) (INTAKE*)) (:NAME)
                       (:MODIFIER (|NEW-ZEALAND'S*|)))))))
         (:MS1
          (FAMILY* (:NOUN) (:MODIFIER (THE*))
                   (:WERE* (RISK* (:NOUN) (:X-WORDS (SECURITY*) (RISK*))
                                  (:MODIFIER (A*))) (:MANNER (NOT*)))))))))
```

```
Stanford output:
(ROOT
  (S
    (NP (NNP Green) (NNP MP) (NNP Keith) (NNP Locke))
    (VP
      (VP (VBD said)
        (SBAR
          (S
            (NP (DT the) (NN family))
            (VP (VBD were) (RB not)
              (NP (DT a) (NN security) (NN risk))))))
      (CC and)
      (VP (VBD met)
        (NP (NNS criteria))
        (PP (IN for)
          (NP
            (NP (NNP New) (NNP Zealand) (POS 's))
            (NNP United) (NNP Nations) (NN refugee) (NN intake)))))
   (. .)))
```

```
Sentence 10:

"But Mr Brown said Mr Zaoui had terrorism-related convictions
from European courts and his case had cost taxpayers millions."
```

```
UGE output:
(BUT* (:MS1
       (SAID* (:ACTOR (BROWN* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))))
        (:MS1
         (AND*
          (:MS2
           (HAD* (:ACTOR (CASE* (:NOUN) (:MODIFIER (HIS*))))
                 (:MS1 (COST* (:ACTOR ?L)
                              (:WHAT (MILLIONS* (:NOUN)
                                                (:X-WORDS (TAXPAYERS*)
                                                          (MILLIONS*))))))))
          (:MS1
           (HAD* (:ACTOR (ZAOUI* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))))
            (:WHAT
             (CONVICTIONS* (:NOUN) (:X-WORDS (TERRORISM-RELATED*) (CONVICTIONS*))
                           (:FROM* (COURTS* (:NOUN)
                                            (:X-WORDS (EUROPEAN*) (COURTS*))
                                            (:NAME)))))))))))
```

```
Stanford output:
(ROOT
  (S (CC But)
    (S
      (NP (NNP Mr) (NNP Brown))
      (VP (VBD said)
        (SBAR
          (S
            (NP (NNP Mr) (NNP Zaoui))
            (VP (VBD had)
              (NP
                (NP (JJ terrorism-related) (NNS convictions))
                (PP (IN from)
                  (NP (JJ European) (NNS courts)))))))))
    (CC and)
    (S
      (NP (PRP$ his) (NN case))
      (VP (VBD had)
        (VP (VBN cost)
          (NP (NNS taxpayers))
          (NP (NNS millions)))))
     (. .)))
```

```
Sentence 11:

"Mr Zaoui was convicted in Belgian and French courts on charges of association
with terrorists, but his lawyers say that those charges were groundless."
```

```
UGE output:
(BUT* (:MS2
       (SAY-THAT* (:ACTOR (LAWYERS* (:NOUN) (:MODIFIER (HIS*))))
                  (:MS1 (CHARGES* (:NOUN) (:MODIFIER (THOSE*))
                                  (:WERE* (?R (:MODIFIER (GROUNDLESS*))))))))
      (:MS1
       (ZAOUI* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))
        (:WAS*
         (CONVICTED* (:WHAT ?R)
          (:IN*
           (BELGIAN* (:UNKNOWN) (:NAME)
                     (AND* (COURTS* (:NOUN)
                                    (:X-WORDS (FRENCH*) (COURTS*)) (:NAME)))
                     (:ON* (CHARGES* (:NOUN)
                                     (:OF* (ASSOCIATION* (:NOUN)))
                                     (:WITH* (TERRORISTS* (:NOUN))))))))))))


Stanford output:
(ROOT
  (S
    (S
      (NP (NNP Mr) (NNP Zaoui))
      (VP (VBD was)
        (VP (VBN convicted)
          (PP (IN in)
            (NP (JJ Belgian)
              (CC and)
              (JJ French) (NNS courts)))
          (PP (IN on)
            (NP
              (NP (NNS charges))
              (PP (IN of)
                (NP (NN association)))))
          (PP (IN with)
            (NP (NNS terrorists))))))
    (, ,)
    (CC but)
    (S
      (NP (PRP$ his) (NNS lawyers))
      (VP (VBP say)
        (SBAR (IN that)
          (S
            (NP (DT those) (NNS charges))
            (VP (VBD were)
              (ADJP (JJ groundless)))))))
    (. .)))



Sentence 12:

"He was also expelled from Switzerland and left Malaysia after reports
the Algerian regime was seeking his extradition."


UGE output:
(AND* (:MS2
       (MALAYSIA* (:UNKNOWN) (:NAME) (:MODIFIER (LEFT*))
                  (:AFTER* (REPORTS* (:NOUN)))
                  (REGIME* (:NOUN) (:X-WORDS (ALGERIAN*) (REGIME*)) (:NAME)
                           (:MODIFIER (THE*)))
                  (:WAS* (SEEKING* (:WHAT (EXTRADITION* (:NOUN)
                                                        (:MODIFIER (HIS*))))))
      (:MS1 (HE* (:PNOUN) (:WAS*
                           (EXPELLED* (:WHAT ?R)
                                      (:FROM* (SWITZERLAND* (:UNKNOWN) (:NAME))))
                           (:MANNER (ALSO*))))))


Stanford output:
(ROOT
  (S
    (NP (PRP He))
    (VP (VBD was)
      (ADVP (RB also))
      (VP
        (VP (VBN expelled)
          (PP (IN from)
            (NP (NNP Switzerland))))
        (CC and)
        (VP (VBN left)
```

```
          (NP (NNP Malaysia))
          (PP (IN after)
            (NP
              (NP (NNS reports))
              (SBAR
                (S
                  (NP (DT the) (NNP Algerian) (NN regime))
                  (VP (VBD was)
                    (VP (VBG seeking)
                      (NP (PRP$ his) (NN extradition))))))))))
    (. .)))
```

```
Sentence 13:

"Mr Brown did agree with Mr Locke that the case had dragged on too long."


UGE output:
(DID* (:ACTOR (BROWN* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))))
      (:MS1
       (AGREE* (:ACTOR ?L)
        (:WITH*
         (LOCKE* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))
          (:THAT*
           (:MS1
            (HAD* (:ACTOR (CASE* (:NOUN) (:MODIFIER (THE*))))
                  (:WHAT (DRAGGED* (:WHAT ?R)
                                   (:ON* (LONG* (:NOUN)))
                                   (:MANNER (TOO*)))))))))))


Stanford output:
(ROOT
  (S
    (NP (NNP Mr) (NNP Brown))
    (VP (VBD did)
      (VP (VB agree)
        (PP (IN with)
          (NP (NNP Mr) (NNP Locke)))
        (SBAR (IN that)
          (S
            (NP (DT the) (NN case))
            (VP (VBD had)
              (VP (VBN dragged)
                (PP (IN on)
                  (ADJP (RB too) (JJ long)))))))))
   (. .)))
```

```
Sentence 14:

"When Mr Zaoui, once elected an MP in Algeria, came to New Zealand,
he sought refugee status on the grounds
he would be tortured or killed if he was sent back to his homeland."


UGE output:
(WHEN* (:MS1
        (CAME*
         (:ACTOR
          (ZAOUI* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))
           (COMMA*
            (ELECTED* (:WHAT (*MP** (:UNKNOWN) (:NAME) (:MODIFIER (AN*))
                                    (:IN* (ALGERIA* (:UNKNOWN) (:NAME)))))
             (:MANNER (ONCE*))))))
          (:TO* (NEW-ZEALAND* (:UNKNOWN) (:NAME)))))
       (:MS2
        (SOUGHT* (:ACTOR (HE* (:PNOUN)))
         (:WHAT
          (STATUS* (:NOUN) (:X-WORDS (REFUGEE*) (STATUS*))
           (:ON*
            (GROUNDS* (:NOUN) (:MODIFIER (THE*))
             (:CLAUSE
              (IF* (:MS2
                    (HE* (:PNOUN) (:WAS* (SENT* (:WHAT ?R) (:MANNER (BACK*))
                                                (:TO* (HOMELAND* (:NOUN) (:MODIFIER (HIS*))))))))
                   (:MS1
                    (WOULD* (:ACTOR (HE* (:PNOUN)))
```

```
                          (:MS1 (BE* (:ACTOR ?L)
                                     (:WHAT (KILLED* (:WHAT ?R)
                                                     (OR* (TORTURED* (:WHAT ?R))))))))))))))))))))


Stanford output:
(ROOT
  (S
    (SBAR
      (WHADVP (WRB When))
      (S
        (NP
          (NP (NNP Mr) (NNP Zaoui))
          (, ,)
          (VP
            (ADVP (RB once))
            (VBN elected)
            (NP
              (NP (DT an) (NNP MP))
              (PP (IN in)
                (NP (NNP Algeria)))))
          (, ,))
        (VP (VBD came)
          (PP (TO to)
            (NP (NNP New) (NNP Zealand))))))
    (, ,)
    (NP (PRP he))
    (VP
      (VP (VBD sought)
        (NP (NN refugee) (NN status))
        (PP (IN on)
          (NP
            (NP (DT the) (NNS grounds))
            (SBAR
              (S
                (NP (PRP he))
                (VP (MD would)
                  (VP (VB be)
                    (ADJP (JJ tortured)))))))))
      (CC or)
      (VP (VBD killed)
        (SBAR (IN if)
          (S
            (NP (PRP he))
            (VP (VBD was)
              (VP (VBN sent)
                (PRT (RP back))
                (PP (TO to)
                  (NP (PRP$ his) (NN homeland)))))))))
   (. .)))



Sentence 15:

"He spent almost two years in prison waiting for his case to be decided as
he fought an SIS security risk certificate and moves to expel him from the country."


UGE output:
(AS* (:MS2
       (AND*
        (:MS2 (MOVES* (:ACTOR ?L)
                      (:TO** (EXPEL*
                              (:WHAT (HIM* (:PNOUN)
                                           (:FROM* (COUNTRY* (:NOUN)
                                                             (:MODIFIER (THE*)))))))))
        (:MS1
         (FOUGHT* (:ACTOR (HE* (:PNOUN)))
                  (:WHAT (CERTIFICATE* (:NOUN)
                                       (:X-WORDS (*SIS**)
                                                 (SECURITY*)
                                                 (RISK*)
                                                 (CERTIFICATE*))
                                       (:NAME) (:MODIFIER (AN*)))))))
      (:MS1
       (SPENT* (:ACTOR (HE* (:PNOUN)))
        (:WHAT
         (YEARS* (:NOUN) (:MODIFIER (TWO*))
          (:IN*
           (PRISON* (:NOUN)
                    (WAITING* (:WHAT ?R)
```

```
                            (:FOR* (CASE* (:NOUN) (:MODIFIER (HIS*))
                                          (:TO** (BE*
                                                  (:WHAT (DECIDED*
                                                          (:WHAT ?R))))))))))))
       (:MANNER (ALMOST*)))))


Stanford output:
(ROOT
  (S
    (NP (PRP He))
    (VP (VBD spent)
      (ADVP (RB almost)
        (NP (CD two) (NNS years)))
      (PP (IN in)
        (NP
          (NP (NN prison))
          (VP (VBG waiting)
            (PP (IN for)
              (NP (PRP$ his) (NN case)
                (S
                  (VP (TO to)
                    (VP (VB be)
                      (VP (VBN decided)
                        (SBAR (IN as)
                          (S
                            (NP (PRP he))
                            (VP
                              (VP (VBD fought)
                                  (NP (DT an)
                                      (NNP SIS)
                                      (NN security)
                                      (NN risk)
                                      (NN certificate)))
                              (CC and)
                              (VP (VBZ moves)
                                (S
                                  (VP (TO to)
                                    (VP (VB expel)
                                      (NP (PRP him))
                                      (PP (IN from)
                                          (NP (DT the)
                                              (NN country)))))
                                 ))))))))))))
    (. .)))



Sentence 16:

"Mr Zaoui was released on bail in December 2004 after a Supreme Court hearing,
and has since lived with the Catholic community in the Dominican Priory
in Auckland awaiting a hearing."


UGE output:
(AND* (:MS2
       (HAS* (:ACTOR ?L)
        (:WHAT
         (LIVED* (:WHAT ?R)
          (:WITH*
           (COMMUNITY* (:NOUN) (:X-WORDS (CATHOLIC*) (COMMUNITY*)) (:NAME) (:MODIFIER (THE*))
            (:IN*
             (DOMINICAN-PRIORY* (:UNKNOWN) (:NAME)
                                (:MODIFIER (THE*))
                                (:IN* (AUCKLAND* (:UNKNOWN) (:NAME)
                                                 (AWAITING*
                                                  (:WHAT (HEARING* (:NOUN) (:MODIFIER (A*)))))))))))))
        (:MANNER (SINCE*))))
      (:MS1
       (ZAOUI* (:UNKNOWN) (:NAME) (:MODIFIER (MR*))
        (:WAS*
         (RELEASED* (:WHAT ?R)
          (:ON*
           (BAIL* (:NOUN)
            (:IN*
             (DECEMBER-2004* (:UNKNOWN) (:NAME)
                             (:AFTER*
                              (HEARING* (:NOUN)
                                        (:X-WORDS (SUPREME-COURT*) (HEARING*)) (:NAME)
                                        (:MODIFIER (A*)))))))))))))
```

```
Stanford output:
(ROOT
  (S
    (NP (NNP Mr) (NNP Zaoui))
    (VP
      (VP (VBD was)
        (VP (VBN released)
          (PP (IN on)
            (NP
              (NP (NN bail))
              (PP (IN in)
                (NP (NNP December) (CD 2004)))))
          (PP (IN after)
            (NP (DT a) (NNP Supreme) (NNP Court) (NN hearing)))))
      (, ,)
      (CC and)
      (VP (VBZ has)
        (ADVP (RB since))
        (VP (VBN lived)
          (PP (IN with)
            (NP
              (NP (DT the) (JJ Catholic) (NN community))
              (PP (IN in)
                (NP
                  (NP (DT the) (NNP Dominican) (NNP Priory))
                  (PP (IN in)
                    (NP (NNP Auckland)))))))
          (S
            (VP (VBG awaiting)
              (NP (DT a) (NN hearing)))))))
   (. .)))



Sentence 17:

"The hearing to review the security risk certificate was due to be held last August
but will not now go ahead until between June and August this year."


UGE output:
(BUT* (:MS2
       (WILL* (:ACTOR ?L)
        (:MS1
         (GO* (:ACTOR ?L) (:MANNER (AHEAD*))
          (:UNTIL*
           (:BETWEEN* (JUNE* (:UNKNOWN) (:NAME)
                             (AND* (AUGUST* (:UNKNOWN) (:NAME)
                                            (:TIME (THIS* (YEAR* (:NOUN)))))))))))
        (:MANNER (NOT*)) (:MANNER (NOW*))))
      (:MS1
       (HEARING* (:NOUN) (:MODIFIER (THE*))
                 (:TO** (REVIEW*
                         (:WHAT (CERTIFICATE* (:NOUN) (:X-WORDS (SECURITY*) (RISK*)
                                                                (CERTIFICATE*))
                                              (:MODIFIER (THE*))))))
                 (:WAS* (DUE* (:NOUN) (:TO**
                                       (BE* (:WHAT (HELD*
                                                    (:WHAT (AUGUST* (:UNKNOWN) (:NAME)))
                                                    (:MANNER (LAST*)))))))))))


Stanford output:
(ROOT
  (S
    (NP (DT The) (NN hearing)
      (S
        (VP (TO to)
          (VP (VB review)
              (NP (DT the) (NN security) (NN risk)
                  (NN certificate))))))
    (VP
      (VP (VBD was)
        (ADJP (JJ due)
          (S
            (VP (TO to)
              (VP (VB be)
                (VP (VBN held)
                  (NP (JJ last) (NNP August)))))))))
      (CC but)
      (VP (MD will) (RB not)
```

```
    (VP
      (ADVP (RB now))
      (VB go)
      (NP
        (ADVP (RB ahead)
          (PP (IN until)
            (PP (IN between)
              (NP (NNP June)
                (CC and)
                (NNP August)))))
        (DT this) (NN year)))))
(. .)))
```

References

Abney, S. P. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research*, *18*, 129-144. Available from `http://dx.doi.org/10.1007/BF01069051` (10.1007/BF01069051) 5, 35, 49

Altmann, G. (1988). Ambiguity, parsing strategies, and computational models. *Language and Cognitive Processes*, *13*, 73-97. 49

Asch, V., & Daelemans, W. (2009). Prepositional phrase attachment in shallow parsing. In *In proceedings of the 7th international conference on recent advances in natural language processing* (pp. 12–17). 63

Baker, C. (1978). *Introduction to generative transformational syntax*. Englewood Cliffs, NJ: Prentice Hall. 3, 18

Berwick, R., & Chomsky, N. (2008). 'poverty of the stimulus' revisited: Recent challenges reconsidered. In *Proceedings of the 30th annual conference of the cognitive science society* (p. 383). Washinton D.C., USA. 19, 23

Bikel, M. D. (2000). A statistical model for parsing and word-sense disambiguation. In *Proceedings of the 2000 joint sigdat conference on empirical methods in natural language processing and very large corpora: held in conjunction with the 38th annual meeting of the association for computational linguistics - volume 13* (pp. 155–163). 49

Bohannon, J., & Stanowicz, L. (1988). The issue of negative evidence: Adult responses to children's language errors. *Developmental Psychology*, *24*, 684-689. 1

Braine, M. D. S. (1971). The ontogenesis of grammar: a theoretical symposium. In D. I. Slobin (Ed.), (chap. On two types of models of the internalization

of grammars). New York: Academic Press. 2

Brent, M. (1997). Computational approaches to language acquisition. In M. R. Brent (Ed.), (p. 1-38). Cambridge, MA: MIT Press. 26

Brown, R., & Hanlon, C. (1970). Cognition and the development of language. In J. R. Hayes (Ed.), (chap. Derivational complexity and order of acquisition in child speech). New York: Wiley. 2

Bruner, J. (1983). *Child's talk: Learning to use language.* New York: Norton. 44

Buchheit, P. (1993). Infant: a modular approach to natural language processing. In *Proceedings of the 1993 acm conference on computer science* (pp. 410–417). 46

Buch-Kromann, M. (2006). *Discontinuous grammar. a dependency-based model of human parsing and language learning.* Unpublished doctoral dissertation, Copenhagen Business School. 53

Carroll, J., Briscoe, T., & Sanfilippo, A. (1998). Parser evaluation: a survey and a new proposal. In *In proc.* LREC98. 116, 118

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press. 3, 18, 21, 48

Chomsky, N. (1968). *Language and mind.* New York: Harcourt, Brace, Jovanovitch. 3, 18, 21, 48

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences*, *5*, 82-88. 33

Clark, A., & Eyraud, R. (2006). Learning auxiliary fronting with grammatical inference. In *Proceedings of the 28th annual conference of the cognitive science society* (p. 125-132). 4, 22

Collins, M. (1999). *Head-driven statistical models for natural language parsing.*

Ph.d.thesis, University of Pennsylvania. 115

Cook, V., & Newson, M. (2007). *Chomsky's universal grammar: an introduction.* Oxford: Blackwell. 21

Costa, F., Frasconi, P., Lombardo, V., & Soda, G. (2003). Towards incremental parsing of natural language using recursive neural networks. *Applied Intelligence*, *19*, 9–25. 49

Crain, S., & Nakayama, M. (1987). Structure dependency in grammar formation. *Language*, *63*, 522-543. 19

Crain, S., & Pietroski, P. (2001). Nature, nurture and universal grammar. *Linguistics and Philosophy*, *24*, 139-186. 19

Crystal, D. (1970). *Prosodic systems and language acquisition* (Prosodic Feature Analysis / Analyse des Faits Prosodiques ed.; G. F. P.R. Leon & A. Rigault, Eds.). Paris. 45

Daniel Cer, D. J., Marie-Catherine de Marneffe, & Manning, C. (2010, may). Parsing to stanford dependencies: Trade-offs between speed and accuracy. In B. M. J. M. J. O. S. P. M. R. D. T. Nicoletta Calzolari (Conference Chair) Khalid Choukri (Ed.), *Proceedings of the seventh conference on international language resources and evaluation (lrec'10).* Valletta, Malta: European Language Resources Association (ELRA). 115

Debusmann, R. (2000). *An introduction to dependency grammar.* 52

Demberg, V., & Keller, F. (2009). A computational model of prediction in human parsing: Unifying locality and surprisal effects. In *In proceedings of the 29th meeting of the cognitive science society (cogsci-09)* (pp. 1888–1893). 35

Demetras, M., Post, K., & Snow, C. (1986). Feedback to first language learners: The role of repetitions and clarification questions. *Journal of Child*

Language, *13*, 275-292. 1

Drescher, G. (1991). *Made-up minds: a constructivist approach to artificial intelligence*. Cambridge, MA: MIT Press. 43

Eisner, J. M. (1998). Three new probabilistic models for dependency parsing: ans exploration. In *In proceedings of the 16th international conference on computational linguistics.* 53

Fodor, J. (1998). Parsing to learn. *Journal of Psycholinguistic Research*, *27*, 339–374. 23

Foraker, R. T. K. N. P. A., S., & Tenenbaum, J. (2009). Indirect evidence and the poverty of the stimulus: The case of anaphoric one. *Cognitive Science*, *33*, 287-300. 3, 4, 22

Gaifman, H. (1965). Dependency systems and phrase-structure systems. *Inforamation and Control*, *8*(3), 304–337. 52

Gibson, E., & Wexler, K. (1994). Triggers. *Language Acquisition*, *25*, 407–454. 23

Gobet, F., Freudenthal, D., & Pine, J. (2004). Modelling syntactic development in a crosslinguistic context. In *In proceedings of the first workshop on psycho-computational models of human language acquisition* (pp. 53–60). 33

Grinberg, L. J., D., & Sleator, D. (1995). A robust parsing algorithm for link grammars. In *In proceedings of the fourth international workshop on parsing technologies.* 56

Hays, D. G. (1964). Dependency theory: A formalism and some observations. *Language*, *40*, 511–525. 52

Hellwig, P. (1986). Dependency unification grammar. In *Proceedings of the 11th coference on computational linguistics* (pp. 195–198). Morristown, NJ,

USA: Association for Computational Linguistics. 48, 50

Hindle, D., & Rooth, M. (1993). Structural ambiguity and lexical relations. *Computational Linguistics*, *19*(1), 103–120. 63

Hirsh-Pasek, K., & Golinkoff, R. (1996). *The origin of grammar: Evidence from early language comprehension*. Cambridge, MA: MIT Press. 1, 3, 5, 22, 35

Hirsh-Pasek, K., Treiman, R., & Schneiderman, M. (1984). Brown and hanlon revisited: Mothers' sensitivity to ungrammatical forms. *Journal of Child Language*, *11*, 81-88. 1

Hohle, B. (2009). Bootstrapping mechanisms in first language acquisition. *Linguistics*, *47*, 359–382. 23

Holt, E. B. (1931). *Animal drive and the learning process*. New York: H. Holt. 2

Hornstein, N., & Lightfoot, D. (1981). *Explanation in linguistics: The logical problem of language acquisition*. London: Longma. 18

Hudson, R. A. (2007). English dialect syntax in word grammar. *English Language and Linguistics*, *1*, 383-405. 48, 50

Ingram, D. (1989). *First language acquisition: Method, description and explanation*. Cambridge: Cambridge University Press. 17

Jarvinen, T., & Tapanainen, P. (1997). Dependency parser demo. In *Proceedings of the fifth conference on applied natural language processing: Descriptions of system demonstrations and videos* (pp. 9–10). 50

Kaplan, M. R. (1971). Augmented transition networks as psychological models of sentence comprehension. In *Proceedings of the 2nd international joint conference on artificial intelligence* (pp. 429–440). 49

Kbler, S. (2005). How do treebank annotation schemes influence parsing results? or how not to compare apples and oranges. In *In proceedings of ranlp 2005.*

Borovets, Bulgaria. 116

Kimball, J. (1973). *The formal theory of grammar*. Englewood Cliffs, NJ: Prentice Hall. 18

Lasnik, H., & Uriagereka, J. (2002). On the poverty of the challenge. *The Linguistic Review*, *19*, 147-150. 3, 22

Lawrence, S., & Margolis, E. (2001). The poverty of the stimulus argument. *British Journal of Philosophy Science*, *52*, 217-276. 3, 22

Legate, J., & Yang, C. (2002). Empirical re-assessment of stimulus poverty arguments. *Linguistic Review*, *19*, 151-162. 3, 19, 22

Lidz, J., & Gleitman, L. (2004). Yes, we still need universal grammar. *Cognition*, *94*, 85-93. 3, 22

Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: experimental evidence for syntactic structure at 18 months. *Cognition*, *89*(3), B65-B73. 19

Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a large annotated corpus of english: the penn treebank. *Comput. Linguistics*, *19*(2), 313–330. 115, 116

McDonald, R. (2006). *Discriminative learning and spanning tree algorithms for dependency parsing*. Unpublished doctoral dissertation, University of Pennsylvania. 53

Melcuk, I. (1988). *Dependency syntax: Theory and practice*. Albany, NY: State Univ. Press of New York. 50

Nivre, J. (2005). *Dependency grammar and dependency parsing* (Technical Report MSI report 05133). Vxj University: School of Mathematics and Systems Engineering. 48, 50, 52

Niyogi, P., & Berwick, R. (1997). Computational approaches to language acquisition. In M. R. Brent (Ed.), (chap. A language learning model for finite parameter spaces). Cambridge, MA: MIT Press. 23

Penner, S. (1987). Parental responses to grammatical and ungrammatical child utterances. *Child Development*, *58*, 376-384. 1

Piaget, J. (1969). *The psychology of the child*. 42

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press. 1, 2, 16, 22, 23, 24, 33

Pinker, S. (1987). Mechanisms of language acquisition. In B. M. whinney (Ed.), (p. 399-439). Hillsdale, N.J.: Erlbaum. 1, 2, 23

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press. 1, 2, 3, 5, 16, 18, 20, 23, 25

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*, 73–193. 23

Pullum, G., & Scholz, B. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, *19*, 9-50. 3, 22

Reali, F., & Christiansen, M. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science*, *29*, 1007-1028. 3, 4, 22

Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes*, *13*, 129-191. 4, 22, 31, 32

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*, 425–

469. 33

Regier, T., & Gahl, S. (2004). Learning the unlearnable: the role of missing evidence. *Language and Cognitive Processes*. 4, 22

Resnik, P. (1992). *Left-corner parsing and psychological plausibility.* 49

Sakas, W., & Fodor, J. (2001). Language acquisition and learnability. In S.Bertolo (Ed.), (pp. 172–233). Cambridge, UK: CUP. 23

Sakas, W., & Fodor, J. (2003). *Slightly ambiguous triggers for syntactic parameter setting.* Poster presented at AMLaP-2003, Glasgow, Scotland. 23

Sampson, G. (1993). The susanne corpus. *ICAME Journal*, *17*, 125127. 116

Sampson, G. (2002). Exploring the richness of the stimulus. *The Linguistic Review*, *19*, 73-104. 3, 22

Schuler, W., AbdelRahman, S., Miller, T., & Schwartz, L. (2010). Broad-coverage parsing using human-like memory constraints. *Computational Linguistics*, *36*(1), 1-30. 49

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, *23*, 569–588. 22

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1-2), 39 - 91. Available from `http://www.sciencedirect.com/science/article/B6T24-3W268SR-2/2/49560a6edf580e187c05e2f293bda994` (Compositional Language Acquisition) 16, 23, 24, 26, 27, 29, 30

Skinner, B. (1957). *Verbal behavior.* NY: Prentice Hall. 2

Sleator, D., & Temperley, D. (1993). Parsing english with a link grammar. In *In third international workshop on parsing technologies.* 5, 48, 55, 56, 57

Small, C. G., S., & Shastri, L. (1982). Toward connectionist parsing. In *In proceedings of the national conference on artificial intelligence* (pp. 247–250). 49

Steedman, M. (1999). *Categorial grammar* (MIT Encyclopedia of Cognitive Science ed.; R. A. W. . F. C. Keil, Ed.). Cambridge, MA: MIT Press. 53, 54, 55

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, *74*(3), 209 - 253. 26

Vodenski, P. (2009). *Relational grammar in computational psycholinguistics.* 48

Wood, M. M. (1993). *Categorial grammars.* Routledge. 5, 48, 55

Yang, C. D. (1999). A selectionist theory of language acquisition. In *Proceedings of the 37th annual meeting of the association for computational linguistics on computational linguistics* (pp. 429–435). 46

Yeap, W. (2005a). A new gofai theory: How language works. In *International lisp conference.* Stanford, CA. 1, 4, 9, 34, 59, 60

Yeap, W. (2005b). Semantics parsing re-visited/how a tadpole could become a frog. In *Proceedings of the 2nd language and technology conference.* Poznan, Poland. 1, 4, 9, 34

Zhao, S., & Lin, D. (2004). A nearest-neighbor method for resolving pp-attachment ambiguity. In *In proceedings of ijcnlp-04.* 63