An English-Learning Application based on extended link grammar formalism

Yuan-Kai Wang
Department of Electronic Engineering Catholic Fu Jen University
E-mail: ykwang@mercury.ee.fju.edu.tw

Ying-Hong Wang
Department of Computer Science and Information Engineering TamKang University Tamsui, Taipei, Taiwan, R.O.C.
E-mail: inhon@mail.tku.edu.tw

Hsiu-Ping Lin
Department of Computer Science and Information Engineering TamKang University Tamsui, Taipei, Taiwan, R.O.C.
E-mail: g7190347@tkgis.tku.edu.tw

Chin-Feng Chao
Department of Computer Science and Information Engineering TamKang University Tamsui, Taipei, Taiwan, R.O.C.
E-mail: g7190263@tkgis.tku.edu.tw

Abstract

In this paper a new method for English learning of Chinese students based on link grammar formalism is presented. Link grammar is a context-free formalism for the description of natural language. The motivation that we develop an English learning system is to estimate what kinds of mistakes that college students will probably make by parsing students' writing English with link grammar. The system will detect students' mistakes whenever they make and give some proper suggestions. By this way, students can not only know why they make a mistake but also learn how to improve their English by taking our suggestions. Furthermore, teachers can use the system to change their teaching strategy by the statistic data of parsing results.

Keywords: *English learning, link grammar, formula form, disjunctive form, Robust parsing*

1. Introduction

Learning English is always a big problem for Chinese students. Some experts suggest that the best way to improve English is to get used to English environment and practice English at any time. This paper provides a good method for Chinese students to learn English by link grammar. They can also get some suggestions when they make a mistake. The system focuses on the application of link grammar and uses it to build a tutoring system for English learning.

The best way to explain the concepts of link grammar is to discuss an example of ***linkage***. Figure 1 shows how a linkage is formed by link grammar.

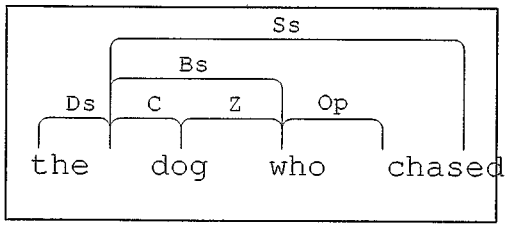


Figure 1

The sentence consists of six words. We may think the linkage as a graph. And the words can be treated as vertices, which are connected by arcs with labels on them, so that the graph is connected and planar. The labeled arcs are called ***links***. However, there are some basic rules to form a valid linkage and will be discussed in the next section.

In Section 2 the link grammar is described more formally and the notation and terminology used throughout the paper will be explained. Section 3 discusses our parsing strategy, which may give some tolerance to students' mistakes, and how fault-tolerance links are used to link some words that are used incorrectly. Our suggestion approach is explained in Section 4. In the last section we summarize the examples we parsed and discuss future work. We also conclude the experiment results and improve the tutoring strategy.

2. Notation and terminology

In this section the concept and notation of link grammar will be described. Link grammar is a context-free formalism for the description of natural language [3]. The figure below represents one of the parses produced by link grammar when the students input the sentence " The big snake the black cat chased bit Mary." As mentioned previously, the labeled arcs connecting words to other words on their left or right are called ***links***. A valid parse is called a ***linkage***.

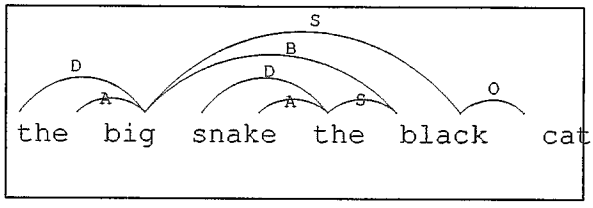


Figure 2

2.1 Meta-rules :

The link grammar dictionary consists of a collection of entries, each of which defines the linking requirements of one or more words. These requirements are specified by means of a formula of connectors combined by the binary associative operators & and *or*. Presidence is specified by means of parentheses. Without loss of generality we may assume that a connector is simply a character string ending in + or - [1].

A sequence of words is a sentence of the language defined by the grammar if there exists a way to draw links among the words so as to satisfy each word's formula, and the following meta-rules:

(1) **Planarity**: The links are drawn above the sentence and do not cross.
(2) **Connectivity**: The links suffice to connect all the words of the sequence together.
(3) **Ordering**: When the connectors of a formula are traversed from left to right, the words to which they connect proceed from near to far. In other words, consider a word, and consider two links connecting that word to words to its left. The link connecting the nearer word (the shorter link) must satisfy a connector appearing to the left (in the formula) of that of the other word. Similarly, a link to the right must satisfy a connector to the left (in the formula) of a longer link to the right.
(4) **Exclusion**: No two links may connect the same pair of words [4].

2.2 Robust parsing algorithm :

As shown in figure 3, the robust algorithm uses the notion of a null link to allow a connection between any pair of adjacent words, regardless of their dictionary definitions.

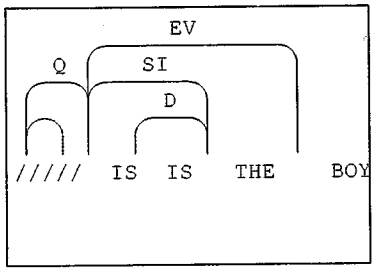


Figure 3

The algorithm proceeds by making three dynamic programming passes through the sentence. In the first pass, the input is parsed using the original algorithm, which enforces the constraints on links to ensure grammaticality. In the second pass, the total cost of each substring of words is computed, where cost is determined by the number of null links necessary to parse the substring. The final pass counts the total number of parses with minimal cost. Memorization together with pruning techniques enable the algorithm to run efficiently, with theoretic time complexity of $O(n^3)$ for an input of n words [5].

2.3 Disjunctive form:

In disjunctive form, each word of the grammar has a set of disjuncts associated with it. Each disjunct corresponds to one particular way of satisfying the requirements of a word. A disjunct consists of two ordered lists of connector names: the left list and the right list. The left list contains connectors that connect to the left of the current word (those connectors end in -), and the right list contains connectors that connect to the right of the current word. A disjunct will be denoted:

$$((L_1,L_2,\ldots,L_m)(R_n, R_{n-1},\ldots,R_1))$$

It is easy to see how to translate a link grammar in disjunctive form to one in standard form. This can be done simply by rewriting each disjunct as:

$$(L_1\&L_2\&\ldots\&L_m\&R_1\& R_2\&\ldots\&R_n)$$

It is also easy to translate a formula into a set of disjuncts. This is done by enumerating all ways that the formula can be satisfied. For example, the formula:

(A- or ( )) & D- & (B+ or ( )) & ((O- or S+))

corresponds to the following eight disjuncts [2]:

((A, D)	(S, B))
((A, D, O)	(B))
((A, D)	(S))
((A, D, O)	( ))
((D)	(S, B))
((D, O)	(B))
((D)	(S))
((D, O)	( ))

Figure 4

3. The Fault Tolerance Parsing

At the beginning of this section, the flow-chart of our system is shown above.

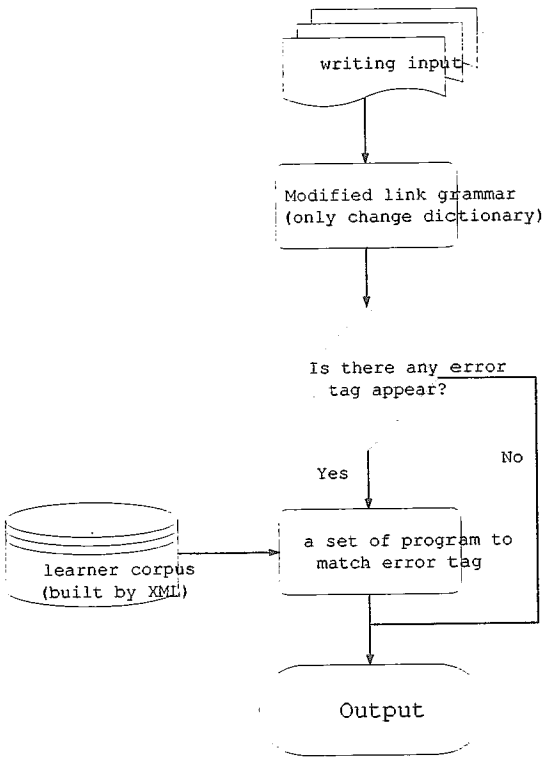


Figure 5

The writing input comes from the writing English of college students. And the modified link grammar will be described in this section. The learner corpus contains the information that we want to show to the students and will be fetched by the program as outputs.

The system that we propose only changes the dictionary of link grammar and the formula form of each word. By this way, the parsing result makes fault tolerance possible.

In the section, this paper discusses two main problems: nouns and verbs. All the examples appearing in this section show that the upper one is correct and the lower one, which proceeds with a star symbol, makes some mistakes.

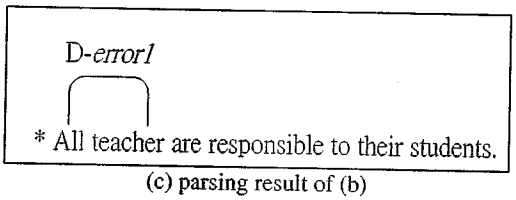
(1) Nouns:

a. Single and plural: As we know, determiners "a" and "an" must be followed by a single noun. Besides, determiners "all" and "many" must be followed by a plural noun. However, some determiners, such as "the", can be followed by both single and plural nouns. If the students make a wrong usage of single and plural, we define it as type 1 mistake.

All teachers are responsible to their students.

(a) Correct sentence

*All teacher are responsible to their students.
(b) Incorrect sentence



(c) parsing result of (b)

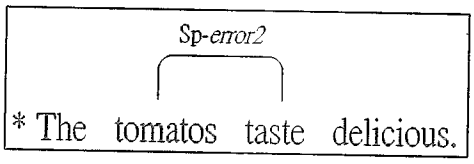
Figure 5

b. The mistakes in plural: The plural nouns may have several different types. Some end with *–s*, some end with *–es*, and some may be irregular. The mistakes of plural nouns are defined as type 2. In this type, several new categories of the dictionary are built and how to build them will be discussed in the next chapter.

The tomatoes taste delicious.

(a) Correct sentence

*The tomatos taste delicious.
(b) Incorrect sentence



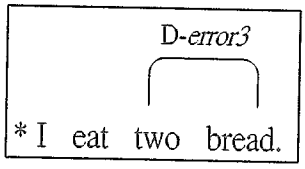
(c) Parsing result of (b)
Figure 6

c. Nouns that are mass: Some nouns that are mass can not count directly. Such nouns include soap, chalk, furniture, bread, information…etc. If it is necessary to count them, several types of unit can be added in front of them. Mistakes in mass noun are defined as error type 3.

I eat two loaves of bread.

(a) Correct sentence

*I eat two bread.
(b) Incorrect sentence



(c) Parsing result of (b)

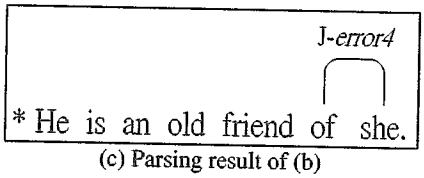
Figure 7

d. The mistakes of pronoun: The students usually make some mistakes in pronoun, such as nominative pronoun and accusative pronoun. Such mistakes in pronoun are defined as error type 4.

He is an old friend of hers.

(a) Correct sentence

*He is an old friend of she.
(b) Incorrect sentence



(c) Parsing result of (b)

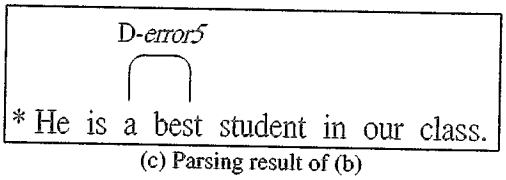
Figure 8

e. The mistakes of determiners: Sometimes, the determiners "a" and "the" are difficult to be distinguished. Some nouns should combine with special determiners. Besides, adjectives and adverbs have also the same features.

He is the best student in our class.

(a) correct sentence

*He is a best student in our class.
(b) Incorrect sentence



(c) Parsing result of (b)
Figure 9

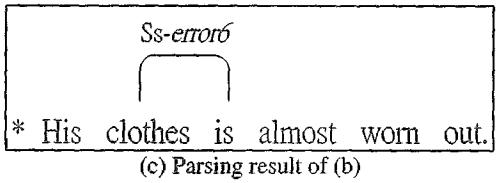
(2) Verbs:

a. Single and plural: If the students make a wrong usage of single and plural of verbs, we set it as type 6 mistake. Type 6 mistake is mostly related to type 1 mistake. Since one makes a mistake of single and plural in noun, the followed verb will be in a big trouble.

His clothes are almost worn out.

(a) Correct sentence

*His clothes is almost worn out.
(b) Incorrect sentence



(c) Parsing result of (b)

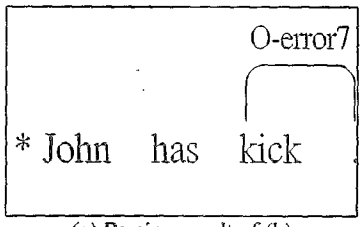
Figure 10

b. Transitive verbs should succeed an object: Transitive verbs have an *O+* connector in the dictionary, and nouns have an *O-* connector [1]. If a transitive verb is used in a sentence, there is no object follow it. This will be defined as type 9 error.

John has kicked the dog.

(a) Correct sentence

*John has kick.
(b) Incorrect sentence



(c) Parsing result of (b)

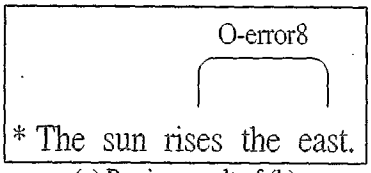
Figure 11

c. Intransitive verb: Some verbs that have not an object are called intransitive verbs, such as *be*, and *rise*. Type 10 error is defined as intransitive verbs followed by an object.

The sun rises from the east.

(a) Correct sentence

*The sun rises the east.
(b) Incorrect sentence



(c) Parsing result of (b)

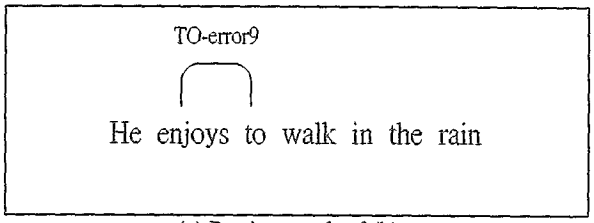
Figure 12

d. Some special verbs: Some verbs must be followed by infinitive *–to*, some must be followed by gerund, and some must be followed by participle. There are some verbs that can accept several different types.

We enjoy walking in the rain.

(a) Correct sentence

*We enjoy to walk in the rain.
(b) Incorrect sentence



(c) Parsing result of (b)

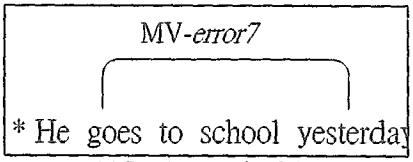
Figure 13

e. The mistake usage in tense: The tense is always a big problem for Chinese students. Therefore, the grammar of English always discusses about the tense. Since the problem of tense consists semantic concept and we leave it discussed in section 5. Of course, this section will not discuss all the tenses. Instead, a simple example is proposed to point out what kinds of mistakes the students may make.

He went to school yesterday.

(a) Correct sentence

*He goes to school yesterday.
(b) Incorrect sentence



(c) Parsing result of (b)

Figure 14

4. Error detection and suggestion

In this chapter, some of the system will be described in detail. We will carry out how we modify the original link grammar and make error detection possible. Some problems and difficulties that we suffer in this period will be mentioned later.
As discussed in section 3, if the students input a sentence and make type 1 mistake, such as the following:

Input: all teacher are responsible to their students

Output will be presented as follows:

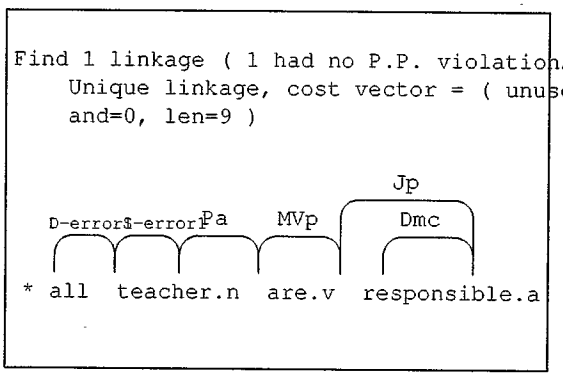


Figure 15

The original link should be **Dmc** and **Spx** if the word "teacher" is changed into "teachers". As the link **D-error1** and **S-error1** appear, the students make a mistake of single and plural in nouns. Such an information is obtained by *error1*.
How the extra link can be obtained? We modify the formula of the word "teacher" in the dictionary so that it can have more flexibility to link to the plural words. And with such an alteration, the system will make fault tolerance possible.
As mentioned in last section, another important method is to construct additional categories. With the same reason the sentence can make more flexibility to compose a valid linkage. All we have to do is collect more examples and conclude the possible mistakes that students will make. This field seems more close to the morphology.
However, the system will show out two examples when the students make any type of mistakes. One is the correct one and the other is the wrong one. Such a review can tell the students right from wrong. Of course, more modification to complete the integrality of the system is necessary.

5. Conclusion and the future work

The system is constructed based on the link grammar. As more and more researches quote from link grammar, its potential seems quite powerful. Now the online dictionary is constructed and the user can parse a sentence in the Web. Of course, it does not contain such an error detection function. However, there still consists some knowledge-based culture backgrounds of language itself. This provides us more topics of research in the future. More and more inspirations about English learning will be proposed by link grammar sooner or later. This paper shows an application of link grammar for English learning. The future work is to combine the system with a tutoring system so that the teachers can get the statistics of students' faults. To sum up, with this system, the students can check out their mistakes to improve their English and the teachers can get the related information to provide them for modifying teaching strategies.

Reference

[1] D. D. K. Sleator and D. Temperley. **Parsing English with a link grammar**. Technical Report CMU-CS-91-196, School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, 1991.
[2] Dennis Grinberg John Lafferty Daniel Sleator **A Robust Parsing Algorithm For Link Grammars** August 1995 CMU-CS-95-125 School of Computer Science Carnegie Mellon University Pittsburgh, PA 15213
[3] John Lafferty Daniel Sleator Davy Temperley **Grammatical Trigrams: A Probabilistic Model of Link Grammar** Presented to the 1992 AAAI Fall Symposium on Probabilistic Approaches to Natural Language.
[4] Lee, Y.-H., Evens, M., Michael, J. A., and Rovick, A. A. 1990. **Spelling Correction for an intelligent tutoring system.** Tech. Rep., Dept. of Computer Science, Illinois Inst. of Technology, Chicago.
[5] R. T. Oehrle, E. Bach, and D. Wheeler, Editors. **Categorial Grammars and Natural Language Structures.** D. Reidel Publishing Company, 1988.
[6] E. Black, J. Lafferty, and S. Roukos. **Development, evaluation, and results for a broad-coverage probabilistic grammar of English-language computer manuals.** To appear in Proceedings of the ACL, 1992.